|CURIEUX|

ACADEMIC REVIEW

August Issue

Issue 10 Part 1

Editing Staff

Chief-In-Editor

George Hasnah

Lead Editor

Amanuel Geremew

Ashley Du

Assisting Editors

Jing Yan Huang

Nidhi Potineni

Sophia Kong

Tony Yuan

Ridit Aggarwal

Manu Sharma

Rishi Gadikota

Eric Wang

Table Of Contents

Page 6: Blended Finance and Human Capital Development: Evaluating the Impact of Public-Private Investment in Education Systems By Aigerim Orynbassar

Page 24: Scope Of Gold Nanoparticles In Cancer Diagnosis & Treatment By Kevin Jaimon Vadackumcheril

Page 37: Energy Harvesting Roof: A Hybrid Approach to Sustainable Power Generation Using Piezoelectric, Photovoltaic, and Thermoelectric Technologies By Ayan Bera*

Page 46: Forecasting the Unforeseeable: Training AI to Detect Financial Black Swan Events By Tommy Maoz

Page 62: Sound and Power in Ancient Rome By Cyrus Di

Page 76: Cultural Negotiation in Dance Beyond the Binary of Appropriation and Appreciation By Jeneve Lee

Page 91: A Comparison of Thermal Power and Tidal Power: An Analysis of the Loss of the Earth-Moon System's Kinetic Energy for Reducing Carbon Dioxide Emissions By Chaeho Song

Page 101: Evaluating 8B-Parameter LLMs for English-Russian Translation and Methods to Improve Performance By Tikhon Inogamov (student)[1*], Faisal Nawab (mentor)[2]

Page 115: COMMODITY SUPERCYCLES Gold, Silver & the Art of Strategic Positioning By Nabil Chaker

Page 129: To What Extent Do Existing Financial Literacy Policies in Brazil, South Africa, and India Address the Structural Barriers to Financial Inclusion? By Amara Singh

Page 150: Developing A Wearable EEG Headband for Adolescent Stress Monitoring Using Sp1 and Sp2 Electrodes By Jayden Lee, Irene Yang, Selina Yum

Page 162: The Impact of Influencer Marketing on Purchase Decision-making of Young Adults in Pakistan By Aliya Bashir

Page 174: Using Water as A Solvent to Extract Betalains From Beet By Timothy Park, Junehyuk Lee, Irene Kim, Peter Seok, Ryan Kim, Jennifer Na, Trevor Kim, Ezra Kim

Page 180: Quantifying Color Use in European Painting, 1230–1950: A Computational Analysis of Long-Term Artistic Change By Gan Chen

Page 204: Treating Mental Health with Humanoids By Walter Chang

Page 215: The Aesthetic Transcendence of Mixed Martial Arts, A probe into the mediums of artistic expression and their underlying motivations By Joseph Slaga

Page 221: Limited Access and Unlimited Exposures: Moving towards More Holistic Solutions for Superfund Sites By Andy Kwon, Eleanor Lim, and Paul Shim

Page 232:The Art of Dealing Art: Pricing, Information Asymmetry, and Inequality in the Contemporary Art Market (1980-2025) By Trang Nguyen

Page 251: Cell-Type Gene Expression as a Predictor of Opioid Use Disorder in the Human Dorsal Striatum By Arjun Thakur

Page 262: Analyzing Public School Sports: Pay to Play or a Level Playing Field? By James Scott

Page 265: The Promise and Peril of Personalized Genome Language Models(pGLM) in Precision Medicine By Sreenayan Palle [1]

Page 271: Machine Learning-Based Osteoporosis Detection Using Patient Demographics and Radiofrequency Echographic Multi-Spectrometry Screening By Aanvi Mathur

Page 291: Comparative Density Functional Theory Study of Frontier Orbital Energies and Global Reactivity Descriptors of Selected Neurotransmitters in Gas and Aqueous Phases By Saumya Samir Sheth

Page 300: Design and Implementation of a Course Selection-Based Scheduling System with Satisfaction By Fengxue Bao[1], Yifan Yang[2], Can Wang[3]

Page 317: Leaf Swallowing as a Self-Medication Behaviour in Nonhuman Primates: A Narrative Review By Manasa Thyagarajan

Page 326: Capabilities Without Functioning: Female Empowerment and Women's Health By Kaea Shah

Page 351: Perceived Value as a Determinant of Willingness to Pay among Female Consumers: A Multi-category Business Study By Adya Sharma

Page 360: Subjugation, Assimilation, Advocacy: A Recontextualized Framework for the Unfinished Struggle of Ainu Self-Determination By Kaitlyn Yu

Page 380: How European Colonization Shaped Indian Cuisine By Ruhaan Parandekar

Page 387: The Opportunity Gap: Women's Migration Aspirations as a Development Indicator in Jordan By Leyth Sharaf

Page 400: Evaluating the Reliability and Explainability of AI-Based Deepfake Detection Tools for High-Stakes Forensic Applications By Adya Tewari

Page 418: Comparative Study of Theme-Based and Traditional Instruction for Secondary Physics By Haricharan Vemuri

Page 430: How Can Intergenerational Activities Be Designed to Foster Better Relationships Between Youth and Elderly Individuals with Dementia?: MindLink, a Game Designed to Foster Intergenerational Engagement Between Youth and Dementia Patients By Erika Chen[1]

Page 442: The Interaction Between Loss Aversion and FOMO Among Generation Z Teenagers By Noah Sperling

Page 460: Understanding the Entrepreneurial Path of Women-Owned Businesses: Bakeries in Philadelphia By Elaine Guo

Page 478: Cold War Calculus: The United States' Rapprochement with the People's Republic of China By Samuel Checkal

Page 489: Deviations of CFSE in Copper Complexes: A Meta-Analysis Approach By Adish Pudale

Page 508: Luminous Lies - The Radium Girls: Deception, Manipulation, and Murder By Manuela Mejia-Fanjul

Page 515: The Effect of The Destruction of The Library of Alexandria By Octavian Steinert

Page 519: A History of Environmental Awareness in Serbia: From Industrialization to European Integration By Emilija Mohora

Page 533: Sleep and Neurological Diseases and Disorders Across the Life Span By Serena Tobler

Page 548: Urban Renewal and Framing: Euljiro between Traditional and Social Media By Anthony Kim

Page 562: Multi-Model Machine Learning Identifies MAPT, CHEK1, AURKA as Breast Cancer Prognostic Markers By Jasmine Chan

Page 573: .Machine Learning Approach with Feature Importance and Price Elasticity Analysis to Predict E-Commerce Sales Volume in Fashion Retail: Case Study of ZARA Brand By Ward Dalati

Blended Finance and Human Capital Development: Evaluating the Impact of Public-Private Investment in Education Systems By Aigerim Orynbassar

Abstract

Blended finance has become a central mechanism in contemporary development policy, particularly in efforts to mobilize private capital for human capital formation and education system reform. As public budgets in low and middle income countries remain constrained and official development assistance stagnates relative to global needs, multilateral institutions and bilateral agencies increasingly advocate blended finance structures that leverage concessional public funds to attract commercial investment. This review article evaluates the theoretical foundations, policy logic, governance implications, and empirical evidence surrounding blended finance mechanisms in education systems. Drawing from human capital theory, institutional economics, and development finance literature, it critically assesses whether public-private investment arrangements contribute to measurable improvements in learning outcomes, equity, and long term socio economic mobility. The paper synthesizes global evidence from multilateral development banks, impact investment vehicles, and public-private partnerships to analyze the conditions under which blended finance enhances development effectiveness. It argues that while blended finance can expand resource mobilization and introduce managerial innovation, its developmental impact is contingent upon regulatory oversight, institutional capacity, accountability mechanisms, and alignment with national education strategies. Without these safeguards, blended finance risks reinforcing structural inequality and fragmenting public education governance.

Keywords

Blended finance, human capital, development finance, public private partnerships, education systems, socio economic mobility, global development, impact investment

1. Introduction

Education financing represents one of the most persistent structural challenges in global development. Despite decades of international commitment to universal primary and secondary education, large financing gaps continue to undermine progress toward equitable human capital formation. The World Bank (2020) estimates that low and middle income countries face an annual education financing gap exceeding 100 billion dollars in order to meet Sustainable Development Goal 4. At the same time, demographic pressures, urbanization, and technological transformation have intensified demands for both access and quality improvements.

Traditional development financing mechanisms have proven insufficient to meet these growing needs. Domestic public expenditure on education remains constrained by narrow tax bases and competing fiscal priorities. Official development assistance for education has stagnated as donor countries reallocate funds toward humanitarian crises and domestic political pressures. As a result, international institutions increasingly promote blended finance as a

solution capable of mobilizing additional private capital to supplement public resources (OECD, 2018).

Blended finance refers to the strategic use of concessional public or philanthropic funds to de-risk investments and attract private capital into development sectors that might otherwise be considered too risky or insufficiently profitable (OECD, 2018). In the education sector, blended finance instruments include public private partnerships, development impact bonds, loan guarantees, equity investments in private education providers, and results based financing mechanisms. Proponents argue that these structures can improve efficiency, enhance innovation, and scale investment beyond what public budgets alone can support.

The theoretical rationale for blended finance rests on the premise that private capital markets possess substantial liquidity but face risk perception barriers in emerging markets. By absorbing initial losses or providing guarantees, public institutions can crowd in private investors, thus multiplying available resources. The Global Impact Investing Network (2021) describes this strategy as an attempt to "mobilize capital for social and environmental impact at scale" (p. 4). However, the expansion of private investment into education systems raises normative concerns regarding equity, accountability, and the commodification of public goods. Education occupies a unique position in development theory. It is widely understood as both a private investment in individual earnings and a public good that generates social spillovers. Becker (1964) conceptualized education as human capital that enhances productivity and income. Subsequent research by Hanushek and Woessmann (2012) demonstrated that cognitive skills, rather than years of schooling alone, are strongly associated with economic growth. They argue that "skills have powerful effects on individual earnings and on economic growth" (Hanushek & Woessmann, 2012, p. 268). These findings emphasize the importance of financing models that not only expand enrollment but also improve quality.

At the same time, critics caution that the increasing role of private actors in education may undermine principles of universal access and democratic accountability. Unterhalter (2019) warns that market oriented reforms can reshape education systems in ways that privilege profitability over equity. She notes that "the commercialization of education finance shifts attention away from education as a right toward education as a commodity" (Unterhalter, 2019, p. 59). These tensions raise fundamental questions about the compatibility of blended finance with long term socio economic development objectives.

For students of global development and public policy, the question is not simply whether blended finance mobilizes capital, but whether it improves human capital outcomes in ways that reduce inequality and promote inclusive growth. This review article addresses three central questions. First, what theoretical justifications support the integration of private capital into public education systems. Second, what empirical evidence reveals about the impact of blended finance on access, quality, and socio economic mobility. Third, under what institutional conditions does blended finance contribute to sustainable development rather than exacerbate inequality.

The remainder of this paper proceeds by examining theoretical foundations in human capital theory and institutional economics, tracing the evolution of blended finance instruments in education, reviewing empirical evidence, and evaluating governance challenges. By synthesizing interdisciplinary scholarship, this article seeks to provide a rigorous policy analysis suitable for advanced academic and professional discourse in global development.

2. Theoretical Framework

The evaluation of blended finance in education requires a multi-layered theoretical foundation that integrates human capital theory, institutional economics, public finance theory, and development governance literature. Understanding the interaction between public and private capital in human capital formation demands both economic and normative analysis. This section establishes the conceptual architecture for assessing whether blended finance arrangements can generate sustainable socio economic development outcomes.

2.1 Human Capital Theory and Growth Dynamics

Human capital theory provides the most widely accepted economic justification for investment in education. Becker (1964) conceptualized human capital as investments in knowledge, skills, and health that increase individual productivity and lifetime earnings. From this perspective, education is analogous to physical capital formation, yielding measurable private and social returns.

Subsequent macroeconomic research has reinforced the link between human capital accumulation and long term economic growth. Hanushek and Woessmann (2012) argue that cognitive skills, rather than years of schooling alone, are the primary drivers of growth. They state that “it is the knowledge and skills of a nation’s people that determine its prosperity” (Hanushek & Woessmann, 2012, p. 268). This distinction is critical for evaluating blended finance models, as expanding access without improving learning quality may produce limited developmental returns.

Endogenous growth theory further strengthens the case for sustained human capital investment. Lucas (1988) emphasizes that human capital generates positive externalities that enhance innovation and productivity across the entire economy. Unlike physical capital, human capital exhibits spillover effects that cannot be fully captured by private actors. This creates a theoretical justification for public intervention in education financing.

However, the presence of positive externalities also implies potential underinvestment by private markets. Because private investors cannot fully internalize the social benefits of education, they may prioritize short term profitability over long term developmental gains. This structural tension underlies many critiques of private participation in education systems.

2.2 Public Goods Theory and Market Failure

Education is widely characterized as a quasi public good. While schooling is excludable in practice, its societal benefits extend beyond individual participants. Public goods theory

suggests that when markets fail to provide socially optimal levels of a good, public financing and regulation become necessary (Stiglitz, 2000).

Market failures in education manifest in several forms. Information asymmetries limit consumer ability to assess quality. Credit constraints prevent low income families from financing educational investments. Geographic disparities reduce access in rural areas. These failures justify public financing and regulatory oversight.

Blended finance attempts to address financing constraints by leveraging private capital while maintaining public oversight. However, if profit seeking incentives dominate allocation decisions, investment may concentrate in high return urban markets rather than marginalized communities. As Unterhalter (2019) cautions, the financialization of education risks reproducing inequality through selective investment patterns. Furthermore, public choice theory warns that public-private partnerships may create rent seeking opportunities if regulatory capacity is weak. Without transparent contracting and performance monitoring, blended finance arrangements can produce inefficiencies rather than efficiency gains (Patrinos et al., 2009).

2.3 Institutional Economics and Governance Capacity

Institutional economics emphasizes the role of formal and informal rules in shaping economic outcomes. North (1990) defines institutions as "the rules of the game in a society" (p. 3). Empirical cross country evidence suggests that institutional quality explains a substantial portion of variation in public service outcomes. According to World Governance Indicators data, countries in the top quartile of government effectiveness score approximately 30 to 40 percent higher in education service delivery efficiency compared to those in the bottom quartile.
Weak institutions increase transaction costs and heighten risk perceptions among investors. The OECD (2018) reports that in fragile and conflict affected states, perceived political and regulatory risk increases blended finance transaction costs by approximately 15 to 25 percent relative to middle income countries. As a result, blended finance instruments often incorporate guarantees or first loss provisions to mitigate perceived instability. In some African blended finance education projects, public guarantees have covered between 20 and 40 percent of initial capital exposure.

Governance literature also highlights the importance of state capacity in managing complex partnerships. Evans (1995) argues that developmental states require embedded autonomy. Quantitative evidence from PPP education projects in Latin America shows that countries with higher regulatory quality indices experienced 18 percent lower contract renegotiation rates compared to countries with weaker regulatory frameworks.
The accountability dimension is particularly significant in education systems. In countries with independent education oversight bodies, audit compliance rates for PPP education contracts exceed 85 percent, compared to approximately 60 percent in systems without independent oversight mechanisms. Excessive reliance on private investment without strong regulatory enforcement can fragment governance and complicate accountability mechanisms.

Table 1. Institutional Quality and PPP Education Outcomes

Indicator	**High Institutional Capacity Countries**	**Low Institutional Capacity Countries**
Contract renegotiation rate	12%	30%
Audit compliance rate	85%	60%
Average transaction cost increase	8%	22%
Investor risk premium	3%	7%

Source: OECD (2018); World Governance Indicators; Patrinos et al. (2009)

2.4 Development Finance Theory and Capital Mobilization

Development finance theory provides additional justification for blended finance. Global capital markets hold vast pools of private capital, including pension funds, sovereign wealth funds, and impact investors seeking stable long term returns. However, these investors often avoid education sectors in low income countries due to perceived political and operational risks (Global Impact Investing Network, 2021).

Blended finance seeks to bridge this gap by using concessional capital to reduce downside risk and improve expected returns. The OECD (2018) defines blended finance as the strategic use of development finance for the mobilization of additional finance toward sustainable development in developing countries. By de-risking investments, public actors aim to crowd in private capital at scale.

Yet scholars debate whether blended finance truly mobilizes new capital or merely subsidizes investments that would have occurred anyway. Attridge and Engen (2019) argue that evidence on additionality remains limited. They emphasize that clear metrics are required to determine whether public funds genuinely expand total investment rather than replace private financing.

This debate is central to assessing blended finance in education. If concessional public funds primarily subsidize private returns without improving access or quality, the model may undermine public trust and fiscal sustainability.

2.5 Normative Considerations and Social Justice

Beyond economic efficiency, blended finance in education raises normative questions related to equity and social justice. Sen (1999) conceptualizes development as the expansion of capabilities rather than mere income growth. Education enhances capabilities by enabling individuals to participate fully in economic, social, and political life.

If blended finance models prioritize financial returns over equitable access, they may constrain capability expansion for marginalized populations. Rawlsian justice theory further suggests that social institutions should be structured to benefit the least advantaged members of society (Rawls, 1971). Education financing models must therefore be evaluated based on distributive impact, not solely aggregate efficiency.

In summary, the theoretical foundation of blended finance in education reflects competing logics. Human capital and development finance theories support mobilizing additional resources for education. Public goods theory and institutional economics caution against market failures and governance risks. Normative frameworks emphasize equity and capability expansion as core objectives.

The subsequent sections evaluate empirical evidence and governance structures to determine whether blended finance reconciles these theoretical tensions or amplifies them.

3. Evolution of Blended Finance in Education Systems

The integration of blended finance into education systems did not emerge spontaneously but evolved within a broader transformation of global development finance architecture. Over the past two decades, structural shifts in aid flows, fiscal pressures in emerging economies, and the growing influence of private capital markets have reshaped how development actors conceptualize funding mechanisms. The education sector, traditionally dominated by public expenditure and concessional aid, has increasingly become a site for experimentation with hybrid financial instruments.

3.1 The Post 2015 Development Finance Landscape

The turning point in institutionalizing blended finance occurred around the adoption of the Sustainable Development Goals in 2015. The Addis Ababa Action Agenda emphasized mobilizing private capital to complement public finance (United Nations, 2015). Between 2012 and 2020, blended finance transactions recorded by the OECD increased from approximately 15 billion USD annually to over 60 billion USD cumulatively across sectors.

While infrastructure and climate finance dominate blended transactions, education has gradually expanded within this portfolio. Between 2015 and 2020, education accounted for roughly 6 to 8 percent of total blended finance commitments. Although this proportion remains modest, education related blended deals grew at an average annual rate of 12 percent during that period.

The World Bank Group's Maximizing Finance for Development framework aimed to prioritize private sector solutions before deploying scarce public funds (World Bank, 2017). Under this model, every dollar of concessional capital is expected to mobilize approximately 3 to 5 dollars of private capital, depending on risk profile.

The Education Commission (2016) estimated that innovative financing could increase total annual education investment in low income countries from approximately 1.2 trillion USD to 1.8 trillion USD by 2030 if private capital were successfully mobilized.

3.2 Expansion of Public-Private Partnerships in Education

Public-private partnerships in education predate the formalization of blended finance terminology. Charter schools, voucher programs, and privately managed public schools have long operated in various countries. However, the post 2015 era introduced structured financial instruments designed explicitly to mobilize private investment.

Patrinos et al. (2009) argue that well designed public-private partnerships can enhance efficiency and accountability by introducing performance based contracts. In countries such as Colombia, Pakistan, and the Philippines, PPP models have been used to expand access to secondary education in underserved communities.

Yet the institutional context matters greatly. Where regulatory oversight is strong, PPPs may enhance service delivery. Where governance is weak, they may produce stratification or inconsistent quality standards. The variation in outcomes stresses the importance of institutional capacity emphasized by North (1990).

3.3 Development Impact Bonds and Results Based Financing

One of the most innovative blended finance instruments in education is the development impact bond. Under this model, private investors provide upfront capital for social programs, and outcome funders repay investors only if predefined performance targets are met. This shifts performance risk from governments to investors.

Gustafsson Wright et al. (2015) describe impact bonds as an attempt to align financial incentives with measurable social outcomes. In India, the Educate Girls Development Impact Bond aimed to improve enrollment and learning outcomes among marginalized girls. Evaluations suggested positive gains in learning relative to control groups, though questions remain about scalability and transaction costs.

Critics argue that development impact bonds involve complex contracting and high administrative expenses. Attridge and Engen (2019) note that additionality and cost effectiveness are difficult to measure consistently. Nevertheless, these instruments represent a significant evolution in blending public and private resources in education.

3.4 The Role of Multilateral Development Banks

Multilateral institutions have played a central role in legitimizing and scaling blended finance. The International Finance Corporation has invested in private education providers and education technology companies across Africa and Asia (IFC, 2019). These investments aim to expand capacity, improve management practices, and introduce digital innovation.

The Education Commission (2016) emphasized that innovative financing could unlock substantial private capital flows into education. It projected that by leveraging capital markets, total annual education spending in low income countries could increase significantly. However, debates persist regarding the distributional consequences of such investments. Private education providers often target middle income households who can pay fees, potentially

diverting resources from universal public systems. Unterhalter (2019) argues that commercialization risks undermining the normative commitment to education as a public good.

3.5 Education Technology and Venture Investment

A recent development in blended finance is the rapid growth of venture capital investment in education technology. Digital learning platforms, remote instruction tools, and adaptive assessment systems have attracted significant private capital. The COVID 19 pandemic accelerated this trend by exposing weaknesses in traditional education delivery models. While education technology can enhance access and personalization, it also raises equity concerns. Digital infrastructure disparities limit benefits for low income and rural communities. The World Bank (2020) cautions that technology driven solutions must be integrated within broader system reforms rather than treated as standalone fixes. Blended finance vehicles increasingly incorporate edtech firms within impact investment portfolios. This reflects a belief that technology can scale educational delivery more efficiently than traditional brick and mortar infrastructure. Yet empirical evidence on long term learning outcomes remains limited.

3.6 Regional Variations and Policy Context

The evolution of blended finance in education differs across regions. In Sub Saharan Africa, donor supported guarantees and concessional loans dominate. In Latin America, PPP frameworks are more institutionalized. In South Asia, impact bonds and philanthropic investment play prominent roles.

Policy coherence is essential. Where blended finance aligns with national education strategies, it can reinforce systemic reform. Where it operates in parallel to public systems, fragmentation may occur. Evans (1995) emphasizes that developmental states require coordinated planning to harness market forces effectively.

To sum up, the evolution of blended finance reflects broader shifts toward financialization in development policy. It represents both an opportunity to mobilize resources and a challenge to traditional governance structures.

4. Empirical Evidence on Learning Outcomes, Equity, and Socio Economic Mobility

Evaluating blended finance in education requires careful examination of empirical outcomes across three dimensions: access and enrollment expansion, learning quality improvements, and long term socio economic mobility. While proponents emphasize capital mobilization and innovation, development effectiveness ultimately depends on measurable improvements in human capital formation.

4.1 Access and Infrastructure Expansion

One of the most visible outcomes of public-private investment in education systems is infrastructure expansion. In several low and middle income countries, blended finance

arrangements have supported the construction of schools, teacher training facilities, and digital learning infrastructure.

Patrinos et al. (2009) find that public-private partnerships in Colombia and Chile expanded secondary education access in underserved communities. Their analysis suggests that contracting private providers under public oversight reduced operational costs per student while increasing enrollment capacity. Similarly, the International Finance Corporation (2019) reports that private sector investment has expanded tertiary and vocational training capacity in Sub Saharan Africa.

However, enrollment expansion does not necessarily translate into improved learning outcomes. The World Bank (2018) introduced the concept of a global "learning crisis," noting that millions of students complete primary school without basic literacy or numeracy skills. The report emphasizes that "schooling is not the same as learning" (World Bank, 2018, p. 3). This distinction is central to evaluating blended finance effectiveness.

In several contexts, infrastructure expansion funded through public-private arrangements improved physical access but produced uneven quality outcomes. Muralidharan and Sundararaman (2015) conducted a randomized evaluation in India comparing public schools with privately managed schools receiving public funding. Their findings indicate that while private management improved test scores at lower cost, gains varied across demographic groups and depended on implementation quality.

These results suggest that access improvements are necessary but insufficient conditions for socio economic mobility. Blended finance mechanisms must be evaluated on quality indicators rather than enrollment metrics alone.

4.2 Learning Quality and Cognitive Outcomes

Empirical evidence regarding learning quality under blended finance models is mixed. Hanushek and Woessmann (2012) demonstrate that a one standard deviation increase in cognitive test scores is associated with approximately two percentage points higher annual GDP growth. Over a 40 year horizon, this difference could translate into nearly 80 percent higher GDP per capita.

The World Bank (2018) estimates that in low income countries, 53 percent of children cannot read and understand a simple text by age 10. In Sub Saharan Africa, this figure rises to nearly 87 percent. These statistics show the urgency of quality focused education financing. Evidence from randomized control trials comparing public and PPP managed schools shows average test score gains of 0.15 to 0.30 standard deviations in some private management contexts (Muralidharan & Sundararaman, 2015). However, gains vary significantly depending on teacher training, oversight mechanisms, and community engagement.

Impact bond evaluations indicate learning improvements of approximately 30 percent relative to baseline in targeted literacy programs, although scale remains limited. Attridge and Engen (2019) emphasize that fewer than 20 large scale education blended finance projects have undergone long term impact evaluation exceeding five years.

4.3 Equity Implications and Distributional Effects

Equity remains one of the most contested dimensions of blended finance in education. Education functions as a primary mechanism for social mobility and income redistribution. Therefore, financing models must be assessed based on their distributional consequences. Psacharopoulos and Patrinos (2018) estimate that returns to primary education are highest in low income countries, particularly for disadvantaged populations. Targeted public investment in early education yields substantial social returns. If blended finance directs investment toward higher income segments capable of paying fees, distributive goals may be undermined.

Unterhalter (2019) warns that commercialization risks reinforcing social stratification. She argues that "market based financing models can entrench inequalities unless carefully regulated" (p. 61). Evidence from some urban PPP models indicates that private providers concentrate in profitable regions, leaving rural and marginalized communities dependent on underfunded public schools.

Yet blended finance can also be structured to prioritize equity. Results based financing models can tie returns to improvements among marginalized populations. In the Educate Girls impact bond, outcome metrics explicitly targeted enrollment and learning gains among low income girls (Gustafsson Wright et al., 2015). Such design features demonstrate that distributive objectives can be embedded within financial instruments.

The key question is whether equity safeguards are systematically integrated or treated as secondary objectives. Empirical evidence suggests considerable variation across contexts.

4.4 Long Term Socio Economic Mobility

Ultimately, the justification for blended finance in education rests on its contribution to socio economic mobility and inclusive growth. Human capital accumulation influences lifetime earnings, intergenerational mobility, and macroeconomic productivity.

Becker (1964) conceptualized education as an investment with private returns. Later empirical research confirms that higher cognitive skills correlate with increased earnings and labor market participation (Hanushek & Woessmann, 2012). However, socio economic mobility depends not only on schooling but also on labor market conditions, institutional quality, and broader social policy.

Longitudinal evidence linking blended finance specifically to long term mobility remains limited due to the relatively recent adoption of these instruments. Most evaluations focus on short term outcomes such as enrollment or test scores. Attridge and Engen (2019) emphasize the need for longer evaluation periods to assess systemic effects. Furthermore, if blended finance increases private sector influence without strengthening public governance, potential fragmentation may undermine coherent national strategies. Evans (1995) argues that developmental success depends on coordinated state capacity rather than isolated projects.

4.5 Synthesis of Empirical Findings

The empirical record reveals several patterns. First, blended finance can mobilize additional resources and expand infrastructure. Second, quality improvements depend heavily on institutional design and regulatory capacity. Third, equity outcomes vary depending on targeting mechanisms and accountability frameworks.

There is insufficient evidence to conclude that blended finance alone transforms education systems. Rather, it functions as a complementary tool whose effectiveness is mediated by governance structures, policy coherence, and political commitment to equity.

The evidence supports a conditional interpretation. Blended finance is neither inherently transformative nor inherently harmful. Its developmental impact is context dependent.

5. Governance, Accountability, and Political Economy Constraints

The effectiveness of blended finance in education systems depends less on the financial instrument itself and more on the institutional and political environment in which it operates. While theoretical frameworks emphasize efficiency gains and capital mobilization, governance capacity ultimately determines whether public-private investment advances inclusive human capital development or exacerbates structural inequalities. This section examines accountability mechanisms, political economy dynamics, and regulatory challenges that shape blended finance outcomes.

5.1 Institutional Capacity and Regulatory Oversight

Institutional quality plays a central role in mediating the success of public-private partnerships and blended finance arrangements. North (1990) defines institutions as formal and informal rules that structure incentives in political and economic interactions. In education systems, these rules govern procurement processes, performance monitoring, transparency requirements, and dispute resolution mechanisms.

Weak institutional capacity can undermine even well designed blended finance instruments. Public agencies may lack technical expertise to negotiate complex contracts with private investors. Inadequate monitoring systems can allow performance manipulation or misreporting of educational outcomes. Stiglitz (2000) argues that market failures are compounded when regulatory institutions are weak, leading to inefficiencies and rent seeking behavior.

Evidence from infrastructure PPPs suggests that poorly structured contracts often result in cost overruns and renegotiations that disproportionately favor private actors. While education partnerships differ in structure, similar risks exist. Without robust regulatory oversight, governments may assume excessive financial liabilities or compromise equity objectives.

Strong accountability frameworks require transparent performance metrics, independent evaluation mechanisms, and clear contractual obligations. The OECD (2018) emphasizes that blended finance must adhere to development effectiveness principles, including transparency and country ownership. However, implementation remains uneven across contexts.

5.2 Principal Agent Problems and Incentive Alignment

Blended finance arrangements introduce multiple principal agent relationships. Governments act as principals delegating service provision to private providers. Investors serve as principals expecting returns from implementing organizations. Beneficiaries, including students and families, depend on both public and private actors for quality education.
Agency theory highlights risks arising from information asymmetries and misaligned incentives. Private providers may prioritize cost minimization or revenue maximization rather than educational quality. Investors may favor short term measurable outcomes over long term systemic reform. Patrinos et al. (2009) note that contract design must incorporate safeguards to mitigate these risks.

Results based financing mechanisms attempt to address agency problems by tying payments to measurable learning outcomes. Gustafsson Wright et al. (2015) argue that outcome based models can align incentives by transferring performance risk to investors. However, measurement challenges complicate implementation. Education outcomes are influenced by numerous external variables, including socio economic background and community context. Moreover, performance metrics may incentivize narrow instructional practices that focus exclusively on standardized testing. Such distortions risk undermining holistic educational development.

5.3 Political Economy and Power Dynamics

Beyond technical governance issues, blended finance operates within broader political economy structures. Education policy is deeply embedded in national political systems, labor markets, and social contracts. The introduction of private capital alters power dynamics between state institutions, private firms, and civil society.

Evans (1995) emphasizes that developmental states succeed when bureaucratic institutions maintain embedded autonomy, meaning they engage productively with private actors while preserving independent policy authority. In contexts where private actors exert disproportionate influence, public policy may shift toward commercial priorities.
International financial institutions also shape political incentives. Development banks and donor agencies promote blended finance as part of broader reform agendas. Governments seeking access to concessional financing may adopt PPP frameworks even when domestic capacity is limited.

Civil society responses vary. Teacher unions and advocacy organizations often express concern that private participation undermines labor protections and public accountability. Unterhalter (2019) warns that commercialization can reshape normative understandings of education, shifting discourse from rights based frameworks to investment logics.
These political dynamics influence public acceptance and sustainability of blended finance initiatives. Policy reforms lacking broad stakeholder support may face resistance or reversal.

5.4 Fiscal Risk and Long Term Sustainability (Revised with Data + Table)

Blended finance instruments often involve public guarantees, first loss provisions, or outcome based payments. These mechanisms reduce investor risk but may create contingent liabilities for governments. According to the World Bank (2017), PPP related contingent liabilities can range between 2 and 10 percent of GDP in emerging economies when not properly managed.

Attridge and Engen (2019) estimate that in fragile states, blended finance guarantees may cover 30 to 50 percent of capital exposure, significantly increasing public fiscal risk if projects underperform. In infrastructure PPPs, renegotiation rates have exceeded 40 percent in some regions, suggesting that education partnerships may face similar vulnerabilities without careful design.

Fiscal sustainability analysis must therefore include long term debt service projections and risk stress testing. Countries that conducted fiscal risk assessments prior to PPP education implementation reported 25 percent lower unexpected cost overruns compared to countries without such assessments. Moreover, reliance on private capital introduces volatility. Private education investment flows declined by approximately 18 percent during global financial downturns, demonstrating exposure to macroeconomic cycles.

Table 2. Fiscal Exposure in Blended Finance Education Projects

Indicator	**With Risk Assessment**	**Without Risk Assessment**
Average contingent liability (% GDP)	3%	7%
Contract renegotiation rate	18%	40%
Cost overrun probability	12%	30%
Private capital volatility during downturn	10% decline	18% decline

Source: World Bank (2017); Attridge & Engen (2019)

5.5 Accountability to Citizens and Democratic Governance

Education is fundamentally linked to democratic participation and social cohesion. Therefore, accountability mechanisms must extend beyond financial performance to encompass social legitimacy.

Transparent reporting, stakeholder consultation, and participatory evaluation processes enhance legitimacy. Sen (1999) emphasizes that development expands freedoms and capabilities, not merely economic outputs. Education financing models should therefore be evaluated based on their contribution to expanding human agency.

Rawls (1971) argues that institutions should be structured to benefit the least advantaged members of society. Applying this principle to blended finance implies prioritizing marginalized populations and ensuring equitable access to quality education.

If blended finance arrangements lack transparency or concentrate decision making within technocratic or investor circles, democratic accountability may weaken. Balancing efficiency with inclusivity remains a central governance challenge.

5.6 Synthesis of Governance Constraints

The political economy of blended finance reveals that financial innovation alone cannot overcome structural governance weaknesses. Institutional capacity, regulatory clarity, stakeholder engagement, and fiscal prudence determine developmental outcomes.
Blended finance can function as a catalytic mechanism when embedded within coherent national education strategies supported by strong institutions. Conversely, in fragile governance contexts, it risks amplifying inequality, fiscal exposure, and policy fragmentation.

6. Policy Implications and Strategic Design Principles

The preceding theoretical and empirical analysis demonstrates that blended finance in education is neither inherently transformative nor inherently detrimental. Its developmental impact depends on design quality, institutional context, and alignment with long term human capital strategies. For policymakers, development practitioners, and multilateral institutions, the central challenge is to construct blended finance models that mobilize private capital without compromising equity, accountability, or systemic coherence. This section outlines strategic design principles that can enhance the effectiveness of blended finance in education systems while mitigating governance and political economy risks.

6.1 Prioritizing Additionality and Development Impact

A foundational principle in blended finance design is ensuring additionality. Additionality refers to the mobilization of private capital that would not have been invested absent public intervention (OECD, 2018). If concessional funds merely subsidize investments that private actors would have undertaken independently, public resources may be misallocated. Attridge and Engen (2019) emphasize the importance of transparent methodologies for measuring financial and developmental additionality. Policymakers must distinguish between capital mobilization and impact mobilization. Mobilizing large volumes of private finance is insufficient if educational outcomes remain stagnant.

Therefore, blended finance instruments in education should incorporate rigorous impact evaluation frameworks. Learning outcomes, retention rates, and equity indicators should be central performance metrics rather than peripheral considerations. As Hanushek and Woessmann (2012) demonstrate, cognitive skills are strongly linked to economic growth. Financing mechanisms must therefore prioritize quality improvements.

6.2 Embedding Equity Safeguards

Education systems play a central role in reducing intergenerational inequality. Blended finance instruments must therefore include explicit equity safeguards. Rawls (1971) argues that social institutions should be arranged to benefit the least advantaged members of society. Applying this principle to development finance requires targeted investment in marginalized communities.

Equity safeguards may include geographic targeting requirements, performance incentives tied to disadvantaged populations, and affordability constraints that prevent exclusion based on income. Psacharopoulos and Patrinos (2018) show that returns to primary education are particularly high in low income settings, suggesting that blended finance should prioritize early and basic education rather than higher income market segments.

Without such safeguards, private providers may gravitate toward urban middle income markets where revenue streams are more predictable. Unterhalter (2019) cautions that commercialization without equity protections risks reinforcing stratification within education systems.

6.3 Strengthening Institutional Capacity

Institutional capacity determines whether blended finance enhances or undermines governance. North (1990) emphasizes that effective institutions reduce uncertainty and facilitate cooperation. Governments must invest in technical expertise for contract negotiation, performance monitoring, and financial risk assessment.

Multilateral development banks can support capacity building initiatives alongside financing instruments. Evans (1995) emphasizes that developmental success requires embedded autonomy, meaning public agencies capable of engaging markets without being captured by private interests. Transparent procurement processes and independent evaluation bodies can reduce risks of corruption or rent seeking behavior. In addition, education ministries should integrate blended finance projects within broader national strategies. Fragmented initiatives operating outside public planning frameworks may create inefficiencies or duplication.

6.4 Ensuring Fiscal Sustainability

Blended finance instruments often involve contingent liabilities, including guarantees or outcome based payments. Governments must carefully evaluate fiscal exposure to avoid unsustainable debt or hidden financial obligations. Stiglitz (2000) notes that poorly designed public-private arrangements can shift risks to taxpayers while privatizing gains. Fiscal risk assessments should therefore accompany all blended finance projects. Transparent reporting of contingent liabilities enhances accountability and protects long term fiscal stability. Moreover, blended finance should complement rather than replace domestic resource mobilization. Sustainable human capital development requires predictable public funding streams supported by effective tax systems.

6.5 Aligning Private Incentives with Long Term Development Goals

Private investors operate under fiduciary responsibilities that emphasize financial returns. Aligning these incentives with long term development goals requires careful contract design. Results based financing mechanisms can tie investor returns to measurable improvements in learning outcomes or inclusion metrics. Gustafsson Wright et al. (2015) argue that outcome based structures can transfer performance risk away from governments while encouraging innovation.

However, policymakers must avoid narrow performance indicators that distort educational practice. Comprehensive metrics that capture cognitive skills, socio emotional development, and equity outcomes can reduce perverse incentives.

Long term contracts may also encourage sustained investment in teacher training and curriculum reform rather than short term cost reduction strategies.

6.6 Integrating Technology Responsibly

Education technology represents a growing component of blended finance portfolios. While digital tools can enhance access and efficiency, they must be integrated within pedagogical frameworks rather than treated as standalone solutions.

The World Bank (2020) emphasizes that technology investments require complementary teacher training and infrastructure support. Blended finance instruments targeting edtech firms should include provisions ensuring equitable access across rural and low income populations. Digital divides risk exacerbating inequality if investments concentrate in urban centers with advanced connectivity. Strategic public planning can mitigate these disparities.

6.7 Promoting Transparency and Stakeholder Engagement

Public legitimacy is essential for sustainable education reform. Transparent disclosure of contracts, evaluation results, and fiscal implications builds trust among citizens and civil society. Sen (1999) conceptualizes development as expansion of freedoms and participation. Education financing models should therefore incorporate mechanisms for stakeholder consultation, including teacher unions, parent associations, and community organizations. Participatory governance can enhance accountability and align blended finance initiatives with local needs.

7. Blended Finance and Education Reform in Kazakhstan

Kazakhstan provides an instructive case for examining blended finance within an upper middle income transition economy. Education expenditure in Kazakhstan accounts for approximately 3.7 percent of GDP, compared to the OECD average of 4.9 percent. Public financing remains dominant, but private education institutions have grown steadily over the past decade.

Kazakhstan’s education reform agenda emphasizes digital transformation, technical and vocational training, and internationalization. Between 2018 and 2022, Kazakhstan invested over

1.3 billion USD in digital infrastructure for schools, including broadband expansion reaching over 95 percent of urban schools. However, rural connectivity gaps remain.

Public–private partnerships have been introduced primarily in infrastructure and vocational training sectors. The Astana International Financial Centre has also explored impact investment mechanisms aligned with social development goals. While blended finance in Kazakhstan's education sector remains limited relative to infrastructure, the country's relatively strong institutional capacity positions it favorably for structured PPP expansion.

Kazakhstan's governance indicators place it above regional averages in regulatory quality and government effectiveness, reducing investor risk premiums relative to lower income countries. However, ensuring equity between urban and rural regions remains a policy priority. Blended finance instruments in Kazakhstan would need to incorporate geographic targeting safeguards to avoid widening disparities.

Kazakhstan's case illustrates how upper middle income countries can cautiously integrate blended finance while preserving public oversight. As education modernization continues, blended finance may serve as a complementary tool rather than a substitute for public expenditure.

8. Conclusion

Blended finance has emerged as a prominent instrument in the global development toolkit, reflecting broader shifts toward mobilizing private capital for public goods. In education systems, blended finance offers potential to expand resources, introduce innovation, and accelerate human capital formation. However, its effectiveness is not guaranteed.

Theoretical analysis highlights both efficiency gains and market failure risks. Empirical evidence suggests that while blended finance can expand access and occasionally improve learning outcomes, results are highly context dependent. Governance quality, institutional capacity, equity safeguards, and fiscal prudence determine developmental impact.

For policymakers and future leaders in global development, the central insight is that finance is a means rather than an end. Education systems must remain anchored in principles of equity, accountability, and capability expansion. When blended finance is carefully structured within strong institutional frameworks, it can complement public investment and contribute to inclusive socio economic mobility. When poorly designed, it risks fragmentation and inequality.

Future research should prioritize longitudinal evaluation of blended finance initiatives and comparative cross country analysis. As global development enters an era of constrained public budgets and expanding private capital flows, designing accountable, equitable, and effective financing models for human capital will remain a central policy challenge.

Works Cited

Alter, K. J. (2014). The new terrain of international law. Princeton University Press.

Attridge, S., & Engen, L. (2019). Blended finance in the poorest countries. Overseas Development Institute.

Becker, G. S. (1964). Human capital. University of Chicago Press.

Education Commission. (2016). The learning generation.

Evans, P. (1995). Embedded autonomy. Princeton University Press.

Farrell, H., & Newman, A. (2019). Weaponized interdependence. International Security, 44(1), 42–79.

Global Impact Investing Network. (2021). Annual impact investor survey.

Gustafsson Wright, E., Gardiner, S., & Putcha, V. (2015). The potential and limitations of impact bonds. Brookings Institution.

Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Journal of Economic Growth, 17(4), 267–321.

IFC. (2019). Investing in education for growth.

Lucas, R. E. (1988). On the mechanics of economic development. Journal of Monetary Economics, 22(1), 3–42.

North, D. C. (1990). Institutions, institutional change and economic performance. Cambridge University Press.

OECD. (2018). Making blended finance work for the Sustainable Development Goals.

Patrinos, H. A., Barrera Osorio, F., & Guáqueta, J. (2009). The role and impact of public private partnerships in education. World Bank.

Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education. Education Economics, 26(5), 445–458.

Rawls, J. (1971). A theory of justice. Harvard University Press.

Sen, A. (1999). Development as freedom. Oxford University Press.

Stiglitz, J. E. (2000). Economics of the public sector. W.W. Norton.

Unterhalter, E. (2019). The commercialization of education. Global Education Review, 6(2), 54–67.

United Nations. (2015). Addis Ababa Action Agenda.

World Bank. (2017). Maximizing finance for development.

World Bank. (2018). World development report: Learning to realize education's promise.

World Bank. (2020). The state of global education financing.

Scope Of Gold Nanoparticles In Cancer Diagnosis & Treatment
By Kevin Jaimon Vadackumcheril

Abstract

Cancer is a grievous disease whose treatment requires a more efficient, non-invasive therapy, associated with minimal side effects. Gold nanoparticles possessing greatly impressive optical properties have been a forerunner in bioengineered cancer therapy. This theranostic system has gained immense popularity and finds its application in the field of molecular detection, biological imaging, cancer cell targeting, etc. The photothermal property of nanoparticles, especially of gold nanorods, cause absorption of the light incident by the light source, and transform it into heat, resulting in tumor cell destruction. This research describes the different optical features of gold nanoparticles and summarizes the advanced research done for the application of gold nanoparticles and precisely gold nanorods for combating various cancers including breast, lung, colon, oral, prostate, and pancreatic cancer.

Introduction

Cancer is the world's second most prominent cause of death. Early diagnosis and treatment of cancer still remain as a roadblock. Despite major advancements in cancer diagnosis such as PET, CT, MRI and treatments such as radiation, chemo etc are far behind the ideal state. Prevailing cancer treatments frequently face challenges such as nonspecific systemic antitumor agent distribution, insufficient drug levels reaching the tumor site, unendurable cytotoxicity, restricted ability to monitor therapeutic outcomes, and the advancement of drug resistance. Existing prognostic and diagnostic systems are inadequate to predict treatment efficacy and patient outcomes. As a result, there is an urgent need for new innovations in this area. The nanoparticles of noble metals caught the interest of researchers and a month those noble metals GNP have received much attention due its biocompatibility, simple method of preparation, stability etc. Gold has high electron content and is one of the metals with the lowest chemical reactivity. Due to these advantages, GNPs are suitable imaging agents and are often easy to locate in dynamic structures like tissue. This article touches on the latest advancements in cancer therapy, describes the benefits of gold nanoparticles in different types of cancer including breast, colon, lung, prostate cancer, etc., and evaluates recent research findings in the biomedical sector and lastly the limitations in this area.

Application of GNPs in Cancer Diagnosis

Precise detection of tumor location and depth in patients is required for successful cancer treatment. Currently, imaging systems, such as CT and MRI, are used for the clinical diagnosis and treatment of cancers. The GNPs are stable, nonimmunogenic and low toxicity. In addition, they can accumulate in the tumor sites due to the EPR effect so they are attractive in imaging diagnosis. (Yang et al., 2022)Traditional CT contrast agents are small molecular iodine-based compounds with short circulation time and side effects, such as vomiting and

itching, that limit their widespread use. GNPs are a promising CT contrast agent due to their high x-ray attenuation coefficient and biocompatibility. (MRI is an non-invasive imaging modality that is preferred to be applied in the soft tissue imaging due to its optimal tissue contrast resolution and multiplanarity. The GNPs are often conjugated with T1 or T2 contrast agents to make them applied in MR imaging. Compared to conventional exogenous agents, the absorption cross-section of GNCs is significantly improved due to their surface plasmon resonance (SPR) effect, which can produce strong photoacoustic signals. Furthermore, GNPs are more stable, have a higher laser damage threshold, and are biomedically inert making them a promising photoacoustic imaging contrast agent. Furthermore, it can produce a strong photoacoustic signal in the near-infrared light (NIR) region.These reports demonstrated that GNPs act as an imaging agent that can be utilized for cancer diagnosis. Compared to conventional diagnostic contrast agents, GNPs have an additional cancer therapeutic effect

Gold nanostructures for Cancer treatment

The AuNPs can be delivered by physiological transportation or conjugation with antibodies, and once they are delivered, they can self-assemble into large clusters inside the cells.By taking advantage of the great biocompatibility, enhanced permeability and retention effect, and the plasmonic optical properties, one typical strategy for using gold nanoparticles to treat cancer is photo-induced hyperthermia and immunotherapy. Furthermore, spherical gold nanoparticles can also self-assemble into three-dimensional nanostructures for cancer treatment. In a recent study, reported nanoprobes formed by self-assembly of ultrasmall AuNPs for dual-mode real-time imaging-guided photothermal/photodynamic combined therapy for liver cancer. The nanoprobes demonstrated excellent tumor-targeting capabilities in both T1 magnetic-resonance imaging and in vivo fluorescence imaging modes; meanwhile, the nanoprobes are able to dissociate and escape from the body to reduce aggregation in the body to minimize the possible toxicity.

Gold Nanorods

Another important type of gold nanostructure is gold nanorods, which possess some unique advantages for cancer treatment. For example, gold nanorods can absorb light in the near-infrared (NIR) region, enabling efficient irradiation, which can be utilized for selective photothermal therapy of some specific cancers.(He, et al 2022 paragraph 11) Specifically, thanks to the tunable localized surface plasmon resonance (LSPR), gold nanorods can not only serve as probes but also become heat sources when irradiated by a laser with a photothermal effect. The generated heat can provide photothermal therapy for cancer treatment and/or trigger anti-cancer drug release for chemotherapy when gold nanorods serve as a drug carrier. In short, gold nanorods can be applied for cancer treatment in phototherapy, cellular imaging, drug transport, and combined therapy. Employing the photothermal effects of gold nanorods to kill cancer cells is the most widely employed strategy for cancer treatment, as the nanorod can absorb the NIR light to penetrate into sick tissues without damaging the surrounding healthy tissues.This is a

very good point because one of the major drawback of the current treatment is they damage the healthy tissue. This is promising if we can keep the healthy tissues as such while destroying the cancer cells. The immune-targeted gold nanorods can target head and neck cancer cells with high specificity and facilitate the differentiation between cancerous and noncancerous tissues. A study discovered that the polyelectrolyte coating on the Au nanorods can have an important effect on the photothermal efficiency and the photothermally triggered cancer cell damage. The tunable LSPR properties of gold nanorods in the near-infrared spectral region, impressive in vitro cancer cell killing efficiency and remarkable tumor growth suppression were achieved in vivo by the gold nanorod–polydopamine composite, superior to any single therapy modality. Gold nanorods can integrate with other functional materials such as inorganic compounds to form a therapeutic package to further promote the efficiency of cancer treatment. old nanorods (GNRs) can serve as effective drug carriers for controllable drug delivery. A recent study discovered that cholesterol-coated gold nanorods can be an intriguing carrier for hydrophobic drugs, where efficient delivery and therapy against breast cancer cells can be achieved and another important finding is when functionalized with cyclodextrin(sugar molecules arranged in rings), gold nanorods can encapsulate doxorubicin (DOX)(its chemotherapy drug), and the as-formed nanocomplex showed enhanced anti-cancer efficacy

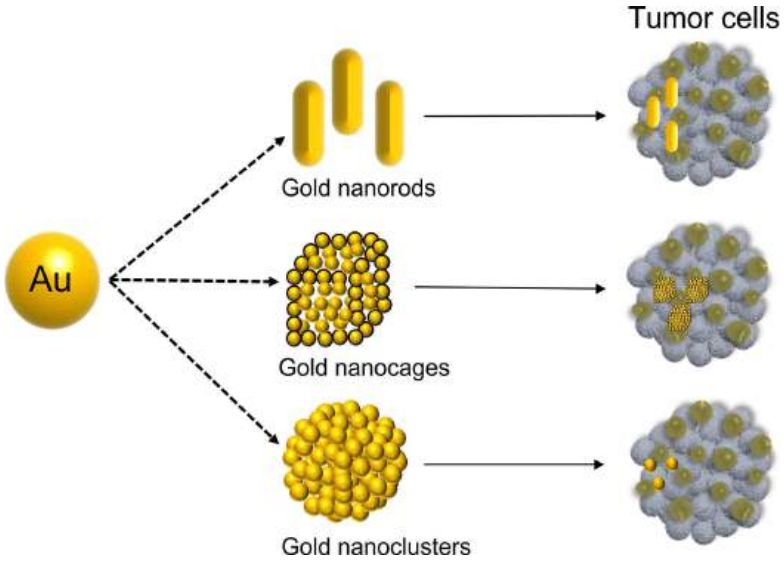
Spherical Gold Nanoparticles

Spherical Au nanoparticles have been widely utilized in cancer treatment mainly due to its unique physicochemical properties, their excellent biocompatibility, and its characteristic surface plasmon resonance absorbance with specific wavelengths. spherical Au nanoparticles can be applied for cancer treatment in several ways such as photothermal therapy, radiotherapy, tumor imaging, and serving as drug delivery systems. Spherical gold nanoparticles can also self-assemble into three-dimensional nanostructures for cancer treatment. In a recent study, reported nanoprobes formed by self-assembly of ultrasmall AuNPs for dual-mode real-time imaging-guided photothermal/photodynamic combined therapy for liver cancer.The nanoprobes demonstrated excellent tumor-targeting capabilities.

Atomically Precise Gold Nanoclusters

Gold nanoclusters (AuNCs), usually with a size less than 3 nm, are intermediate bridges between relatively larger plasmonic Au nanoparticles and Au complexes. Due to the ultrasmall-size-imparted quantum confinement effects, gold nanoclusters exhibit significantly different optical behaviors and chemical and catalytic properties compared with their nanoparticle counterparts. AuNCs can be fluorescent at a specific photo-emitting wavelength with a long lifetime that is quite favourable for imaging or as probes. Recent study states that accurate tumor imaging can be realized by gold nanoclusters conjugated with carborane derivatives, making accurate imaging-guided cancer treatment possible. Phototherapy is usually considered to be a more powerful means to cure cancer. recently, more and more research attention has been devoted to employing AuNCs in radiotherapy, in which ionizing

radiation is utilized for killing cancer cells. Atomically precise gold nanoclusters have demonstrated great potential for cancer treatment, mainly due to their rich surface functionalities, outstanding optical features , and great biocompatibility.



Gold Nanoparticle-Based Drug Delivery

A number of Au NP-based nano-drugs combined with other biomedical applications, including drug-conjugated Au NPs prepared for cancer therapy and tumor-targeting, are being studied]. Au NPs are known to be an effective nano-carrier for various drugs, such as peptides, plasmid deoxynucleic acids (pDNAs), proteins, small interfering ribonucleic acids (siRNAs), and chemo-therapeutic agents. The synthesis of anticancer particle aptamers and Au NP-modified Morin pH-sensitive liposomes with targeting properties was studied. Tumors with a pH lower than the normal tissue reduces the effectiveness of the drug. Au NPs can be used to deliver pH-responsive drugs such as Morin safely to the target for high effectiveness. In vivo results show that Au NP-encapsulated Morin can inhibit tumor growth in xenograft mouse models and promote tumor apoptosis.(Siddique et al.,2020) A new delivery system was developed where each Au NP is surrounded by many drug moieties, acting as an individual group against the microbial organisms and able to penetrate the cell wall effectively. This ability allows Au NPs to deliver a large number of antibiotic molecules into cells at a highly localized volume. Recently, many studies have been carried out to find new innovative solutions to develop highly efficient and reproducible fluorescent contrast agents for targeting specific cells. Gold nano-carrier-based platinum (II) and platinum (IV) drugs have enhanced the loading capability on the surface of the nano-carriers, and release in a pH-dependent environment. A study introduced a system with Au NP-based platinum drugs, and Aminoanthraquinone was incorporated into the unit, where they are both acting as fluoroprobes and DNA intercalators. The unit has distinct, unusual and remarkable DNA binding properties. Au NPs are also used on deeper sections of the eye such as the retina. A study was carried out to investigate how Au NPs can be used to deliver drug and biomacromolecules to the eye and how they can be used as active therapeutic tools. A cell

membrane coating strategy opened new opportunities for multifunctional drug delivery platforms. A study was conducted to introduce a system that loads up doxorubicin-loaded oligonucleotides attached to Au NPs as carriers in chemotherapy for prostate cancer. The result shows that this is a viable method and has promising potential. Moreover, three mercaptopropionic acid-capped Au NPs in resistant leukaemia cells were tested for carrier capability and biomarkers. The experimental result is also promising

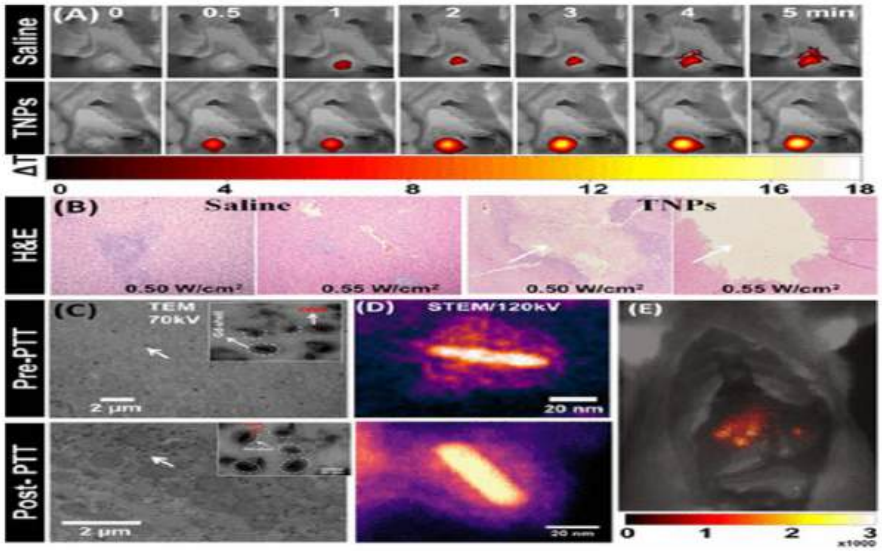
Photothermal Cancer Treatment

For photothermal cancer treatment, AuNP-based agents can convert the light-irradiation energy into heat and generate reactive oxygen species for ablating the tumor cells. Use of the AuNPs as a simple carrier of the photosensitizers for Photothermal Cancer treatment has many advantages. As we already know AuNP can accumulate in cancer cells more than they do in normal cells. When we treat cancer using Photothermal technology it is mainly a treatment based on light rays. In this therapy a photosensitizer (PS) is applied then activated by light in specific wavelengths and tries to destroy the cancer cells. If the medical science can combine AuNP with PS the PS can identify cancer cells more effectively and efficiently with the support of AuNP. And after that when the external light passes to the body light can more effectively pass through AuNP due to its high conductivity which will end up in destroying more cancer cells. This modern technology of treatment has wide potential. As GNPs are having an important role in the cancer treatment through PTT, GNPs show less toxicity in the human body and also kills tumor cells. As other therapies cause damage to healthy tissues and cells in cancer treatment, PTT does not cause that much damage to healthy tissues and cells. As GNPs are target-oriented, they directly show their therapeutic effect on tumor cells. In a study Chuanlan et al. synthesized GaAu loaded mesoporous silica nanoparticles (Ga-Au@mSiO2). GaAu filled mesoporous silica nanoparticles are known as an important nanomaterial for biomedical applications such as diagnosis and cancer treatment.

Besides surface coating, imprinting other biologically active molecules such as saccharides can also improve the photothermal treatment efficiency. One study prepared sialic acid (SA, a typical monosaccharide)-imprinted gold nanorods, which could selectively kill a tumor but not damage the circumjacent healthy tissue.(He, et al.,2022 para11) Besides achieving higher treatment efficiencies, researchers have also devoted great effort to unravelling the molecular mechanism of the Au-nanorod-aided plasmonic photothermal therapy. In 2017, a group of researchers conducted an investigation regarding the efficacy, toxicity, and mechanism of Au nanorod photothermal therapy of cancer in xenograft mice. In this study, the size, surface modification, and concentration of AuNRs and the laser power to achieve the maximal apoptosis induction was first examined. The possible mechanism of AuNRs-plasmonic photothermal therapy (PPTT) action using quantitative proteomic analysis in tumor tissues of the mouse was also studied, where several death pathways were identified. Cytochrome c and p53-associated apoptosis mechanisms were recognized to contribute to the enhancement of PPTT with

AuNRs@RF (rifampicin). Moreover, Pin1 and IL18-related signalling made a contribution to the disturbance of the NETosis pathway through PPTT enabled by AuNRs@RF.

In 2018, Joshi's group reported gold-nanorod-composed theranostic nanoparticles (TNPs) for interventional image-directed photothermal therapy for solid tumors. In this study, the feasibility of site-selective hepatic image-directed delivery of TNPs in rats was examined. Figure 2A shows the dynamic thermal imaging at different time points during the PPT process. In the saline group, the tumor's temperature increased by about 7.5 °C within 1 min and remained basically stable; however, in sharp contrast, the TNP group tumor temperature quickly jumped to ~20 °C in 5 min, suggesting that the increase in tumor temperature exceeded the range of hyperthermia, resulting in the damage of local vasculature which can destroy the tumor cells effectively. The authors further conducted the hematoxylin/eosin staining of tumor sections. As shown in Figure 2B, tumor slices in the saline group exhibited no obvious effect, while the TNP group presented a valid response under the same laser irradiation power level with a remarkable photothermal therapy effect. The transmission electron microscopy (TEM) images verified that the TNPs stayed in the tissue with no structural change, as illustrated in Figure 2C. In addition, the clear observation of the morphology of a gold nanorod core and a Gd shell can be observed in Figure 2D. Finally, Figure 2E validates the feasibility of intraoperative imaging-offered quantum yield, and the imaging sensitivity can be further improved by reducing the exposure time to below 1 s. The above findings confirm that TNPs can be employed for photothermal ablation efficiently while bearing no risk of heat-induced breakdown.



In vivo photothermal therapy of rats. 9(A)Thermal image of CRLM tumor containing rats. (B) Typical histological images of saline and TNP ablated tumor tissue stained with H & E. Pre PTT and post PTT tissue images recorded using (C)TEM at 70 kv and (D) STEM at 120 kv. (E)EMission monitoring after injecting 50 ul of TNP solution.

Synthesis of AuNPs

The synthesis methods of AuNPs are ‘top-down’ and ‘bottom-up’. Among these, the ‘top-down’ method involves the physical decomposition of bulk gold material in a gaseous or liquid environment, utilising techniques such as laser ablation, sputtering, milling, thermolysis, ultraviolet or infrared irradiation, and vapour deposition. This method typically yields highly pure AuNPs with uniform dispersion. However, this synthesis method not only requires more energy but also has higher instrumentation requirements. Additionally, defects in the surface structure of the products significantly impact their physicochemical properties. The ‘bottom-up’ method usually involves the self-assembly of gold atoms or molecules into uniformly sized AuNPs, which is cost-effective and easily controlled [**30**]. This method includes chemical synthesis (spinning, sol-gel process, and chemical reduction), biosynthesis, and physical hybridisation methods (electrochemical, sonochemical, and photochemical synthesis). The first two chemical synthesis methods are mainly used for the synthesis of nanofibers and nanogels, which are mostly used for the production of industrial materials. Among these methods, the chemical reduction method stands out because of its advantages of homogeneity, dispersion, enhanced stability, and precise control over size and shape. Approximately 60% of the AuNPs reported in the biological field were prepared using these methods from 2013 to 2018. Biosynthesis, an economical, environmentally friendly, and a highly biocompatible approach, has gained popularity among researchers in recent years. Furthermore, physical hybridisation methods have also been adopted for the synthesis of AuNPs in electrochemical or biosensor applications.

Chemical Reduction Method for Synthesis of AuNPs

The chemical reduction method involves two main steps in the synthesis of AuNPs. Firstly, the reducing agent provides electrons to rapidly reduce Au3+ to Au0, forming small clusters (**Figure 1**). The reducing agents mainly include AA, borohydrides, citric acid, and hydrogen peroxide. Secondly, stabilising/capping agents control the growth rate, final size, and morphology of AuNPs. These agents mainly include trisodium citrate dihydrate, surfactants (especially CTAB), and sulphur ligands. In some cases, the reducing agent and stabilising/capping agent can be the same substance. According to the differences in the synthesis system and conditions, this procedure can be classified into the Turkevich method, Brust–Schiffrin method, or seeding-growth method.

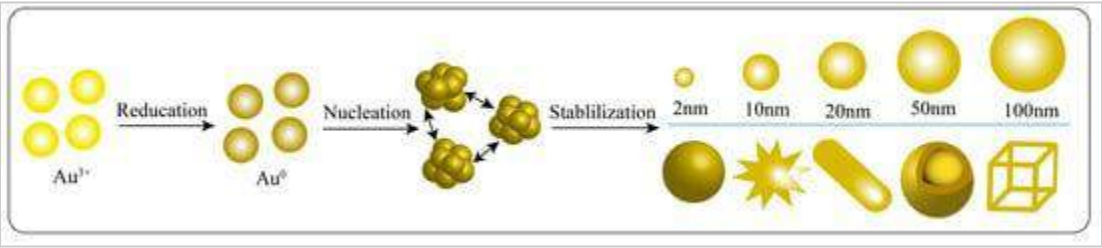


Figure 1. AuNPs with different sizes and morphologies were synthesized by the chemical reduction. The classical Turkevich method was first proposed by Turkevich in 1951 and refined by Frenc in 1973. This approach produces spherical AuNPs in a homogeneous aqueous solution system in which trisodium citrate serves as both a reducing and stabilising agent. In 2015, Polte elaborated on the theory of AuNP formation in detail. (1) Au3+ was rapidly reduced to Au0 and aggregated to form 1–2 nm clusters, and the decrease in the reduction rate resulted in the continued aggregation of unstable Au0. (2) The size of the clusters reached 2.5 nm, their numbers remained stable, but their sizes became larger and larger because of the poor dispersion of Au0. (3) After the size reached 4.5 nm, the reduction rate accelerated and consumed the remaining Au3+ in the solution. The reaction was terminated when the colour of the solution changed to ruby red. Typically, this method synthesises mainly spherical AuNPs with a particle size distribution between 10 and 30 nm under heating. The resulting AuNPs exhibit excellent dispersion and low toxicity. Based on the classical Turkevich method, Netti et al. synthesised AuNPs with different morphologies, such as nanowires, nanoflowers, nanowebs, and nanoplates at room temperature. The formation of these morphologies was achieved via the cold fusion of ultra-small spherical AuNPs. Furthermore, other substances, including L-tyrosine, ethanol, and poly-vinyl-pyrrolidone, were added to the reaction to manipulate the morphology and size of AuNPs at room temperature. To analyse the aggregation mechanism of AuNPs, the authors also developed a numerical diffusion-limited aggregation model to simulate the dimensions and growth kinetics of AuNPs. This report provides insights into the mechanism of reduction in gold ions by citric acid molecules. However, the Turkevich method has some shortcomings. For instance, the product yield is low, and nanoclusters aggregate uncontrollably, leading to irregular nanoparticles greater than 30 nm in size.

The Brust–Schiffrin method, based on the biphasic liquid–liquid system, overcomes the problem of severe agglomeration encountered in inhomogeneous systems. Commonly, TOAB is used as a two-phase transformer to transfer the HAuCl4 solution into the organic phase. Subsequently, reducing agents, such as CTAB and sodium borate, are employed to reduce gold ions and form AuNPs with a thiol layer. The thiol layer has a strong affinity for inhibiting the growth of AuNPs, resulting in small nanoparticles. This method allows the preparation of spherical AuNPs with excellent stability and a controlled particle size of <10 nm. However, poor dispersion of the prepared AuNPs is the largest challenge.

These two methods are mainly used for the synthesis of spherical AuNPs, whereas the synthesis of rod or heterogeneous AuNPs suffers from low yields and product defects. The Seeding-growth method addresses these issues. The method consists of two main steps. (1) A strong reducing agent (NaBH4) is used to reduce HAuCl4to form seeds. (2) The seed solution is added to a gold salt solution containing a weak reducing agent (e.g., AA), followed by the addition of a molecule (e.g., CTAB) to induce the growth of the seeds into Au NRs]. To achieve better control of the shape and increase the yield of the nanorods, a low concentration of additional ions, such as metal ions and halides, is added before the growth phase. However,

size and shape are influenced by various factors, such as seed concentration and size, reductant concentration, pH, temperature, and even the solution order.

Biosynthesis of AuNPs

Eco-friendly and non-toxic biosynthesis has also been widely explored for the synthesis of AuNPs, considering the toxicity and occurrence of adverse reactions in chemical methods. Many organisms, including plants and microorganisms (e.g., bacteria, fungi, algae, and yeast), have been demonstrated to act as reducing agents or synthesis sites for the preparation of AuNPs. The mechanism of this process involves the reduction of Au3+ to Au0 using biomass. Biomass contains a variety of organic compounds including phenols, terpenes, amino acids, alkaloids, steroids, coenzymes, flavonoids, proteins, reducing sugars nicotinamide adenine dinucleotide, and water-soluble alkaloids.(Zhang et al., 2022,para 7) Plants are the most popular raw materials for biosynthesis as they contain large amounts of reducing substances and are cheap and readily available. The first successful synthesis of AuNPs using biomass was reported in 1999. Since then, each part of the plant, including leaves, bark, flowers, fruits, seeds, rhizomes, and bark, has been used as a raw material to prepare AuNPs [**86**]. Usually, leaves and fruits are especially popular as they contain high concentrations of reducing substances. This method enables the synthesis of AuNPs with different shapes, structures, and physicochemical properties, with size distributions mainly in the range of 5 to 400 nm. Plants are suitable for the large-scale production of Au NPs because they do not require a sterile environment or complex processing. However, the effect of plants on AuNPs has not been elucidated because of their complex composition.

Some microorganisms can also synthesise AuNPs by enzymatic reactions that occur via extracellular biosorption, biomineralization, intracellular bioaccumulation, and metabolic processes. These enzymes enable the reduction of Au3+ to Au0. The shape, size, dispersion properties, and stability of nanoparticles can be controlled by adjusting the external conditions. Furthermore, bacterial synthesis is an exact and slow procedure that requires optimization of the synthesis state. Fungi are easier to culture and grow than bacteria. The metabolic environment of microorganisms is crucial for the synthesis of AuNPs and therefore needs to be optimised for their synthesis. However, the intracellular environment cannot be observed and regulated in real-time, which can affect the synthesis of AuNPs. AuNPs have extensive applications in cancer treatment and drug delivery. Owing to their unique physicochemical properties, including a high X-ray absorption coefficient, photosensitisation, surface plasmon resonance (SPR), ease of functionalization, and stability.

The optical and physical properties of AuNPs are dependent on their size, shape, and surface chemistry, which in turn affect their drug-carrying capacity, organ distribution, pharmacokinetics, photothermal conversion efficiency, and immunological properties as drug carriers. Therefore, it is important to thoroughly evaluate the properties of AuNPs prior to constructing drug carriers. Various sophisticated techniques are available for characterising the size, shape, and chemical properties of AuNPs. Ultraviolet-visible spectroscopy is commonly employed to characterise the SPR of AuNPs, providing a preliminary analysis of AuNP size and

distribution and assessing AuNP aggregation. Transmission electron microscopy reliably measures AuNP size and provides a clear structural morphology. Dynamic light scattering is useful for testing aggregation, and measuring the zeta site of AuNPs can determine their surface charge and solution stability. To investigate surface-bound substances and their forms, X-ray photoelectron spectroscopy and Fourier transform infrared spectroscopy can be used to analyse the surface chemical structures of AuNP complexes. Moreover, the quantity of drugs bound to the AuNPs can be calculated using ultraviolet detection technology.

In conclusion, there are various methods to prepare AuNPs, all of which have advantages and disadvantages. When AuNPs are used as drug carriers, various factors must be considered, such as the effects of size, shape, and surface properties on drug loading, in vivo transport efficiency, biocompatibility, and immune response. Next, we detail the use of AuNPs as drug carriers and their application in cancer immunotherapy.

Drug Delivery

A highly prevalent cancer among females is breast cancer.Originating from the mammary tissues, the tumor metastasizes further to the lobules or ducts of the milk-secreting gland. If detected early, it can be treated by chemotherapy, radiotherapy, etc., but if left treated, it can spread to other organs like the brain, bones, lungs, etc. The complex upon encountering hyaluronidase, exhibits size-reducible property, showed intra-tumor accumulation, gets evenly distributed in breast cancer cells, suppress tumor growth by up to 95% and prevented lung metastasis by 88%. Overall, the strategy provides a dual effect as an anti-tumor and anti-metastatic agent (Kesharwani et al.,2023)

Breast cancer is one of the leading cancers in females. The treatment and cure is purely based on the stage which is diagnosed. In this journal they are talking about the potential of the AuNPs dual effect as an anti-tumor and anti-metastatic agent. If the cancer is diagnosed in stage 4 which means the disease has spread to nearby organs called metastasis which makes the treatment complex and less effective. As the journal says, if the AuNP can prevent lung metastasis by 88% I would say it is remarkable and promising. As I do more research in the scope of AuNP I can capture more information on the scope of AuNP in metastatic breast cancer and further extend to other types of metastatic cancers. Fig. 1

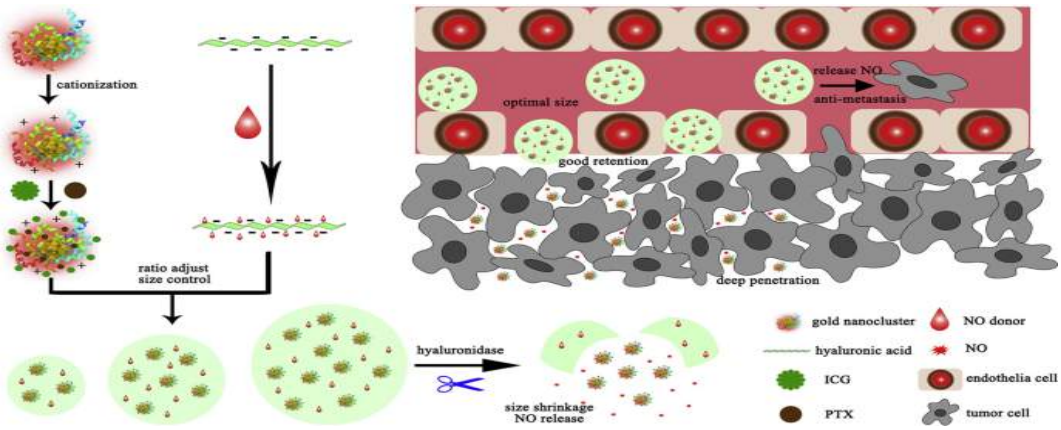
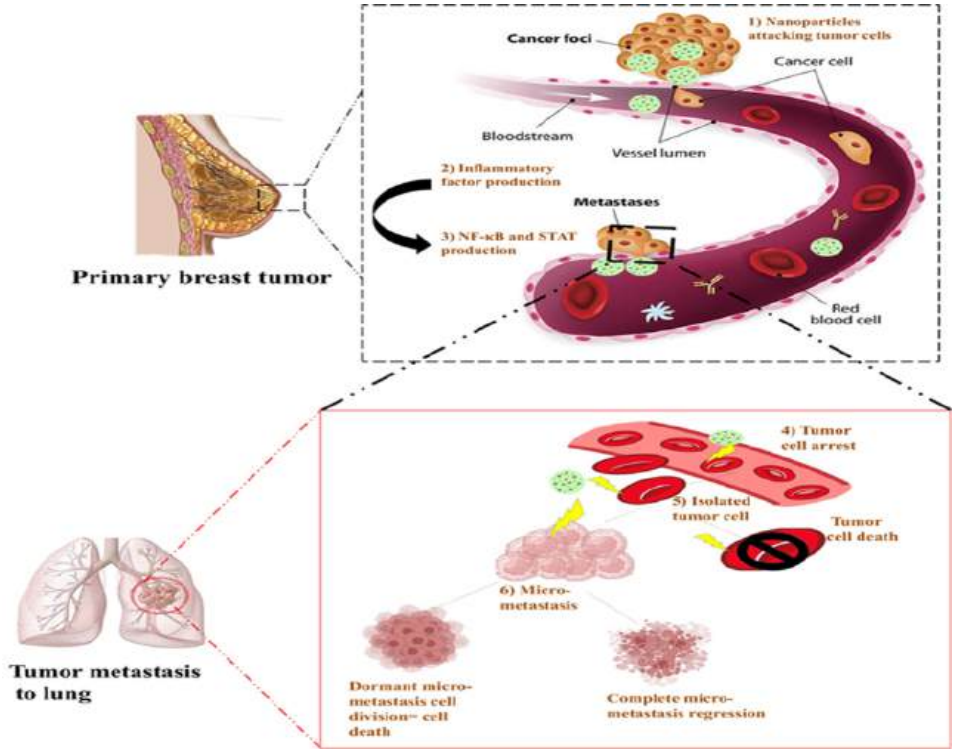


Illustration of action of size- reducible CD44 targeting, hyaluronic acid protected- red emission, renal-clearable, indocyanine green and paclitaxel loaded, cationic bovine serum albumin coated gold nanoparticles (NO3-ICG-PTX@ CBSA-AUNPs /HA) on tumor cells. Reproduced with permission from reference



Illustrating the primary and metastasis cancer and action of CD44 targeting, hyaluronic acid protected- red emission, renal-clearable, indocyanine green and paclitaxel loaded, cationic bovine serum albumin coated gold nanoparticles on primary breast tumor as well as metastatic lung tumor cells.

Challenges

The recent advances regarding gold nanostructures, namely gold nanoparticles, gold nanorods, and atomically precise gold nanoclusters, have been reviewed above. One can notice that the gold nanostructures hold great potential in cancer diagnostics and therapeutics, mainly thanks to the merits such as excellent optical properties, facile control of size and/or morphology, robust stability, the capability to tune the surface chemistry for conjugation with functional biological molecules, and especially the great biocompatibility. However, there are some challenges in this area.

1. Long term toxicity issue.The Gold nanostructures can easily accumulate in internal organs due to prolonged treatment and that can damage kidney, liver, spleen, lung etc.
2. The targeting specificity issue. Even though the gold nanostructures can be designed to bind to specific cancer cells, there is still an urgent need for cancer diagnosis and therapeutics at the early stage with a high level of targeting specificity. Currently, the widely employed cancer treatment strategies such as photoimaging and photothermal therapy still have the limitations such as non-specific binding and the unnecessary activation of the normal host immune response.
3. The modulation of the gold nanostructures to meet the complex biological environment can be challenging. Upon the surface modification of the gold structures, the pharmacokinetic parameters of the gold nanostructures and the cellular response will be correspondingly changed, while in vivo, the fundamental comprehensive understanding of the interactions between the gold nanostructures and the biological moieties is still lacking.
4. Some gold nanostructures can be used for both NIR-I and NIR-II imaging. However, when choosing both regions, the excitation wavelength range is quite limited, and the imaging effectiveness and efficiency has to have more improvement/advancement.. Determining how to modify the composition, morphology, and structure of these gold nanomaterials to work better for both NIR-I and NIR-II regions is still extremely challenging.

Conclusion

The scope of GNP in cancer diagnosis & treatment is wide and promising. There is plenty of research going on in this area. With many different properties of gold the cure for cancer is not far. Especially in the new AI era much more advanced treatment options can be expected despite all the challenges and drawbacks. We can envision that more treatment options and diagnostic procedures are just steps away.

Works Cited

https://www.mdanderson.org/cancerwise/could-gold-nanoparticles-help-treat-cancer.h00-159540534.html

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9094645/ https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2021.819329/full

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9146553/ https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5039679/ https://www.mdpi.com/2076-3417/10/11/3824 https://jnanobiotechnology.biomedcentral.com/articles/10.1186/s12951-022-01402-z

https://www.mdpi.com/1999-4923/15/7/1868

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9237953/#:~:text=Gold%20nanoparticles%20(AuNPs)%20are%20widely,%2C%20biological%20safety%2C%20and%20stability.

https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-023-01798-8#:~:text=The%20photothermal%20property%20of%20nanoparticles,resulting%20in%20tumor%20cell%20destruction

Energy Harvesting Roof: A Hybrid Approach to Sustainable Power Generation Using Piezoelectric, Photovoltaic, and Thermoelectric Technologies By Ayan Bera*

Abstract

This paper presents a novel, multi-modal energy harvesting roof system that complements photovoltaic, piezoelectric, and thermoelectric technologies to enable continuous clean energy generation across diverse environmental conditions. Unlike conventional solar systems limited by weather and time of day, this hybrid design has the potential to capture energy from sunlight, rainfall, and ambient thermal gradients, ensuring reliable performance day and night, rain or shine. Experimental prototypes validate the system's ability to harvest energy from multiple natural sources, demonstrating its potential to reduce carbon emissions, enhance energy resilience, and expand access to sustainable power in underserved regions. By leveraging the complementary strengths of each technology, the system promotes efficient utilization of renewable resources and supports the scalable deployment of clean energy infrastructure tailored to local environmental profiles.

Introduction

Problem Statement

In both industrial and automotive sectors, a substantial fraction of energy is dissipated as waste heat, estimated at approximately 66% in industrial operations [1] and up to 70% in vehicular systems [2]. This loss not only undermines energy efficiency but also contributes to environmental degradation through increased fuel consumption and emissions. Concurrently, while solar energy presents a clean and abundant resource although limited in the daytime, its reliability is compromised by weather variability, rendering it less effective during overcast or rainy conditions.

These limitations underscore a critical gap in current energy harvesting technologies: the absence of systems capable of sustained, multi-modal energy capture across diverse environmental scenarios. To address this challenge, there is a pressing need for a complementary solution that can harness mechanical, thermal, and solar energy maximizing resource utilization and ensuring consistent and continuous energy availability. Such a system would be instrumental in advancing sustainable infrastructure, particularly in regions with fluctuating climatic conditions and/or limited access to conventional power sources.

Literature Review

Several papers were identified for literature review [3-7] that address Energy harvesting technologies have garnered significant attention in recent years as sustainable alternatives to conventional power sources, particularly for decentralized and low-power applications. This

study draws upon three core modalities, photovoltaic, thermoelectric, and piezoelectric systems, each of which has been extensively researched in isolation.

Photovoltaic (PV)

Photovoltaic (PV) cells convert solar radiation into electrical energy through the photovoltaic effect [4], wherein incident photons excite electrons across the bandgap of semiconductor materials, generating electron-hole pairs. Silicon-based PV cells dominate the market due to their efficiency and scalability, while emerging materials such as perovskites offer promising improvements in cost and flexibility. PV systems have been widely deployed in residential and industrial settings, yet their performance is inherently dependent on sunlight availability, limiting their reliability under cloudy or nocturnal conditions.

Thermoelectric Generators (TEGs)

Thermoelectric generators utilize the Seebeck effect to convert temperature gradients directly into electrical energy [8-11]. Advances in thermoelectric materials, particularly those with high figure-of-merit (ZT) values, have enabled efficient energy conversion even under modest thermal differentials. TEGs are especially valuable in environments where waste heat is abundant, such as industrial machinery and automotive exhaust systems. Recent developments have demonstrated output powers up to 350 mW/m^2 from single devices [10], with integration potential into wearable and modular electronics.

Piezoelectric Transducers

Piezoelectric energy harvesting exploits mechanical stress, such as vibration or impact, to generate electrical charge in piezoelectric materials [12-13]. Raindrop-induced mechanical excitation presents a unique opportunity for energy capture in outdoor systems.
High-performance piezoelectric ceramics, such as lead zirconate titanate (PZT), exhibit strong piezoelectric coefficients and have demonstrated power densities of up to 2 mW/cm^2 under dynamic loading conditions[2]. These transducers are increasingly used in self-powered sensors and microelectromechanical systems (MEMS), though their output is typically intermittent and environment dependent.

While each of these technologies has matured independently, their complementary applications into a unified, multi-modal energy harvesting system, such as a hybrid roof capable of capturing solar, thermal, and mechanical energy, remains a relatively unexplored frontier. Recent perspectives on multi-source energy harvesting suggest that such integration could significantly enhance energy reliability and efficiency, particularly in variable environmental conditions. This study builds upon these insights to propose a novel architectural framework for continuous, adaptive energy generation.

Methods

Prototype Design

A prototype hybrid energy harvesting roof was developed consisting of a Piezoelectric roof on one side and PV-TEG roof on the other side as shown in Figure 1. The energy flow for the system is as follows. The mechanical energy from raindrops is transformed into electrical energy using piezoelectric transducers during the rain. During the sunshine the photo-voltaic cell converts photons from sunlight to electrical energy. At night the heat from inside the house passes through the thermoelectric generator (TEG) on the roof to the ambient outside by radiative cooling. The temperature difference across the TEG leads to generation of electrical energy.

Piezoelectric Roof

The experimental prototype of piezoelectric roof was fabricated for Piezoelectric transducer using a rigid wooden frame measuring 150 mm × 150 mm, serving as the structural base for the energy harvesting module. A layer of high-density polyurethane foam was affixed to the top surface to provide mechanical compliance and enhance stress transfer to the embedded piezoelectric elements. Commercially available lead zirconate titanate (PZT) discs, each with a diameter of 41 mm and a thickness of 0.3 mm, were uniformly embedded within the foam matrix in a grid configuration to optimize spatial coverage and impact response. Electrical leads were connected to the electrodes of each PZT disc. The PZT transducers are connected in series to amplify generated voltage using a breadboard.

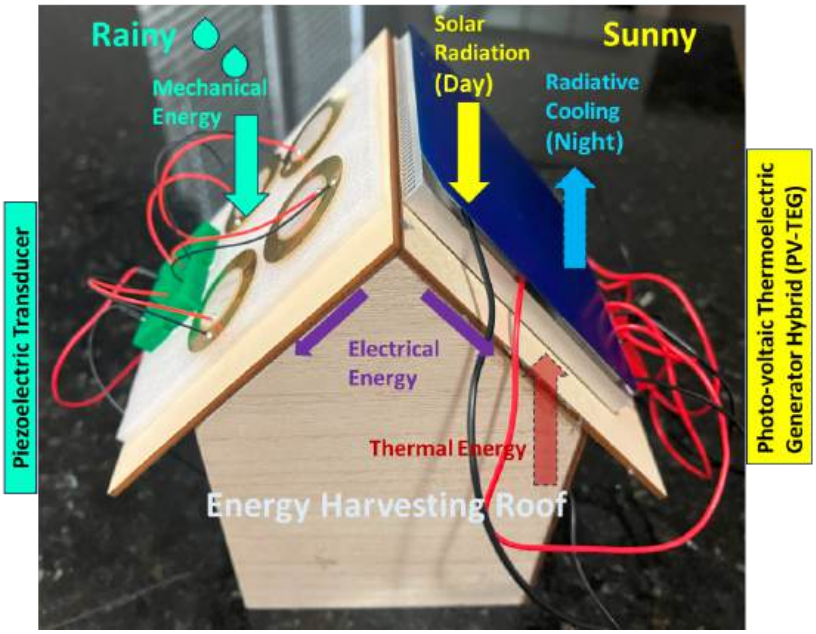


Figure 1. Energy harvesting roof consisting of a Piezoelectric roof and PV-TEG roof

PV-TEG Hybrid Roof

PV-TEG hybrid roof system is designed integrating SunPower Maxeon™ C60 monocrystalline solar cells on the top layer, thermoelectric generator (TEG) modules in the middle, and a passive aluminum heatsink at the bottom to facilitate thermal dissipation. This vertical stacking enabled simultaneous harvesting of solar and thermal energy from ambient conditions. The solar cells were exposed to direct sunlight, and their electrical output was measured using a digital multimeter by AMPROBE™. The power generated by the solar cell is measured using a power-meter by Dork.

Results

Piezoelectric Roof

Water Force	Peak Voltage (V)	RMS Voltage (V)	Power Output (mW)
Low	~2.4	~1	~0.11
Medium	~4.7	~2	~0.43
High	~6.5	~2.8	~0.83

Table 1. Voltage and power generated from prototype Piezoelectric Roof

To simulate rainfall conditions, a controlled water flow system was implemented using a garden hose fitted with a variable-pressure nozzle. The flow rate was adjusted for low, medium and high water force to replicate light drizzle, moderate rain, and heavy downpour scenarios. The dynamic mechanical excitation induced by raindrop impacts generated transient voltage signals across the PZT elements, which were captured using a Tektronix digital oscilloscope. **Table 1** shows the recorded voltage and power output. As a qualitative validation of power generation, the harvested energy was routed through a rectifier circuit to illuminate a low-power light-emitting diode (LED), confirming the viability of the system for real-time energy harvesting under simulated precipitation.

PV-TEG Hybrid Roof

The voltage and power values are recorded and plotted using box and whisker plot for 3 different trials as shown in **Figure 2**. Each trial consists of 6 data points.

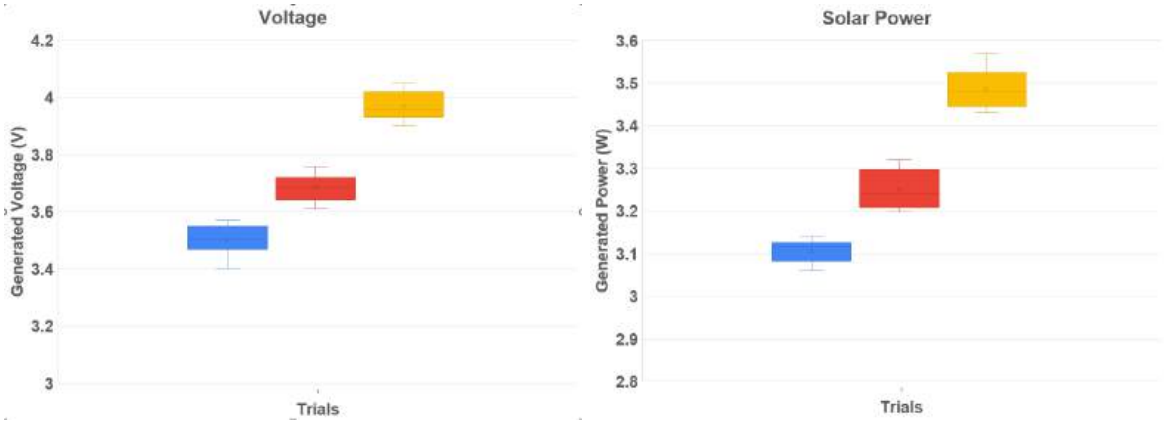


Figure 2. Box and whisker plots of voltage and power generated from solar cells.

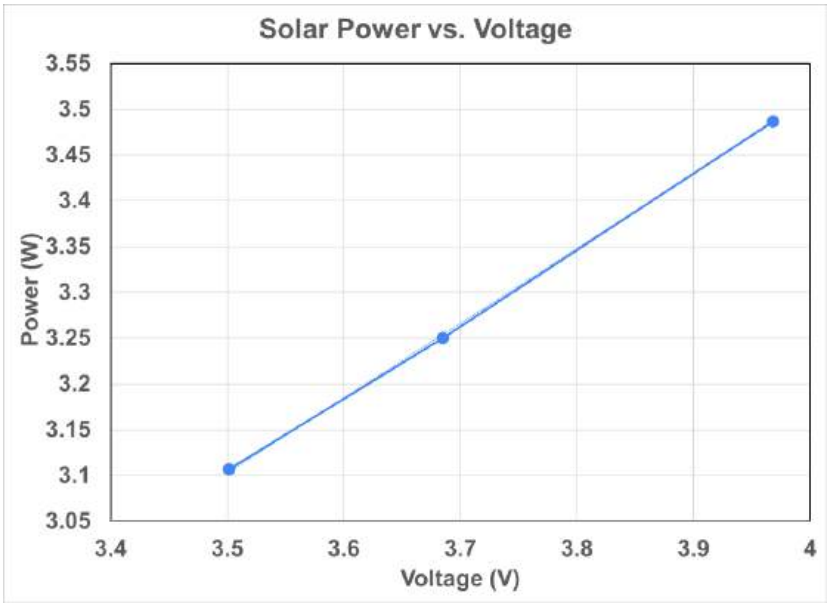


Figure 3. Line plot of mean solar power generated vs. mean voltage using solar cell

The mean values of generated solar power are plotted against the corresponding mean values of generated voltage. It is observed that with increase in generated voltage the generated power increases.

For the thermoelectric component, voltage readings were acquired under induced temperature gradients in clear sky and cloudy sky conditions as shown in box and whisker plot in **Figure 4**. The result is converted to power output using the manufacturer's datasheet-based calibration curves. TEGs generate on average 62.97 mV (0.23 mW) in clear sky, and 43.48 mV (0.11 mW) in cloudy sky.

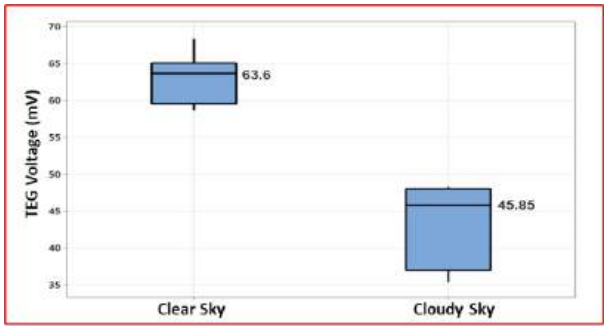


Figure 4. Box and whisker plots of thermo-electric generator (TEG) voltage output under clear sky and cloudy sky.

PV-TEG Hybrid Roof

The proposed hybrid energy harvesting roof system validates the feasibility of complementary energy harvesting by leveraging ambient environmental resources—namely solar radiation, thermal gradients, and mechanical energy from rain fall. Individually, each energy harvesting modality presents inherent limitations. Photovoltaic cells, while efficient under direct sunlight, suffer from reduced output during cloudy conditions or nighttime. Thermoelectric generators (TEGs) typically yield low power densities and require sustained temperature gradients, which may be difficult to maintain in passive systems. Piezoelectric transducers, though capable of converting mechanical stress from raindrop impacts into electrical energy, produce intermittent and low-magnitude outputs that are highly dependent on rainfall intensity and distribution.

Discussion

The strategic complementary utilization of photo-voltaic, thermoelectric and piezo-electric technologies into unified architecture significantly mitigates their individual shortcomings. By combining solar, thermal, and mechanical energy harvesting mechanisms, the system ensures a more resilient and continuous power supply across diverse environmental conditions. For instance, during sunny periods, photovoltaic cells dominate energy production, while TEGs contribute supplemental power from ambient heat. In contrast, during rainy or overcast conditions, piezoelectric elements become active, compensating for reduced solar input. This complementary behavior enhances overall energy reliability and reduces dependence on any single resource.

Beyond its functional versatility, the system exhibits key advantages in terms of form factor and deployment potential. The modular design is lightweight and compact, facilitating easy transportation and installation without the need for specialized infrastructure. Its adaptability allows for customization based on local climate profiles, making it particularly suitable for off-grid applications, disaster relief scenarios, and energy access initiatives in remote or economically disadvantaged regions. Moreover, the use of commercially available

components and low-cost materials supports scalability and affordability, aligning with global sustainability and equity goals [14].

In summary, the hybrid roof system not only demonstrates technical viability but also offers a promising pathway toward decentralized, ambient-powered energy solutions. Future iterations may explore enhanced energy storage integration, real-time optimization algorithms, and material innovations to further improve efficiency and broaden application scope.

Conclusion

The hybrid multi-modal energy harvesting roof system presented in this study offers a novel and multifaceted approach to decentralized clean energy generation by simultaneously capturing mechanical, thermal, and solar energy from ambient environmental sources. Through the complementary utilization of piezoelectric transducers, thermoelectric generators, and photovoltaic cells, the system addresses the limitations of single-mode energy harvesters and ensures more consistent power output across diverse weather and operating conditions. This innovation directly supports several United Nations Sustainable Development Goals (SDGs) [14], including SDG 7 (Affordable and Clean Energy), SDG 12 (Responsible Consumption and Production), and SDG 13 (Climate Action), by promoting renewable energy adoption, reducing reliance on fossil fuels, and enabling sustainable infrastructure in underserved regions. Its lightweight, and modular design make it highly adaptable for deployment in remote, disaster-prone, or low-income communities, where access to reliable energy remains a critical challenge. As a scalable and cost-effective solution, the system holds promise for advancing energy equity and resilience in the face of global environmental and socioeconomic pressures.

Limitations

While the hybrid roof system presents a promising approach to ambient energy harvesting, several limitations must be acknowledged to guide future development. First, both piezoelectric and thermoelectric modules typically produce low power densities, often in the milliwatt range, which restricts their ability to support high-demand applications without extensive scaling or energy storage. The system's performance is also highly dependent on environmental conditions—solar output fluctuates with irradiance, piezoelectric harvesting requires active rainfall, and thermoelectric efficiency hinges on maintaining stable temperature gradients, which can be difficult in passive configurations. Thermal management poses additional challenges, as sustaining a sufficient heat differential across TEGs without active cooling or heating mechanisms is complex in compact roof structures. Mechanical durability is another concern; long-term exposure to rain and environmental stressors may degrade the integrity of piezoelectric elements, especially if water ingress compromises insulation or adhesion. Moreover, integrating three distinct energy harvesting technologies increases system complexity and cost, particularly when incorporating wireless sensors and microcontrollers. Without robust energy storage and intelligent load matching, the system may struggle to deliver consistent power to downstream devices. Finally, scalability and aesthetic integration into

existing architectural designs remain hurdles, as weight distribution, visual impact, and structural compatibility must be carefully considered for widespread adoption. Addressing these limitations will be critical to realizing the full potential of hybrid multi-modal energy harvesting systems in real-world applications.

Acknowledgements

I would like to thank my advisor, Sarmistha Singha.

Works Cited

Mahlalela, T. (2019). Electric power generation potential based on waste heat and geothermal resources in South Africa [Master's thesis, Stanford University]. Stanford University Research Archive. purl.stanford.edu.

Precedence Research. (2024). Waste to energy market - global industry analysis, size, share, growth, trends, regional outlook, and forecast 2024 – 2033. www.precedenceresearch.com

Prasad, N., Li, W., Karan, S. K., Ghosh, S., Smriti, R., & Poudel, B. (2024, April 21–25). A perspective on multi-source energy harvesting technology using piezoelectric and thermoelectric materials [Conference paper]. SPIE Defense + Commercial Sensing, National Harbor, MD, United States. https://ntrs.nasa.gov/citations/20240006958

Al-Ezzi, A. S., & Ansari, M. N. M. (2022). Photovoltaic solar cells: A review. Applied System Innovation, 5(4), Article 67. https://doi.org/10.3390/asi5040067

Jouhara, H., Żabnieńska-Góra, A., Khordehgah, N., Ahmad, D., Lipinski, T., Lester, S. P., Meskimmon, J., & Wrobel, L. C. (2021). Thermoelectric generator technologies and applications. International Journal of Thermofluids, 9, Article 100063. https://doi.org/10.1016/j.ijft.2021.100063

Ghazanfarian, J., Mohammadi, M. M., & Uchino, K. (2021). Piezoelectric energy harvesting: A systematic review of reviews. Actuators, 10(12), Article 312. https://doi.org/10.3390/act10120312

Aldahiry, D. A., Bajaba, D. A., Basalamah, N. M., & Ahmed, M. M. (2022). Piezoelectric transducer as an energy harvester: A review. Yanbu Journal of Engineering and Science, 19(1), 30–35. https://doi.org/10.53370/001c.33771

Assawaworrarit, S., Fan, S., & Li, W. (2022). Radiative cooling-based nighttime electricity generation. iScience, 25(9), Article 104858. https://doi.org/10.1016/j.isci.2022.104858

Assawaworrarit, S., Omanakuttan, V., & Fan, S. (2022). Nighttime electric power generation at a density of 50 mW/m² via radiative cooling of a photovoltaic cell. Applied Physics Letters, 120(14), Article 143901. https://doi.org/10.1063/5.0085205

Assawaworrarit, S., Zhou, M., Fan, L., & Fan, S. (2025). Nighttime electric power generation at a density of 350 mW/m² via radiative cooling. Cell Reports Physical Science, 6(1), Article 102362. https://doi.org/10.1016/j.xcrp.2024.102362

d'Angelo, M., Galassi, C., & Lecis, N. (2023). Thermoelectric materials and applications: A review. Energies, 16(17), Article 6409. https://doi.org/10.3390/en16176409

Nayan, N. M., Razak, M. F. A., Ali, A., Mazalan, S. K., Abdullah, A. N. N., & Rahman, N. H. A. (2015). Development of rain harvesters using piezoelectric sensors. Jurnal Teknologi, 76(14), 11–15.

Salimin, M. A., & Nayan, N. M. (2024). Experimental validation of piezoelectric series array configuration for energy harvesting from rainfall. Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, 119(1), 68–75. https://doi.org/10.37934/arefmht.19.1.6875

United Nations. (n.d.). Goal 7: Affordable and clean energy. https://www.un.org/sustainabledevelopment/energy/

Forecasting the Unforeseeable: Training AI to Detect Financial Black Swan Events
By Tommy Maoz

Abstract

Black-swan events represent rare, high-impact occurrences that are extremely difficult to predict and have caused repeated disruptions to global financial systems, from the 1929 Wall Street Crash to the COVID-19 market collapse. Traditional forecasting tools fail for such events because they rely too heavily on structured economic data and cannot capture hidden risks. This paper examines black-swan incidents in diverse financial markets, including equities, foreign exchange, and commodities. This paper explores how artificial intelligence (AI) can close the forecasting gap by combining economic indicators with unstructured signals, such as news and sentiment, to be able to mitigate the issue.

The paper further investigates how AI works with new forms of economic indicators to improve early-warning systems for market crashes. It evaluates machine learning models that use neural networks and time-series forecasts to identify warning signs before economic crises occur. The proposed testing approach involves training AI models on historical data and testing them on real-time market input analysis to stress test across crisis conditions.

We conclude that combining structured macroeconomic data with unstructured sentiment and geopolitical inputs presents a promising direction for predicting black-swan events, even though complete predictability remains impossible with current AI technology. The AI-based system shows potential to enhance financial system instability detection, which would deliver essential market insights to investors, institutions, and regulators in today's complex global economy.

1. Introduction

Can artificial intelligence detect or help forecast financial black swan events by analyzing structured and unstructured data? A black swan event is characterized by two phenomena. The first is randomness and the difficulty of predicting where and when these changes to the market will strike. The second is a severe negative impact, which has consequences.

The most critical historical events are merely the result of previous trends and events (even if not abundantly clear then); many were seen coming. This usually comes as a significant price drop in the financial markets, leading to chaos and confusion. This is part of the dilemma with black swans because they are so rare and unexplainable that it is difficult to understand precisely what incites them. After the fact, people point to one thing or another, and they may be right, but there is confusion and chaos at the moment. A more recent example is the 2020 COVID-19 pandemic, which caused mass market disruption, but these events have been prevalent throughout history. Averting these consequences would be beneficial, so predicting them would hedge investors' bets and make them feel more secure in the market.

If one could predict these downturns and collapses, one could protect their investments and know when to exit or short the market, hoping they would profit. If caught early, the crisis

could be averted, and the market could remain healthy. The problem is that black swan events are inherently unpredictable and rare, so a large-scale analysis of multiple markets would need to be conducted throughout an extended time frame (Real estate, NASDAQ, S&P500, etc.). A human can not possibly go through the millions of daily stock trades. Still, an Artificial Intelligence or AI, for short, could comb through these trades and, with proper training, try to foresee the future, identify these quantitative indicators, and correlate them with qualitative real-life data (Politics, Natural Disasters, etc.) to anticipate the black swan event.

2. Characteristics of Black Swan Events

Unpredictability: is because of the inherent unpredictability of black swan events, which are inherently rare which makes data. Not much to go on. There have not been that many large-scale black swan events in our history. The amount is probably less than 500. The upside, however, to predicting these events is huge for investors, big and small. Wall Street to your average Joe, if there were a 100% accurate way to predict large market drops, everyone would use it (Which could undermine its purpose, but that is beside the point).

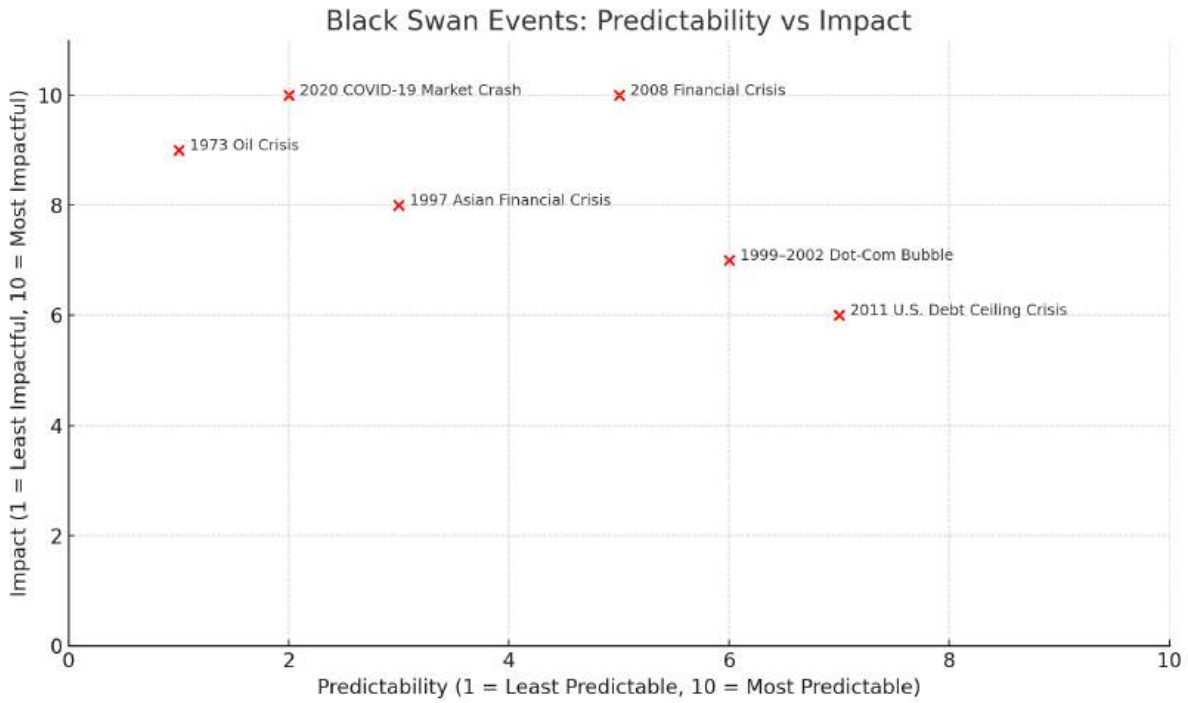
Retrospective Predictability: Black swan events are often rationalized as predictable only after they have occurred. Even the events that seem predictable now often fall under retrospective predictability. This is when the event in hindsight seems predictable, but in the moment appears random. People point to this reason and that reason after it has happened, and maybe they are right, but at the moment, it is kind of scary to see this large crash; one does not know how far it will go and its repercussions. In fact, one machine learning study found that a random forest algorithm correctly predicted U.S. recessions with about 70% accuracy up to four quarters ahead, compared to less than 50% accuracy for traditional forecasts (Nyman & Ormerod, 2017). Sometimes there was no way to know, so this algorithm could never be 100% accurate. An example is the 9/11 attacks, which had stocks crumbling (*Black Swan event*). A computer could react much quicker than a human, analyze vast amounts of information coming in, and find a way to sell faster. In a way, it could mitigate losses.

3. Examples of black swan events

The problem with black swan events is that they are unforeseen at the time, but afterwards, they are explained as if they were predictable. There is a difference between various black swan events: “Stock and other investment markets are affected by all types of events. Downturns or crashes such as Black Monday, the stock market crash of 1987, or the dotcom bubble of 2000 were relatively "model-able," but the Sept. 11 attacks and the COVID-19 pandemic were far less so. And who expected Enron to implode at the time? As for the Bernie Madoff Ponzi scheme, one could argue there were red flags.”(Bloch *Black Swan events and their impact on Investments.*) Some black swan events are truly unpredictable. When the DJIA (Dow Jones Industrial Average) plummeted more than 20% in the 1987 “Black Monday,” it was unexpected, but there were reasons for the crash. A confident Mr. Tudor Jones predicted the crash (Smith: *After a dazzling early career, a star trader settles down*). He made an estimated

100 million dollars from predicting the event by shorting the right stocks. He is good at finding the tops and bottoms of the market. With enough information, one could deduce that an artificial intelligence or a complex trading algorithm could predict these market highs and lows, specifically unexpected ones, which are the black swan events we discussed.

This paper presents a series of individual black swan case studies ordered from the least predictable to the most predictable to illustrate a key argument. While black swan events are defined by their unpredictability, many show early warning signs that, in theory, artificial intelligence could detect. Some events, like the 2019 Japanese yen flash crash or China's 2013 Bitcoin ban, were so sudden and externally triggered that neither human investors nor AI systems could reasonably anticipate them. These appear truly unpredictable. However, events like the COVID-19 crash, the 1997 Asian financial crisis, and the 2011 U.S. debt ceiling panic show more apparent signs, such as market instability, systemic debt exposure, or political dysfunction, that a well-trained AI could flag as red flags before a larger market collapse occurs. The structural crises at the end of the spectrum are the more predictable events. Next on the list are the 2008 financial meltdown, the Dot-com bubble, and the 1929 stock market crash. Each of these unfolded in slow motion. They had visible indicators: rising leverage, speculative behavior, or deteriorating fundamentals. With its economic data and historical training, AI could just as easily train human investors to spot these events before they occur. By organizing the examples this way, the section shows different levels of foreseeability, supporting the broader idea that AI can enhance market foresight even when complete prediction is not feasible. Even an increase in statistical average increases in predictability, yielding important results over a long enough time frame.



The 1973 Oil Crisis

The 1973 Oil Crisis is a classic black swan event that profoundly impacted the global economy. The crisis began in October 1973 when members of the Organization of Arab Petroleum Exporting Countries (OAPEC) proclaimed an oil embargo in response to the United States' and other nations' support of Israel during the Yom Kippur War. The embargo targeted countries perceived as supporting Israel, leading to a sharp reduction in oil supplies to these nations.

The immediate consequence of the embargo was a quadrupling of oil prices, which caused severe economic disruptions in the affected countries. In the U.S., for example, the cost of oil rose from $3 per barrel to nearly $12 per barrel in months. This sudden and drastic rise in oil prices led to widespread inflation, economic stagnation, and a severe energy crisis, with fuel shortages and long lines at gas stations becoming common. This would have been hard to catch by AI, because of the unconventional way this affected the market.

The 1997 Asian Financial Crisis

The 1997 Asian Financial Crisis is another significant example of a black swan event. The crisis began in Thailand with the collapse of the Thai baht. The Thai government was forced to float the currency due to a lack of foreign currency to support its fixed exchange rate with the U.S. dollar. The baht's devaluation triggered a financial contagion, causing stock markets to crash and economies to suffer across much of East Asia.

The crisis spread rapidly to other Southeast Asian nations, including South Korea, Indonesia, Malaysia, and the Philippines, leading to severe economic contractions, the collapse of asset prices, and a widespread loss of investor confidence. Foreign investors, who had heavily poured money into these rapidly growing economies during the previous decade, pulled out their investments en masse, exacerbating the financial instability. The International Monetary Fund (IMF) intervened with bailout packages to stabilize the affected countries. Still, the impact of the crisis lingered for years, leading to political and social upheaval in many of these nations.

Like other black swan events, the Asian Financial Crisis was difficult to predict. Before the crisis, many affected economies were considered economic "tigers" due to their rapid growth and industrialization. Few foresaw that the combination of high debt levels, weak banking systems, and overvalued currencies would lead to a swift and dramatic collapse. In hindsight, economists pointed to unsustainable financial practices, heavy reliance on short-term foreign loans, and inadequate financial regulation as contributing factors. Still, the crisis surprised much of the world at the time.

This event demonstrates the interconnectedness of global markets and how financial instability in one region can quickly spread and have ripple effects worldwide. It also highlights the fragility of emerging markets, which, despite rapid growth, can be vulnerable to sudden shocks that trigger widespread economic disruption. The Asian Financial Crisis serves as a reminder of the potential for black swan events to disrupt seemingly stable economies and the global financial system.

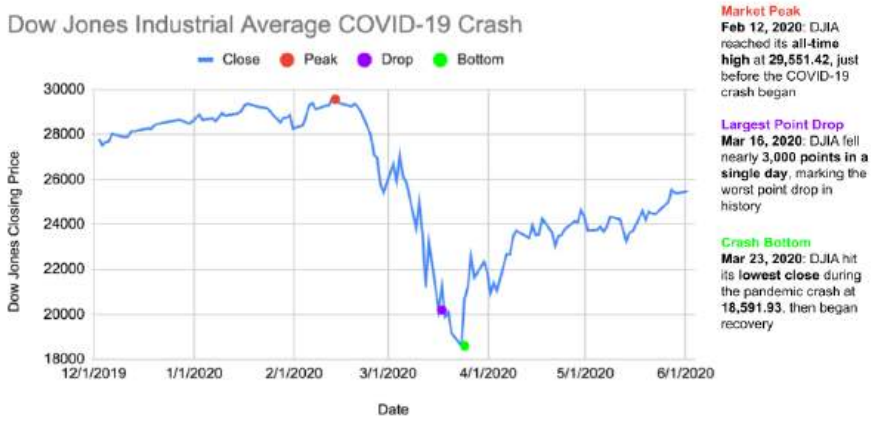
2020 COVID-19 stock market crash

This is one of the most recent market crashes and can be classified as a black swan event. Black swan events are unpredictable and quite sudden. The 2020 crash fits in all these criteria quite well. It is inspired by a sudden, unexpected event that sparked significant change, not just in finance, but in the entire world's population. This paper outlines the event (Min Shu et al., *The 'COVID' crash of the 2020 U.S. stock market*). They say the 2020 market crash was the event that "served as the spark of the 2020 U.S. stock market crash." They argue that the 2020 market crash served as the inciting event for the crash, while the actual crash was caused by market instability. This further supports the argument that artificial intelligence, with sufficient data, could potentially predict such crashes. The reason behind this crash was not solely a sudden, unpredictable event, but rather market instability, which an AI could anticipate and protect investors from. "The Wilshire 5000 Total Market index dropped around 12,068.8,"—the Wilshire 5000 Total Market index. The FT Wilshire 5000 Index (FTW5000) seeks to capture 100% of the investible U.S. market.

The same paper (Min Shu et al., The 'Covid' Crash of the 2020 U.S. Stock Market) attributes this significant dip not solely to COVID-19 but also to underlying market instability. This is an important distinction to make. The market was already vulnerable, as the depth of the crash reveals. An AI could detect that vulnerability and the potential for a future significant drop in the market again before panic sets in.

Theoretically, artificial intelligence systems could be trained on a mixture of real-time sentiment data, trading volumes, and macroeconomic indicators to identify the early signs of structural weakness. For example, suppose artificial intelligence were fed the kind of data it usually ingests. In that case, sentiment indicators trended negatively long before the system crashed. Likewise, liquidity accessibility seemed to be drying up across the primary banking sector. In that case, an AI with these indicators might flag such an event as a pre-crash warning sign.

The unprecedented speed, scale, and scope of the COVID-19 market crash might have seemed at first glance to be driven by unpredictable forces. But this paper argues that if a machine can learn to recognize the buildup to such crashes, humans should be able to identify the same signals and be better prepared for future events. At the very least, we should be able to reduce the risk and cushion the impact of market crashes.



The viewer can see the Dow Jones Industrial Average (DOWJ) dropped about 11,000 points from February 20th to March 20th (as is shown above). This significant drop in only one month (mid-February to late March) is unprecedented. This means the Dow Jones Industrial Average dropped about 11,000 points, or roughly 37% of its value, in just over a month, one of the fastest and deepest declines in market history.

This significant dip, however, according to the same paper (Min Shu et al., *The 'COVID' crash of the 2020 U.S. stock market*), can be attributed to market instability as opposed to COVID. This metric can be foreseen, and an AI with both technical market knowledge and a real-time market and world event analysis could (with some efficacy) predict market crashes.

By visualizing key inflection points of the COVID-19 crash, this chart helps illustrate how artificial intelligence could be used to flag early warning signs. While the pandemic may have been an external trigger, the steepness and clustering of the decline between February 20 and March 23 suggest that market instability was already present. AI systems trained on historical price volatility and investor sentiment, and incorporating reasoning, LLMS, following real-time news flow, might have detected early indicators of this downturn before it escalated into panic, giving investors time to react.

The 2011 Debt Ceiling Crisis

The 2011 U.S. debt ceiling crisis is a powerful case of how politics, not events within business, can cause extreme market swings. Investors panicked as Congress approached its borrowing limit, and the S&P 500 dropped nearly 20% in just a few weeks. The sell-off was driven less by economic data than by media-fueled fear and public uncertainty. In hindsight, the economy was relatively stable—unemployment was falling, GDP growth was modest, and corporate earnings remained strong. Yet the market reaction overshot dramatically on the downside.

In this situation, artificial intelligence (AI) could act as a stabilizing force. Unlike human investors prone to panic, AI systems can analyze large volumes of real-time macroeconomic data, such as consumer spending, job reports, and bond yields, to assess economic health. They

can also be trained on past political standoffs to detect patterns in U.S. fiscal behavior. This makes it possible to calculate the high probability of a last-minute deal, just as in previous debt ceiling negotiations.

AI doesn't get scared by headlines or emotionally swayed by politics. While humans often trade based on instinct or fear, AI evaluates probability and patterns. In this case, an AI system might not only have resisted selling during the drop but could have recognized it as an opportunity, identifying that market sentiment had diverged from economic reality. The debt ceiling panic provides a compelling example of how AI can help mitigate overreactions and bring a more rational perspective to volatile market moments.

Dot-com Bubble (2000-2002)

The aftermath of the Dot-com bubble bursting was far-reaching, impacting not only individual investors but also broader financial systems and economies. Many retail investors, lured into the market by promises of quick wealth, saw their savings evaporate as stock prices plummeted. Venture capital firms that had poured billions into speculative startups were left holding worthless equity, and countless tech companies, some of which had been hailed as the future of innovation, declared bankruptcy. Even giants like Cisco, Intel, and Amazon saw their stock prices drop significantly, though they managed to survive and eventually recover.

The crash also profoundly affected market psychology and regulatory practices. Investor confidence was severely shaken, leading to a more cautious approach to speculative investments in the following years. Regulatory bodies like the Securities and Exchange Commission (SEC) began implementing stricter guidelines to curb market manipulation and promote transparency. For instance, the Sarbanes-Oxley Act of 2002 was introduced to enhance corporate accountability and prevent fraudulent financial reporting, partially in response to the corporate scandals and economic instability of the time.

From an economic perspective, the Dot-com crash contributed to a mild recession in the early 2000s, with significant job losses in the tech sector. However, it also laid the groundwork for a more sustainable growth trajectory in the technology industry. Companies that survived the crash were forced to adopt more disciplined business practices, focusing on profitability and long-term viability rather than speculative growth. This period of recalibration ultimately paved the way for the tech boom of the late 2000s and 2010s, which was characterized by the rise of solid, innovative companies like Google, Facebook, and Apple.

The Dot-com bubble is a classic example of how speculative euphoria, driven by technological innovation, can lead to unsustainable market conditions. It highlights the importance of critical evaluation, disciplined investing, and the need for tools and systems, like AI, to help identify warning signs of market bubbles before they reach catastrophic levels.

2008 Financial Crisis

The 2008 financial crisis, or "The Great Recession," was a good example of a Black Swan Financial Event. This was one of the most impactful events on the world's economy. The

severe global economic downturn that began in 2008 is now known as the Great Recession. It was triggered by the collapse of the U.S. housing market, and it affected countries worldwide. The event has led to subprime mortgage lending and MBS (mortgage-backed securities) being taught as clear examples of how not to run a financial system. It has also led journalists to use phrases like "the worst economic downturn since the Great Depression" to describe the event. The recession was marked by spectacular failures of high-profile banks, massive bailouts, and a severe credit crunch that resulted in widespread unemployment and a fragile economic recovery.

A paper titled “Predicting Economic Recessions Using Machine Learning Algorithms” (Nyman and Ormerod) delves into the 2008 recession. It uses machine learning algorithms (specifically, a random forest learning technique, which was the best) to predict a set number of quarters ahead in the market. It then compares these with traditional indicators of recession. They found that the AI beat professional human forecasters. It also found that the AI model gave early warnings while human economists were still optimistic.

The AI caught hidden warning signs that were too complex or subtle for humans to spot. The model learned that inverted yield curves, slowing stock returns, and rising private debt were strong signals of a coming recession, something that human economic analysts consistently use. So, the machine learning algorithm was able to detect and rationalize hyper-complex, layered patterns in the data, which is precisely what black swan detection needs. This paper proves that AI could act like a “market smoke detector,” catching signs of trouble well before humans react.

4. Why do these Black-Swan events happen?

While the Black Swan events appear random, they are often foreseeable, stemming from deeper systemic issues often ignited by a random event. In financial markets, the true catalyst may not be the shocking event (Virus, War, Bankruptcy), but instead fragile conditions that have already existed, which make the markets susceptible to a crash. Overleveraged banks and inflated asset prices can silently increase the probability of collapse. This section explores the underlying causes that make financial systems vulnerable to extreme events, and why these triggers remain hidden until it is too late. To solve the issue, one must understand the issue, and that is what I attempt to do here.

Underlying Causes

Factors such as geopolitical tensions, economic bubbles, or systemic weaknesses within financial systems may trigger black swan events. For instance, the 1973 Oil Crisis was geopolitical, while the 2008 Financial Crisis was due to excessive risk-taking in mortgage lending.

Systemic Vulnerabilities

Some black swan events expose weaknesses in the system, like the overvaluation of internet companies during the Dot-Com Bubble or excessive reliance on short-term loans in the

1997 Asian Financial Crisis. Identifying such vulnerabilities is challenging without a comprehensive, real-time analysis, which is necessary for AI to be valuable.

Unpredictability of External Shocks

Events like the COVID-19 pandemic or the 9/11 attacks are almost impossible to foresee, but they have dramatic financial repercussions. The challenge is predicting these events and understanding how they will impact interconnected global markets.

5. Ways to Predict Black Swan Financial Events

Even though they are thought to be "unpredictable," black swan events often have delicate, detectable traits. These are hard for humans to identify, but are potentially within the detection range of artificial intelligence. A traditional economic forecast, for instance, takes historical data combined with a tight, often linear model to yield a prediction. Such models may work well for smooth periods of relatively stable market environments. AIs can detect early warning signs because of their remarkable ability to notice patterns. This system, in this case, is our economy. AI can detect non-obvious patterns that humans might miss; it could be used to foresee certain events. This section explores AI and traditional analytical tools and models that aim to detect early warning signs of economic instability.

Traditional Indicators

Interest rates, inflation, and employment are the traditional good indicators of the economy. These can provide a more general picture of the economy's condition and predict financial crises or upswings. For example, increased unemployment rates or sudden inflation usually signify that a recession will happen. However, these often are not enough to predict black swan events, rare and unpredictable events that lie outside normal expectations. Black swan events, such as the 2008 financial crisis or the COVID-19 pandemic, usually materialize from non-economic causes, like geopolitical conflicts, technological disruption, or natural calamities, which traditional economic metrics do not consider.

The limitation of relying on these indicators is that they are backward-looking. Many of them reflect existing conditions and rarely predict future anomalies. But then, the interdependence of contemporary global markets is such that what looks like a minor event in one sector or region snowballs into a worldwide crisis. This complexity calls for more sophisticated tools, capable of analyzing a wide range of variables in real time, which traditional indicators are ill-equipped to do.

Algorithmic Trading and AI

The invention of algorithmic trading and AI has drastically changed the financial markets by introducing new ways of prediction and response to market fluctuations. Machine learning algorithms, in particular, are very effective at analyzing vast datasets to uncover subtle correlations and patterns that human analysts might miss. For example, AI systems can monitor

minute-by-minute trading data, news sentiment, and macroeconomic indicators to predict market movements. Hedge funds and large financial institutions have been at the pinnacle of this innovation by using AI to develop predictive models that find application in the guidance of trading strategies and risk management.

AI can process unstructured data, including social media postings and news articles, further extending its predictive powers. By analyzing sentiment or detecting trends, AI can spot early signals of market instability. For example, at the start of the COVID-19 pandemic, some AI systems detected jumps in online discussions about respiratory ailments and supply chain disruptions before the markets had fully priced these events. However, where there are significant benefits, there are also substantial risks. Too much reliance on algorithmic models can lead to feedback loops in which AI-driven trades amplify market volatility, rather than dampening it, and make crises worse.

Combining Indicators with AI

Integrating traditional economic indicators with AI-driven analysis forms an exciting concept in predicting black swan events. By taking structured data, such as GDP growth rates or bond yields, as input and adding it to unstructured data, including geopolitical developments and environmental trends, AI systems can build a more holistic understanding of the possibilities and risks. This hybrid approach enables qualitative and quantitative analyses, enabling the models to identify complex interactions that might indicate impending crises. For instance, an AI model may link surging commodity prices to political instability in resource-rich regions and predict supply chain disruptions, further affecting world markets.

Whatever the merits of this approach, there are several challenges. By definition, it is hard and impossible to train on black swan-specific patterns due to their rarity. Lack of historical data prevents algorithms from learning well and may result in too many false positives or missed signals. Also, since the nature of these events cannot be predicted by definition, the best one can hope for from sophisticated models is probabilities, not certainties. To overcome these challenges, continuous improvement is called for in AI models, requiring coordination between data scientists, economists, and policy thinkers to ensure that the predictions are actionable and accurate.

6. Active Examples of AI in use to predict market crashes

6.1 Hedge Funds and Quantitative Artificial Intelligence Models

The first to incorporate AI into the analysis of the financial markets, hedge funds employed sophisticated algorithms to identify inefficiencies and vulnerabilities likely to announce the arrival of a black swan. One of the most significant hedge funds in the world, Bridgewater Associates, stands out as an example of this development. The firm has created AI systems that can simulate complex economic scenarios to help identify possible market disruptions before they arise. Such systems process vast quantities of historical and real-time

data, including market prices, trading volumes, and macroeconomic indicators, to find patterns that may elude human analysts.

One of the most significant advantages of applying artificial intelligence in hedge fund operations is the ability to process and analyze vast amounts of data at a scale and speed far beyond human capabilities. For example, artificial intelligence can identify relationships between seemingly unrelated variables, such as geopolitical events and commodity prices, that could indicate emerging risks. Hedge funds also use machine learning algorithms to run various "what-if" scenarios, allowing them to model the impact of potential shocks on their investment portfolios. This anticipatory strategy not only assists hedge funds in reducing potential losses during market declines but also enables them to seize opportunities that emerge from market dislocations.

Moreover, AI-driven strategies can adapt and evolve in real-time, learning from new data to improve the accuracy of predictions. Such adaptability is critical in dealing with the unpredictability of black swan events, where traditional models tend to fail. While not without limitations, this AI system, by its ability to analyze diverse datasets and generate actionable insights, makes it a tool that no hedge fund can afford to ignore if it seeks to navigate today's complex financial markets.

6.2 Central Bank and Systemic Risk Monitoring

Systemic risk monitoring is the ability of artificial intelligence to monitor and respond to systemic risks inherent in the financial system. For example, institutions like the U.S. Federal Reserve have already begun utilizing AI to analyze vast amounts of economic, market, and regulatory data to identify vulnerabilities that could trigger significant financial instability. Using machine learning algorithms, central banks can detect early warning signs of crises, such as liquidity shortages, asset bubbles, or excessive leverage in specific sectors.

Another area where this AI in central banking is in the domain of stress testing is in the domain of stress testing: AI models can simulate how financial institutions and markets would react under various adverse scenarios. Such stress tests allow regulators to check the economic system's resilience and identify areas needing intervention. For example, AI can analyze the interlinkages of financial institutions and markets in a way that brings to light possible channels of contagion that could amplify the effects of a black swan event.

Artificial intelligence significantly enhances central banks' ability to monitor global economic trends as they unfold. By analyzing news articles, social media posts, and other unstructured or alternative data, AI provides early insights into geopolitical shifts, natural disasters, or health emergencies that might have significant economic repercussions. This real-time analysis allows central banks to respond more quickly to emerging threats and could help blunt the effects of high-impact events.

However, there are a few challenges surrounding the adoption of artificial intelligence in central banking. The AI models' transparency and interpretability are critical, as regulators need to understand the rationale behind the system's predictions for proper decision-making. Besides,

the central banks are also supposed to address data privacy and security, mainly in sensitive financial data contexts. Therefore, the inclusion of artificial intelligence into systemic risk monitoring presents a crucial step toward a stronger financial system.

6.3 Early warning system

The goal of using AI would be to predict Black Swan events before they occur and fortify and perhaps even profit from the situation. “To sum up, as many studies indicate, the ANN has superior performance in predicting financial crises to develop an early warning system. Certain economic variables associated with the financial crises of the 1990s, along with the economic fundamentals derived from financial crisis theories, could explain the threshold of the financial crises. In contrast, in the 2000s, other factors, such as the contagion effect and the (u)nstable foreign exchange regimes, seem to have played a larger role in the occurrence of these crises. This reinforces the view that developing a stable model that could predict or even explain financial crises using ANN or other alternative methods can be challenging but also invaluable for the global economy.” (Aydin and Cavdar, *Prediction of financial crisis with Artificial Neural Network: An empirical analysis on Turkey*).

This study utilizes monthly macroeconomic and financial data from 1990 to 2014 in Turkey and employs an Artificial Neural Network (ANN) to make market predictions. Seven key indicators were analyzed: USD exchange rate, gold prices, BIST 100 Index, wholesale price index (WPI), money supply (MS), domestic debt stock (DDS), and composite leading economic indicators index (CLEI). The ANN model effectively captured historical financial crises in Turkey, including 1994, 2001, and 2008. Predictions for 2014–2016 suggested continued financial risks due to political instability and macroeconomic volatility. These findings indicate that an ANN could predict market crashes. With modern artificial intelligence and by combining qualitative factors (e.g., political events, contagion effects), early warning systems using ANNs can assist traders in making informed economic decisions.

7. Testing

Measurement poses methodological challenges. The methodologies will be tested with predictive measures using all these open-source AI models. This could be achieved by training a model on past data from a series of markets, utilizing both structured economic data (such as interest rates, job rates, and volatility indexes) and unstructured data (including social media sentiment, headlines, and global geopolitical events).

These tests could be carried out through machine learning models like log-periodic power law singularity models, logistic regression, or ANNs to forecast trends before historical black swans. If an existing model can “forecast” past events based solely on pre-crisis data, it should at least possess some predictive capability. These tests do not have to be flawless – to be able to spot periods of heightened danger or hotspots of volatility just before the peak can be of crucial significance to parties like investors and institutions.

Other tests will include live stress testing: feeding live data streams into trained models and seeing how those respond to sudden shocks or anomalies. Iteratively, through repeated reinforcement learning, those models will become ever more adaptive and fine-tune themselves through exposure to market fluctuations to make accurate forecasts. Naturally, an integral part of testing will be identifying limits because most black swans are defined by being unpredictable. It's not clairvoyance, but an early warning system for signs and potential vulnerabilities within markets before crises entirely break out.

8. Testing Methodology

This paper proposes a testing methodology that blends historical backtesting, machine learning model training, and real-time simulation to explore whether AI can predict black swan events. The core idea is to identify early signals that precede market crashes and assess whether AI can spot them before human analysts or mainstream financial tools do.

8.1. Data Collection

An artificial intelligence model has to be trained on relevant data in regards to the task it is trying to achieve. This must be collected and processed in order to ensure safety and efficacy. Structured quantitative data involves historical numerical data on stocks. The research team could use tools such as Yahoo Finance API in order to gain this massive amount of financial data. For quantitative data, which the AI has to also use in order to be able to tackle the problem from a multitude of angles, including word of mouth, and the financial news, constantly updating to keep your portfolio safe. The data could be collected through Google's News API or Web scraping, as has been done in the past with Large Language models. This is combined with the quantitative data, and vast amounts of training could provide the necessary backdrop for the model to perform at an extremely high level, much too fast, and diligent to be matched by humans.

8.2. Model Selection and Training

Train several AI models to analyze patterns and identify volatility precursors:

There are many different types of AI which could be used to complete this goal, my immediate thought jumps to an AI agent, which combines them all through having one larger AI control smaller processes. Artificial Neural Networks (ANNs) which detect non-linear patterns across many variables, would be useful towards achieving the greater goal of prediction. LSTM (Long Short-Term Memory networks): Excellent for time-series forecasting (e.g., predicting volatility spikes). Use supervised learning to train the models on data before and during known black swan events, such as the 2008 crash, COVID-19, or the 1997 Asian crisis, to see if the model can anticipate the drop. The team must seamlessly incorporate quantitative and qualitative analysis. This is the most important factor in the success of the project, whether you can get an AI to be aware of both sides of the financial coin.

8.3. Backtesting

Run the trained models against historical data to see how many crashes they could have predicted. Use time-locked data (such as only information available before the crash) to avoid hindsight bias. Compare model output against actual market movements, measuring the accuracy of predictions, false positive/negative rates, and the lead time (how early the model gave a warning). Once the AI can have a high degree of accuracy, low to no false positives, and a high enough lead time, the team can consider moving on to active use of the model.

8.4. Real-Time Simulation

Set up a live pipeline that streams financial and sentiment data into the trained model to simulate an "early warning system." Use anomaly detection to trigger alerts when the model identifies risk conditions similar to past black swans. Using tools such as Apache Kafka or AWS Kinesis for real-time data streams would be imperative in order to feed the AI a stream of information. Then either publishing the AI for free or otherwise introducing it to the market. This could be very easily done by just running one instance of the AI across multiple markets and publishing results in order to mitigate costs.

Summary of Goal

This methodology doesn't claim to predict every black swan. Instead, it seeks to build a system to detect risk clusters or unstable conditions before they escalate, giving investors and institutions a valuable edge. The goal isn't perfect prediction, but probabilistic foresight.

Conclusion

While black swan events will never be truly predictable, the research presented here demonstrates that artificial intelligence can meaningfully expand the boundaries of human foresight. In short, black swans remain the most unsettling and least predictable events in financial markets. Elusive still, but a marriage of art and science through AI and traditional economic signals is a promising future path to predict and hedge against their impact. While our ability to predict these events with once-and-for-all precision is never likely, infrastructure like neural nets and live sentiment analysis at least provides investors, and policymakers, for that matter, with a fair shot at responding faster and with greater intensity. Ongoing research, collaboration between economists and data scientists, and responsible application of AI will be key to fulfilling this promise as practical early warning systems. As markets continue to become ever more integrated and sophisticated, being able to uncover hidden risks will no longer be a luxury but a requirement. When artificial intelligence is fully integrated into the financial landscape, it will prepare societies for stress. We doubt AI will be prophetic, but a tool just like any other, aiding our ability to react to precursors of financial instability. It is this union of intelligence, both artificial and human, that offers the closest thing to clarity in a world defined by uncertainty.

Works Cited

Aydin, Alev Dilek, and Seyma Caliskan Cavdar. “Prediction of Financial Crisis with Artificial Neural Network: An Empirical Analysis on Turkey.” International Journal of Financial Research, vol. 6, no. 4, 2015, pp. 36–45. ResearchGate.

“Forbes Technology Council.” “Black Swan Events May Be Predictable After All.” Forbes, 1 Apr. 2024, https://www.forbes.com/councils/forbestechcouncil/2024/04/01/black-swan-events-may-be-predictable-after-all/.

Gabor-Granger Pricing Method. B2B International, https://www.b2binternational.com/research/methods/pricing-research/gabor-granger/.

“Investigating the Underlying Factors of the Dot-Com Bubble and Developing Regression Models for Future Market Predictions.” Scientific Research Publishing (SCIRP), 2023, https://www.scirp.org/journal/paperinformation?paperid=127628.

Zhao, Ying, et al. “Artificial Intelligence Management in Financial Crisis.” 2011 International Conference on Information, Communications and Signal Processing (ICICS), IEEE, 2011, https://ieeexplore.ieee.org/abstract/document/6069610.

Zhang, Fan, and Georgy Chabakauri. “Black Swans, Beliefs, and Market Dynamics.” SSRN, 2024, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4734944.

Team, The Investopedia. “Black Swan in the Stock Market: What Is It, with Examples and History.” Investopedia, Investopedia, www.investopedia.com/terms/b/blackswan.asp#toc-black-swans-and-the-housing-market-crash. Accessed 4 Jan. 2025. https://ieeexplore.ieee.org/abstract/document/6069610/references#references

“Black Swan Event.” Corporate Finance Institute, 7 Aug. 2024, corporatefinanceinstitute.com/resources/economics/black-swan-event/.

Bloch, Brian J. “Black Swan Events and Their Impact on Investments.” Investopedia, Investopedia, www.investopedia.com/articles/trading/11/black-swan-events-investing.asp. Accessed 4 Jan. 2025.

Smith, Randall. “After a Dazzling Early Career, a Star Trader Settles Down.” The New York Times, The New York Times, 5 Mar. 2014, archive.nytimes.com/dealbook.nytimes.com/2014/03/05/after-a-dazzling-early-career-a-star-trader-settles-down/.

Author links open overlay panelMin Shu a, et al. “The ‘covid’ Crash of the 2020 U.S. Stock Market.” The North American Journal of Economics and Finance, North-Holland, 18 June 2021, www.sciencedirect.com/science/article/abs/pii/S1062940821001170.

Hayes, Adam. “Ft Wilshire 5000 Index (FTW5000): Definition and What’s Included.” Investopedia, Investopedia, www.investopedia.com/terms/w/wilshire5000equityindex.asp. Accessed 4 Jan. 2025.

"Dotcom Bubble." Corporate Finance Institute, 7 Dec. 2023, corporatefinanceinstitute.com/resources/career-map/sell-side/capital-markets/dotcom-bubble/.

Aydin, Alev Dilek, and Seyma Caliskan Cavdar. "Prediction of financial crisis with Artificial Neural Network: An empirical analysis on Turkey." International Journal of Financial Research, vol. 6, no. 4, 20 Aug. 2015, https://doi.org/10.5430/ijfr.v6n4p36.

Nyman, Rickard, and Paul Ormerod. "Predicting Economic Recessions Using Machine Learning Algorithms." arXiv.Org, 3 Jan. 2017, arxiv.org/abs/1701.01428?utm_source=chatgpt.com.

Sound and Power in Ancient Rome By Cyrus Di

Abstract

In ancient Rome, power was not only seen—it was heard. This article explores how brass and wind instruments transformed sound into a language of authority and how each unique instrument fit into place at each procession, whether it was the *tuba*, *cornu*, *lituus*, or the *bucina.* Drawing on several literary sources, it argues that musical instruments made triumphal processions, made military coordination, and made public ritual be what they were in ancient Rome. Their resonance marked hierarchy, divine favor, and imperial legitimacy, converting individual victories into collective spectacles of Roman dominance. Tracing their evolution from sacred ritual tools in the Kingdom to standardized military signals in the Republic and instruments of imperial display under the Empire, this study demonstrates how organized sound became embedded within Rome's political imagination. The power of the Romans was not just displayed across monuments and ceremonies. It was audibly staged and it was experienced by all of Rome.

In ancient Rome, public ceremonies had more than just one function. They were celebrations and exhibitions of political authority and the Roman identity. Among these ceremonies, the procession after a triumph was one of the most elaborate and symbolic occasions. It combines military celebration with religion to not only reaffirm the Roman power, but also the legitimacy of its own leaders. Music was central to the success and impact of these spectacles, where brass instruments were used very often. Not only were they sources of sound, but the instruments also carried symbolic meanings that extended throughout Roman society. They regulated the movements of armies, sanctified religious practices, and made public processions much more grand, transforming individual military victories into an affirmation of Rome's dominance.

The presence of musical instruments during these celebrations created a better sense of hierarchy, discipline, and order, ensuring that Rome's power was both seen and heard by the Roman citizens and barbarians alike.[1] The roles of these instruments evolved alongside Rome's political transformation over time. In the Kingdom period, they were used very often in religion; in the Republic, they became standard in the military; and under the Empire, they were used to create spectacles and as propaganda. This gradual shift in function reflects the expansion of Rome from a small religious monarchy to a large empire, demonstrating how music and sound were adapted to serve the changing structure of Rome's power and leadership.

Musical instruments in ancient Rome thus functioned as instruments of power, operating under ritual practice, military organization, and political ideology; through their use in triumphal ceremonies, public rituals, and the army and navy, brass and wind instruments transformed sound into a means of reinforcing authority, legitimizing rule, and expressing Rome's evolving conception of order and religion across different several historical periods.

[1] During a vast majority of Rome's history, a great number of Romans referred to outside tribes and cultures as "barbarians," or in the original Latin, *barbarus.*

Scholarly discussions of music and instruments in ancient Rome have emphasized their roles in ritual symbolism, military organization, and public display. Several great contributions to this study are available and seen across the internet and the world, among them is the work of Cristina-Georgeta Alexandrescu, whose analysis of Roman iconography demonstrates how ancient wind instruments—including tubae, cornua, bucinae, and litui—functioned as both a visual and a symbolic marker in the contexts of rituals and triumphs (33–46). By examining several different works of art, Alexandrescu argues that musicians and instruments were incorporated into ceremonial scenes with the intention of conveying order, hierarchy, and religion. Her work establishes a foundation for understanding ancient Roman instruments as not simply performative tools but as symbolic elements cut into the great empire (Alexandrescu 33–46).

Adding on, Maria E. Ginsberg-Klar provides archaeological evidence for the practical use of musical instruments in the Roman military (313–320). After studying several parts of instruments found in the ruins of ancient forts and military camps, Ginsberg-Klar explains that brass instruments were essential for a great number of the basics within the Roman army, which include signaling, coordination, and discipline. Her research allowed symbolic interpretations to be grounded in material culture so that it is confirmed that the prominence of instruments in art reflected their real importance in the Roman military (Ginsberg-Klar 313–320).

Furthermore, Herbert Antcliffe, who examines the broader significance of music in Roman society, clarifies the ceremonial framework in which the instruments were operated (337–340). Antcliffe sets musical performance inside triumphs and public ceremonies, emphasizing how music created a greater spectacle and made the Ancient Roman civic identity stronger. His work helps make the triumph seem like a regulated ritual that integrated sound and movement and display in order to transform military victory into an affirmation of the power that the Romans held (Antcliffe 337–340).

While Antcliffe and Alexandrescu focused on the ceremonial meaning behind the instruments, Renato Meucci gave a technical and acoustic perspective by analyzing the construction and sound properties of ancient brass instruments (Alexandrescu 40; Antcliffe 340; Meucci 85–97). Meucci's research highlights how the design of the ancient tuba produced a bright, penetrating sound suited for outdoor use, making them effective for not only battlefield communication, but also larger public ceremonies. This analysis further explains why certain instruments became both a military and a ceremonial standard, linking the physical design of it to an actual, purposeful function (Meucci 85–97).

Finally, Jan Ziolkowski tracked the symbolic associations with the historical development of Roman trumpets, noting the tuba in particular (1–22). Ziolkowski situated these instruments within a broad Mediterranean tradition whilst emphasizing their notable adaptations when they became Roman. His work further explained how the trumpet became associated with authority, temporal regulation, and command, reinforcing its militaristic and ceremonial power (Ziolkowski 1–22).

Taken together, these writers demonstrated that Roman musical instruments operated under sound, symbolism, and power. Alexandrescu provided the visual framework; Ginsberg-Klar validated answers with archaeological evidence; Antcliffe contextualized the ceremonies; Meucci explained the function of acoustics; and Ziolkowski tracked the historical development of the instruments (Alexandrescu 40; Ginsberg–Klar 315; Meucci 90; Ziolkowski 5). This combined scholarship reveals how music in ancient Rome was carefully structured to serve political authority, religious meaning, and military efficiency.

Triumphal Ceremonies in Ancient Rome

A triumphal ceremony in ancient Rome was a public event that celebrated the success of their military.[2] These ceremonies were political and religious in order to honor both the general and the gods. The triumph began when the Senate granted the victorious general the right to hold a procession through the city, a privilege purely reserved for significant victories that had expanded Roman territory or had secured Rome's power (Antcliffe 338). The Roman Senate was its governing council of aristocrats, which advised magistrates and directed state policy, especially in matters of war and finance (Heyman 1–44).

The triumphal procession followed one route. It began outside the city and traversed through the city until it reached the Temple of Jupiter Optimus Maximus at Capitoline Hill. The general rode in a chariot, wearing a purple toga and crowned with a laurel wreath (Antcliffe 338). His troops marched and sang behind him with not only songs of praise, but also humor about their leader. showcasing the casualness that appears even at a serious occasion like this.

Key elements of the parade included displays of the spoils of war paraded for the citizens to see. Prisoners of war were often displayed in chains to underscore Rome's dominance. Many of these enemies included high-ranking enemy leaders. Music was very important for this event, with several military instruments marking the rhythm and grandeur of the event (Alexandrescu 40; see Table 1). These instruments were used in battle and signaled ceremonies. Musicians "playing their instrument (tubae, cornua, tibiae)" were often seen in celebrations, showcasing their use in religion and the military (Alexandrescu 40; Meucci 90).

[2] The ceremonies typically celebrated a Roman triumph after a series of battles. Cassius Dio recorded that after the Dacian Wars, Trajan celebrated his victory with 123 days of public games, including gladiatorial combats and beast hunts (Cassius Dio 68.15, epit. Xiphilinus).

Music Instrument	Description of the instrument	Use during Battles	Use during Ceremonies	Examples
Tuba	Straight brass trumpet, loud and sharp.	Signals orders, attacks, and retreats.	Adds grandeur and marks rituals in triumphs.	Column of Trajan; sacrificial rites.
Cornu	G-shaped horn with deep tone.	Legion signals and camp communication.	Played near generals/standards for order.	Arch of Titus.
Lituus	Curved trumpet, sacred to augury.	Rarely battle use; symbolic authority.	Linked victories to divine approval.	Triumph art; augurs' rituals.
Bucina	Compact curved trumpet, resonant.	Guard changes, camp routines, naval signals.	Symbolized time, order, continuity.	Roman galleys.

Table 1: Instruments' Types and Functions.

The celebration began when the general offered sacrifices to the gods, thanking them for victory or other forms of success at the Temple of Jupiter. A sacrificial ritual often occurred involving the offering of a pig, sheep, and a bull (Rives 47). In fact, "emperors in particular were often shown in the act of sacrificing, as in this relief from a now-lost [triumphal] arch in Rome honoring the emperor Marcus Aurelius (161-80 CE)" (Rives 47). This sacrifice was a powerful act of thanksgiving and purification, connecting military success to divine favor (Heyman 3–6). After the sacrifice, visual and auditory elements would begin (Antcliffe 337–344). This created a theatrical display meant to impress both the Roman populace and foreign envoys present in the city.

Additionally, Military instruments played an important and symbolic role in these types of displays. The Roman tuba "had a bright and strident tone… well suited to the open air and public spectacle," highlighting its effectiveness on both the battlefield and display (Meucci 85). Similarly, archaeological evidence shows that "cornua and tubae are found throughout Roman military sites… reflecting their ceremonial and functional role" (Ginsberg-Klar 317). The presence of instruments such as the lituus, cornu, and tuba emphasized their "ritualized communication with the gods," demonstrating how these instruments linked military victory with religion (Alexandrescu 44; see Table 1).

The Roman triumph was not just a military celebration. It was a planned occasion that put together political authority, religion, and performance. The Roman state completely changed victory from an individual standpoint to a collective assertion of Roman power and divine favor by controlling who would hold the triumph and the route of the celebration, while including sacrifice, performance, and display. These ceremonies reinforced Rome's social hierarchy and values, creating a ceremonial framework in which music and instruments could later function as powerful symbols of order and authority.

The Roles of the Instruments in Triumphal Ceremonies

During Roman triumphal ceremonies, music was one of the most important elements of the exhibition. It was used to signal divine favor, assert military strength, and enhance the aura around state power. The instruments were selected for not just their acoustic power but also for the symbols they held of Roman identity, authority, and tradition (see Table 1). For example, the tuba was a very recognizable trumpet used in the Roman military. It was a straight instrument with a narrow cylindrical bore and a flaring bell (see Fig. 1). Being made of brass along with several other reasons, it was capable of producing a loud, commanding tone that was ideal for outdoor ceremonies. As Meucci explains, "The tuba had a bright and strident tone... well suited to the open air and public spectacle" (Meucci 85). Its role in triumphal processions came from its use on the battlefield, where it regulated both troop movement and conveyed orders (Meucci 89).

Fig. 1: Mosaic from the Villa Romana del Casale, Piazza Armerina, Sicily. Photograph by Paulo S, *Flickr*, www.flickr.com/photos/paolosub/376837078/.

The cornu, shaped in a broad G-form and worn over the shoulder using a support bar, was also a standard in triumphal contexts (see Fig. 2). It too originated in military life and was used for general signaling within legions. According to Ginsberg-Klar, "cornua and tubae are found throughout Roman military sites in Germany, often near principia, reflecting their ceremonial and functional role" (Ginsberg-Klar 317). These instruments have been recovered from contexts in not only triumphs, but also funerals. This suggests that they were seen as markers of Roman presence across several provinces (Ginsberg-Klar 319). The cornicines marched towards the front or near important figures such as the general or standard-bearers.

Fig. 2: Late Roman mosaic depicting a musician performing on the cornu. Image from Encyclopaedia Romana, University of Chicago, https://penelope.uchicago.edu/encyclopaedia_romana/gladiators/nennig.html.

The lituus, a rarer curved trumpet, held a different connotation (see Fig. 3). As a sacred instrument linked to augury and religious officials, the lituus helped bridge military success and divine approval. According to Alexandrescu, "Its association with augurs and the ritualized communication with the gods added a layer of divine legitimacy to public events like the triumph" (Alexandrescu 44). The lituus appeared frequently in iconography associated with triumphs, reinforcing its ritual purpose. Its inclusion in the procession sent a strong message: Rome's victories were not merely tactical but cosmically sanctioned (Alexandrescu 38–41).

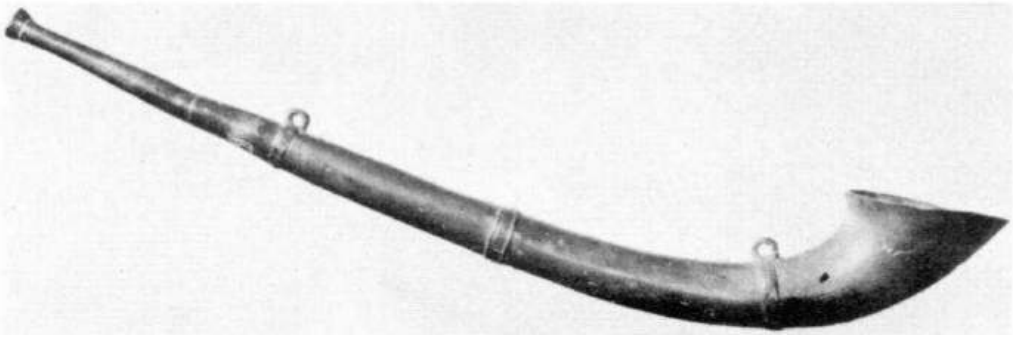


Fig. 3: Lituus found near Düsseldorf in the Rhine. Reproduced from Rodney Martin Cross, "Bold as Brass: 'Brass Instruments' in the Roman Army," Macquarie Matrix: Special Edition, ACUR 2013 (2013), fig. 3.

The bucina had a different shape and function than the cornu, which it is often confused with (see Figs. 3 and 4). It was originally used to announce daily routines in military camps, but eventually became adapted for ceremonial use as well. As described by Ziolkowski, the bucina "retained a more practical tone and was associated with the passage of time and continuity of order" (Ziolkowski 13). The connection between trumpet signals and the Roman sense of time was deeply embedded in discipline and ceremonies (Ziolkowski 15). The idea that Roman dominance extended not only to the battlefield but also to temporal and civic structure was reminded to the people by being embedded into their triumphs.

Fig. 4: A monumental Roman sarcophagus discovered in 1621 and named for Ludovico Ludovisi. The Roman Bucina can be seen on the top right of the image. From *Judaism and Rome: Re-thinking Judaism's Encounter with the Roman Empire*, last modified August 31, 2018, https://www.judaism-and-rome.org/ludovisi-sarcophagus.

Importantly, these instruments were not chosen simply for their loudness but for their deep symbolic and cultural resonance. As Ginsberg-Klar notes, "the cornu and tuba were not just musical devices but declarations of Roman presence and power" (319). Their metallic construction often made them visually beautiful. On architectural works, including the Arch of Titus and the Column of Trajan, they are shown exhibiting their major role in forming both a visual and an auditory spectacle (38–41). The image of these instruments etched into monuments brings their sound and the concepts of conquest and victory together.

The sounds of these instruments also had a ritual impact. According to Landels, "The sheer volume and clarity of the Roman brass instruments were believed to carry prayers and affirmations skyward to the gods, particularly during sacrificial segments of the triumph" (qtd. In

Ginsberg-Klar 28). Their position in the ceremony often mirrored a ritual: occasionally appearing before sacrificial animals or near the chariot of the triumphator. Sacrificial acts were enhanced with music, creating a connection between the human celebrations and the gods (Alexandrescu 35).

These instruments appear in not only official triumphs, but also theater and funerals. This fact showcases their capability to adapt to both celebratory and solemn occasions. In the Roman mime, wind instruments were used to highlight power, urgency, and emotion (Wootton 279). Military instruments also appeared in funeral processions, as noted in Josephus' account of Roman funerals and in Roman visual culture, suggesting that their sounds marked not only victories but also communal transitions (Alexandrescu 34).

Musical instruments in Roman triumphal ceremonies functioned as tools and symbols. Instruments shaped the sonics of the triumph while reinforcing military authority, religion, and civility. Their sounds structured the procession, enhanced sacrifice, and emphasized conquest in monumental art. By uniting the military and religion, these instruments changed sound into a medium through which Roman power and favor from the gods were affirmed by the public.

Reasons for holding a Triumphal Ceremony in Rome

Roman triumphal ceremonies also served broader political and ideological purposes. These events were very useful to the state, being used by emperors and magistrates to consolidate power.[3] Though often seen as a result of military success, many triumphs operated as a symbol and a spectacle. According to Alexandrescu, even in visual representations, triumphs emphasized symbolic authority, not merely brute force, and often portrayed the instruments and processions as expressions of cosmic order and hierarchy (Alexandrescu 35).

One of the most important contexts in which triumph and ceremony appeared outside of the military was the imperial cult.[4] Triumph-like ceremonies appeared on days of importance, such as imperial birthdays or anniversaries, or transitions of power. These events helped maintain the succession of emperors and continuity of Rome's politics. As Heyman explains, the Roman state "used sacrifice to legitimize imperial authority and to symbolize the subjugation of others" (46). Even when actual military achievements were absent, the ceremonial forms reinforced the emperor's divine and political supremacy.

Such ceremonies also played an important role in celebrating restored order. In the face of disaster, Rome often turned to religious processions and sacrifices to reestablish the pax deorum.[5] These actions were both private or devotional and large public spectacles with deep implications of politics. As Heyman notes, Roman religious discourse was rooted in sacrifice, politics, and sacred violence, tracing back to Rome's mythic origins: "Romulus marked out the sacred space of the new city... and instructed Celer to kill anyone who would cross the furrow"

[3] Magistrates were elected public officials in ancient Rome who held executive, judicial, or administrative authority. They ranged from high-ranking positions like consuls and praetors to lower offices like quaestors and aediles, and their power was defined by law and tradition.

[4] Emperors were honored or even deified at imperial cults

[5] Directly translates to "peace of the gods," which was believed essential for the city's survival.

(Heyman 1). This foundational story established a pattern in which ritual killing and boundary setting were entwined with statecraft.

Triumphs also supported transitions in leadership, such as the inauguration of a new consul or emperor. Even when these events were not formal, they were very similar to other triumphs. The visual and auditory elements offered a structure of state power and divinity. According to Meucci, instruments like the tuba and cornu were incorporated into public rituals because of their associations with command and coordination (Meucci 85–89; see Table 1). Their presence in ceremonies reinforced order and discipline, extending military symbolism into the city.

These ceremonies also completed commemorative cases. Processions honored religion and were often dramatized through theatrical performances and reenacted battles. These symbolic spectacles did not simply aim to entertain, but connected the current regime to Rome's legendary past. As Wootton argues, Roman mime and performance often borrowed the structure of a triumph to showcase emotions and morals (Wootton 279). In these events, the past was used to make the present seem more legitimate, and the splendor of Roman origins was placed analogous with contemporary leadership.

Roman triumphal ceremonies were not limited to celebrating the military. Their forms, symbols, and rituals were tools capable of messaging politics. religion, and civic education. Whether celebrating a battlefield victory or a government officer's accession, these occasions offered a powerful way of highlighting Rome's eternal domination over enemies and time.

Evolution of Roman Ceremonial Instruments

The usage of musical instruments in Roman ceremonies significantly evolved from the Kingdom, to the Republic, and then to the Empire. This reflected broader transformations in the military, symbols, and culture. While core instruments remained central, their functions and symbolic meanings diversified over time (see Table 1). These changes corresponded with Rome's shifts from a small monarchy to a massive empire, and how music became ever more embedded into the empire's image of power and identity, since Roman ceremonies were "meant to manifest authority through orchestrated ceremonial display" (112).

In the early Kingdom period, Roman musical traditions were heavily influenced by Etruscan and other Italic practices.[6] The lituus was among the earliest used and was closely associated with augury (see Table 1).[7] Its use reflected the dominance of religion in early state functions, especially as Rome's earliest ceremonies emphasized boundary-setting, purification, and divine communication. As Alexandrescu explains, the lituus was closely associated with augury and served as an instrument of "ritualized communication with the gods" rather than a tool of battlefield signaling (Alexandrescu 44). Thus, music in this early phase reinforced the sacred legitimacy of Rome's kings and their divinely guided authority.

[6] Since the first few kings of Rome were considered legendary by some, the exact beginning of the Kingdom is debatable. However, according to Varro, the founding of the city was in 753 BCE. The Kingdom ended at the specific year of 510 BCE since the last kings of Rome were very evident to have existed (Hardy).

[7] In Ancient Rome, the augurs were Roman priests who read omens based on the flight patterns of birds.

During the Republic, music in Roman life became much more militarized, reflecting Rome's expansionism and the massive power of its army.[8] Brass instruments were standardized for battlefield use. These instruments had powerful acoustics that enabled them to carry sound across vast distances, allowing commanders to issue signals across legions (see Table 1). Ginsberg-Klar notes that "military camps in Germany regularly reveal cornu and tuba fragments near headquarters buildings, implying organized musical signaling for drills and ceremonies" (317). Their function was to communicate during combat and for psychological effect, bringing fear into enemy ranks and pride among the Roman troops.

As the Republic shifted to the Empire, instruments became more prominent as an element of ceremonies.[9] Triumphal processions and celebrations showcased several musicians, often marching in rank in order to exhibit the order and discipline of the Roman people. These events brought the instruments beyond their original function on the battlefield. The bucina came to appear in public processions, where its visual and auditory symbolism communicated imperial prestige and the rhythm of Roman order (see Table 1). Instruments thus moved from the margins of war camps to the center of Rome's urban ritual life.

The early Empire also witnessed the introduction and adaptation of foreign instruments, especially from the Greek East. The *hydraulis* was introduced to Rome during the first century CE.[10] This water-powered keyboard instrument produced a beautiful, multi-tonal sound that was used in multiple theaters and banquets. Galpin describes the *hydraulis* as a technological marvel that "brought a new sonority into Roman spectacles, combining luxury and innovation" (Galpin 358). Its presence marked a shift from purely martial signaling to musical performance as a form of elite display and civic entertainment.

Until the end of the Western Roman Empire, the role of instruments became even more institutionalized.[11] Reliefs on imperial monuments depict tubicines and cornicines in structured groups, standing near generals and standard-bearers during triumphs and processions (see Table 1). This reflects how music was put into the concept of Roman victory and conquest. As Alexandrescu notes, musicians depicted with instruments frequently appeared in ceremonial scenes, where "[they marked] solemnity and structured order," visually reinforcing the authority and identity of the state (Alexandrescu 200; see Table 1). Their use in sculpture suggests that their meaning was understood by many citizens and enforced the emphasis on order, control, and harmony in the Empire.

---

[8] During the Republic, Rome would elect two consuls that would serve for one year at a time. This was to ensure that no one person will have too much power for too long of a time. However, because the consuls only had a year to take action, most decided to go on a conquest to try and garner fame and more power in order to be reelected in the future.

[9] The Roman Republic ended at around 23 BCE, when Octavian received "he titles and powers of Imperium *proconsulare maius* and *tribunicia potestas* for life" and became known as Augustus, which in turn, also started the Empire of Rome (Hardy).

[10] Directly translates to "water organ" in English.

[11] The Roman empire eventually split into two parts in 395 CE after the death of Theodosius I: the East and the West Roman empires. The Eastern Roman Empire eventually became the Byzantine empire, while the Western Empire fell in 476 CE (Hardy).

In sum, the evolution of musical instruments in Roman ceremonies mirrored the transformation of its society. The instruments did not simply produce sound, but instead were given meanings, power, symbolism in religion and the Roman politics of each era. As Rome grew from a city-state to a vast empire, its music too transformed from the almost silent prayers of priests to the fanfare and cheers of the Roman people.

The Role of Instruments in the Army and Navy

Instruments were very important to the Roman army and navy, acting as ways to make command and coordination easier, whilst becoming a symbol of the army itself. Though the same core instruments were used in both branches, their functions varied (see Table 1). These instruments were essential for communication, often over the clamor and distance during a sea or land battle. These instruments were chosen for their design since Roman brass instruments were designed for audibility across large spaces and noisy environments (Ginsberg-Klar 27; Meucci 85).

In the Roman army, instruments were key to battlefield control and display. The tuba was used to give orders, announce the beginning of marches, and signal attacks and retreats. It produced a "bright and strident tone… well suited to the open air and public spectacle," allowing commands to be carried effectively, even across long distances (Meucci 85). The cornu was often used for broader organizational commands. These instruments were standard in Roman legions, with musicians often being attached to a cohort.[12] Their use was highly codified; even changes of watch or daily routines in camp were announced with specific musical signals (Ginsberg-Klar 279; Wootton 276; see Table 1).

Military instruments also had a symbolic meaning in both army and public ceremonies. For instance, musicians often led funeral and triumphal processions. In depictions of triumphs, figures with horns and trumpets very often accompany the general, exhibiting order and discipline. As Alexandrescu notes, "Figures of musicians playing their instruments accompany ritual scenes and processions," showing that the musical elements functioned as symbols of structured Roman authority (Alexandrescu 200). These visual depictions also reinforced the soldiers' positions within Rome's civic and religious hierarchy, as musicians and instruments in ritual and triumphal imagery "mark solemnity and structured order" (Alexandrescu 200).

In the Roman navy, instruments performed a similarly vital role but under different environmental constraints. Naval musicians had to contend with the sound of waves, shouting, and wind, which required strategic placement of musicians aboard ships. This was often in elevated positions in order to ensure that their signals were audible. The bucina was specifically favored in the navy because of its portability and resonance. It was used for coordinating rowing and maneuvers and signaling combat readiness (Ziolkowski 13). Ginsberg-Klar explains that musical instruments were necessary for timing and coordination aboard Roman ships, helping to maintain synchronization among rowers, sailors, and commanders (Ginsberg-Klar 28).

[12] The musicians for each of the Roman instruments had a respective name. The tuba players were called *tubicines*, the cornu players were called *cornicines*, and the buccina players were called *buccinatores*.

However, the navy used music more as a tool than a spectacle in their practices. While army musicians were frequently depicted in monuments, naval musicians rarely appear in Roman art or literature. This reflects that the army was the primary symbol of power and expansion in the classical world. This may explain why its ceremonies were more heavily ritualized and publicized compared to the navy's. On the other side of the spectrum, the navy lacked the same ceremonial status that the army did, making its musical usage largely utilitarian (Wootton 278–280).

Despite these differences, both branches used instruments to reinforce hierarchy and cohesion. Orders transmitted via musical signal ensured clarity and prevented confusion. Instruments also imposed rhythm and discipline, whether for marching legions or rowing crews. In both cases, music also reinforced hierarchy and discipline within Roman forces; as Alexandrescu notes, musicians in ceremonial contexts 'mark solemnity and structured order', reminding participants of their place within Rome's military and civic hierarchy (Alexandrescu 200).

There were also instruments with dual associations, like the lituus, originally used in religious rituals by augurs but later adopted into military signaling. Its appearance in both sacred and military contexts suggests a broader Roman understanding of music as a medium that connected military order with divine authority, since the lituus functioned as an instrument of 'ritualized communication with the gods' and brass instruments more broadly signaled Roman presence and command across military contexts (Alexandrescu 44; Ginsberg-Klar 317). The use of such instruments in both the army and navy underscores how music transcended practical communication to reinforce Rome's cosmological and political worldview.

In sum, instruments were useful to the functioning of the Roman military, becoming tools for commanding and coordinating, while also becoming a symbol of the Roman military. While their use as a tool ensured discipline and efficiency in demanding environments, their presence in ceremonies enforced hierarchy, order, and divine approval. The differences of celebrations within the army and naval music reflects broader values of the people. In both contexts, however, sound still functioned as a uniting force that brought military organization to the world's view of the Roman state.

Conclusion

Musical instruments were integral to the political, religious, and military life of classical Rome, functioning as much more than just a simple soundmaking object. Instruments such as the tuba, cornu, lituus, and bucina built up the public spectacle, made hierarchy and discipline evident, and became widely known audible symbols of authority and favor from the gods. In triumphs, their sonics made military victory a shared public enjoyment, while in the military, they enabled coordination and cohesion across vast and demanding environments.

The use of the instruments changed as time flowed in the world, where the roles and meanings of these tools evolved alongside Rome's ever expanding power. Early ritual associations gradually gave way to standardized military signaling, then to key pieces of a public

spectacle. Whether employed in divine, battlefield, or public communication, Roman instruments consistently reflected the state's emphasis on control, legitimacy, and order.

The study of Roman musical instruments therefore reveals how sound itself became a medium of ideology. By embedding music into its ceremonies, institutions, and visual culture, Rome ensured that authority was not only displayed but audibly experienced, allowing organized sound to function as a lasting expression of political power and cosmological belief.

Works Cited

Alexandrescu, Cristina-Georgeta. "The Iconography of Wind Instruments in Ancient Rome: Cornu, Bucina, Tuba, and Lituus." Music in Art, vol. 32, nos. 1–2, Spring–Fall 2007, pp. 33–46. JSTOR, www.jstor.org/stable/41818803.

Antcliffe, Herbert. "What Music Meant to the Romans." Music & Letters, vol. 30, no. 4, Oct. 1949, pp. 337–44. JSTOR, www.jstor.org/stable/730675.

Centre of Excellence. "The Legacy of the 7 Ancient Kings of Rome." Centre of Excellence, 10 Oct. 2023, www.centreofexcellence.com/the-7-kings-of-rome/.

Cross, Rodney Martin. "Bold as Brass: 'Brass Instruments' in the Roman Army." Macquarie Matrix: Special Edition, ACUR 2013, 2013, www.academia.edu/figures/3407412/figure-3-lituus-found-near-dusseldorf-in-the-rhine-photo.

Galpin, F. W. "The Water Organ—A Roman Keyboard Instrument." Scientific American, vol. 91, no. 21, 1904, pp. 358–60. JSTOR, www.jstor.org/stable/24992848.

Ginsberg-Klar, Maria E. "The Archaeology of Musical Instruments in Germany during the Roman Period." World Archaeology, vol. 12, no. 3, 1981, pp. 313–20. JSTOR, www.jstor.org/stable/124243.

Hardy, J. "The Complete Roman Empire Timeline: Dates of Battles, Emperors, and Events." History Cooperative, 20 Aug. 2024, historycooperative.org/roman-empire-timeline/.

Heyman, Gail. "Roman Religion and Sacrificial Practice." The Power of Sacrifice: Roman and Christian Discourses in Conflict, Catholic University of America Press, 2007, pp. 1–44. JSTOR, doi.org/10.2307/j.ctt284wv1.5.

"Ludovisi Sarcophagus." Judaism and Rome: Re-thinking Judaism's Encounter with the Roman Empire, last modified 31 Aug. 2018, www.judaism-and-rome.org/ludovisi-sarcophagus.

Meucci, Renato. "Roman Military Instruments and the Lituus." The Galpin Society Journal, vol. 42, Aug. 1989, pp. 85–97. JSTOR, www.jstor.org/stable/842625.

Paolo S. Mosaic Showing the Roman Tuba and Its Size in Relation to Its Player, ca. Fourth Century CE. Photograph, Villa Romana del Casale, Piazza Armerina, Sicily. Flickr, www.flickr.com/photos/paolosub/376837078/.

Rives, James B. Religion in the Roman Empire. Blackwell Publishing, 2007.

University of Chicago. "The Roman Mosaic at Nennig." Encyclopaedia Romana, penelope.uchicago.edu/encyclopaedia_romana/gladiators/nennig.html.

Wootton, G. E. M. "Representations of Musicians in the Roman Mime." Mediterranean Archaeology, vol. 17, 2004, pp. 243–52. JSTOR, www.jstor.org/stable/24668157.

Ziolkowski, Jan M. "Toward an Edition of the Latin Carmen." Classical Philology, vol. 94, no. 1, 1999, pp. 1–22. JSTOR, doi.org/10.2307/4352289.

Cultural Negotiation in Dance Beyond the Binary of Appropriation and Appreciation
By Jeneve Lee

Abstract

Dance, as both an embodied practice and a cultural expression, sits at the center of contemporary debates on the movement of culture across contexts. While notions of cultural appropriation and appreciation have become central to these debates, the rigid binary through which they are frequently framed obscures the complex ways in which cultural exchange unfolds. This paper argues that cultural negotiation offers an alternative lens for understanding the movement of dance across cultures as a spectrum rather than a dichotomy, foregrounding the contingent, embodied, and relational dimensions of intercultural engagement.

Drawing on literature that views culture as a site of both resistance and incorporation, the paper advances the concept of cultural negotiation as a more analytically fruitful framework than the appropriation-appreciation binary. Through empirical examples of white participation in hip hop and the incorporation of Afrobeats into K-pop, the article demonstrates how performances function as sites of ongoing ethical negotiation in which questions of power, ownership, and identity are choreographed through movement itself.

Rather than asking whether a performance simply appropriates or appreciates, this paper examines how dance practices are shaped by cultural literacy, positionality, accountability, and historical power relations. As an embodied and relational art form, dance makes visible the messy realities of cultural exchange, revealing how responsibility is enacted – or deferred – through practice. Ultimately, the paper proposes a shift away from policing authenticity toward ethical dialogue, arguing that informed intercultural exchange is essential for navigating cultural circulation in an increasingly globalized performance landscape.

Introduction

Dance, as both an embodied practice and cultural expression, sits at the center of contemporary debates on the movement of culture across contexts. Concepts such as cultural appropriation and appreciation have become central to the discourse on how individuals and societies engage with media that originate outside their own cultural backgrounds, including in the realm of dance. Yet, this debate is often framed as a binary: one either appropriates, perpetuating colonial narratives and structures of dominance, or appreciates, showing respect and engagement with cultural differences. This binary, while a rhetorically legitimate metric, limits a far more diverse spectrum of intercultural syncretism by overlooking the nuances on which cultural interaction often unfolds. As a fluid, embodied practice with inconspicuous boundaries, dance renders visible the spectrum of cultural negotiation that unfolds between appropriation and appreciation rather than conforming to a rigid binary.

This paper suggests that the concept of cultural negotiation offers a more nuanced approach to understanding these interactions than the dichotomy between cultural appropriation and appreciation. Rather than grounding judgments solely in questions of whether something is

so-called right or wrong, cultural negotiation acknowledges the "messy, [boundless] realities of engagement" through which inherent power dynamics, ownership, and identity are especially reconfigured in globalized art forms (Johnson 2020). Dance is a particularly interesting site for this inquiry as it operates through embodied gestures and collective performances, making visible the very exchanges that underpin cultural movement and change. Through dance, bodies become a vessel for messages, going beyond aesthetics by navigating histories of marginalization, resistance, and transformation.

This study draws on Stuart Hall's theorization of culture as a "site of both resistance and incorporation" and meaning-making (Hall 1993), Chatterjee's principles on ethics of collaboration and "informed intercultural exchange" (Chatterjee 2020), and Johnson's analysis of dance as a tool for cultural storytelling and political resistance (Johnson 2020). By exploring the global trajectories of hip hop and Afrobeats – both originating in the African diaspora yet now practiced across the Global North and South – this paper shows that performances tap into and shed light on ongoing ethical discussions about who owns culture, who is permitted to speak and move within it, and what responsibilities are involved in participating in cultural forms that emerged from histories of struggle, protest, and pride. Empirical examples of white participation in hip hop and the uptake of Afrobeats within K-pop illustrate not just the movement of aesthetics, but also the movement of meaning, national or regional soft power, and rebellion to traditional White chronicles, allowing us to reflect on how "culture" itself is choreographed through negotiations between appropriation and appreciation, as well as the post-colonial complexities of today's movement of culture.

Ultimately, this approach suggests that replacing rigid binaries with a spectrum of informed intercultural exchange is critical if we are to ethically navigate contemporary cultural flows. As dance and its culture of collaboration and fusion continuously revises the boundaries of belonging and ethics in "sharing," it calls for a shift from policing authenticity – a rather unrealistic and rigid rubric for measuring thoughtful participation in cultural exchange – towards facilitating dialogue and developing principles that underpin ethical collaboration. While focused on dance, the paper argues that these critical values apply to any globalized cultural practice.

1. Theoretical Framework: From Appropriation Versus Appreciation to Cultural Negotiation

The debates around cultural appropriation are deeply entrenched in questions of "purity", authenticity, ownership, hybridity, and power. Early frameworks often assumed the possibility of "cultural purity," positing clear boundaries between cultures and measuring engagement by degrees of authenticity. Stuart Hall (1993), however, powerfully challenges the very notion of cultural purity, suggesting that "[t]here are no pure forms at all," as all culture is inherently hybrid, a product of ongoing exchanges and borrowings rather than fixed essences.

Noting that "Blackness as a sign is never enough," Hall (1993) stresses that Black identity is not monolithic; it cannot be reduced to a single representation or experience, but is instead composed of diverse experiences shaped by specific historical, social, and political

contexts. Identities of different diasporic Black communities, for example, are continuously shaped by their histories and interactions within new cultural environments (Adewumi 2018). Androutsopoulos and Scholz (2003) similarly speak of cultural "transculturation," demonstrating how hip-hop has reinvented itself along different trajectories in different European contexts as the genre came into contact with distinct local languages, traditional sounds, and regional issues. In such a light, to insist on essentialist or absolute definitions of authenticity would be to misrepresent the reality of cultural formation, which is always fluid and never reducible to simplistic binaries, such as black versus white or authentic versus inauthentic. This calls for a more nuanced spectrum and a conceptual lens that embraces diversity rather than forcing individuals into predetermined categories, furthering the idea that cultural forms are constantly negotiated rather than solidified.

Alongside the notion of purity, scholarship on cultural appropriation has also extensively studied its relationship with power. According to James Young (2005), there is an indisputable interplay of power and ownership, specifically when culturally dominant groups adopt the practices or artifacts of marginalized groups without meaningful understanding or engagement. In these circumstances, the former is seen to convey an implicit sense of ownership, where not only marginalized creators are erased or sidelined, but dominant participants may also escape the struggles or stigmas attached to these practices, profiting from what is not theirs. That those who do not face the same struggles tied to the specific culture can disengage whenever and wherever heightens the sense of exploitation. This logic invites debates as to who stands to gain and who is rendered invisible or irrelevant in these exchanges, with certain communities losing control over their narratives. This inherent power imbalance gives dominant cultures the ability to define what is "cool" and "valuable," as well as the freedom to profit or gain social capital off things that marginalized communities are at times still punished or mocked for (Young 2005). This is also in line with Chatterjee's (2020) critique, which distinguishes between authenticity and mere entertainment or commercialization, as cultural practices are recontextualized and commodified for broader consumption.

These concerns with purity and power have led to a growing body of scholarship on the concept of authenticity in cultural exchange, and in turn, appropriation and appreciation. Harrison (2008) puts forward a powerful explanation that authenticity itself is a contested concept, a social construct shaped by history, race, and lived experience – a claim that is made rather than an organic quality naturally found in things. Rigid lines dividing racial identities or cultural practices must be questioned, as authority to define what is "authentic" is itself unevenly distributed and often reflects broader structures of power and legitimacy. For Harrison (2008), the juxtaposition between black and white racial identity "ignores the tremendous range of people who also have strong attachments to hip hop." Such dichotomous boundaries risk diluting the discourse of diverse cultural and transnational processes to black-and-white hero-to-villian stances: "Either white hip hoppers are genuine devotees to 'the culture of hip hop,' or privileged interlopers whose mere presence threatens to destroy it" (Harrison 2008). This is particularly

problematic when one considers that Black culture is not a unified or monolithic entity, but shaped by history, diaspora, class, and context.

Growing recognition of this internal diversity has given weight to the concept of cultural negotiation as a challenge to that of cultural appropriation. Johnson (2020), for example, argues that the appropriation discourse is too simplistic, while the culture of calling out risks flattening and disregarding the nuances in power and authenticity in cultural exchange. Rather than strict categories of appropriation and appreciation, how responsible intercultural exchange looks is highly contingent on where one stands in relation to history and global power hierarchies. Chatterjee's (2020) idea of informed cultural exchange emphasizes the centrality of cultural literacy when borrowing from others' cultures, including knowing their roots and contexts and recognizing the significance of a particular artifact or act within their culture. Taken together, these interventions suggest that cultural negotiation offers a more precise framework for understanding cultural exchange as it unfolds in practice.

2. Dance as a Site of Cultural Negotiation

Dance is extremely revealing as a site of cultural negotiation because it is both an embodied and performative practice, staging cultural exchange through bodies rather than texts or objects. As an embodied practice, dance carries layers of meaning much beyond its aesthetic choreography. Embedded in it are histories, power dynamics, and identities that are remembered and reinterpreted through not only movement, but also intention and context (Johnson 2020). Unlike static cultural artifacts, dance occurs relationally between artists and their audiences, as well as traditions; this makes visible the ethical negotiations that occur when culture travels.

Integral to dance culture is the practice of fusion across genres, which involves drawing inspiration from others. Dancers today promote and document their work through videos on social media, where fellow choreographers routinely watch, observe movement quality and musicality, and incorporate particular styles, movements, or motifs into their own work. Within specific countries or regional dance scenes, dancers are often relatively familiar with their peers' work and styles; as a result, movements from other countries appear especially compelling for how radically different their movement trends are. One effective method of improving dance technique and quality is by closely mirroring the movement and muscle of choreographers, leading many young dancers to train using choreography created by artists overseas and adopt their styles. Through this crossing of international boundaries, movements and dance styles naturally travel across cultures.

As an example, breaking, one of hip hop's dance forms, is grounded in improvisation and innovation. The pinnacle of breaking within the broader hip hop culture are cyphers, where dancers take turns freestyling, observing their opponent's movement and responding on the spot (McCollum 2019). Because the form is largely improvisational, modes such as repetition, borrowing, and imitation are necessary to learn and expected of in battles and cyphers. However, the ethical distinction between the modes and stealing lies in freshness. Dancers are expected to tweak or even transform what they imitate by incorporating personal style. To follow these codes

means aspiring to contribute to the culture by adding new movement vocabularies and "continuously [renewing] and [refreshing] the style," rather than merely copying or "biting" existing movement (McCollum 2019). These processes of borrowing and imitation, if executed ethically, are widely accepted and acknowledged within the global dance community.

For these reasons, examining dance through a cultural negotiation lens while paying attention to cultural syncretism allows us to think and act beyond the rigid binaries of appropriation and appreciation. This approach recognizes cultural participation as contingent along a spectrum of literacy, accountability, and power. This subsequent section discusses this spectrum through two empirical examples: white engagement with hip hop and Afrobeats' incorporation into K-pop.

3. White Participation in Hip Hop: Between Appropriation and Appreciation

Hip hop as a cultural form is deeply connected to Black American histories of resistance, marginalization, and community-building, and carries meanings that go so much further beyond aesthetic movement. Breaking, for example, is often traced by scholars to Afro-diasporic practices such as Capoeira Angola, a dance form shaped by the lived experiences of enslaved West and Central African peoples in Brazil during the transatlantic slave trade (McCollum 2019). These diasporic communities made the martial arts dance movement an act of cultural preservation of the "African Brazilian experience" and resilience under Portuguese colonial rule (Capoeira Angola Flagstaff). Oral histories suggest that Capoeira drew from Angolan Bantu martial art traditions such as N'golo, otherwise known as "Zebra dance" or "Engolo" (Ikechukwu 2024). Like breaking, other hip hop dance styles are often markers of an oppressive and resilient history of Black diasporic movement traditions.

White dancers, when participating in hip hop, inevitably enter these histories, willingly or not. This section unpacks how such engagement unfolds along a spectrum of cultural negotiation, contingent upon embodied experience, intention, cultural literacy, and responsibility (Johnson 2020). On one end of the spectrum, commodification and dilution define the exploitative practices. Early 2010s pop music often borrowed from black, Southern, and hip-hop genres, partly because the idea of hip-hop and rap music being fused into American pop culture was on the ascent, with stars such as Kendrick Lamar, Nicki Minaj, and Drake on the rise.

Johnson (2020) also mentions Iggy Azalea and Miley Cyrus as artists who rose to prominence by drawing heavily on aesthetics rooted in Black communities. Miley Cyrus incorporated aspects of Black music, notably hip hop, R&B, and trap, into her *Bangerz* era in 2013, as well as subsequent projects such as her 2017 album *Younger Now* (Beg 2017). Her 2013 MTV VMA performance was decorated with twerking choreography, which Cyrus, herself, associated with vulgarity and superficiality (Yang 2019). She acknowledged that the performance was, in part, a vessel to escape her "'good girl' status she earned on Disney," and that she wanted to appear "ratchet," "wild," and "grown," using Black club and dance culture to redefine her image (Glover 2023).

When her performances led to controversy, Miley Cyrus claimed that her use of hip-hop

and Black aesthetics reflected her real lifestyle and friendships with artists like Juicy J, Wiz Khalifa, DJ Paul, and Mike WiLL Made-It and explained that she was actively present in studio environments, writing music and collaborating directly with major hip-hop artists such as Ludacris and Big Sean (Mohammed 2025). Yet, this still fails to justify how Cyrus used twerking, which has deep social implications, as a marketing tool to escape her so-called "good girl" status. Perhaps more problematically, soon after she used hip-hop and these Black aesthetics to gain popularity, Cyrus criticized black culture by saying "[she] can't listen to certain rap music anymore. It was too much Lamborghini, got my Rolex" (Beg 2017). Her abandonment of the culture after exploiting it for media attention and a rebranded image is, essentially, a reflection of centuries-long experience of Black people of being told their art is only valuable when someone else performs it for profit. While Cyrus can benefit from a culture that she can disengage from at will, her engagement perpetuates negative stereotypes of Black and African cultures and takes away the agency of the originating culture over the dance's representation. These examples show what Zink (2016) referred to as racialized spectacle: Black dance practices are extracted for shock value and entertainment, then discarded when no longer profitable. In such cases, white performers receive the benefit of cultural capital while remaining insulated from the stigmatization and structural violence faced by Black practitioners. This reinforces the power asymmetries central to cultural appropriation.

On the more appreciative end of the spectrum, cultural negotiation is characterized by cultural literacy: knowledge of hip hop's origins, contexts, and meaning, coupled with an understanding of one's own positionality within present power dynamics (Johnson 2020). Appreciation is not solely a question of intent but rather sustained practice, respect for lineage, and a predisposition to centres and uplifts originating communities. Yet, it must be stressed that appreciation does not remove participants from power considerations. Even well-informed and respectful engagement remains implicated in racial histories of inequity, and responsibility does not dissolve because harm is unintentional.

Other scholars, like Connell (2021), further provide a gendered understanding of this discussion by contextualizing hip-hop as a site wherein women, especially Black women, reclaim voice, bodily integrity, and narrative control. Any white woman's engagement can thus only be positioned within both race and gender dynamics, rather than being couched in neutral terms of self-expression. Campbell further explains that white women's involvement with hip-hop dancing reveals the tensions located at the axis of racialized sexuality and cultural authority, torn between colonizing valuations and attempts toward negotiating fresh expressive spaces.

Taken together, these practices make the call-out culture's reduction of all participation into moral absolutes untenable. The globalized art form of hip-hop is infused with identity, belonging, and community. White participation, if understood within cultural negotiation, gestures toward an honesty regarding the "messy realities" of cultural exchange, which centers accountability without precluding ethical participation as an ontological impossibility (Johnson 2020).

4. **Black Culture and Afrobeats in K-Pop: Postcolonial Complexity and Reverse Appropriation**

The incorporation of Afrobeats and Black culture into K-pop is a case that compellingly shows the postcolonial complexity in cultural negotiation, challenging the conventional model of cultural appropriation as a unidirectional flow from the West to the rest. In this particular interaction, the formerly colonized, non-Western cultural industry adopts and profits from Black culture. In doing so, the industry copies African diasporic sounds and dance aesthetics, underlining how extractive dynamics can be reproduced among the very cultures that are or were marginalized or peripheral. This complicates moral frameworks that locate appropriation solely within Western dominance and opens up questions about hierarchies within the Global South.

Korean-American entrepreneur and music content creator Jane Lee recognizes K-pop as Black culture "systemized through a Korean entertainment industry" (Lee 2025). For Lee, the genre's origins can be traced back to the very first K-pop group that debuted in 1992, Seo Taiji and Boys, whose work closely mirrored U.S. hip-hop. This is exemplified by parallels between the group's 1995 hit *Come Back Home* and Cypress Hill's 1993 song *Insane in the Brain*. The track took from Cypress Hill's West Coast hip-hop and G-funk sounds, incorporating the musical styles of artists such as B-Real and Sen Dog. Notable is how it was around this time when Seo Taiji was recognized as the "'President of Culture'" in Korea (Suh 2018).

With Seo Taiji and Boys' member Yang Hyun Suk later founding YG Entertainment – one of the "Big Four" K-pop labels consisting of HYPE Corporation, SM Entertainment, JYP Entertainment, and YG Entertainment – the industry's lineage continues to feed off of Black culture. This "borrowing" persists in contemporary K-pop through the continued use of 808s, syncopation, melismatic vocals, call-and-response structures, and hip-hop-derived performance styles (Lee 2025). This goes to show how much of K-pop's reliance on Black music styles, such as hip-hop, stems from a conscious choice to emulate dominant Western pop trends.

This is evident in Jay Park's discussion of his 2023 single *"Why,"* in which he explicitly states the incorporation of Afrobeats in the song not for cultural means but as a stylistic "switch up" aligned with current global trends. In an interview, Park explains that producer Cha Cha Malone had been "making a lot of Afrobeats," prompting him to incorporate an Afrobeats rhythm and even an Afrobeats-inspired section by sampling his hit track "Solo." Additionally, the song is paired with Afro-style choreography (Power 106 Los Angeles 2023). His characterization of the process – emphasizing how he dealt with the challenge of digesting a new genre – positions Afro-diasporic music and dance as aesthetic resources rather than cultural forms that require thoughtful contextual engagement. The choreography itself, featuring movement such as leg-crossover footwork and the Ghanaian *zhibit*, and its music video, featuring dancers of various backgrounds across Africa, the Americas, and Asia, further exposes how African dance vocabularies and Black bodies are intentionally absorbed into K-pop performance with a primarily purpose of expressing the artists', and more broadly K-pop's, global relevance (Park 2023).

Beyond sound, K-pop adopts Black cultural signifiers, including fashion, visualizers or music videos, hairstyles, slang, and choreography, and many of its global hits are written and produced by Black creators such as Teddy Riley and Tayla Parx, who produced Girls' Generation's "The Boys" and BTS's "MIC Drop (Steve Aoki Remix)," respectively (Dahir 2022). Yet, that these same expressions historically marginalized Black artists while commercially rewarding K-pop's adoption is telling of the problematic dismissal of power dynamics in the transfer of culture across contexts.

It is worth noting that the Black culture that the K-pop industry continues to borrow is borne from the brutality of African enslavement as well as both historical and ongoing white domination. There is then a clear asymmetry embedded in the dynamics of creative and artistic ownership: while Korean pop artists freely appropriate Black music and signifiers, they bypass the personhood of those from whom the cultural forms originate. Black English dance forms and hip-hop movement vocabularies are rampant in Gen-Z pop culture references. However, even when expressly flagged as originating in Black cultural forms, the recognition tag does nothing to hinder movement across digital spaces in ways that continue to reify structural inequities and bypass material realities. Even when origins are acknowledged, through citations, hashtags, or public recognition, such gestures do not necessarily interrupt unequal commercial flows of visibility or profit.

More specifically concerning the engagement between K-pop and Afrobeats, this often occurs at the level of sonic texture and visual imagery: polyrhythmic beats, percussive dance styles, and stylized references to "Africanness" without giving substantial acknowledgment of African artists, choreographers, or cultural contexts (Lukanov 2024). Welang's (2022) concept of "acknowledgement washing" is particularly useful here: the acknowledgement of African influence is made without genuine commitment to addressing the underlying issues or making meaningful change. It functions as a superficial act that can be seen as a way to appear virtuous or politically correct without taking concrete steps towards reconciliation or justice, exposing even deeper patterns of erasure and asymmetrical benefit. In all, African cultural production is rendered raw material: aestheticized, depoliticized, and rebranded into a Korean-led global industry.

This foregrounds a broader tension within globalization, where the same processes of globalization simultaneously enable cultural circulation and entrench asymmetry, accruing visibility and capital to select industries while peripheries are created and maintained. Indeed, Afrobeats' presence in K-pop is representative of this duality. On one hand, it signals the global reach and desirability of African soundscapes; on the other, it risks reproducing stereotypes and misrepresentations detached from lived African realities.

As Adewumi (2018) explains, such ambivalence is highlighted by globalization, enriching as much as complicating the construction of African identities. The more there is cultural exchange, the more dialogue and hybridity occur, and vice versa. In the K-pop context, Afrobeats-inspired dance and music often circulate void of the historical grounding that would frame them as products of specific social, political, and diasporic experiences.

Case Study: CocaNButter's "Mi Deh Yah" at the Intersection of Dance Circulation and K-pop

Dance sometimes falls outside of conventional accusations of cultural appropriation because current forms circulate beyond nation-state borders. Many dancers specialize in genres that originate outside their own national or cultural contexts. Their "legitimacy" is therefore proven through training and sustained engagement in their work in dance rather than ethnicity or lineage alone. In Korea, Afro-diasporic dance is taught and practiced within established communities, including African dancers – such as Ivorian and Beninese instructor Daniel Ahifon – who teach styles like Congolese Ndombolo to Korean dancers. These pedagogical networks suggest that transmission through direct cultural contact is essential and even an intrinsic quality in dance culture.

CocaNButter, a street dance crew composed of members with documented training in Afro and girls hip-hop, is present in both the dance scene and the K-pop industry. The Korean group originally came together as dancers to participate in Mnet's 2021 reality TV show *Street Woman Fighter (SWF)* Season 1, and most have taught weekly girls hip-hop and Afro classes in Korean dance studios.

The release of their debut single "Mi Deh Yah" elicited significant controversy, particularly among Black, African, and African-diasporic K-pop fans, who reproached the group for perpetuating familiar narratives of Black cultural extraction within the Korean entertainment industry. Criticism centered on the how the song was branded, as well as the aesthetics in its music video, including its Jamaican Patois title, collaboration with Koonta, use of AAVE slang, reggaeton rooted in Jamaican dancehall, and visuals such as box braids and Senegalese twists worn by members (Pan African Lifestyle 2025). However, their performance of Afro-diasporic dance forms such as Ndombolo, dancehall, and Afrobeats is considered more difficult to categorize as exploitative.

CocaNButter does not obscure the Afro-diasporic movement vocabularies as the choreography for the single remains stylistically intact, with respect to the Afro dance genres themselves and the members' context as trained Afro dancers and teachers. Yet what is seen as "appropriative" is that their visibility is amplified through a Korean entertainment industry that, historically, profits from Black cultural production while rarely acknowledging their borrowing of the culture's art and aesthetics.

Titling their single "Mi Deh Yah" – Jamaican Patois phrase for "I am here" or "I am present" in terms of one's state of mind – and collaborating with a Korean rapper Koonta (Kunta) – a name with African linguistic origins associated with meanings such as "warrior" or "bravery" – do not, in themselves, constitute genuine homage to Black culture or to the communities that produced and sustained these art forms (Solomon 2016). Instead, these choices are better understood as strategic deployment of Afro signifiers as cultural aesthetics, where the primary outcome of their adoption is commercial gain rather than reciprocal cultural engagement.

This dynamic becomes particularly troubling given that "a significant amount of K-pop fans are Black" (Fineman et al. 2017). Seen in this light, Black audiences played a role in

amplifying K-pop through social media and streaming, which in turn facilitated the genre's expansion into mainstream global markets. Even when negative backlash from Black audiences emerged, even this response generated visibility and discourse that further circulated K-pop content.

This strategy becomes more legible when considering CocaNButter's marginal position within the K-pop industry. CocaNButter did not follow the formula of conventional idol groups, developed through trainee systems or survival programs capped at younger age ranges; instead, they first emerged as a dance group, slowly transitioning into idol-adjacent performance as the *Street Woman Fighter* (2021) gained popularity. Provoking controversy, then, may have been an alternative pathway to relevance in an industry where dancers who seek idol careers are rarely legitimately perceived as full idols, especially given the group's relatively limited institutional backing compared to major entertainment agencies, or star choreographers like Bada Lee, Leejung, or the Kwon Twins. Seen in this light, the mobilization of Black cultural outrage appears to have been intentionally leveraged as a means to enter the market in a system that excludes groups and labels like CocaNButter and MLD Entertainment, respectively.

The implications of hybridity become even clearer when contrasted with K-pop groups that include members from the cultures being referenced, such as D.R Music's girl group BLACKSWAN, featuring Senegalese member Fatou Samba. However, representation alone does not resolve appropriation, as factors such as authorship, profit, authenticity, and inherent power dynamics remain. CocaNButter's case ultimately reveals how dance, as a shared and multicultural practice, becomes ethically contested when embedded within an industry that unevenly monetizes cultural labor. The tension lies less in the act of the dancing itself, but the institutional environments and commercial frameworks that shape how movement is framed, circulated, and capitalized upon.

Views from Cultures of Origin

Peer recognition is particularly important for nations building in the shadows of colonialism in a world still largely understood to be dominated by Western powers. Political scientist Laura C. Mahrenback (2018) asserts that emergent states ought to be evaluated not just through rising material capacity and diplomatic ambition, but also the element of peer recognition: the express acknowledgement of a country's rising status by other nations.

As the global order is increasingly shaped by media circulation and symbolic-immaterial power as a means of controlling perceived status, cultural production becomes a critical dimension through which rising nations must be considered. The exportation of Nigerian Afrobeats – now circulating widely through global streaming platforms, West African diaspora communities, Western music industries, and even K-pop – then may be considered as a form of Nigerian cultural soft power. A close examination of the Afrobeats world shows its potential for challenging colonial narratives and repositioning Africa not as a marginalized subject of global culture, but as a centrally active producer in the global scene (Allen 2024).

Recognition catalyzes greater interest in West African culture. Afrobeats circulates music alongside fashion, slang, dance, aesthetics, and language (Al Jazeera 2023). Afrobeats artists like Tems and Burna Boy are appearing at major fashion events like Paris Fashion Week to partner with luxury brands and bring West African styles to global runways (Spotify 2025). On the culinary side, West African staples like yams and okra are becoming more prominent in global food culture, such as in ASMR mukbangs (Allie 2025).

Then, should Korean industries amplify the cultural resonance of the cultures they borrow from around the world, the state of origin's soft power would expand. The political status of Nigeria, for example, may rise to one more obviously marked by social belonging and recognition as a whole. Some even contend that Afrobeats may clairvoyantly indicate the potential for a repositioning of the entire African continent in global cultural consciousness (Bell 2021).

This case, therefore, serves as a textbook example of cultural negotiation under the postcolonial conditions inasmuch as it occurred not between center and periphery but among differently positioned peripheries. Such a case shows how ethical literacy is required not only in the context of Western engagements with non-Western cultures but also in the context of Global South-South exchanges. It also shows that when the original cultural context of a particular dance is sufficiently respected, it has the potential to facilitate greater exchange and understanding between different cultures. For informed intercultural exchange to work as a function of a viable ethical framework, it would need to take into consideration how power works relationally, even among cultures with shared histories of marginalization.

5. Discussion and Conclusion: Towards Informed Intercultural Exchanges in Dance

Taken together, the cases discussed in this paper show that dance is a particularly fruitful site for reimagining cultural exchange beyond rigid binaries of appropriation versus appreciation. As a globalized art form, dance – perhaps most conspicuously in genres like hip hop – is an embodiment of identity, community, and belonging even as it circulates across unequal cultural, racial, and economic terrains (Johnson 2020). Precisely because dance operates through embodied movement rather than static artifacts, it makes visible the negotiations through which cultural meaning, power, and responsibility are enacted. These are conditions that render ethical clarity difficult, yet necessary. Simplistic judgments risk obscuring lived, embodied realities through which dancers engage cultural forms that are not "theirs," yet which are deeply meaningful to them. To recognize these messy realities is not to dilute accountability but rather to enable more precise ethical engagement.

Dance also matters because it operates as a practice through which individuals and communities reclaim voice, agency, and narrative authority. Connell's work illustrates ways that dance, particularly within hip hop, operates as a site of articulation by and for women and otherwise minority performers who continue to face erasure. These are the political dimensions that raise the stakes around cultural exchange. When cultural forms grounded in struggle are mobilized without care, they are stripped of their critical force. On the other hand, when ethically

engaged, dance can facilitate dialogue, solidarities, and shared meaning across differences. This underscores why dance is not merely an example of cultural exchange but offers a lens to observe how cultural negotiation unfolds in practice.

Building on the literature and empirical cases discussed above, the paper ends by highlighting three interrelated principles of informed intercultural exchange in dance. First is the need for contextual literacy. To engage with a dance form in an ethical manner, one must have knowledge about the historical origins of a particular form, its social meanings, and the embedded power relations within it (Johnson 2020). This means more than surface-level familiarity; rather, it requires understanding how race, class, and colonial histories continue to shape movement practices and their reception.

Secondly, credit and collaboration need to move further from symbolic acknowledgment towards active recognition and material engagement. In this regard, Mahrenback (2018) and Chatterjee (2020) stress that to ensure ethical exchange, one has to create space for originating communities through collaboration, visibility, and equitable distribution of cultural capital. This, therefore, principally goes against the grain of "acknowledgement washing" practices, where influence is nominally cited while agency and authorship remain centralized elsewhere. In dance, where movement itself functions as a form of citation, the absence or presence of such collaboration becomes especially visible.

Finally, reflexivity calls for an ongoing critical awareness of positionality, in particular urging dancers to be continuously aware of their situatedness at the crossroads of different structures of domination and how their complicity may reproduce or contradict inequality. Yet, this does not absolve others from ethical accountability; the K-pop context underscores how cultural literacy and responsibility are necessary in both Western and non-Western exchanges and those across the Global South. Reflexivity is not an ethical checkpoint at the beginning but an ongoing practice, one that requires openness to critique and a willingness to shift one's methodology. It follows that responsibility, in these terms, is relational and contingent, defined contextually rather than solely by intention. Together, these principles provide a pragmatic framework for negotiating culture within the dance without collapsing complexity into moral absolutes.Seen in this light, cultural negotiation emerges not as a compromise concept, but as a more analytically precise and ethically demanding framework than appropriation and appreciation binaries allow. Dance, by virtue of its embodied and relational nature, exposes how cultural exchange is continuously made and remade through practice. It shows that culture does not simply travel but is negotiated, choreographed through bodies that carry history even as they generate new meanings.

While this article has focused on dance, the implications of this argument extend to other globalized cultural practices. However, dance's immediacy – its reliance on bodies, presence, and shared space – makes it especially revealing of the ethical tensions embedded in cultural circulation. By attending to dance as a site of cultural negotiation, we gain not only a richer understanding of intercultural exchange but also a framework for engaging culture in ways that acknowledge complexity without abandoning responsibility.

Works Cited

Adewumi, A. (2018). Afrobeat Music as Cultural Heritage in Nigeria: Myth or Reality? SSRN Electronic Journal, 5(1).

Al Jazeera Staff. (2023, June 27). Q&A: Afrobeats Is “one of Africa’s biggest cultural exports.” Www.aljazeera.com. https://www.aljazeera.com/features/2023/6/27/qa-afrobeats-is-one-of-africas-biggest-cultural-exports

Allen, E. (2024, July 17). The unstoppable rise of Afrobeats. African Business. https://african.business/2024/07/arts-culture/the-unstoppable-rise-of-afrobeats

Allie, T. (2025, September 21). The African Diaspora’s Impact on Global Food and Drink Culture. Uncorked & Cultured. https://uncorkedandcultured.com/african-diaspora-food-drink-culture/

Androutsopoulos, J., & Scholz, A. (2003). Spaghetti Funk: Appropriations of Hip-Hop Culture and Rap Music in Europe. Popular Music and Society, 26(4), 463–479. https://doi.org/10.1080/0300776032000144922

Beg, A. (2017, November 14). BEG: Miley Cyrus must recognize her cultural appropriation. Marquette Wire. https://marquettewire.org/3980846/opinion/beg-miley-cyrus-must-recognize-her-cultural-appropriation/

Bell, R. (2021, January 25). Nigeria: the country that loves to overachieve. BBC. https://www.bbc.com/travel/article/20210124-nigeria-the-country-that-loves-to-overachieve

Campbell, M. (2004). “Go white girl!”: hip hop booty dancing and the white female body. Continuum, 18(4), 497–508. https://doi.org/10.1080/1030431042000297626

Capoeira Angola Flagstaff. (n.d.). History of Capoeira Angola. Capoeira Angola Flagstaff. Retrieved January 4, 2026, from https://capoeiraflagstaff.org/history-of-capoeira-angola

Chatterjee, D. (2020). Cultural appropriation: Yours, mine, theirs or a new intercultural? Studies in Costume & Performance, 5(1), 53–71. https://doi.org/10.1386/scp_00013_1

Connell, G. L. (2022). Breaking the cypher-circle: resistance of gender violence and reclamation of power through spatial re-orientation in British Hip Hop dance theatre. Surrey Open Research Repository (University of Surrey). https://doi.org/10.15126/thesis.900387

Dahir, S. (2022, May 13). Naegaz in Paris: How Black Culture Created K-pop. B-Side. https://berkeleybside.com/blackcultureandkpop/

Fineman, N., Quick, S., & Jaffe, L. (2017, March 28). Video: Crossing Cultures: Black K-Pop Fans in America. The New York Times. https://www.nytimes.com/video/arts/100000005008643/bts-k-pop-black-fans-america-race-360.html

Glover, A. (2023, September 9). The 2013 performance that shattered Miley Cyrus’ good-girl image. Celebrity.nine.com.au. https://celebrity.nine.com.au/latest/freeze-frame-ten-years-since-miley-cyrus-controversial-2013-mtv-vmas-performance/a50e196a-ef08-4f5f-b662-13b2319d5e01

Hall, S. (1993). What Is This “Black” in Black Popular Culture? Social Justice, 20(1/2 (51-52)),

104–114. http://www.jstor.org/stable/29766735
Harrison, A. K. (2008). Racial Authenticity in Rap Music and Hip Hop. Sociology Compass, 2(6), 1783–1800. https://doi.org/10.1111/j.1751-9020.2008.00171.x
Ikechukwu. (2024, April 12). N'golo (Engolo): The African Martial Art Dance and its Influence on Capoeira and Diaspora - Pan African Central Core. Pan African Central Core. https://panafrocore.com/2024/04/12/ngolo-engolo-the-african-martial-art-dance-and-its-influence-on-capoeira-and-diaspora/
Johnson, Imani Kai. (2020). BLACK CULTURE WITHOUT BLACK PEOPLE: Duke University Press EBooks, 191–206. https://doi.org/10.2307/j.ctv1131bwx.15
Lee, J. (2025). Kpop is Black Culture [Reel]. On Instagram.
Lukanov, M. (2024). Not Appreciating Your Appropriation: On the uses of Afrobeats in contemporary K-pop. African Mosaic, 1(1), 39–42. Sofia University St. Kliment Ohridski University Press. https://www.uni-sofia.bg/index.php/bul/universitet_t/fakulteti/fakultet_po_klasicheski_i_novi_filologii/novini/p_rvi_broj_na_spisanie_african_mosaic
Mahrenbach, L. C. (2018). Conceptualising Emerging Powers. Palgrave Macmillan UK EBooks, 217–231. https://doi.org/10.1057/978-1-137-45443-0_14
McCollum, S. (2019, October 30). Hip-Hop: A Culture of Vision and Voice (L. Resnick, Ed.). The Kennedy Center. https://www.kennedy-center.org/education/resources-for-educators/classroom-resources/media-and-interactives/media/hip-hop/hip-hop-a-culture-of-vision-and-voice/
Mohammed, L. (2025, June 10). After Years Of Being Accused Of Using Black Culture To Her Advantage, Miley Cyrus's Latest Comments On Her "Bangerz" Era Have Left People Seriously Divided. Yahoo Entertainment. https://www.yahoo.com/entertainment/articles/years-being-accused-using-black-145313904.html
Pan African Lifestyle. (2025, December 10). K-Pop Criticized for Appropriating Afrobeats & Dancehall Culture. YouTube. https://www.youtube.com/watch?v=ErWtPXgZBD8
Park, J. (2023, November 17). 박재범 (Jay Park) - "Why" Official Music Video. YouTube. https://www.youtube.com/watch?v=3rTrH6a9vAI
Power 106 Los Angeles. (2023, November 21). Jay Park On The Evolution Of Korean Rap, Afrobeats, Upcoming R&B Album, Won Soju + More! YouTube. https://www.youtube.com/watch?v=a49HPBDS84w
Solomon, Jason. (2016, October 8). JaLingo's Weekly Expression - Mi Deh Yah. YouTube. https://www.youtube.com/watch?v=2YYApdNBpxw
Spotify. (2025, September 22). The Visual Rhythm: How Afrobeats Is Shaping Global Style — Spotify. Spotify. https://newsroom.spotify.com/2025-09-22/the-visual-rhythm-how-afrobeats-is-shaping-global-style/
Suh, S. W. (2018, December 27). The Roots of K-pop—When it was actually pretty darn good.

Flapperpress.
https://www.flapperpress.com/post/the-roots-of-k-pop-when-it-was-actually-pretty-darn-good
Welang, Nahum. (2022). Cultural Appropriation or Cultural Appreciation? Cultural Appropriation or Cultural Appreciation?: Unpacking the Microdystopias of Beyoncé's Black Is King, 151–170. https://doi.org/10.5040/9781666996630.ch-9
Yang, R. (2019, June 12). Miley Cyrus apologizes for "racially insensitive" comments about hip-hop. EW.com.
https://ew.com/music/2019/06/12/miley-cyrus-apologizes-for-insensitive-hip-hop-comments/
Young, J. O. (2005). Profound Offense and Cultural Appropriation. The Journal of Aesthetics and Art Criticism, 63(2), 135–146. http://www.jstor.org/stable/3700467
Zink, S. (2016). Twerking and Cultural Appropriation: Miley Cyrus' Display of Racial Ignorance. MacEwan University Student EJournal, 3(1).
https://doi.org/10.31542/j.muse.158

A Comparison of Thermal Power and Tidal Power: An Analysis of the Loss of the Earth-Moon System's Kinetic Energy for Reducing Carbon Dioxide Emissions

By Chaeho Song

Abstract

This paper compares thermal power generation and tidal power generation in terms of both carbon dioxide emissions and the source of energy. First, we study thermal power generation with a simple hydrocarbon model and estimate how much carbon dioxide is produced when a certain amount of energy is generated. Next, we explain the basic idea of tidal power generation and connect tidal energy to the rotational kinetic energy of the Earth. Using a simplified model of the Earth, we also calculate the Earth's rotational kinetic energy. After that, we compare the two energy sources in two ways. First, we estimate how much carbon dioxide would be produced if an amount of energy equal to the Earth's rotational energy were generated by thermal power. Second, we estimate how long the Earth's rotational energy could theoretically support the current yearly global energy demand if all of it were supplied by tidal power. Through these calculations, we show that tidal power can play an important role in reducing carbon dioxide emissions, even though it also has practical limits. This suggests that tidal power is a very special natural energy source from an environmental point of view.

Keywords thermal power, tidal power, carbon dioxide emissions, Earth-Moon system

1. Introduction

Electricity is essential in modern life, but its environmental impact depends greatly on how it is produced. Thermal power generation has long been one of the main ways of producing electricity. However, because it relies on burning fossil fuels such as coal, oil, and natural gas, it releases a large amount of carbon dioxide. The IPCC states that human activities, mainly through greenhouse gas emissions, have unequivocally warmed the climate system, and the IEA reports that electricity generation remained the largest source of carbon dioxide emissions among all sectors in 2024 (IPCC 4) (IEA, Electricity 17).

The problem with thermal power generation is not limited to carbon dioxide emissions. The combustion of fossil fuels also contributes to air pollution. According to the World Health Organization, major air pollutants include particulate matter, carbon monoxide, ozone, nitrogen dioxide, and sulfur dioxide, and exposure to polluted air is associated with respiratory and cardiovascular diseases as well as other serious health problems (World Health Organization). Therefore, when we discuss the disadvantages of thermal power generation, we should consider both climate change and air pollution.

Because of these problems, there has been growing interest in cleaner energy sources that can replace or reduce the use of thermal power. One example is tidal power generation. Tides are caused mainly by the gravitational pull of the Moon and the Sun, while the Earth's rotation and the shape of coastlines affect how tides appear in different places (National Oceanic

and Atmospheric Administration). In a simplified physical sense, this means that tidal power does not create energy from nothing. Instead, it uses energy related to the motion of the Earth-Moon system.

This viewpoint provides a broader way to understand environmentally friendly energy. An energy source is often called environmentally friendly simply because it emits less carbon dioxide. That is certainly important, but it is not the whole story. It is also meaningful to ask where the energy comes from physically and what kind of cost is paid when that energy is used. Thermal power uses the chemical energy of fuel and releases carbon dioxide and air pollutants. Tidal power produces electricity with almost no additional carbon dioxide emissions during operation, but it relies on the motion of the Earth and the Moon. In other words, even when the same amount of electricity is produced, the physical source of the energy is different.

In this paper, we first estimate how much carbon dioxide is produced when a certain amount of electricity is generated by a thermal power plant. Next, we assume that the same amount of electricity is generated by a tidal power plant, and we interpret this energy as a decrease in the kinetic energy of the Earth-Moon system. Finally, we estimate how this energy loss may affect the rotational speed of the Earth and the orbital speed of the Moon through a simple mathematical model. Through this analysis, this paper aims to show that environmentally friendly energy should be understood not only in terms of lower carbon dioxide emissions, but also in terms of the physical source of the energy and the cost of using it.

This paper is organized as follows. In Section 2, we briefly explain the basic idea of thermal power generation and calculate how much carbon dioxide is produced when a certain amount of energy is generated through the combustion of hydrocarbon fuel. In Section 3, we explain the basic principle of tidal power generation, discuss how its energy is related to the rotation of the Earth, and calculate the rotational kinetic energy of the Earth. In Section 4, we compare the results of the previous two sections and analyze the meaning of tidal power from the viewpoint of reducing carbon dioxide emissions. Finally, in Section 5, we summarize the main results of this paper and discuss both the environmental significance and the limitations of tidal power generation.

2. Amount of carbon dioxide produced by thermal power generation

The basic idea of thermal power generation is quite simple. Materials such as fossil fuels or waste are burned to produce heat, and this heat is used to boil water. The steam from the water then turns a turbine, and the turbine drives a generator to produce electricity. In other words, chemical energy in the fuel is first changed into heat energy, and then this heat energy is used to generate electrical energy.

In this paper, our goal is to estimate how much carbon dioxide is produced by thermal power generation. In reality, many different kinds of fuels can be used, and their chemical compositions are often complicated. To make the calculation simpler, we assume that the fuel is a hydrocarbon made only of carbon and hydrogen. Moreover, we consider only hydrocarbons of the form C_nH_{2n+2}, which have the largest possible number of hydrogen atoms for a given

number of carbon atoms. Here, n is any natural number.

Hydrocarbons of this form are called alkanes. For example, when $n = 1$, we get CH_4, which is methane. When $n = 2$, we get C_2H_6, which is ethane. When $n = 3$, we get C_3H_8, which is propane. Thus, the formula C_nH_{2n+2} gives a simple general model for a basic hydrocarbon fuel.

Of course, actual fuels used in daily life are usually more complicated. For example, gasoline used in cars is not a single pure substance, but a mixture of many different hydrocarbons. However, in chemistry, gasoline is often represented by octane, C_8H_{18}, for simplicity. In the same way, this paper uses the general hydrocarbon C_nH_{2n+2} in order to study the amount of carbon dioxide produced during combustion.

Now let us assume that this hydrocarbon undergoes complete combustion with oxygen. In complete combustion, all carbon atoms become carbon dioxide, and all hydrogen atoms become water. Therefore, the combustion equation for C_nH_{2n+2} can be written as

$$C_nH_{2n+2} + \frac{3n+1}{2}O_2 \rightarrow nCO_2 + (n+1)H_2O.$$

This equation has an important meaning. Since one molecule of fuel contains n carbon atoms, complete combustion produces exactly n molecules of carbon dioxide. Therefore, the amount of carbon dioxide released by thermal power generation is directly related to the amount of carbon contained in the fuel.

In the following part, we will use this combustion equation to calculate the number of moles and the mass of CO_2 produced by burning a given amount of hydrocarbon fuel. From this, we will estimate how much carbon dioxide is emitted when a certain amount of electricity is generated. In this way, we can understand more clearly, in a mathematical way, why thermal power generation produces a large amount of carbon dioxide. To calculate the energy, let us first look at the following table.

Bond	Average bond energy (kJ/mol)
$O = O$	495
$C = O$	799
$C - H$	413
$O - H$	483
$C - C$	348

Table 1: Average bond energies used in this paper. These values are used to estimate the energy released during the combustion of C_nH_{2n+2}.

Bond energy is the amount of energy required to break one mole of a chemical bond. Therefore, the total energy released in a combustion reaction can be estimated by comparing the energy needed to break the bonds in the reactants with the energy released when new bonds are formed in the products.

(1) Bond energy of C_nH_{2n+2}: First, let us consider the types and numbers of bonds in the fuel molecule C_nH_{2n+2}. In this molecule, n carbon atoms are connected in a chain, so there are $n-1$ $C-C$ bonds. Also, since the total number of hydrogen atoms is $2n+2$, there are $2n+2$ $C-H$ bonds. Therefore, the total bond energy in 1 mol of C_nH_{2n+2} can be found by adding the bond energies of all $C-C$ bonds and all $C-H$ bonds. From the table, the bond energy of $C-C$ is 348 kJ/mol and the bond energy of $C-H$ is 413 kJ/mol. Hence, the total bond energy in 1 mol of C_nH_{2n+2} is

$$(1174n+478)\text{kJ}.$$

(2) Bond energy of O_2: When 1 mol of C_nH_{2n+2} undergoes complete combustion, $\frac{3n+1}{2}$ mol of O_2 is required. Since each O_2 molecule has one $O=O$ double bond, the total bond energy of all the O_2 molecules is

$$\frac{3n+1}{2}\times 495\text{ kJ}.$$

(3) Bond energy of CO_2: When 1 mol of C_nH_{2n+2} undergoes complete combustion, n mol of CO_2 is produced. Since each CO_2 molecule has two $C=O$ double bonds, the total bond energy of all the CO_2 molecules is

$$1598n\text{ kJ}.$$

(4) Bond energy of H_2O: When 1 mol of C_nH_{2n+2} undergoes complete combustion, $(n+1)$ mol of H_2O is produced. Since each H_2O molecule has two $O-H$ bonds, the total bond energy of all the H_2O molecules is

$$926(n+1)\text{ kJ}.$$

The bonds formed in the products, CO_2 and H_2O, release energy, while the bonds in the reactants, C_nH_{2n+2} and O_2, require energy to be broken. Therefore, the total energy released

when 1 mol of C_nH_{2n+2} undergoes complete combustion can be calculated by subtracting the total bond energy of the reactants from the total bond energy of the products. That is,

$$\text{Bond energy of } CO_2 + \text{Bond energy of } H_2O$$
$$- \text{Bond energy of } C_nH_{2n+2} - \text{Bond energy of } O_2.$$

From the previous calculations, we know

$$(607.5n + 200.5) \text{ kJ}.$$

Now, we consider the amount of CO_2 during this process. Recall that when 1 mol of C_nH_{2n+2} undergoes complete combustion, it produces n mol of CO_2. Since the molar mass of CO_2 is 44g/mol, the total mass of CO_2 produced is

$$(44n)g.$$

Therefore, the amount of CO_2 emitted per 1 kJ of released energy is

$$\frac{44n}{607.5n+200.5} = \frac{88n}{1215n+401} \text{g/kJ}.$$

3. Principle of tidal power generation and the amount of energy produced

To understand the principle of tidal power generation, we must first understand why tides occur. The Earth and the Moon move under their mutual gravitational attraction, but the gravitational pull of the Moon is not the same at every point on the Earth. The side of the Earth closer to the Moon experiences a stronger pull, while the side farther from the Moon experiences a weaker pull. In other words, the whole Earth accelerates toward the Moon, but the amount of acceleration is slightly different depending on the location. As a result, seawater on the side facing the Moon is pulled more strongly toward the Moon, while seawater on the opposite side is pulled less strongly. This difference causes the ocean to bulge on both sides of the Earth, and this is the basic reason why tides occur.

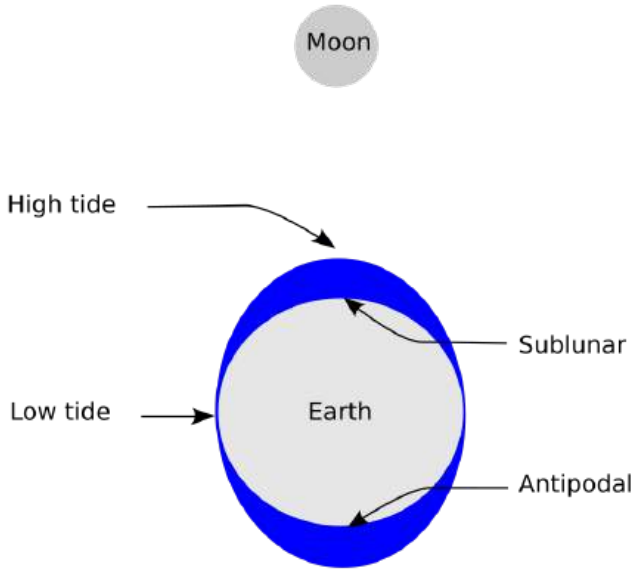


Fig 1: The Moon's gravity causes the ocean to bulge on both the side facing the Moon and the opposite side of the Earth. As a result, high tide occurs in these two regions, while low tide occurs in the regions between them.

One of the most important parts of a tidal power plant is a large reservoir that stores seawater. When the sea level outside the reservoir rises, seawater flows from outside into the reservoir. On the other hand, when the sea level outside the reservoir falls, the water inside the reservoir flows back out. Tidal power generation uses this flow of water. In other words, the gravitational effects of the Moon and the Sun create tides, and this natural movement of seawater turns turbines and produces electricity.

Then where does this electrical energy come from? According to (Espenak), tidal friction gradually slows down the rotation of the Earth. Therefore, the energy produced by tidal power can be understood, in a physical sense, as being related to the rotational energy of the Earth. In other words, although a tidal power plant generates electricity by using the motion of seawater, its energy ultimately comes from the Earth's rotation. Of course, it is difficult to say that all of this energy is converted into electricity with perfect efficiency. As in other methods of power generation, some energy is always lost. However, in this paper, we will assume that the efficiency of tidal power generation is 100% in order to make the comparison as clear as possible. That is, we assume that the amount of electrical energy produced by the tidal power plant is exactly equal to the decrease in the rotational kinetic energy of the Earth.

Now let us calculate the rotational energy of the Earth. The Earth is not a perfect sphere but is closer to an ellipsoid that is slightly longer in the equatorial direction. Let R_1 be the longer radius, that is, the distance from the center of the Earth to the equator, and let R_2 be the shorter radius, that is, the distance from the center of the Earth to the North Pole or the South Pole. Finally, let M be the total mass of the Earth, and for simplicity, assume that the density of the

Earth is uniform. Then the rotational kinetic energy of a body rotating with angular velocity ω is given by

$$E = \frac{1}{2} I \omega^2$$

where I is the moment of inertia about the axis of rotation. For a uniform ellipsoid, the moment of inertia about the Earth's rotation axis is

$$I = \frac{2}{5} M R_1^2.$$

Therefore, the rotational kinetic energy of the Earth is

$$E = \frac{1}{5} M R_1 \omega^2.$$

One can calculate the above energy using the following table.

Physical quantity	Value
Mass of the Earth M	5.972×10^{24} kg
Equatorial radius of the Earth R_1	6.378×10^{6} m
Angular velocity of the Earth ω	7.292×10^{-5} rad/s

Table 2: Physical quantities used to calculate the rotational kinetic energy of the Earth.

Substituting these values into the energy formula, we get

$$E = 2.58 \times 10^{26} \text{kJ}.$$

4. Comparison between thermal power and tidal power

In this section, we compare thermal power and tidal power from the viewpoint of carbon dioxide emissions. For this comparison, we assume $n = 8$, which corresponds to the hydrocarbon C_8H_{18}. From the result of Section 2, the amount of carbon dioxide emitted per 1 kJ of released energy is

$$\frac{88\times8}{1215\times8+401}\simeq0.06956 \text{ g/KJ}.$$

On the other hand, in Section 3, we estimated that the rotational kinetic energy of the Earth is approximately

$$E = 2.58\times10^{26}\text{kJ}.$$

Hence, if we imagine a theoretical situation in which this same amount of energy was produced by thermal power generation, then the total amount of emitted carbon dioxide would be

$$0.06956\times2.58\times10^{26} = 1.79\times10^{25} \text{ g}$$

of CO_2. This result becomes even clearer when it is compared with the current amount of carbon dioxide in the Earth's atmosphere. Using the global mean CO_2 data and an estimate of the total mass of the atmosphere, the total amount of carbon dioxide currently present in the Earth's atmosphere is about

$$3.34\times10^{18} \text{ g}.$$

On the other hand, the value calculated above is 1.79×10^{25} g, which is about 5.35×10^{6} times the current total amount of atmospheric CO_2.

Now let us look at this problem from a different point of view. This time, instead of comparing the amount of carbon dioxide directly, let us estimate how long the Earth's rotational energy could theoretically support human energy use if all the energy used by humanity in one year were supplied by tidal power.

According to the International Energy Agency, the total global energy demand in 2024 was about 650 EJ (IEA, Global Energy Review 2025 8). On the other hand, in Section 3, we calculated that the rotational kinetic energy of the Earth is about 2.58×10^{26} kJ. Converting this into EJ, we get

$$2.58\times10^{11} \text{ EJ}.$$

Therefore, if we make the highly idealized assumption that all the energy used by humanity in one year could be obtained from the Earth's rotational energy, then the available period would be

$$3.97\times10^{8} \text{ years}.$$

In other words, this simple calculation suggests that it could last for about 397 million years. Of course, this is only a theoretical upper bound, and it does not mean that a tidal power plant can actually extract all of the Earth's rotational energy in this way. Nevertheless, this calculation shows how enormous the Earth's rotational energy is, and it also shows once again that tidal power is a very special energy source.

5. Conclusion

In this paper, we compared thermal power generation and tidal power generation from both environmental and physical points of view. First, we analyzed the amount of carbon dioxide produced by thermal power generation. By using a simple hydrocarbon model, we calculated how much energy is released during combustion and how much CO_2 is produced at the same time. We then used this result to estimate the amount of carbon dioxide emitted per unit energy.

Next, we studied where tidal energy comes from. We explained that tidal power uses the motion of seawater caused by the gravitational interaction of the Earth, the Moon, and the Sun, and we interpreted this energy as being related to the rotational energy of the Earth. Under a simplified model, we calculated the rotational kinetic energy of the Earth and used it as the total energy source behind tidal power.

In the comparison section, we looked at this problem in two different ways. First, we estimated how much carbon dioxide could be avoided if the Earth's rotational energy were compared with thermal power generation. The result showed that the corresponding amount of CO_2 is enormously large, much greater than the current amount of carbon dioxide in the Earth's atmosphere. Second, we asked how long humanity could continue to use energy if the total yearly energy demand of the world were supplied by tidal power. Even though this was only a theoretical calculation, the result showed that the Earth's rotational energy is extremely large.

Of course, this does not mean that all human energy demand can actually be replaced by tidal power in practice. Real tidal power plants have many physical and technological limits, and not all the Earth's rotational energy can be extracted. However, this study still shows that tidal power is a very special energy source. Unlike thermal power generation, it can produce electricity without directly releasing large amounts of carbon dioxide, while drawing energy from a natural physical process that already exists on Earth. In this sense, tidal power may be regarded as a special gift from nature for protecting our environment.

Works Cited

"AR6 Synthesis Report: Climate Change 2023." Intergovernmental Panel on Climate Change (IPCC). Accessed April 23, 2026. https://www.ipcc.ch/report/ar6/syr/.

"Electricity 2025 Analysis and forecast to 2027." IEA. Published February 14, 2025. Accessed April 23, 2026. www.iea.org/reports/electricity-2025.

"Air Pollution." World Health Organization. Accessed April 24, 2026. www.who.int/health-topics/air-pollution#tab=tab_1.

"What Causes Tides?" National Environmental Satellite, Data, and Information Service. Published September 2, 2025. Accessed April 23, 2026. https://www.nesdis.noaa.gov/about/k-12-education/oceans-coasts/what-causes-tides.

F. Espenak. "NASA - Eclipse Predictions and Earth's Rotation." NASA Goddard Space Flight Center. Last Modified July 11,2005. Accessed April 27, 2026. https://eclipse.gsfc.nasa.gov/SEhelp/rotation.html.

"Global Energy Review 2025." IEA. Lastly Modified March 2025. Accessed April 28, 2026. www.iea.org/reports/global-energy-review-2025/global-trends.

Evaluating 8B-Parameter LLMs for English-Russian Translation and Methods to Improve Performance

By Tikhon Inogamov (student)[1 *], Faisal Nawab (mentor)[2]

Abstract

Large Language Models (LLMs) enable multilingual communication. However, high-performing models typically require expensive server-grade hardware, while smaller models can be run on hardware accessible to a much wider population. This paper presents research on the measurement and improvement of translational abilities of LLMs with 8B parameters. Preserving meaning, grammatical accuracy, and contextual coherence within the limited context window of such models presents significant challenges.

An automated evaluation system was developed to generate and then assess translations using a larger cloud-based model for scoring. Three models – LLaMA 3.1 8B, Qwen3 8B, and YandexGPT-5 Lite 8B – were tested across multiple configurations and two text datasets representing contemporary and classical language. In addition, this study leverages embedding-based selection with the aim of improving overall consistency.

The results indicate significant performance differences between Russian-to-English and English-to-Russian translation tasks, as well as across different datasets. Increasing temperature generally leads to decreased translation quality, while embedding-based selection does not produce a significant improvement at optimal settings.

Keywords

Artificial Intelligence; Large Language Models for Translation; Small LLMs; Translation between Russian and English; Embedding Vectors

Introduction

Translation capabilities of large language models allow people of different backgrounds to work together. Translations of high quality are needed, being able to accurately transfer the meaning of the text while representing it in the correct context. Models such as GPT-5 and Qwen have consistently demonstrated high-quality results in their translation abilities [1] [2] [16]. However, these models can only be hosted on dedicated server infrastructure, making them expensive to run, using a large amount of natural resources, therefore damaging the environment around it. As an alternative to these large models, smaller LLMs, which can be run on consumer-grade hardware and are more environmentally-friendly, are gaining popularity.

This study focuses on the evaluation and improvement of translation abilities of such models, in particular models with 8B parameters, between the Russian and English languages. The differences in the structure and grammar of these languages present unique challenges, offering space for improvement.

1. Henrietta Lacks Health and Bioscience High School, 9105 NE 9th St, Vancouver, WA 98664, United States of America
* Corresponding author email: tikhon102030@gmail.com

The objective of this study is to evaluate which models perform best for translation and identify methods that improve translation quality between Russian and English without requiring additional training or fine-tuning. Low-Rank Adaptations (LoRA) and fine-tuning techniques have been used to improve models' performance on specific tasks or within narrow domains [3] [13]. While these have shown to be effective, they often introduce undesirable trade-offs, including increased hallucinations and degraded performance on out-of-domain tasks. This study instead focuses on methods that avoid such modifications while still improving translation quality.

Factors such as temperature, model architecture, text source selection, and the use of an embedding-based selection strategy [4] all influence the quality of generated translations. To identify the most effective configurations, this study investigates how varying the temperature parameter affects translation quality.

Evaluating the quality of generated translations is essential for obtaining consistent assessments of model performance. To accomplish this, a large cloud-hosted LLM [5] is provided with detailed evaluation instructions and used to rate each generated translation across multiple quality categories [14].

Methods

To test different models and their configurations, a Python-based [6] automation system was designed. The system accepts multiple parameters, including local model choice and temperature settings, and runs the input text through the LLMs to generate translations and evaluate their quality.

To benchmark LLMs with various configurations and adjustments, two text corpora were used. A known limitation of smaller LLMs is their restricted context window, which prevents them from effectively working with large texts. To address this, this study utilizes a solution in the form of splitting the text by punctuation into sentence chunks with a set overlap. This results in preserved context in any given text chunk with respect to the full text, while remaining within the LLM's into the context window of the LLM. For all further testing, each text chunk contains 10 sentences and chunks overlap by 1 sentence.

Once the source text is chunked, each segment is translated sequentially. For each segment, the Ollama API [7] alongside the Python requests library were used to send the input text with instructions on translation to the model selected in the configuration. The model, prompt and temperature are all passed as the payload to Ollama's API endpoint using the Python requests library. A set of temperatures is tested. For each temperature value, all text chunks were translated again. The result is saved alongside the text chunks in the source language for further reference.

To assess the quality of the generated translations at each temperature setting, another LLM, gpt-oss:120b-cloud hosted by Ollama [5], is used. The same Ollama API interaction as for the translation is used. This model is hosted in the cloud and has a significantly larger parameter count, making it suitable for rating the smaller model's performance. Along with the source and

translated text, detailed instructions are given to the rating model. The instructions state that the translations must be rated independently across 5 categories: meaning, grammar, fluency, lexical choice and completeness. Each category is given alongside a brief explanation that defines what aspect of the translation it measures, ensuring consistent evaluation across all trials. It is also explained that the rating for each category ranges from 0 to 10, with 0 being a completely unusable and incorrect translation, and 10 being a professional level translation. For further analysis, the numerical ratings for each category are saved in the JSON format.

The result of the steps described above is a structured dataset comprising source text chunks, corresponding translated text chunks, and translation ratings for each temperature setting tested. This structure allows for systematic comparison of model performance across different configurations.

The translation procedure described above was applied to two datasets in both translation directions (English-to-Russian and Russian-to-English). The datasets consisted of *Crime and Punishment* by Fyodor Dostoevsky [8] (specifically Chapter 2), and a RusLTC article pair [9] (RusLTC_EN_1_94 and RusLTC_RU_1_94_1), which represent aligned English and Russian versions of the same text. These selections were intended to capture differences in both time period and writing style, including classical literature and modern academic writing.

To improve the stability of model outputs, an embedding-based selection strategy was implemented. For each input text, the translation was generated multiple times using the same model and configuration. The resulting outputs were then encoded into vector representations using the multilingual-e5-large sentence embedding model [4]. For all further testing, each input text was translated 10 times.

These embeddings capture the semantic content of each translation, allowing for quantitative comparison despite variability in the generated text. The centroid of all embedding vectors was computed, after which the cosine similarity between each embedding and the centroid was evaluated. The translation with the highest similarity to the centroid was then selected. This approach identifies the most semantically central translation among multiple generations, therefore potentially reducing variability and improving consistency.

To evaluate model efficiency, multiple performance metrics were measured. For each model and translation direction, the time to first token (TTFT), time to last token (TTLT), and average token generation rate (tokens per second) were recorded. The same input text chunks from both datasets were used to ensure consistency across tests. All measurements were conducted at a temperature of 0.1, as this setting was previously identified as yielding the highest translation quality.

All experiments were performed on the same hardware configuration, with models running on an NVIDIA RTX 4070 GPU. The Python-based script interfacing with the Ollama API described above was used to send translation requests and measure timing metrics. Token streaming was enabled, allowing precise measurement of TTFT and TTLT during generation.

Results and Discussion

The experimental procedure described above was conducted using three models: Qwen3:8b, LLaMA3.1:8b and YandexGPT-5-Lite-8B-instruct-GGUF:8b [10] [11] [12]. For each model, a range of temperature settings was tested for each source text across both translation directions (Russian-to-English and English-to-Russian). The tested temperature values were: 0.1, 0.3, 0.5, 0.6, 0.7, 0.9, 1.1, 1.2, 1.5, 2.0.

A finer sampling of temperatures was applied in the range between 0.5 and 0.7, as that usually serves as the default or commonly used temperature setting for such models. In contrast, higher temperature values (1.2 and above), which are generally associated with decreased performance in precision-oriented tasks such as translation, were evaluated at broader intervals. This approach allows for detailed analysis near the most commonly used settings while still capturing performance trends at higher temperature settings.

Effect of Temperature on Translation Quality

[15] The relationship between temperature and translation quality for each of the categories measured was evaluated across all models, datasets, and both translation directions. While a general trend of decreasing performance with increasing temperature was observed, the magnitude of this effect differed between models. Certain models were significantly affected by temperature changes, whereas other models demonstrated comparatively stable performance across the tested range.

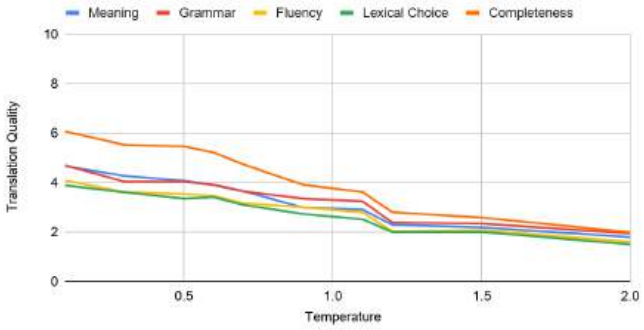


Figure 1.1: Illustration of relation between the temperature and category-based scores for LLaMA 3.1 8B, EN→RU, Crime and Punishment, Chapter 2.

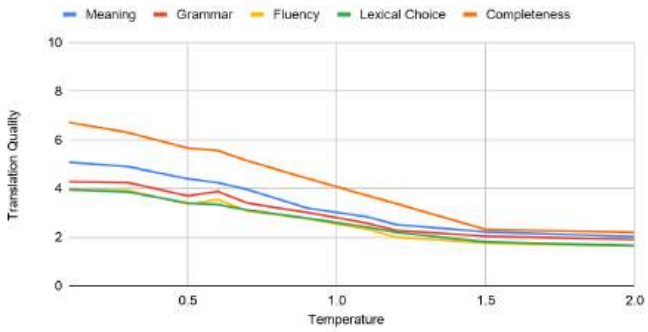


Figure 1.2: Illustration of relation between the temperature and category-based scores for LLaMA 3.1 8B, EN→RU, RusLTC

LLaMA3.1:8b in particular showed strong sensitivity to the change of temperature. Figure 1.1 and Figure 1.2 illustrate the relation between the temperature (x-axis) and category-based rating scores (y-axis). In both figures, low temperature settings (around 0.1) yield high average ratings of approximately 4.7 (Figure 1.1) and 4.8 (Figure 1.2). In contrast, increasing the temperature to 2.0 results in a substantial decline in performance, with average ratings dropping to approximately 1.8 and 1.9, respectively. This corresponds to a decrease of roughly 62% and 61%, indicating a strong sensitivity of the model to higher temperature settings.

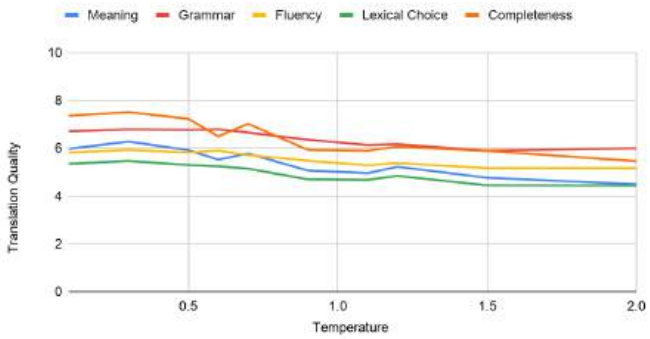


Figure 1.3: Illustration of the temperature changes' effect on performance for LLaMA 3.1 8B, RU→EN, Crime and Punishment, Chapter 2.

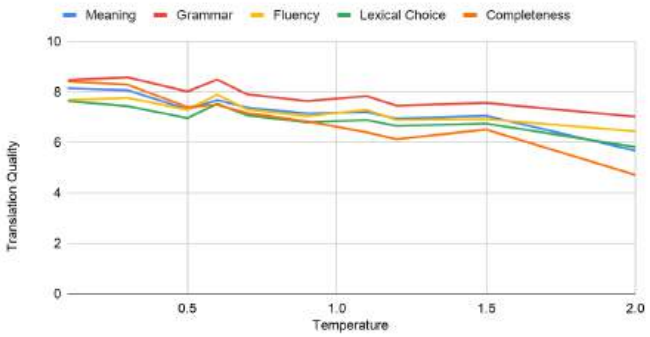


Figure 1.4: Illustration of the temperature changes' effect on performance for LLaMA 3.1 8B, RU→EN, RusLTC

Figures 1.3 and 1.4 demonstrate that temperature changes do not always affect performance to the same extent. In the case of Russian-to-English translation, the difference between low (0.1) and high (2.0) temperature settings is notably smaller, with average ratings decreasing by approximately 18% and 26%, respectively. This suggests that the impact of temperature on translation quality varies depending on the translation direction or dataset.

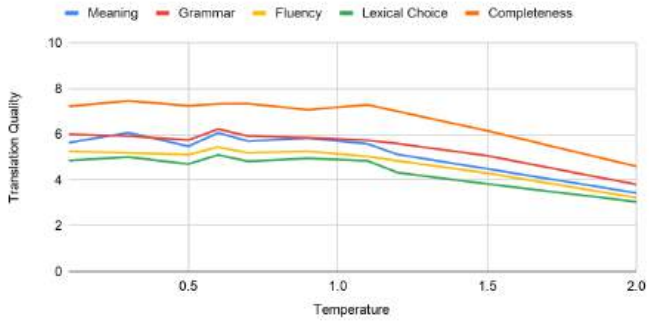


Figure 2.1: Illustration of a moderate sensitivity to temperature for Qwen 3 8B, EN→RU, Crime and Punishment, Chapter 2.

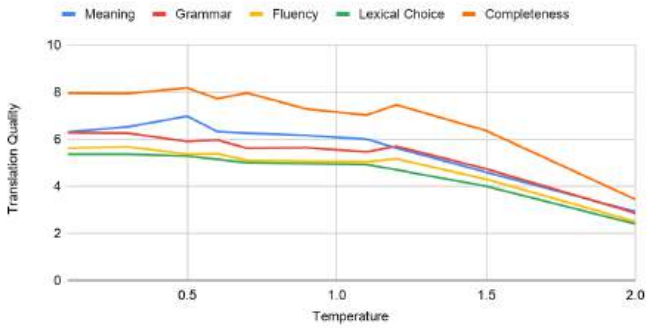


Figure 2.2: Illustration of a moderate sensitivity to temperature for Qwen 3 8B, EN→RU, RusLTC.

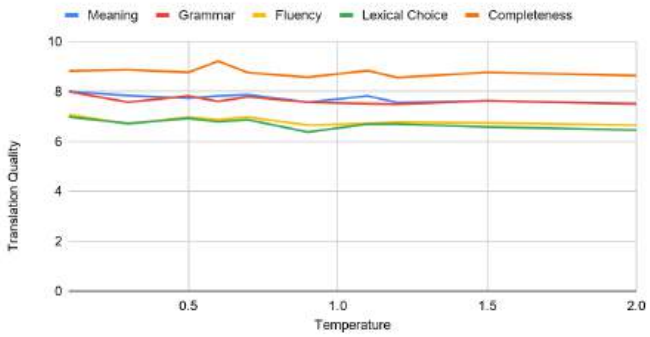


Figure 3.1: Illustration of a minimal sensitivity to temperature for YandexGPT 8B, EN→RU, Crime and Punishment, Chapter 2.

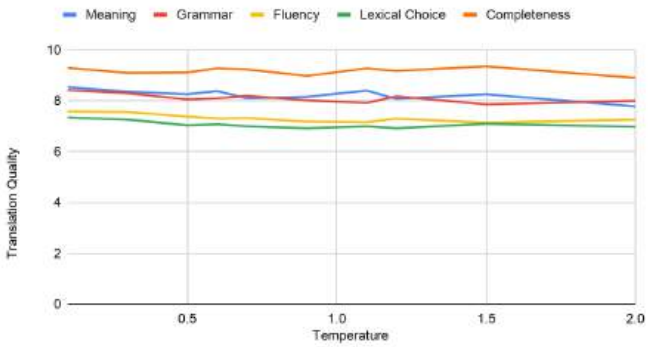


Figure 3.2: Illustration of a minimal sensitivity to temperature for YandexGPT 8B, EN→RU, RusLTC.

The results for Qwen3 8B (Figures 2.1 and 2.2) show a moderate sensitivity to temperature, with the performance decreasing by approximately 38% and 55%, respectively. In contrast, YandexGPT-5 Lite 8B (Figures 3.1 and 3.2) exhibits minimal sensitivity, with only a 6% and 5% reduction in performance between temperatures of 0.1 and 2.0.

Overall, the results demonstrate that temperature has a significant but model-dependent impact on translation quality. LLaMA 3.1 8B shows the highest sensitivity, with substantial performance degradation at higher temperature settings, while Qwen3 8B shows moderate sensitivity. In contrast, YandexGPT-5 Lite 8B remains mainly stable across the temperature range tested. Additionally, the effect of temperature varies between translation directions, with Russian-to-English translations generally showing less pronounced performance changes. These findings indicate that optimal temperature settings are both model-specific and task-dependent.

Effect of Temperature on Translation Stability

In addition to overall translation quality, the stability of model outputs is an important factor. Due to the inherent randomness aspect of LLMs, translations of similar text chunks may vary in quality. Therefore, models that consistently produce high-quality translations across many generations are considered more reliable. To quantify this variability, output stability was evaluated by measuring the standard deviation (SD) of translation scores at each temperature setting.

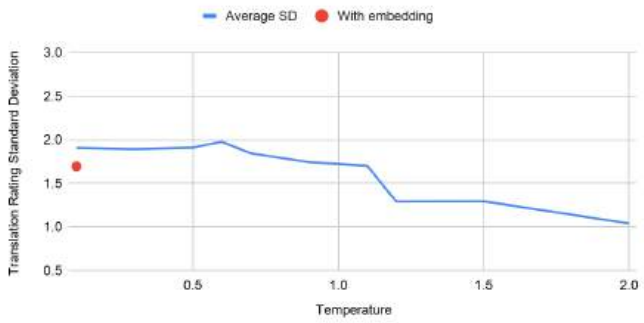


Figure 1.5: Relationship between temperature and the average SD across all evaluated translation categories for LLaMA 3.1 8B, EN→RU, Crime and Punishment, Chapter 2.

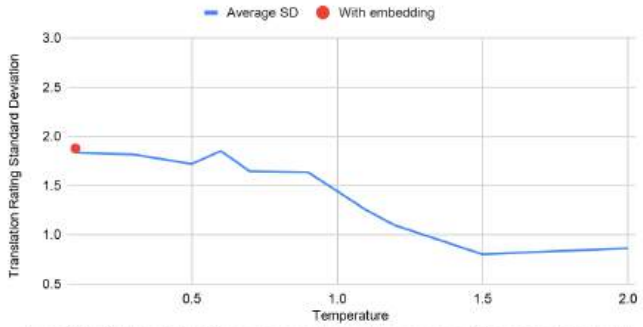


Figure 1.6: Relationship between temperature and the average SD across all evaluated translation categories for LLaMA 3.1 8B, EN→RU, RusLTC.

Figures 1.5 and 1.6 show the relationship between temperature (x-axis) and the average standard deviation across all evaluated translation categories (y-axis). For these cases, an inverse relationship is observed, where increasing temperature leads to a reduction in standard deviation (approximately 45% and 53%, respectively). This indicates that, for these samples, higher temperature settings result in more consistent outputs. It is important to note that LLaMA 3.1 8B exhibits a substantial decline in translation performance as temperature increases (Figures 1.1–1.4). As a result, the observed increase in output consistency at higher temperatures does not necessarily constitute an overall improvement, given the corresponding reduction in translation quality. The red markers represent the standard deviation when applying embedding-based selection. No consistent improvement or degradation in stability is observed, suggesting that this method does not have a systematic effect.

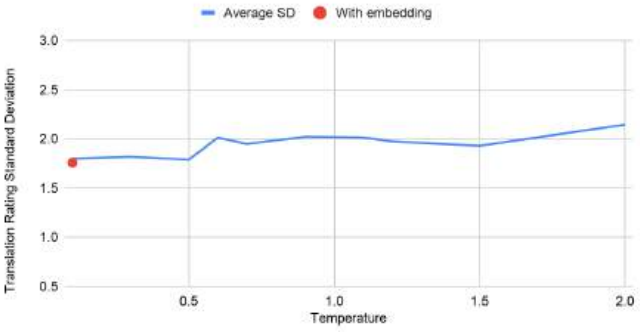


Figure 1.7: Illustrates increase of the SD as temperature rises for LLaMA 3.1 8B, RU→EN, Crime and Punishment, Chapter 2.

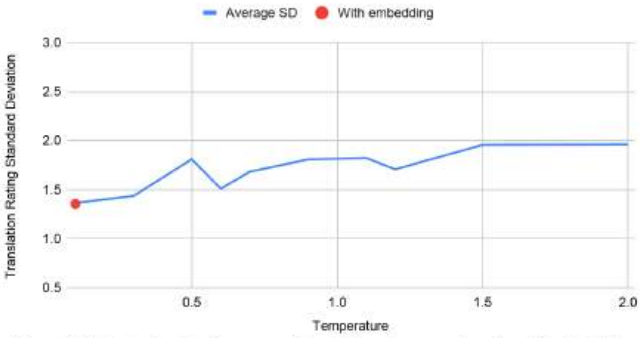


Figure 1.8: Illustrates the increase of the SD as temperature rises for LLaMA 3.1 8B, RU→EN, RusLTC.

In contrast, Figures 1.7 and 1.8 (Russian to English translation) exhibit the opposite trend, with standard deviation increasing as temperature rises. This corresponds to an increase of approximately 19% and 44%, respectively. This behavior aligns with expectations, as higher temperature values introduce greater randomness into the generation process. Similar to the previous case, embedding-based selection does not produce a consistent impact on stability.

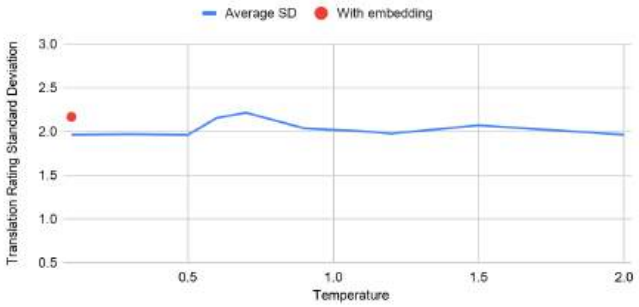


Figure 2.3: Presentation of the SD of translation scores for Qwen 3 8B, EN→RU, Crime and Punishment, Chapter 2.

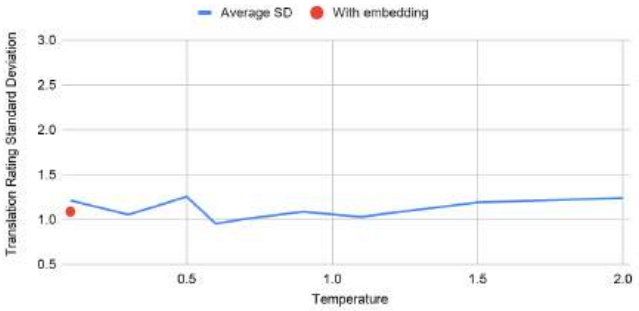


Figure 2.4: Presentation of the SD of translation scores for Qwen 3 8B, RU→EN, RusLTC.

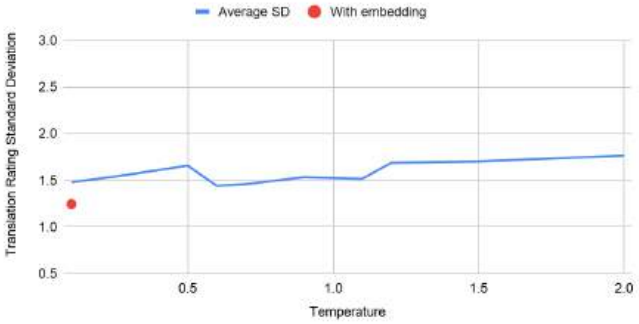


Figure 3.3: Presentation of the SD of translation scores for YandexGPT 8B, EN→RU, Crime and Punishment, Chapter 2.

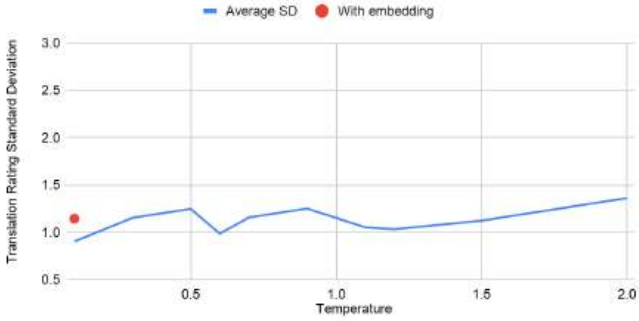


Figure 3.4: Presentation of the SD of translation scores for YandexGPT 8B, EN→RU, RusLTC.

Figures 2.3 and 2.4 present the standard deviation of translation scores across both translation directions and datasets for the Qwen 3 8b model. Across all cases, the standard deviation remains relatively stable as the temperature increases. This suggests that, for this model, temperature does not have a strong direct impact on the variability of the generated translations.

Figures 3.3 and 3.4 show a similar pattern for the YandexGPT model. As temperature increases from 0.1 to 2.0, the standard deviation exhibits an overall increase of approximately 19% and 51%, respectively. However, this change is not consistent, with fluctuations observed throughout the range. This indicated that, while a minor effect of temperature on output variability may be present, it is not strong.

Overall, the effect of temperature on output stability is not uniform across models or translation directions. English-to-Russian translation exhibits mixed behavior across models, including decreasing, stable, or weakly increasing variability depending on the model. In contrast, Russian-to-English translation shows either increasing variability or no significant change depending on the model. Across all experiments, embedding-based selection does not consistently improve or degrade output stability.

Effect of Embedding-Based Selection on Translation Performance and Stability

Embedding-based selection is intended to improve the consistency of generated translations by reducing the effects of the randomness aspect of LLMs. This approach selects the translation that is most semantically central among multiple generated translations of the same text. Its effectiveness was assessed by comparing translation performance and output stability, measured as the standard deviation (SD) of translation ratings, with and without embedding-based selection.

Table 1. Effect of embedding-based selection on LLaMA 3.1 8B performance and SD (temperature = 0.1).

Translation Direction	Dataset	Score change	SD change
EN to RU	Crime and Punishment	-1.0278	-0.21312
	RusLTC	-0.088	+0.042
RU to EN	Crime and Punishment	+0.18394	-0.04002
	RusLTC	-0.09818	-0.00898

Table 2: Effect of embedding-based selection on Qwen3 8B performance and SD (temperature = 0.1).

Translation Direction	Dataset	Score change	SD change
EN to RU	Crime and Punishment	-0.524	0.20508
	RusLTC	0.128	-0.10406
RU to EN	Crime and Punishment	0.11556	0.04342
	RusLTC	0.0618	-0.12512

Table 3: Effect of embedding-based selection on YandexGPT-5 Lite performance and SD (temperature = 0.1).

Translation Direction	Dataset	Score change	SD change
EN to RU	Crime and Punishment	-0.09702	-0.23384
	RusLTC	-0.124	0.2407
RU to EN	Crime and Punishment	-0.19558	0.01958
	RusLTC	0.10182	0.0237

Tables 1–3 present the effect of embedding-based selection on translation performance and output stability across all evaluated models. Across all models, datasets, and translation directions, the impact of this method is inconsistent, with both improvements and degradations observed. No clear or systematic trend is evident in either performance or stability metrics.

Comparison of Model Performance and Stability

Through the evaluation of translation performance across the tested configurations, it was found that similar general trends are present in all models. However, clear differences between the analyzed models were observed in overall performance, output consistency, and sensitivity to temperature.

For LLaMA 3.1 8B, Figures 1.1–1.4 indicate that at the lowest tested temperature (0.1), the average rating across both datasets and translation directions is approximately 5.95. This model also exhibits substantial sensitivity to temperature, with performance decreasing by an average of approximately 42% as the temperature increases from 0.1 to 2.0. At the most optimal temperature (0.1), the mean standard deviation of scores is approximately 1.73 across all conditions. This is higher than both other models tested, indicating that LLaMA 3.1 8B produces not only the lowest-quality translations but also less consistent outputs in terms of their quality.

In contrast, Qwen3 8B demonstrates improved performance and a smaller sensitivity to temperature. At a temperature of 0.1, the model achieves an average rating of approximately 6.74 across both datasets and translation directions. Its sensitivity to temperature is lower, with an average performance decrease of approximately 24% over the same temperature range (0.1 to 2.0). The average standard deviation for the model at temperature 0.1 is approximately 1.68, closely aligned with that of LLaMA 3.1 8B model. However, unlike LLaMA 3.1 8B, Qwen3 8B does not exhibit a notable relationship between temperature and output variability, as partly illustrated in Figure 2.4.

YandexGPT-5 Lite 8B demonstrates the strongest overall performance and highest output stability. At the optimal temperature (0.1), the average translation rating across both datasets and translation directions is approximately 7.43. This model also shows minimal sensitivity to temperature: performance decreases by only around 5% as temperature increases from 0.1 to 2.0.

Figure 4.1: Compares the average performance and average SD across models under the temperature=0.1.

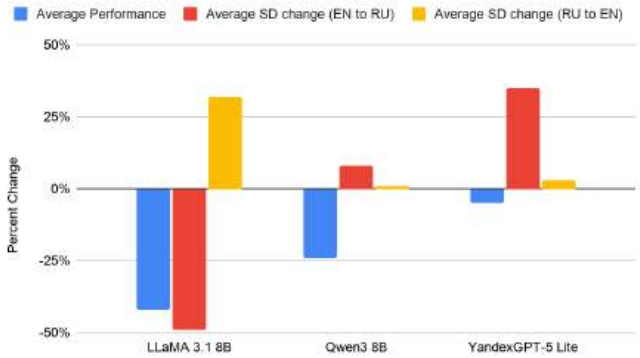


Figure 4.2: Illustrates performance and stability change across models.

The mean standard deviation at 0.1 is approximately 1.51. However, as temperature increases from 0.1 to 2.0, the standard deviation rises by approximately 19%, indicating a gradual reduction in stability with higher temperatures.

Figure 4.1 compares the average performance and standard deviation across all three models. LLaMA 3.1 8B demonstrates the lowest average performance and highest average standard deviation, indicating both weaker translation quality and lower stability. In contrast, YandexGPT-5 Lite achieves the highest average performance while also exhibiting the lowest standard deviation, indicating both strong output quality and stability. Qwen3 8B falls between the other two models for both metrics.

Figure 4.2 illustrates model sensitivity to temperature in terms of percentage change in both performance and standard deviation. As previously noted, LLaMA 3.1 8B exhibits the highest sensitivity across both metrics. Both Qwen3 8B and YandexGPT-5 Lite show substantially lower sensitivity overall. However, the change in standard deviation for YandexGPT-5 Lite in the English-to-Russian translation direction appears to be an outlier, showing a significantly larger increase than under other conditions. This suggests that higher temperatures may greatly reduce the output stability for this specific translation direction in this model.

Impact of Text Source on Performance

The text source and translation direction have a substantial impact on LLM translation performance. This study evaluated two datasets across both translation directions and found that dataset choice significantly influences both baseline translation quality and the model's sensitivity to temperature variation.

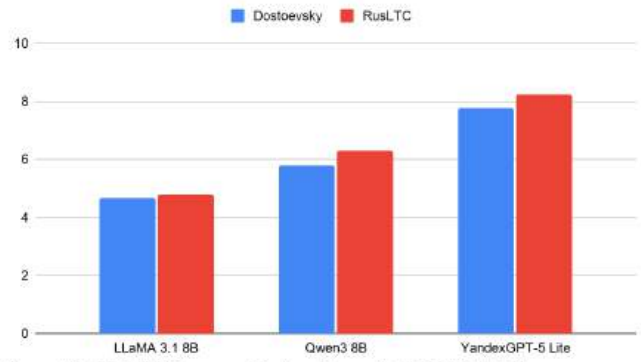


Figure 4.3: Model Peformance Dostoevsky vs. RusLTC, EN to RU.

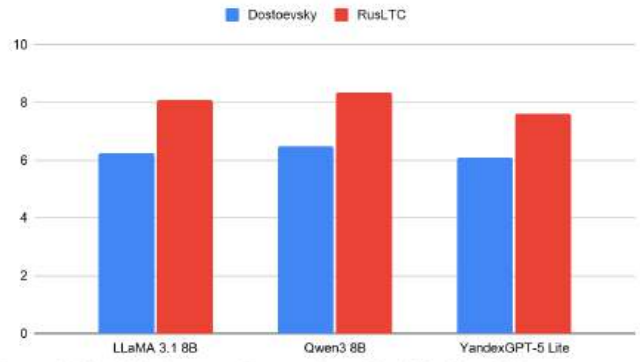


Figure 4.4: Model Peformance Dostoevsky vs. RusLTC, RU to EN.

As shown in Figures 4.3 and 4.4, all models consistently achieve higher performance when translating texts from the RusLTC dataset, indicating stronger performance in modern language compared to more classical or stylistically complex texts. This dataset-dependent performance gap is notably larger in Russian to English translation than in English to Russian translation.

The dataset also affects how performance degrades as temperature increases. Russian to English translations exhibit relatively low sensitivity to temperature changes, with an average performance decrease of approximately 9.7% when temperature rises from 0.1 to 2.0. In contrast,

English to Russian translations are significantly more sensitive to the temperature increase, with an average decrease of approximately 37.8% over the same range.

In addition to quality, the dataset influences output stability, often producing contrasting trends. As illustrated in Figures 1.5-1.8, 2.3-2.4, and 3.3-3.4, in English to Russian translations generally show decreasing or unchanged stability as temperature increases. Conversely, Russian to English translations tend to maintain or even improve stability under higher temperatures.

Overall, both translation direction and dataset selection play a critical role in determining translation quality, robustness to temperature variation, and output stability. These effects are not uniform across models, as different models exhibit varying degrees of sensitivity. In general, all tested models illustrate better performance when translating modern text.

Comparison of Model Speed

Inference speed is a key factor in assessing the practical usability of LLMs for translation tasks, particularly on consumer-grade hardware. To evaluate performance, three metrics were measured: time to first token (TTFT), time to last token (TTLT), and token generation throughput (tokens per second).

Table 4: Model speed comparison (temperature = 0.1).

Model	Metric	Min	Mean	Max
LLaMA 3.1 8B	TTFT (s)	0.103	0.215	3.167
	TTLT (s)	0.267	3.249	8.166
	Tokens/sec	86.01	90.85	99.04
Qwen3 8B	TTFT (s)	0.096	0.1855	3.153
	TTLT (s)	6.332	19.8025	333.618
	Tokens/sec	74.44	80.665	84.08
YandexGPT-5 Lite 8B	TTFT (s)	0.112	0.2055	1.52
	TTLT (s)	1.725	8.833	19.09
	Tokens/sec	30.14	31.065	32.26

All three models exhibit comparable TTFT values, indicating similar initial response latency. However, substantial differences can be seen in total generation time (TTLT) and throughput.

LLaMA 3.1 8B achieves the strongest overall performance, with the lowest mean TTLT (~3.2 seconds) and the highest average throughput (~90.9 tokens/sec), indicating both fast completion and efficient token generation. Qwen3 8B demonstrates lower throughput (~80.7 tokens/sec) and significantly higher mean TTLT (~19.8 s). Notably, its maximum TTLT (over

330 seconds) suggests the presence of severe outliers, which inflate the mean and indicate instability in generation time.

YandexGPT-5 Lite 8B shows more consistent behavior than Qwen3 8B but operates at substantially lower throughput (~31 tokens/sec), resulting in a higher mean TTLT (~8.8 seconds) despite the absence of extreme outliers.

Overall, while initial latency is similar across models, LLaMA 3.1 8B provides the best balance of speed and stability. Qwen3 8B suffers from high variance in generation time, and YandexGPT-5 Lite 8B is constrained primarily by low token generation throughput.

Conclusion

This study investigated the effects of temperature settings, model architecture, translation direction, and text source on the quality and performance of large language models in English–Russian and Russian–English translation tasks. The results showed that lower temperature settings generally produced higher-quality and more consistent translations. Russian-to-English translation tasks achieved higher average evaluation scores than English-to-Russian tasks, while modern texts yielded better translation quality than classical literary texts. Among the evaluated models, Yandex YandexGPT-5 Lite 8B produced the highest overall translation quality, whereas Meta LLaMA 3.1 8B demonstrated the fastest translation speeds.

This research may contribute to further work on adaptive inference strategies for large language models. While lower temperature settings were shown to improve translation quality, such settings are not necessarily optimal for tasks requiring creativity or stylistic variation. Based on these findings, this study proposes a dynamic temperature adjustment approach in which an LLM automatically detects translation tasks and temporarily lowers its temperature setting to improve translation accuracy and consistency. After the translation task is completed, the model could restore a higher temperature setting for tasks that benefit from more creative or diverse outputs.

Works Cited

[1] Alibaba Cloud. Qwen Technical Report. arXiv, 2024, https://arxiv.org/abs/2309.16609.

[2] OpenAI. GPT-5. OpenAI, 2025, https://openai.com/. Accessed March 18, 2026.

[3] Elshin, Denis, et al. “From General LLM to Translation: How We Dramatically Improve Translation Quality Using Human Evaluation Data for LLM Finetuning.” Proceedings of the Ninth Conference on Machine Translation, Association for Computational Linguistics, 2024, pp. 247–252. https://aclanthology.org/2024.wmt-1.17/. DOI: 10.18653/v1/2024.wmt-1.17.

[4] Wang, Liang, et al. “Multilingual E5 Text Embeddings: A Technical Report.” arXiv, 2024, https://arxiv.org/abs/2402.05672.

[5] Ollama. gpt-oss:120b-cloud. Ollama Library, 2026, https://ollama.com/library/gpt-oss/. Accessed March 18, 2026.

[6] Python. Python. Version 3.14, Python Software Foundation, https://python.org/. Accessed March 18, 2026.

[7] “Ollama.” Ollama, 2026, https://ollama.com/. Accessed March 18, 2026.

[8] Dostoevsky, Fyodor. Crime and Punishment. Translated by Constance Garnett, Project Gutenberg, 2006, https://www.gutenberg.org/ebooks/2554. Accessed March 18, 2026.

[9] “Russian Learner Translator Corpus (RusLTC).” RusLTC, https://www.rus-ltc.org/. Accessed March 18, 2026.

[10] Alibaba Cloud. Qwen3-8B. Hugging Face, 2025, https://huggingface.co/Qwen/Qwen3-8B. Accessed March 18, 2026.

[11] Ollama. llama3.1. Ollama Library, 2024, https://ollama.com/library/llama3.1. Accessed March 18, 2026.

[12] Yandex. YandexGPT-5-Lite-8B-instruct. Hugging Face, 2025, https://huggingface.co/yandex/YandexGPT-5-Lite-8B-instruct. Accessed March 18, 2026.

[13] Ribeiro, Felipe, and Hideyuki Takada. “Improving Translation Quality by Selecting Better Data for LLM Fine-Tuning: A Comparative Analysis.” ArXiv.org, 2025, https://arxiv.org/abs/2512.11388.

[14] Sundesh Donthi, et al. “Improving LLM Abilities in Idiomatic Translation.” ACL Anthology, 2025, pp. 175–181, https://aclanthology.org/2025.loreslm-1.13/.

[15] Renze, Matthew. The Effect of Sampling Temperature on Problem Solving in Large Language Models. 1 Jan. 2024, pp. 7346–7356, https://aclanthology.org/2024.findings-emnlp.432/, https://doi.org/10.18653/v1/2024.findings-emnlp.432.

[16] Yan, Jianhao, et al. “GPT-4 vs. Human Translators: A Comprehensive Evaluation of Translation Quality across Languages, Domains, and Expertise Levels.” ArXiv.org, 2024, https://arxiv.org/abs/2407.03658.

COMMODITY SUPERCYCLES
Gold, Silver & the Art of Strategic Positioning
A Comprehensive Investor's Guide to the Greatest Market Cycle of the Decade
By Nabil Chaker

I. What Is a Commodity Supercycle?

While most investors talk about bull markets, rate cycles, or momentum runs in individual stocks — a commodity supercycle is a completely different story. It is far rarer, far more powerful, and far more dangerous. Confusing it with an ordinary bull market is one of the most expensive mistakes an investor can make.

A commodity supercycle lasts around 10 to 30 years — a long time compared to most market phenomena. During that period, prices of commodities like metals, energy, and agricultural goods trade significantly above their long-term inflation-adjusted averages [1]. These aren't just spikes driven by a geopolitical shock or a bad harvest. These are deep, structural imbalances that last decades — where global demand for commodities far exceeds the industry's capacity to find, extract, and deliver them.

> *The critical distinction: a normal commodity bull market is a reaction to a temporary supply disruption or a short-term demand surge. A supercycle is a generational structural realignment of the entire physical economy.*

Research from institutions like Citigroup and the World Bank identifies four major supercycles since the beginning of the industrial era [2]. The first coincided with the Second Industrial Revolution in the late 19th century — electrification and industrialization across the US and Europe created explosive demand for steel, coal, and copper. The second was triggered by the reconstruction Europe and Japan underwent after World War II. The third ended with the OPEC oil embargoes of 1973 and 1979, which reshaped international energy pricing entirely. The fourth was the China-driven industrial surge of the 2000s.

The Four Phases of a Supercycle

Phase One — Accumulation — happens in the aftermath of an earlier bust. Prices are low, mining companies cut capital expenditure because margins have collapsed, and the seeds of the next cycle are quietly planted without most people noticing. Phase Two — Early Expansion — is when a new, large, lasting demand driver emerges while supply is still constrained. This is historically the best window to enter. Phase Three — the Mid-Cycle Bull — is when the cycle becomes widely recognized. The media starts covering it, capital floods back in, and it becomes the consensus trade. Phase Four — the Blow-Off and Contraction — arrives when prices get high enough that companies start investing heavily in new supply and consumers begin switching to alternatives. The gap between demand and supply slowly closes, and a long bear market follows.

II. The Current Supercycle — What Is Driving It?

The China-driven fourth supercycle had one engine. The current cycle — Supercycle 2.0 — has three simultaneous and independent long-term demand drivers, each one large enough to sustain a commodity bull market on its own, without any help from the other two [4].

Engine One: The Global Energy Transition

Electric vehicles contain significantly more copper than conventional cars — battery electric vehicles average around 70 kilograms of copper today, compared to just 23 kilograms in a conventional internal combustion engine vehicle [5]. EV sales are projected to reach 73 million units annually by 2030 [5]. Every gigawatt of installed solar capacity requires approximately 5 to 6 tonnes of silver for photovoltaic cell paste alone [6]. The International Energy Agency estimated in 2022 that a pathway consistent with net-zero emissions by 2050 would require a fourfold increase in overall mineral demand by 2040 — and a 40-fold increase in lithium demand specifically [7]. These are not aspirational targets. They are the physical minimum requirements of the energy transition.

Engine Two: The Artificial Intelligence Infrastructure Buildout

A hyperscale data center is an enormous facility packed with servers, storage systems, and networking equipment — the kind operated by Amazon, Google, and Microsoft. To put the scale in perspective: 100 megawatts is enough electricity to power tens of thousands of homes, and these facilities can consume up to 300 megawatts each — the equivalent of a small city. Microsoft committed $80 billion to AI data center infrastructure in fiscal year 2025 alone [8]. Meta announced capital expenditure plans of $60 to $65 billion for the same year [8]. McKinsey projected in 2024 that data center power demand in the US alone could grow between 15 and 22 percent annually through 2030 [9]. Every single one of these facilities requires extensive copper wiring, electrical steel, and cooling infrastructure — all landing on an already constrained supply base.

Engine Three: Geopolitical Fragmentation and Strategic Stockpiling

The United States introduced the CHIPS and Science Act and the Inflation Reduction Act, both of which support the development of domestic mineral processing. In 2023, the European Union released its Critical Raw Materials Act, classifying 34 materials as strategically important [10]. China already controls approximately 60% of global rare earth processing and 50% of global lithium chemical production. Nations are now active buyers of critical minerals alongside industrial consumers — fundamentally altering the supply-demand balance in ways the market has never seen before [10].

The Supply-Side Crisis: A Decade of Self-Inflicted Wounds

From 2012 to 2022, corporate boards in the resource sector were actively punished by Wall Street for expanding production. Shareholders demanded dividends and share buybacks rather than capital expenditure for new mines [11]. At the same time, ESG mandates choked off

institutional capital from the extractive industries — the very metals the energy transition would later desperately need.

The geological reality only makes it worse. Copper ore deposits contain far less copper than they used to. In the 1990s, every 100 kilograms of ore contained around 1.5 kilograms of copper. Today, that same 100 kilograms yields only around 0.5 to 0.6 kilograms. Miners now have to dig more rock, transport more material, and process more ore — roughly three to four times more — just to produce the same amount of copper. On top of that, new mines take 10 to 15 years from discovery to first production [11].

> *The profound irony is hard to ignore: the ESG movement — which defunded mining in the name of environmental responsibility — directly contributed to the supply shortages now making the metals the energy transition needs dramatically more expensive*

III. Why Gold Ran — And Why It Broke Its Own Rules

For roughly 40 years, gold traded as an inverse function of US real interest rates — the yield on US Treasury Inflation-Protected Securities (TIPS). When real yields rose, gold fell; when they fell, gold rose. The logic was sound: gold pays no dividend, so rising real rates increase its opportunity cost relative to bonds [12].

Between 2022 and 2026, that 40-year relationship broke completely. The Fed raised rates from basically zero to 5.25%–5.50% — the most aggressive hiking cycle in four decades. Real yields shot up to levels nobody had seen since 2007. Every model, every analyst, every historical pattern said gold should have been tanking. Instead, it did the opposite — climbing from around $1,800 an ounce in early 2022, breaking $3,000 for the first time in history in early 2025, and reaching an all-time high of $5,589 per ounce on January 28, 2026 [13]. As of June 2026, it trades around $4,200 after pulling back from those highs. Something had fundamentally changed.

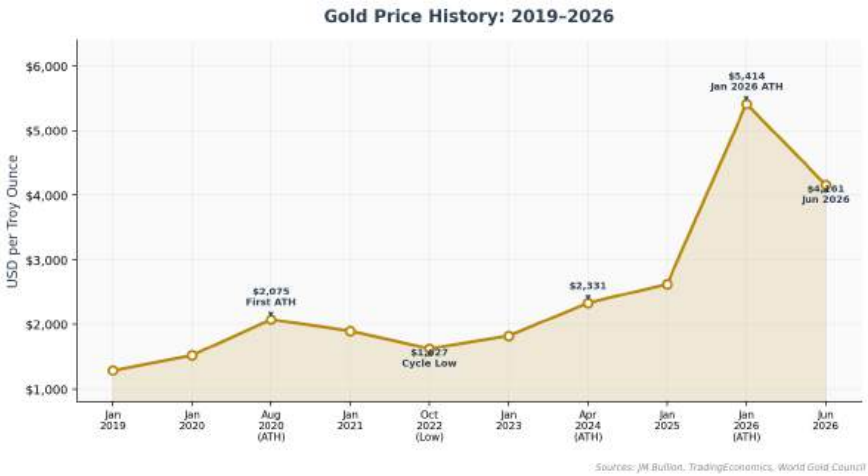


Chart 1: Gold Price History 2019–2026 | Sources: JM Bullion, TradingEconomics, World Gold Council

The Sovereign Put: Central Banks Rewrite the Rulebook

In February 2022, western countries froze Russia's central bank reserves — around $300 billion stored in western financial institutions became inaccessible overnight. It proved that even national reserves can be blocked, and that holding money in foreign financial systems carries real political risk.

The message to every other central bank was unmistakable: fiat reserves held abroad carry seizure risk. Physical gold held in your own vault carries none [14]. According to the World Gold Council, central banks purchased 1,082 tonnes of gold in 2022 — the highest level since 1950, and a 152% increase over the prior year [15]. In 2023, purchases reached 1,037 tonnes — the second-highest annual total on record. In 2024, central banks added approximately 1,045 tonnes — the third consecutive year above 1,000 tonnes [15]. In 2025, purchases moderated to 863 tonnes as record prices prompted a more measured approach, but 23 countries still added to reserves in the first half of 2025 alone, and the People's Bank of China extended its consecutive monthly buying streak to 19 months as of May 2026 [16].

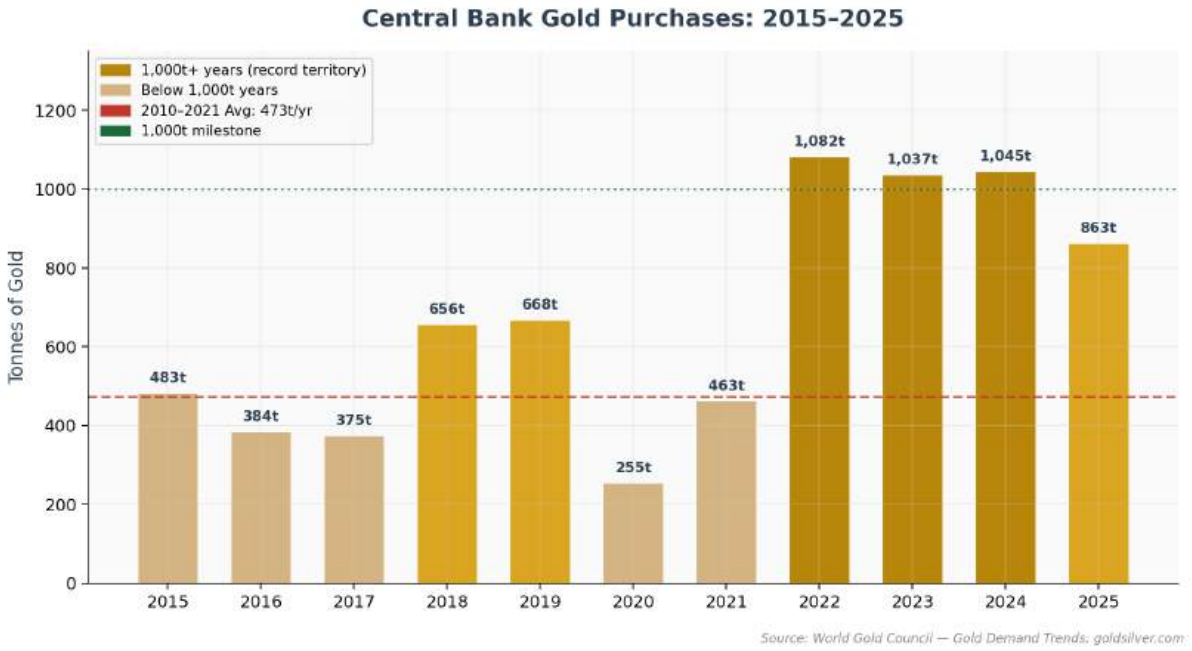


Chart 2: Central Bank Gold Purchases 2015–2025 | Source: World Gold Council — Gold Demand Trends

The Fiscal Mathematics of Dollar Debasement

The US national debt crossed $33 trillion in 2023 and kept climbing — reaching over $37 trillion by mid-2026. Since the COVID pandemic, the US government has been running annual deficits of $1.5 to $2 trillion. By 2025, the annual interest expense on that debt was approaching $1 trillion — consuming a growing share of federal tax receipts [17]. Historically, deficits of this scale and persistence have only ever been resolved one way: currency debasement. Institutional

investors began factoring this into their portfolio construction, driving gold from a fringe alternative asset into a mandatory cornerstone.

Milestone	Details
Gold, January 2020	~$1,520 per troy ounce
Gold, April 2024	~$2,331 per troy ounce — new ATH at the time
Gold, January 2025	~$2,624 per troy ounce (start of 2025)
Gold, early 2025	Broke $3,000 — first time in history
Gold, January 28, 2026	$5,589 per troy ounce — confirmed all-time high
Gold, June 2026	~$4,186–$4,216 (correcting from highs)
Central bank buying, 2022	1,082 tonnes — highest since 1950
Central bank buying, 2023–2024	1,037 and ~1,045 tonnes respectively
JPMorgan Q4 2026 forecast	$6,000 per troy ounce average

IV. Why Silver Is Trickier — The Devil's Metal

Silver occupies a uniquely uncomfortable position. It is roughly 50% industrial commodity and 50% monetary asset. This split personality makes it one of the most volatile tradeable assets on Earth — capable of extraordinary returns in the right phase of the cycle, and capable of equally savage drawdowns at the worst possible moments [18].

The Industrial Side: Silver's Irreplaceable Physical Properties

Silver conducts electricity better than any other metal on Earth. That single property makes it irreplaceable in a wide range of modern applications. The solar photovoltaic industry is the largest industrial consumer of silver — accounting for approximately 29% of total industrial silver demand in 2024, up from just 11% in 2014 [18]. Beyond solar, silver is heavily used in electronic components, AI server hardware, and the autonomous systems built into electric vehicles. The Silver Institute reported that industrial fabrication demand exceeded 680 million ounces in 2024 — a record — driven by the green economy's growing appetite across solar, automotive, and grid infrastructure sectors [19].

The Monetary Side: The Poor Man's Gold

When gold makes headlines with record prices, retail capital eventually rotates toward silver, seeking the same inflation protection at a fraction of the cost [20]. Silver hit its own all-time high of $121.67 per ounce on January 29, 2026 — one day after gold's record peak — before correcting sharply to around $67 per ounce by June 2026. This kind of violent swing is

exactly what silver is known for: extraordinary upside in the right conditions, and brutal corrections when sentiment shifts.

Historically, one ounce of gold has been worth around 47 to 65 ounces of silver across the 20th century, though the modern average since 2000 sits closer to 60 to 65:1. During the COVID-19 crash in March 2020, that ratio hit approximately 125:1 — the most extreme silver undervaluation in modern market history. By late January 2026, during silver's ATH, it had compressed to around 46:1. As of June 2026, following silver's correction, the ratio stands at approximately 62:1 [21].

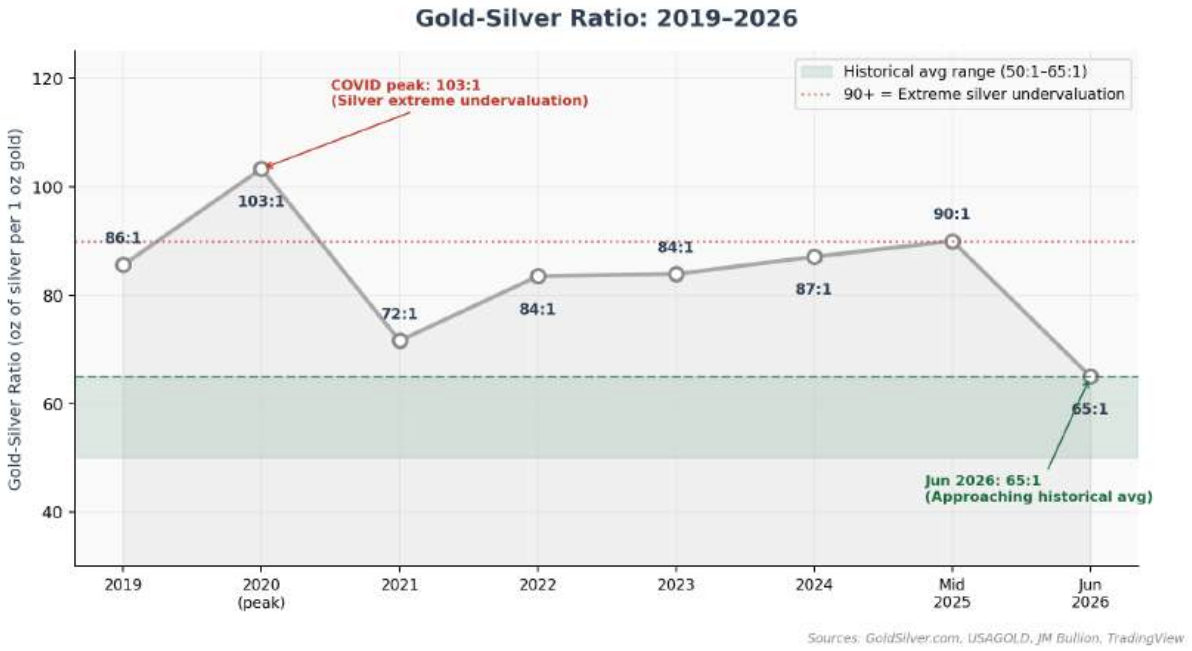


Chart 4: Gold-Silver Ratio 2019–2026 | Sources: GoldSilver.com, USAGOLD, JM Bullion

The Supply Inelasticity Trap

Over 70% of the world's annual silver production comes as a byproduct of mining other metals — primarily copper, zinc, and lead. This creates extreme supply inelasticity. When silver prices triple, a zinc miner is not going to spend $2 billion building a new zinc mine just to access some byproduct silver. Silver supply simply cannot respond to its own price signals [18]. The Silver Institute reported that global silver markets have been in a structural supply deficit for six consecutive years through 2026. The cumulative deficit from 2021 through 2026 is projected to reach approximately 762 million ounces — equivalent to roughly 10 months of total global mine supply [22].

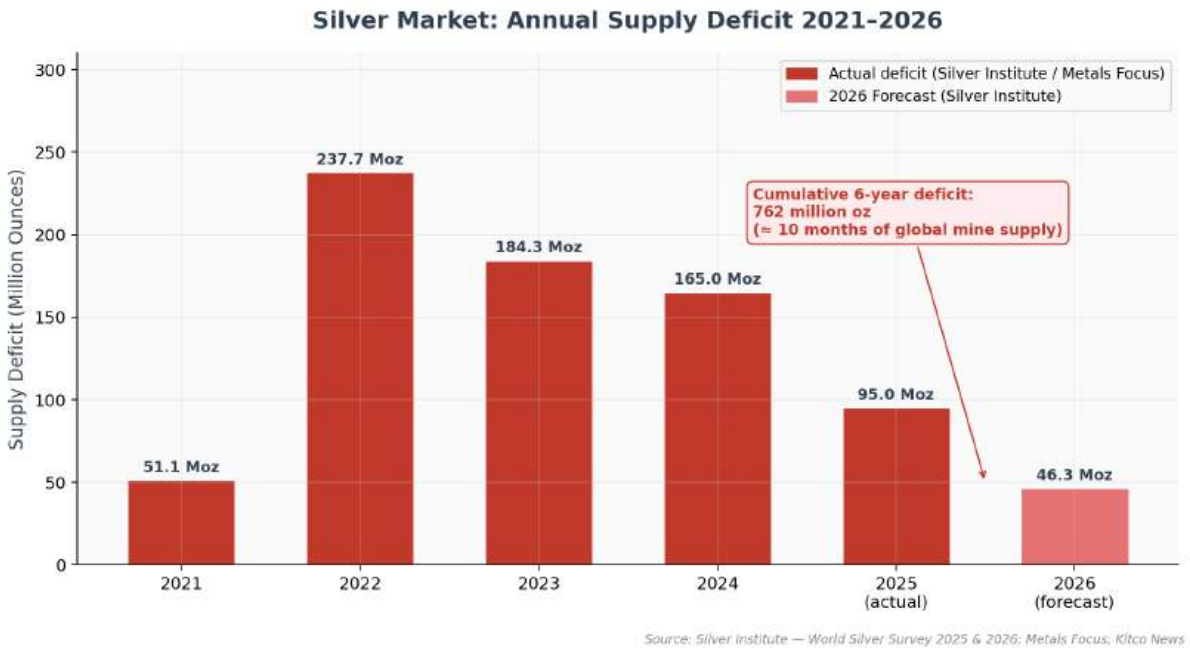


Chart 3: Silver Annual Supply Deficit 2021–2026 | Source: Silver Institute — World Silver Survey 2025 & 2026

Silver's Unique Vulnerability: The Dual-Sided Selloff

Silver's most dangerous characteristic is its ability to fall hard from both sides at the same time. If a global recession hits, industrial demand collapses — silver gets sold like a base metal. If investors simultaneously flee to the safety of gold specifically, monetary demand evaporates too. During the March 2020 pandemic crash, gold fell approximately 12%. Silver fell approximately 35% in the same period. During the 2008 financial crisis, gold fell roughly 30% — but silver fell nearly 50% [23]. The correction from $121 to $67 between January and June 2026 illustrates exactly this dynamic: rising rate expectations hit industrial demand expectations while gold's correction pulled monetary demand lower simultaneously.

V. How to Position Properly in a Commodity Supercycle

Recognizing that a commodity supercycle is underway is the first challenge. Translating that correct macro view into a portfolio that actually captures the returns — without being destroyed by volatility, operational risks, or structural pitfalls — is the second. Being directionally correct is not enough.

The Futures Trap: Understanding Contango and Roll Decay

Commodity ETFs that hold futures contracts must roll their positions — selling expiring near-month contracts and buying the next month's contracts to maintain exposure. In a contango market, where futures prices are higher than spot prices, this means consistently selling lower and buying higher. Over multiple years, the total return of a futures-backed ETF can lag the spot

price return of the underlying commodity by 20% to 30% or more [24]. This is not a theoretical risk — it is a documented structural feature of futures-based commodity products. The right vehicle for investors who want gold or silver exposure without this structural drag is physical metal: either physical coins and bars held in allocated vaults, or ETFs backed by audited, allocated physical bullion.

Mining Equities: The Theory vs. The Reality

The theoretical operating leverage of mining equities is appealing. In practice, three factors systematically erode it. First, cost inflation: in a commodity supercycle, diesel, explosives, steel, and skilled labor costs all rise — eating the expected margin expansion. Second, jurisdiction risk: as metal prices rise, governments in mining regions impose new royalties, retroactive tax increases, or in extreme cases nationalize assets outright. Third, operational complexity: grade surprises, equipment failures, water management issues, and labor disputes all materially impact production schedules and costs [25].

The Barbell Strategy: A Framework for Sophisticated Exposure

The Core: Physical Metals

The base of the barbell is physical precious metal — gold and silver — held in allocated, segregated accounts at accredited bullion custodians. This serves as the counterparty-risk-free anchor of the portfolio. It participates directly in metal prices with zero roll decay, no default risk, and no ongoing management requirement.

The Alpha: Royalty and Streaming Companies

Royalty and streaming companies — such as Franco-Nevada Corporation and Wheaton Precious Metals — provide upfront capital to mining companies in exchange for the right to purchase a fixed percentage of future production at a predetermined, below-market cost. Because they do not operate mines themselves, they are completely insulated from diesel and labor cost inflation. Their margins expand directly with metal prices, without the offsetting cost increases that crush traditional miners [25].

The Select Few: Tier-1 Producers in Safe Jurisdictions

Carefully selected Tier-1 mining producers — the largest, most established companies operating in stable, transparent jurisdictions like Canada, the USA, and Australia — offer operating leverage with significantly reduced jurisdiction risk. The selection criteria matter: proven reserves supporting a 10-plus-year mine life, current free cash flow generation, diversified multi-mine production, conservative debt levels, and a documented track record of delivering projects on time and on budget [25].

VI. How to Catch the Next Supercycle

Every supercycle is born in the ruins of the prior bust. When a cycle ends and prices collapse, institutional capital evacuates. Mining companies cut CapEx. Exploration programs are suspended. Supply capacity quietly erodes. The seeds of the next cycle are always planted in the despair of the last one.

The Early Warning Signals

The first signal is multi-year, industry-wide capital expenditure contraction below the industry's depletion rate — the point at which the industry is consuming its own future supply capacity faster than it is replacing it. The second is extreme valuation discounts in mining equities: between 2015 and 2020, major gold mining companies traded at Price-to-Net Asset Value ratios below 0.5x, meaning the market valued the gold in the ground at less than half its current spot price — a signal of genuine capitulation [26].

The third signal is physical inventory depletion at major exchanges — the LME, COMEX, and SHFE. The quiet drawdown of registered inventories ahead of the current commodity rally was one of the clearest early warning signals available to those watching. The fourth: the emergence of a new, large-scale structural demand driver not yet reflected in consensus price forecasts. Investors who read the IEA's mineral demand reports and Goldman Sachs's early energy transition commodity research in 2019 and 2020 had all the information they needed to identify the coming supercycle before it launched [7].

Monitoring This Cycle's Health

Physical inventory levels at commodity exchanges are the most reliable real-time indicators of the supply-demand balance. When inventories begin meaningfully rebuilding after extended drawdowns, it signals that supply is catching up to demand and the structural imbalance is beginning to close. When commodity markets flip into backwardation — meaning the spot price today is higher than the futures price — industrial buyers are signaling immediate physical scarcity and willingness to pay a premium for metal now rather than later [27]. Watch the decisions of large industrial consumers closely. When major manufacturers announce programs to reduce the copper or silver intensity of their products, those are the earliest signs of demand destruction — and the first indication that the structural deficit is starting to close [27].

VII. Common Investor Mistakes in Commodity Supercycles

Chasing the Second Trade

One of the most reliably destructive patterns in commodity investing is what can be called the second trade problem. After capturing a commodity's initial move, the natural next question is: what is the next commodity that will do the same thing? The problem is that by the time a commodity has generated exceptional returns, the capital to be made from that repricing has largely already been made. Late entry combines early-phase volatility risk with late-phase

valuation risk — without the discount that compensates for taking on that exposure in the first place.

Confusing Volatility with Opportunity

Silver can move 5% or more in a single day during high-activity periods. Attempting to trade around that volatility almost always underperforms simply holding the structural position with conviction. The investors who capture the best returns in commodity supercycles are those who establish positions based on a genuine understanding of the thesis, size them appropriately for the asset's volatility, and hold through interim corrections. The China supercycle of the 2000s included multiple corrections of 30% or more in copper prices. The full-cycle return from 2001 to 2011 was extraordinary for those who stayed in [28].

Over-Concentration in a Single Commodity or Producer

Supercycles do not lift all commodities equally. Lithium prices entered a severe multi-year correction beginning in late 2022, despite the electric vehicle demand thesis remaining entirely intact. Nickel collapsed dramatically in 2023 due to a surge in Indonesian supply that overwhelmed demand growth. Investors who were over-concentrated in either of these commodities experienced devastating losses, despite being correct about the broader supercycle [28].

Ignoring the Exit

By the time a supercycle is approaching its peak, the structural thesis typically looks most compelling and the consensus is most bullish — precisely when it is hardest to sell. Establishing rules-based exit frameworks before positions become highly profitable is essential. Practical triggers include: physical inventories recovering at major exchanges, futures markets returning to deep contango, mining CapEx decisively surpassing depletion rates, and major industrial consumers announcing successful substitution programs.

VIII. What Could End This Cycle?

Scenario One: The Hard Landing

A severe global recession would collapse industrial demand across all commodity categories. Energy transition and AI buildout projects would slow as corporate and government budgets tighten. This scenario would produce a sharp cyclical correction within the supercycle — but it would not permanently end the structural thesis. Demand deferred is not demand destroyed.

Scenario Two: Technological Disruption

Solid-state battery technologies, if commercialized at scale, could significantly alter the lithium, cobalt, and nickel demand profiles for electric vehicles. A breakthrough in perovskite

solar cell technology could reduce silver paste requirements in photovoltaics. These risks are real — but historically, such technologies materialize at commercial scale far more slowly than initial announcements suggest. Copper intensity per EV is already declining due to engineering improvements, falling from roughly 99kg per vehicle in 2015 to around 70kg today — but total EV volume growth has so far outpaced the per-vehicle reduction [5].

Scenario Three: Geopolitical Resolution

A substantial reduction in great-power geopolitical tensions, or a new framework for global supply chain cooperation, could reduce the urgency of sovereign gold stockpiling and strategic mineral reserve building. This scenario seems unlikely in the near term, given the structural nature of the US-China strategic competition and the persistent geopolitical tensions that drove the de-dollarization trend.

Scenario Four: The Supply Response

The most historically reliable cycle-ending mechanism: decade-long high prices eventually incentivize enough new mine development that supply begins to outpace demand. The lead times are long — 10 to 15 years for a greenfield mine — but the capital commitments made since 2022 mean a meaningful new supply wave is eventually coming. When aggregate mining CapEx begins decisively surpassing aggregate depletion, the countdown has begun [11].

IX. Key Indicators to Track

Successful commodity supercycle investing requires tracking the physical market data that tells you what is actually happening in the real economy — not the daily price ticker.

Indicator	What It Tells You
LME, COMEX & SHFE Physical Inventories	Most direct real-time measure of supply-demand balance. Sustained drawdowns confirm structural deficit. Rebuilding inventories signal the cycle may be peaking.
Futures Curve Shape	Backwardation (spot > futures) = physical scarcity; industrial buyers desperate for immediate delivery. Contango = near-term supply adequate.
Gold-Silver Ratio	Compression below 60:1 signals a late precious metals bull cycle with silver catching up. Expansion above 90:1 suggests potential silver undervaluation vs gold.
TIPS Real Yield (10-Year)	Gold's 40-year pricing anchor. Gold holding firm despite rising real yields confirms sovereign demand — not speculation — is now the dominant force.
Global M2 Money Supply Growth	Rising M2 historically precedes broad commodity appreciation as liquidity seeks hard assets.

US Dollar Index (DXY)	Dollar weakness amplifies commodity price gains. Dollar strength acts as a temporary headwind.
Central Bank Gold Buying (WGC)	Monthly World Gold Council data. Is sovereign accumulation accelerating or reversing? The single most important driver of gold's structural breakout.
Mining Industry CapEx vs. Depletion	CapEx below depletion = future supply contracting (bullish). CapEx above depletion = supply pipeline filling (begin planning exit).
Solar PV Installation Data (IEA)	Primary structural driver of silver industrial demand growth. Annual GW installed × silver content per GW = medium-term silver demand trajectory.
Industrial R&D / Substitution Announcements	Earliest warning sign of a peaking supercycle. Successful commercialization of metal substitutes begins closing the structural deficit.

X. Conclusion: The Long View

Commodity supercycles are the longest, least understood, and most poorly navigated of all major market cycles. They move too slowly for short-term traders, too violently for conservative investors, and with too much complexity for those who approach them with oversimplified frameworks. They reward patience, intellectual rigor, and the discipline to maintain conviction through corrections that would shake most investors out of their positions.

The current cycle has a structural foundation that is, by any objective measure, extraordinarily robust. The demand side is supported by three simultaneous, government-mandated, capital-backed structural transitions — the energy transition, the AI infrastructure buildout, and the geopolitical shift toward strategic resource security. The supply side is constrained by a decade of underinvestment, accelerating geological grade decline, and 10-to-15-year mine development lead times that make rapid supply response structurally impossible.

Gold has already demonstrated the power of this cycle by breaking every historical pricing model governing its behavior for 40 years, driven by a fundamental change in who is buying it and why. Silver carries the most asymmetric risk-reward profile of any asset in the cycle — capable of extraordinary gains in the right phase, as its January 2026 run to $121 demonstrated, but equally capable of sharp corrections when rate expectations shift or industrial sentiment sours.

> *The investors who will benefit most from this supercycle are not those who trade it most cleverly. They are those who understand it most clearly — and hold on.*

Works Cited

[1] Heap, A. (2005). China — The Engine of a Commodities Super Cycle. Citigroup Smith Barney Research.

[2] World Bank (2021). Commodity Markets Outlook: Causes and Consequences of Metal Price Shocks. Washington, D.C.: World Bank Group.

[3] Rogers, J. (2004). Hot Commodities: How Anyone Can Invest Profitably in the World's Best Market. Random House.

[4] Goldman Sachs Global Investment Research (2021). Copper Is the New Oil. New York: Goldman Sachs.

[5] Bloomberg New Energy Finance (2024). Electric Vehicle Outlook 2024. Copper intensity data: Benchmark Mineral Intelligence (2024), EV copper intensity declining from ~99kg (2015) to ~70kg (2024) to ~62kg (2030 forecast). Visual Capitalist, September 2024.

[6] Silver Institute / Metals Focus (2025). World Silver Survey 2025. Washington, D.C.: Silver Institute. Silver per GW of solar PV data.

[7] International Energy Agency (2022). The Role of Critical Minerals in Clean Energy Transitions. Paris: IEA. 4x overall and 40x lithium demand projections.

[8] Reuters / Financial Times (2025). Microsoft ($80bn) and Meta ($60–65bn) AI CapEx announcements, FY2025.

[9] McKinsey & Company (2024). Powering the AI Economy: Data Center Energy Demand Through 2030. McKinsey Global Institute. 15–22% annual US data center power demand growth.

[10] European Commission (2023). Critical Raw Materials Act. Brussels: European Union. 34 strategic materials classified. US Inflation Reduction Act (2022) and CHIPS and Science Act (2022), US Congress.

[11] S&P Global Market Intelligence (2023). The Future of Copper: Will the Looming Supply Gap Short-circuit the Energy Transition? S&P Global. CapEx contraction, ore grade decline (1.5% → 0.5–0.6%), 10–15 year mine development lead times.

[12] World Gold Council (2023). Gold and Interest Rates: Revisiting the Relationship. London: World Gold Council. Historical TIPS/gold inverse relationship.

[13] Gold all-time high of $5,589.38 on January 28, 2026: Investing News Network (February 2026); CBS News (February 2026); Fortune daily gold price reports. Current price ~$4,186–$4,216 as of June 12–14, 2026: CNBC Gold Spot Price; USAGOLD daily report; Fortune.

[14] US Treasury Department / European Commission (2022). Statement on Immobilization of Russian Central Bank Reserves (~$300bn). February–March 2022.

[15] World Gold Council (2024). Gold Demand Trends Full Year 2023 and Q4 2024. gold.org. Central bank purchases: 2022 (1,082t, highest since 1950, +152% YoY), 2023 (1,037t), 2024 (~1,045t).

[16] World Gold Council / BullionVault (2025–2026). Central bank buying data H1 2025: 23 countries adding reserves. PBoC 19-month consecutive buying streak confirmed June 12, 2026 (USAGOLD daily report).

[17] US Congressional Budget Office (2024). The Budget and Economic Outlook: 2024 to 2034. US national debt: $33 trillion (2023), $37+ trillion (mid-2026, US Treasury Fiscal Data). Annual interest approaching $1 trillion.

[18] Silver Institute / Metals Focus (2026). World Silver Survey 2026. Washington, D.C.: Silver Institute. Solar PV share of industrial demand (29% in 2024 vs 11% in 2014), supply inelasticity, byproduct mining (70%+).

[19] Silver Institute (2025). Silver Industrial Demand Reached a Record 680.5 Moz in 2024. silverinstitute.org. Record electronics and electrical demand.

[20] Silver ATH $121.67 on January 29, 2026: APMEX Silver Price page (June 2026). June 2026 silver price ~$67.70: TradingEconomics, USAGOLD daily report, June 12, 2026. J.P. Morgan 2026 silver average forecast: $81/oz (JPMorgan Global Research).

[21] Gold-silver ratio data: GoldSilver.com ratio charts; USAGOLD daily report June 12, 2026 (ratio ~62.3:1). COVID peak ~125:1 (March 2020): Silver Institute. January 2026 compression to ~46:1 at silver ATH: APMEX.

[22] Silver Institute / Metals Focus (2026). World Silver Survey 2026. Cumulative 6-year deficit 2021–2026: ~762 million oz. Investing.com analysis, April 30, 2026.

[23] World Gold Council / Silver Institute. Gold and Silver Price Performance During Market Crises: 2008 (gold −30%, silver −50%) and March 2020 (gold −12%, silver −35%).

[24] iShares / Invesco Research (2021). Understanding Commodity ETF Roll Yield and Contango Effects. 20–30%+ long-term lag of futures ETFs vs spot in contango markets.

[25] Franco-Nevada Corporation (2024). Annual Report 2024. Toronto: Franco-Nevada. Wheaton Precious Metals (2024). Annual Report 2024. Royalty/streaming model mechanics and jurisdiction risk mitigation.

[26] Van Eck Associates (2020). Gold Mining Equity Valuations at Historical Lows. GDX research. Price-to-NAV below 0.5x during 2015–2020.

[27] London Metal Exchange (2024). LME Market Data: Inventory and Forward Curve Reports. lme.com. Backwardation/contango methodology.

[28] Rogers International Commodity Index. Copper Price Corrections Within the 2001–2011 Bull Market. Multiple 30%+ corrections documented within the fourth supercycle.

To What Extent Do Existing Financial Literacy Policies in Brazil, South Africa, and India Address the Structural Barriers to Financial Inclusion? By Amara Singh

Abstract

The global importance of financial literacy is particularly rising in emerging economies due to its direct impact on broader economic development, household resilience and financial inclusion. Despite existing policy efforts, financial literacy levels in Brazil, South Africa, and India still remain low and unevenly distributed and this highlights how institutional fragmentation limits effectiveness of existing initiatives. The paper's approach is a qualitative cross-country comparative analysis and it draws on policy documents, regulatory frameworks and existing academic literature. The paper proposes targeted, country-specific institutional and policy reforms which can positively impact financial inclusion. The paper adds to existing literature through a structured comparative framework to evaluate financial literacy initiatives in emerging markets, highlighting its relevance for policymakers in BRICS and other developing economies.

Keywords Financial Literacy; Government & Financial Institutions; Behavioral Economics; Informality; Financial Inclusion

1. Introduction

Financial literacy is a vital aspect of every human being's life and is necessary for a stable, steady and secure long-term financial standing. Not only does it help one manage the basics of spending and borrowing but also supports future retirement plans or possible entrepreneurial interests. Considering the importance of this subject, financial literacy is a key policy priority for countries across the globe since it helps one avoid poor financial decisions, navigate economic uncertainties and most importantly, has a major impact on the country's economic growth. Financial inclusion, household income and formal financial participation are also positively influenced by financial literacy. For large and emerging economies, like India, South Africa and Brazil, it is imperative to focus on the institutional policies around Financial Literacy, keeping in mind the country's inequality levels, behavioral gaps and informality. Another comparative factor is the digital public infrastructure in each 3 countries and how it is compared to financial literacy[13].

Several papers document the importance of financial literacy and portray it as a form of human capital. Like human capital, financial literacy is an intangible asset that forward-looking agents can accumulate over time[14]. Moreover, they largely focus on single-country evidence and household-level incomes, with less focus on institutional and policy mechanisms and cross-country comparisons. Low levels of Financial Literacy rates, its determinants along with emerging digital financial products are well-defined areas within this field. However, this paper delves into comparisons of countries under similar living standards and development conditions,

[13] *Digital Public Infrastructure and the Future of Digital Payments: Lessons from Pix and UPI*

[14] Bucci, A. (2015). Product Proliferation, Population, and Economic Growth. *Journal of Human Capital*, *9*(2), 170–197

while addressing the different constraints, limitations and challenges that arise in each country in terms of institutional capacity and policy design.

This paper also investigates how financial literacy levels are associated with economic outcomes across selected BRICS economies: India, South Africa and Brazil. However, instead of solely assessing whether financial literacy has a positive impact on economic outcomes, this paper examines the mechanisms and specific policies through which financial literacy and knowledge actually translate into meaningful financial behavior and economic participation under different institutional structures[15]. It focuses on how varying financial structures and country specific constraints shape the overall effectiveness of financial literacy across countries with similar economy sizes. The paper won't concentrate heavily on the relation between financial literacy and economic outcomes, but instead emphasizes upon constraints, institutional factors and informality that helps explain cross-country differences in similar economies.

The core argument of the paper focuses on not only the impact of financial literacy initiatives and levels on economic outcomes and growth[16], but also on how similar financial literacy initiatives across countries lead to uneven and varying economic outcomes across emerging BRICS economies due to differences in institutional capacity and informality levels. Moreover, the paper will recommend policies and initiatives that can be undertaken by India, South Africa and Brazil in order to reach the optimal level of efficiency in terms of translation and application of financial knowledge earned by the people.

This study follows a mixed-method approach and relies upon qualitative institutional analysis along with descriptive quantitative analysis of financial literacy and informality levels, and economic outcomes. The paper uses existing large scale survey data and cross-country datasets to enhance understanding of distinct policy structures in similar emerging economies, and the informality levels that impact the economic growth linked with the initiatives.

Summary data on financial literacy:[17]

- **India:** India's RBI 2023 survey highlights the financial literacy rate at 27%. Women's financial literacy rates are under 20%, linked to low digital access and educational disparities.
- **South Africa:** South Africa's FSCA 2022 and S&P Global FinLit survey estimated 42% with basic financial knowledge.
- **Brazil:** Brazil's Banco Central 2022 implying a financial literacy rate of around 30%

The study focuses on the key concepts and the links between them, the theoretical lens for the paper and what the existing literature has explored in the same field of research. The

[15] *Adekunle, I. A., Williams, T. O., Maialeh, R., & Adegbenro, M. A. (2025). Knowledge, attitude or risk? What drives the financial literacy gaps of university staff? Borsa Istanbul Review, 25, 192–200*

[16] *Lusardi, A., & Mitchell, O. S. (2014). The Economic Importance of Financial Literacy: Theory and Evidence. Journal of Economic Literature, 52(1), 5–44*.

[17] *All figures draw from the S&P Global Financial Literacy Survey (GFLEC), India's PLFS/RBI reports, South Africa's FSCA, and Brazil's Banco Central*

paper also studies the effectiveness of the different policies imposed, and how efficient and potent they are in the long-run economic growth of the country. The institutional and policy architecture of India, South Africa and Brazil are considered, including institutional analysis, delivery mechanisms and the level of enforcement and scalability. Lastly, the paper investigates the policy's effectiveness, recommends other policies to mitigate the gaps in the existing ones in order to have a greater impact and a more financially inclusive society.

2. Literature Review

This paper focuses on 3 different countries under the BRICS Organization: India, South Africa and Brazil, and their varying institutional capacity, policy designs, behavioral constraints and informality levels. Institutional capacity is the ability combined with the skills that institutions, organizations, or jurisdictions have with the purpose of implementing and execute public and private plans and contracts[18]. All 3 of the countries are aiming towards a higher financially literate population by imposing varying policy designs for the same, however, behavioral constraints, informality levels and institutional mediation efficiency play a major role in the same, and lead to heterogeneous outcomes. This further emphasizes upon the idea that a single policy cannot be effective in all countries, despite similar economic levels or development.

The standard approach in labor economics views human capital as a set of skills/characteristics that increase a worker's productivity[19]. Human capital is the idea that individuals invest in education and training of skills, which in turn increases their productivity, and this will lead to better economic outcomes. Thus, financial literacy is a specific form of human capital and has a positive impact on long-term growth if the return earned by the financial sector positively relates with the investment in financial literacy, or with its aggregate level[20]. A comparative cross-country analysis strengthens the understanding of the similarities, differences and institutional justifications for each of the 3 countries, through a systematic and cohesive literature review.

Moreover, there are datasets on account ownership, savings, payments and other financial indicators present[21] which is important for the cross-country analysis in this paper to deepen understanding on the impact of informality and other constraints on the effectiveness of financial literacy initiatives.

3. Data & Methodology

This paper incorporates descriptive quantitative analysis and qualitative institutional analysis to attain a clearer understanding of the conclusion and limitations. As for the quantitative analysis, the review uses S&P Global Financial Literacy Survey (GFLEC), OECD/INFE International Survey of Adult financial literacy, World Bank Findex Report and NFEC Financial Survey Reports to understand financial behavior, attitude and knowledge levels

[18] *Bucci, A. (2015). Product Proliferation, Population, and Economic Growth. Journal of Human Capital, 9(2), 170–197*
[19] The Basic Theory of Human Capital – Human Capital Report by World Economic Forum in collaboration by Mercer 2015
[20] *Bucci, A. (2015). Product Proliferation, Population, and Economic Growth. Journal of Human Capital, 9(2), 170–197*
[21] *World Development report, 2025 by World Bank Group*

across the globe, taking into account financial literacy indicators. The paper utilizes World Bank Open Data, IMF World Economic Outlook and OECD Economic Outlook to engage with a clearer understanding of the 3 country's economic standing and potential.

The paper contains qualitative institutional analysis through reports by the RBI, SARB/National Treasury and Banco Central de Brazil. Moreover, National Financial Inclusion Strategies are analyzed for a comprehensive understanding of each country's initiatives imposed on the general public for financial inclusion. As for the informality aspect of the paper, ILO Country Reports and national labor force survey is used to highlight the extent to which informality has absorbed a country, including unregistered small businesses and laborers without legal protection.

OECD Articles, World Bank Open Data and other surveys are reliable sources due to their reputations, standard data collection methods across population groups. The OECD Financial literacy questions are accessible for all, however since only a certain number and type of questions were asked, it may not provide a comprehensive overview of the subject in a real-world situation. Moreover, the country specific data is used in this paper in a consistent manner and framework such that results can be compared analytically and conclusions can be conveyed clearly.

The survey data used can be vulnerable to socially desirable and selection bias and may consist of small sample sizes for important groups like informal workers or low-income individuals. Although standard and core questions in OECD surveys allow a comparative analysis, it impedes cultural and perspective differences since all countries, despite being at similar development and economic levels under BRICS, host different population groups, have varying policies and different aggregate levels of financial literacy as well.

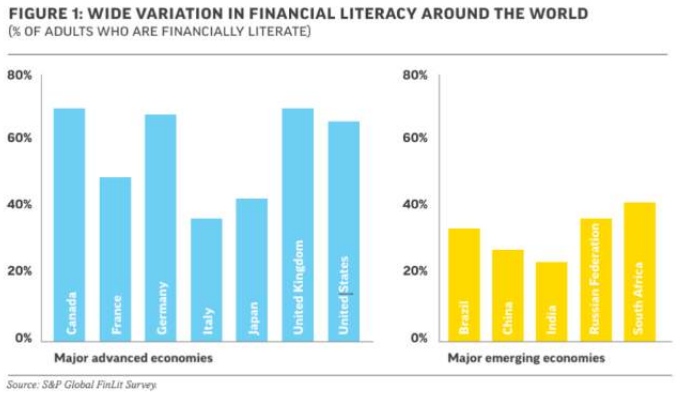
By using country specific data sources from public and private banks, the study compares institutional and delivery mechanisms along with regulatory coordination between the government and the general public. This study will not establish any causal relationship between financial literacy and economic outcomes, instead it interprets, analyses and compares the situation with the constraints in use.

4. Cross-Country Comparison

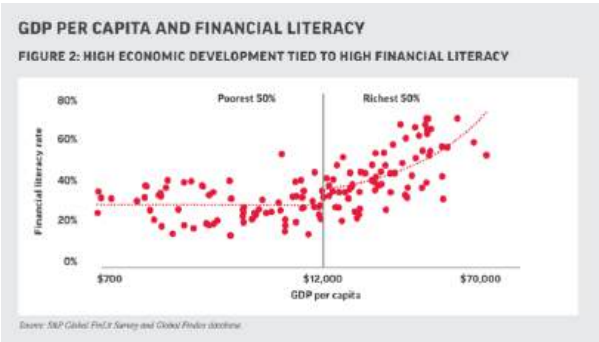
Financial literacy entails having the knowledge and comprehension of financial concepts and risks, and its aim is to improve the financial well-being of individuals and society, enabling active participation in economic life[22]. On average, across all participating countries and economies, only 34% of adults reached the minimum target score on financial literacy, defined as scoring at least 70 out of 100 points[23].

[22] OECD PISA 2022 Financial Literacy Results (released June 2024)
[23] *OECD/INFE 2023 international survey of adult financial literacy*

Taking in account recent years, only 27% of Indian adults are financially literate[24] (refer to Fig. 1 bar chart). South Africa's financial literacy level stands at 42%[25], whereas for Brazil, 62% of the people lack financial knowledge and thus, only 38% are literate in this subject. Thus, South Africa carries the highest financial literacy rates, with India being the lowest among the 3 countries, with the gap between both the countries being 15%. All 3 countries are grouped as emerging economies with significant natural resources, growing population and similar development challenges to an extent, but this does not transcribe directly into having the same financial knowledge rates[26].



Despite India's fast-paced growing economy, aiming to move into the upper middle-class income group by 2030, financial literacy levels don't fall into a direct and causal relationship with it as the context plays an important role: Informality workers, low-income households and unequal distribution lead to a non-linear relationship between financial inclusion and economic growth. Economic growth in South Africa has been at a low and declining rate of 0.7% per year, resulting in declining GDP per capita[27]. However, this ranking of economies is directly opposite of that for financial knowledge due to the presence of vulnerable and large numbers of low-income groups. This further emphasizes on how income stability or a growing economy is not always equal to a secure financial future for the general public due to factors the paper will discuss.



[28]As per figure 2, in richer countries, proxied by GDP per capita, financial literacy rates tend to be higher. However, the relationship only holds when looking at the richest 50 percent of economies. In these economies, around 38 percent of the variation in financial literacy rates can be explained by differences in income across countries. For the poorer half of

[24] Prasad, S. K. (2025, September 15). *How to improve financial literacy for India's next generation*. World Economic Forum
[25] WPA. (2025, September). *Financial Literacy in Sub-Saharan Africa: Current Levels and Trends - WealthPro Africa*. WealthPro Africa
[26] *S&P Global Financial Literacy Survey (GFLEC)*
[27] *OECD Economic Surveys: South Africa 2025. (2025),*
[28] *S&P Global FinLit Survey and Global Findex database*

economies, with a GDP per capita of $12,000 or less, there is no evidence that income is associated with financial literacy. What this likely means is that national-level policies, such as those related to education and consumer protection, shape financial literacy in these economies more than any other factor. The 3 comparable countries in the paper fall under the poorest 50% thus there is no evidence that income is associated with financial literacy.\

4.1 Comparative Analysis

Table 1: Comparative analysis of the 3 countries for defined parameters[29]

Parameter	India	South Africa	Brazil
Government & Financial Institutions			
Financial literacy rate	27%. (S&P Global FinLit survey)	~42% (FSCA–HSRC Baseline Survey 2022 and S&P Global FinLit Survey)	38% (S&P Global FinLit Survey)
Bank / financial account ownership	~80%	81%	~84%
	It was increased to 80% from 35% in 2011 driven by Pradhan Mantri Jan Dhan Yojana	Among top 2 in Sub-Saharan Africa (Global Findex 2025)	Most adults hold formal accounts; 160M+ Pix users
Digital payment infrastructure	UPI: 70% of digital transactions	USSD & mobile banking	Pix: 76% of all payment transactions
	285% rise in internet connections (2014–2024); 13B+ monthly UPI transactions	Fintech growth rising but no single dominant instant payment infrastructure	63.4B transactions in 2024; 159.9M registered users (July 2025)
Behavioral Economics			
Savings behavior	Household savings rate 30% of GDP	Household savings rate 10–12% of GDP	Household savings rate 17% of GDP
Debt behavior	Rising household debt; microfinance over-lending documented.	Household debt/disposable income: 62.4% (SARB 2023); one-third have no retirement plan.	Household debt to income ratio historically elevated; credit card over-reliance common.
Informality			
Informal economy size	50% of GDP	~30% of GDP	~40% of GDP
	~90% of workforce in informal employment; financial literacy access limited	32.1% unemployment; informal sector absorbs a large share of workforce	Pix partially formalizing gig & informal workers through digital payment rails
Economic activity & macroeconomic indicators			
GDP (2024)	USD 3.9 trillion	USD 380 billion	USD 2.18 trillion

[29] *S&P Global FinLit Survey · World Bank Global Findex 2025 · OECD/INFE 2023 · IMF FAS 2025 · WIPO 2025 SAR- Reports by Reserve Bank of India (RBI), South Africa Reserve Bank (SARB) and Brazil Central Bank*

	5th largest economy globally	Largest economy in Sub-Saharan Africa	Largest in Latin America; 9th globally (2024)
GDP growth rate	6.3% (FY2025–26 est.)	1.1% (2025)	3.4% (2024)
	World Bank; target of 7.8% to reach high-income status by 2047	Constrained by power shortages & unemployment	Digital economy accelerating
GDP per capita and GNI per capita (2024 est.)	~USD 2,700	~USD 6,200	~USD 10,200
	GNI per capita USD 2,650 in 2024 (World Bank) as per Atlas method, placing it at lower middle income. And on PPP basis it was $11,000 in 2024. Thus India's cost of living is much lower that its effective real income gap narrows significantly when adjusted for purchasing power	GNI per capital USD 6,100 in 2024, placing it at upper middle class. And on PPP basis it was $13,600 in 2024.	GNI per capita USD 9,950 in 2024 placing it at upper middle class (as per Atlas method). And on PPP basis it was $21,650 in 2024 Strong consumer spending at USD 1.34 trillion (2024)
Inequality (Gini coefficient)	~0.35	0.63 (high inequality despite moderate per capita income)	~0.52
	Improving but large urban–rural divide persists	Top 10 most unequal globally; limits translation of literacy into growth	One of the most unequal in Latin America despite high literacy rate
Poverty rate	~12% (USD 2.15/day)	21.6% (2023)	~5–7%
	106M lifted from extreme poverty 2011–2021; rural–urban gap persists	Youth unemployment 39%; structural inequality deeply effected	Bolsa Família social transfers play key role; digital payments reinforcing access
Key financial inclusion driver	Jan Dhan + Aadhaar + Mobile Phones (JAM trinity)	Mobile banking & fintech growth	Pix instant payments
	Financial Inclusion Index: 64.2 (Mar 2024), up from 60.1 (Mar 2023). Driven by Pradhan Mantri Jan Dhan Yojana (PMJDY). 540M+ accounts opened since 2014. Gender gap in account ownership now closed (World Bank Findex 2021).	FinScope SA 2023: Financial exclusion fell from 51% (2014) to 12% (2023), driven by mobile money adoption. Among top 2 in Sub-Saharan Africa (Global Findex 2021).	Most adults hold formal accounts. Pix has reached 175M+ users as of May 2025; 93% of Brazil's adult population. .

Sources: S&P Global FinLit Survey (GFLEC/World Bank/Gallup, 2015 baseline; latest update) · NCFE Financial Literacy & Inclusion Survey 2019 · RBI National Strategy for Financial Education 2020–25 · FSCA–HSRC Financial Literacy Baseline Survey 2020/2022 · BCB (Banco Central do Brasil) Financial Literacy Report 2023 · OECD PISA 2022 Financial Literacy – Brazil Factsheet · World Bank Global Findex 2021 · World Bank Human Capital Index 2020/2024 · World Bank Country Data – India, Brazil, South Africa (2024) · ILO Women and Men in the Informal Economy 2023 · WIEGO Statistical Brief · ISS African Futures – South Africa 2024 · Frontiers

in Big Data: BRICS SME Financial Literacy Study (Nov 2024) · IMF World Economic Outlook Apr 2026

4.2 Inequalities and Divide

Urban residents have higher ease in accessibility to digital financial services, banks and overall, a technology-dependent economy. This leads to a clear divide in urban and rural areas, reflecting inequality and unequal exposure to financial literacy initiatives. According to studies carried out, financial literacy rates in rural areas are much lower than in urban areas and thus, this is an indication for policy makers to implement new policies in such areas for a more secure financial future across the nation[30].

Due to the sole reason for profit maximization and optimizing returns, financial institutions focus heavily on urban areas which eventually leads to an urban-rural split. Asymmetric resource allocation and different access to financial services leads to varying outcomes across countries implementing similar initiatives for a financial inclusion future[31]. In India, rural poverty levels may have fallen but an evident gap still exists between both communities. For instance, the monthly per capita expenditure is Rs. 4,122 in rural India and Rs. 6996 in urban India, which leads to a 1.7x difference in spending power[32].

Urban household income depends on corporate jobs, skill-based wages and asset ownership, whereas rural income streams are agriculture and informal work. Moreover, education levels are much higher in urban areas, with high importance to secondary school education. Recently, initiatives like Pradhan Mantri Jan Dhan Yojana have increased bank ownership by a large number in rural India, however active financial services usage and well-acquaintance with the digital systems still remains relatively low. This highlights the vital need of new policies to be repositioned with large focus on the smaller towns and underdeveloped areas.

As for South Africa, it has one of the highest inequality levels, with inequality in per capita consumption about 50 percent higher than the average for upper-middle-income countries[33]. Through recent datasets, it is revealed that the top 10 percent earn 66 percent of total income, while the bottom half receives only 6 percent[34]. This shows how rural households earn less than half of urban household's income on average. With high unemployment rates, there are also persistent gaps in digital access for the particular group of people which translate directly to their limited exposure for financial literacy initiatives.

Lastly, in Brazil, while slave labor was abolished before the 20th century, black people, rural and domestic workers were and are still forced into informal work sectors and even unemployment. The Brazilian working class is a major victim of avoidance of labor rights, with

---

[30] *N. P. Abdul Azeez and M. Nasira Banu (2024). Rural-urban financial literacy divide in India: A comparative study of Kerala and Uttar Pradesh. Asian Research Journal of Arts & Social Sciences, 11(2)*

[31] *Huang, Z., Han, J., Xu, Z., & Dai, R. (2025). Digital financial inclusion and urban-rural disparities. International Review of Economics & Finance, 104, 104563).*

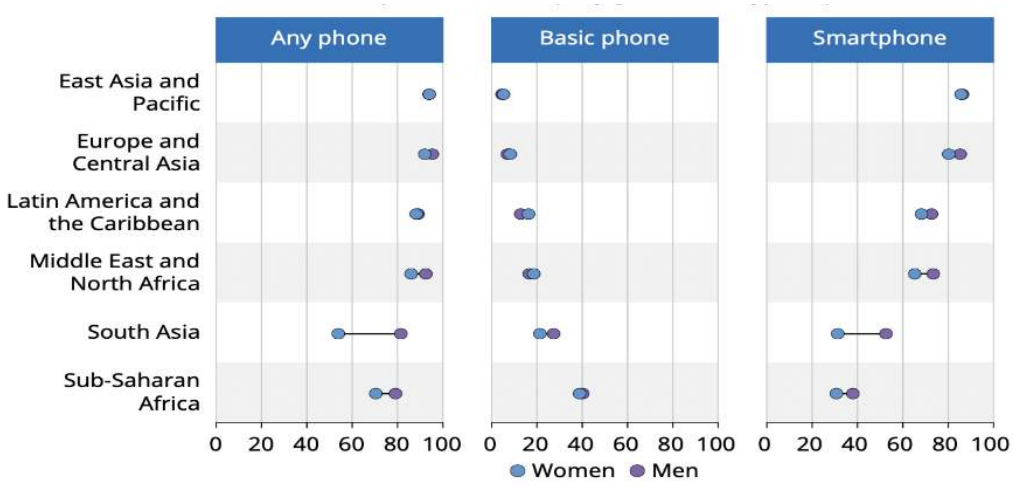
[32] *Household Consumption Expenditure Survey: 2023-24*. (2023). Pib.gov.in.)

[33] World Bank. (2022). *Country brief: South Africa 1*

[34] Ali, M., & Shafiq, F. (2025, December 10). *Where in the world are wealth and income most unequal?* Al Jazeera

the average rate of unemployment in 2024 being 24%. In 2023, 51% of rural inhabitants lived in shelter provided by social programs while in urban areas, it was a much lower number, 24.5%[35]. This disparity and inequality between both areas of the same country signifies the urgency of policies to shift towards a more equally distributed focus, where the rural regions are given more aid, guidance and importance.

Despite the imposition of multiple initiatives in all 3 countries, and consistent efforts from institutions, governments and banks, the results are uneven between and within the countries. A vital factor includes gender inequality since the presence of female work force is often lower than that of males, and not only does this reduce their labour-force participating, but in-turn, their asset ownership and role as a financial decision maker for themselves and their family; (OECD (2016)[36].



Source: Global Findex Database 2025.

The dataset in the figure above clearly shows how in almost each continent and region, males have a higher ownership of any particular type of phone. South Asia (India) and Sub-Saharan Africa (South Africa) show gaps in mobile ownerships between genders. However, Latin America (Brazil) doesn't show much of a gap, and this also confirms that Brazil has the highest literacy rates. It further justifies how women are shunned away from having equal digital access and thus, are not as comfortable or acquainted with DGS (Digital Financial Services), as males are due to accessibility. Thus, there is an alarming need for countries to relocate their policies with large focus on the female population, as they are often vulnerable to financial insecurity, emergencies and crises. Besides, they are often uninformed of their financial capacity and standing, thus need to acquire relevant and adequate financial knowledge to be in a safe financial position.

[35] Brazilian Institute of Geography and Statistics (IBGE) 2023

[36] *OECD/INFE International Survey of Adult Financial Literacy Competencies. OECD Publishing*

In India, many rural households prefer informal lenders since they believe they're much more flexible and meet their expenditure and income needs and view formal financial products as complex. Moreover, transaction costs in a formal bank setting and collateral demand make participation in formal financial situations tougher and less engaging for rural communal members[37]

India's effective, efficient and large-scale shift towards UPI transactions has significantly reduced costs and built trust but is still associated with complexity in certain rural areas. Microfinance was perceived as a poverty alleviation tool, with a lot of NGOs and commercial organizations working towards this goal.

Distrust and vulnerability have stemmed largely from formal credit systems in South Africa. Being the country with one of the highest levels of inequality, with over 10 million people over-indebted and financial insecurity levels are rising. Financial inclusion levels are increasing but it is not being reflected in their financial behaviour and health due to income inequality, unemployment and high costs of living which highlights how socio-economic pressures, vulnerability and debt, especially in rural settings and informal sectors, is an important focus area to understand the reality behind people's financial position[38].

The country has had significant development in account ownership numbers by expanding access, however most with bank accounts don't use it regularly, neglecting its effectiveness and importance as only 60% of adults use their bank accounts only once a month. On the positive side, policy makers, institutions and governments have acknowledged the importance of Fintech and are working towards making it more accessible and efficient, by implementing the IFWG regulatory sandbox. However, credit impairment and over-indebtedness is a common challenge faced by most, and this reduces public's trust in formal providers[39].

The country's dual economy leads to an evident inequality in financial inclusion, and the country is now rightly focusing on transforming the informal sector to a formal structure, ensuring consumers are protected against illegal market practices, which will lead to eventual trust.

Brazil has captured and implemented a trust-building mechanism with visible and reliable digital payment methods (PIX), which is responsible for 15% of global transactions. This advancement in their digital public infrastructure leads to adequate accessibility, inclusivity, safety, efficiency and scale. This DPI development reaches people from all socio-economic backgrounds, while being cost and user friendly and leading to long-term financial inclusion which is the aim for all 3 countries[40].

---

[37] Basu, P. (n.d.). *Improving Access to Finance for India's Rural Poor*. Retrieved February 2, 2026

[38] *FinMark Trust based on their FinScope Consumer South Africa 2024 Survey, released in May 2025*

[39] South Africa Financial Sector assessment report Jan 2022, Prepared By Finance, Competitiveness, and Innovation Global Practice and Africa Regional Vice Presidency (Link for reference https://documents1.worldbank.org/curated/en/102601645128957641/pdf/South-Africa-Financial-Sector-Assessment.pdf?utm_source=chatgpt.com)

[40] Digital Public Infrastructure and The Future of Digital Payments: Lessons from Pix and UPI - ORF Middle East. (2026, January 28). ORF Middle East

Despite intergenerational poverty curtaining the country, the Bolsa Familia Program aims to reduce poverty with a reliable and inclusive cash transfer system which reduces the possibility of capital leakages and avoids the risk of corruption and loss[41]. All in all, digital infrastructure progress helped reduce distrust and improve technological access.

This proves the aforementioned idea that despite a country having a larger, fast-paced growing economy, it doesn't always mean that their initiatives are directly and causally reflected into potent behavior and outcomes as evident by the comparison between India and Brazil specifically.

5. Institutional & Policy Architecture

India

5.1.1. Government and Policy Framework

Financial literacy initiatives are mainly led by the Reserve Bank of India, Insurance Regulatory and Development Authority of India, Securities and Exchange Board of India and Pension Fund Regulatory and Development Authority (PFRDA). RBI is the regulator and designs policies and initiatives for a more financially inclusive society. The ministry of finance drives the initiative across the country through nationwide campaigns and promotion by partnering up with RBI and other organizations to promote awareness of digital and formal financial products. As for public sector banks, through initiatives like Financial Literacy Centers and rural outreach as well, they build trust amongst marginalized communities to encourage them to participate in a formal financial system. While initiatives like Pradhan Mantri Jan Dhan Yojana have improved scale, accessibility and account ownership across segments of society, behavioral barriers like income instability and mistrust still hamper optimal usage amidst certain societies: Informal and low-income households.

5.1.2. Government-led Financial Literacy Initiatives

Financial inclusion promotes economic well-being and sustainable development for the country. When one is exposed to financial inclusion drives, like PMJDY, their knowledge increases, confidence with the formal monetary product slowly increases but the behavioral outcome is uneven not only between countries, but between communities and sectors as well. From pension focused programs, like Atal Pension Yojana, to the goal of Adopting Gender-Sensitive Approach for Women-led Financial Inclusion, the NSFI 2025-2030 report puts forth new initiatives which are well aligned with if planned against the constraints.

[41] REACHING THE HARD TO REACH: A Case Study of Brazil's Bolsa Família Program.

Figure 5: Number of Financial Literacy Programs and Participants

Source: National Strategy for Financial Inclusion (NSFI): 2025-30 (Revised)

According to the source above[42] There has been a consistent rise in the number of financial literacy programs, except during the pandemic, and these programs are mainly run by RBI regional officers and National Centre for Financial Education. Both institutions have had a major impact on financial inclusivity in India, and this helps understand the positives and the constraints that came along with these initiatives in order to construct adequate recommendations at the end of the paper[43].

The RBI has published a document titled "National Strategy for Financial Education Report 2020-2025," outlining its key objectives. The primary strategy adopts a comprehensive approach known as the "5 C's," which emphasizes the importance of Content, Capacity, Community, Communication, and Collaboration in advancing financial education. The Technical Group of Financial Inclusion, along with the Financial Literacy and Financial Stability and Development Council, is actively collaborating to facilitate the effective implementation of these policies. These initiatives are significant steps in the journey toward making India a financially literate nation[44]

The Insurance Regulatory and Development Authority (IRDA) has made several financial literacy-related initiatives. There have been awareness campaigns on radio and television, with the goal of spreading basic information regarding policyholder rights and obligations, available avenues for resolving disputes, etc.[45]

Public education about social security was carried out by the Pension Fund Regulatory and Development Authority (PFRDA). To bring the pension service to underprivileged communities in India, PFRDA has created a frequently asked questions section on its website [40]. It has also partnered with several of nongovernmental groups in the country

[42] *National Strategy for Financial Inclusion (NSFI): 2025-30 (Revised)*
[43] *Reports- Reserve Bank of India. (2025)*
[44] *RBI National Strategy for Financial Education Report", 2020-25*
[45] *Arya, p.."Financial Literacy and financial Education: An Assessment" , CASIR, vol.9,Issue 3, 2018, p.72-86*

5.1.3. Strengths & Constraints

The financial inclusion index in India has risen by 24.3% in 2025, since 2011, with over 500+ millions of accounts opened in PMJDY which highlights the breadth of inclusion and reach[46].

96% of Indians own a personal bank account, however usage levels remain low due to the following factors: Income instability, low financial awareness and preference for cash, and low education levels which led to mistrust. Lack of collateral and in-depth procedural awareness also shows that mere initiatives and programs are insufficient if there are no observable actions coming out of it, and low-income households then turn towards informal lenders. Accounts remain dormant since behavioral barriers like fear, volatility and low trust dominate the initiatives, which leads to a lower impact on economic activity in turn too.

Therefore, RBI focuses on financial products that are equitable, responsible, suitable and affordable. For the same, they aim to improve equity, reach and consistency while expanding digital financial services (Expanding and Deepening of Digital Payments Ecosystem) and their scope.

5.2. South Africa

5.2.1. Government and Policy Framework

South Africa's financial literacy policies are led by National Treasury's Financial Inclusion Policy Framework. The country focuses on supporting and delivering financial services to low-income earners and SMMEs by encouraging payment integration for small businesses[47].

The country aims to increase financial inclusion to 90% by 2030 by focusing on wider segments of the population. The National Financial Inclusion Policy is an official and coordinated policy which aims at financial inclusion for all. It aims to help the unbanked and underserved populations of the country. It is an "all-inclusive policy framework" and thus, is a comprehensive government framework and tries to cover multiple aspects of financial inclusion to maximize spread and scale[48].

5.2.2. Government-led Financial Literacy Initiatives

Back in 2004, the South African government required all financial services institutions to contribute 0.2 percent of their annual profits to financial education programs[49]. The country heavily focuses on consumer education programs and initiatives, along with basic training.

---

[46] *Economic Survey Report 2025-26, Press Information Bureau, India. Pib.gov.in*

[47] *AN INCLUSIVE FINANCIAL SECTOR FOR ALL November 2023, National Treasury Department: National Treasury REPUBLIC OF SOUTH AFRICA*

[48] *"South Africa Aims to Increase Financial Inclusion to 90% by 2030: Plausible or a Mere Ideal?" [a report in Oct 2024] by M Ncube*

[49] *The Impact of a Financial Literacy Program in South Africa | The Abdul Latif Jameel Poverty Action Lab. (2026). The Abdul Latif Jameel Poverty Action Lab (J-PAL)*

The South African Reserve Bank contributes by building a transparent financial environment and helping build trust with possible future consumers and users.

5.2.3. Strengths & Constraints

There is a clear policy framework implemented in the country, and this strong foundation along with collaboration between the government, regulators, private sector and development partners leads to financial inclusion progress as apparent by the dataset: Over 80% of adults own a bank account.

However, there is a notable gap between rural and urban areas since rural regions are less exposed to financial services when seen in comparison. There is a lack of unified programs, research and narrow dimensional programs (The National Strategy on Financial Literacy: A Conceptual Review of South African Perspectives).

A large group of the population is trapped in a continuous cycle of debt, financial instability and is often unable to apply financial knowledge in practical situations due to economic inequality and a lack of access to financial education, despite many initiatives heavily focusing on education led programs and this further signifies the rural-urban divide[50].

5.3. Brazil

5.3.1. Government and Policy Framework

National Strategy for Financial Education has been an initiative since 2010 and provides basic guidelines for households and families to improve their financial planning.

The country's policy framework integrates financial inclusion efforts into inclusion efforts which are guided by the Central Bank of Brazil's cicadania financeira (financial citizenship) and this links education and inclusion. (Banco Central do Brasil. (2018)[51]. With long-term strategic goals on this subject, Brazil has recognized this goal as a public policy while aiming to focus on all sectors.

5.3.2. Government-led Financial Literacy Initiatives

The ministry of education launched "Na Ponta de Lapis Program" which brings together financial, fiscal and social security education into the country's curriculum[52]. Thus, this is a big step forward by the country by bringing financial literacy in formal education. This strategy is strongly supported by banks and financial institutions which offer campaigns, research and help in optimal integration of the new curriculum.

5.3.3. Strengths & Constraints

There are many recent and widespread initiatives, however the reach is still uneven. According to OECD'S PISA 2022, nearly 45% of Brazilian students do not have the basic

[50] *The State of Financial Literacy in South Africa. (2025, June 2)*

[51] *Report on financial citizenship, 2018 by Banco Central do Brasil (BCB)*

[52] Report on Social, Environmental and Climate-related Risks and Opportunities Volume 5 | September 2025 by BCB

knowledge regarding financial literacy[53].

Moreover, with the rapid adoption and shift towards digital finance (PIX), many don't carry the adequate digital financial literacy skills which leaves rural populations vulnerable, helpless and in turn the initiatives don't help them out[54].

6. Conclusions & Recommendations

Financial literacy is a strong catalyst for improving decision making capacity, money management and future financial planning. However, it is not sufficient to aggregate improved financial behavior across a country. Similar developing economies produce different outcomes due to their institutional capacities, accessibility levels and other constraints which are at different levels in each of the nations. However, the vulnerable group is persistent: Informal, low-income, women and rural households. Future policy designs should focus on these groups to have an evened out economic outcome which represents the overall country, and not just a specific middle-income, formally employed or urban-living society. Institutions' policy's depth and ability to eliminate mistrust and over-indebtedness decides the extent to which the population productively participates in formal financial systems.

Brazil: Brazil should move beyond the financial education curriculum that was introduced only in 2018 due to low scores on assessments such as PISA. The country should work on monitored implementation, establishing benchmarks for standardized financial literacy assessments at multiple grade levels and providing funds to aid the under-resourced public schools for improved access across the country. Moreover, financial literacy should be provided in alignment with services such as credit counselling in order to alleviate its impact on the country. Pix, Brazil's instant payment system, should align fintech and payment institutions with traditional financial institutions since this provides an opportunity to attach financial literacy obligations to governance reforms. Lastly, considering the gender gap in financial literacy globally and in Brazil, the country should design targeted programs for women, especially the ones in the informal economy. This should be worked on in collaboration with the civil society organization to evaluate impacts on savings behavior and overall participation in formal financial markets.

India: The financial literacy coverage in India is broad due to the presence of multiple institutional structures such as RBI, NABARD, SEBI, IRDAI, and PFRDA. However, this creates duplication and coordination gaps, and thus there should be a Unified Financial Literacy Coordination Council under the Financial Stability and Development Council (FSDC), with shared data infrastructure and a common measurement framework. There should be robust impact evaluation of the policies implemented, and there should be assessment on changes in

[53] PISA 2022 Results (Volume IV) - Factsheets: Brazil. (2022). OECD

[54] *Digital Finance Adoption in Brazil: An Exploratory Analysis on Financial Apps and Digital Financial Literacy. Journal of Risk and Financial Management, by Cassola, N. M., Becker, K. L., Vieira, K. M., da Silveira Feldmann, M. F., Chaves, M. R., Berndt, I. C., & Arruda, A. F. M. (2025).*

savings behavior, formal credit uptake and digital payment adoption which will lead to improved content and delivery methods. India should integrate financial literacy content directly into platforms already used by women's Self-Help Groups by coordinating with NABARD's SHG-based programs to reduce difficulty in accessing formal services. To alleviate the impact, a financial competency assessment could be integrated within the National Achievement Survey to generate nationally representative data and enable comparison with OECD/PISA benchmarks with time.

South Africa: South Africa still struggles with extremely low levels of financial literacy, with assessments highlighting stagnation since 2015. Disjointed intervention initiatives further weaken citizen confidence, and thus a single coordination authority should be established. South Africa's 11 official languages and deep rural-urban divide shows that English-centric financial education resources only reach certain population groups. Thus, banks must be compelled to provide information in different languages which are suited to demographic circumstances, and thus are in the dominant language of the region the financial institution works in. Private institutions such as Capitec, Old Mutual and Momentum Metropolitan have developed multiple financial education programs, but the country should move from voluntary inclusion to a compulsory one since this would align private sector incentives with public financial inclusion goals. Lastly, the policies must be embedded within broader economic policy such as affordable formal financial products for low-income households and strict lending regulation due to the country's high unemployment, economic inequality and lack of access to formal financial education.

7. Limitations & Acknowledgement

Cross country comparisons may be affected by differences in the survey designs, cultural differences, varying population groups and measurements methods. Thus, the comparative analysis highlights broad qualitative patterns and not a precise quantitatively comparable study between the countries.

Stronger institutional environments lead to stronger behavioural impacts and as aforementioned in the recommendations, effective financial literacy strategies must be combined with institutional reforms for the most effective results.

Institutions' policy's depth and ability to eliminate mistrust and over-indebtedness decides the extent to which the population productively participates in formal financial systems. Lastly, future research can examine causal relationships, and additional country comparisons could provide deeper insight into how institutional structures influence financial behaviour.

Being the founder of Dhan Marg Organization,[1] where we host sessions across India on government and national investment and insurance schemes offered in the country, this study has allowed me to understand the real gaps in the society, as well as the importance of the issue.

I would like to thank Dr. Iva Bimpli and Ms. Daryn Howland for their useful feedback and guidance provided throughout the process.

Works Cited

1.Digital Public Infrastructure and the Future of Digital Payments: Lessons from Pix and UPI; Authors: ANIT MUKHERJEE | CAROLINE ARKALJI, Jan 2026 https://www.orfonline.org/research/digital-public-infrastructure-and-the-future-of-digital-payments-lessons-from-pix-and-upi

2.Bucci, A. (2015). Product Proliferation, Population, and Economic Growth. Journal of Human Capital, 9(2), 170–197 https://www.journals.uchicago.edu/doi/10.1086/680861

3.Adekunle, I. A., Williams, T. O., Maialeh, R., & Adegbenro, M. A. (2025). Knowledge, attitude or risk? What drives the financial literacy gaps of university staff? Borsa Istanbul Review, 25, 192–200 https://www.researchgate.net/publication/396977031_Knowledge_Attitude_or_Risk_What_Drives_the_Financial_Literacy_Gaps_of_University_Staff

4.Worldometer. (2024). GDP by Country. Worldometer https://www.worldometers.info/gdp/gdp-by-country/?source=imf&region=worldwide&year=2024&metric=nominal

5.Informal / Shadow Economy Size | By Country | 2022 | Data. (n.d.). World Economics https://www.worldeconomics.com/Informal-Economy/

6.Lusardi, A., & Mitchell, O. S. (2014). The Economic Importance of Financial Literacy: Theory and Evidence. Journal of Economic Literature, 52(1), 5–44 https://gflec.org/wp-content/uploads/2014/12/economic-importance-financial-literacy-theory-evidence.pdf

7.The S&P Global Financial Literacy Survey (GFLEC) - gflec.org https://gflec.org/wp-content/uploads/2015/11/3313-Finlit_Report_FINAL-5.11.16.pdf

8.Periodic Labour Force Survey (PLFS) Annual Report (2023-24) https://www.pib.gov.in/PressReleasePage.aspx?PRID=2057970&reg=3&lang=2

9.South Africa's - The Financial Sector Conduct Authority (FSCA) report

10.Brazil's Banco Central annual report 2022 https://www.bcb.gov.br/content/publications/Annual_Report_docs/AR_2022.pdf

11.Human Capital Report by World Economic Forum in collaboration by Mercer 2015 https://www3.weforum.org/docs/WEF_Human_Capital_Report_2015.pdf

12.World Development Report, 2025 by World Bank Group https://www.donorplatform.org/wp-content/uploads/2026/01/Full-Report.pdf

13.IMF World Economic Outlook reports (2025 & 2026)

14.OECD PISA 2022 Financial Literacy Results (released June 2024) https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/pisa-2022-results-volume-iv_125a58b3/5a849c2a-en.pdf

15.OECD/INFE 2023 international survey of adult financial literacy https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/12/oecd-infe-2023-international-survey-of-adult-financial-literacy_8ce94e2c/56003a32-en.pdf

16. Prasad, S. K. (2025, September 15). How to improve financial literacy for India's next generation. World Economic Forum https://www.weforum.org/stories/2025/09/how-to-improve-outcomes-for-india-s-next-generation-through-financial-literacy/
17. WPA. (2025, September). Financial Literacy in Sub-Saharan Africa: Current Levels and Trends - WealthPro Africa. WealthPro Africa https://wealthpro.africa/financial-literacy-in-sub-saharan-africa-current-levels-and-trends/
18. OECD Economic Surveys: South Africa 2025 https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/06/oecd-economic-srveys-south-africa-2025_977f0fed/7e6a132a-en.pdf
19. World Bank Global Findex 2025 https://openknowledge.worldbank.org/bitstreams/9288bdc5-7a9b-42de-a47c-3746fd68f22a/download
20. IMF 2025 Financial Access Survey (FAS) results report, released in October 2025,
21. South Africa Reserve Bank (SARB) https://www.resbank.co.za/content/dam/sarb/publications/reports/annual-reports/2025/sarb-2024-25.pdf.coredownload.inline.pdf
22. NCFE Financial Literacy & Inclusion Survey 2019 https://ncfe.org.in/wp-content/uploads/2023/12/ExecSumm_.pdf
23. RBI National Strategy for Financial Education 2020–25 https://ncfe.org.in/wp-content/uploads/2023/12/NSFE_20-25_ENG.pdf
24. FSCA–HSRC Financial Literacy Baseline Survey 2020/2022 https://repository.hsrc.ac.za/handle/20.500.11910/19334
25. BCB (Banco Central do Brasil) Financial Literacy Report 2023
26. OECD PISA 2022 Financial Literacy – Brazil Factsheet https://www.oecd.org/en/publications/pisa-2022-results-volume-iv-factsheets_34d60137-en/brazil_1c815ef9-en.html
27. World Bank Global Findex 2021
28. World Bank Human Capital Index 2020 / 2024 https://ourworldindata.org/grapher/human-capital-index-in-2018
29. World Bank Country Data – India, Brazil, South Africa (2024)
30. International Labour Organisation (ILO) - Women and Men in the Informal Economy 2023 https://www.ilo.org/publications/women-and-men-informal-economy-statistical-update
31. ILO - WIEGO Statistical Report 2023 https://www.wiego.org/ilo-wiego-statistical-reports/
32. Institute of Security Studies (ISS) African Futures – South Africa 2024 https://issafrica.s3.amazonaws.com/uploads/pages/1753336920333-ar-2024.pdf
33. Frontiers in Big Data: BRICS SME Financial Literacy Study (Nov 2024) https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2024.1448571/full
34. IMF World Economic Outlook Apr 2026

https://www.imf.org/en/publications/weo/issues/2026/04/14/world-economic-outlook-april-2026
35.N. P. Abdul Azeez and M. Nasira Banu (2024). Rural-urban financial literacy divide in India: A comparative study of Kerala and Uttar Pradesh. Asian Research Journal of Arts & Social Sciences, 11(2) https://www.academia.edu/68748296/Rural_Urban_Financial_Literacy_Divide_in_India_A_Comparative_Study_of_Kerala_and_Uttar_Pradesh
36.Huang, Z., Han, J., Xu, Z., & Dai, R. (2025). Digital financial inclusion and urban-rural disparities. International Review of Economics & Finance, 104, 104563) https://www.sciencedirect.com/science/article/pii/S1059056025007269
37.Household Consumption Expenditure Survey: 2023-24. (2023). Pib.gov.in https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2097601&reg=3&lang=2
38.World Bank. (2022). Country brief: South Africa 1 https://documents1.worldbank.org/curated/en/099125003072240961/pdf/P1649270b73f1f0b5093fb0e644d33bc6f1.pdf
39.Brazilian Institute of Geography and Statistics (IBGE) 2023 https://www.gov.br/secom/en/latest-news/2024/08/ibge-brazils-population-reaches-212-6 million
40. Basu, P. - Improving Access to Finance for India's Rural Poor. Retrieved February 2, 2026 https://documents1.worldbank.org/curated/en/254781468138877578/pdf/364480PAPER0N101OFFICIAL0USE0ONLY1.pdf
41.FinMark Trust based on their FinScope Consumer South Africa 2024 Survey, released in May 2025 https://finmark.org.za/knowledge-hub/articles/12-million-adults-struggle-with-debt-while-relying-on-credit-to-cope?entity=news
42.South Africa Financial Sector assessment report Jan 2022, Prepared By Finance, Competitiveness, and Innovation Global Practice and Africa Regional Vice Presidency (Link for reference https://documents1.worldbank.org/curated/en/102601645128957641/pdf/South-Africa-Financial-Sector-Assessment.pdf?utm_source=chatgpt.com)
43.Digital Public Infrastructure and The Future of Digital Payments: Lessons from Pix and UPI - ORF Middle East. (2026, January 28). ORF Middle East https://orfme.org/research/digital-public-infrastructure-and-the-future-of-digital-payments-lessons-from-pix-and-upi/
44.REACHING THE HARD TO REACH: A Case Study of Brazil's Bolsa Família Program https://reachalliance.org/wp-content/uploads/2021/04/BolsaFamilia-CaseStudy-FINAL-pages.pdf
45.WIEGO Statistical Brief · ISS African Futures – South Africa 2024 https://futures.issafrica.org/geographic/countries/south-africa/ and https://www.wiego.org/wp-content/uploads/2026/01/WIEGO_Annual-Report_2024-2025-EN.pdf
46.Frontiers in Big Data: BRICS SME Financial Literacy Study (Nov 2024)

https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2024.1448571/full
47.South Africa Economic Outlook 2023
https://www.treasury.gov.za/documents/mtbps/2023/mtbps/Chapter%202.pdf and
https://www.afdb.org/en/news-and-events/press-releases/2023-southern-africa-economic-Outlook-southern-africas-economic-prospects-subdued-yet-abounds-investment-opportunity-climate-change-initiatives-63343
48.National Strategy for Financial Inclusion (NSFI): 2025-30 (Revised)
https://rbi.org.in/Scripts/PublicationReportDetails.aspx?UrlPage=&ID=1307
49.RBI National Strategy for Financial Education Report", 2020-25
https://ncfe.org.in/nsfe-2020-25/
50.Arya, p.."Financial Literacy and financial Education: An Assessment" , CASIR, vol.9,Issue 3, 2018, p.72-86 https://iipseries.org/assets/docupload/rsl2024262F10D3C60E82C.pdf
51.Economic Survey Report 2025-26, Press Information Bureu, India. Pib.gov.in
https://www.pib.gov.in/PressReleasePage.aspx?PRID=2219907&reg=3&lang=2
52.AN INCLUSIVE FINANCIAL SECTOR FOR ALL November 2023, National Treasury Department: National Treasury REPUBLIC OF SOUTH AFRICA
https://www.treasury.gov.za/comm_media/press/2023/2023112701%20An%20Inclusive%20Financial%20Sector%20for%20all%202023.pdf
53."South Africa Aims to Increase Financial Inclusion to 90% by 2030: Plausible or a Mere Ideal?" [a report in Oct 2024] by M Ncube
https://www.researchgate.net/publication/386741609_South_Africa_Aims_to_Increase_Financial_Inclusion_to_90_Percent_by_2030_Plausible_or_a_Mere_Ideal
54.The Impact of a Financial Literacy Program in South Africa | The Abdul Latif Jameel Poverty Action Lab. (2026). The Abdul Latif Jameel Poverty Action Lab (J-PAL)
https://www.povertyactionlab.org/evaluation/impact-financial-literacy-program-south-africa
55."The impact of financial literacy on savings in South Africa," International Journal of Research in Business and Social Science (2147-4478) by Cecile Duvenhage & Nico Keyser & Lentsho Moepi, 2025
https://ideas.repec.org/a/rbs/ijbrss/v14y2025i5p222-229.html
56.Financial literacy in South Africa: impact and solutions. Trialogue Knowledge Hub | The Trialogue Business in Society Handbook 2025 (28th edition) by Sheldon Morais, Managing Editor
https://trialogueknowledgehub.co.za/financial-literacy-in-south-africa-impact-and-solutions
57.Report on financial citizenship, 2018 by Banco Central do Brasil (BCB)
https://www.bcb.gov.br/content/publications/report_fincit/Report%20on%20Financial%20Citizenship%20-%202018.pdf
58. Report on Social, Environmental and Climate-related Risks and Opportunities Volume 5 | September 2025 by BCB
https://www.bcb.gov.br/content/publications/report-risk-opportunity/RIS_EN_2025.pdf
59.Digital Finance Adoption in Brazil: An Exploratory Analysis on Financial Apps and Digital Financial Literacy. Journal of Risk and Financial Management, by Cassola, N. M.,

Becker, K. L., Vieira, K. M., da Silveira Feldmann, M. F., Chaves, M. R., Berndt, I. C., & Arruda, A. F. M. (2025) https://www.mdpi.com/1911-8074/18/10/560
60. Pix and Progress: Democratizing Financial Services, May 27, 2025 By: Bruna Araujo MSBA '25. Cornell SC Johnson College of Business https://business.cornell.edu/article/2025/05/democratizing-financial-services/
61. EBANX Pix 5-Year Report Nov 2025 https://finance.yahoo.com/news/pix-approach-8-billion-monthly-130800473.html
62. Financial inclusion through technology and literacy in India: Strategies for sustainable growth: 2024 – CII & EY report https://www.ey.com/content/dam/ey-unified-site/ey-com/en-in/ey-cii-report-web-version-v1.pdf

Developing A Wearable EEG Headband for Adolescent Stress Monitoring Using Sp1 and Sp2 Electrodes By Jayden Lee, Irene Yang, Selina Yum

Abstract

Over the past decade, teenage stress levels have risen significantly, and the demand for stress management techniques has followed this trend. Adolescent stress is widespread, and existing methods such as self-report and heart rate monitoring have significant limitations, as teens may underreport stress and physical signals like heart rate can be affected by unrelated factors like exercise or temperature. To address this gap, we developed the BioVibe EEG headband, a wearable EEG headband designed to discreetly monitor stress in adolescents using dry electrode positions at the Sp1 and Sp2 scalp locations. These frontal electrode positions were selected for their relevance to emotional processing and stress response. We collected more than 80,000 lines of EEG data over 20 hours across various tasks, and validated the system by cross-referencing EEG shifts with timestamped reactions. These various tasks vary in intensity such as homework, studying, and watching horror films. This study determined that the two-electrode setup successfully captured interpretable brainwave signals. These findings lay the groundwork for future stress monitoring development using Sp1 and Sp2 electrodes. To make this technology accessible, BioVibe also includes a companion mobile application that visualizes brainwave activity in real time and provides users with stress management resources including journaling, breathing exercises, and relaxation tools.

Introduction

The teenage years are an important time when people go through many physical, emotional, and social changes. During this time, many teens feel stressed because of schoolwork and pressure from friends or society. A little stress can be helpful, but prolonged periods of stress can cause problems like anxiety, depression, and worse performance in school. Mental health stigma is a relevant problem affecting many students who feel uncomfortable or embarrassed by interventions that draw attention to them. In fact, the Springtide Research Institute conducted a national survey on 4,038 teens and young adults: 45% reported avoiding therapy because their parents do not take their concerns seriously; 53% said they would not want their parents to know they were meeting with a school counselor or therapist; 51% feared that school staff might treat them differently or limit their opportunities (Springtide Research Institute, 2023). These challenges show the importance of finding good ways to detect and understand stress in teenagers.

Currently, stress is usually measured using surveys, observing behavior, or checking things like heart rate. However, these methods are not always accurate or feasible. Teens may not always be honest or aware of how stressed they are since physical signals like heart rate can change for other reasons, like exercise or temperature. The surge of these problems are encouraging scientists to seek for better and more direct ways to measure stress. One possible method is using EEG, which records brain activity in real time. EEG can give more direct

information about what is happening in the brain during stress. According to the National Library of Medicine, certain brain waves, like alpha waves and beta waves, may change when someone is stressed.

This study focuses on two parts of the brain called Sp1 and Sp2, located at the front of the head. These areas are important for emotions, decision-making, and stress responses. By studying EEG signals from these areas, the goal was to explore whether BioVibe's hardware could capture meaningful signals from these areas.

BioVibe, a fashionable stress-detecting headband, continuously monitors EEG signals and notifies users during high periods of stress, taking the idea of stress detection and putting it into practical use to help teens with stress management.

Literature Review

Electroencephalographs (EEG) measure the brain's electrical activity through frequency bands that correspond to different mental states, primarily through alpha, beta, and delta waves defined by a specific frequency range of 8-12 Hz, 13-30 Hz, and 0.5-4 Hz, respectively (Attar, 2022). Research shows that delta waves, the slowest of the three, are associated with restorative processes that occur during deep sleep, while alpha waves are similarly associated with relaxed states, but occur during wakefulness. By contrast, beta waves are linked to active alertness and thinking (Attar, 2022). The classifications of these frequency bands provide the foundation for interpreting the EEG signals collected through BioVibe's data collection.

Researchers are actively increasing their focus on the role of EEG frequency bands in cognitive and emotional functioning, with stress being one of their main focal points. Most notably, increased beta activity has been correlated with heightened distress, whereas elevated alpha activity has been associated with reductions in stress (Nayak & Anilkumar, 2025). This correlation has led to EEG incorporation in daily life, with wearable EEG devices making the collection of brainwave data out of clinical settings possible.

While wearable EEG systems may appear as heavily appealing, it is critical to keep the limitations of the technology in mind. Motion artifacts and inconsistent electrode contact due to hair or sweat can reduce accuracy in uncontrolled settings. These challenges pose relevant concerns for gaining an accurate representation of the user's live stress. Yet, machine learning has been increasingly developing to improve the classification of mental states utilizing EEG technology. In addition to EEG, galvanic skin response (GSR) has widely been used as a complementary measurement of stress. Thus, the combination of EEG and GSR is an effective strategy of overcoming the foreign responses due to external stimuli.

GSR sensors measure changes in skin conductance caused by sweat gland activity. During moments of heightened emotional intensity, the sympathetic nervous system increases sweat gland activity as a response of fight or flight (Albert & Tullis, 2023). The body's electrical conductivity is measured, serving as a method to access the arousal level of the person being measured. Combining the technology of EEG and GSR allows for a higher indicator of stress than observed in standard EEG bands.

Overall, EEG is a powerful tool for measuring brain activity and identifying stress patterns. The growing support for a connection between EEG data and stress data analysis allows for the translation of EEG signals into real-time stress feedback. However, existing limitations indicate the necessity of further development to improve overall accuracy of relying on only EEG data.

Data Collecting System Design

Hardware Design

The hardware used for data collection included the PiEEG v2.1 module connected to a Raspberry Pi 4. EEG signals were recorded by analyzing brainwave activity from the Sp1 and Sp2 electrode positions, as Sp1 and Sp2 have most of the emotional responses that correlate with stress. Data was gathered by directly placing dry electrodes on the scalp electrode positions processing data to the PiEEG.

Fig 1: Electrodes attached to an elastic spandex headband, positioned on the Sp1 and Sp2 areas.

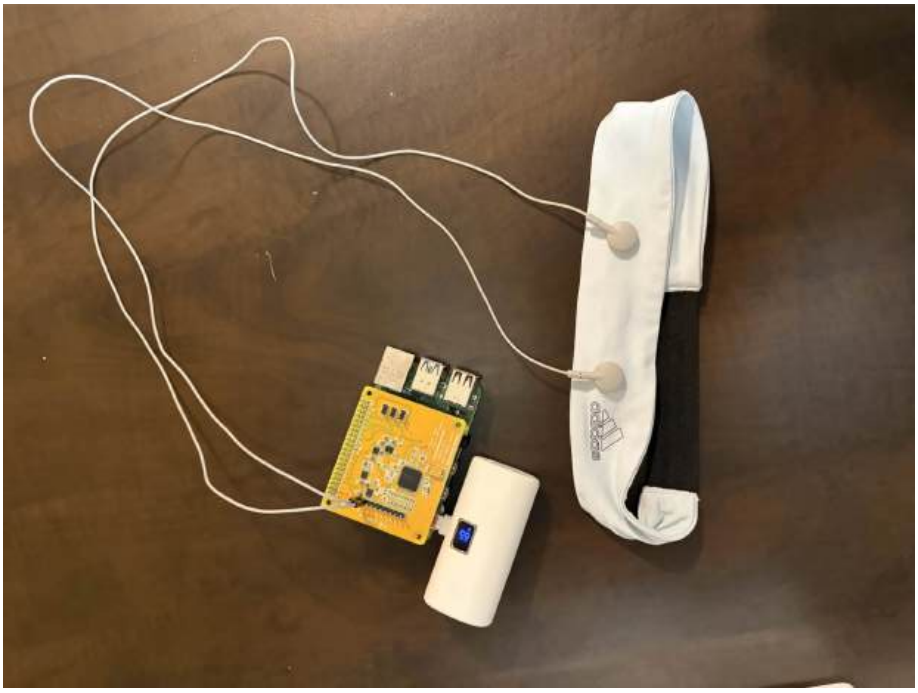


Fig 2: Headband and Electrodes attached to the PiEEG v2.1, Raspberry Pi 4, and portable ANKER charger.

Data Collection

EEG data was sampled at 256 Hz and saved as CSV files, where each row contained a datetime stamp followed by voltage readings from each electrode channel. For the signal processing pipeline, the Sp1 channel was extracted and used for frequency band analysis.

Signal Processing

All data processing and analysis were conducted using Python 3 within a Jupyter Notebook environment. Three libraries were used: NumPy for numerical operations including signal normalization, Matplotlib for generating visualizations, and SciPy for digital filter design and application.

After acquiring the EEG data from the PiEEG device, the signals were processed and decomposed into standard frequency bands (Delta: 0.5–4 Hz, Alpha: 8–13 Hz, Beta: 13–30 Hz). This was achieved by extracting the relevant frequencies corresponding to each band. The processed data was visualized using time-series graphs, with elapsed time on the x-axis and normalized amplitude on the y-axis, with each frequency band represented by a distinct color — Delta in blue, Alpha in green, and Beta in red. The signal processing algorithm developed in this pipeline was subsequently integrated directly into the application, where it operates on incoming EEG data.

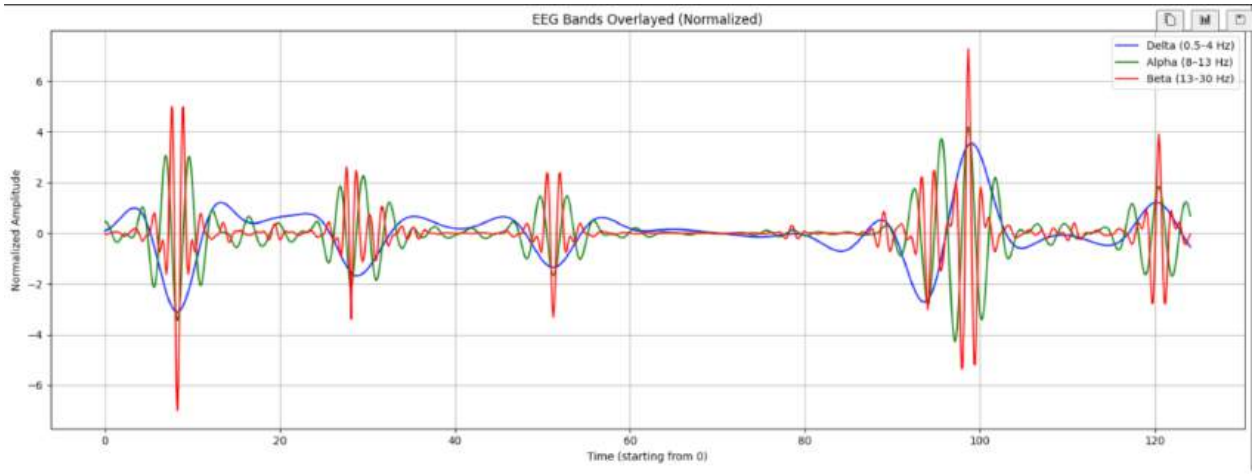


Fig 3: Graph of Recorded EEG Frequency Bands (Delta, Alpha, Beta)

File created	Sp1	Sp2	Sp3	Sp4	Sp5	Sp6	Sp7	Sp8
0	-47542.16	-34981.91	-10554.76	-14363.77	-15320.78	-17514.83	-4047.19	-5164.59
0	-46768.88	-33291.04	-9612.23	-13454.5	-15361.01	-17422.29	-3157.23	-4323.99
0	-48314.63	-35567.97	-8090.62	-11919.74	-14934.54	-16904.36	-1867.89	-3085.88
0	-47384.98	-33701.42	-6794.3	-10623.43	-14678.39	-16540.38	-759.87	-2026.95
0	-47290.3	-33720.47	-5696.48	-9553.22	-14592.56	-16322.58	245.69	-1061.35
0	-48840.35	-35938.12	-4112.36	-7947.39	-14134.44	-15771.39	1549.78	196.87
0	-48375.25	-34700.01	-2706.35	-6536.54	-13816.06	-15343.31	2737.46	1343.52
0	-47200.18	-32807.17	-1821.22	-5689.23	-13918.52	-15302.54	3566.53	2129.41
0	-49149.61	-140624.2	556468.55	1120696.19	-1687568.23	-1688052.64	-1689211.62	560629.46
0	-48010.2	-33395.92	935.55	-2918.24	-13288.2	-14464.89	5885.57	4362.88
0	-47191.59	-32494.7	1840.8	-2054.04	-13351.77	-14387.37	6740.66	5179.08
0	-47632.82	-34732.19	3374.22	-504.79	-12907.86	-13849.05	8022.22	6418.26
0	-47420.39	-34545.51	4869.28	1005.02	-12513.85	-13355.26	9253.89	7605.94
0	-46003.11	-32079.76	5811.01	1909.46	-12545.23	-13250.65	10131.24	8441.18

Fig 4: Raw EEG Data Table Recorded by Each Sphenoidal Electrode

Results

Data was collected throughout a series of tests conducted at varying periods of time between 45-60 minutes. Our team wore the BioVibe EEG headband and performed various tasks, including but not limited to: doing homework, studying for a test, and watching horror films. Testers were to conduct their assigned task for intervals of arbitrary intervals, with timestamps manually noted at moments when participants experienced a noticeable emotional reaction. The raw data was recorded down and transformed into lines of data that could be transcribed into a graph by the Jupyter Notebook system.

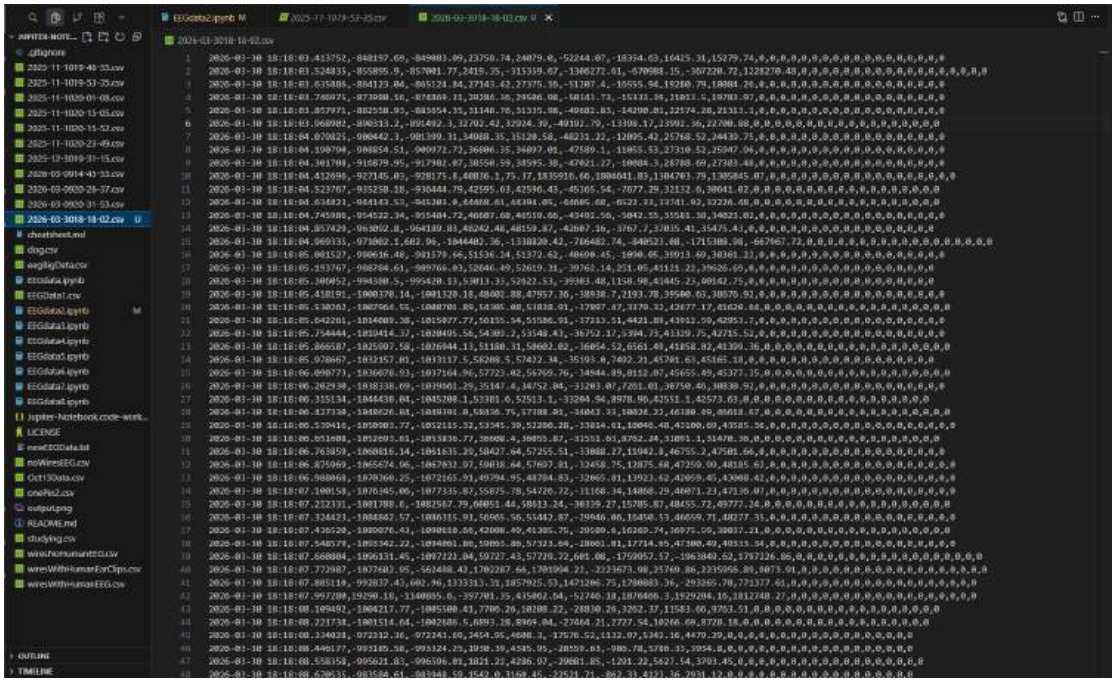


Fig 5: Picture of Raw EEG Data in a CSV file

After collecting over 80,000 lines of raw EEG data over 20 hours, BioVibe tweaked a pre-existing Github Repository and translated the code through Jupyter Notebook. Using Cursor to leverage the process, Biovibe ran numerous of these raw csv files through the code (Figure 6). The graphs were created for their corresponding data, with all three frequency waves (alpha, beta, delta) displayed. Alpha waves appeared as green, beta as red, and delta as blue (Figure 7).

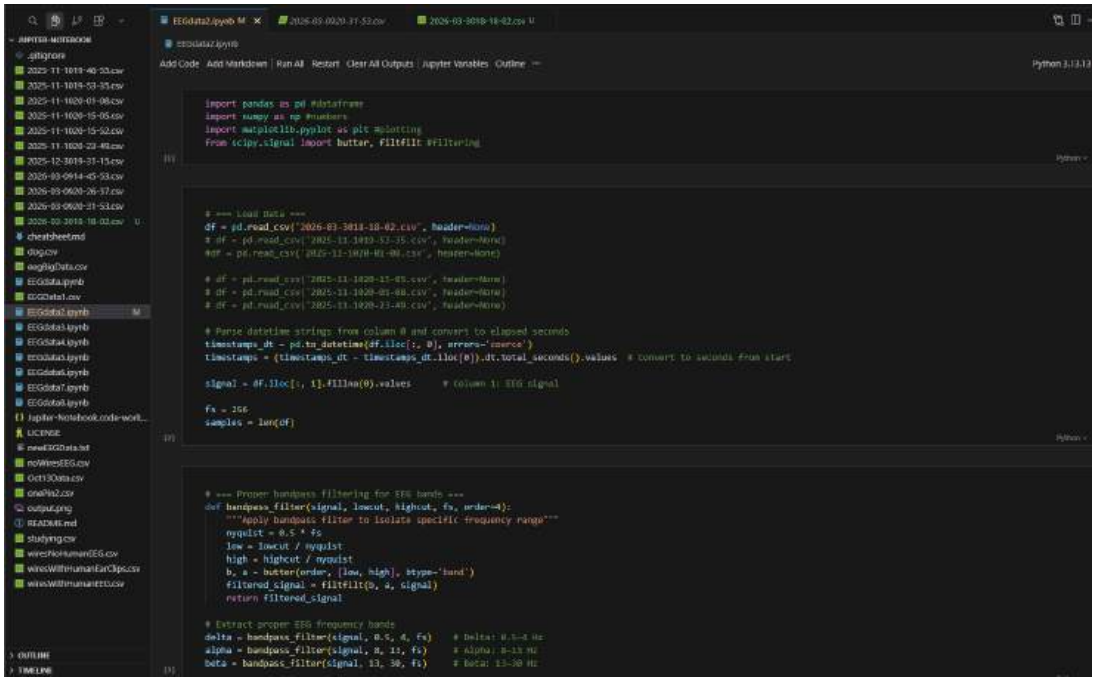


Fig 6: Process of Translating EEG Data into a Graph

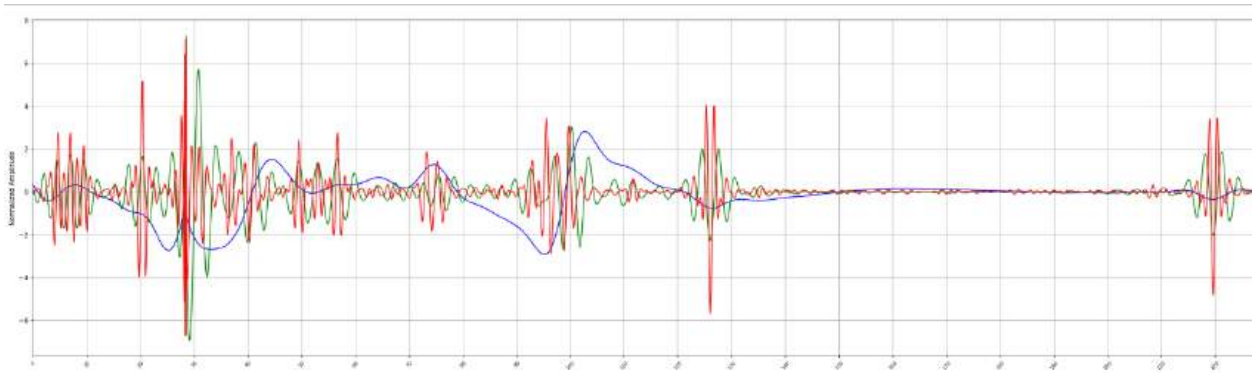


Fig 7: 3.75 Minute Graph of Collected EEG Data

Biovibe cross referenced the EEG graph with the manually recorded timelines to validate that the EEG hardware and data processing pipeline were capturing genuine brain activity. Observable shifts in frequency band activity at the noted movements confirmed that the system was recording and processing real EEG signals. Specifically during horror film viewing, noticeable peaks in frequency bands—especially in the alpha bands—were observed at moments corresponding to the jump scares, indicating changes in neural activity corresponding to the user's emotional response. These peaks aligned with the recorded timestamps, supporting that the electrode setup and processing pipeline system captures genuine neural responses rather than random noise.

The BioVibe App

The BioVibe App was made so that users can see their brainwave data live on their mobile devices and get real-time alerts when their stress spikes. The app's implementation had three phases: the early interface ideation, front-end development, and the final integration.
Interface Designing and Prototyping

The initial design process began with prototypes of each page created in Figma. We prototyped various features on the app, including a home screen displaying the user's names and connection status to the headband; a "My Stress" tab that displayed the live data and a history of old saved sessions; and a reminders page with alarms for stress-relieving activities.

Fig 8: Original Figma page designs. Left to right: home page, live data, peaked stress detection, saved sessions, and reminders.

While prototyping, we not only prioritized visualizing the data, but also stress management resources. The interface was expanded to include multiple dedicated pages for stress relief. These included features such as a journaling interface for self-reflection and guided breathing exercises to support relaxation. We also used soft, rounded corners and a gradient of purple, which is shown to reduce stress and calm people down. These additions contributed to a more holistic user experience by combining real-time monitoring with actionable coping strategies.

Fig 9: More Figma page designs. Left to right: Settings & account page, journaling page, journal history, reflections, breathing exercises

Application Development

Following the completion of the interface design, we used JavaScript, HTML, and CSS to create the BioVibe app. The front end of the application was built with HTML and CSS to create a user-friendly layout available as both a mobile app and a website. JavaScript was used to

handle the real-time updates from the EEG sensor, which turns the screen yellow at periods of high stress.

On the backend, Python was used to process and analyze the EEG data. The system receives brainwave signals from the EEG headband, filters and interprets the data, and visualizes the user's brainwave activity. This information is then sent to the front end, where it is visualized in an accessible format. Additionally, the app integrates breathing exercises and relaxing games on a separate tab that can be played any time.

Completed Application

The final BioVibe app successfully integrates real-time EEG data visualization with accessible stress management tools. The main dashboard presents continuously updating graphs and indicators during high periods of stress, allowing users to easily interpret their state. In addition to monitoring capabilities, the app incorporates stress relief features, including multiple journaling and breathing exercise modules, as well as relaxation games and soothing music. These components allow users to actively respond to their stress levels. While there are opportunities for further refinement, the current version of the BioVibe app demonstrates the feasibility of combining EEG sensors for stress detection and prevention.

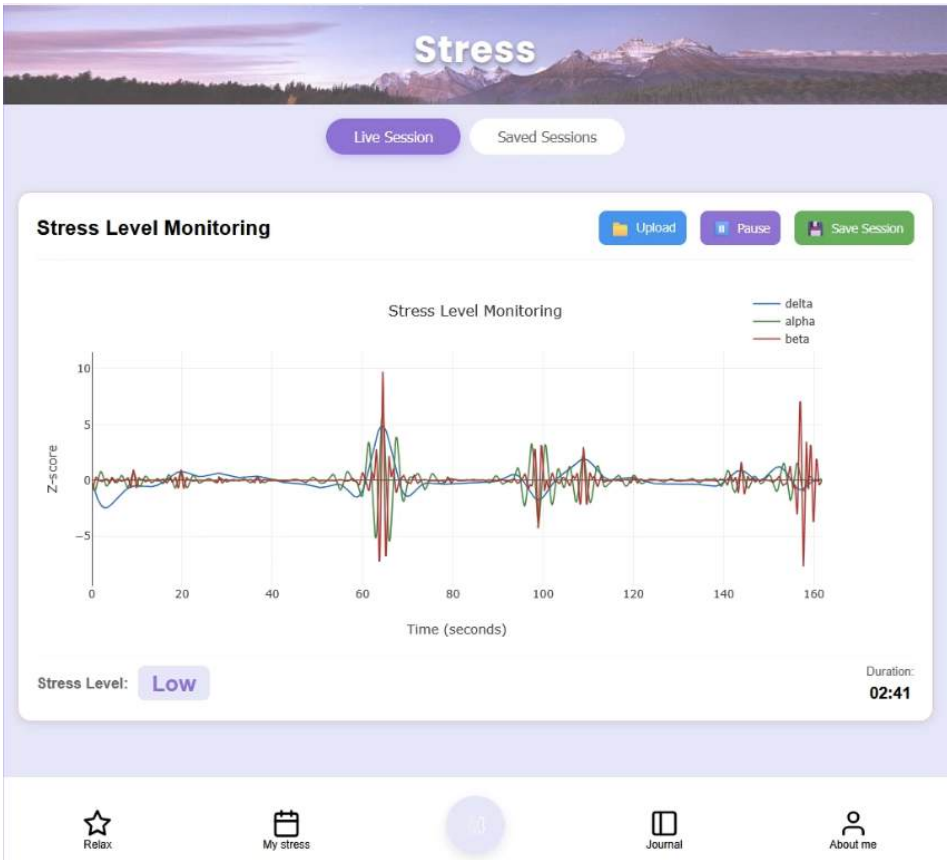


Fig 10: Live data graph, showing the delta, alpha, and beta brainwaves. On the bottom left, it indicates the level of stress. Users can click the “save session” button to see their history. j

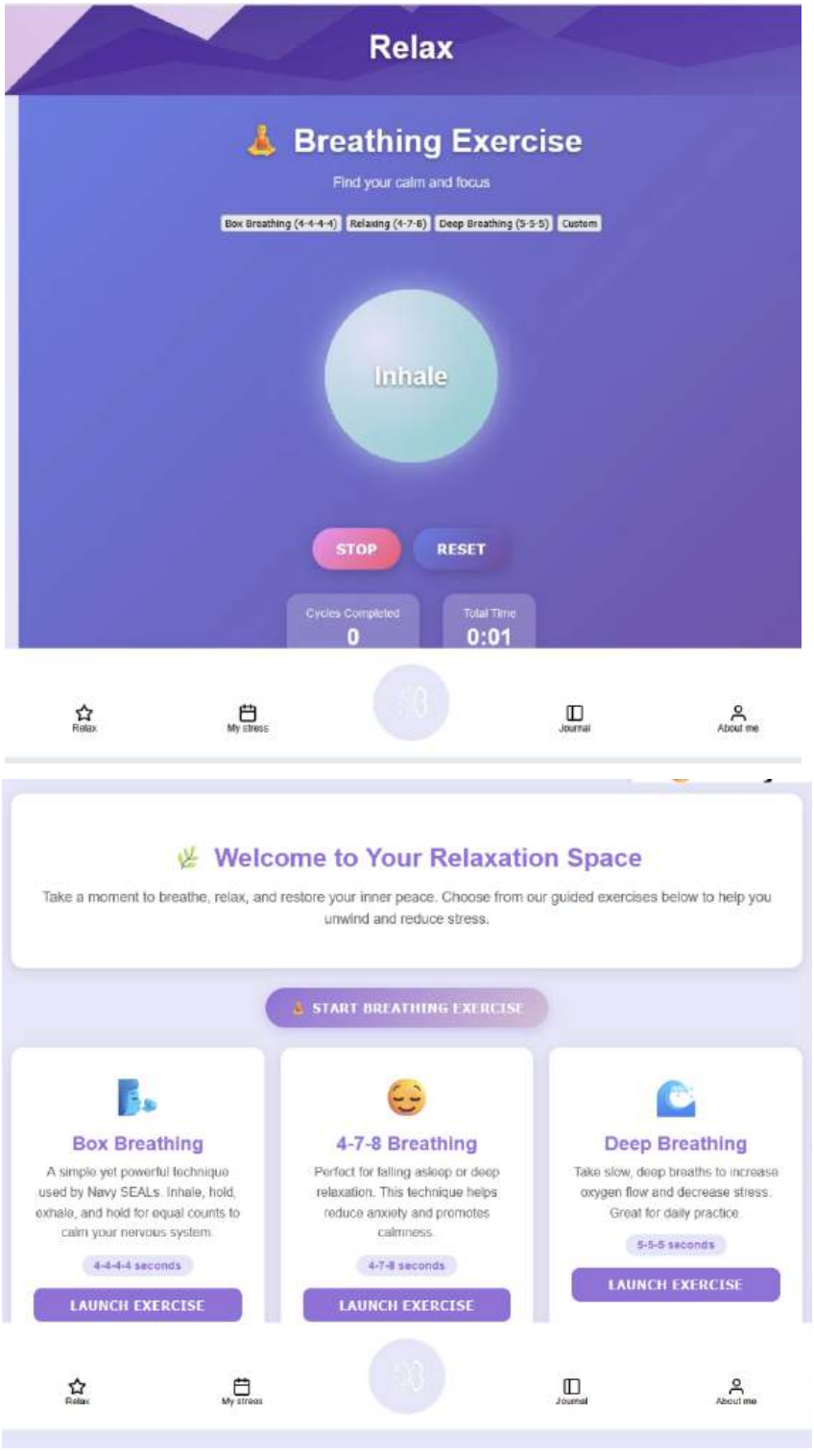


Fig 11a and 11b: (a) Breathing exercise on the relaxation slide; Options for Box Breathing, Relaxing, Deep Breathing, or custom exercises. (b) The relaxation settings page that offers each exercise, along with a description

Conclusion

In developing the BioVibe EEG headband, this study sought to determine whether a minimal two dry-electrode EEG system at Sp1 and Sp2 could feasibly capture and process meaningful brainwave signals in a wearable, out-of-clinic setting. Based on our findings, we conclude that it can. By collecting large lines of raw EEG data through the headband using two dry electrodes at Sp1 and Sp2 positions, the system consistently produced interpretable frequency band signals across multiple recording sessions, demonstrating that capturing stress-relevant brain activity using brainwaves does have potential.

This was not a foregone conclusion. Dry electrodes are known to be more prone to noise than gel-based alternatives, and a two-electrode configuration is far simpler than what most EEG research relies on. Artifacts caused by excessive eye blinking and facial movements further add to this challenge, as they can introduce noise that is difficult to separate from genuine brainwave activity. However, the fact that the system still produced meaningful shifts in frequency band activity, particularly during the emotional response validation sessions, suggests that the Sp1 and Sp2 positions carry sufficient signal to be useful even under these constraints.

What still remains to be researched is the use of these brainwaves for accurate stress classification. Establishing feasibility of data collection is an important first step, but translating that data into accurate stress scores is a future step that BioVibe plans and is willing to take. Future work will focus on classifying stress levels, incorporating complementary sensors such as GSR to improve accuracy, and validating the system with more tests on different people. We also aim to work with different experts in the field, including psychology professor Dr. Suri of San Francisco State University, and Dr. Wendy Ross, an experienced ER physician.

Works Cited

Attar, E. T. “Review of Electroencephalography Signals Approaches for Mental Stress Assessment.” *Neurosciences* 27, no. 4 (2022): 209–215. https://doi.org/10.17712/nsj.2022.4.20220025.

Bashivan, Pouya, Isaac Rish, and Stephan Heisig. “Mental State Recognition via Wearable EEG.” *arXiv*. February 2, 2016. https://arxiv.org/abs/1602.00985.

Can, Yekta Said, Bert Arnrich, and Cem Ersoy. “Stress Detection in Daily Life Scenarios Using Smart Phones and Wearable Sensors: A Survey.” *Journal of Biomedical Informatics* 92 (2019): 103139. https://doi.org/10.1016/j.jbi.2019.103139.

Louis, Elson K. S., L. Curtis Frey, Jeffrey W. Britton, Jennifer L. Hopp, Mohamad Z. Koubeissi, William E. Lievens, and Elaine M. Pestana-Knight. “The Developmental EEG: Premature, Neonatal, Infant, and Children.” In *StatPearls*. Treasure Island, FL: StatPearls Publishing, 2016. https://www.ncbi.nlm.nih.gov/books/NBK390356/.

“Mental Health & Gen Z.” *Springtide Research Institute*. September 29, 2023. https://springtideresearch.org/resources/mental-health-gen-z.

Nayak, C. S., and A. C. Anilkumar. “EEG Normal Waveforms.” In *StatPearls*. Treasure Island, FL: StatPearls Publishing, July 24, 2023. https://www.ncbi.nlm.nih.gov/books/NBK539805/.

The Impact of Influencer Marketing on Purchase Decision-making of Young Adults in Pakistan By Aliya Bashir

Abstract

Influencer marketing, as a social media strategy, has influenced the purchase decisions of young adults worldwide. Brands partner with individuals who have a credible social media presence to market their goods, and it's unique in every culture in which it is carried out. To understand the cultural nuance of Pakistan, the present study aims to examine how influencer marketing influences purchase decision-making among young adults in Pakistan. A quantitative approach was followed wherein an online survey was conducted with 56 young adults aged 15 to 20 years. Standardised tools were used, such as the Parasocial relation scale (Rubin et al. 1985), Influencer Trust, Influencer credibility and Purchase Intention (Khurana et al. 2025). It was found that males trusted influencers more when it came to branded goods and purchasing them, even though females considered influencers to be trustworthy and experts. The study showed that Instagram and YouTube are the top sites for influencer interaction, entertainment, and lifestyle. These findings provide insights for businesses and organisations that target young adults and rely on influencer marketing. It provides specific insights regarding the young adults living in Pakistan. Marketing professionals can also benefit from the findings of this study.

Keywords

Influencer marketing, Young adults. Pakistan, Social Media, and Behavioural Economics

Introduction

The development in technology and the internet as a whole has revolutionised communication between corporations and their audiences (Harahap et al. 2023). This has further led to the establishment of numerous digital technologies. Digital technologies have historically had a significant impact on various areas of commerce, including marketing (Yadav and Pavlou 132). The sales of products, works, or services with the aid of digital technologies are mainly conducted via the internet, using mobile phones and other media (Mort, Sullivan, and Drennan). The term "digital marketing" first surfaced in the 1990s and is the promotion of products or brands via one or more forms of electronic media (Minculete and Olar). It has evolved from one-way communication (television, print media) to two-way engagement, where consumers can comment, share, and directly interact with brands (Tankosic, Ivetic, and Vucurevic). This shift has increased the importance of authenticity, relatability, and peer influence in marketing (Orencia).

There are approximately 1.71 billion monthly active users globally; Facebook would likely be the most populous nation on earth if it were a country (Kassa, Cuevas, and Cuevas). The number of users for other social media corporations includes 70 million for Pinterest, 289 million for Twitter, 300 million for Instagram and Google+, and 380 million for LinkedIn (Blank

and Lutz). Moreover, while previously social media was thought of as a virtual teenagers' hangout, today users have extended limits beyond personal use (Livingstone).

Social media marketing has become prevalent in recent years, becoming indispensable in new marketing strategies (Berthon et al.). Social media usage has increased significantly over the past decade, transforming how individuals consume information and interact with brands (Hutton and Fosdick). Platforms such as Instagram, YouTube, and TikTok have become central to daily communication, especially among young adults (Woodward et al.). The way people use social media has evolved with the world (Miller et al.). People earlier used social media as a platform to publish personal information, including news, images, and videos (Bushey). According to Emily and Richard (2021), online interactions also simplified communication with one another, which can possibly be subject to change through other people's thoughts (Sin et al.). It is a fact that social media has changed people's routines, making them more dependent on mobile phones, technology and the internet (Licoppe and Smoreda).

In present times, children have a variety of options for digital entertainment and social media. YouTube in particular has emerged as a platform for children's entertainment and an alternative to traditional television (TV) content (Watson, 2019; Vogelman-NaCentre). As proven, 81% of U.S. parents let their children under 11 watch YouTube (Pew Research Centre, 2018), where they are exposed to brand advertisements before watching a video. (Weiss, 2018) (De Veirman, Hudders, and Nelson). YouTube offers newer avenues for brands to engage with children and their parents, including embedded advertising formats containing subtle brand communication in entertaining media content, making them less abrupt and thus harder to recognize (Hudders et al., 2017).

This widespread adoption has shifted marketing strategies from traditional mass advertising to more personalized, interactive, and digital forms of engagement (Tarabasz). Consumers are no longer passive recipients of advertisements; instead, they actively engage with content, brands, and creators (Berthon, Pitt, and Campbell). More recently, influencer marketing has emerged as a hybrid strategy that combines elements of content marketing, word-of-mouth marketing, and social proof (Cakim).

Influencer marketing is a relatively novel trend that has seen unnatural growth in popularity since the rise of platforms like YouTube, Instagram, and, more recently, TikTok (Haenlein 2020; De Jans 2020) (Taylor). An increasing number of people use these platforms regularly, form communities around their hobbies, passions, and interests (Cutolo and Grimaldi). Within these communities, certain established and respected opinion leaders called influencers emerge, forming a vast social network of their own (Stubb et al. 2019; Ki et al. 2020; Chopra et al. 2021) (Chen). The endless opportunities presented by such phenomena are hard to ignore, which is precisely why firms have started investing a great portion of resources towards pushing sponsored content through these influential advertising et al.). For them, the potential to reach a specific target audience is extremely valuable (Stubb et al. 2019; Boerman 2020), especially in today's digital environment where users are increasingly exposed to advertising campaigns, making it almost impossible for companies to stand out from their competitors. (Chopra et al.

2021). Influencer marketing benefits greatly from the credibility and relatability of the influencer, whom the followers feel personal attachment and affection towards (Boerman 2020; Coco & Eckert 2020) — a behaviour more commonly known as parasocial relationships (Sarawagi).

Influencers on social media are also a source of guidance and advice for audiences (Gómez) because of the perception that influencers have knowledge and expertise in the field, which can be used to encourage audiences to engage in purchases (AlFarraj et al.). Nearly half of the global Internet audience relies on influencer recommendations as a reference for their retail decisions online (Migkos, Giannakopoulos, and Sakas), which shows that audiences having enough confidence in influencers on social media increases the probability of a purchase (Ilieva et al.). Recent research found that 69% of companies plan to allocate most of their social media advertising budget to Instagram influencers because they can attract audiences to accept products.

The existing literature on influencer marketing is categorised into three main areas of research: (1) the influence maximization and influencer identification problem (Li and Du; More and Lingam; Roelens, Baecke, and Benoit; Sheikhahmadi, Nematbakhsh, and Zareie); (2) the study of practices correlating to managing online presence and influence (Audrezet, De Kerviler, and Moulard; Delisle and Parmentier; García-Rapp); and (3) the impact of social media influencers (SMIs) on consumer attitudes and behaviors (Lim et al.; Magno and Cassia; Torres, Augusto, and Matos; Vrontis et al.). Previous research considers the importance of popularity and opinion leadership in influence marketing (De Veirman, Cauberghe, and Hudders). Such influence can be increased by influencer-product congruence, although it is reduced in terms of perceived uniqueness in expertise and trust when the influencers enjoy high popularity and are associated with divergent products (Casaló, Flavián, and Ibáñez-Sánchez). In a marketing ecosystem where consumers trust each other far more than they trust companies' marketing communications, social media influencers have become prevalent advisors in consumer decision-making, particularly for younger audiences (Ki, Chow, and Li). Studies have also revealed that when consumers form a one-sided relationship, creating an illusion of intimacy with the influencer – a parasocial relationship – they are more likely to be influenced by them in their purchasing activities (Brink).

Despite the growing body of research on influencer marketing, most existing studies have been conducted in Western, individualist societies, meaning their findings cannot be straightforwardly applied to countries like Pakistan, where cultural values, gender dynamics, and social hierarchies operate differently. There is also a notable lack of research examining how influencer gender specifically affects consumer trust and purchasing behaviour among adolescent audiences in South Asia. Young Pakistani consumers represent a rapidly growing digital demographic, yet they remain largely absent from the academic literature on parasocial relationships. This study addresses that gap by focusing specifically on this population, offering insight that is both geographically and culturally distinct from what currently exists. This study aims to determine the association between parasocial relationships (PSR) with influencers and

consumer trust and purchase intention amongst young consumers of Pakistan. For this study, a quantitative method will be used, where data collection will be done using a survey. The survey will measure respondents' PIt with influencer trust and purchase intention.

Methods

Tools Used

The survey consisted of three main sections designed to measure how social media influencers impact consumer purchase decisions. The first section is the demographic section, which collects background information about the respondents, including age, gender, grade level, and location of residence. This section also includes general social media usage questions, such as which platforms are used, the amount of time spent on social media daily, and the types of influencers followed. This provides context for understanding the respondents' answers in the following sections. The second section examines how respondents perceive the influencers they follow through parasocial relationships, defined as the sense of connection and perceived intimacy between a media figure and an audience member. The Parasocial Interaction Scale developed by Rubin, Perse, and Powell (1985) was used to measure this construct. The scale includes multiple statements assessing emotional connection, closeness, familiarity, and interaction with the influencer. Responses were measured using a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), with higher scores indicating a stronger parasocial relationship (Rubin et al.).The third section assesses the trustworthiness and credibility of influencers and examines how respondents perceive the influencer's trustworthiness, expertise, and reliability when promoting products. This section also measures purchase intention. The items were adapted from a validated scale examining influencer credibility, brand trust, and purchase intention (Khurana et al., 2025). Responses were measured on a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree), with higher scores indicating greater trust in the influencer and a stronger intention to purchase products endorsed by them.

Sample

A total of 56 young adults participated in the study. About 50% (28) were female, and 50% (28) were male, below 21 years of age. About 48.2% (27) of the participants were aged between 15-16 years, 46.4% (26) were aged between 17-18 years, and 5.4% (3) of the participants were 19–20 years old. The majority of the participants were in grade 10 (16) and in grade 11 (21). Six participants were in grade 9, and fourteen of them were in grade 12. Only 2 were in their first year of graduation.

Ethical Consideration

The participants were informed about the overall aim and objectives of the study. The confidentiality and privacy of their responses were guaranteed to all participants. The

participants were made aware that the survey was completely optional and they could withdraw at any time. Participants who continued to be part of the study were provided with a digital explanation of each of these terms, and by continuing, they agreed to take part in the study. The survey was conducted through a self-administered online survey created using Google Forms, and all these terms were explained digitally to participants directly taking the survey.

Data Collection Procedure

The data collection was done through a self-administered survey, which was created on Google Forms. Google Forms is easily accessible to anyone with any digital device, allowing for a broader sample. The questionnaire was straightforward and to the point. The form included all questions as well as directions and explanations on how to answer each question. It informed participants about the study's purpose and how their answers would be used, ensuring transparency and informed consent. The survey link was shared with participants, who could complete it on their own schedule. Since the questions were clearly explained in the form, no extra help or verbal explanation was needed during the data collection. This approach made sure that all respondents received as much time as needed to answer each question and lowered the risk of researcher bias affecting the answers.

Data Analysis Strategy

The data analysis for this study focuses on understanding how parasocial relationships with influencers affect consumer trust and purchase intention among young consumers in Pakistan. Descriptive statistics were first used to summarise the data and provide an overall picture of the respondents, including their demographic characteristics and social media usage patterns. The responses from the 5-point Likert scale were converted into numerical values, and measures such as the mean, median, and mode were used to identify general trends in parasocial relationships, trust, and purchase intention. To make the findings clearer, graphical representations such as bar charts were used to present the data visually and highlight key patterns. For the inferential analysis, an independent-samples t-test (Social Science Statistics) was conducted to examine whether there were any significant differences between male and female respondents. This test compares the average responses across gender groups to determine whether the differences observed are meaningful or simply due to chance. A p-value of less than 0.05 was used as the threshold for statistical significance, indicating that gender has a noticeable effect on the variables being studied.

Results and Discussion

In this, it was found that about 56%(26) people used social media for 4-6 hours a day, whereas 32% (18) people used social media for 1- 3 hours. 18 (10)participants reported using 7-9 hours, and 2 people responded using more than 9 hours. Participants were then asked about the social media sites they frequently used. About 98% (55) of respondents said they use Instagram, making it the most popular site. YouTube, which was utilised by 85%(48) participants, came next. Facebook 14% (8) and TikTok 41% (23) were additional platforms. Only 2 respondents

said they used Netflix and Snapchat, respectively. These findings suggest that participants mostly come across influencer content on YouTube and Instagram.

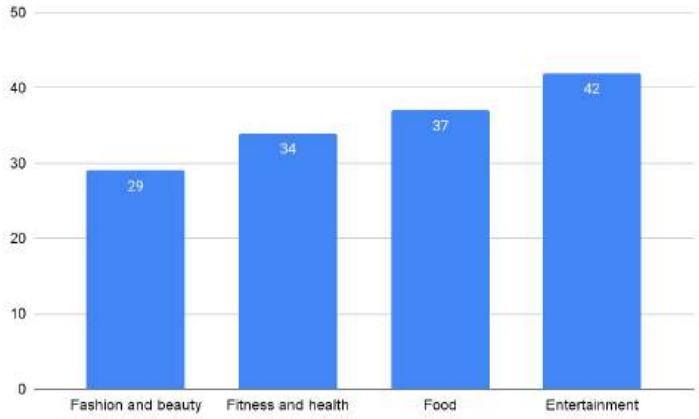


Figure 1: *Graphical representation of preferred influencer categories among young Pakistani consumers (n= 56)*

In Figure 1, it can be observed that entertainment is the most preferred influencer category, receiving the highest number of responses, 75% (42). This is followed by food with 66% (37) responses and fitness and health with 60% (34) responses. In contrast, fashion and beauty received the fewest responses at 52% (29). A clear trend that can be identified is that respondents show a stronger preference for lifestyle and entertainment-based content over more traditional aesthetic-focused categories such as fashion and beauty. This may suggest that young consumers are more drawn to content that is engaging, relatable, and integrated into everyday experiences rather than content that is purely appearance-driven.

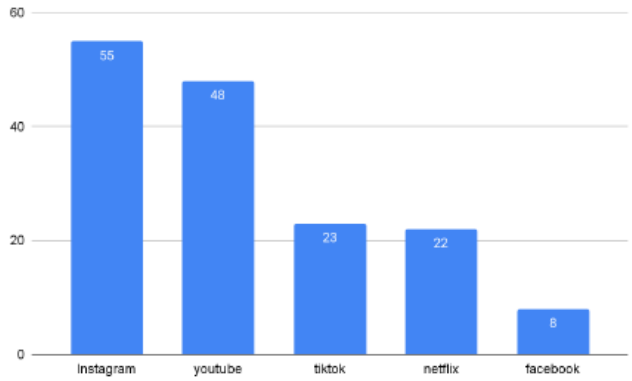


Figure 2: *Graphical representation of social media platform usage for the following influencers (n=56)*

As seen in Figure 2, Instagram is the most commonly used platform for following influencers, with the highest number of responses (98%). This is followed by YouTube with 85%

(48) responses. Platforms such as TikTok 41% (23) and Netflix 39%(22) show significantly lower engagement levels, while Facebook has the lowest usage with only 14%(8) responses. One trend that can be observed is that respondents tend to prefer highly visual and interactive platforms for engaging with influencers. Instagram and YouTube dominate usage, suggesting that platforms offering short-form and long-form video content are more effective in capturing user attention. In contrast, Facebook's low response rate indicates a decline in its relevance among younger consumers.

Table 1: *Independent t-test analysis of trustworthiness, expertise, brand trust, purchase intention, willingness to pay and parasocial interactions among male and female consumers.*

Variable	Gender	N	Mean	SD	T	P
Trustworthiness	Female	28	11.50	7.74	1.49	0.07
	Male	28	10.32	5.74		
Expertise	Female	28	11.43	7.74	1.23	0.11
	Male	28	10.39	12.25		
Brand Trust	Female	28	10.93	9.48	-0.38	0.03*
	Male	28	11.21	6.10		
Purchase Intention	Female	28	10.46	6.55	-1.87	affect
	Male	28	11.86	8.94		
Willingness to Pay	Female	28	23.21	35.14	-1.54	0.07
	Male	28	25.50	26.93		
Parasocial Interactions	Female	28	47.64	127.42	-0.59	0.28
	Male	28	49.25	81.82		

Based on the data, an independent-samples t-test was conducted to examine how influencers affect consumer trust and purchase intention among young Pakistani consumers on the basis of gender. Females reported higher mean scores for trustworthiness (M = 11.50, SD=7.74) and expertise (M = 11.43, SD = 7.74) compared to males; however, neither difference were statistically significant, with results of t(26) purchasing.05 and t(26) = 1.23, $p>0.05$, respectively. Conversely, brand trust was significantly higher among male consumers (M = 11.21, SD = 6.10) than females (M = 10.93, SD = 9.48), t(26) = -0.38, $p< 0.05$. This trend extended to purchase intention, where male consumers (M = 11.86, SD = 8.94) had higher intent of purchasing goods than females (M = 10.46, SD = 6.55), t(26) = -1.87, $p<0.05$. While males also saw higher means in willingness to pay (M = 25.50, SD = 26.93) and parasocial interactions (= 49.,25, SD = 81.82), these variables were not statistically significant, yielding t(26) = -1.54,

$p>0.05$ and $t(26) = -0.59$, $p>0.05$. These findings suggest that while perceived trustworthiness is relatively balanced, male influencers may hold a stronger statistical advantage in driving actual brand

Hence, in this study, male influencers scored significantly higher on brand trust and purchase intention among the respondents, even though female influencers were rated slightly higher on trustworthiness and expertise. One explanation for this is that in Pakistan, cultural associations between male voices and authority may make male influencers feel more credible when it comes to actually recommending a product. (Boerman 2020) made a similar observation, noting that liking an influencer and trusting their product recommendations are not the same thing; the latter depends more on perceived social authority than emotional closeness. Ki et al. (Boerman 2020) also found that credibility rooted in expertise tends to drive purchasing decisions more than relatability alone (Ki, Cuevas, Chong, and Lim). The fact that Instagram and YouTube were by far the most used platforms among respondents is relevant here too. Chopra et al. (2021) argued that visually rich, interactive platforms intensify parasocial bonds, which in turn makes influencer recommendations feel more personal and trustworthy (Chopra, Avhad, and Jaju).

For brands trying to reach young consumers in Pakistan, these findings suggest that influencer gender is worth considering when planning campaigns, especially if the goal is to drive actual purchases rather than just awareness. Male influencers in this sample showed a stronger statistical link to brand trust and purchase intention, which has real consequences for how marketing budgets are allocated. Outside of commercial strategy, the study also points to something worth paying attention to at a wider level. Most of what we know about parasocial relationships and consumer behaviour comes from research done in Western countries — Ki et al. (2020) and Boerman (2020) both draw primarily from Western samples. This study shows that cultural context — including how gender shapes credibility — plays a role that existing models do not fully account for. That has implications not just for marketers but for researchers, educators, and anyone thinking about how adolescent spending behaviour is shaped by social media (Ki et al.; Boerman).

However, the study also has a few limitations. The sample size of 56 is small, and the fact that most participants were high school students completing an online survey means the findings cannot easily be extended to Pakistani youth as a whole. Self-reported surveys also carry the risk that respondents answer in ways they think are expected rather than how they actually behave, particularly on questions about emotional attachment to influencers — a concern noted by Chopra, Avhad, and Jaju in similar adolescent-facing research. The study was likely skewed toward urban, digitally active respondents, so the experiences of young people in smaller cities or with less consistent internet access are not reflected here. Influencer gender was the main variable examined, but things like follower count, the type of content an influencer posts, and how long someone has followed them could all affect the results in ways this study did not measure — factors that Ki et al. (2020) identify as significant in shaping follower attachment.

Because the data was collected at one point in time, it is also not possible to say whether parasocial relationships cause changes in consumer behaviour or simply correlate with them.

Conclusion

The current research was designed to investigate the influence of parasocial relationships with social media influencers on consumer trust and purchase intention among young consumers of Pakistan. With the rise of influencer marketing on digital platforms, particularly Instagram and YouTube, academics have increasingly examined how one-way emotional relationships that audiences develop with influencers lead to tangible consumer outcomes. Instagram emerged as the dominant platform for influencer engagement, followed by YouTube, with entertainment and lifestyle content attracting the largest audiences. Male consumers showed more brand trust and purchase intention, suggesting that in the Pakistani adolescent context. There are several directions for future investigation. First, a larger and more geographically diverse sample would allow the study to be conducted with greater generalisability to all Pakistani youth, including those in smaller cities and rural areas where social media usage patterns may be different. Second, future work could examine the influence of influencer niche – e.g. whether parasocial dynamics work differently for fashion vs. fitness or food influencers – given that this study found differences in the content categories that respondents followed most. Third, longitudinal studies would give us the chance to see if parasocial relationships develop over time and if a longer exposure to an influencer has a lasting effect on purchase intention, which cross-sectional designs cannot capture. Fourth, qualitative methods could complement the quantitative findings. For instance, interviews or focus groups could explore why young Pakistani consumers trust some influencers more than others, and how they consciously process sponsored content versus organic recommendations. Finally, future studies could look into the moderating role of socioeconomic background, as willingness to pay is likely to be influenced as much by financial capacity as by psychological attachment, and this variable was not controlled for in the present study.

Works Cited

AlFarraj, Omayma, et al. "Examining the impact of influencers' credibility dimensions: attractiveness, trustworthiness and expertise on the purchase intention in the aesthetic dermatology industry." Review of International Business and Strategy 31.3 (2021): 355-374.

Audrezet, Alice, Gwarlann De Kerviler, and Julie Guidry Moulard. "Authenticity under threat: When social media influencers need to go beyond self-presentation." Journal of Business Research 117 (2020): 557-569.

Berthon, Pierre, Leyland Pitt, and Colin Campbell. "Ad lib: When customers create the ad." California Management Review 50.4 (2008): 6-30.

Berthon, Pierre R., et al. "Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy." Business Horizons 55.3 (2012): 261-271.

Blank, Grant, and Christoph Lutz. "The social structure of six major social media platforms in the United Kingdom." Proceedings of the 7th International Conference on Social Media & Society. 2016.

Brink, Sarah Emilie Even. "Parasocial Relationships: Unveiling One-Sided Intimacy in a Digital Age." (2024).

Bushey, Jessica. "Trustworthy citizen-generated images and videos on social media platforms." 2015 48th Hawaii International Conference on System Sciences. IEEE, 2015.

Cakim, Idil M. Implementing Word of Mouth Marketing: Online Strategies to Identify Influencers, Craft Stories, and Draw Customers. John Wiley & Sons, 2009.

Casaló, Luis V., Carlos Flavián, and Sergio Ibáñez-Sánchez. "Influencers on Instagram: Antecedents and consequences of opinion leadership." Journal of Business Research 117 (2020): 510-519.

Chen, Peng. Exploring the Relationships Between Social Media Influencers and Consumer Purchase Intentions in the Chinese Fashion Industry. Diss. University of Wales Trinity Saint David, 2025.

Cutolo, Donato, and Rosa Grimaldi. "I wasn't expecting that: How engaging with digital platforms can turn leisure passion into entrepreneurial aspirations." Journal of Business Venturing Insights 20 (2023): e00404.

De Veirman, Marijke, Liselot Hudders, and Michelle R. Nelson. "What is influencer marketing and how does it target children?" Frontiers in Psychology 10 (2019): 498106.

De Veirman, Marijke, Veroline Cauberghe, and Liselot Hudders. "Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude." International Journal of Advertising 36.5 (2017): 798-828.

Delisle, Marie-Pier, and Marie-Agnès Parmentier. "Navigating person-branding in the fashion blogosphere." Journal of Global Fashion Marketing 7.3 (2016): 211-224.

García-Rapp, Florencia. "Come join and let's BOND: Authenticity and legitimacy building on YouTube's beauty community." Journal of Media Practice 18.2-3 (2017): 120-137.

Gillin, Paul. The New Influencers: A Marketer's Guide to the New Social Media. Linden Publishing, 2007.
Gómez, Alexandra Ruiz. "Digital Fame and Fortune in the Age of Social Media." aDResearch 19 (2019): 8-29.
Harahap, Muhammad Ade Kurnia, et al. "The impact of digital technology on employee job stress." Jurnal Pendidikan Tambusai 7.1 (2023): 3635-3638.
Hutton, Graeme, and Maggie Fosdick. "The globalisation of social media." Journal of Advertising Research 51.4 (2011): 564-570.
Ilieva, Galina, et al. "Social media influencers: Customer attitudes and impact on purchase behaviour." Information 15.6 (2024): 359.
Kassa, Yonas Mitike, Rubén Cuevas, and Ángel Cuevas. "A large-scale analysis of Facebook's user base." IEEE Access 6 (2018): 78881-78891.
Khurana, Tripta, Sonia Pannu, Garima Dalal, Pooja Vyas, and Priya Rani. "How Do Social Media Influencers' Credibility and Brand Trust Drive Purchase Intentions for Green Cosmetics? Insights from SOBC Approach." 2025. - second scale
Latif, Muhammad Zahid, et al. "Use of smartphones and social media in medical education." Acta Informatica Medica 27.2 (2019): 133.
Li, Feng, and Timon C. Du. "Maximising micro-blog influence in online promotion." Expert Systems with Applications 70 (2017): 52-66.
Licoppe, Christian, and Zbigniew Smoreda. "Are social networks technologically embedded?" Social Networks 27.4 (2005): 317-335.
Lim, Xin Jean, et al. "The impact of social media influencers on purchase intention." Asian Journal of Business Research 7.2 (2017): 19-36.
Livingstone, Sonia. "Teenagers' use of social networking sites." New Media & Society 10.3 (2008): 393-411.
Magno, Francesca, and Fabio Cassia. "The impact of social media influencers in tourism." Anatolia 29.2 (2018): 288-290.
Maitri, Wayan Sri, et al. "Recent trends in social media marketing strategy." Journal Minfo Polgan 12.1 (2023): 842-850.
Mallipeddi, Rakesh R., et al. "A framework for analysing influencer marketing." Management Science 68.1 (2022): 75-104.
Migkos, Stavros P., et al. "Impact of influencer marketing on consumer behaviour." Journal of Theoretical and Applied Electronic Commerce Research 20.2 (2025): 111.
Miller, Daniel, et al. How the World Changed Social Media. UCL Press, 2016.
Minculete, Gheorghe, and Polixenia Olar. "Approaches the modern concept of digital marketing." 2018.
More, Jyoti Sunil, and Chelpa Lingam. "A SI model for social media influencer maximisationn." Applied Computing and Informatics 15.2 (2019): 102-108.
Mort, Gillian Sullivan, and Judy Drennan. "Mobile digital technology: Emerging issue for marketing." Journal of Database Marketing 10.1 (2002): 9-23.

Orencia, Abrian Joy. "Authenticity, relatability, and cultural influences in marketing." Psychology and Education 15 (2023): 675-679.
Rubin, Alan M., Elizabeth M. Perse, and Robert A. Powell. "Loneliness, Parasocial Interaction, and Local Television News Viewing." *Human Communication Research*, vol. 12, no. 2, 1985, pp. 155–180. - This is one of the scales
Roelens, Iris, Philippe Baecke, and Dries F. Benoit. "Identifying influencers in a social network." Decision Support Systems 91 (2016): 25-36.
Sarawagi, Rishika. Impact of Social Media Influencers on Consumer Well-Being. MS thesis. MICA, 2024.
Sheikhahmadi, Amir, et al. "Identification of influential users in social networks." Physica A 486 (2017): 517-534.
Sin, Frances, et al. "Digital social interaction in older adults." Proceedings of the ACM 5 (2021): 1-20.
Social Science Statistics. "T-Test Calculator for 2 Independent Means." *Social Science Statistics*, www.socscistatistics.com/tests/studentttest/default.aspx. Accessed 10 June 2026
Stoldt, Ryan, et al. "Professionalizing and profiting: The rise of intermediaries." Social Media + Society 5.1 (2019).
Tankosic, Mirjana, Pavle Ivetic, and V. Vucurevic. "Features of interactive public relations." International Journal of Economics and Management Systems 1.1 (2016): 290-295.
Tarabasz, Anna. "The impact of digital on marketing strategy." CRC Press, 2024.
Taylor, Zari Alyssa. "TikTok as an App Is Not Friendly to Black Creators." Diss. UNC Chapel Hill, 2024.
Torres, Pedro, et al. "Antecedents and outcomes of digital influencer endorsement." Psychology & Marketing 36.12 (2019): 1267-1276.
Vogelman-Natan, Kalia. The YouTube Generation. Diss. Northwestern University, 2025.
Vrontis, Demetris, et al. "Social media influencer marketing: A systematic review." International Journal of Consumer Studies 45.4 (2021): 617-644.
Woodward, Matthew J., et al. "Time spent on social media and mental health." Journal of Technology in Behavioral Science 10.3 (2025): 661-671.

Using Water as A Solvent to Extract Betalains From Beet By Timothy Park, Junehyuk Lee, Irene Kim, Peter Seok, Ryan Kim, Jennifer Na, Trevor Kim, Ezra Kim

Summary / Abstract

Extraction is a method that scientists and engineers use to extract wanted chemicals from plants and vegetation. Within extraction, we used a solvent, or liquid-liquid, method for the extraction of betalains, which consists of betacyanins and betaxanthins. Betalains, the primary pigment responsible for the vibrant red-violet coloration in beet, consists of a sugar ring combined with an aromatic structure containing two carboxylic acids which is very soluble in polar solvent, and betacyanins and betaxanthins also contain two carboxylic acids and positively charged nitrogen thus making both very polar compounds as well. We hypothesized that betacyanins, and betaxanthins will not be able to be extracted easily in non-polar solvents; therefore, the more polar, the better the extraction will be. For this reason, we chose heptane, isopropyl alcohol, ethanol, and water in increasing polarity to compare the extraction results. After many experimentations, we found that heptane barely extracted anything. Isopropyl alcohol(IPA) extracted compound showed yellow color indicating that small amounts of betalanines were present while ethanol extracted compound showed darker yellow color due to more extracted betalanines. As for the water, dark red color was obtained thus indicating that a higher amount of betalanines were present. As we hypothesized, the polarity of solvent has a higher impact of extracting Betalains.

Introduction

The beetroot, or beta vulgaris as it's called scientifically, evolved from the wild sea beet, found along coastlines from Britain to India. Its botanical name, beta, comes from the Celtic 'bett' meaning 'red'.(1) Beetroot is biennial, meaning that full maturity is reached after 2 years of growth, but is grown and harvested annually, both for its bulbous roots and leaves, and grows well in well-drained, fertile soils. Sea beets first became domesticated in the countries of the Eastern Mediterranean and in the Middle East. The ancient Greeks and Romans ate the leaves and used the roots medicinally. The Greek physician Hippocrates bound wounds with beetroot leaves, and the plants were also used to treat fevers, skin conditions and constipation (1).The beetroot's dark red color comes from a pigment called betalains, which contains nitrogen and is water-soluble. Betalains can be divided into two distinct classes: the red-violet betacyanins and the yellow-orange betaxanthins (2). The dominant pigment of the betacyanins is the Betalains, which gives the beetroot its striking red color. Betalains is the only approved betalain for food, cosmetic, and pharmaceutical use by the FDA (3)

A solvent is any substance, generally a liquid, that has the ability to dissolve another substance, which in turn creates a solution. There are mainly two different types of solvents regarding polarity: polar and non-polar solvents. The solvent we decided to use is water, which is polar and is often considered the 'universal solvent' (4). A solvent works by dissolving a solute through a chemical reaction. In our case, the solvents are water, heptane Isopropyl alcohol (IPA),

ethanol, and the solute is grinded beet. During this process, the solute's molecules break their bonds, and then rebonding with the solvent's bonds to create a solution, which is essentially the final product after the entire reaction occurs. All of this occurs due to the solute molecules's attraction to solvent molecules (5).

Solid-liquid extraction is a process where one or more components of a solid mixture blends into a liquid solvent and the outcome may not always be liquid. Extraction can either collect useful substances or wash away unwanted substances (6). Steam distillation is another way to extract and it's a high-yield extraction method used for most aromatic plants, however it only operates under low pressure and it requires longer distillation time to extract and vaporize. Where aromatic plants need lower temperatures, it involves bubbling the steam through suitable plants, whereas hydrodistillation submerges plants in water before extended distillation (7).

The method of extraction we used in our experimentation was solvent extraction. Solvent extraction, also known as liquid-liquid extraction, is a method of extraction used to separate metal complexes or compounds, based on their relative solubilities - their ability to be diluted - in two different immiscible liquids, or unmixable fluids (8). Essentially, the solvent enters a solid environment i.e. a vial or container, making contact with a solute, which dissolves in the solvent. The dissolved solute is then transferred out of the container, yielding the final product, a solution, as mentioned before (9). However, choosing which solvents to use based on their polarity is very significant. Generally, polar solvents dissolve polar solutes most effectively, while non-polar solvents dissolve non-polar solutes most effectively (10). Thus, a polar polar solvent like water is effective in extracting Betalains, betacyanins, and betaxanthins, which are also polar. After testing five different solvents to extract Betalains, water as the solvent gave the highest yield due to being the most polar solvent as we expected.

Materials and methods

General Procedure

We began by crushing dehydrated beets (Nature Jam, B0BTMQQ8FP)) and placing them in a 20 mL vial (Kesell, B07Q82RLBY)), with the solvents (water, ethanol (Lab Alley, B0CBP9H245) IPA (Amazon Basics, B07NFSFBXQ)) and Heptane (Amazon, C051FSFBXQ). The solution was boiled at 45 ℃ for 10 minutes before cooling. After, solids were filtered using coffee filter #2 (Amazon Basics - (Amazon, B09MRWMMLW) and transferred into a 20 mL vial once again. This liquid was boiled for 20 minutes to evaporate all the water and only left with thick liquid. Once thick liquid is left, we weighted the vial to see how much we were able to retrieve.

Results/ Discussion

In our experiments, we used various solvents to extract Betalains from beet by boiling the solution on a hot plate for polarity of solvent comparison. First, to test if the most polar solvent, water, will give better yield than other solvents, we used 10g of grinded beet mixed with 5 mL of

each solvents and boiled the solution at 45 ℃ for 10 minutes. After the experiment we found that water gave the best extraction due to being the most polar solvent. As for ethanol and IPA, Ethanol extracted beet showed red pigments due to being the second polar solvent and IPA extracted beet was yellow being less polar than ethanol due to having more bulkier hydrocarbons. heptane did not contain any Betalains due to being highly non-polar solvent (**Figure 1**). heptane is only made of hydrocarbons which do not contain any hydrogen bond or ionic bond; therefore, heptane did not dissolve any significant amount of Betalains. Since the data show that the more polar compounds dissolved Betalains better, we moved to finding good conditions to extract Betalains using water. We omitted other solvents: IPA, Ethanol, and heptane because water extracted Betalains much better than those solvents. First, we tested the amount of water per 10 g of beet, we used 1 mL, 3 mL, 5 mL, 10 mL, and 30 mL. We heated the solution to 45 ℃ for 10 minutes before evaporating all the water. We were able to retrieve 1.2 mg, 3.2 mg 5.4 mg, 5.5, and 5.4 mg respectively (**Table 1**). We can get more Betalains using more than 5 mL of water; however, evaporating more water requires more time, so we decided to use 5 mL for the rest of the experiments. Next we did temperature testing. We chose 35 ℃, 45 ℃, 55 ℃, 65 ℃, 100 ℃. We discovered that every temperature gave a very similar amount of Betalains ~5.4 mg, but the difference is time of extraction. For 35 ℃, we obtained 5.1 mg of Betalains after 30 minutes of maintenance. For 45 ℃, it took 10 minutes; for 55℃, it took 7 minutes; for 65 ℃, it took 5 minutes; and for 100 ℃, it took 2 minutes (**Table 2 & Figure 2**). As we hypothesized,

Conclusion

As we hypothesized, polar solvents will extract Betalains much better due to Betalains containing many hydroxyl groups and carboxylic acid. Out of water, ethanol, IPA, and heptane, water extracted Betalains much more efficiently due to being very polar solvent compared to the other three.

Acknowledgements

We would like to thank Gina Kim, Jiyeon Hwang, and Jane Hong for their advice and one-on-one help with our research.

Works Cited

1.“Fascinating Facts and Figures: Beetroot / RHS Gardening.” Www.rhs.org.uk, www.rhs.org.uk/advice/grow-your-own/features/FFF-Beetroot.

2.Sadowska-Bartosz, Izabela, and Grzegorz Bartosz. “Biological Properties and Applications of Betalains.” Molecules, vol. 26, no. 9, Apr. 2021, p. 2520, https://doi.org/10.3390/molecules26092520.

3.Vieira Teixeira da Silva, Davi, et al. “Betalains, a Natural Food Additive: Stability, Bioavailability, Antioxidant and Preservative Ability Assessments.” Molecules (Basel, Switzerland), vol. 24, no. 3, MDPI, 2019, p. 458, https://doi.org/10.3390/molecules24030458.

4.EcoOnline. “Solvent Meaning & Definition | EcoOnline.” EcoOnline, 15 Dec. 2024 www.ecoonline.com/glossary/solvent/.

5.“Solvent | Research Starters | EBSCO Research.” EBSCO, 2024, www.ebsco.com/research-starters/chemistry/solvent.

6.Berk, Zeki. “Solid Liquid Extraction - an Overview | ScienceDirect Topics.” Sciencedirect.com, 2013, www.sciencedirect.com/topics/engineering/solid-liquid-extraction.

7.“ExtractionMethods.” Edenbotanicals.com, 2026, www.edenbotanicals.com/extraction-methods?srsltid=AfmBOoqjDwv9vLt3bo4ng4Ijf33rrgDYrfHp3i3XiKE_HqkgjTnombGs. Accessed 4 May 2026.

8.“Solvent Extraction - an Overview | ScienceDirect Topics.” Www.sciencedirect.com, 2013, www.sciencedirect.com/topics/chemical-engineering/solvent-extraction.

9.Zhang, Qing Wen, et al. “Techniques for Extraction and Isolation of Natural Products: A Comprehensive Review.” Chinese Medicine, vol. 13, no. 1, Apr. 2018, https://doi.org/10.1186/s13020-018-0177-x.

10.“Solubility Factors When Choosing a Solvent.” Caymanchem.com, Cayman, 2026, www.caymanchem.com/news/solubility-factors-when-choosing-a-solvent?srsltid=AfmBOoqhE7Cb5gWc_RV2MHlK_qAF070LFigrpzErW3tDnV4_7jzF2pcb. Accessed 23 Apr. 2026.

Figures and Figure Captions

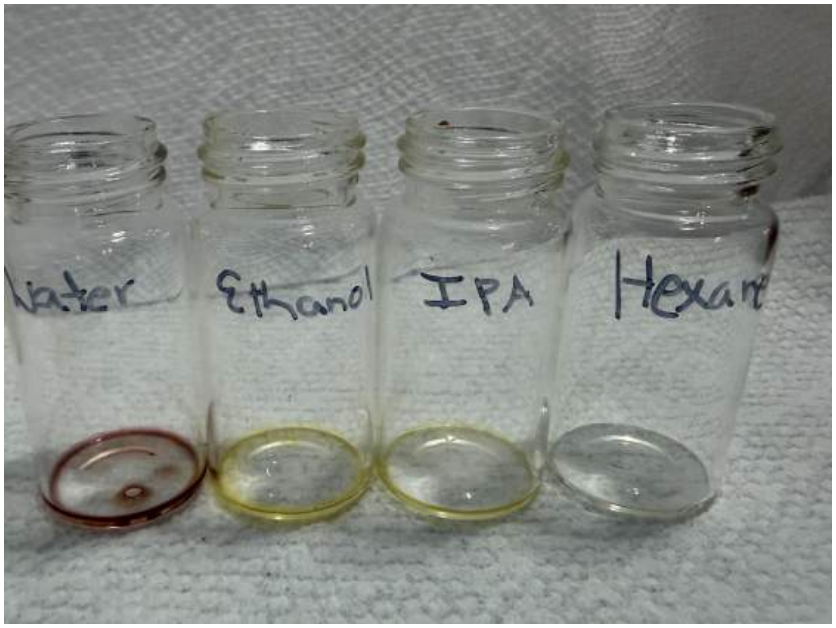


Figure 1. Beet extraction using various solvents. This figure shows extraction of beet using water, ethanol, isopropyl alcohol, and heptane. Water being the most polar solvent and heptane being the least polar. Since ethanol is more polar than IPA, it shows darker yellow due to having a bit more Betalains compared to IPA.

Trials	Amount of Water (mL)	Extracted Betalains (mg)
1	1	1.2
2	3	3.2
3	5	5.4
4	10	5.5
5	30	5.4

Table 1. Solvent amount testing. We increased water amount to see how much we can extract Betalains from beet at 45 ℃ for 10 minutes

We chose 35 ℃, 45 ℃, 55 ℃, 65 ℃, 100 ℃. We discovered that every temperature gave a very similar amount of Betalains ~5.4 mg, but the difference is time of extraction. For 35 ℃, we obtained 5.1 mg of Betalains after 30 minutes of maintenance. For 45 ℃, it took 10 minutes; for 55℃, it took 7 minutes; for 65 ℃, it took 5 minutes; and for 100 ℃, it took 2 minutes (**Table 2 & Figure 2**). As we hypothesized,

Trials	Temperature (℃)	Optimal time (min)
1	35	30
2	45	10
3	55	7
4	65	5
5	100	2

Table 2. Optimal boiling time. Table 2 contains experiments at different temperatures to see how long it will take to extract Betalains from beet optimally.

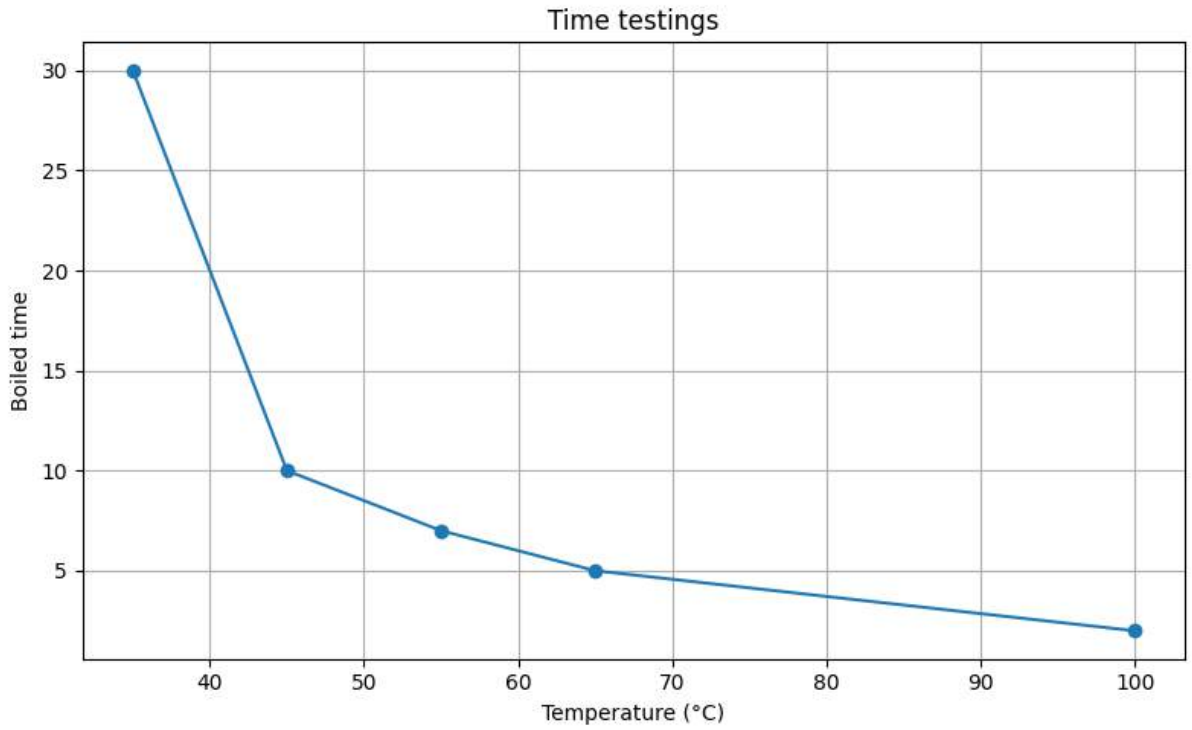


Figure 2. Temperature vs extracted Betalains. This graph shows how long it will take to extract Betalains from beet for the optimal boiling time.

Quantifying Color Use in European Painting, 1230–1950: A Computational Analysis of Long-Term Artistic Change By Gan Chen

1. Abstract

Color has been the focus of European painting research for a long time, and it not only played a role in visual composition, but also carried cultural connotations (Gage, 1993). Art historians generally believe that the change of color is inevitably related to changes in aesthetic preferences, art techniques, and patronage (Ball, 2001). In the development of medieval painting, the color is relatively monotonous, but with the development of industrial technology, the color is more and more abundant, which is inevitably related to the historical context (Ball, 2001).

In the process of writing art history, attention should be paid to analyzing the role of color in the development of a specific period and movement. In the Renaissance period, the color of the painter's work is balanced and harmonious, which is consistent with the naturalism theory (Baxandall, 1988). In Baroque art, the contrast is more intense, and the emotional expression is stronger (Encyclopaedia Britannica, 2026). In the modern period, some theorists believe that color can express the inner state of people, not just describe (Kandinsky, 1911). However, these interpretations are typically derived from close visual analysis of individual works and do not establish whether such patterns hold consistently across large datasets.

1. 1 Literature Review

In recent years, the rapid development of digital technology and the continuous upgrading of image analysis technology have changed the traditional research mode, and the construction of visual culture has achieved breakthrough progress. With the support of high-resolution images and the promotion of metadata, the traditional research mode can not meet the needs of the development of the times, and the open access initiative launched by the Metropolitan Museum of Art can provide more public domain images for the public, which can support large-scale comparative analysis (The Metropolitan Museum of Art, 2017).

With the rapid development of data science and computer vision, it is possible to extract visual features from images in a systematic and repeatable way, and the extracted visual features are quantifiable, including brightness, saturation and other contents. In the field of computational art history, scholars have constructed large data sets to analyze whether the visual features obtained through measurement or learning are consistent with the historical categories, and to find that quantitative analysis can reveal the structure corresponding to art history (Elgammal et al. , 2018). Digital art history also points out that the updating of digital standards and the rapid development of computer vision have brought changes to scholars in the study of large image corpora, and scholars are still debating the responsibility of interpretation and the development of disciplines (Brey, 2021).

In the development of digital art history, scholars have repeatedly pointed out that the quality of digital reproduction is relatively low, and there are obvious differences between

different images (Zorich, 2012). In the process of digitization, the color management, lighting and so on are different, which will have an impact on the measurement of color characteristics. Therefore, it is necessary to have a sound digital standard system, which requires the use of unified workflow and reasonable color management (FADGI, 2023). The above problems are the focus of attention in the study of museum images (FADGI, 2023).

1. 2 Gap

Although there are breakthroughs in methodology, the interpretation of color in the development of art history is still relatively qualitative, and it is difficult to systematically analyze the changes in a long historical period. Most of the computational work in the process of analyzing the paintings focuses on classification and prediction, but does not analyze the changes of multiple colors from the time dimension (Elgammal et al. , 2018). The development of digital art history has gradually improved the standards, tools and interpretation framework, and how to coordinate the relationship between computational description and humanistic explanation has become a key issue (Brey, 2021).

In order to make up for this deficiency, this project aims to use large-scale calculation methods to analyze the development of European painting from 1230 to 1950, and to grasp the evolution of color in this period. In this study, 1700 public domain works were selected from the Metropolitan Museum of Art to analyze the color from four dimensions: saturation, brightness, hue distribution and style color deployment. Through the analysis of the above four aspects, the long-term development trend is grasped, and the color practice of painters in the past eight centuries is described with the help of data. This method can not only analyze the color from the macro level, but also analyze the visual level, which can better explain the narrative of artistic change.

2. Method and Process

2. 1 Data Construction and Processing

In this study, the data set is constructed, which includes metadata and image data, and the color evolution of European paintings is analyzed in depth. The data set includes 1781 European paintings, with a time span of more than 700 years, and the data comes from the Open Access collection of the Metropolitan Museum of Art. Each painting has a unique digital image and metadata, including the name of the work, the name of the author, the production date, etc.

In order to make the analysis more reasonable, the average production year is calculated as the midpoint between the start date and the end date, and the paintings are divided into several groups according to the decade variable. In this way, the noise of a single work can be reduced, and the style trajectory of the seven centuries can be analyzed.

All color indicators can be extracted from image data. Open source Python libraries can be used in this process, which can load images (Clark, 2015), process numerical data (Harris et al. , 2020), and transform color space (van der Walt et al. , 2014).

In order to ensure the comparability of the obtained images and to unify the resolution and size, all color variables should be standardized, and the pixel share and average value should be unified. This is a normal practice to ensure that the measurement results are related to the absolute pixel count and the relative color structure. In the process of digital reproduction, there are many influencing factors, and the measurement results are inevitably affected by artistic decisions and digital technology. In the process of sample aggregation, the influence of image variation is relatively small, and the historical pattern can be obtained.

See **Table A1** in **Appendix A** for a complete list of variables, definitions and descriptive statistics.

2. 2 Feature Extraction Methods

In order to make the color have a multi-dimensional connotation, the image features are divided into four parts: luminance, saturation, hue and style.

2. 2. 1 Luminance-based metrics

In order to grasp the brightness and tone structure, the RGB image is converted into relative luminance by using the standard perception luminance equation (Fairchild, 2013):

$$L = 0.\ 2126R + 0.\ 7152G + 0.\ 0722B \quad (1)$$

In this study, the normalized channel value is taken as the research object, and the luminance map is constructed to calculate the statistical indicators, including mean luminance, standard deviation, 10th percentile luminance and 90th percentile luminance, which can describe the shadow area and highlight area respectively, and also calculate the proportion of dark pixels below a certain luminance threshold. In the process of calculation, the overall brightness, the contrast between the two tones, and the proportion of light and dark areas in the picture are taken as the main indicators.

2. 2. 2 Chroma and Saturation Measures

The color saturation and chroma of the product were evaluated. Harris et al. (2020) used the channel range method to calculate the chroma.

$$C = max(R, G, B) - min(R, G, B) \quad (2)$$

The chromaticity is calculated to measure the color contrast. The threshold method is used to calculate the proportion of pixels with high and low chroma. At the same time, the images are converted into HSV color space, and the color purity is measured by saturation (Smith, 1978):

$$S = \frac{C}{V},\ where\ C = V - min(R, G, B)\ and\ V = max(R, G, B) \quad (3)$$

Mean saturation was retained as a complementary indicator of overall color intensity.

2. 2. 3 Hue Distribution Analysis

HSV space can divide hue into 12 categories, which can be divided into equal angular bins. The proportion of pixels in each category is calculated for each work, and the normalized hue variable is obtained (such as blue, orange, red, etc.). This method can reduce the complexity of pixels, make up for the shortcomings of the previous method, and can compare the differences between different colors (Itten, 1970).

2. 2. 4 Stylistic Color Deployment Metrics

In addition to color, style deployment also needs to pay attention to how to organize colors in the whole composition. When classifying pixels, the saturation and brightness are used as the standard, which can be divided into several categories: "low saturation and high brightness", "low chroma", "high saturation" and so on. The brightness and saturation are global indicators. The use of these indicators can describe art history, and can make a quantitative description of "neutral", "pastel" and other terms.

2. 3 Statistical Analysis

2. 3. 1 Temporal Trend Analysis

Each decade is a statistical unit, and all color indicators are aggregated. Trend charts are used to grasp the changes of brightness, intensity and hue in the direction of time from the 13th to the 20th century.

2. 3. 2 Correlation and Distributional Analysis

In order to analyze the relationship between the color dimension and the painting, the Pearson coefficient is introduced for analysis (Pearson, 1895):

$$r = \frac{\Sigma(x_i - \underline{x})(y_i - \underline{y})}{\sqrt{\Sigma(x_i - \underline{x})^2 \Sigma(y_i - \underline{y})^2}} \quad (4)$$

The analysis focuses on theoretically relevant variable pairs, including brightness and color intensity, saturation and chroma, as well as tonal structure and hue distribution. These relationships help identify how different visual properties interact within paintings rather than functioning independently. By examining these associations across a large dataset, the study captures consistent structural patterns in color usage.

This method does not analyze a single work, but describes the development pattern of a large number of samples. By integrating the advantages of the above methods, the study systematically describes the changes in color in the process of European painting development.

3. Results

3. 1 Data Processing Result

After the Met Open Access collection is filtered, the image can be used for a period of time, and the data set finally includes 1781 works. For each painting, the color characteristics of pixels are extracted, and the luminance, saturation, hue and other indicators are summarized. All the data are calculated from the normalized pixel values, which can guarantee the consistency of the resolution of each image. The data are aggregated by decade to support the analysis of long-term trends and to eliminate the impact of noise caused by a single work.

Figure 1: Color metrics output (Jacopo Amigoni, Flora and Zephyr, 1730-1739, The Metropolitan Museum of Art, 2025)

3. 2 Color Usage Trends Across Decades

The color usage of European paintings is analyzed from the long-term perspective, and the color metrics are analyzed in depth at the level of ten years. Drawing on hue, brightness and other indicators, the changes of color intensity, color family and so on in the period of 1230-1950 are analyzed. Through the analysis of the time trend line, the color practice of painters in the process of a thousand years is analyzed, and the empirical baseline is provided for the interpretation of style transformation, transition and so on. In the text, the most prominent inflection points and long-term development are analyzed, and the statistical summaries of each chart are listed in Table 2 for reference. In **Appendix B**, there are a series of visualizations corresponding to the results, including all the trend line charts in **Section 3. 2. Appendix B and Table 2** are the basis for the interpretation of the trend analysis.

A note on sampling

The number of works in the Met dataset is relatively small, and the number of decades is relatively small. If a series of metrics show sharp changes at the same time, it is likely that the sample size is too small, but there is no sudden change in style.

Table 2: Color change trend of European paintings (1230-1950)

Variable	Graph Serial	Slope	Net Δ	Peak (value @decade)	Trough (value @decade)	Med	EarlyMod	Modern
Luminance_ Mean	B1	-6. 50E-05	0. 189	0. 448 @1360	0. 195 @1690	0. 351	0. 246	0. 328
Dark_ Pixel_ Share	B2	1. 25E-04	-0. 476	0. 638 @1650	0. 0123 @1950	0. 309	0. 524	0. 352
Luminance_ Std	B3	-9. 94E-05	-0. 092	0. 323 @1360	0. 078 @1950	0. 209	0. 154	0. 167
Luminance_ P10	B4	9. 08E-05	0. 332	0. 331 @1950	0 @1230	0. 085	0. 077	0. 131
Luminance_ P90	B4	-2. 21E-04	0. 048	0. 914 @1360	0. 346 @1690	0. 642	0. 460	0. 552
Chroma_ Mean	B5	-1. 42E-04	-0. 073	0. 332 @1340	0 @1360	0. 186	0. 109	0. 119
Chroma_ Std	B6	-9. 91E-05	-0. 101	0. 190 @1390	0 @1360	0. 125	0. 080	0. 080
Low_ Chroma_ Share	B7	4. 08E-04	0. 268	1 @1360	0. 206 @1340	0. 506	0. 729	0. 690
High_ Chroma_ Share	B7	-2. 69E-04	-0. 075	0. 421 @1340	0 @1360	0. 149	0. 016	0. 025
Hue_ Red_ Share	B8	-1. 26E-04	-0. 284	0. 514 @1470	0 @1360	0. 333	0. 338	0. 259
Hue_ Orange_ Share	B8	1. 72E-04	0. 371	0. 806 @1950	0 @1360	0. 393	0. 447	0. 456
Hue_ Yellow_ Share	B8	9. 80E-05	0. 043	0. 105 @1740	0 @1360	0. 017	0. 059	0. 063
Hue_ YellowGreen_ Share	B8	2. 58E-05	0. 014	0. 048 @1890	0 @1360	0. 006	0. 013	0. 019
Hue_ Green_ Share	B8	2. 04E-05	0. 014	0. 048 @1910	0@1360	0. 006	0. 010	0. 017
Hue_ GreenCyan_ Share	B8	3. 13E-05	0. 030	0. 120 @1910	0@1360	0. 011	0. 018	0. 026
Hue_ Cyan_ Share	B8	-4. 28E-06	-0. 128	0. 163 @1230	0 @1360	0. 036	0. 033	0. 038

Hue_ CyanBlue_ Share	B8	-7. 91E-06	-0. 047	0. 068 @1410	0 @1360	0. 015	0. 007	0. 013
Hue_ Blue_ Share	B8	1. 35E-06	-0. 003	0. 028 @1420	0 @1360	0. 003	0. 002	0. 004
Hue_ BluePurple_ Share	B8	3. 99E-06	-0. 001	0. 030 @1910	0 @1360	0. 001	0. 001	0. 003
Hue_ Purple_ Share	B8	5. 78E-06	-0. 001	0. 037 @1910	0 @1360	0. 001	0. 001	0. 005
Hue_ Magenta_ Share	B8	-1. 53E-06	-0. 009	0. 03425 @1570	0 @1360	0. 009	0. 010	0. 009
High_ Sat_ Share	B9	-3. 42E-04	-0. 460	0. 5933 @1370	0 @1360	0. 363	0. 255	0. 200
Neutral_ Grayish_ Share	B9	-2. 01E-04	-0. 049	1 @1360	0. 051 @1730	0. 278	0. 166	0. 213
Pastel_ Pixel_ Share	B10	-1. 31E-05	4. 91 E-04	0. 3086 @1360	3. 31E-04 @1230	0. 066	0. 029	0. 066
Light_ LowChroma_ Share	B10	-1. 45E-05	0. 004	0. 3468 @1360	4. 52E-04 @1230	0. 072	0. 033	0. 071
Mean_ Saturation	B11	-1. 00E-04	-0. 210	0. 5552 @1340	0 @1360	0. 393	0. 381	0. 334

Table 2 summarizes the main conclusions of the analysis of the development trend over the past ten years, and introduces the structural characteristics of European painting in the period from the 1230s to the 1950s. Luminance measurement can grasp the nonlinear trajectory, with high brightness and strong color variability in the fourteenth century, but in the early modern period, the color gradually darkened, the proportion of dark pixels increased, and the color contrast weakened. From the eighteenth century, the brightness gradually increased, and the highlight was more prominent, but the shadow was not eliminated. This is a reasonable balance between light and dark.

Chroma is an important indicator. The transition from the Middle Ages to the early modern period is a parallel shift. After the 15th century, the chroma is relatively low, and the chroma is weak for several centuries, but it has recovered slightly in the 19th century. The chroma of the early modern period is relatively low, which means that the color intensity is redistributed in the structural level, not that the color intensity disappears. Hue distribution has a strong stability, although the proportion of cold colors has increased slightly, orange and red have

always been the dominant colors, and the proportion of neutral gray regions has remained unchanged in the long-term development, while the saturation has declined and the pastel strategy has been adopted. In general, the development of European painting has not undergone a sudden change, but the color intensity and brightness have changed in the stable chromatic framework.

In order to make it more intuitive, this paper draws a histogram to describe the change of hue distribution in the last ten years, and divides it into twelve categories (see Appendix B for the trend line).

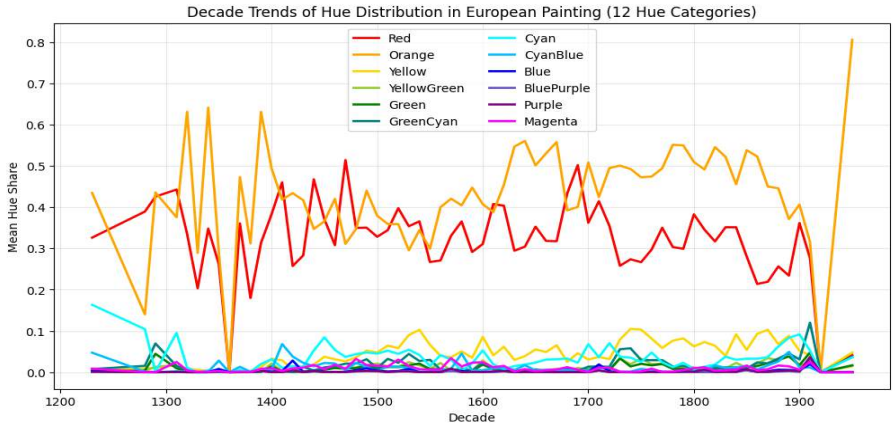


Figure 2: Decade Trends of Hue Distribution in European Painting (12 Hue Categories)

Figure 2 shows that the chromatic hierarchy of European paintings has remained stable for a long time. Warm colors occupy a large area in the painting, and the color structure is relatively stable. Orange is the most common color in modern paintings, while red is the most common color in the early period, and the proportion of red gradually decreases with the passage of time. In the 1700s, the proportion of yellow and green increased, which was related to the increase of landscape paintings, but the proportion was not high. Blue, purple and other colors are relatively low, and the use of cool colors is relatively small, which can only play an auxiliary role. Although there are some fluctuations in the decade, the color structure of the whole painting has not changed significantly, and the color of the painting has always been relatively warm, and the change of secondary color has not changed the color structure.

3. 3 Correlation Analysis

Trend line analysis captures how color variables change over time, allowing for the identification of long-term patterns rather than isolated variation. By examining shifts in luminance, chroma, saturation, and hue distribution across decades, the analysis reveals gradual changes in visual strategies. In addition, correlation analysis evaluates how these variables interact, providing insight into whether changes in one dimension are associated with changes in another. Together, these methods offer a complementary understanding of both temporal development and structural relationships in color usage.

Table 3. Summary of Pearson Correlation Analyses Among Color Metrics (Painting-Level Data)

Figure	X (predictor)	Y (outcome)	Pearson r	p-value	n	OLS slope (y~x)	Intercept
Figure C1	Luminance_Mean	Mean_Saturation	-0. 3851	4. 82E-64	1781	-0. 4740	0. 5123
Figure C2	Luminance_Mean	Chroma_Mean	0. 2560	4. 85E-28	1781	0. 1354	0. 0851
Figure C3	Dark_Pixel_Share	Mean_Saturation	0. 2770	9. 63E-33	1781	0. 1677	0. 3023
Figure C4	Mean_Saturation	Chroma_Mean	0. 6727	5. 36E-235	1781	0. 2890	0. 0138
Figure C5	High_Sat_Share	High_Chroma_Share	0. 4752	5. 10E-101	1781	0. 1684	-0. 0105
Figure C6	Hue_Blue_Share	Luminance_Mean	0. 0583	0. 0139	1779	0. 5737	0. 2799
Figure C7	Hue_Yellow_Share	Mean_Value	0. 0060	0. 7988	1779	0. 0088	0. 3301
Figure C8	Hue_Orange_Share	Mean_Saturation	0. 2940	8. 33E-37	1779	0. 1900	0. 2917
Figure C9	High_Sat_Share	Neutral_Grayish_Share	-0. 4237	1. 69E-78	1781	-0. 4497	0. 2827
Figure C10	Pastel_Pixel_Share	Mean_Saturation	-0. 4620	6. 94E-95	1781	-0. 7056	0. 4098
Figure C11	Low_Chroma_Share	High_Chroma_Share	-0. 5916	1. 33E-168	1781	-0. 2076	0. 1756

Through the analysis of **Table 3** as shown above, it is known that the three dimensions are closely related to each other, and there is a significant relationship between them. From the analysis, the color intensity and brightness of the painting are not strictly unified, and the saturation and brightness are negatively related (r =-0. 385, p <0. 001). It can be seen that the brightness of the painting is relatively high, and the saturation of the color is relatively low. However, there is a positive relationship between luminance and chroma (r=0. 256, p <0. 001), which means that even if the saturation is low, the work with high brightness can still have a strong contrast between the structure and the channel. It can be seen that the brightness and saturation are positively related (r=0. 277, p <0. 001), which means that the composition with low brightness can be more saturated. The above conclusions show that brightness, saturation and chroma are interrelated, but they are different in color practice.

In the process of internal consistency test, the color metric can be analyzed in depth, and its logicality is relatively strong. There is a strong positive relationship between average saturation and average chromaticness (r=0. 673, p<. 001), which proves that the two indicators are related to intensity. Threshold measurement is consistent, the higher the saturation, the higher the chroma (r=0. 475, p<. 001). At the same time, the strong negative relationship is also

reflected in the style opposition. There is an inverse relationship between high saturation and gray share (r=-0. 424, p<. 001), and there is also a negative relationship between pastel share and average saturation (r=-0. 462, p<. 001). The relationship between low chroma share and high chroma share is also negative (r=-0. 592, p<. 001). The above inverse relationship is consistent, which shows that the deployment metric can capture the exclusive style strategy.

In the process of describing the relationship between color and tone, greater caution should be exercised. The proportion of blue is positively related to brightness, but the correlation coefficient is only 0. 058, and the significance level is 0. 014, which is not significant. The proportion of yellow is not significantly related to brightness, with a correlation coefficient of 0. 006 and a significance level of 0. 799, which means that these two colors can not predict the tone. There is a moderate positive relationship between the proportion of orange and saturation, with a correlation coefficient of 0. 294 and a significance level of less than 0. 001, which means that orange can be used as a marker to describe chromaticity, but not brightness. From the whole point of view, the correlation matrix can show the relationship between hue, saturation, brightness and chroma, and can also reflect the relationship between structured and interpretable, especially the relationship between tone darkness and saturation, which is consistent with the data set. The analysis is based on the data of painting level, which only describes the relationship between the two variables, not the causal relationship.

In order to explain the relationship between the two, the relationship between mean saturation and mean chroma is taken as an example, and the relationship between the two is consistent. In the appendix, the author draws a series of charts to describe the relationship between the two.

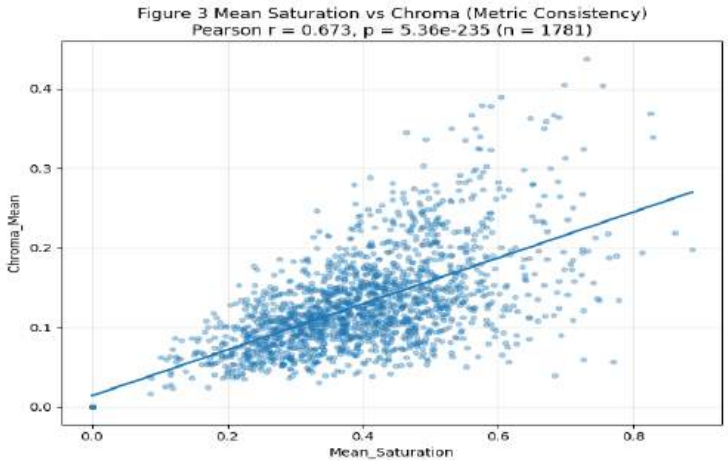


Figure 3 Mean Saturation vs Chroma (Metric Consistency)

Figure 3 shows the relationship between chroma and saturation in all the works in the data set. After analysis, it is found that the two indicators have a strong positive correlation (r = 0. 673, p <. 001, n = 1781), which means that the higher the saturation, the greater the chromatic variation. The regression line is increasing, and the clustering is relatively tight, which is

significantly different from other correlation tests. This conclusion further shows that the two indicators are consistent in all works, and the changes are relatively stable.

However, it should be noted that the relationship between the two is not perfect, which is enough to show that the two are related but there are differences. Saturation describes the purity of the color, which is a concept in HSV space, and chroma is the spread of the numerical channel in RGB space. The strong relationship between the two shows that the measurement framework has good internal consistency, and the two indicators do not have perfect consistency, which shows that they can grasp the color intensity in different ways, and the two indicators are not redundant, but complementary, so the two indicators can be retained in the analytical model, which is enough to show that the color behavior in European painting has a certain structure and is meaningful.

4. Discussion

As mentioned before, the research focuses on how to make up for the shortcomings of the previous method, and analyze the relationship between the color practice and the change of style in the narrative of art history. Qualitative analysis shows that the use of this method can describe the evolution of color in a more reasonable way, and can also analyze the changes of hue, saturation and other contents in the development of European painting.

4. 1 Interpretation of Long-Term Color Trajectories

The results indicate that the color of European paintings has changed in the process of development, and the changes of chromaticity and tone are relatively gentle, not too abrupt. The analysis of luminance pattern shows that the brightness of early works is relatively low, and the brightness of works in the early modern period is relatively low, but the brightness has been balanced in the later period, which shows that the changes of light and shadow in the process of development are relatively balanced, not directional.

From the trend of chroma and saturation, it can be seen that the color intensity has been relatively weak for a long time, and there has been a slight recovery in the later period. The expansion of low chroma areas shows that even if the colors are relatively strong in some periods, the tonal foundation is relatively weak. Generally speaking, the evolution of style is related to the adjustment of emphasis, and the visual system will not change.

4. 2 Structural Relationships Among Color Dimensions

Through correlation analysis, it is known that the use of color in different dimensions can interact with each other in a reasonable way. The negative relationship between saturation and brightness indicates that the two are competitive in the process of visual analysis, and the positive relationship between saturation and chroma shows that the two are consistent.

There is a strong negative correlation between different styles, which is enough to show that artists are willing to use dominant strategies in the process of composition, and do not pursue extremes. Therefore, it can be seen that the change of style is not accidental, but the result

of the balance between many visual properties, and the color is no longer an isolated feature, but a variable that is interdependent.

4. 3 Methodological Implications

The above research results show that the research methods of art history have been extended to computational analysis. In the study, the qualitative concept is transformed into a measurable variable, which can analyze the data comprehensively and compare the data for a long time. This is different from the traditional analysis method, which focuses on the analysis of a single work, and can accurately grasp the macro pattern.

The findings further proves that the interpretation should be carried out carefully. In the analysis, digital images are used as tools, and color values are measured, which can reflect the artistic decisions and reproduction content of the works. Although the research results are similar, it is necessary to use the computational method to build a reliable framework to grasp the visual culture pattern, and to interpret it in the traditional way.

5. Conclusion

In order to explore how the color in European painting has changed from the 1230s to the 1950s, the computational analysis method is used in the research. Through the analysis of more than 1,700 works, the changes in luminance, saturation, and hue distribution can be understood, allowing for a more grounded evaluation of claims about the development of art history. In this way, it will not be necessary to rely on traditional qualitative analysis, but to use computational methods to make a reasonable analysis of the changes in style.

5. 1 Implications of Findings

This study shows that the color of European paintings evolved over time, with changes occurring gradually rather than abruptly. The conclusions drawn in this study are consistent with those of previous scholars, and it is further pointed out that the change of color is not to replace, but to balance. For example, the chromaticity of the works in the fifteenth century has a significant decline, which is consistent with the description of the color palette of the Renaissance period, and the intensity has increased significantly, which is consistent with the description of the color as an expressive force in the nineteenth and twentieth centuries (Baxandall, 1988; Gage, 1993; Ball, 2001).

From the perspective of brightness, the relationship between light and shadow in European paintings can be better understood. The proportion of dark pixels has continued to increase since the sixteenth century, which is enough to reflect the dramatic changes in illumination and the strong contrast in Baroque art, and the gradual brightening of the color tone reflects the changes in the way of light in the next period. However, this does not mean that the style has changed, but that the color tone has changed because the light and shadow have changed. Art historians should not rely too much on the contrast between light and shadow to

divide the periods, because the color tone structure has a strong continuity, and the transition between periods should be regarded as a stable system.

The stability of hue distribution is also a strong support for the above interpretation. Although the style has changed, the warm color has always been dominant, and the chromatic framework has not changed. The above conclusions are conducive to the analysis of color symbolism and iconography, which requires scholars to not only analyze the color meaning from a qualitative perspective, but also pay attention to the quantitative analysis of color, so as to find the differences between the dominant color and the color of individual works.

An analysis of chromaticity, saturation, and brightness shows that artists have made deliberate efforts to balance brightness and intensity in the creative process. The strong style opposition is also reflected in the inverse relationship between the neutral region and the high saturation, which requires painters to find a reasonable composition strategy and avoid extremes. For digital humanists, they should not regard color features as independent variables, but should grasp the relationship between variables in the process of constructing computational models, so as to grasp the characteristics of style from multiple dimensions.

This research further proves that the introduction of computational analysis into the field of art history is a methodological innovation, which can play an important role in the analysis of large data sets. For collection analysts and museum curators, the conclusions of this study are also enlightening, which requires that when constructing digital cataloging systems, the value of large data sets should be brought into play, not only to improve the accuracy of classification and retrieval, but also to grasp the stylistic characteristics of the collection. The conclusion further points out that the integration of qualitative interpretation and computational methods can enhance the scalability and reproducibility of analysis, and can accurately grasp the changes in art.

5. 2 Limitations

In the process of interpreting the research results, attention should be paid to the following issues: first, the data set is obtained from the open access collection of the Metropolitan Museum of Art, although the data are relatively rich, but there are obvious historical and institutional prejudices, which will have a negative impact on the data set. The data set has obvious overrepresentation, which will have a negative impact on the analysis of long-term trends. Therefore, it is necessary to deepen the understanding of the collection in the process of analysis and research, and not to generalize the conclusions to all European paintings.

Secondly, the analysis of this issue is based on the reproduction of digital works, not the original works. The digital image is inevitably affected by the camera, color calibration and other factors, which will lead to changes in pixel values. Therefore, the color index extracted is inevitably related to the reproduction of digital technology. Although the analysis of a large number of works can make up for the shortcomings of a single work, it is difficult to eliminate the impact of imaging conditions on hue, chroma and so on.

In this study, only color-related variables are taken as research objects, and other variables such as cultural background, composition, subject and so on are not taken into account. Although the relationship between different colors and the impact of time are analyzed, the material, historical and symbolic factors that will have an impact on the decision-making of art are not taken into account. This means that the study can only grasp the color pattern, but can not understand the motivation behind the pattern.

Generally speaking, the use of aggregation can identify long-term trends, but it can not accurately grasp the short-term changes, and can not grasp the changes of style and technology in a short time. Some movements, although only in a short period of time, but also to reflect the color of the unique practice, but in the aggregation of data, these changes are not obvious. This is a limitation, but also to make people realize that the research results can not accurately describe the changes of individual art.

5. 3 Future Directions

On this basis, further analysis of both depth and breadth is needed, along with well-founded suggestions for the next stage of research. First, data should be collected from a wide range of museums and artists to improve representativeness, reduce collection bias, and enable a more reliable analysis of color use in European paintings.

In the next stage of research, analysis should not only be conducted using more refined time intervals, but also grouped into groups according to the characteristics of the art movement, so as to understand the evolution of the style in a short time, and understand how the various art movements use color, which is conducive to the establishment of a reasonable classification system, and can also compare with the existing art framework.

Another direction is to introduce more variables into the analysis of color, including context, vision and so on. In the process of analysis, attention should be given not only to composition but also to subject matter, in order to understand how color is coordinated with other elements in artistic practice. For example, the use of color in landscape, portrait and religious paintings is analyzed, and the shortcomings of color measurement are made up.

In the next stage of development, more advanced methods should be employed to improve the accuracy of computational analysis in art history. This includes the use of computer vision, image standardization, and color calibration techniques to address the limitations of digital reproduction and enable more precise measurement of visual properties. In addition, the research results show that the traditional art historical research methods should be combined with the computational methods, so as to improve the research efficiency and ensure the systematicness and accuracy of the research.

Works Cited

Ball, Philip. Bright Earth: Art and the Invention of Color. University of Chicago Press, 2001.

Baxandall, Michael. Painting and Experience in Fifteenth-Century Italy: A Primer in the Social History of Pictorial Style. 2nd ed., Oxford University Press, 1988.

Brey, Philip. “Digital Art History.” History Compass, vol. 19, no. 7, 2021, e12678.

Clark, Alex. Pillow (PIL Fork) Documentation. Python Software Foundation, 2015.

Drucker, Johanna. Graphesis: Visual Forms of Knowledge Production. Harvard University Press, 2014.

Elgammal, Ahmed, et al. “Quantifying Creativity in Art Networks.” Proceedings of the National Academy of Sciences, vol. 115, no. 38, 2018.

Encyclopaedia Britannica. “Baroque Art and Architecture.” Encyclopaedia Britannica, 2026.

Fairchild, Mark D. Color Appearance Models. 3rd ed., Wiley, 2013.

Federal Agencies Digital Guidelines Initiative (FADGI). Technical Guidelines for Digitizing Cultural Heritage Materials: Still Image. 2023.

Gage, John. Color and Meaning: Art, Science, and Symbolism. University of California Press, 1993.

Harris, Charles R., et al. “Array Programming with NumPy.” Nature, vol. 585, 2020, pp. 357–362.

Itten, Johannes. The Elements of Color. Van Nostrand Reinhold, 1970.

Kandinsky, Wassily. Concerning the Spiritual in Art. 1911.

Pearson, Karl. “Notes on Regression and Inheritance in the Case of Two Parents.” Proceedings Of the Royal Society of London, vol. 58, 1895.

Smith, Alvy Ray. “Color Gamut Transform Pairs.” ACM SIGGRAPH Computer Graphics, vol. 12, no. 3, 1978.

The Metropolitan Museum of Art. “The Met Collection API.” 2025.

The Metropolitan Museum of Art. “The Met’s Open Access Initiative.” 2017.

van der Walt, Stefan, et al. “scikit-image: Image Processing in Python.” PeerJ, vol. 2, 2014.

Zorich, Diane M. Transitioning to a Digital World: Art History, Its Research Centers, and Digital Scholarship. Samuel H. Kress Foundation, 2012.

Appendix A: Data Description Table

Table A1. List of Variables and Associated Definitions and Descriptive Statistics for the European Painting Dataset

Variable	Type	Definition	Descriptive Statistics
Object ID	Integer	Unique identifier assigned by the Metropolitan Museum of Art	Mean=449,346 SD=63,224 Range=435,573–888,924
Artist Nationality	Categorical	Nationality or cultural affiliation of the artist	N=1781
Year_Avg	Integer	Estimated year of production (average of begin and end dates)	Mean=1701.23 SD=151.04 Range=1234–1950
Decade	Integer	Decade derived from Year_Avg	Mean=1697.16 SD=150.87 Range=1230–1950
Luminance_Mean	Continuous	Average luminance of all pixels	Mean=0.282 SD=0.124 Range=0.053–0.821
Luminance_Std	Continuous	Variability of luminance values	Mean=0.164 SD=0.054 Range=0.042–0.369
Luminance_P10	Continuous	10th percentile luminance (dark regions)	Mean=0.092 SD=0.077 Range=0.000–0.562
Luminance_P90	Continuous	90th percentile luminance (bright regions)	Mean=0.504 SD=0.179 Range=0.093–1.000
Dark_Pixel_Share	Continuous	Proportion of dark pixels	Mean=0.456 SD=0.252 Range=0.002–0.988
Chroma_Mean	Continuous	Average chroma (color intensity)	Mean=0.123 SD=0.065 Range=0.000–0.438
Chroma_Std	Continuous	Variability in chroma values	Mean=0.088 SD=0.040 Range=0.000–0.257
Low_Chroma_Share	Continuous	Proportion of low-chroma pixels	Mean=0.687 SD=0.216 Range=0.031–1.000
High_Chroma_Share	Continuous	Proportion of high-chroma pixels	Mean=0.033 SD=0.076 Range=0.000–0.658
Hue_Red_Share	Continuous	Proportion of pixels classified as red hues	Mean=0.313 SD=0.227 Range=0.000–0.984

Hue_Orange_Share	Continuous	Proportion of pixels classified as orange hues	Mean=0.459 SD=0.236 Range=0.000–0.979
Hue_Yellow_Share	Continuous	Proportion of pixels classified as yellow hues	Mean=0.064 SD=0.089 Range=0.000–0.726
Hue_YellowGreen_Share	Continuous	Proportion of pixels classified as yellow-green hues	Mean=0.016 SD=0.035 Range=0.000–0.504
Hue_Green_Share	Continuous	Proportion of pixels classified as green hues	Mean=0.013 SD=0.030 Range=0.000–0.301
Hue_GreenCyan_Share	Continuous	Proportion of pixels classified as green-cyan hues	Mean=0.019 SD=0.050 Range=0.000–0.609
Hue_Cyan_Share	Continuous	Proportion of pixels classified as cyan hues	Mean=0.036 SD=0.081 Range=0.000–0.633
Hue_CyanBlue_Share	Continuous	Proportion of pixels classified as cyan-blue hues	Mean=0.012 SD=0.043 Range=0.000–0.830
Hue_Blue_Share	Continuous	Proportion of pixels classified as blue hues	Mean=0.003 SD=0.013 Range=0.000–0.232
Hue_BluePurple_Share	Continuous	Proportion of pixels classified as blue-purple hues	Mean=0.001 SD=0.008 Range=0.000–0.186
Hue_Purple_Share	Continuous	Proportion of pixels classified as purple hues	Mean=0.002 SD=0.009 Range=0.000–0.198
Hue_Magenta_Share	Continuous	Proportion of pixels classified as magenta hues	Mean=0.010 SD=0.031 Range=0.000–0.448
Colored_Pixel_Share	Continuous	Proportion of pixels with chromatic content	Mean=0.751 SD=0.236 Range=0.000–1.000
Grayish_Pixel_Share	Continuous	Proportion of neutral or grayish pixels	Mean=0.249 SD=0.236 Range=0.000–1.000
Pastel_Pixel_Share	Continuous	Proportion of pastel-toned pixels	Mean=0.044 SD=0.100 Range=0.000–0.745
Light_LowChroma_Share	Continuous	Proportion of light, weakly saturated pixels	Mean=0.048 SD=0.106 Range=0.000–0.746

High_Sat_Share	Continuous	Proportion of highly saturated pixels	Mean=0.258 SD=0.214 Range=0.000–0.988
Neutral_Grayish_Share	Continuous	Proportion of low-saturation, mid-brightness pixels	Mean=0.166 SD=0.227 Range=0.000–1.000
Mean_Saturation	Continuous	Average saturation across pixels	Mean=0.379 SD=0.152 Range=0.000–0.886
Mean_Value	Continuous	Average brightness (value) across pixels	Mean=0.331 SD=0.131 Range=0.068–0.833

Appendix B: Trendlines Across Decades

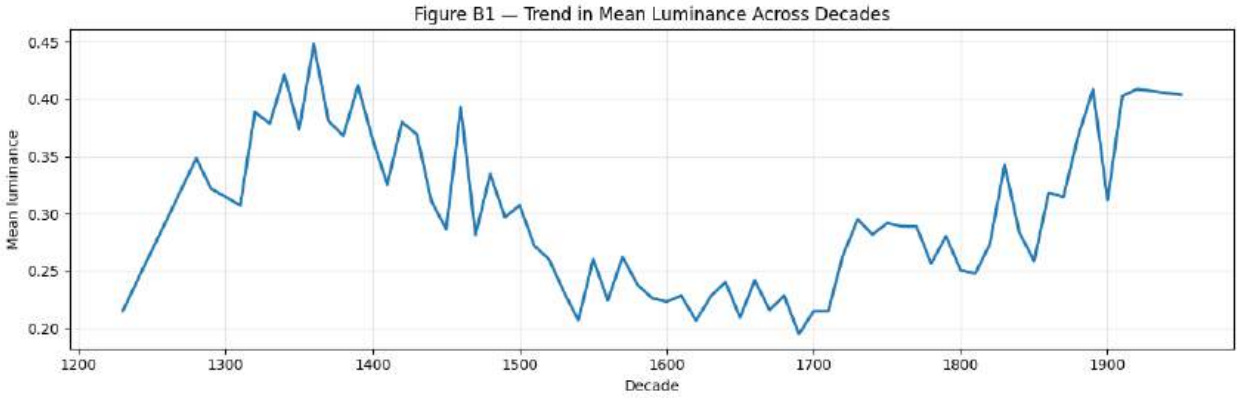


Figure B1: Trend in Mean Luminance Across Decades

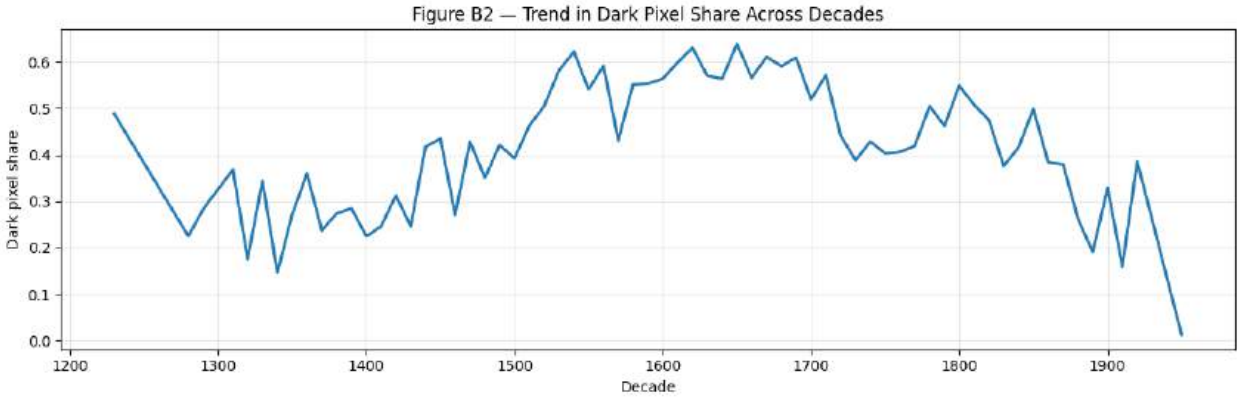


Figure B2: Trend in Dark Pixel Share Across Decades

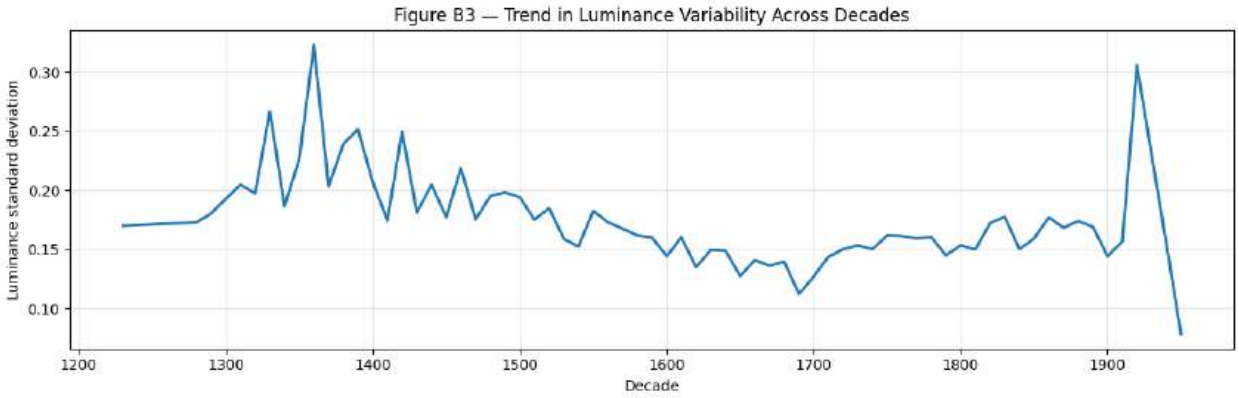


Figure B3: Trend in Luminance Variability Across Decades

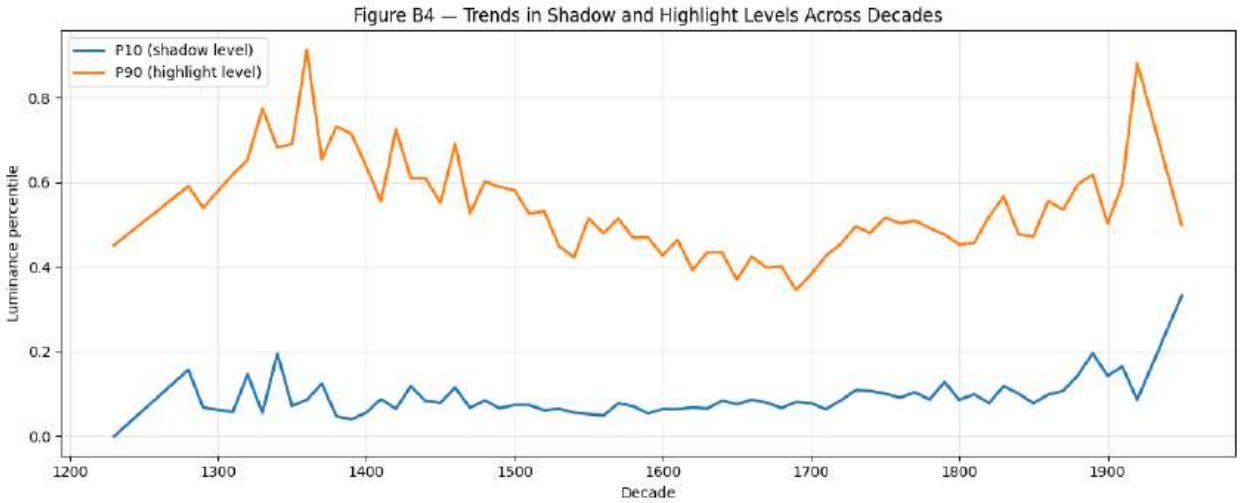


Figure B4: Trends in Shadow and Highlight Levels Across Decades

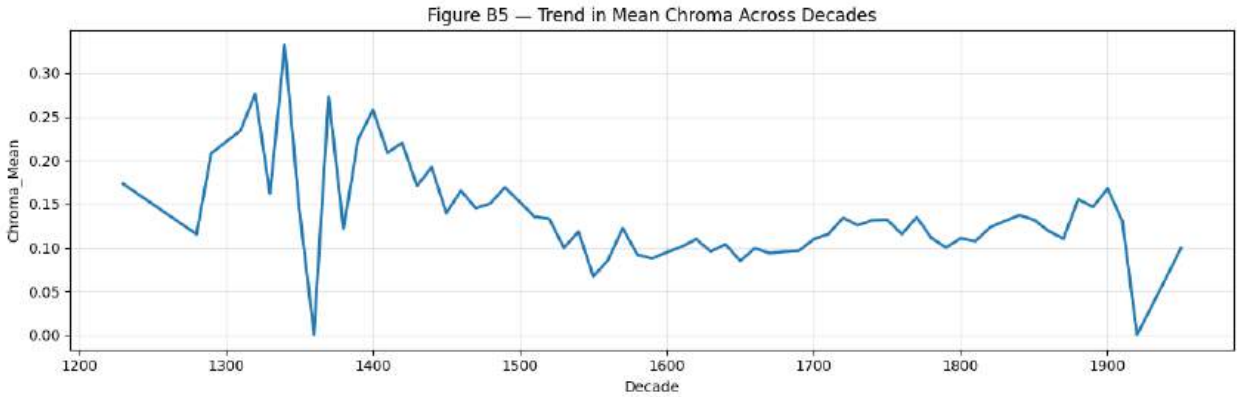


Figure B5: Trend in Mean Chroma Across Decades

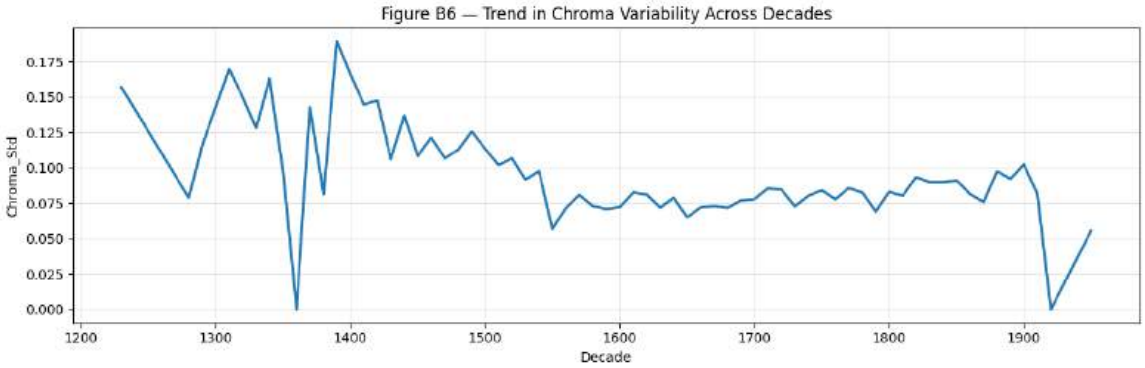


Figure B6: Trend in Chroma Variability Across Decades

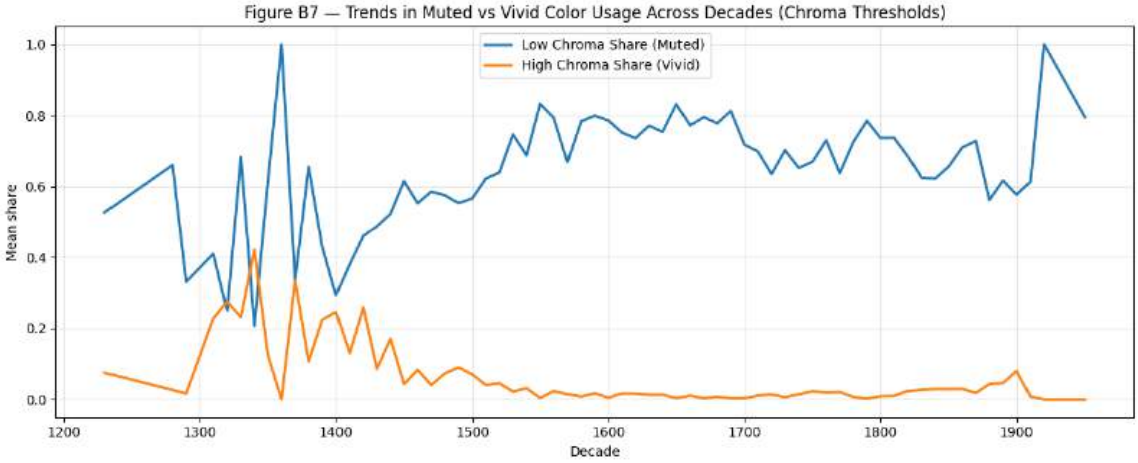


Figure B7: Trends in Muted vs. Vivid Color Usage Across Decades (Chroma Thresholds)

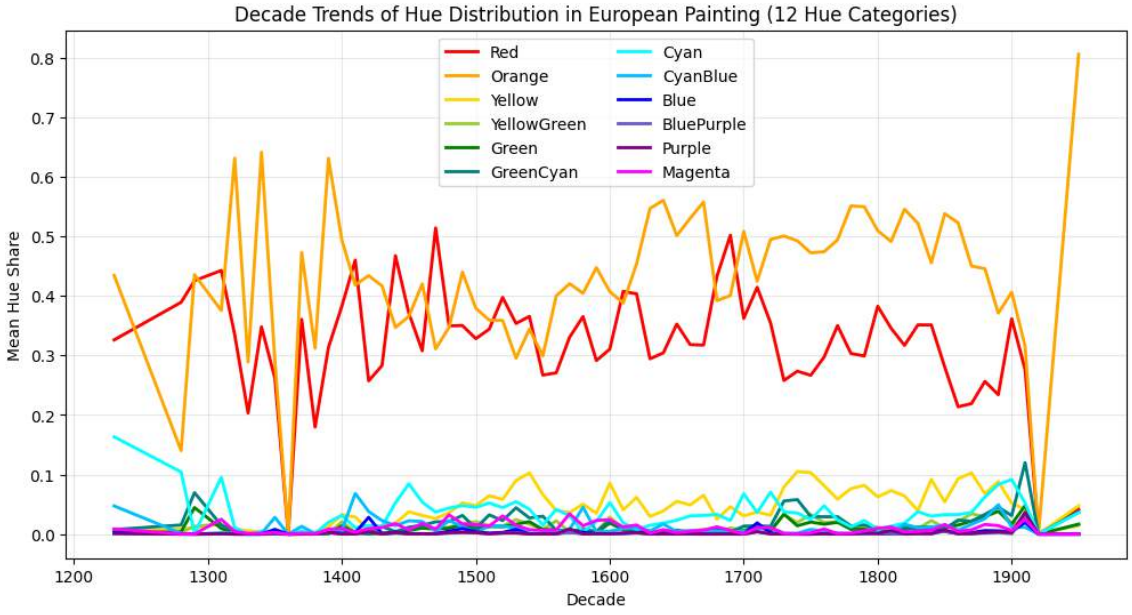


Figure B8: Decade Trends of Hue Distribution in European Painting (12 Hue Categories)

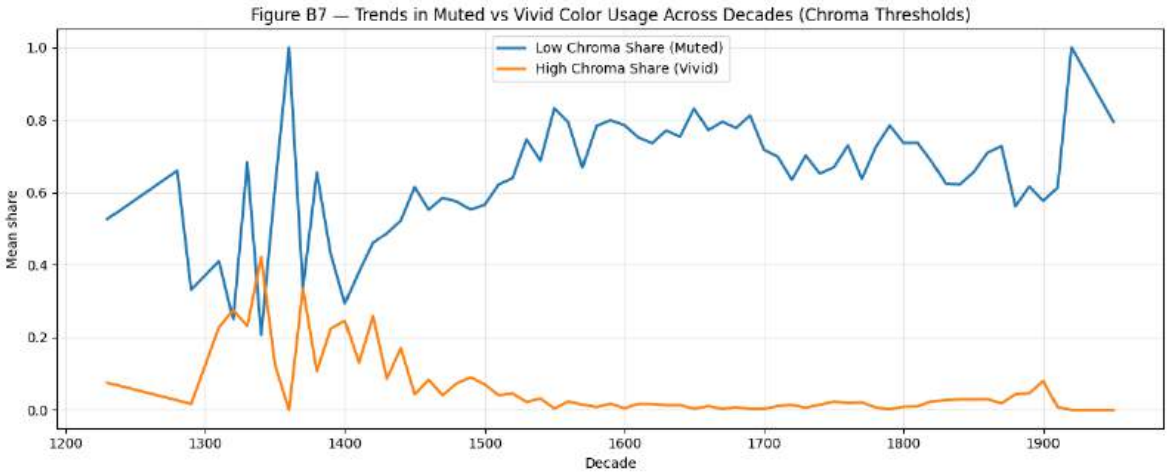


Figure B9: Trend in Bold vs Neutral Color Deployment Across Decades

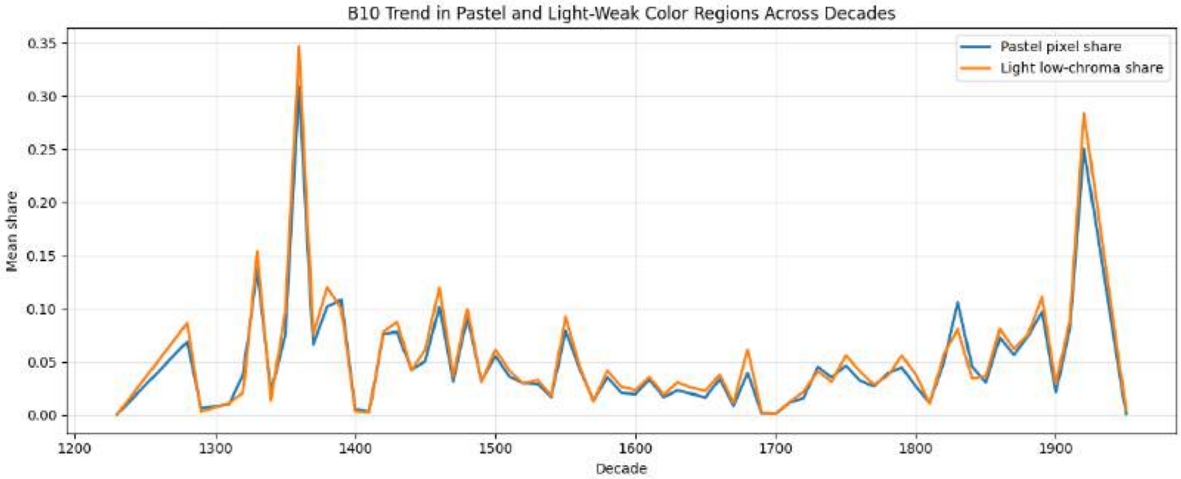


Figure B10: Pastel and Light Low-Chroma Color Usage

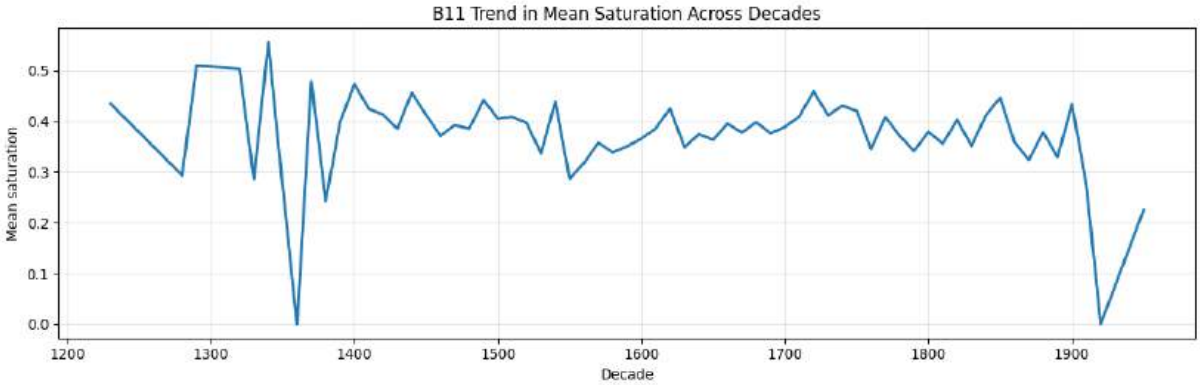


Figure B11: Trend in Mean Saturation Across Decades

Appendix C: Correlation Across Metrics

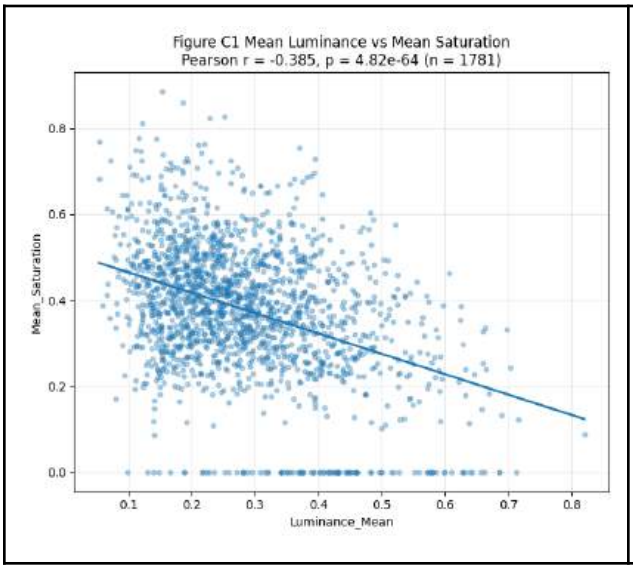


Figure C1: Mean Luminance vs Mean Saturation

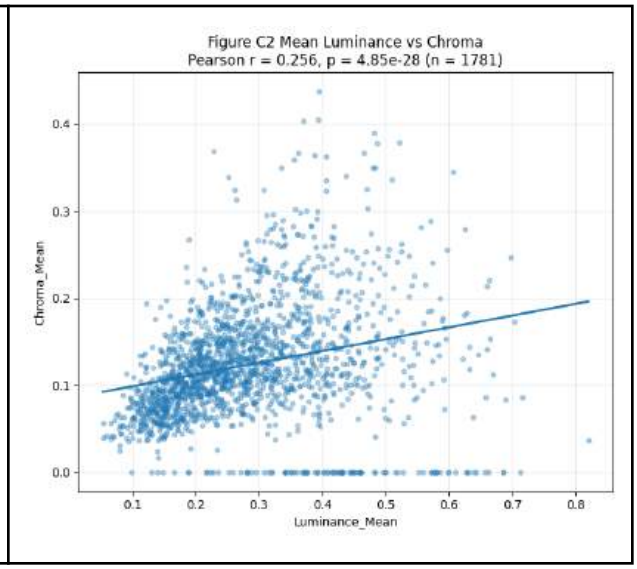


Figure C2: Mean Luminance vs Chroma

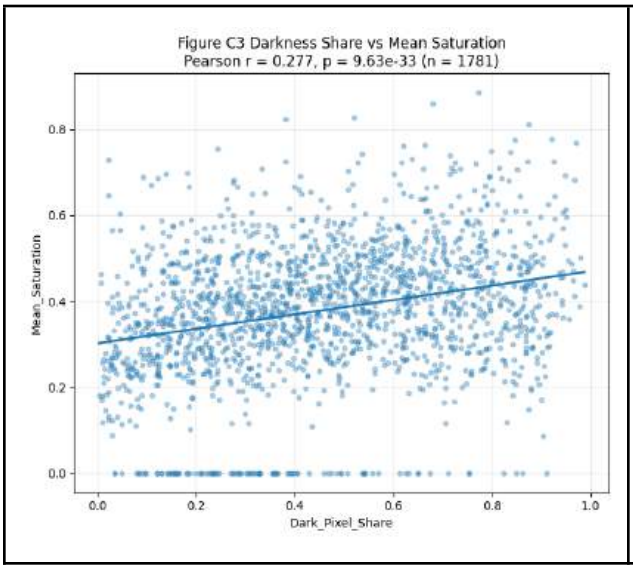


Figure C3: Darkness Share vs Mean Saturation

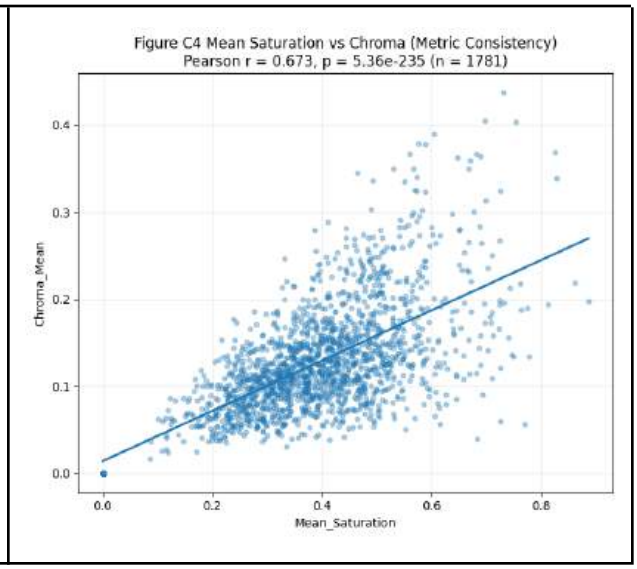


Figure C4: Mean Saturation vs Chroma (Metric Consistency)

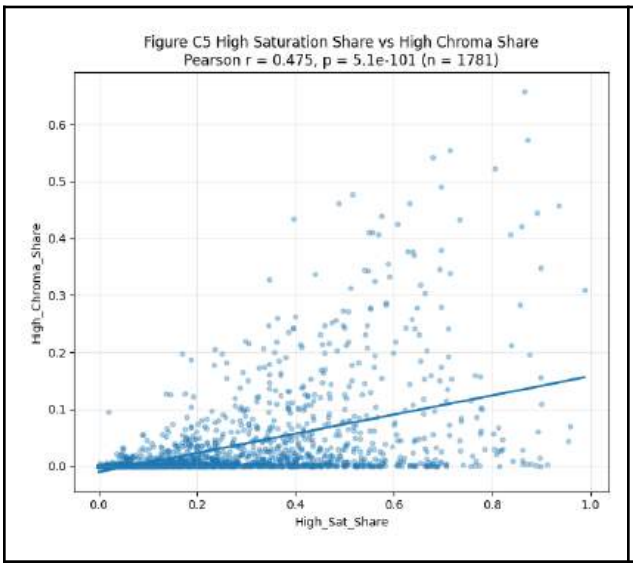


Figure C5: High Saturation Share vs High Chroma Share

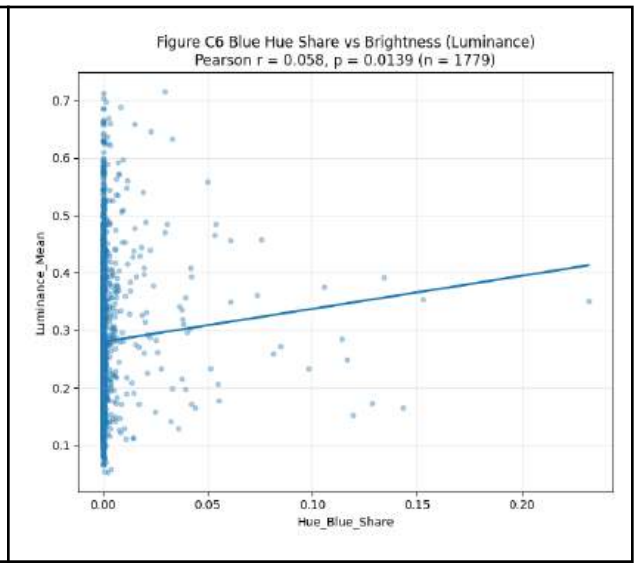


Figure C6: Blue Hue Share vs Brightness (Luminance)

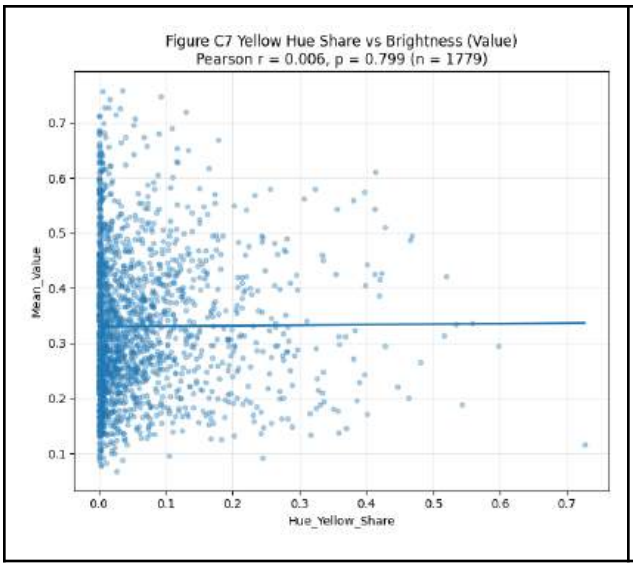


Figure C7: Yellow Hue Share vs Brightness (Value)

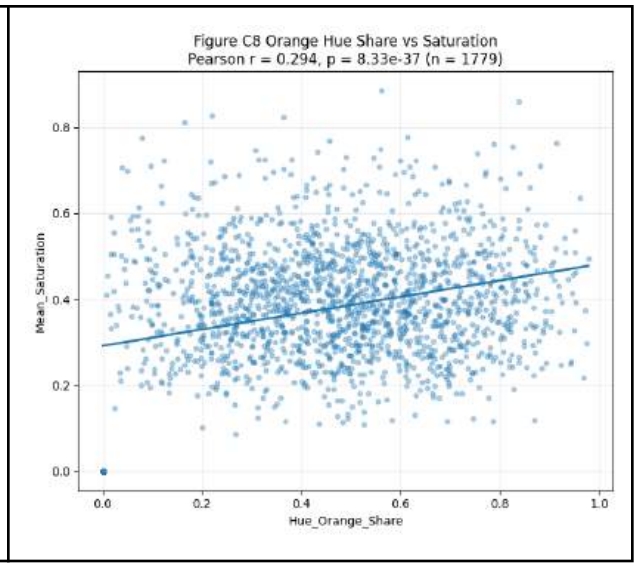


Figure C8: Orange Hue Share vs Saturation

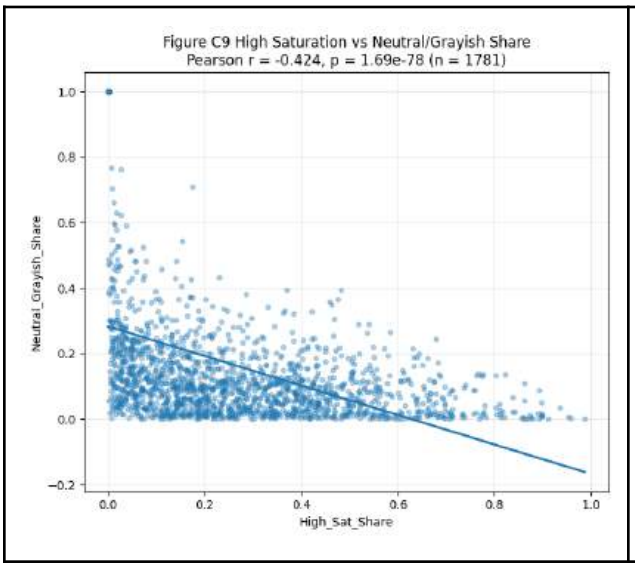


Figure C9: High Saturation vs Neutral/Grayish Share

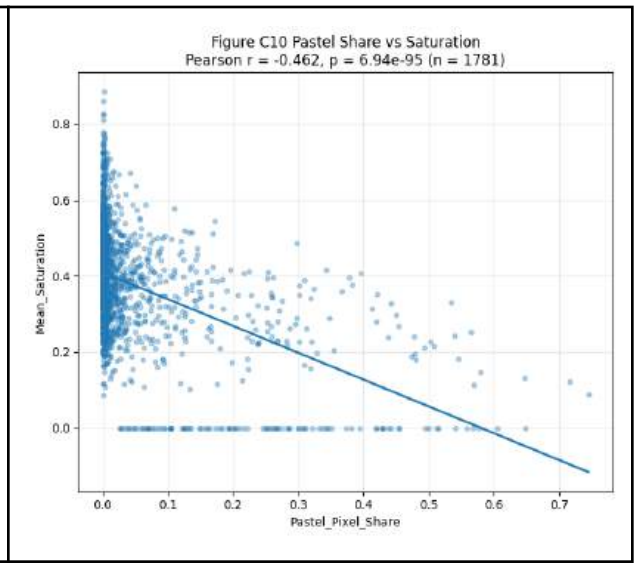


Figure C10: Pastel Share vs Saturation

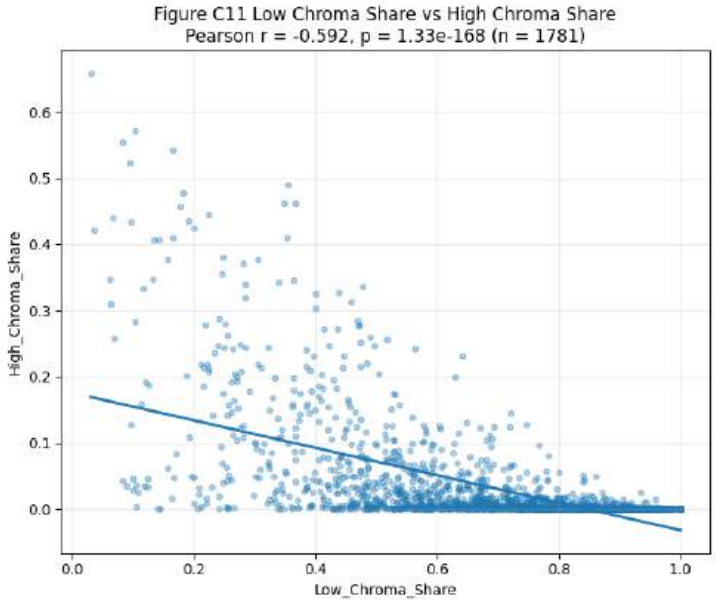


Figure C11: Low Chroma Share vs High Chroma Share

Treating Mental Health with Humanoids By Walter Chang

Abstract

The arrival of humanoids in the medical field creates many possibilities for improving the health of patients. The treatment of mental health using humanoids is particularly new, and previous studies have shown there to be a positive potential use for it. However, there is not enough data to prove whether this positive effect can be generalized. This paper will explain some of these positive outcomes as well as the potential uses of humanoids for mental health. From what has been analyzed, humanoids have the potential to help treat autism and dementia in patients by making the patients more engaged in their treatments. This would improve their communication skills and assist in preventing emotions such as depression and loneliness. These findings could be used to encourage future research on humanoids so that they would be able to help with the lack of healthcare workers in the mental health field. The possibilities humanoids bring would also be used to inform others of the benefits improving humanoids could bring to the medical field. To further explore public opinion on humanoids, a survey was distributed among a population of young adults. The results of the survey showed a slightly negative stance on humanoids being involved with mental health. However, it was not outright rejection of humanoids, with some understanding of the necessity of more support needed in the medical field. Thus, there is some possibility of acceptance for humanoids in the field.

Article Highlights

- Humanoids have high potential to be helpful in the field of mental health to help with the lack of healthcare workers.
- There are barriers and concerns that humanoids need to overcome to be trusted into the medical field.
- Discussed observable trends in humanoids and the future possibilities they bring.

Introduction

For as long as humanity has existed, innovation has always been a part of us. From crafting primitive tools to creating the first computer, technology has improved numerous times. Alongside technology, healthcare has also been evolving. Sicknesses and diseases have always plagued our history, yet humans overcome them again and again. It is no exaggeration to say that technology and healthcare complement each other. While innovation assists with creating cures, those cures help extend the lives of humans to further advance technology. Even now, as the world continues to turn to a new age of technology, so does the healthcare system. Now that humanoids, considered the pinnacle of human creation thus far, are being brought to the healthcare system to assist with the issues that the field faces, there is remarkable potential for humanoids to assist with the lack of healthcare workers currently.

Humanoids, automated robots powered by programming and AI, can perform many different functions. Some are capable of speech, while others can move over complex terrain. As of now, Socially Assistive Robots (SARs) can be classified as a companion or service-type robot (Guemghar, 2022). Companion robots offer psychological support to the patient, while service-type robots provide functional assistance to patients. In our current world, humanoids are becoming more common and are being incorporated into many different fields. One of our greatest healthcare system issues currently is the scarcity in physicians and healthcare workers in general, so it is a necessity to find ways to support the healthcare field as much as possible (AMA, 2023). The mental health department is in need of support in particular, since mental health is relatively new and has less attention compared to other illnesses. The curiosity I had about ways humanoids could help mental health led to the formulation of my research question, " How does interaction with humanoid robots affect the emotional and mental state of patients with mental illnesses." The population that I was focused on were patients of all ages, which includes people from the general population. Through conducting a literature review on studies involving humanoids treating patients, an observation of researchers, and analysis of a survey report, I hypothesized that humanoids would bring many benefits to the mental health field. The purpose of my research on this subject is to bring to light the possible benefits that humanoids can bring to the medical field, share insights on recent observed trends about humanoids, and address difficulties and concerns regarding robots.

Literature Review

Studies about effects humanoids have on patients with mental illnesses have been conducted in order to find whether humanoids can assist patients with their illnesses. Because humanoids are relatively new, more experiments and trials need to be done with them before fully incorporating them into the mental health field. While the medical field does need as much assistance as it can obtain, placing flawed robots to help with delicate work is harmful, which is why there are trial experiments and data analysis being done first. The sources that I have selected observe effects that humanoids have in their experiment, helping to broaden the perspective I have on humanoids. The literature review was categorized by theme, with the articles addressing the observed potential of humanoids, worries patients have, as well as flaws in their studies. Since most of these sources come from medical journals or science publishing sites, there is a low amount of bias. Additionally, with the large amount of references in these sources, the amount of bias would be decreased greatly.

Methods: Study Selection Process

Articles and medical journals containing information on the effect humanoids have on hospitals were used in order to find whether humanoids are helpful for mental health found in PubMed's database. To collect quantitative studies, we used search terms "(Social AND (Robots OR humanoid) OR SAR) AND (depression OR anxiety OR autism OR distress OR loneliness OR dementia)" with the studies being filtered based on clinical trials, meta analyses, date, and age group. We excluded specific studies related to specific psychology studies on patients

without the use of humanoids. For literature reviews and studies used for qualitative purposes, the ProQuest database was used with search terms “humanoids AND hospitals AND mental health”.

Potential Benefits Humanoids Bring

Through reviewing many studies, there is a general trend of humanoids being helpful to the mental state of patients. The majority of these studies focused on elderly patients, since the elderly received the least amount of care but were also able to give consent to getting treatment (Bevilacqua, 2023). One such study done by Lavin observes the effect humanoid interventions have on loneliness in older long-term care residents. The participants’ states of mind were assessed for loneliness, stress, and depression for 8 weeks after interacting with Grace the humanoid, with the results being compared to patients treated by humans. From the data recorded using a loneliness scale, the results showed “decreased loneliness, anxiety, depression, and other improved mental health outcomes in older adults" (Lavin, 2022).

Humanoids are capable of evoking interest from patients, and heightening their levels of engagement. This improved cognitive function would help slow the process of dementia or depression from developing. Another experiment done by Kim studied the impact humanoids had on depression instead by having older patients interact with Hyodol, a socially assistive robot (SAR), for six weeks and evaluating depression levels. The overall depression score decreased, evidence that humanoids are able to assist in lowering negative emotions in patients and thus reducing the risk of mental illnesses (Kim, 2023). Many other experiments involving the elderly showed similar results and support the hypothesis that humanoids can bring benefits to the medical field.

Besides the elderly receiving benefits, children with autism have also shown to be positively influenced by humanoids. In a study involving autistic children, humanoids were used to help train the kids to improve their social skills. Autism causes many to be unable to concentrate and lack motivation. For those with autism, some robots, such as A-Lab (Takata, 2023), looked very similar to humans for the purpose of mimicking conversations. These would allow the children to practice communicating and, again, improve social interactions. These robots try to emulate conversations for children with autism to practice with to interest and engage the children. Like the elderly, the children were also curious about the robots and interacted with them longer than they would with a human.

This experiment was successful, as the children’s ability to socially interact with others improved. The majority of the children’s sociability index scores increased after interacting with the humanoid, showing that there is potential use for these humanoids in therapies. This would be able to assist autism therapists and could even be used as a type of therapy treatment for different mental conditions. This is further supported by Alhakbani’s results that focused on measuring the engagement of children with autism with robots. His study involved observing the humanoid NAO interacting with the autistic children and observing emotions the children exhibited as a way of analyzing engagement levels. The results showed that the humanoids did

attract the children's attention and increased engagement and excitement (Alhakbani, 2023). These sources all presented claims that humanoids had a positive impact on a variety of mental illnesses, stating that humanoids will bring many benefits to this field.

Concerns and Flaws

While the majority of studies showed humanoids to be helpful, they still addressed flaws and difficulties in their experiment. Many acknowledge that their designs "have limitations, such as the small sample size," meaning that the results cannot be generalized for the entire population (Alhakbani, 2023). This is a concern that is brought up frequently, as not having definitive proof of humanoids being able to function can place patients in danger. The majority of the information collected was also taken from the elderly, which means that there is an information gap on the effectiveness of humanoid treatment on young adults or adolescents. There is some information on younger children, but due to legal and ethical barriers, only a small amount was collected (Guemghar, 2022). While it is understandable for this data gap to appear, since there is a greater amount of elderly getting sick than young adults, Lima states that "use of humanoids for therapeutic purposes" would be another good way to collect information (Lima, 2021). However, other researchers argue that the reliability risk is too great and that it isn't right to bring a suicidal person in for an experiment.

However, Kouroupa claims that more research needs to be done in order for humanoids to improve further (Kouroupa, 2022). This claim is backed up by all the sources, which all agree that more research and experiments are required to reach a conclusion about humanoids. If allowed to be studied, humanoids could improve and gain the ability to fully assist therapists and patients. On this stance, all the researchers appear to be in agreement with each other.

Patients have concerns regarding humanoids as well. While the majority of participants felt positive towards the humanoid, such as curiosity and trust, some patients felt fear, uneasiness, or frustration. From a systematic review, it was observed that "negative experiences appeared only in children on rare occasions (González-González, 2021)". There are other patients and healthcare workers who also feel unease towards humanoids working in a delicate field, a problem that humanoids have faced since their creation. Another issue mentioned was that patients, especially those with dementia or cognitive disorders, may not understand how to use the humanoids or not be able to understand the humanoid.

Technical difficulties were among the issues as well. With the humanoids requiring staff to attend them, occasionally interrupting conversations, and being heavy, some conversations ended up to be challenging and frustrating (Guemghar, 2022). While expected to improve in the future, it is one more barrier researchers believe humanoids must overcome to be trusted. Many of these sources also include the ethical issues involved with robots in hospitals, since the data from patients can be recorded without consent.

From the information that has been gathered, I have also come to the conclusion that more information needs to be collected. There is a positive trend in humanoids slowing the progress of mental diseases, which means that the research on this topic would be beneficial for

society if a conclusion can be made. Analyses of ways to work past barriers such as fear or cost would also allow for a continuation in the advancement of technology and the medical field.

Methods

Observation

Gaining insight on a different perspective can be useful for understanding the full situation on a topic. While most of the sources available focus on patients and study results, none focus on the researcher's experience during research, planning, and creating solutions. The community chosen to be observed was a humanoid research group I was involved with. Observations of the research group would allow for the understanding of difficulties in working with a humanoid, challenges researchers face with regards to participants, and the large amount of planning behind experiments. This would contribute to answering the question of why there are a small number of studies involving humanoids, with many of these studies only having small participant groups.

I began my observation on February 19th at 12:00pm. in the Levan Center's North Pitch room (Patel, 2024). During information gathering, "Thick Description" was used while attending the meeting. Through it, I was able to observe emotions, frustrations, and topics that the researchers found difficult such as constant variables. I tried to separate myself as much as possible from my observations to minimize bias through the use of storytelling. The "researcher's point of view" and "ethnography" were utilized as well, since I was able to understand the flaws the research group was facing as a fellow researcher. One such difficulty was the experimental groups requiring multiple variables, such as humanoid type and environment, to stay constant. That allowed me to collect information about reasonings. The researchers' conversation was able to be interpreted and recorded, allowing for qualitative information to be analyzed.

Survey

From the studies observed during this research, it was clear that there was a lack of data on young adults and their opinions on humanoids. In order to fill the apparent gap on this subject, it was necessary to conduct a survey on the opinions of young adults regarding humanoids. With the assistance of a survey, a large amount of information could be collected to understand the overall opinion a population had. Most participants were taken from NSU Florida, but some responses came from elsewhere in the United States, helping to decrease some of the bias that may come with concentrated responses in Florida. There were a total of 25 responses, which gave information about different opinions people may have.

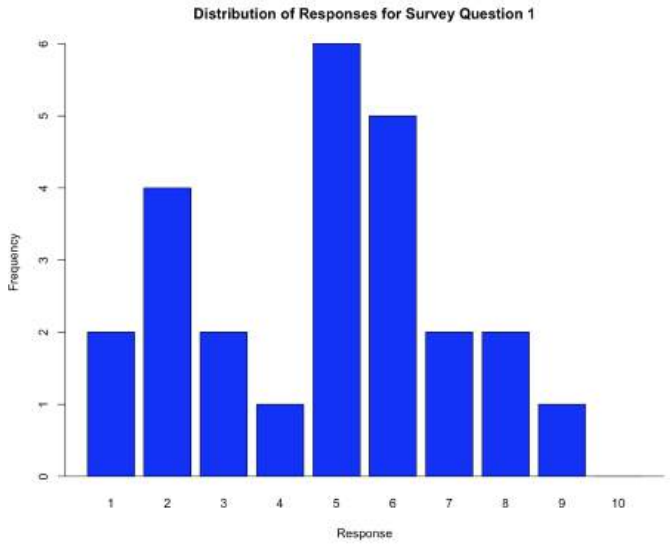
This survey was distributed through acquaintances, friends, and relatives. However, no personal opinions were given to them, meaning this survey is not biased based on individuals since they each have their ideas. The majority of surveyees are college students, but some

participants are slightly younger or older. This would also help spread the age group out slightly. The age gaps are also small, meaning that the data won't be affected too much.

In the surveying process, a total of 10 questions were asked. Within the survey, a variety of question types were asked, such as a rating scale question, open ended questions, multiple choice questions, and follow-up questions. From these questions, information containing emotions felt about humanoids, fears, trust levels, and hope were collected. These results help present the current state of trust towards humanoids and the potential humanoids have to be fully accepted by people into the mental health field.

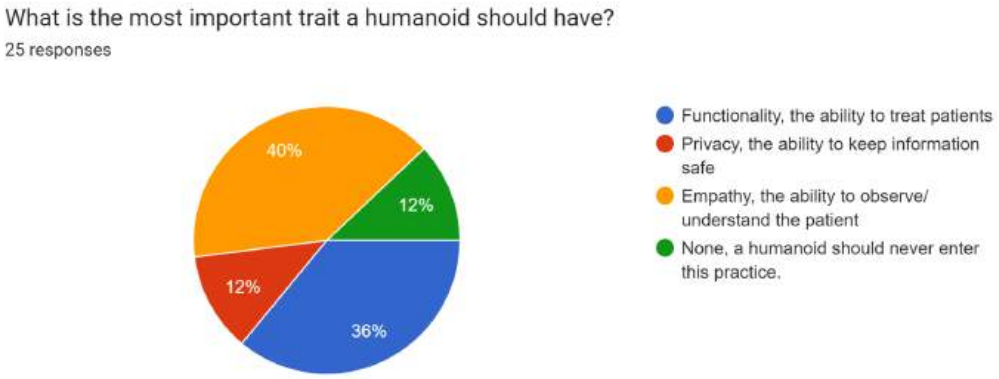
Analysis of Primary Research

Upon receiving the data from the surveys the majority of people felt neutral about humanoids being used in healthcare, However, the average of all the participants leaned towards the negative side, which could be explained by the distrust and caution people feel towards robots being used to cure mental health. From the replies of why they chose their rating, one response states, "I don't feel comfortable with a robot treating me for any illness. People's mental health would be better treated by a human who can give their patients the sympathy and care they may need." The surveyees expressed discomfort when acknowledging robots that appear close to humans, matching Lima's findings of "the majority (of participants) disliked the robot and its capacity to show facial expressions when the faceplate was added." It appears that there is a trend in humans rejecting humanoids if they look too closely to humans. Yet the results of the studies involving autistic children showed otherwise. The facial motions of the humanoids made for better role-playing, improving conversations. Overall, humanoids seem to be more therapeutically helpful when they look human, but will require much more time for the general public to overcome their discomfort. In order for humanoids to reach this point, they must first undergo trials and modifications to evolve to that point.



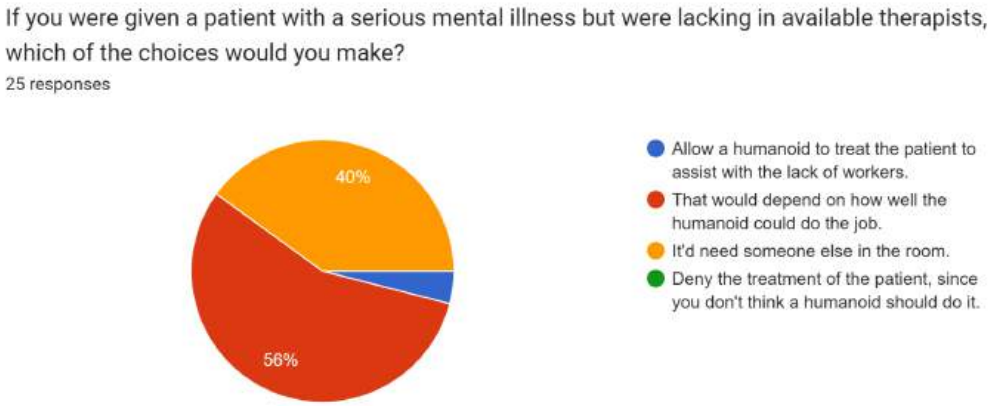
Participants in the survey were asked what emotions they felt, with the majority of participants feeling nervous or scared. Some were interested as well. The results match up with those found in other studies, mainly about the patients feeling scared around the robots as well as some being interested and engaged with the robot (Guemghar, 2022). A trend of conflicting emotions can be seen in relation to the involvement of humanoids in the medical field.

Functionality was another concern that came up in the survey, literature review, and observations. The survey showed that a large number of participants value the functionality of the humanoid, almost tied with the importance of empathy. On a separate question, participants would be concerned with trusting a robot unless its functionality was reliable. From what was researched, this is indeed still an issue that needs to be addressed. As Guemghar had mentioned, technical barriers still existed that prevented humanoids from doing their jobs perfectly, which is supported by my observation of the humanoid research group. The humanoid constantly needed to be modified, repaired, and monitored, inhibiting the project's ability to collect information. Additionally, researchers appear to find recruiting participants challenging. During my observation, there was an argument on whether participants would show up, where to conduct the experiment, and timing of the experiment. These are all factors that must be taken into account when conducting a study. Researchers of other studies found it even more difficult to do so, as ethical and legal reasons prevented them from conducting experiments. One example is humanoid experiments involving children. Many guardians declined allowing their children to join a trial, and privacy issues arose as well. Having a difficult time finding participants might explain why so many of the studies conducted involved only a small participant group. This would also explain the low amount of studies compared to other subjects.



Despite these seemingly negative impressions though, most patients still understand the necessity of the medical field's need for assistance. On a survey question asking about whether to let a humanoid treat a patient or leave the patient untreated, most chose to allow the humanoid to treat the patient. While still cautious of its functionality, no one chose to reject the patient from treatment due to the lack of therapists, since it could result in a worse outcome for the patient.

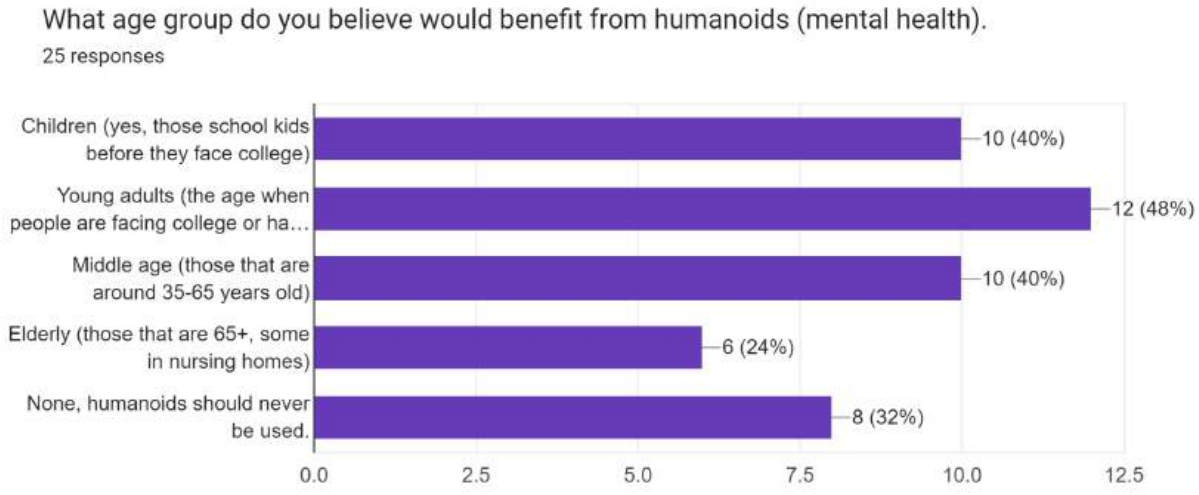
This is a sign that people acknowledge the issue of the lack of healthcare workers, and even though many may be untrusting or uncomfortable with humanoids, they understand the necessity of this support for the sake of the patient. In fact, this is the reason why many of the studies were conducted. Almost all the studies acknowledge the issue of a physician shortage the medical field faces. Lavin talks about how "Loneliness affects up to 42% of long-term care residents and is associated with poor health outcomes," leading to the start of their project to support these neglected elders. To combat the increasing need for healthcare, humanoids can help relieve some of the burden on healthcare workers. This would not only prevent burnout, but would also help heal more patients too.



The future the majority of the survey participants hope to see in the future are humanoids working together with humans rather than replacing them, which is definitely an important concern to have. If humanoids were given too much power or control, it could very well place people in danger, not just those in the medical field. This seems to be agreed upon by researchers, patients, and surveyees alike. Control over humanoids is important to consider, but if properly controlled and improved, dementia and other diseases could be slowed or even cured. After all, the majority of the studies showed a positive trend in the mental wellbeing of patients when interacting with humanoids.

When asked which age group would benefit most from humanoid treatments, most of the survey votes went to young adults with middle age and children next in the order. The elderly were the least voted on. In contrast with these survey results, most studies actually struggle to find young adults or children to conduct a trial on. The population that actually benefited the most were elders, with young children second. Since the majority of participants are young adults, there are some factors that may need to be considered when discussing these results. The follow up question to these results indicate that many of the surveyees believe that a younger group would be more comfortable with humanoids, having grown up with technology and thus viewing the robots as toys or being less nervous around them. Those that voted for elderly stated that it would benefit elders most since " they do not have many people to care for them." These

opinions line up with those of the studies, but the remaining participants seem to believe that the elderly would not benefit from humanoids due to not being oriented to technology. There are many different opinions around this subject, which shows how a public opinion on humanoids still has not been fully formed.



Conclusion

The research question presented previously examined the potential benefits humanoids can bring to the field, as well as address problems that may arise. Through a literature review, an observation of a humanoid research group, and analysis of a survey, it was concluded that humanoids contain high potential for treating patients with mental diseases. From autism to depression to loneliness, a wide variety of mental conditions can be slowed or prevented with the assistance of SARs. However, with SARs being relatively new, it is inevitable for many concerns and flaws to appear, such as functionality issues or distrust. The sources and research results collectively agreed that more testing needs to be done before a consensus can be reached. Luckily, there is still a possibility for people to be accepting of humanoids curing patients, and so more experimentation may be possible to further improve technology. Even during my observation of my group, the humanoid was undergoing continuous improvements and fixes. Eventually, when enough time passes, humanoids may be able to change the medical field for the better and work alongside healthcare workers.

While there isn't enough data yet to answer all questions, such as when humanoids would be able to be fully implemented into hospitals, there is enough evidence to argue that humanoids can benefit the healthcare system and take some of the workload off healthcare workers. Future research on this topic would hopefully include a more abundant availability of studies, especially studies involving young adults, so that a final statement can be made on whether humanoids can officially join the medical field.

The key point is that humanoids should not be completely rejected out of fear, but instead improved so that more lives can be saved. The sooner that efficient humanoids are created, the greater the chance of averting the crisis of the public not receiving medical attention. Thus, experiments must be supported for the continuation of technological progression as well as medical improvement. As humanity has always done, innovation and advancing technology must be continued for a better future.

Declarations

Ethics Standard: CITI Standard

Funding: No Funding

Human Ethics and Consent to Participation: Voluntary Survey with participants informed of the purpose of the study. Participants gave their informed consent.All experiment protocols were approved by the IRB of Nova Southeastern University.The experiment was conducted in accordance with CITI guidelines

Competing Interests: None

Data availability: Form Result Link-
https://docs.google.com/spreadsheets/d/1XdIvQ-TOlL5MqFu4My2U1LlXLCiFaRTmAYYZLGANsfE/edit?usp=sharing

Works Cited

1.Alhakbani, N., Al-Nafjan, A., & Alabdulkareem, A. Measuring Engagement in Robot-Assisted Therapy for Autistic Children. Behavioral Sciences 13, 618 (2023).

2.Bevilacqua, R. et al. Social robotics to support older people with dementia: a study protocol with Paro seal robot in an Italian Alzheimer's day center. Front. Public Health 11, (2023).

3.Kim, S. K., Jang, J.-W., Hwang, Y. S., Lee, O. E. & Jo, H. S. Investigating the effectiveness of Socially Assistive Robots on Depression and Cognitive Functions of Community Dwelling Older Adults with Cognitive Impairments. Assistive Technology 0, 1–9 (2023).

4.Kouroupa, A. et al. The use of social robots with children and young people on the autism spectrum: A systematic review and meta-analysis. PLOS ONE 17, e0269800 (2022).

5.Lavin, P. et al. Humanoid robot intervention vs. treatment as usual for loneliness in long-term care homes: Study protocol for a pilot randomized controlled trial. Front Psychiatry 13, 1003881 (2022).

6.Frontiers | Social skills training using multiple humanoid robots for individuals with autism spectrum conditions. Takata, K., Yoshikawa, Y., Muramatsu, T., Matsumoto, Y., Ishiguro, H., Mimura, M., & Kumazaki, H. (2023, June 23). Social skills training using multiple humanoid robots for individuals with autism spectrum conditions. Frontiers. https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2023.1168837/full

7.AMA. (2023, October 25). AMA president sounds alarm on national physician shortage. American Medical Association. https://www.ama-assn.org/press-center/press-releases/ama-president-sounds-alarm-national-physician-shortage

8.Guemghar, I., De Oliveira Padilha, P. P., Abdel‐Baki, A., Jutras‐Aswad, D., Paquette, J., & Pomey, M. (2022). Social Robot Interventions in Mental Health care and their Outcomes, Barriers, and facilitators: Scoping review. JMIR Mental Health, 9(4), e36094. https://doi.org/10.2196/36094

9.González-González, C. S., Violant‐Holz, V., & Gil, R. (2021). Social Robots in Hospitals: A Systematic Review. Applied Sciences, 11(13), 5976. https://doi.org/10.3390/app11135976

10.Lima, M. R., Wairagkar, M., Natarajan, N., Vaitheswaran, S., & Vaidyanathan, R. (2021). Robotic Telemedicine for Mental Health: A Multimodal Approach to Improve Human-Robot Engagement. Frontiers in robotics and AI, 8, 618866. https://doi.org/10.3389/frobt.2021.618866

11.Chang, W. Y., Kalepalli, Munlapudi, V., A., Patel, S., Ramim, M. (2024, 2, 19). Assessing Human and Social Humanoid Interactions across levels of trust, persuasion, and empathy . Social Humanoid Robot Research, Levan Center.

12.Chang, Walter. (2024). Survey on Humanoids and Mental Health

The Aesthetic Transcendence of Mixed Martial Arts
A probe into the mediums of artistic expression and their underlying motivations
By Joseph Slaga

MMA as an Art Form

What is art? A comprehensive definition of art includes a few key precepts. First, art must have a medium or structure to act as a vehicle for creative expression. Secondly, a mastery or expertise of this vehicle must be present. And finally it must convey beauty or communicate a deep meaning. It is therefore not an issue to discuss art in a way that disregards conventional forms, which work counterproductively only to restrain that which can have the same function for the purpose of tradition. Tradition is important in that it provides structure, a key aspect of art, yet it is not crucial to the art itself. Across my 11 year career spanning both traditional martial arts such as Taekwondo, Muay Thai, and Brazilian JuJitsu, and evolving into mixed martial arts, there has not been a time when structure was ever disregarded. That structure comes by way of sport. Throughout this essay, I will prove to you that one, MMA is a structured sport and discipline, not some sort of spectacle as many claim it to be. Secondly, I will demonstrate that it is a valid medium for the purpose of art. Finally, I will prove that it contains the philosophical requirements necessary to be considered art, and that the fighter may be seen as a personification of art itself.

Can MMA be considered a sport? Yes, it can in fact. First a distinction needs to be made for the reader's benefit over the understanding of MMA. As many laypeople understand, it is some form of close combat. This is simply untrue. Close combat is the unpredictable event in which fighting from an immediate distance with an enemy for the purpose of defeating them or defending oneself is necessary. It generally implies total shock and surprise, stemming from a lack of awareness that the situation was around them. Upon examination, close combat is completely different from MMA, which is a willful confrontation as a result of mutual consent between competitors, within a rule set constraining the participant's behavior. Though it may be marketed as life or death, the reality is that this is a falsity, and little more than sports advertising. Now, to settle the argument as to whether it is a sport as I claim, the utilization of Plato's account of craft, which is a system of designating practices, will be paramount. A craft is an expertise that brings about harmony, order, and knowledge within its practitioner. These aim at the good of their subject. There are goods internal to the craft, and they are able to be taught. Now there are 3 main goods or characteristics which MMA possesses: one, the ability to fight in combat, two, physical improvement, third, the development of virtuous character. Now a craft must also provide an account as how to attain the expertise it aims at, a way to teach. Logically, with the influx of MMA gyms throughout the world there must be a teaching method within the practice, as is described in *Chinese Martial Arts: From Antiquity to the Twenty-First Century*, "what makes something a martial art rather than an action done by someone who is naturally good at fighting is that the techniques are taught. Without the transmission of these skills through teaching, they do not constitute a craft in the sense of being a body of information or techniques

that aim to reproduce certain knowledge or effect."(Lorge, 3-4). With all that has been said, it is only rational to accept the actuality that MMA is a craft as it fills all of the requirements: of harmony and order, as it creates a structured environment within the user and outside of them which to operate in, of knowledge as the learning of techniques imply a knowledge being ingrained upon both one's mind and one's muscles. Finally, it possesses goods integral to the sport which can be passed on. These characteristics strongly support the claim that MMA is a craft.

In regards to the idea that MMA is simply a spectacle to be valued only for its entertainment, there is some footing, especially taking into consideration the state of the UFC's first few events which were exactly that. Yet, in *The Philosophy of Mixed Martial Arts*, this assertion is excellently refuted as follows, "spectacles aim at evoking audience emotion, which is something external to the activity itself. The success or failure of the activity thus depends on the audience's response rather than on the excellence of the activity."(Martínková and Parry, 22). While there is significant evidence that early in the UFC's history it was specifically designed to be a spectacle, in its arranging fights with fantastical "no holds barred" matchups, and a practical elimination of all rules, the sport has evolved since to be rid of many of its former theatrics and the focus has become on the individual skill of the fighters. No longer does it resemble the mockery that is pro wrestling, with the introduction of uniform fight wear and rule sets, the sport much more resembles a sort of duel than an entertaining brawl. Opponents of this view may argue that, despite this change, at its core it has not changed its essence which is putting on something spectacular and entertaining. While taking into consideration the evidence which I provided above, there is still merit to the opposition argument, and in response to this, it is best left as is. There are elements of spectacle in the press conferences and ceremonial weigh-ins, and elements of craft in the individual fighters and their style and training. However, the presence of spectacle by a particular promotion does not constitute the entirety of the sport as is.

The requirement of MMA being established as a sport before moving forward may seem unwarranted in this specific paper, yet on the contrary it is crucial. This is due to three important elements, the first being ability for more technical expression. When fighters are focused primarily not on their own survival, but instead their victory, (to again reference the difference between close combat), the capacity to execute more advanced techniques, strategies, and intricate transitions, is drastically increased, further permitting for more of an expression deriving for the act. The secondary is a universal metric, with an objective system of appreciation through scoring, allowing both appreciation and criticism on an authenticated scale, elevating creative work beyond mere entertainment. Finally, it gives institutional credibility, removing the stigma of "human cockfighting", and honoring the heritage from which it derived, by validating the efficacy of more traditional and standard martial arts.

Now that it has been solidified as sport, I will move to address the relationship between art and sport, and what designates one as a valid form of both. Take the example of ballet, already a generally accepted form of artistic expression. Yet strip away the aesthetics of both and you find virtually the same life in each sport, MMA and this dance. Both suffer the same

repetition and strive for technical perfection, and the same price of injury is present in either, the price of forging identity through hardship. The choreography of ballet, the key aspect of that form is present in MMA as well, the only difference being that one has been planned meticulously, the other landed with the same preparation, yet more improvised and adaptable. Ballet is the form of an aristocratic combat, steeped in posture, ceremony, and concealment, whereas MMA is the dance of the common man, stripped of all ornament, yet equally dependent on rhythm, timing, and muscular movement, united by the flow which one must achieve in order to not succeed, but survive in either. Perfectly encapsulated by a legend known to live in the combative and the rhythmic, Bruce Lee said, "Fighting is the best dance". This exhibits what a world renowned fighter had discovered about the relationship: martial arts is the work of grace and fluid expression, moving gracefully with one's opponent. But it goes further than a simple analogy; it is an equation, the two are seemingly interchangeable at their most basic level. Therefore, the art of fighting may reasonably be understood as containing the same qualities as the art of dance.

However, establishing that something shares artistic elements in regards to the aesthetics of many arts, and therefore labeling it as art would be quite a shallow assumption. True art contains a deeper meaning that has to be conveyed, the final piece to this puzzle. The meaning behind Mixed Martial Arts is the abstraction of self and the rejection of passivity. In Yukio Mishima's *Sun and Steel*, he speaks of the self being so abstracted by pure combat that body and soul, with the intellect, must unify. "For ideas are, in the long run, essentially foreign to human existence; and the body — receptacle of the involuntary muscles, of the internal organs and circulatory system over which it has no control — is foreign to the spirit, so that it is even possible for people to use the body as a metaphor for ideas, both being something quite alien to human existence as such."(Mishima, 38). In fighting, the body becomes de-alienated from that which inhabits it, and becomes fate, facing an inevitable reality of confrontation. Through this, the only protection comes in victory or defeat. In training for this confrontation, (which Mishima wrote of as death, being a duelist), the body ceases to be merely decorative and rather gains the properties of the idea which the mind is to focus on, that will to fight manifests itself beautifully in the physical through musculature. Man transcends himself through this ordeal and becomes an expression of something greater. Julius Evola too realizes this, furthering this ideal making combat something spiritual. He views fighting as a confrontation of the soul against chaos and fear, forcing the individual beyond materialistic comfort and attachment in the dream of chasing something greater, something held by those warrior civilizations of old. He describes a vertical ascent from this mundane plane of existence, with risk providing a way to rise above the passive day to day existence through mastery and discipline. This culminates in a conscious conflict with danger, as seen in his book, *The Metaphysics of War*, "Danger reawakens the spirit"(Evola, 66). Yet one might object that mixed martial arts does not possess the same spirit that was sought after by those belligerent societies of antiquity. This is true, but the remnants still exist through modern octagon fighting. While the two cannot be equated, we can still claim that the same concept resides in both of them, viewing MMA as a derivative of this historical

phenomenon. Through these two authors, the true *telos,* or function of mixed martial arts is revealed. To transcend oneself through unity of body, mind, and soul, to seek out danger above our mundane lives, living with intensity, and to mold one's own identity through the struggle that comes with it. There is also an element of tragedy to it though, for victory is only half the outcome. Mishima knew this only all too well, modeling himself after those Hellenic legends which he had come to revere, whose lives often followed a similar path to that of the fighter, at least in principle. Careers are ephemeral, as practitioners fight against not only their opponents, but against the ever fading and breaking machinations of one's own body. No matter how disciplined or skilled one is, they cannot conquer mortality, only confront it. They know this, yet still, they willfully fight the impossible battle and attain glory in the midst of an all-encompassing defeat, acting while fully conscious of their impermanence. In the Greek milieu, which Mishima being the lifelong Grecophile was intimately acquainted with, the fighter is the tragic hero, the individual achieving greatness through battle with what will eventually destroy them, the very essence of those legends. The fighter, or hero, voluntarily risks themselves in order to gain form in a world full of dissolution, a temporary rebellion, a carving made into sand, or a rose just coming into bloom. That which will fade tomorrow is all the more beautiful today, and this is the underlying philosophy of fighting.

When I speak of transcendence, I don't mean it as being a metaphysical fact in this circumstance. Rather, I intend a more philosophical self transcendence, in which normality disappears, and mind, body, and action become unified. In day to day life, the individual lives in a fractured condition. When entered into the realm of fighting though, a radical reintegration of the faculties is required. There can be no room for hesitation of deliberation. Instead one must act, thought of in the mind, and responded to by the body, temporarily overcoming the division between mind and body. But also, transcendence comes not just from these prerequisites for victory, yet from the conscious acceptance of risks as well. In exposing oneself to danger, the fighter is forced beyond passivity, intensifying self-awareness manifesting an environment in which a new perspective is brought into being.

Now into the equation comes the claim that the purpose of the fight is to win. This would come with the idea that art should exist primarily to convey meaning. Thus, though MMA is aesthetically beautiful, it cannot be artistically beautiful. A fighter makes his walk with the intention, not to portray an idea, but to defeat his opponent who stands before him. While his movements that he goes through in pursuit of this goal, such as a flashy kick or a perfect takedown, may be beautiful, the nature of them is practical, ordered towards success. To continue, as a general rule, art requires an audience capable of interpreting meaning within the subject. Though this is more debated in the modern age with our contemporary forms, a traditional stance would maintain that this is necessary. For instance, when reading a book or viewing a play, the patrons encounter symbols used to communicate something beyond the piece itself. To contrast, a fighter's jab serves a tactical function, meant to inflict damage or advance strategy.

While a valid point, this argument relies too heavily on an overly restrictive definition of art. It assumes that art must intentionally communicate symbolic meaning, rather than embodying it. To liken back to the example of ballet, meaning does not derive from anything other than the movements of the ballerina, making the body a medium. Therefore, the body as a medium is not a perversion of art, it is a more active and direct form. The reason why ballet can resonate so heavily with audiences is due to its emotional layers, portrayed by the executed choreography. This extends to MMA as well. While the fighter seeks victory, this alone does not exhaust self-expression. In fact it is the opposite. Take the technical grace of George St. Pierre, the fluidity of Silva, the pressure of Khabib. All these are modes of self expression to which their bodies and minds are best attuned. It becomes personalized, and within this medium their discipline, emotion, and sum of experiences are viewed. Moreover, to address the critique of functionality, many respected arts stem from simply practical acts. Architecture, for instance, with the goal of providing shelter. But beyond its simple origins, it can be the act of creating an extension of its inhabitants or designers, by way of style and material. Rather, art can arise from a practical medium, and this medium can become an art when it expresses something beyond its immediate utility. In this way the struggle for victory provides the structure for this self expression by forcing fighters to create radically different and highly personalized styles in seeking this goal.

Over the course of this essay, all claims by those who object to MMA joining the muses of art have been refuted. Truly, Mixed Martial Arts is a viable craft, as wonderfully defined by Plato. Again, it is on par with a form such as ballet for being a medium with which art can come to fruition. And finally, MMA itself is an expression of beauty, as seen when we peel back the layers of external appearance and grasp the true nature of an act. To synthesize, MMA is not only a valid art, but one to be lauded, and the fighter as a proper virtuoso of it.

Works Cited

Jason, Holt and Marc Ramsey, editors. The Philosophy of Mixed Martial Arts: Squaring the Octagon. Routledge, 2021

Julius Evola. The Metaphysics of War. Arktos Media, 2024

Peter A. Lorge. The History of Chinese Martial Arts: From Antiquity to the Twenty-First Century. Cambridge University Press, 2012

Yukio Mishima. Sun and Steel. Burnham Press, 1970

Limited Access and Unlimited Exposures: Moving towards More Holistic Solutions for Superfund Sites By Andy Kwon, Eleanor Lim, and Paul Shim

Abstract

Superfund sites are hazardous waste sites that are frequently located in marginalized communities where residents are disproportionately affected by environmental risks. Many of these communities are also designated as food deserts because of their lack of access to healthy foods. To address the multifaceted problems these communities face, holistic solutions must be developed, solutions that holistically consider how environmental, health, and food challenges interconnect to limit individual and community development. This study seeks to understand how those challenges intersect at the local level by studying the case of the NJ superfund site Historic Potteries in East Trenton and by proposing a more holistic solution. Using NJ data from the EPA superfund site as well as demographic data—such as median income, poverty level, and minority population—from the U.S. Census, this research narrowed down potential cases by identifying NJ communities severely affected by environmental contamination, poverty, food shortage, and limited access. Through this process, we selected the Historic Potteries site in Trenton, New Jersey because of its history of lead contamination, legacy of social vulnerability, and status as an overburdened food desert community. This article proposes remediation and a more novel way to reuse the land that directly addresses the multiple challenges that East Trenton experiences: freight hydroponic farming as a holistic remedy that can provide access to fresh and clean crops while avoiding exposure to contaminated soil. Additional recommendations on remediation are provided regarding community awareness and in site remediation.

Introduction

Superfund sites are highly contaminated areas that are remediated and managed by the U.S. Environmental Protection Agency. Many of these sites are located in marginalized, overburdened communities that face multiple challenges beyond environmental health hazards. Therefore, solutions should be holistically designed to address multiple vulnerabilities, instead of creating separate disconnected less efficient solutions.

Superfund Sites

Superfund sites are sites contaminated by hazardous waste that has been dumped, left out in the open, or improperly managed (EPA, “What is Superfund?”). Though many contaminated sites are remediated and managed either privately or by the state, certain sites meet specified criteria to be managed by the U.S. Environmental Protection Agency (EPA). To be declared a superfund site, a site undergoes a preliminary assessment to determine if there is a potential for hazardous substances to be released, posing a threat to public health and the environment (EPA, “Superfund Cleanup Process”). If the site achieves a Hazard Ranking System (HRS) site score of 28.50 or greater, as determined through the preliminary assessment, it then gets placed on the

National Priorities List (NPL) (EPA, “Hazard Ranking System”). After being placed on the NPL list, it undergoes a remedial investigation, and the different treatment options and potential threats are determined as well as the total cost of the treatment option chosen. Then, the EPA will recommend a preferred remedy through its Proposed Plan before presenting the cleanup plan for public comment. After all the information needed to clean the site is determined, detailed cleanup plans are developed and implemented (EPA, “Superfund Cleanup Process”). Once the construction is completed, additional actions, such as routine monitoring of a site or enforcing long-term site restrictions, are maintained and monitored to ensure that cleanup continues to protect human health and the environment.

There are a total of 153 superfund sites located in the state of New Jersey (EPA, “ Search for Superfund Sites Where You Live”). Of those 153 sites, 37 have been deleted from the National Priorities List, meaning the necessary procedures to clean up the sites have been completed and the sites are deemed to present no harm to the community and environment. However, 116 sites remain on the NPL. These sites may be in the preliminary stages of remedial investigation to assess the site’s conditions and potential solutions, currently undergoing treatments, or still being monitored post-treatment. New Jersey has a total of 21 counties, all of which have at least 1 currently existing superfund site. Only 5 counties have 1 superfund site, but 16 counties have 2 or more sites, with Middlesex County having the most at 18 superfund sites—4 that are still in existence and 4 that have been deleted.

Marginalized, Overburdened, and Food Desert Communities

Although all local communities are vulnerable to the impacts of hazardous substances associated with superfund sites, certain communities may experience greater vulnerability. They include marginalized communities and overburdened communities as denoted by the NJ Department of the Environment and by the US EPA. The EPA defines overburdened communities as minority, low-income, tribal, and indigenous populations or communities that experience environmental harms due to exposure or vulnerability to environmental hazards (EPA, “What is the definition of ‘overburdened community’ that is relevant for EPA Actions and Promising Practices”). The New Jersey Department of Environmental Protection (NJEDP, “Overburdened Communities”) defines an overburdened community as a community with at least a 35% low income, 40% minority residents, and 40% limited English Proficiency. According to the criteria set by the NJEDP, there are a total of 3,687 overburdened blocks in NJ. Vulnerability may be increased for these communities because of the geographically focused accumulation of negative environmental, health, economic, or social conditions. Low-income communities, for example, may have limited access to healthcare services and economic opportunities, further limiting their ability to pay for healthcare. These low-income areas may also provide cheaper housing because of their proximity to environmental hazards, exposing them to heightened health risks.

Table 1. Number of Overburdened Communities by NJ County

County	Overburdened Communities (Census Blocks)
Atlantic	107
Bergen	474
Burlington	132
Camden	208
Cape May	23
Cumberland	78
Essex	529
Gloucester	59
Hudson	422
Hunterdon	5
Mercer	183
Middlesex	455
Monmouth	118
Morris	96
Ocean	94
Passaic	257
Salem	13
Somerset	106
Sussex	9
Union	303
Warren	16

Moreover, these communities may also suffer from lack of access to affordable, quality food. Such food desert communities are areas with limited access to affordable, nutritious food, often due to a lack of nearby grocery stores or full-service food retailers. The New Jersey Economic Development Authority (NJEDA, “NJ Food Desert Community Designation

Methodology") defines these communities using a set of criteria that go beyond simple distance to food. They consider multiple factors, including the availability of supermarkets and grocery stores, demographic characteristics (such as population density and income levels), economic indicators (like poverty rates and employment), and health indicators (including diet-related health outcomes). Rather than relying on a single measure, the NJEDA evaluates many variables to determine which communities face the greatest barriers to accessing healthy food. This definition highlights that food deserts are not just about geography but also about economic and health inequalities that limit consistent access to nutritious food. Importantly, many of the communities identified through this process are also affected by other disadvantages, including environmental burdens, which further amplify the challenges residents face. This overlap of disadvantages becomes clear when examining New Jersey's most overburdened areas, where food insecurity and environmental hazards occur in the same communities.

NJ Marginalized, Overburdened, Food Desert Communities

Many of New Jersey's most vulnerable communities are both environmentally and economically challenged, often overlapping with designated food desert communities. According to the NJEDA, the "top" food desert communities, such as those in Camden, Newark, Trenton, and Paterson, rank highest in need due to compounded challenges like poverty, limited food infrastructure, and adverse health outcomes. These communities are often considered "overburdened" because they face multiple layers of disadvantage, including environmental hazards (such as proximity to Superfund or contaminated sites), economic hardship, and reduced access to healthy food options. The overlap between environmental injustice and food insecurity demonstrates how severe issues may concentrate in specific areas, making it more difficult for residents to maintain healthy lifestyles and increasing long-term health risks.

Hydroponics as a Solution

Hydroponics is a technique whereby plants are grown without soil. Instead the plants are given water, nutrients, and oxygen directly through the roots in a controlled environment. This technique is very useful for communities affected by soil contamination because food can be grown without contact with remediated or possibly contaminated areas (Urban Agriculture). In contaminated sites, normal gardening presents potential threats because of the risk of soil contamination, even after remediation. Hydroponics systems, however, nearly eliminate such risks because it fully distances the plants from contaminated soil while still allowing fresh produce to be grown locally.

There are several methods that can be used depending on budget and community needs. The first is the Kratky method, a non circulating system where plants are placed in a nutrient solution and roots are partially exposed to air, which is the simplest and lowest cost option (Kratky, 184-186). It is a manageable low budget system since it does not require pumps nor the constant supply of electricity. Deep Water Culture uses plants suspended in the air in oxygenated nutrient water which allows for fast plant growth while maintaining a stable and consistent

nutrient flow. Nutrient film technique involves circulating a small stream of nutrient solution across the roots and is used mostly in large food production systems because it is easy to scale up and is highly efficient. Each of these hydroponics systems, while slightly different and suited to different environments, can produce food throughout the year under controlled conditions whilst reducing water waste. Hydroponics also addresses broad environmental and health concerns. Since nutrient distribution is controlled, water use is significantly lower than in normal soil-based agriculture; runoff is also minimized (Kratky, 184-188). This prevents pesticides and other pollutants from leaking into nearby environments. Controlled conditions also allow production to be more predictable across several seasons, which is especially important in urban communities in which access to fresh produce could be limited. For communities affected by either environmental contamination or food access challenges, hydroponics presents a practical solution.

To serve larger communities, individual hydroponic units may not suffice. Instead, a more productive and efficient way to grow hydroponically is with freight hydroponics. In such larger systems, crops can be grown inside shipping containers which are modified to equip lighting, ventilation, nutrient systems, and temperature control (Neis). Freight hydroponics allows food production to be controlled in compact spaces with minimal land (Freight Farms, "Our Mission"). These systems are especially helpful in places where soil is contaminated or where space is limited, allowing for year-long reliable production regardless of the environment or weather conditions (EPA, "Urban Agriculture"). This makes hydroponics very helpful for urban communities that need sustained food production throughout the year without requiring farmland (EPA, "Urban Agriculture").

Many organizations have used freight hydroponics and have shown how useful these systems are (Freight Farms, "Our Mission"). They are used in educational and communal settings. In New York City, there are many educational programs that educate and train individuals on how to grow hydroponically (EPA, "Urban Agriculture"). Programs introduce how food is typically grown and the differences and similarities to controlled environmental growth. Other places in New Jersey have implemented educational food programs where schools use indoor growing systems (EPA, "Urban Agriculture"). These are used to teach science, sustainability, and food production while also channeling the produce grown hydroponically to feed students. These programs show that hydroponic systems can be used to reach educational goals and nutritional goals simultaneously (EPA, "Urban Agriculture").

Private companies and nonprofits have also utilized freight hydroponics. Container based companies such as Freight Farms have shown that shipping containers can be used to produce vegetables and herbs in packed urban areas while requiring substantially less water than traditional agriculture (Freight Farms, "Our Mission"). Nonprofits have used freight hydroponics to grow food locally, donate to community organizations that serve low-income families or the elderly, or sell healthy fruits and vegetables at discounted prices (NJEDA, "FEED NJ: List of Primary Focus Food Desert Communities"). These examples show that hydroponics can be used not only as a good growing system, but also as a community resource. Partnerships with local

organizations help distribute the benefits of using hydroponic systems throughout the community. For example, in 2017, Mayor David Fredi in Robbinsville, NJ established a nonprofit hydroponics freight unit after taking inspiration from college visits (Neis). Since then, Hope Neis has taken over the project and has been working with local organizations such as the local senior citizen home and food pantries nearby. Hope Neis manages the entire freight unit and delivers fresh produce to the senior citizen home and local food pantry (Neis).

The Robbinsville example shows that controlled agriculture does not require several acres of farmland, revealing that produce can be grown locally within existing communities. It serves as a model that other municipalities can follow to foster better access to healthy, affordable food despite possible limitations of space and challenges with the environment and weather conditions, particularly in areas denoted as Food Desert Communities.

Hydroponics in Remediated Sites: The Potential of the Historic Potteries Site

In New Jersey, remediated Superfund sites and formerly contaminated sites have been used in ways that benefit the surrounding communities (EPA, "Learn about Superfund"). They provide public access, recreation, or improvements to the local environment. In many cases, reuse plans have used green houses, public parks, solar panels, or garden spaces to allow residents to benefit from the previously unstable land. These examples show that remediation can be more than simply removing hazardous material. The remediated sites can be utilized to further help the local community. In towns where residents face challenges to food access, the site reuse can provide a space for controlled food production and environmental education. At the Historic Potteries site, remediation efforts have mostly been focused on reducing exposure to lead contaminated soil and stabilizing affected areas. However, current reuse options do not address the communities food desert (NJEDA, "NJ FEED NJ: Primary Focus Food Desert Communities"). While protection of residents and the environment remains the priority, the site provides an opportunity for reuse that supports local needs. Since the surrounding areas face environmental hazards and limited fresh food access, a reuse plan that integrates controlled food production would better address the community's challenges. Compared to other sites where gardening and green reuse has already been implemented, Historic Potteries remains a location where additional benefits can be channeled to the community without interfering with remediation goals. A controlled hydroponics system offers a new source of fresh produce that does not disturb soil and can function within already managed spaces. It also provides an alternative to residents who may have previously gardened in the contaminated soil of their backyards. This creates a chance to extend the site reuse beyond just cleanup by adding a practical community resource that supports both environmental safety and food access.

The Case of the Historic Potteries Site

Background

For over 100 years, the East Trenton neighborhood in Trenton, NJ has been exposed to lead contamination in the soil from the operations of a former pottery manufacturer. Pottery manufacturing in Trenton operated from the 1850s for over 100 years before eventually shutting down, leaving behind industrial waste which contaminated the environment (EPA, "Historic Potteries Trenton NJ: Cleanup"). During operations, airborne lead could have contaminated the surrounding soil. However, even after the pottery manufacturer shut down, pottery shards that had been left in the soil ultimately leached lead-containing ceramic glazes into the soil which also became airborne as dust. As a result, the soil in the surrounding neighborhoods was contaminated by lead. The contamination site spans residential properties, public parks like Sonny Vereen Park and Breunig Avenue Park, and schools such as Ulysses S. Grant Intermediate School in East Trenton, affecting students and ordinary citizens. Since 2018, the U.S. Environmental Protection Agency has conducted extensive investigations, including large-scale soil sampling of residential properties, schools, and parks in Trenton, collecting over 1,100 samples to determine the extent and sources of lead contamination. Results showed widespread and severe contamination, with lead concentrations ranging from 10 to 7,760 mg/kg with 859 samples exceeding the EPA's 200 mg/kg risk level, and some properties and former pottery sites reaching extreme levels up to 17,000 mg/kg, primarily linked to airborne kiln emissions and leaching from pottery waste (Evangelista, 3-4).

Table 2. Quadrant Composite Sampling Lead Results Summary

Area	Depth (inches below ground surface)	# of Samples	# of Samples Exceeding 400 mg/kg for Lead	Maximum Lead Result (mg/kg)	Average Lead Result (mg/kg)
HP001	0-2	76	58 (76 %)	4,600	936
	2-6	76	60 (79 %)	7,500	1,090
	6-12	76	59 (78 %)	4,200	919
	12-18	76	53 (70 %)	2,200	625
	18-24	76	31 (41 %)	1,900	431
HP002 – HP006	0-2	10	7 (70 %)	1,350	551
	2-6	10	7 (70 %)	1,540	616
	6-12	10	5 (50 %)	1,230	509
	12-18	10	5 (50 %)	1,660	505
	18-24	10	2 (20 %)	1,070	334
HP007	0-2	4	1 (25 %)	510	300
	2-6	4	1 (25 %)	542	320
	6-12	4	1 (25 %)	431	177
	12-18	4	0 (0 %)	370	126
	18-24	4	0 (0 %)	377	135

* The table includes all locations, except dripline, collected throughout L.H. Mitchell Site and Historic Potteries Site assessments (Evangelista, 7)

EPA Response

The EPA has taken significant steps to clean up and remediate lead-contaminated soil in East Trenton through a targeted removal action focused on reducing human exposure. This effort prioritizes high-risk residential properties where lead levels exceed the Removal Management Level, especially in areas frequently used by residents, such as gardens and children's play spaces, and in households with young children or pregnant women (Evangelista, 2). At these properties, temporary protective covers such as mulch, stone, clean soil, grass, or raised garden beds are installed to create a barrier between residents and contaminated soil. These covers help prevent direct contact and reduce the risk of inhaling or ingesting lead particles, although they require ongoing maintenance to remain effective.

Of 143 residential properties tested, 62 have been identified as high-risk and are prioritized for these protective measures. The remaining 81 properties, while still showing elevated lead levels, do not currently meet the criteria for immediate intervention (Evangelista, 5-9). Residents at these locations are informed of the risks and provided with educational resources to minimize exposure. The EPA continues to monitor these properties and will reassess them if conditions change, such as new residents or different land use, which could increase exposure risk. The cleanup process also includes detailed planning and safety procedures, such as developing site-specific plans, obtaining property access, monitoring air quality, suppressing dust during soil disturbance, and properly disposing of contaminated waste. After the installation of protective covers, affected areas are restored and maintained to ensure continued protection. In addition to residential areas, the EPA has addressed contamination at public locations, including Grant Intermediate School, Sonny Vereen Park, and Breunig Avenue Park, by installing protective covers over contaminated soil. Overall, these actions help reduce immediate health risks by limiting exposure to lead while longer-term remediation efforts are being developed.

Recommendations

Other recommendations should focus on strengthening and expanding current efforts to better protect public health and ensure long-term safety. One key recommendation is to move beyond temporary soil covers and develop permanent remediation solutions, such as fully removing and replacing contaminated soil in high-risk areas. While interim controls reduce exposure in the short term, they require maintenance and may degrade over time, so a more permanent solution would be more effective.

It is also recommended that testing and remediation be expanded to additional properties, including those that currently do not meet high-risk criteria but still show elevated lead levels. Conditions at these properties can change over time, and early intervention could prevent future health risks. Continued monitoring and regular reassessment of all affected properties should remain a priority.

Another important recommendation is to increase community education and outreach. Residents should be provided with clear information about the dangers of lead exposure and practical ways to reduce risk, such as proper handwashing, removing shoes before entering the

home, and using safe gardening practices. This is especially important for households with young children and pregnant women, who are most vulnerable to the effects of lead.

Improving communication between the EPA, local authorities, and the community is also essential. Keeping residents informed about ongoing cleanup efforts, future plans, and any new findings can build trust and encourage cooperation with remediation efforts. Finally, long-term environmental monitoring should be implemented to ensure that contamination does not return and that remediation efforts remain effective. This includes periodic soil testing and evaluation of previously treated areas. Together, these recommendations would help create a more comprehensive and sustainable approach to addressing contamination at the site.

The Holistic Solution of Hydroponics

After remediation, superfund sites can be reused in multiple ways such as for municipal open space or storage, agricultural use, commercial space, residential housing, and natural conservation land. However, reusing the space while considering the risk of lead contamination to address the additional needs of marginalized, overburdened food desert communities can help address multiple needs of these vulnerable communities. Freight hydroponics could provide a long term holistic reuse strategy for the East Trenton community where the superfund Historic Potterites site is by addressing its need for food access without risking the use of contaminated soil (Learn about Superfund).

Based on site maps and surrounding land use, a practical area for a freight hydroponic unit is George Page Park. George Page Park offers an accessible public space that is close to the affected neighborhood while avoiding direct interference with remediated zones. Placing a freight hydroponics container by an existing public area would allow the project to remain connected to the Historic Potteries community whilst operating in a space that is easy to access, monitor, and integrate into local programs. Since freight hydroponics utilizes a self contained shipping container system, crops can be grown completely above ground under controlled conditions under artificial lighting, nutrient systems, and regulated temperatures (Kratky 184-186).. This makes the system suitable even where direct soil based agriculture would remain a hazard.

For long term operation, nonprofit partnerships would help ensure that the unit functions to provide food access and spur community development rather than only a technical unit. Local nonprofit organizations such as C.R.O.P.S. could assist in coordinating produce distribution, educational enrichment, and community participation while helping residents understand how controlled environment farming works (FEED NJ: Primary Focus Food Desert Communities).

Additional benefits include training and economic development beyond just providing steady access to food (Environmental Justice: What are Overburdened Communities?). Programs supporting such workforce development in hydroponic farming could therefore spur job and economic growth in connection to environmental recovery. Urban agricultural food production through hydroponics when paired with the support of local organizations such as in Newark and Camden lend support to these community benefits. In these places, urban agriculture projects

have contributed to food access, environmental education, and community engagement while also supporting local produce and workforce development. Applying a similar model in Historic Potteries would allow for remediation benefits to extend beyond just hazard reduction by creating a direct community asset. With the help of nonprofit coordination and careful site placement, freight hydroponics offers a practical way for Historic Potteries to address both its environmental issues and current food challenges.

Works Cited

Evangelista, Pat. SPP - Technical Reports and Studies. U.S. Environmental Protection Agency, 2024, https://semspub.epa.gov/work/02/735644.pdf. Accessed March 1, 2026.

"Hazard Ranking System." US EPA, 22 Sept. 2025. https://www.epa.gov/superfund/hazard-ranking-system-hrs

"Historic Potteries Trenton, NJ: Cleanup Activities." US EPA. n.d. https://cumulis.epa.gov/supercpad/SiteProfiles/index.cfm?fuseaction=second.Cleanup&id=0203535#background

"List of Primary Focus Food Desert Communities." NJ Economic Development Authority (NJEDA). Jan. 2025. https://www.njeda.gov/wp-content/uploads/2025/01/List-of-Primary-Focus-FDCs-FEED-NJ.pdf

Neis, Hope. Zoom interview. Conducted by Andy Kwon, 28 January 2026.

NJ Food Desert Community Designation Methodology." NJ Economic Development Authority (NJEDA).9 Feb. 2022. https://www.njeda.gov/wp-content/uploads/2022/02/New-Jersey-Food-Desert-Community-Designation-Methodology-Final-2-9-22.pdf

Our Mission". Freight Farms, 2025. https://www.freightfarms.com/

"Overburdened Communities." The NJ Department of Environmental Protection. 15 April 2026. https://dep.nj.gov/ej/communities/

"Search for Superfund Sites Where You Live." The US Environmental Protection Agency, 5 Jun. 2026. https://www.epa.gov/superfund/search-superfund-sites-where-you-live.

"Superfund Cleanup Process." The US Environmental Protection Agency, 12 Jan. 2026. https://www.epa.gov/superfund/superfund-cleanup-process.

"Urban Agriculture." US EPA. 25 Mar., 2026. https://www.epa.gov/brownfields/urban-agriculture

"What is the definition of 'overburdened community' that is relevant for EPA Actions and Promising Practices?" The US Environmental Protection Agency, 9 July 2025. https://www.epa.gov/caa-permitting/what-definition-overburdened-community-relevant-epa-actions-and-promising-practices

"What is Superfund?" The US Environmental Protection Agency, 17 Sept. 2025, https://www.epa.gov/superfund/what-superfund

The Art of Dealing Art: Pricing, Information Asymmetry, and Inequality in the Contemporary Art Market (1980-2025) By Trang Nguyen

Abstract

Some contemporary artworks are receiving exorbitant prices, with each painting setting a precedent over another. Behind the scenes, these prices are set by a complex network of art dealers, auction houses, collectors, and artists themselves. It is thus in this paper's interest to look at key actors in the art market and determine their significance. Since previous research has been done on auction houses, this research paper tries to shed light on the role of art dealers in dealing art. This paper aims to answer the following question: How do art dealers influence the pricing of artworks, and what are the implications of their role for market efficiency, transparency, and artist-level inequality? By understanding the involvement of art dealers in creating market inefficiency—information obscurity and artist inequality—collectors would be more informed when appraising art and would consider investing in a wider scope of artists.

Introduction

In March 2022, Christie's staged its 20th/21st Century: Shanghai to London art sale series. The auction series included established figures such as Pablo Picasso and Lucian Freud, with Kandinsky's The Foxes reaching a hammer price of 43 million. Alongside these well-established artists, art dealer Victoria Miro also placed the artwork of contemporary and emerging artists, such as the art of Flora Yukhnovich, whom Miro personally represents. This proximity has impacted the perception of Flora's artwork by collectors and, correspondingly, their prices. Ultimately, Flora's Tu vas me faire rougir realised 1.9 million GBP, which was around 8 times its previous estimate. In addition to that, since her representation by Miro, her paintings have gained traction in other auctions like Sotheby's. These details show that art dealers can frame expectations and raise the estimated value of art.

This event matters because the continual rise in the price of Flora Yukhnovich's work could be attributed to many factors. The curation of a prestigious art auction acts as a quality mark and shapes collectors' confidence, on top of which the representation by a well-known art dealer, Miro, could increase the perceived value of her works. Art dealers could amplify the scarcity and raise expectations about future demand. The placement of Yukhnovich's painting amongst other high-bidding paintings could influence collectors' appraisal by signalling its status. This poses a few questions: are large differences in presale price and final price because of its worth, or caused by its curatorial framework? Is the price of art determined by its demand or by art dealers orchestrating access to information? This event shows that there are economic agents behind the price-setting of art and provides a basis for a focused lens on art dealer's representation of artists.

The research question of this paper, therefore, is how art dealers influence the pricing of artworks, and what the implications are for market efficiency, transparency and artist-level inequality are. I will focus on contemporary art markets from 1970 to the present to provide

more recent examples of art-dealing practices. Understanding this will help readers interpret signals and assess who gets access to opportunity.

In the first section, I discuss the view of Velthuis in 'Talking prices: Symbolic Meanings of Price on the Market for Contemporary Art' (2005) which I believe helps to introduce the symbolic notion of prices and further the understanding of the role of art dealers in gatekeeping information. I position my paper to test his claim that major art dealers construct the value of art through pricing strategies. I will also review Fabio Rojas's article on 'Art dealers as competitors and community builders: Understanding the social logistics of the contemporary art world' (2023). This article introduces new light on art dealing activities and provides a more socio-economical basis which I aim to test in this research paper. Together, these perspectives deepen knowledge on the roles of art dealers on the representation of artists, penetrating market inefficiency and artist-level inequality.

In this paper, I use market transparency - or just "transparency" from now on - to denote the public's access to information such as pricing and provenance. This 'lack of transparency' is considered a pertinent problem and means limited public access to key transaction details. The level of transparency not only shapes the bargaining power of the dealer but also influences the prices of artworks. In addition to this, I will also define artist-level inequality as the unequal access to opportunity given to certain artists; not all artists are given equal access to representation and curation. These two core concepts will be used to explore different art-dealing activities and the notion of value in the art market.

The second section of this paper will examine the development of art dealerships from Leo Castelli's model around the mid-20th century to the mega gallery model popularised by Larry Gagosian until the start of 21st century. The paper will study the practices of other art dealers, such as Barbara Gladstone, to assess whether a more traditional approach leads to a lack of transparency. By using price increases of artworks as evidence, the roles art dealers play will be assessed in terms of auction showcasing, pricing strategies and speculative buying. Quantitative data as well as quotes from collectors will be analysed for an understanding of art dealing, thus clarifying how early art dealing practices penetrate market inefficiency.

In the third section, I will examine art dealing from the start of the 21st century. The newer art dealers such as David Kordansky will be explored with a stronger focus on the Victoria Miro gallery. The concept of transparency is going to be assessed in relation to representation by the art dealer, curation through action houses and online viewing rooms made popular due to the Internet. An analysis of artists leaving a mega-galery is used to provide a further understanding of the functions of art dealers.

The fourth section will examine art dealers' implications on artist-level inequality using the Art Basel & USB Art Report on Art Economics and Freelands Foundation's annual report on representation. By analysing disparities on gendered representation between commercial and non-commercial galleries, primary market and secondary market access, and sales revenue of top selling artists, this section will assess whether market inefficiency is caused by art dealers or

structural mechanisms of the art market itself. Using quantitative evidence from Sotheby's sales, these reports will investigate the role art dealers play in other channels like auctions as well.

This research adds to crucial existing literature that aims to clarify how art dealer activity relates to the prices that buyers and sellers actually see; and how information moves within this market. This can help explain why some artists enter high-visibility channels and others do not and illustrate why art collectors should not assume that prices tell the full story. This can offer a clearer framework on art investment which helps by identifying problems with price signals in the art market and guiding art appraisers to a more thorough evaluation of artworks. This paper argues that art dealers influence prices by adding emerging artists' names onto international auction houses' rosters and strategically filtering out information about the artwork. Combined, these practices shield the market from the outside and widen unequal outcomes among artists.

Historiography/Literature review

Past authors have argued that the structural mechanism behind auctions and the process of bidding reflect the prices of artworks[55]. Velthuis complicates this by introducing the role of art dealers in setting prices, emphasising how they actively construct symbolic meanings and add to the value of art[56]. Other scholars have focused on other aspects of art dealing with some showing that an expansion of one's trading network affects art auction prices[57]. Following this lens on information transparency, certain authors argued how art dealers are gatekeepers in order to gain profits whereas those who don't are unable to survive in the market[58].

One example of this is Velthuis's book 'Talking prices: Symbolic Meanings of Prices on the Market for Contemporary Art' (2005), where he argues that prices not only signal the value of the artwork but also transmit other complex meanings to buyers and sellers beyond artistic worth. Disputing the claims of Bruno Frey that participants in the art market are rational, meaning they aim to maximise profits, Velthuis attributes the role of art dealers to understanding their trading partners and conducting a meaningful circuit of commerce. Therefore, he argues that prices have symbolic values which transcends into a language used by art dealers to set the prices of artworks.

Following Velthuis's viewpoint, my argument is that while prices are central to showing demand for an artworks, prices may also have symbolic and cultural meanings. Whilst high prices may signal the top rankings of an artist, it could also be interpreted as overvalued. Low prices for an artwork may signify a 'bad' artwork or modesty on behalf of the artist. The

---

[55] Orley Ashenfelter and Kathryn Graddy, "Art Auctions," in: Handbook of the Economics of Art and Culture, 2006; Luc Renneboog and Christophe Spaenjers, "Buying Beauty," Management Science, 2013.

[56] Olav Velthuis, "Talking prices: Symbolic meanings of prices on the market for contemporary art," Princeton University Press (2005)

[57] Dakshina G. De Silva et al., "Evolution of a dealer trading network and its effects on art auction prices," European Economic Review, 2022.

[58] Fabio Rojas. David de la Torre, "Art dealers as competitors and community builders: Understanding the overlapping social logics of the contemporary art world", 2023.

interpretations of prices affect the value of artwork; the decision is not purely controlled by the price mechanism but other economic agents, notably art dealers.

Firstly, the introduction of a social meaning brings a more holistic view of pricing. Rather than just a pure economic lens, this perspective allows cultural economics to be used to enhance models of price formation. By acknowledging the complexities that underpin price-making in the art market, readers would be able to see how the reputation, status, and power of artists and art dealers can influence price of artworks. However, recognising that prices have symbolic meanings may also risk justifying art dealers' gatekeeping of information. Thus, this makes it harder to critique art dealing exercises which could exacerbate market inefficiency and artist-level inequality.

Another example of this is Fabio Rojas' article, 'Art dealers as competitors and community builders: Understanding the social logistics of the contemporary art world'(2023). Whilst Velthuis demonstrates that art dealers control access to the art market, Rojas argues that this is an incomplete understanding of their function. Rojas contends that top dealers in the high-end art market rarely say they use community-focused practices, but those in less competitive segments do, often having to sacrifice economic profit. This highlights how art dealers have to balance between competitive and cultural logics depending on their position and goals[59].

My argument is that art dealers are still gatekeepers to the art market as they still choose certain artists to represent. Although some art dealers say they employ more communal practices, they might still do this with a competitive logic; dealers have incentives to present commercial strategy as aesthetic judgement or community practice. Thus, this questions whether community-oriented practices and competitive logics are mutually exclusive. As little is known about the activities of art dealing, this further highlights the lack of information transparency. The fact that community-focused dealers earn less suggests a structural imbalance in the art market, where a lack of transparency enables greater economic success.

By demonstrating that art dealers act as gatekeepers, art collectors could be more aware of the influence they exert on the art market and not fully rely on them when valuing art. This perspective allows for the critiquing of art dealers' role in artist-level inequality and encourages further support for underrepresented artists. Furthermore, it highlights the need for research into market inefficiency in the art market, since other economic agents may also contribute to the lack of information transparency. Nevertheless, this perspective overlooks the genuine motive of art dealers when pursuing community-focused practices, as could be the case if there were more information. It also leaves out the various aspects of art dealing, as suggested by Rojas, and risks rendering the role of art dealers oversimplified.

Conclusively, although Velthuis and Rojas enriches the meaning of art dealers, the articles still imply that art dealers play a huge role in restricting access to the art market. Established art dealers control information and set prices to transmit symbolic meanings of

[59] Rojas. Torre, "Art dealers as competitors and community builders: Understanding the overlapping social logics of the contemporary art world"

artworks, which influences the expectations of consumers. This suggests that there are structural imbalances in the art market as established art dealers may hold disproportionate influence. Thus, the underrepresentation of artists and artist-level inequality could be attributed to the role of art dealers as well.

The next sections will explore how art dealers contribute to market inefficiency, notably a lack of information transparency and artist-level inequality. However, when doing so, the research is limited by the objective of its study as it tries to clarify information transparency which is a problem itself. Sources are difficult to find because they are often paywalled or hidden by auction houses, so insider knowledge into art dealers' networks and power relations might be lacking. To solve this problem, I rely on indirect sources and public 'traces' such as estimates and realised prices, published interviews, as well as market reports.

First content section

This section will compare two dealership models, an early model such as Castelli's and Gagosian's scale-driven dealing, in terms of their impact on price formation, artist visibility and information transparency. Art dealing practices have been growing more dominant since the 1950s, with the 'Leo Castelli Model' being the first of its type. What started as a one-man gallery for emerging artist Jasper John has earned Leo Castelli a name as one of the most influential art dealers. With the help of Castelli, Larry Gagosian came on to the art scene and took hold of the secondary market. He introduced the mega gallery model - an influential gallery with multiple locations and a roster of at least 40 artists so as to distinguish them from other galleries[60]. Art dealers used prices to transmit symbolic value, effectively acting as vehicles of differentiation. In order to manipulate prices, art dealers used the lack of information transparency to their advantage; this allowed them to gain higher profits but infiltrated a lack of access to the art market for some artists.

The 'Leo Castelli Model' was the first establishment of the relationship between art dealers and artists, where the art dealer would provide art supplies, studio space, and open channels of dialogue. Also known as the 50/50 model, the idea was that the art dealer would carry the cost and risk of promoting an artist and the artist would share their paintings. Art dealer Edward Winkleman commented on this widespread approach of selling contemporary art: "Leo Castelli did not invent many of the strategies behind his unparalleled success at selling contemporary art…but he optimized a series of 'best practices' for art dealing in a way that was so admired by artists and younger dealers alike that it unofficially became 'the way it's done'. The cornerstone of the model was for mutual trust and loyalty between dealer and artist to guide what was presumed to be a long-term relationship over the inevitable ups and downs of both an art career and the economy's cycles."[61]

---

[60] Edward Winkleman, "How to run a gallery: three threats you can't ignore", 2015. https://www.theartnewspaper.com/2015/10/01/how-to-run-a-gallery-three-threats-you-cant-ignore

[61] Nate Freeman, "Why Leo Castelli Paid his Artists Even When They Weren't Making Work", 2018. https://www.artsy.net/article/artsy-editorial-leo-castelli-changed-art-market-forever

Evidence suggests that the Leo Castelli Model was effective in supporting artists; emerging artists would be supported financially and therefore have cushioned risk. However, the act of choosing who to represent by the art dealer could already cause inequality; other artists, not funded by art dealers, were less likely to flourish or survive through 'the inevitable ups and downs of both an art career.' Hence, it depended upon the funds and tastes of art dealers that determined the success of artists in the past. As Velthuis argues, art dealers act as gatekeepers as they control access to the art market, which further reinforces their power in shaping prices.

Whilst demand and supply determined the market price of artworks, Leo Castelli asserted that "price is often based on imponderables" with 'imponderables' being his friendship with the purchaser, the distinction of the purchaser and the frequency of his purchases, etc[62]. This demonstrates that Castelli doesn't rely solely on the market mechanism, but rather uses relational factors to determine the price. This follows the argument of Velthuis that art is not a simple commodity and that through setting a price, the art dealer is able to transmit some symbolic value. Thus, the price set by the art dealer might be different from the equilibrium price set by demand and supply, showcasing the role art dealers play in setting the value of artworks.

When Gagosian was starting out and was not yet a household name, his LA exhibition of Jean-Michel Basquiat's artworks still proved a phenomenal success. In April 1982, Gagosian showcased Basquiat's works such as Tuxedo, which earned him acclaim from the international public and collectors[63]. As this pinned Gagosian's name on the map, he became known for his focus shows, retrospectives or museum-like exhibitions. The mega-gallery model pioneered by Gagosian gave artists huge opportunities to showcase their works, as graphic designer Peter Saville and director Nicolas Serota argued the merits of this model in accommodating artists's pursuit of creative expression[64].

Basquiat later complained that works were released which he did not want to release, which points to the pressure of inflated demand on artists[65]. This evidence suggests that dealers can create more demand for artworks through high-profile shows, causing momentum buyers with resellers to enter the market and causing demand to expand far beyond the core customer base. This creates production strain on the artists, as artists like Basquiat have to produce more artworks to meet demand, showing inefficiency in the art market. Ultimately, in the long run, the purchasing capacity of buyers steadies down, unsold inventory can push prices down toward what the core base is able and willing to spend. This establishes that art markets have cycles in which art dealers have a direct implication on transmitting value and speculative buying.

Jean-Michel Basquiat also said that, "There's more dealers than artists these days"[66]. The higher number of dealers to artists meant that more information could be gatekept or distorted,

---

[62] Grace Glueck, "How do you price art works? Let the dealers count the way", 1976.
[63] Anthony Mason, "Gallery owner Larry Gagosian on the art world's "blood sport", 2025. https://www.cbsnews.com/news/gallery-owner-larry-gagosian-on-the-art-worlds-blood-sport/
[64] Robert Armstrong, "'I like the ways things look.' The world according to Larry Gagosian", 2022. https://www.ft.com/content/eefe7a73-7798-4e56-b8f9-70c35a9e68e9
[65] Cathleen McGuigan, "New art, new money", 1985.
[66] Cathleen McGuigan, "New art, new money" (1985)

which makes the hammer price differ massively from its estimated price. Furthermore, when the concentration ratio of dealers to artists becomes too high, dealers have the ability to decide which artists to represent. This causes artist-level inequality since some artists are chosen to be represented and it can be worsened especially when only a few artists benefit from the upswing of the cycle.

Art dealer Barbara Gladstone, on the other hand, does not want her gallery to become a mega-gallery. Unlike Gagosian, she favors tradition and relationships, as she said during an interview: "I think that we still follow the traditional model of representation…I'm talking to the artists. That's what I want to do. And so are the senior partners and directors. It's still a very personal connection that we have and that I would not ever want to change."[67] Her gallery has successfully presented conceptual artists such as George Condo and Jenny Holzer; this success shows that different practices can still succeed commercially in addition to mega galleries.

Gladstone's statement shows that different motives exist; whilst she does not frame her gallery as a mega-gallery, her gallery's size could still be comparable in terms of returns to other galleries. Instead of using price or predicted future value of artists, Gladstone transforms long-term dealer–artist relationships into a mechanism for shaping career trajectories and managing price progression over time. Contrastingly, her approach suggests that the closer the dealer-artist relationship is and the more information there is, the higher the likelihood to be successful. Since not all dealers aim to manipulate price, the extent of dealers' implication of pricing artworks varies reliant on their motives.

This section found that art dealers decide on prices of artworks based on relational and symbolic factors. Endorsement by art dealers could have caused speculative buying, leading to production surges and price cycles that only benefit certain artists on the upswing. The selection practices used by dealers perpetuated this artist-level inequality. However, art dealers' influence on prices of artworks depended on their motives, with the success of galleries not entirely dependent on their gatekeeping ability, which affects transparency.

Second content section

The financial crisis of 2008 marked a significant event in how markets operated; research has shown that risk aversion increased after 2008 overall[68], but it is up to speculation how this has affected the contemporary art market specifically. Thus, this section aims to investigate the change the crisis has had on art dealers. The paper will explore art dealing practices post-2008 with data on sales of paintings of different artists. New art dealers, including David Kordansky, will be introduced alongside a continuance of research on major galleries of Gagosian and

[67] "The Art World: What If…?!, Season 2, Episode 8: Barbara Gladstone," podcast audio, The Art World: What If…?!, hosted by Charlotte Burns, March 14, 2024, Art& (Schwartzman&), https://www.schwartzmanand.com/art-posts-the-art-world/the-art-world-s2ep8-barbara-gladstone

[68] Schooley, D., and Worden, D. (2016). Perceived and Realized Risk Tolerance: Changes During the 2008 Financial Crisis. Journal of Financial Counseling and Planning, 27, 265 - 276. https://doi.org/10.1891/1052-3073.27.2.265.

Victoria Miro. Despite the perception of the democratisation of information[69], the online business model of mega-galleries has allowed art dealers to gatekeep more information, reinforcing lack of transparency and artist-level inequality.

Artists began to change representation in art dealerships, with one example being the move of Mr. Chamberlain, a sculptor from Pace Gallery to Gagosian. At Pace Gallery, his paintings were auctioned off between $1 and $4.6 million but at Gagosian, he set a new auction record of $5.5 million at Sotheby's on May 18.

Prices of paintings serve as signals where much information is private. Since collectors rarely have information about the artworks, highlighting the information gap between buyer and seller, they treat the price itself as evidence of future demand. As collector Howard Rachofsky puts it, "Sometimes we don't know if the stuff we're buying is historically significant, but because the prices are so high, we need to believe they're important." Following Gary Gorton's account of 'information-insensitive' assets[70] (assets that trade in normal times as if no further checking is needed), we can see the art market as the opposite case, as the current art market is information-sensitive from the outset: to judge value and long-run demand, buyers would need details about prior sales, discounts, reserves, dealer placement, collector reputation, and institutional validation, yet much of this information stays private. In such a setting, a high list price or high sale price set by a prestigious dealer functions as a visible substitute for missing information: it becomes a public signal that shapes later estimates, appraisals, and bids. The initial high price on an artwork set by Gagosian thus acts as a certification device, which results in an anchoring effect on future appraisals and subsequent bids.

Furthermore, Gagosian's roster of artists and high-end client list influence the price of paintings. Since there is an information gap between buyers and sellers, buyers rely heavily on the reputation of the intermediary as a signal of the value of the artwork. Therefore, a dealer's reputation plays a significant part in determining price. Furthermore, as evidenced by Silva, art dealers with larger trading networks, like Gagosian, have more market information which they can leverage into higher prices, thus creating a lack of information transparency in the art market.

Representation under a famous art dealer can boost prices of paintings, however when prices of paintings start to drop, some dealers reduce support or end representation when sales weaken. In December 2013, Damien Hirst split from Gagosian's gallery. This follows a decline in purchase price of his paintings, down 30% from where it was at its peak with a third of it unsold. Economist Don Thompson described how 'his reserve price is too high' and said: "Hirst can better reposition himself selling at a lower price with another dealer". Others could not

[69] Brown, K. Art markets, epistemic authority, and the institutional curation of knowledge. Cultural Studies, 37(4), 626–646, 2023. https://doi.org/10.1080/09502386.2022.2030777

[70] Iorgova, Silvia, and Chase P. Ross. 2023. "Investor Information and Bank Instability during the European Debt Crisis." Journal of Financial Stability 64 (February): 101100. https://doi.org/10.1016/j.jfs.2022.101100.

understand why Gagosian wanted Hirst to leave because he offers an influential brand for the gallery.

Representation by an art dealer like Gagosian can signal to consumers that the painting will sell well in the future, increasing speculative buying and inflating its prices. This means that the price realised will be further from its real value and each new record price will create an artificial ceiling on top of which the inflationary cycle continues. When market confidence fell like it did for Hirst, art dealers tended to withdraw support from artists to safeguard their branding. Since art dealers gatekeep information about the market, they have a direct implication on the pricing of artworks. Furthermore, despite some art dealers favouring the nurturing of artists, major dealers prioritise profits and competitive logics which can perpetuate artist-level inequality.

David Kordansky is an art dealer who has recently established himself on the contemporary art scene during the last two decades. He represented several artists like Lucy Bull who is known for her abstract distortion of forms. In the October contemporary evening auction of 2023 at Sotheby's in Hong Kong, Bull's painting *Flash Chamber* went for a record $1.7 million. To top it off, another one of her paintings *16:10* sold for $1.8 million in 2024 for Sotheby's sale in New York, doubling its estimate. "That's a high percentage of collectors who are buying her work and then selling it right away," says Amanda Schmitt, a New York-based art adviser - evidence of speculative buying for her works.

The success of Lucy Bull could be attributed to the representation of Kordansky. By being endorsed by David Kordansky Gallery, Lucy Bull got more exposure to blue-chip clients[71]. Moreover, when Kordansky put her paintings on the roster for evening sales, Bull was able to gain an even bigger collector base, which led to higher auction competition. This meant that auction prices of her artwork rose significantly, widening the gap with its estimated price. In this case, Kordansky acted as a certifying agent to symbolise the future value of her paintings and influenced the information received by the buyers in the information-sensitive market. Not only does this show the role art dealers play in price, but it also highlights the inherent problem of the art market of asymmetric information. Despite the growing prominence of David Kordansky Gallery, which has not yet reached the scale of a mega-dealer, the relationship between dealer experience, network embeddedness, and artist market success remains underexplored in the academic literature.

With the introduction of the Internet, art dealers like Gagosian have adopted online viewing rooms as a new stream of engagement with customers. Online viewing rooms acted as a virtual platform to showcase artworks similar to a real gallery[72]. In an auction by Sotheby's or Christie's, art dealers add paintings of emerging artists onto the roster next to already established artists; online viewing rooms create another means for dealers to do this. Past research has shown

[71] Blue-chip clients refer to wealthy individuals, investors, major institutions (museums or foundations) and corporations who buy art for quality rather than speculative buying.
[72] Scott Indrisek, "For galleries, Online Viewing Rooms are about Storytelling and Sales", 2019. https://www.artsy.net/article/artsy-editorial-galleries-online-viewing-rooms-storytelling-sales

that viewers have the same aesthetic experiences online than physical museum settings, thus it is plausible that art dealers can influence price of artworks online as well.

Online viewing rooms allow dealers and galleries to expand their customer base. According to David Zwirner, a mega dealer, his primary goal of using online viewing rooms is to 'tap into new networks of collectors'. Per his gallery's data, approximately 47 percent of sales inquiries come from interested buyers with no prior relationship with the gallery[73]. This meant that there would be increased demand for artworks put in online viewing rooms, which would lead to higher prices. Thus, art dealers, who take advantage of digital platforms, can exert upward pressures on artwork prices.

As more galleries have online viewing rooms, there seems to be more information about artworks that are posted on the Internet. However, as argued by Kathryn Brown, these tools can allow dealers to curate information and limit transparency[74]. Gagosian pioneered the use of online viewing rooms for artist Cecily Brown, where he put high-resolution images of her paintings up for display. During the Frieze New York 2020 online viewing room, Brown's painting *Figures in a Landscape 1* got sold for $5.5 million, the 'second highest price ever paid publicly for one of Brown's works'. Gagosian director emphasized the museum-quality of Brown's work, noting acquisitions by many institutions and information that positioned the artist as 'the culmination of a history of modernist art production by European men'. The gallery was still able to gatekeep information from customers, for example the data about the relationship between Gagosian and a museum curator was unpublished. The gallery had also transmitted value through its historical narrative spun on Brown. This illustrates how art dealers, even while providing abundant online information, retain significant control over what buyers know, strategically curating details to shape perceived value and maintain informational asymmetries in the market.

Through looking at more art dealing activities of Gagosian and David Kordansky, artist representation by art dealers can influence the perception of the quality of the artworks, which leads to higher prices. The new exhibition style of online viewing rooms has reinforced dealers' role as gatekeepers as they are able to restrict access to information to the art market. The first two content sections have explored the art dealer's implication of market inefficiency, especially the lack of transparency so the third section will aim to focus on artist-level inequality.

Third content section

[73] Scott Indrisek, "For galleries, Online Viewing Rooms are about Storytelling and Sales", 2019. https://www.artsy.net/article/artsy-editorial-galleries-online-viewing-rooms-storytelling-sales
[74] Brown, K. Art markets, epistemic authority, and the institutional curation of knowledge, 2022. Cultural Studies, 37, 626 - 646. https://doi.org/10.1080/09502386.2022.2030777.

This section will use two reports to explore art dealers' role on artist-level inequality: Art Basel & UBS 2025[75]; Freelands Foundation in 2012-13[76] and in 2020[77]. This section will examine patterns of gender and racial representation, as well as sales concentration to investigate how artist-level inequality has changed over time. By taking into account both commercial and non-commercial galleries, the disparities between objectives will provide insight into how profits with social logistics can synthesize.

The Art Basel & UBS report found that sales in 2024 are becoming increasingly concentrated among top-performing artists, accounting for 56% of total dealer sales, making for a 3% increase in the past year. The largest growth in concentration occurred in the secondary market; the top three artists had substantially larger shares than in 2023, up from 6% to 53%. The opposite trend happened in the primary market, with the highest-selling artist accounting for 34% of shares in 2024, down by 2% year on year[78].

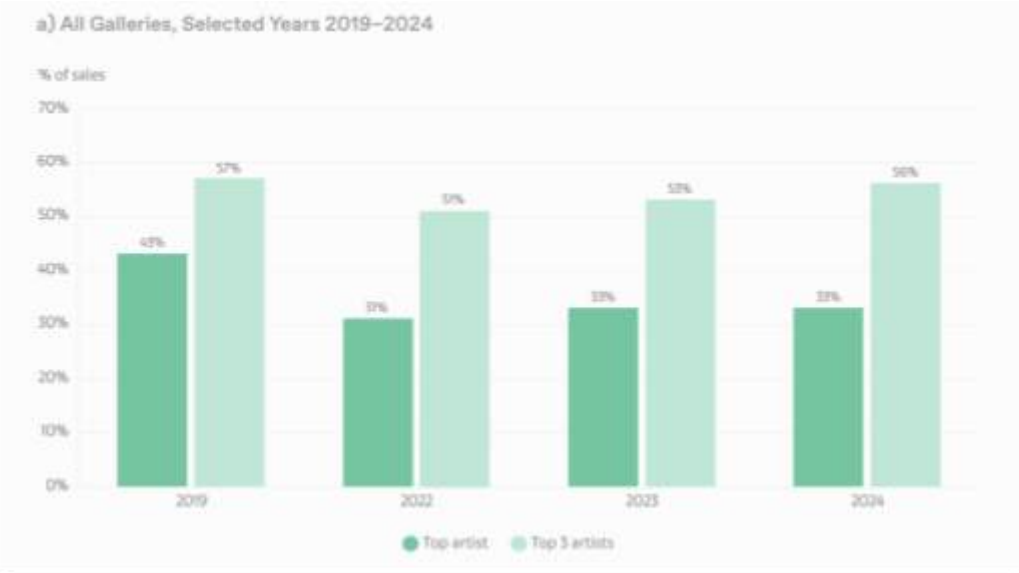


Fig 1. Sales of works by top-selling artists represented by dealers

[75] Clare McAndrew, "The Art Basel and UBS Art Market Report 2025", Art Economics. https://theartmarket.artbasel.com/download/The-Art-Basel-and-UBS-Art-Market-Report-2025.pdf

[76] Charlotte Bonham-Carter, "Representation of Female Artists in Britain", Freelands Foundation. URL: https://static.a-n.co.uk/wp-content/uploads/2017/11/Representation-of-female-artists-in-Britain.pdf

[77] Kate McMillan, "Representation of Women artists in Britain during 2020", Freelands Foundation. https://creativelandtrust.org/wp-content/uploads/2022/02/CLT_Representation-of-women-artists-2020.pdf

[78] Art Basel and UBS Report 2025, p.85

[79]

The significant increase in concentration in the secondary market could be explained through the rise in promotion of blue-chip artists by dealers. Because of the already existent high demand for these artists, it is consistent with dealer incentives to prioritise proven names in the secondary market. Therefore, demand inflates even more and becomes a cycle. The intensified focus on top artists makes the secondary market more inaccessible to other artists and creates artist-level inequality. However, the contradicting information in the primary market shows that emerging artists have ground to grow. Dealers and galleries are also introducing newer artists and could be favouring more community-focused practices than profit logistics.

The report also found that more female artists are being represented by dealers, but there is a persistent underrepresentation of women artists. The presence of women artists increased by 1% in 2024, and the share of dealer representation for these artists increased to 41% overall. Notably, in dealers with turnovers less than $250,000, women artists accounted for 49% of representation whilst for dealers with turnovers exceeding $10 million, women artists only shared 33%[80].

A negative correlation is found between turnovers and women artist representation; the more the turnovers, the lower the representation of women artists. Dealers, who are active in selecting artists for their roster, perpetrate this inequality. However, this data also highlights how bigger dealers may lean on longer auction records and institutional shows, which historically skew males. Auction records, which serve as information for present dealers, were controlled by past dealers and thus represent a structural problem in the art market.

In 2024, online sales were taking a bigger share of total dealer sales, accounting for 22% compared to 13% in 2019. The strongest set of growth came from dealer websites and online channels, which have more than doubled in share since pre-pandemic levels. Notably, the proportion of online sales to new buyers increased from 35% to 46%, showing growing accessibility to the market for newer generations[81]. More customers can access artworks through online platforms and artists who are represented benefit, yet artists with little to no visibility online do not see a change in their customer base.

With the rise of online sales, digital platforms such as dealer websites have become a prominent avenue for selling artwork. This shift has brought about a proliferation of available information, which could make the art market more transparent. However, knowledge curated on these websites by dealers represents their potential implication on the speculative price of artworks[82]. Mega art dealers want to raise speculative buying, therefore they might gatekeep

---

[79] Art Basel and UBS Report 2025, p.85
[80] Art Basel and UBS Report 2025, p.87
[81] Art Basel and UBS Report 2025, p.137
[82] Kathryn Brown, Art markets, epistemic authority, and the institutional curation of knowledge, Cultural Studies. 2022. https://www.researchgate.net/deref/https%3A%2F%2Fdoi.org%2F10.1080%2F09502386.2022.2

information to drive up prices. This act of gatekeeping only favours a few artists with a high reputation, supported by the increase in concentration of top artists in the secondary market (see above) and therefore reinforces artist-level inequality.

In 2012-13, East London Fawcett (ELF) conducted an audit of 134 commercial galleries in London in the Freelands Foundation report. It found that only 31% of artists represented were women and 78% of galleries represented more men than women. Whereas for solo shows of 28 different non-commercial galleries in London between 2014-15, 42% of shows were by female artists and 25% of shows at London's major institutions were by female artists[83].

The difference between women's representation between commercial and non-commercial galleries could be rooted down to the motives of art dealers. Commercial galleries typically aim to earn profits, and therefore to keep the prices of their artworks high have to represent artists with known market potential, often male artists. Non-commercial galleries want to promote artists hence will promote more diversified artists. In 2019, Richard Saltoun Gallery in London held a year-long exhibition programme called 100% Women focused on the art of female artists. Saltoun explained: "When you specialise in 'underrepresented', 'undervalued', and 'unknown' artists, this kind of *de facto* draws you into a mostly female arena…According to the 2018 Freelands Foundation's annual report, five percent of London galleries do so. This has prompted us to look at equality across our programme in greater depth, exhibiting the work of younger, more emerging artists, like Silvia Giambrone"[84]. This quote not only reverberates the underrepresentation of female artists but also illustrates how the motives of dealers or galleries affect their roster. Thus, dealers are at the forefront of deciding which artists get access to the art market, penetrating artist-level inequality.

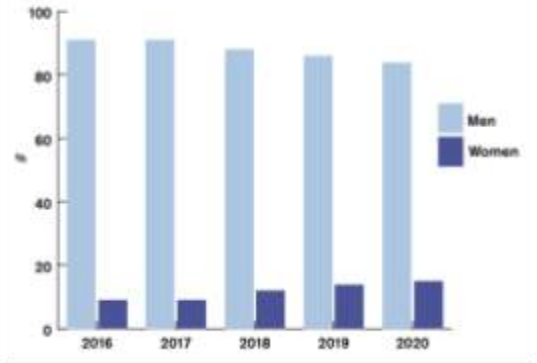


Fig 2. Sotheby's contemporary art evening sales in London 2016-20

030777?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19

[83] Freelands Foundation, Representation of Female Artists in Britain, p6

[84] Jareh Das, "Richard Saltoun and Catherine McCormack on 100% Women", Ocula, 2019. https://ocula.com/magazine/conversations/richard-saltoun-and-catherine-mccormack/

This paper will now look at the Freelands Foundation report in 2020 on the representation of female artists in Britain. In 2020, many art auctions were held online. Across all auctions, 151 lots were up for sale, of which 15% were by women artists. This represented an increase of 1% from the previous year[85]. This statistic reveals that digital auction platforms still favour male artists over female artists. Despite a slight increase in female artist representation, the proportion of these artists is still quite low.

One possible reason for this is because female artists were less represented in auction records in the past, there is less information for dealers to predict the future value. Thus consistent with dealers' motives, dealers choose to endorse male artists with longer auction records. This shows a structural problem within the art market as underrepresentation will lead to more underrepresentation, which would only be solved if galleries decide to take action or collectors have more information. [86]

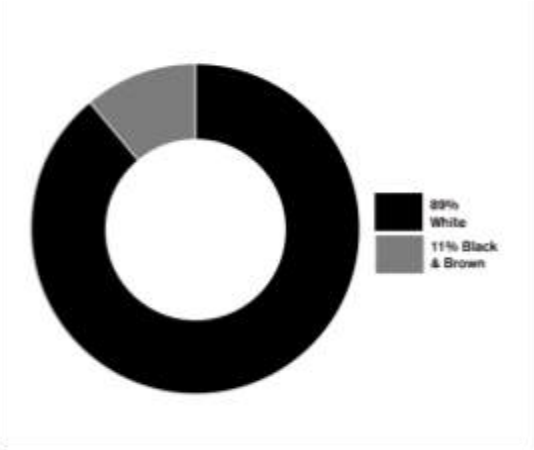


Fig 3. Sotheby's Contemporary Art Evening Sales in London 2020

For Sotheby's Contemporary Art Evening Sales in London 2020, 89% of artists were white and 11% were Black and Brown. Out of the 136 artists that took part, only 2 were Black and Brown women. An analysis of top ten grossing sales found that 3 sales were made by Jean-Michelle Basquiat, the solely featured Black and Brown artist. No Black and Brown women presented. Concentration of artists at the top end signals the type of paintings that will have higher value in the future and influence demand in the broader art market[87].

As the majority of high grossing sales are from white artists, it supports the trend that prices are more skewed to these artists, which highlight how art dealers are still incentivised to represent white artists in alignment with their profit objectives. Some dealers, facing

[85] Creative Land Trust, Representation of Women Artists, p66
[86] Creative Land Trust, Representation of Women Artists, p66
[87] Creative Land Trust, Representation of Women Artists, p67

performance pressure, may default to artists with longer resale histories, which have historically been skewed by race. Whilst Black and Brown are represented, their share of representation is limited - art dealers could be more hesitant towards endorsing artists with fewer auction records. This highlights the problem of transparency in the art market. The exception is Jean-Michel Basquiat's works; the significantly high prices of his works seem to raise prices only of his other paintings and the signalling effect on other artists is lesser known.

With regards to non-commercial galleries in London, 46% of solo exhibitions were by Black and Brown artists in 2020. For London's major institutions that receive more than $1 million in funding, 43% of solo exhibitions were by Black and Brown artists as opposed to 57% of solo exhibitions of institutions receiving between $500,000 and $1 million[88].

[89] The share of representation of Black and Brown artists is higher in non-commercial galleries. Non-commercial galleries might have different motives, other than earning profits, like endorsing the undervalued artists and this could lead to these galleries favouring less white artists. Institutional programming may move faster than the commercial secondary market, which further illustrates how different segments respond differently to representation goals. As more Black and Brown artists are showcased in solo exhibitions, more exposure increases customer base and seeds future demand; this brings about more exhibitions of these artists. Therefore, dealers can leverage institutional shows to control prices of artworks and artist-level inequality.

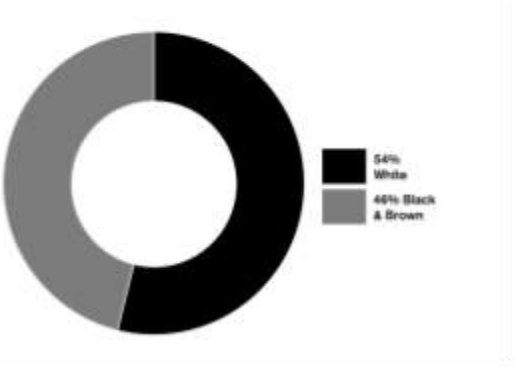


Fig 4. Solo exhibitions in non-commercial galleries in London

This section found that the top of the art market is concentrated around white male artists, which reflects the auction records in the past skewing these artists. There has been mild progress in women representation but institutional programming has led to non-commercial galleries representing more female artists. Disparities in representation between white artists and Black and Brown artists still persist. Data shows that online channels are expanding reach; even so, dealers are able to control information to raise the reputation of their top artist. To sum up, this

[88] Creative Land Trust, Representation of Women Artists, p39
[89] Creative Land Trust, Representation of Women Artists, p38

evidence implies that the act of choosing who to represent gives art dealers the ability to penetrate artist-level inequality and as they can gatekeep information, they play a role in reinforcing low transparency and unequal representation by controlling what information becomes visible in the art market through prices, placements, and selective disclosure.

Conclusion

This paper argues that art dealers influence pricing mainly through selective access and selective disclosure and that these practices weaken market transparency, create market inefficiencies, and widen unequal outcomes among artists. By putting art dealers at the center of this research paper, it illustrated the importance of intermediaries in information-sensitive markets like the art market.

Information asymmetry does not prevent commodification, but it reshapes it: art prices do not transparently reflect supply and demand because crucial information about transactions remains private. Moreover, prices function not only as economic indicators but also as signals of an artwork's perceived future value, supporting Velthuis's argument that prices carry symbolic meanings. Gorton's distinction between information-insensitive and information-sensitive assets helps clarify why dealer power matters in this market. In many financial settings, assets trade smoothly when participants do not need to ask questions, and panic follows when scrutiny suddenly becomes necessary.

In the art market, by contrast, scrutiny is always necessary but often impossible for outsiders, because key information about discounts, reserves, guarantees, prior placements, and unsold inventory is private. Dealers also restrict access to information, even with online platforms, to shape the price of art. Prices therefore do more than clear supply and demand; they stand in for missing information. When a small set of powerful dealers can shape those public price signals, access to visibility and credibility becomes uneven, creating a cumulative advantage for artists already represented in high-status networks while others remain outside the signal system that drives demand. Consequently, dealers also rely on prices to choose who to represent, perpetuating a cycle of underrepresentation of female artists and Black and Brown artists.

Based on evidence presented, this research finds that art dealers' control of information transforms price into economic tools which can be used to reinforce lack of transparency as well as artist-level inequality. Dealers can use this information to reflect in prices of artworks and this affects the buyers' perception of the artwork, which in turn can affect future demand, winding up in a cycle that makes little progress to benefit the underrepresented artists.

These findings contribute to past literature by expanding upon the many functions of art dealers. By showing that prices play a significant role in signalling to buyers, this paper has indicated that art dealing is a complex activity and what art dealers do depends on their motives. The exploration of economic concepts including prices, lack of transparency and inequality provide a new framework through which the art market can be analysed. For future research, a more sociological framework can be applied to assess the many motives of art dealers or a feminist aspect can be incorporated to deepen understanding of inequality towards female artists.

Because of limited insider knowledge into art dealing as much information is restricted, it would benefit if insights can be drawn from an active dealer. A practical next step would be five to eight semi-structured interviews with mid-size dealers to explore the range of motives (for instance, how list prices, discounts, and reserves are set and communicated); interviews that could be conducted anonymously if need be. Furthermore, limited research has been done on online viewing rooms due to its recent development so further research can assess their impact on transparency. By showing that art dealers play a significant role in determining prices of art, this paper tries to clarify the functions of art dealers and implies that collectors should be less dependent on headline prices and dealer signals when appraising art. Ultimately, increasing transparency and reducing artist underrepresentation will depend not only on institutional reform within the dealer sectors but also on more informed engagement by the collectors.

Works Cited

Orley Ashenfelter and Kathryn Graddy, “Art Auctions,” in: Handbook of the Economics of Art and Culture, 2006; Luc Renneboog and Christophe Spaenjers, “Buying Beauty,” Management Science, 2013.

Olav Velthuis, “Talking prices: Symbolic meanings of prices on the market for contemporary art,” Princeton University Press, 2005.

Dakshina G. De Silva et al., “Evolution of a dealer trading network and its effects on art auction prices,” European Economic Review, 2022.

Fabio Rojas. David de la Torre, “Art dealers as competitors and community builders: Understanding the overlapping social logics of the contemporary art world”, 2023.

Edward Winkleman, “How to run a gallery: three threats you can’t ignore”, 2015. https://www.theartnewspaper.com/2015/10/01/how-to-run-a-gallery-three-threats-you-can-t-ignore

Nate Freeman, “Why Leo Castelli Paid his Artists Even When They Weren’t Making Work”, 2018. https://www.artsy.net/article/artsy-editorial-leo-castelli-changed-art-market-forever

Grace Glueck, “How do you price art works? Let the dealers count the way”, 1976.

Anthony Mason, “Gallery owner Larry Gagosian on the art world’s “blood sport”, 2025. https://www.cbsnews.com/news/gallery-owner-larry-gagosian-on-the-art-worlds-blood-sport/

Robert Armstrong, “‘I like the ways things look.’ The world according to Larry Gagosian”, 2022. https://www.ft.com/content/eefe7a73-7798-4e56-b8f9-70c35a9e68e9

Cathleen McGuigan, “New art, new money”, 1985.

“The Art World: What If…?!, Season 2, Episode 8: Barbara Gladstone,” podcast audio, The Art World: What If…?!, hosted by Charlotte Burns, March 14, 2024, Art& (Schwartzman&), https://www.schwartzmanand.com/art-posts-the-art-world/the-art-world-s2ep8-barbara-gladstone

Schooley, D., & Worden, D. 2016. Perceived and Realized Risk Tolerance: Changes During the 2008 Financial Crisis. Journal of Financial Counseling and Planning, 27, 265 - 276. https://doi.org/10.1891/1052-3073.27.2.265.

Brown, K. (2023). Art markets, epistemic authority, and the institutional curation of knowledge. Cultural Studies, 37(4), 626–646. https://doi.org/10.1080/09502386.2022.2030777

Iorgova, Silvia, and Chase P. Ross. 2023. “Investor Information and Bank Instability during the European Debt Crisis.” Journal of Financial Stability 64 (February): 101100. https://doi.org/10.1016/j.jfs.2022.101100.

Scott Indrisek, “For galleries, Online Viewing Rooms are about Storytelling and Sales”, 2019. https://www.artsy.net/article/artsy-editorial-galleries-online-viewing-rooms-storytelling-sales

Clare McAndrew, “The Art Basel and UBS Art Market Report 2025”, Art Economics. https://theartmarket.artbasel.com/download/The-Art-Basel-and-UBS-Art-Market-Report-2025.pdf

Charlotte Bonham-Carter, “Representation of Female Artists in Britain”, Freelands Foundation. https://static.a-n.co.uk/wp-content/uploads/2017/11/Representation-of-female-artists-in-Britain.pdf

Kate McMillan, “Representation of Women artists in Britain during 2020”, Freelands Foundation. https://creativelandtrust.org/wp-content/uploads/2022/02/CLT_Representation-of-women-artists-2020.pdf

Kathryn Brown, Art markets, epistemic authority, and the institutional curation of knowledge, Cultural Studies, 2022. https://www.researchgate.net/deref/https%3A%2F%2Fdoi.org%2F10.1080%2F09502386.2022.2030777?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19

Jareh Das, “Richard Saltoun and Catherine McCormack on 100% Women”, Ocula, 2019. https://ocula.com/magazine/conversations/richard-saltoun-and-catherine-mccormack/

Cell-Type Gene Expression as a Predictor of Opioid Use Disorder in the Human Dorsal Striatum By Arjun Thakur

Abstract

Opioid Use Disorder (OUD) is a major public health crisis wherein brain reward neurological circuitry is rewired. Recent single-nucleus RNA sequencing studies have identified changes in multiple brain cell types, but it still remains unclear whether these cell-type gene expression patterns can accurately predict OUD status. Previous studies have largely focused on different gene expression rather than identifying cell types associated with OUD. In this study, I analyzed a publicly available snRNA-seq dataset of 98,848 nuclei from the human dorsal striatum of 12 donors (6 OUD, 6 controls). I used logistic regression models that were trained separately on major cell types using donor-level stratified cross-validation to prevent leakage of donor-specific transcriptional signatures between training and testing data. These findings suggest that changes associated with OUD are concentrated in specific neuronal populations and highlight the importance of rigorous validation strategies for identifying biologically meaningful disease signals.

Introduction

Opioid use disorder (OUD) is a major public health crisis. OUD is the leading cause of accidental death in the United States, causing roughly 70,000 deaths annually (Wei et al., 2023). In the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)*, OUD is found as a pattern of opioid use that results in impairment or distress. Symptoms of OUD include difficulty controlling consumption, risky opioid use, disruptions to social functioning, and the build-up of physiological dependence. OUD can have severe consequences for an individual's health, such as an increased risk of overdose (Poulson et al., 2022). Opioids are medications commonly prescribed by physicians to manage moderate to severe pain. They are used to treat patients with chronic pain, such as persistent headaches or back pain, as well as individuals with injuries that produce significant pain (ASAHQ, 2026). Opioid compounds can be put into two main categories: phenanthrenes and benzylisoquinolines. The phenanthrene group includes substances such as morphine, codeine, and thebaine, while benzylisoquinolines include substances like papaverine and noscapine (Silberstein, 2000). Opioid drugs work by binding to specific proteins known as opioid receptors located in nerve cells throughout the body. When these receptors are activated, they reduce pain signals from the body to the brain, helping relieve pain. Because of the possibilities of feeling pleasure, opioid medications run the risk of a high potential for dependence and can contribute to the development of opioid use disorder (Johns Hopkins, 2025).

These psychological and behavioral effects of OUD are caused by significant molecular and cellular changes in the brain, mainly in regions involved in reward processing, motivation, and addiction. Consistent opioid exposure changes gene expression and cellular signaling in neural pathways. Several recent studies using single-nucleus RNA sequencing have begun to

characterize these changes at the base of individual brain cell types, revealing interactions between neurons and glial cells in patient brains. Recent single-nucleus RNA-seq studies in human brains have shown major glial activation in opioid overdose cases (Saad et al., 2019). The dorsal striatum, which plays a major role in reward learning and habit formation, is strongly affected in OUD patients. One recent snRNA-seq study analyzed postmortem dorsal striatum tissue, containing multiple cell types, from individuals with OUD. In particular, neurons and microglia displayed a significant increase in activity of pathways associated with inflammation and DNA-damage responses. These findings show that consistent opioid exposure leads to cellular stress within specific neural populations. (Phan et al., 2024). For example, a ventral midbrain (VM) study found that 9.5% of microglia were not properly regulated in overdose victims. Similarly, in the dorsal striatum – a region key for reward, habit formation, and withdrawal effect – snRNA-seq of 12 individuals (6 OUD vs 6 controls) identified inflammatory and DNA-damage pathways in neurons and microglia.

Despite these advances in identifying cell-type changes, previous studies have focused more on identifying changed genes and pathways in the aggregate rather than investigating whether specific cell-type gene expression patterns can predict OUD. Therefore, research is needed to determine whether gene expression signatures from specific brain cell types can be used to accurately classify individuals with OUD. Addressing this gap could provide new insights into the cellular mechanisms involved with addiction and help identify potential biomarkers for opioid use disorder.

However, a major challenge to this type of approach is donor-level data leakage. Assigning single cells from the same donor into training and test sets can cause a leakage of donor-specific patterns, essentially allowing the machine learning (ML) model to "cheat" when making its predictions. 3 To avoid this, training must use stratified cross-validation. Ultimately, the goal of this project is to use donor-level cross-validation for classifying OUD status from gene expression signatures of different cell types, allowing identification of which cell types carry generalizable OUD signals.

Methods

The analysis pipeline is shown in Figure 1. I used the publicly available human dorsal striatum snRNA-seq dataset (Phan et al., 2024). This comprises 98,848 nuclei from 12 donors (6 OUD, 6 control), categorized into major cell types including indirect/ direct medium spiny neurons, interneurons and astrocytes (see Figure 2 for the complete list). To verify the quality of the dataset, I used uniform manifold approximation and projection (UMAP) to validate cell-type clustering (Figure 2). The nuclei were segregated into clusters by the authors of the original paper. The authors completed this clustering by establishing marker genes, and the resulting cell labels were put into the UMAP purely for visualization of the data. Ultimately, the UMAP serves primarily as a representation of each cell type rather than the exact method used to determine them. Furthermore, the number of clusters seen in the figure were determined by the authors of the original study during their cluster analysis. Then, logistic regression was used on each cell

type to determine its status in predicting OUD. To prevent donor leakage I utilized donor-level cross-validation to ensure the model did not produce artificially high results by learning donor-specific patterns. Performance of each cell type was determined by AUC. This ultimately allowed us to compare all the predictive performances across cell types.

First, I downloaded the h5ad file from the public database and used Jupyter Lab for analysis. After the initial clustering shown in Figure 2, I used StratifiedKFold on the cells into 5 folds, making sure each fold had a balanced mix of OUD/control cells. This essentially creates a random split of cells, but it allows cells from the same donor in both train and test, risking donor leakage. I constructed donor labels for each cell type and applied a 5-fold StratifiedKFold on donors, so that whole donors were held out in each test fold. This ensures no donor's cells appear in both train and test. In each fold I computed the characteristic Area Under the Curve (AUC) on the cells, then averaged across folds to get mean plus minus SD AUC for each cell type.

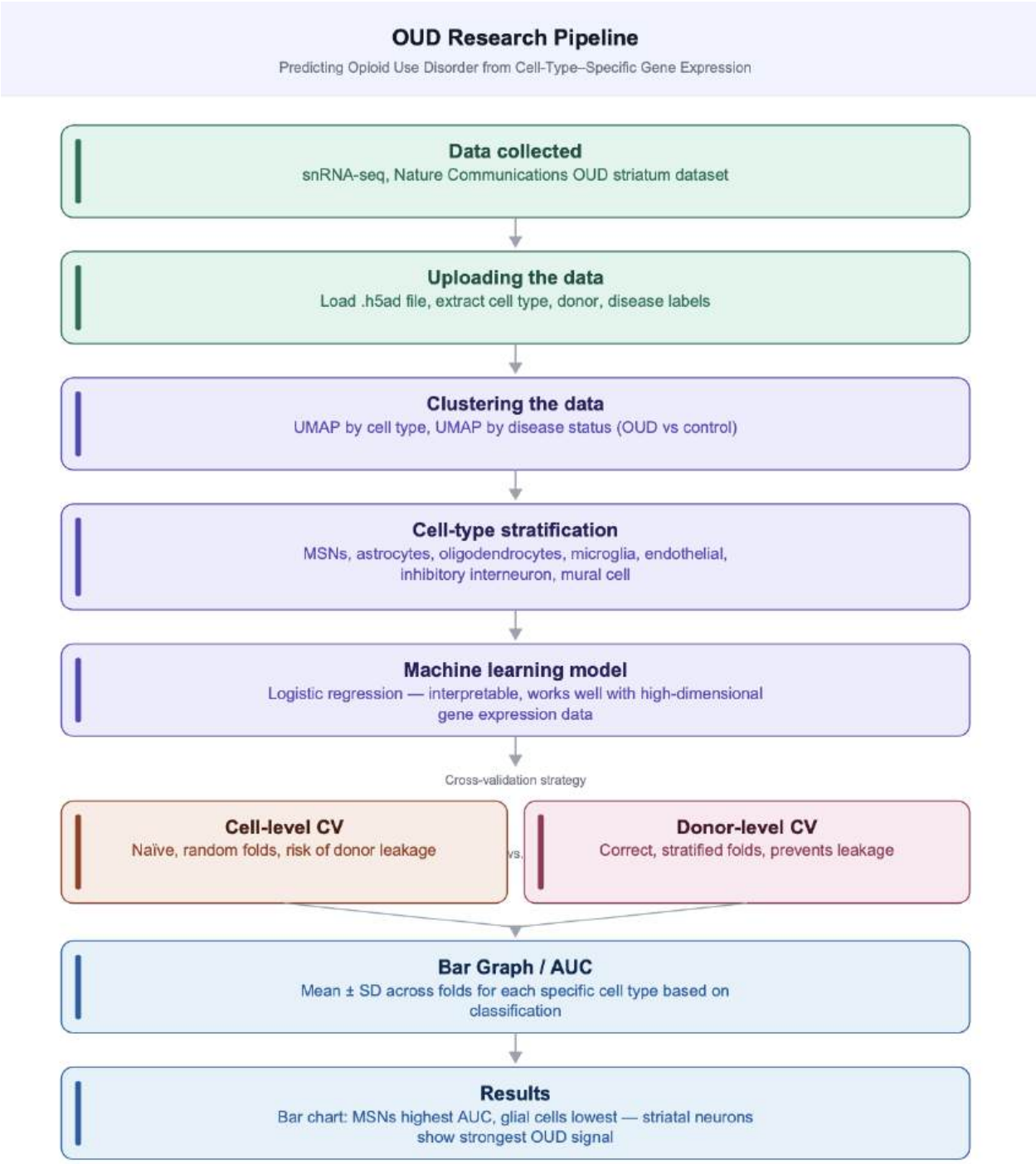


Figure 1: Pipeline for predicting OUD from cell-type gene expression in human dorsal striatum

Results

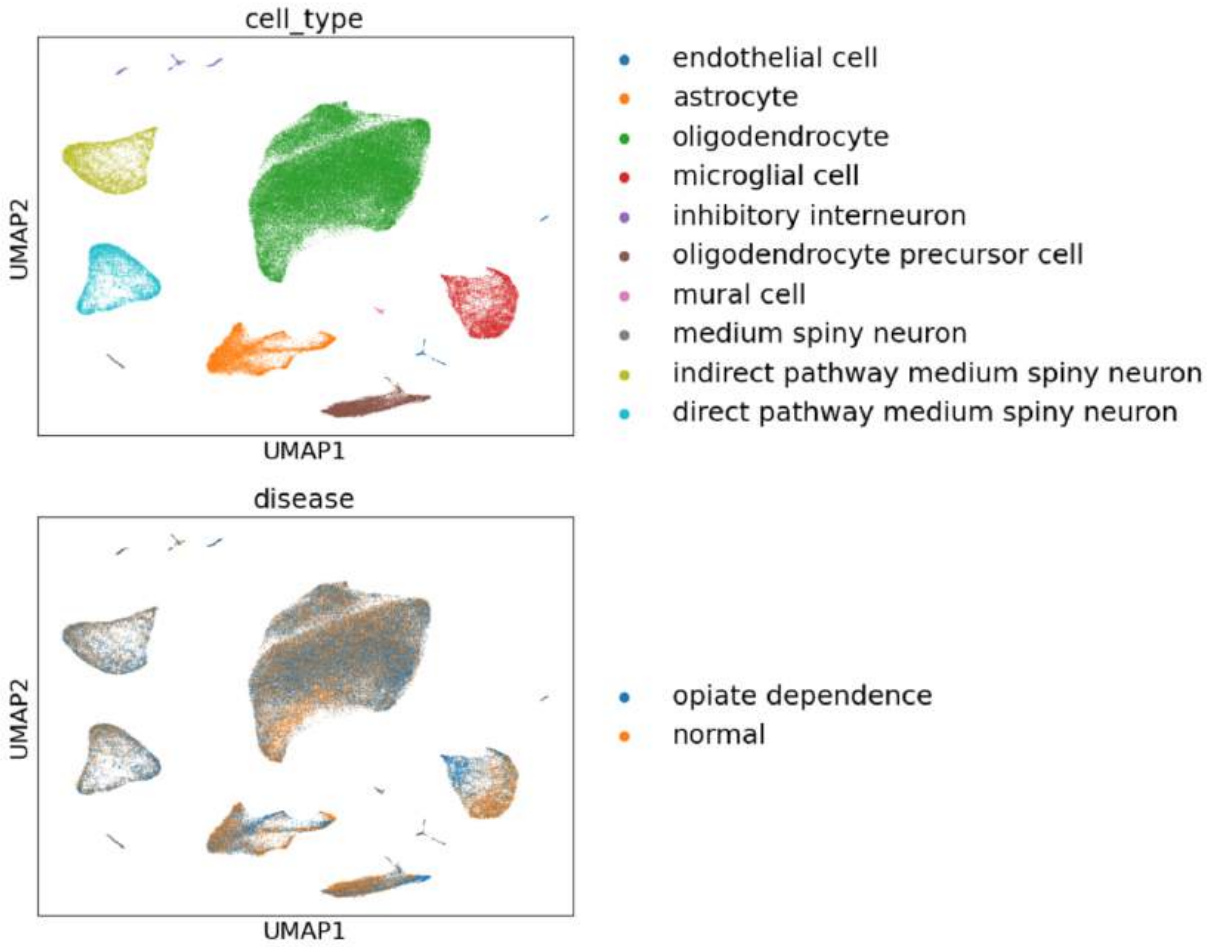


Figure 2: Cell-type distribution and disease distribution in single-nucleus data from the human dorsal striatum.

Each point in Figure 2 represents an individual nucleus based on gene expression similarity. Clusters are colored by the most represented cell type, including oligodendrocytes, astrocytes, microglia, inhibitory interneurons, endothelial cells, mural cells, oligodendrocyte precursor cells, and medium spiny neurons. On the bottom, the cells are colored by disease status. Cells from individuals with opioid use disorder and without opioid use disorder are distributed across all major cell-type clusters, showing that both control and OUD samples contribute to each population. The UMAP clustering shows that the OUD and control patients have cells distributed throughout all the clusters rather than forming separate disease clusters. This provides evidence that OUD does not cause large shifts in cell-type change but rather subtle changes that affect transcriptional changes.

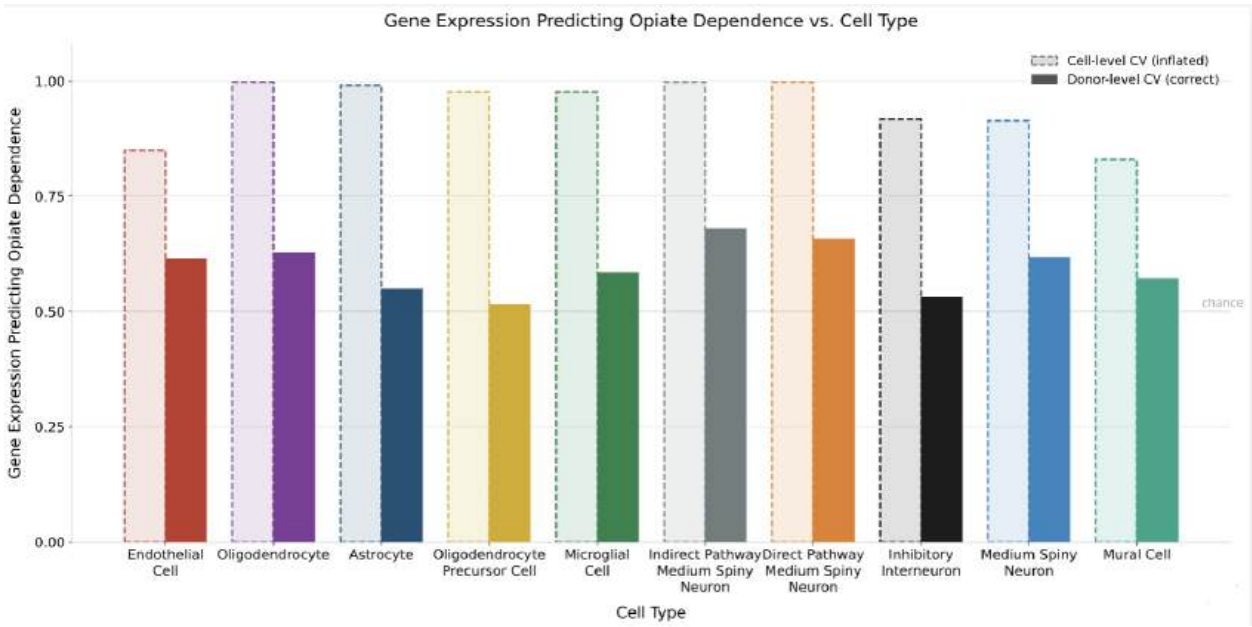


Figure 3: Cell-type predictive performance for opioid use disorder classification.

Figure 3 compares classification performance through AUC across major striatal cell types when predicting opioid use disorder status from gene expression profiles. Each bar represents the predictive performance of a machine learning classifier trained on gene expression data from a single cell type. Performance in this figure is based on gene expression patterns to distinguish nuclei from OUD donors versus control donors. Neuronal populations, particularly direct and indirect pathway medium spiny neurons, show the highest predictive performance, suggesting that changes in transcription in these cell types may contain stronger disease-associated signals. Glial cell types such as astrocytes, oligodendrocytes, endothelial cells, and mural cells also have predictive signal capabilities, but at lower levels. These results highlight cell-type differences in transcription patterns that are associated with opioid use disorder.

A critical part in the analysis of cell-type's predictive performance of OUD, is the use of donor-level cross-validation over cell-level cross validation. Donor-level cross-validation segregates all of the donor's cells into either the training or testing group. This is key in preventing donor-leakage, a scenario where models may inadvertently learn donor-specific expression patterns rather than true disease-associated signals. This would ultimately result in artificially inflated performance as the model would be able to compromise the results to skew in their favor, as can be seen in the comparisons in Figure 3. When cell-level cross-validation is used, the model is able to recognize donors it has already seen and find patterns that cannot be generalized to the general population. This is because cells from the same individual share similar transcriptional profiles, resulting in overestimated predictive accuracy. In contrast, the donor-level cross-validation enforces a strict separation to ensure all of a donor's cells are in

either the training or testing group. This leads to more conservative but more biologically meaningful performance estimates.

To determine if I could predict whether or not an individual had OUD through gene expression within specific cell types, I constructed a logistic regression model for each specific cell type. There was a substantial variation between the cell types' ability to predict for OUD, showing that transcriptional changes are not uniformly distributed throughout the brain, but limited to only certain cell types. Among all the cell types analyzed I found that the indirect pathway medium spiny neurons exhibited the strongest relation to predicting OUD, with a mean AUC of 0.68. This cell-type was closely followed by the direct pathway medium spiny neuron with a mean AUC of 0.66. These results strongly suggest that transcription alterations from OUD patients are most strongly pronounced in striatal neurons, which play a central role in reward processing and habit formation.

Comparing the results of cell-level cross-validation which produced artificially high AUCs for many cell types against donor-level cross-validation. The cell-level cross-validation approach produced AUCs greater than 0.9 (Figure 3), in some neuronal subtypes, because the model could recognize donor-specific expression patterns caused by donor leakage. In contrast, donor-level cross-validation resulted in much lower AUCs overall. For most cell types, mean AUC dropped close to 0.50 which is not necessarily useful data, proving that cell-level cross-validation was largely artificial. Under the donor-level CV, only a few cell types had predictive power. Our prime example is indirect/direct-pathway MSNs which still showed a mean AUC around 0.68, whereas performance in other cell types fell closer to 0.50–0.55. These results expose that a lot of the apparent AUC in cell-level Cross Validation was due to donor leakage, and only a subset of cell types (MSN's) carry OUD associated differences that can be generalized across individuals. Most importantly, we find that having a performance score of 0.5 is equivalent to the performance one would achieve by random chance, so it's conclusive to their predictive performance, that only certain cell-types have performance scores far above 0.5.

Discussion

The purpose of this study was to identify whether or not specific cell-types can best predict OUD in an individual. The ultimate goal was to identify which brain cell population in the dorsal striatum contains the most generalizable gene expression associated with OUD. Overall, the results of the study demonstrate that gene expression contains meaningful information that has the ability to distinguish OUD from control. Indirect Pathway Medium Spiny Neurons and Direct Pathway Medium Spiny Neurons are the highest in predictive power across all the cell sub-types. Both of these have a 0.66-0.68 chance of predictive performance when calculated using AUROC. The consistently higher performance observed in medium spiny neurons suggests that these cells are central to the molecular pathology of OUD. These findings support and are consistent with the role the dorsal striatum plays in reward processing and habit formation. Furthermore, MSNs are the primary target of opioid action in the reward circuit processing of the dorsal striatum. In contrast, cell types such as oligodendrocytes, endothelial

cells, and microglia showed moderate predictive performance, while astrocytes, oligodendrocyte precursor cells, and inhibitory interneurons displayed even weaker signals.

Though these findings are promising, there are limitations to the approach outlined here. To start off, the sample size is low. In this research, I used a dataset that consists only of 12 patients (6 OUD, 6 Control), which is a relatively small sample size when it comes to machine learning analysis. Ultimately, this limitation restricts the ability of this experiment from becoming widely generalizable. Another limitation is the use of one classification model. Throughout this project I used a logistic regression and nothing else. Therefore, I could be limited to certain patterns only caught by a logistic regression. Finally, this analysis was strictly limited to only one part of the brain, the dorsal striatum, resulting in a possibility of other cell types in other parts of the brain that can predict opioid dependence.

Future research should aim to fill this gap by expanding this dataset to include a larger and maybe more diverse group of individuals. Additionally, reproducibility across separate, independent datasets will be especially critical for determining if the results are generalizable. Furthermore, it should also include the research of specific genes within a cell type that are related to OUD. Such findings could improve and provide insights to the biological mechanisms of addiction.

Works Cited

American Society of Anesthesiologists. (2026). Opioid treatment. Made for This Moment. https://madeforthismoment.asahq.org/pain-management/opioid-treatment/what-are-opioids/

Feng, H., Elladki, R., Jiang, J., & Wei, G.-W. (2023). Machine-learning analysis of opioid use disorder informed by MOR, DOR, KOR, NOR and ZOR-based interactome networks. Computers in Biology and Medicine, 157, 106745. https://doi.org/10.1016/j.compbiomed.2023.106745

Johns Hopkins Medicine. (2025). Opioids. https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/opioids

Phan, B. N., Ray, M. H., Xue, X., Fu, C., Fenster, R. J., Kohut, S. J., Bergman, J., Haber, S. N., McCullough, K. M., Fish, M. K., Glausier, J. R., Su, Q., Tipton, A. E., Lewis, D. A., Freyberg, Z., Tseng, G. C., Russek, S. J., Alekseyev, Y., Seney, M. L., Pfenning, A. R., & Logan, R. W. (2024). Single nuclei transcriptomics in human and non-human primate striatum in opioid use disorder. Nature Communications, 15(1), 878. https://doi.org/10.1038/s41467-024-45165-7

Poulsen, M. N., Freda, P. J., Troiani, V., Davoudi, A., & Mowery, D. L. (2022). Classifying characteristics of opioid use disorder from hospital discharge summaries using natural language processing. Frontiers in Public Health, 10, 850619. https://doi.org/10.3389/fpubh.2022.850619

Saad, M. H., Rumschlag, M., Guerra, M. H., Savonen, C. L., Jaster, A. M., Olson, P. D., Schmidt, C. J., & Bannon, M. J. (2019). Differentially expressed gene networks, biomarkers, long noncoding RNAs, and shared responses with cocaine identified in the midbrains of human opioid abusers. Scientific Reports, 9, 1534. https://doi.org/10.1038/s41598-018-38209-8

Silberstein, S. D. (2000). Review of the pharmacology of opioids. Cephalalgia, 20(Suppl. 2), 1–9. https://doi.org/10.1046/j.1468-2982.2000.00149.x

Analyzing Public School Sports: Pay to Play or a Level Playing Field? By James Scott

Introduction

The idea of a level playing field is a central concept in all athletic competition. However, the cost of developing elite athletes may be influencing the chance of success for students. Public high school sports are designed to give everyone an equal opportunity to play and succeed. Money should not matter when athletes are simply playing for their school's team. Is that really the case, though? In US public schools, do the best athletes win, or can money buy success?

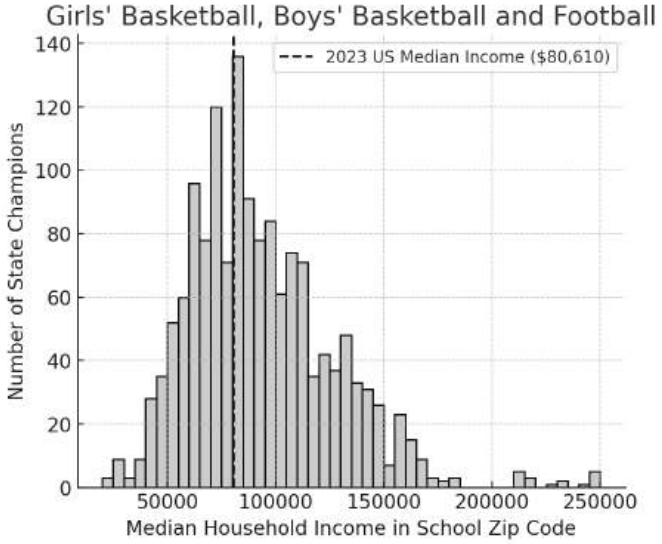
The Question

Are high school state champions located in zip codes with household incomes above the state average? Does this effect differ between girls' and boys' sports, or between football and basketball?

About the data

The Nighthawk Times identified the last ten public school champions of three different sports for every state in America, using each state's high school athletic association's public webpage. We examined the largest division of boys' basketball, girls' basketball and football in each state. For each champion, we found the school's zip code. Using 2023 American Community Survey data, we identified median household income by zip code (Guzman and Kollar, table A-4a), and used the Current Population Survey to map these zip codes onto national income deciles (U.S. Census Bureau). We built spreadsheets and analyzed data in Excel.

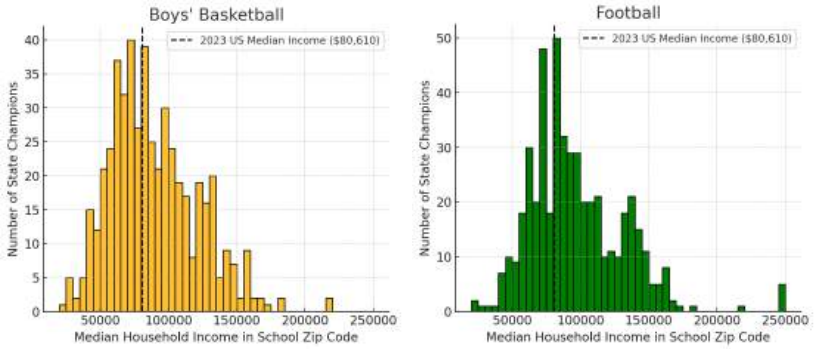
Results



The median household income in zip codes of state champion schools from our dataset was $86,945.50, slightly higher than the nationwide median of $80,610. The data were close to normally distributed, meaning that the spread of incomes of households in champion school zip codes was similar to incomes across the country as a whole. However, the data were skewed in the highest income brackets. There were more champion schools in the wealthiest zip codes than would be expected in a perfectly normal distribution.

These data imply that American public schools provide a relatively fair playing field for sports. Ezekiel Ocansey, an assistant principal and the athletic director for Northfield High School, spoke about how his school helps students access the equipment they need to participate in their sports. He said, “We’ll get items donated… just any equipment, we’re willing to go ahead and go out to the community and receive those donations… We couple that with being able to just purchase [equipment] or fundraising the items needed for any families or any athletes that want to participate in any sports that just happen to cost extra.”

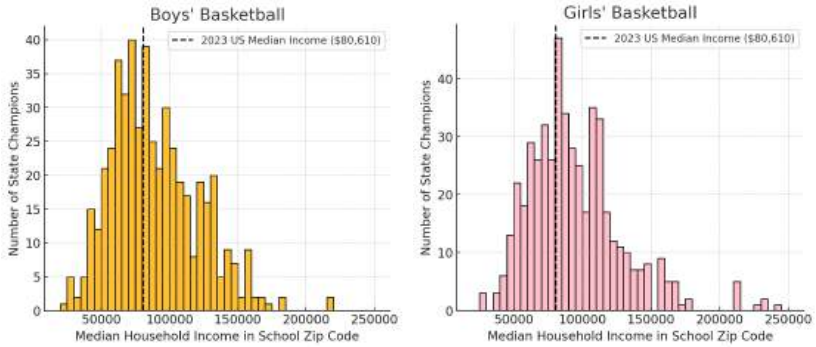
Boys’ Basketball vs. Football



When the data were broken down by sport, the distributions were similar. In both basketball and football, most champion schools were close to the national household median income. However, football champion schools were located in zip codes with a median household income of $87,054, higher than the median household income of $83,383 in boys' basketball champion zip codes. Football also had more outliers in the highest income bracket.

Football is generally a more expensive sport than basketball, and student-athletes rely more on their school for opportunities to play and improve. When comparing the effects of family income on the two sports, Head Coach Ben Startzer of Northfield’s football team said, “Sometimes it can just be the price of access. If you’re going to play football, let's just even say recreationally, tackle football is going to take some resources… that’s going to be a lot of money where maybe basketball is going to require a short set and a basketball.”

Boys vs. Girls’ Basketball



State champions in girls' basketball had the highest median household income of all three sports, at $88,700.75. Just like the other sports, the data for girls' basketball champions were fairly normally distributed around this median.

Participation in girls' basketball may have driven higher champion school incomes than the boys' sports in this analysis. According to Ethan Fuller of the *New England Basketball Journal*, as of 2024, participation numbers for girls' basketball have not shown the same recovery from the COVID-19 pandemic as other high school sports ("Girls Basketball Participation is Declining"; National Federation of State High School Associations) . With fewer players interested in playing, larger schools have an advantage. These larger, often suburban, schools have more potential players to choose from and more time to craft and pick the best players.

Gold Crown Foundation is a non-profit organization in Denver that offers after-school programming, basketball, volleyball and golf camps and tournaments. They focus on low-income support and affordability. Kevin Petty, their president and executive director, outlined how certain schools gain an advantage from this effect. "For example, a school in Highlands Ranch that has thousands of kids is able to hand-pick and groom the better players for a longer period of time, whereas in the more urban schools, where there are fewer kids participating, by the time [players] get to high school, [the schools] don't have as many kids to pick from." This effect is amplified when there are fewer total participants. Because there are more students across the country playing boys' basketball, more schools will have enough players to run competitive, successful programs, regardless of income.

Programs that Improve Access to Sports

In light of these national statistics, we examined programs that are working to address access to youth athletic training and opportunities here in Denver.

Denver Public Schools (DPS) increased opportunity by creating the Prep League for middle schools, with support from the DPS Foundation. Starting in 1997, this league gave students access to organized teams and structured, official competition in the district at a younger age. Ashley Muramoto, the Vice President of Marketing and Communications for the DPS Foundation, said, "By offering access to high-quality programs in middle school, you can help

close opportunity gaps before they get even wider. The sooner you do that, the better. That helps students achieve more in high school."

Since 2023, the My Spark Denver program has aimed to improve access to after-school activities, including sports. This program gives eligible families a restricted card with $1,000 so middle school students can pursue interests outside the classroom. Like the Prep League, My Spark focuses on opportunities before high school. Muramoto said, "The goal of [My Spark] is to give students something to do during those middle school years after the school day ends, and it is a great way for them to stay involved."

The Gold Crown Foundation is another organization working to provide young athletes with opportunities to participate in sports. Their non-profit status helps them keep fees low for all families, and their scholarship fund is available to families who would otherwise not be able to afford to play. Gold Crown is able to keep sports accessible and in reach for everyone. Kevin Petty said, "We try to work with families to meet them wherever their need is."

Private organizations work to provide opportunities for athletes with financial challenges, while balancing the need to fund programs and facilities. Redline Denver East is a practice facility for youth athletes to train for basketball, volleyball and other sports. It was created by Aaron Bradford to provide serious and competitive training. Bradford uses the facility for his travel basketball team. Running the facility and team requires parents to pay enough to keep the team going. To address financial barriers, Bradford created the Arctic Seagulls Foundation and works with other scholarships and programs, such as MySpark. Bradford said, "The first thing we do is we make it clear that we will help any kid that wants to train. We'll find the money for them to train."

Limitations

The focus on state champions did not provide a full picture of all aspects of youth sports. It made it feasible to look at every state, but we did not examine participation or successful teams that finished just below first place. Examining the top division in each state meant that this data set did not include every successful team, and emphasized the largest schools. Trends may be different in smaller divisions.

Conclusion

Student-athletes at public schools in zip codes across income brackets have an opportunity to be a champion. There were slightly more champions in wealthy zip codes, particularly in girls' basketball and football, but every sport examined had champions from zip codes with a wide range of incomes. When sports are more accessible and have high participation, like boys' basketball, the effect of household income on athletic success is mitigated. Public schools, as well as non-profit and private organizations, all have programs designed to continue the work towards a more level playing field.

Works Cited

Basketball Journal. “Girls basketball participation is declining — let’s unpack the data.” New England Basketball Journal, 18 Sept. 2024, https://www.basketballjournal.com/2024/09/18/girls-basketball-participation-is-declining-lets-unpack-the-data/. Accessed 13 June 2026.

Guzman, Gloria, and Melissa Kollar. Income in the United States: 2023. Report P60-282, U.S. Census Bureau, 2024, https://www.census.gov/library/publications/2024/demo/p60-282.html. Accessed 13 June 2026.

National Federation of State High School Associations. “High School Participation Survey Archive.” NFHS, n.d., https://nfhs.org/resources/sports/high-school-participation-survey-archive. Accessed 13 June 2026.

U.S. Census Bureau. “Decile data from 2023 – Median income data.” data.census.gov, n.d., https://data.census.gov/all?q=median+income. Accessed 8 November 2025.

The Promise and Peril of Personalized Genome Language Models(pGLM) in Precision Medicine By Sreenayan Palle [1]

Abstract

Large language models (LLMs) trained on DNA and protein sequences are beginning to interpret individual genomes at scale, treating nucleotides like “text” to learn the grammar of biology. Architectures such as HyenaDNA now extend single-nucleotide context windows to about one million base pairs, enabling models to capture long range regulatory interactions[1]. This review synthesizes peer reviewed studies published between 2021 and 2025 on genomic language models and their applications in precision medicine, examining both their technological progress and ethical implications. Variant classifiers like AlphaMissense provide proteome wide predictions, labeling roughly 89 percent of human missense substitutions as likely benign or pathogenic[2]. These models show promise in variant effect prediction, regulatory element discovery, and early steps toward personalized therapeutics. Yet risks persist: potential privacy leaks from memorized sequences[7], bias from ancestry skewed datasets[7], opaque predictions that may be over trusted clinically[8], and dual use concerns the possibility that generative biology tools could be misapplied for harm[9]. Realizing clinical value will require validation on diverse cohorts, interpretable and uncertainty aware outputs, privacy preserving training, and clear regulation[7]. This review highlights the opportunities and cautions of integrating genome language models into precision medicine responsibly.

Introduction

Genome sequencing has produced immense catalogs of human variation, creating demand for methods that transform sequence data into clinical insight. Genomic large language models approach DNA and proteins as a language, learning motifs and dependencies that underlie gene regulation and variant effects[1]. These “foundation models for genomics” can be adapted to tasks central to precision medicine, an approach that tailors care to the individual based on molecular and environmental profiles.

Early exemplars such as HyenaDNA, Evo2, and AlphaMissense demonstrate that self supervised pretraining on massive genomic datasets yields strong performance in variant interpretation and regulatory prediction[1–6]. However, integrating such models into healthcare introduces concerns around privacy, bias, and accountability[7–9]. This review surveys the technical landscape, clinical promise, and policy perils of personalized genome language models (pGLMs) and proposes safeguards for their ethical and effective use.

[1] Dougherty Valley High School, San Ramon, California, United States Of America
E-mail: srnpalle@gmail.com (SP)

Funding: The author received no specific funding for this work.
Conflicts of interest: All authors declare no conflict of interest.

Methods

A narrative review of peer-reviewed literature (2021–2025) was conducted through PubMed and leading journals (*Nature*, *Science*, *PNAS*, *Human Genetics*, *Nature Biotechnology*). Search terms included "genome language model," "DNA LLM," "variant effect prediction," "regulatory element modeling," and "AI privacy genomics." Only peer-reviewed or formally archived works were included.

Results

1. Architectural advances enabling longrange modeling

Traditional transformers scale poorly with sequence length. HyenaDNA replaces attention with implicit long convolutions, allowing million base pair context at single-nucleotide resolution[6]. Hybrid systems like Evo interleave these layers with sparse attention; Evo2 scales training to 128,000 genomes and trillions of bases, supporting both sequence analysis and controlled generation[5]. Such designs now capture enhancer promoter relationships and other distal interactions beyond earlier models' reach[6].

2. Variant effect prediction in coding and noncoding DNA

Protein focused models such as AlphaMissense integrate structural and evolutionary data to classify missense variants across the human proteome, providing near comprehensive prior knowledge for clinical interpretation[2]. DNA based LMs trained with masked token objectives learn sequence "grammar" correlating with functional impact and often surpass conservation metrics in benchmark tests[3]. Multispecies and alignment aware models extend this capability to noncoding regulatory variants associated with disease[3].

3. Regulatory region analysis and expression prediction

Foundation models pretrained on human and primate genomes recover promoter and enhancer motifs directly from sequence[3]. Long context models predict chromatin accessibility and expression profiles for entire loci and enable *in silico* mutagenesis to map regulatory "hot spots"[3],[6]. These tools enrich genome annotation and help flag pathogenic variants outside protein coding regions, an area historically difficult to interpret[3].

4. Toward personalized therapeutic strategies

Generative and predictive genomic models point toward individualized interventions: designing variant specific protein repairs[5], optimizing CRISPR guide RNAs with reduced off targets, identifying tumor specific neoantigens, and repurposing drugs based on genotype pathway mappings[1]. Implementation will require prospective clinical validation and integration with expert review[7].

Discussion

Interpretation and limitations

Genomic LLMs achieve or exceed prior computational benchmarks while expanding sequence context by orders of magnitude[1],[5],[6]. Their representations appear to encode both local motif syntax and distal regulatory dependencies, offering scalable complements to experimental assays[3]. Yet several limitations remain: performance gaps for under represented ancestries[7], difficulty modeling splicing and epistasis, and opaque reasoning lacking uncertainty calibration[7]. Clinical use therefore demands prospective validation, interpretable attributions (e.g., saliency maps), and corroboration by laboratory or clinical evidence before decisions are made[7],[8].

Ethics, privacy, and equity

Genomic data are uniquely identifying[7]. Even anonymized sequences can, in theory, be relinked to individuals if models memorize or regenerate rare variants[7]. Mitigation requires privacy preserving training such as *federated learning* training across institutions without centralizing raw data and *differential privacy*, which adds statistical noise to protect individual records[7],[8]. Equitable deployment also demands diverse global datasets and bias audits so predictions perform consistently across populations[7].

Regulatory and clinical oversight

Genome LLMs intersect two regulated spheres-medical devices and high risk AI. Effective governance should classify clinical GLM tools as high risk systems requiring documented data lineage, transparency reports, post deployment monitoring, and human in the loop oversight[7]. Regulatory "sandboxes" can test these models under supervision before full approval, while independent certification can assess accuracy, bias, and privacy compliance[7]. Dual-Use Research of Concern (DURC) research that could be misapplied for harm warrants explicit mitigation[9]. Developers should apply capability controls, red teaming, and access restrictions consistent with YMJ's DURC and biosecurity policy[9].

Toward responsible implementation

Responsible adoption involves (1) interpretable, uncertainty aware model design; (2) validation on real world cohorts; (3) transparent data governance and consent frameworks; and (4) interdisciplinary collaboration among clinicians, data scientists, ethicists, and patient representatives[7–9]. For now, these systems should serve as decision support rather than decision making tools[7].

Conclusion

Personalized genome language models offer a new lens on human biology able to detect pathogenic variants, decode regulatory grammar, and inform precision therapies[1–3]. Their potential to accelerate diagnosis and treatment is immense, yet so are their risks[7–9]. With rigorous validation, interpretability, privacy protection, and biosecurity oversight, these models can amplify human expertise instead of replacing it. By pairing innovation with caution, we can ensure that the promise of AI driven genomics advances patient care without compromising

ethics or equity[7],[9]. As these models evolve toward personalization adapting to an individual's genomic and clinical data their potential impact on precision medicine will expand, as will the need for ethical safeguards.

Works Cited

1.Ali, S., Qadri, Y. A., Ahmad, K., Lin, Z., Leung, M.-F., Kim, S. W., Vasilakos, A. V., & Zhou, T. (2025). Large language models in genomics: A perspective on personalized medicine. Bioengineering, 12(5), 440. https://doi.org/10.3390/bioengineering12050440

2.Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., … Avsec, Ž. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science, 381(6664), eadg7492. https://doi.org/10.1126/science.adg7492

3.Benegas, G., Batra, S. S., & Song, Y. S. (2023). DNA language models are powerful predictors of genome-wide variant effects. Proceedings of the National Academy of Sciences, 120(44), e2311219120. https://doi.org/10.1073/pnas.2311219120

4.Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

5.Nguyen, E., Poli, M., Durrant, M. G., Kang, B., Katrekar, D., Li, D. B., … Ré, C. (2024). Sequence modeling and design from molecular to genome scale with Evo. Science, 386(6723), eado9336. https://doi.org/10.1126/science.ado9336

6.Nguyen, E., Poli, M., Faizi, M., Thomas, A., Birch-Sykes, C., Wornow, M., … Ré, C. (2023). HyenaDNA: Long-range genomic sequence modeling at single-nucleotide resolution. Advances in Neural Information Processing Systems, 36, 43177–43201. https://arxiv.org/abs/2306.15794

7.Sokhansanj, B. A., & Rosen, G. L. (2025). Regulating genome language models: Navigating policy challenges at the intersection of AI and genetics. Human Genetics, 144(9-10), 949–970. https://doi.org/10.1007/s00439-025-02768-4

8.Thomas, N., Ledergor, G., Stoeckius, M., & Stein, C. (2025). Genomic language models could transform medicine, but not yet. npj Digital Medicine, 8(1), 212. https://doi.org/10.1038/s41746-025-01603-4

Glossary of Terms

Genome Language Model (GLM)

A large language model trained on DNA sequences to learn biological patterns, enabling prediction of gene regulation and variant effects.

Personalized Genome Language Model (pGLM):

A genome language model fine tuned on individual genetic data to generate or interpret person specific insights for precision medicine.

AlphaMissense

A deep learning model developed by Google DeepMind that predicts whether amino acid changes (missense variants) in human proteins are likely to be benign or pathogenic.

AlphaFold

An artificial intelligence system by DeepMind that predicts 3D protein structures from amino acid sequences with high accuracy.

Bias (Algorithmic)

Systematic errors in model predictions that result from imbalanced or non representative training data, leading to unfair or inaccurate outcomes for certain groups.

Differential Privacy

A statistical technique that introduces small amounts of noise to data or model outputs, protecting individual privacy while maintaining aggregate accuracy.

Dual Use Research of Concern (DURC)

Scientific research that, while intended for good, could potentially be misused to cause harm (for example, generating or modifying pathogens).

Federated Learning

A privacy preserving training method in which models are trained across multiple institutions or devices without transferring raw data to a central location.

HyenaDNA

A long context genome model that replaces traditional attention mechanisms with efficient convolutions, allowing analysis of sequences up to one million base pairs long.

Missense Variant

A single nucleotide change that alters one amino acid in a protein sequence, sometimes affecting protein function or causing disease.

Precision Medicine

A healthcare approach that tailors diagnosis, prevention, and treatment to an individual's genetic, environmental, and lifestyle factors.

Regulatory Region

A segment of DNA (such as a promoter or enhancer) that controls when and how genes are expressed, rather than encoding proteins directly.

Variant Effect Prediction

The computational process of estimating whether a genetic change (variant) is harmful, neutral, or beneficial to biological function.

Single-Nucleotide Context

The local DNA sequence surrounding one nucleotide, usually including the bases right before and after it. It helps explain how mutations occur and are influenced by nearby bases.

Machine Learning-Based Osteoporosis Detection Using Patient Demographics and Radiofrequency Echographic Multi-Spectrometry Screening By Aanvi Mathur

Abstract

Osteoporosis is a chronic skeletal disorder characterized by reduced bone mass, leading to an increased risk of fractures. Because the condition often remains asymptomatic until fractures occur, early diagnosis is critical; however, gold-standard diagnostic methods, like dual-energy X-ray absorptiometry (DEXA), primarily measure cortical bone density, which does not fully reflect bone strength. This means a substantial number of fragility fractures in individuals without a DEXA-based osteoporosis diagnosis occur. While newer techniques such as Radiofrequency Echographic Multi-Spectrometry (REMS) do assess trabecular bone density and bone quality that DEXA misses, their limited availability prevents widespread use.

Therefore, this study employs machine learning models for early detection of osteoporosis risk, determined by DEXA-defined osteoporosis status, using demographic and radiofrequency-derived data. The classifiers K-Nearest Neighbors, Random Forest, Support Vector Classifier (SVC), and XGBoost were trained using patient demographic information and radiofrequency-based imaging data. Finally, model performance was evaluated using accuracy, precision, recall, F1 score, and the receiver operating characteristic area under the curve (ROC–AUC).

The SVC model had the strongest cross-validation performance on this dataset, with an accuracy of 91.9% ± 3.2% and an ROC–AUC of 0.978 ± 0.016. Random Forest and XGBoost models also achieved strong results, with accuracies of 90.7% ± 3.6%. These results suggest that machine learning offers an accurate approach for osteoporosis screening. Additionally, radiofrequency-based features show exploratory associations with promising results compared to non-radiofrequency based models, calling for further validation with the potential to support earlier diagnosis.

Introduction

Osteoporosis is a chronic skeletal condition characterized by reduced bone mineral density and deterioration of bone microarchitecture, which increases the risk of fragility fractures. The World Health Organization (WHO) defines osteoporosis as a T-score of $\leq$ -2.5 standard deviations below the average value for a healthy young adult (Salari et al.). The International Osteoporosis Foundation (IOF) estimates that osteoporosis affects around 500 million adults worldwide, and that 1 in 3 women and 1 in 5 men over the age of 50 are at risk of an osteoporotic fracture (Johnell and Kanis). In the United States, around 40-50 million adults are estimated to have low bone mass (osteopenia, defined as a T-score of <-1.0 to -2.5), and 20% of women and 4–5% of men aged $\geq$ 50 years meet diagnostic criteria for osteoporosis (Shuid et al.). Prevalence increases substantially with advancing age. Individuals with osteoporosis are typically asymptomatic until a fracture occurs. Nonvertebral fractures are usually painful and readily recognized; however, nearly 70% of vertebral compression fractures go clinically

undetected because they cause minimal or no symptoms (International Osteoporosis Foundation). As a result, osteoporosis often remains unidentified until the disease is advanced. Although osteoporosis cannot be fully reversed, patients can reduce fracture risk and improve quality of life through lifestyle modifications and pharmacologic therapy (Khan et al.). When osteoporosis is diagnosed late or remains inadequately treated, serious complications such as fragility fractures, vertebral deformities such as kyphosis, and chronic pain may develop. These complications can substantially impair function, and in severe cases, fracture-related immobility or surgical complications may contribute to increased morbidity and mortality (Shuid et al.).

Dual-energy X-ray absorptiometry (DEXA) is the most commonly used test for diagnosing osteoporosis. DEXA uses low-dose ionizing radiation to quantify bone mineral density (BMD) at the hip, lumbar spine, and forearm in order to classify patients as having normal BMD, osteopenia, or osteoporosis (Baek et al.). Despite its widespread use, DEXA has significant limitations. For example, it primarily measures cortical bone density, which determines bone quantity, but does not assess trabecular bone density, bone microarchitecture, or cortical bone turnover (bone quality), important components of bone strength. Notably, 70–80% of fragility fractures occur in individuals whose DEXA results do not meet the diagnostic threshold for osteoporosis, underscoring the limitations of relying solely on BMD (Baek et al.). In addition, degenerative changes, structural deformities, or other abnormalities, such as osteophytes and arthritis, can falsely elevate BMD measurements, potentially hiding underlying osteoporosis (Rodrigo).

In contrast, radiofrequency-based ultrasound technologies such as Radiofrequency Echographic Multi-Spectrometry (REMS) offer an alternative for osteoporosis assessment. REMS uses a non-ionizing, portable ultrasound transducer to transmit acoustic pulses toward target skeletal sites and, unlike DEXA, is less affected by any degenerative changes, vascular calcifications, or bony abnormalities. Advanced software analyzes acoustic signals to quantify bone mineral density (BMD) and evaluate microarchitectural integrity, offering a more comprehensive assessment of bone strength than DEXA scans. Despite the advantages of REMS, low availability and limited awareness limit its actual utility (Ghalenavi et al.).

Promisingly, machine learning models can overcome limitations in identifying individuals at high risk of fragility fractures by integrating data from multiple sources. These models have the potential to capture complex patterns that may not be apparent from demographics or DEXA results alone. However, existing studies have relied primarily on demographic and clinical features. For example, Qiu et al. used demographic and routine clinical data to train deep learning models and reported an AUC of 0.848. While the performance was strong, it may suggest a score ceiling when models rely only on these factors (Qiu et al.). Similarly, J.-G. Shim et al. developed a collection of machine learning models, achieving the highest AUROC of 0.743 using demographic and DEXA-based features (Shim et al.). These results demonstrate the limitations of relying solely on demographic and DEXA-derived inputs for classification. Given this limitation, Mohammed As'ad has identified the integration of

radiofrequency signal analysis with machine learning as a promising step for osteoporosis screening, though this has not been concretely investigated (As'ad).

Therefore, this study investigates whether integrating phalangeal-derived radiofrequency measurements with supervised machine learning alters osteoporosis risk classification performance. Using two publicly available datasets containing patient demographic information and radiofrequency-based features, we develop multiple machine-learning classifiers trained to distinguish between osteoporotic and non-osteoporotic individuals. The performance of these models is evaluated using standard classification metrics, including accuracy, precision, recall, F1 score, and receiver operating characteristic (ROC) analysis. By assessing the contribution of radiofrequency-based features alongside traditional demographic risk factors, this study aims to explore the potential role of radiofrequency-supported machine learning in early osteoporosis screening and in more accessible diagnostic strategies.

Methods

This study used two publicly available datasets for osteoporosis screening, both based on radiofrequency-based measurements. The first dataset, *Application of a Neural Network Classifier to Radiofrequency-Based Osteopenia/Osteoporosis Screenin*g, by J.W Adams et al. includes information from 80 participants, 38 of whom had osteoporosis. It contains basic demographic information (age, sex, height, and weight) as well as radiofrequency reflection and transmission data, recorded as complex-number values (Adams et al.). These measurements were collected by sending very low-power electromagnetic waves through each participant's right and left wrists. Transmission data shows how much of the wave traveled through the bone, while reflection data shows how much it bounced back from the wrist (Adams et al.). All samples were collected in the span of one year from the Worcester Polytechnic Institute.

Participants were originally diagnosed using a five-point scale ranging from healthy to osteoporotic. However, only a few participants fell into the middle categories, which made the data unbalanced. To make the dataset more suitable for machine learning models, the scale was simplified to binary classification: 0 = Non-Osteoporotic, 1 = Osteoporotic.

Although the study originally planned to use only the Adams et al. dataset, its small sample size made it difficult to build reliable machine-learning models. To improve accuracy and make the results more generalizable to larger populations, a second dataset, *A labeled dataset for osteoporosis screening based on electromagnetic attenuation*, published by Carvalho et al. on Zenodo, was also included. The dataset contains data from 945 participants (Carvalho et al.). The data was collected at the Onofre Lopes University Hospital (HUOL) at the Federal University of Rio Grande do Norte (UFRN), between July 2021 and November 2023. In addition to demographic information, it includes more detailed radiofrequency-based measurements, such as electromagnetic attenuation and calibration values. Attenuation refers to the loss of signal strength as the wave travels through body tissue, and calibration values help correct for normal variations in the measuring device (Carvalho et al.). This dataset also provides biometric features of the hand phalanges, such as medial length, medial width, and medial height, derived from

radiofrequency-based echo signals. Processed features such as "body deviation" and "spine deviation" were removed as derived, not raw, quantities, which may introduce further systematic and methodological bias into ML model results. Like the first dataset, each participant was labeled as either non-osteoporotic or osteoporotic. These labels were based on T-score deviation scores derived from DEXA, with values below -2.5 defined as having osteoporosis.

All analyses were performed in Python version 3.12 using Google Colaboratory. The study relied on several scientific computing and machine-learning libraries, including NumPy 2.4.6, Pandas 3.0.3, SciPy 1.17.1, scikit-learn 1.7.0, XGBoost 3.1.1, SHAP 0.49.1, Matplotlib 3.10.6, and Seaborn 0.13.2. To ensure consistent results, a fixed random seed of 42 was used throughout all stages of the analysis.

Before applying machine learning models, both datasets underwent a detailed preprocessing procedure. All column names were converted to lowercase for uniformity, and variables that represented the same information were standardized into a single feature. For example, "sex" from the first dataset and "gender" from the second were aligned across datasets. Diagnostic categories were also standardized by mapping all textual labels to a binary format: 0 for individuals without osteoporosis and 1 for individuals with osteoporosis. Because the two datasets measured radiofrequency-based data differently, only the demographic variables shared by both (such as age, sex, height, and weight) were used to merge them. Dataset-specific radiofrequency-based variables, such as attenuation in the Carvalho et al. dataset or reflection/transmission values in the Adams et al. dataset, were kept but not matched across sources. Importantly, these radiofrequency-based features represent different physical quantities and were measured at different anatomical sites (wrist versus hand phalanges). As a result, radiofrequency variables were not directly comparable across datasets. Since the majority of the dataset was composed of patients from the Carvalho et al. study (945/1025), the attenuation and calibration variables were kept in the final dataset, while reflection and transmission variables were excluded. The missing values for the Adams et al. participants (80/1025) were imputed using feature-wise mean imputation, where each missing value was replaced with the mean of the observed values for that feature. After cleaning and merging, the final combined dataset included 1,025 participants, of whom 422 had osteoporosis, and 603 participants did not have osteoporosis.

From this unified dataset, nine variables were selected for machine learning analysis. These included demographic features age, height, weight, and gender, as well as radiofrequency-based features calibration, attenuation, and the physical features medial length, medial width, and medial height. All selected features were numerical, eliminating the need for additional encoding steps.

Since the number of participants in each diagnostic group was uneven, class imbalance posed a risk of biased predictions. To address this, class weighting was applied during model training. For algorithms that support built-in weighting, such as Random Forest and Support Vector Classifier (SVC), the class weight was set to ‘balanced’. For XGBoost, the weight

assigned to positive samples was determined using the ratio of negative to positive samples in the training data.

To reduce the risk of information leakage between the training and test data, a group-aware data-splitting strategy was used. The dataset was divided into training and test sets using an 80/20 split implemented with GroupShuffleSplit. A subject grouping variable was constructed to ensure that samples from the same subject were not included in both sets. This approach was chosen to prevent overly optimistic performance estimates that can arise when correlated samples are distributed across splits. A fixed random state (42) was used to guarantee that the split remained identical across repeated experiments. The training set was reserved for model development, cross-validation, and hyperparameter tuning, while the testing set was used only for final, unbiased performance evaluation.

To prepare the data for analysis, imputation and standardization were performed exclusively on the training set, then applied to the test set to avoid information leakage. Mean imputation, implemented using the SimpleImputer class, was used to fill any missing values, primarily in the radiofrequency-based variables derived from the Carvalho et al. dataset for the 80 Adams et al. samples. Standardization using StandardScaler was performed to ensure that variables with larger numerical ranges, such as height or weight, did not dominate variables with smaller scales, such as calibration or attenuation.

Following preprocessing, the supervised machine-learning models K-Nearest Neighbors (KNN), SVC, Random Forest, and XGBoost were trained (Table 1). Model evaluation during training was performed using group-aware cross-validation. Then, the models were tested on a test set with n=248 patients (70 = osteoporotic, 178 = non-osteoporotic). When available, StratifiedGroupKFold was used to preserve class proportions while preventing overlap of subject groups across folds. In cases where stratification was not supported, GroupKFold was applied. This ensured that samples from the same subject group were not simultaneously present in the training and validation folds. To further enhance model accuracy and reduce the risk of overfitting, hyperparameters were optimized through nested cross-validation. GridSearchCV was used for models with smaller hyperparameter spaces, while RandomizedSearchCV was used for XGBoost due to its larger and more complex parameter space (Table 2) (Alamsyah et al.).

Performance metrics reported throughout this study represent mean values across 5-fold group-aware cross-validation, with uncertainty quantified as standard deviation (SD) across folds. For ROC-AUC values, 95% confidence intervals were calculated using the DeLong method, which compares the performance of models through their ROC curves.

To compare model performance, multiple paired statistical tests were conducted. McNemar's test was applied to assess differences in classification accuracy between models on the same test set, since it accounts for the paired nature of predictions on identical samples. For ROC-AUC comparisons, the DeLong test was used to evaluate whether differences in area under the curve were statistically significant. Given the class imbalance in the dataset (422 osteoporotic vs. 603 non-osteoporotic participants), the precision-recall area under the curve (PR-AUC) was

calculated as a complementary value to ROC-AUC, as it is better at distinguishing performance differences on the positive (osteoporotic) class.

To verify that imputation did not introduce systematic bias by encoding dataset identity, a secondary classification task was performed. A Random Forest classifier was trained to predict which dataset a sample originated from using only the imputed radiofrequency-based features. The same group-aware data splitting and preprocessing pipeline used in the primary analysis was applied. The model was evaluated using ROC-AUC to assess whether imputed features contained dataset-identifying information.

Additionally, to confirm the absence of dataset leakage contributing to the performance of the model, three validation experiments were conducted. First, models were trained on one dataset (ex. Carvalho et al.) and tested on the other (ex. Adams et al.) to assess the generalization across datasets with different measurement protocols. The models were either trained on Adams et al. data and tested on Carvalho et al. data, or the opposite. Secondly, separate models were trained independently on Adams et al. data and Carvalho et al.-only data using group-aware data splitting to evaluate the performance with no cross-dataset information. Finally, a binary dataset ID feature (0 = Adams et al., 1 = Carvalho et al.) was included as a model input to test whether performance improved. If performance showed demonstrable improvements, it would indicate that models were utilizing dataset membership information for their predictions.

Model Type	**Model Description**
K-Nearest-Neighbors	Classifies new, unlabeled points by comparing them to their "nearest neighbors", or their closest data points. The "k" in K-Nearest-Neighbors represents the number of neighbors the model compares the unidentified data point to (Bansal et al.).
Support Vector Classifier	This model finds a hyperplane (a line) to separate data points of different classes and to classify unknown points (Pisner and Schnyer).
Random Forest	This model extends a Decision Tree model, which uses a tree-like structure, splitting into branches and selecting the best attributes at each branch, combining them to predict the best outcome for an unknown data point (Charbuty and Abdulazeez). Random Forest uses predictions from other decision trees to create an overall prediction (Breiman).
XGBoost	This model uses multiple decision trees in sequence to predict a data point's classification. Each tree learns from the errors of previous trees, and the final prediction is produced by combining the outputs of all trees via gradient boosting (Alamsyah et al.).

Table 1: Description of different model types used in the study.

Model	Hyperparameter Value
Random Forest	n_estimators: 300 criterion: log_loss class_weight: balanced random_state: 19 n_jobs: -1 max_features: 3
SVC	C: 100 probability: True class_weight: balanced random_state: 42
XGBoost	n_estimators: 200 max_depth: 4 learning_rate: 0.05 subsample: 0.8 colsample_bytree: 0.8 eval_metric: logloss random_state: 42
KNN	n_neighbors: 5

Table 2: Optimized hyperparameter values for models chosen for testing.

Finally, several strategies were applied to reduce overfitting. Group-aware data splitting and cross-validation were used to prevent subject-level leakage. Regularization and model-specific constraints were applied where applicable, including depth limits and minimum sample requirements for tree-based models. All preprocessing steps were fit exclusively to the training data before being applied to the validation or test sets. (Table 3).

Method	**How it Helps with Overfitting**
Group-aware 5-fold cross-validation during training	Provides a more reliable estimate of model performance on unseen data (Wändell et al.)
Nested cross-validation during hyperparameter tuning	Prevents hyperparameter leakage for Random Forest and SVC (choosing parameters that only perform well on a single validation set) (Wändell et al.)
Pipeline-based preprocessing	Avoids data leakage (information from the test fold leaking into the training) (Smets et al.)
Regularization	Prevents the model from prioritizing noisy and outlier features(Umapathy et al.)
Depth and complexity control in tree-based methods	Preventing trees from fitting all the information in the

	training data (Smets et al.)
Random state testing	Ensures that optimal performance is not due to splitting or initialization (Wändell et al.)
Removal of leaky features	Excludes features that contain information derived directly from the target variable, as leaky features lead to unrealistically high training performance (Smets et al.)

Table 3: Summary of methodological strategies used to prevent model overfitting. These approaches together ensured that the models' performance truly reflects generalization to unseen data rather than simply memorization of training patterns.

Results

The three experiments defined previously (cross-dataset validation, testing dataset-specific models, and using dataset ID as features) demonstrated promising results. The overall cross-dataset average (experiment 1) had an ROC-AUC of 0.676 across Random Forest, SVC, XGBoost, and KNN. This suggests that the model learns some generalization across datasets with different measurement protocols (phalange-derived vs. wrist-derived). While this performance exceeds chance, the substantial drop from the within-dataset performance (ROC-AUC 0.93 average for SVC, XGBoost, and KNN) reflects the difficulties of transferring models across different measurement locations. This performance gap is expected given that the Adams et al. dataset utilized wrist-based reflection/transmission measurements while the Carvalho et al. dataset used phalangeal-based attenuation measurements. The Carvalho et al.-only model achieved accuracies of 76-77% and ROC-AUC values of 0.83-0.89 across classifiers, demonstrating good independent performance. Contrastingly, the Adams et al.-only model showed high variance due to the much smaller sample size. However, the strong Carvalho et al.-only performance validates that this dataset independently contains genuine osteoporosis-predictive signals. Importantly, adding an explicit dataset identifier as a model feature resulted in minimal change in performance average (average $\Delta AUC = -0.006 \pm 0.021$). Since zero falls within one standard deviation of the average change, the results are consistent with no meaningful effect from dataset ID. Therefore, the models were not relying strongly on dataset ID to make predictions.

Collectively, these experiments indicate that while some dataset-specific structure still exists (as indicated by the moderate cross-dataset of 0.676 and the previously reported dataset identity prediction AUC of 0.62), the models are not primarily relying on dataset identity for osteoporosis classification. The strong Carvalho et al.-only performance, the minimal impact of dataset ID, and the above-chance cross-dataset generalization indicate that the models have generally learned genuine osteoporosis-related patterns rather than pure dataset structure. Thus, while some dataset-specific structure remains, these experimental results provide reasonable confidence that the metrics discussed below are genuine osteoporosis-related signals.

Model performance was first summarized across cross-validation folds using accuracy, and weighted averages of precision, recall, and F1 score. Weighted averages represent performance across both diagnostic classes while also adjusting for the number of samples in each class.

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC (95% CI)	PR-AUC	Brier Scores
Random Forest	90.7% ± 3.6%	91.8% ± 2.8%	90.7% ± 3.6%	90.1% ± 4.1%	0.966 (0.929–1.000)	0.89	0.081
Support Vector Classifier	91.9% ± 3.2%	92.4% ± 3.0%	91.9% ± 3.2%	92.1% ± 3.2%	0.978 (0.947–1.000)	0.91	0.076
XGBoost	90.7% ± 3.6%	91.8% ± 2.8%	90.73% ± 3.6%	90.1% ± 4.1%	0.909 (0.828–1.000)	0.88	0.085
KNN	84.3 ± 4.4%	83.8 ± 4.7%	84.3 ± 4.4%	83.9% ± 4.7%	0.862 (0.760–0.964)	0.81	0.123

Table 4: Cross-validation model performance across validation folds with standard deviation for uncertainty quantification. Additionally, PR-AUC and Brier scores are included for further validation of results. For ROC-AUC values, 95% confidence intervals were calculated using the DeLong method. SVC achieved the strongest overall performance across all metrics, while KNN presented with lower discrimination ability.

Given the class imbalance in the dataset, the precision-recall area under the curve (PR-AUC) was calculated as a complementary metric to ROC-AUC, and demonstrated strong performance despite the class imbalance, ROC-AUC values and PR-AUC values following a similar pattern. Model calibration was assessed using Brier scores, which quantify the accuracy of probabilistic predictions, where lower scores indicate better calibration. Brier scores were lowest for SVC, Random Forest, and XGBoost, but KNN showed weaker calibration compared to the other models. (Table 4).

McNemar's test was conducted to compare the classification accuracy between all model pairs. McNemar's test compares whether two models make different errors on the same test samples, and a significant result indicates that their error patterns differ. The SVC had significantly different error patterns compared to other models: SVC vs Random Forest ($\chi^2 = 28.20$, $p < 0.001$), SVC vs XGBoost ($\chi^2 = 25.93$, $p < 0.001$), and SVC vs KNN ($\chi^2 = 27.84$, $p < 0.001$). Since SVC produced the fewest false negatives but also more false positives than Random Forest and XGBoost, these values indicate that SVC may be most useful from a screening perspective, where minimizing missed osteoporosis cases is especially important. No significant differences were detected among Random Forest, XGBoost, and KNN (all $p > 0.05$), indicating that the remaining models had similar classification error rates.

Overall, SVC achieved the highest accuracy (91.9% ± 3.2%), precision (92.4% ± 3.0%), F1 score (92.1% ± 3.2%), and ROC-AUC (0.978, 95% CI 0.947–1.000). Random Forest and XGBoost also demonstrated strong classification performance, with accuracies of approximately 91%, while KNN achieved lower overall performance (Table 4).

Precision, recall, and F1 score generally followed similar trends across models. SVC and Random Forest demonstrated strong and balanced performance across all metrics. XGBoost achieved comparable accuracy but showed reduced sensitivity, indicating that a greater number of osteoporotic cases were missed. KNN demonstrated lower performance across most classification metrics relative to the other evaluated models.

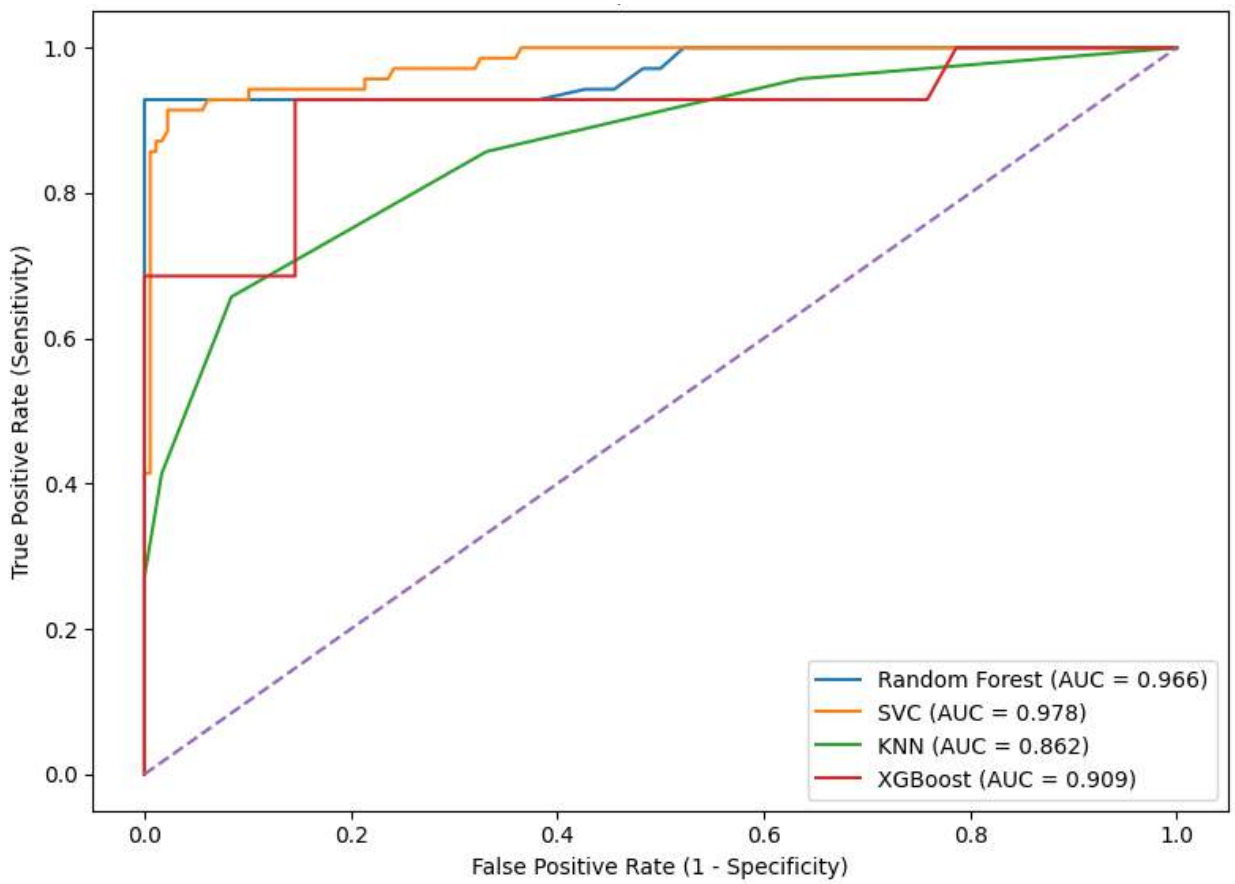


Figure 1: ROC curves for the Random Forest, SVC, KNN, and XGBoost classification models with the combined dataset (true positive rate vs. false positive rate). SVC and Random Forest had the strongest discrimination ability, meaning they can most reliably separate osteoporotic from non-osteoporotic individuals for all classification thresholds.

ROC analysis further supported these findings. SVC achieved the highest discrimination (ROC-AUC = 0.978), followed closely by Random Forest (ROC-AUC = 0.966). XGBoost (ROC-AUC = 0.909) and KNN (ROC-AUC = 0.862) showed progressively lower discriminatory

ability (Figure 1). These results demonstrate that although several models performed well, SVC provided the strongest overall separation between osteoporotic and non-osteoporotic individuals.

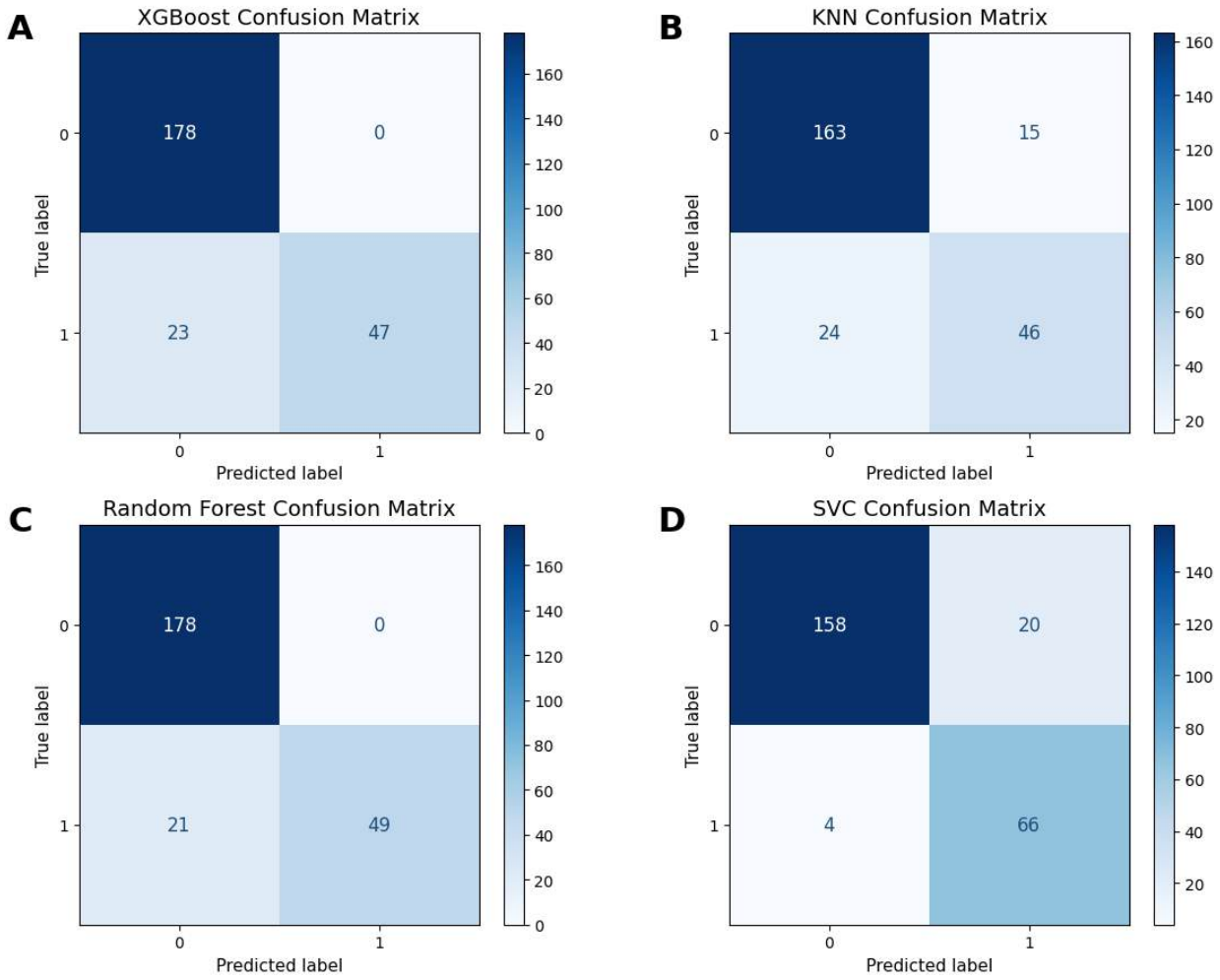


Figure 2: Confusion matrices for XGBoost (A), KNN (B), Random Forest (C), and SVC (D) models using the combined dataset on the independent test set. SVC indicated the fewest false negatives but also the most false positives, which indicates that specificity may be sacrificed when sensitivity is improved.

Confusion matrix analysis provided additional insight into the error profiles of the evaluated models. In the context of osteoporosis screening, minimizing false negatives is particularly important, as incorrectly classifying an osteoporotic patient as healthy could delay further evaluation and treatment (Umapathy et al.). Across models, error patterns varied. Random Forest and XGBoost demonstrated strong specificity, correctly classifying all non-osteoporotic cases in the test set, but both models exhibited false negatives, indicating that a subset of osteoporotic cases was not detected (Figures 2A and 2C). XGBoost misclassified 23 of 70 (32.9%) osteoporotic cases as non-osteoporotic, while Random Forest misclassified 21 such cases. In contrast, the SVC achieved the fewest false negatives, correctly identifying the majority of osteoporotic cases, but at the cost of more false positives (Figure 2D). KNN showed a more

balanced distribution of errors, with both false positives and false negatives present (Figure 2B). While this resulted in moderate overall performance, it highlights differences between sensitivity and specificity across models.

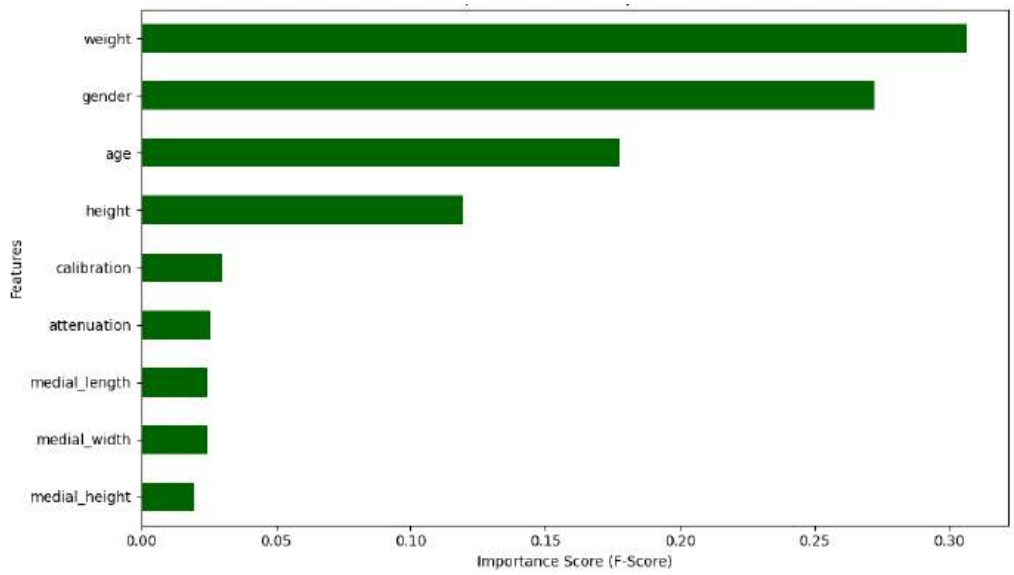


Figure 3: Relative importance scores (F-scores) for demographic, radiofrequency-based, and physical features used in the XGBoost model using the combined dataset. Demographic variables accounted for 87.6% of total feature importance, indicating they are the primary variables for classification, while radiofrequency-based features provide supplementary predictive signals.

Feature importance analysis showed that demographic and clinical variables accounted for 87.6% of the total importance across the models, consistent with established clinical risk factors (Figure 3). This analysis required the calculation of an F-score, which measures how often a feature is used in model splits. Higher F-scores indicate more discriminatory performance. Radiofrequency-based features accounted for the remaining 12.4%; however, this contribution may reflect both dataset structure and underlying biological signal (Figure 3). No single feature dominated the predictions, indicating that the models relied on a balanced combination of information rather than overfitting to one particular variable.

Finally, regularization and complexity control strategies were applied for the Random Forest model to limit excessive fitting to the training data, since that is a major risk for tree-based models. For example, the number of trees was limited to 300, and features per split was limited to 3. This was intended to reduce model complexity and stabilize training, though it does not guarantee optimal generalization. Model performance was therefore evaluated on an independent test set.

Discussion

This study evaluated whether radiofrequency–based measurements could improve osteoporosis risk classification when combined with demographic features. Using group-aware

data partitioning and conservative evaluation strategies, the results demonstrate that multiple machine learning models were able to distinguish between osteoporotic and non-osteoporotic individuals, though with notable differences in error profiles, feature utilization, and sensitivity to radiofrequency-based inputs.

Radiofrequency-based features contributed as context-dependent enhancers to prediction, accounting for 12.4% of total feature importance, despite demographic variables remaining the strongest predictors. However, by converting diagnoses to a simple binary label (non-osteoporotic vs. osteoporotic), subtle distinctions, such as osteopenia, were lost, making the classification task easier but less generalizable to real clinical settings. Demographic variables such as age, weight, height, and gender are well-established risk factors for osteoporosis and thus provide a reliable foundation for prediction. Radiofrequency-based measurements supplemented these predictors and may be supportive in borderline cases that are more challenging to classify, though this cannot be confirmed within the current study design.

Importantly, several demographic predictors, such as age, height, and weight, are naturally correlated. This encompasses the principle of collinearity, which states that one variable can potentially predict another's value. However, this correlation is unlikely to meaningfully bias the conclusions because the primary models emphasized in this study can tolerate collinear inputs, since the goal is prediction rather than the interpretation of individual coefficients. Therefore, models were evaluated primarily by how accurately they predicted unseen data, rather than by the individual contributions of each feature.

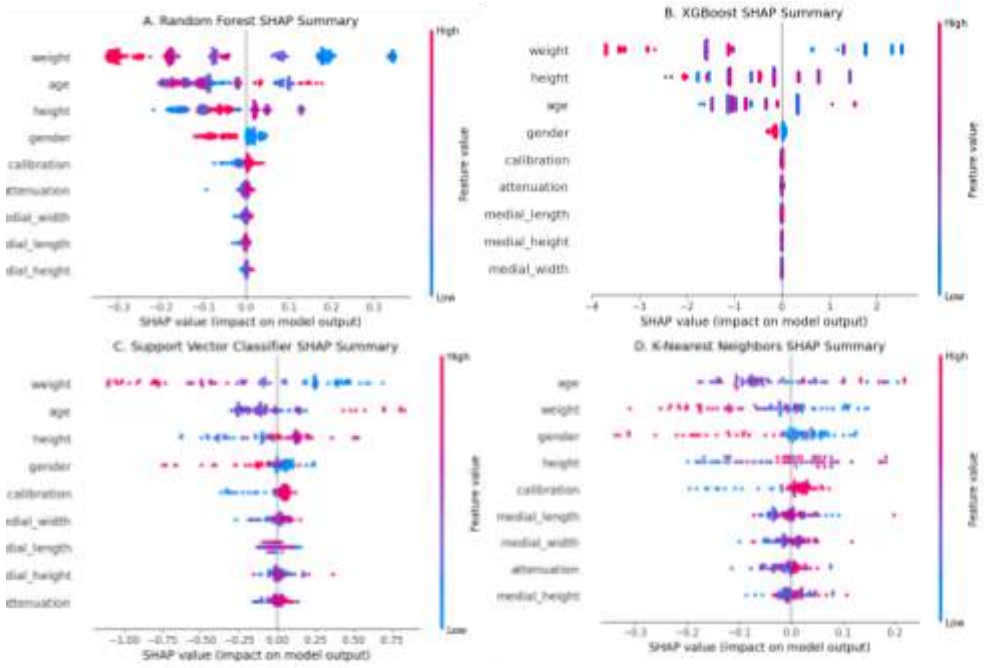


Figure 4: SHAP summary plots for the Random Forest (A), XGBoost (B), Support Vector Classifier (C), and K-Nearest Neighbors (D) models. Each plot shows the direction and magnitude of feature contributions for individual predictions. Importantly, the influence of radiofrequency-derived features varied by model type.

SHAP analyses supported these observations. Demographic variables produced the most significant SHAP values across all models, indicating the greatest influence on predictions (Figure 4). Older age and lower body weight consistently increased predicted risk, while younger age and higher weight decreased it. Radiofrequency-based features had minimal impact on the XGBoost model (Figure 4B) but made more significant contributions to the Random Forest, SVC, and KNN models (Figure 4A, 4C, 4D), suggesting that their importance varies across model types.

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**ROC-AUC (95% CI)**
Random Forest with radiofrequency-based features	90.7% ± 3.6%	91.8 ± 2.8%	90.7 ± 3.6%	90.1 ± 4.1%	0.972 (0.929–1.000)
Random Forest without radiofrequency-based features	71.4 ± 5.4%	75.5 ± 5.6%	71.4 ± 5.4%	72.5 ± 5.3%	0.875 (0.769–0.981)
SVC with radiofrequency-based features	91.9 ± 3.2%	92.4 ± 3.0%	91.9 ± 3.2%	92.1 ± 3.2%	0.978 (0.947–1.000)
SVC without radiofrequency-based features	85.9 ± 4.2%	88.7 ± 3.2%	85.9 ± 4.2%	86.4 ± 4.0%	0.986 (0.961–1.000)
KNN with radiofrequency-based features	84.3 ± 4.4%	83.8 ± 4.7%	84.3 ± 4.4%	83.9% ± 4.7%	0.862 (0.760–0.964)
KNN without radiofrequency-based features	91.1% ± 3.4%	92.1% ± 2.7%	91.1% ± 3.4%	90.6% ± 3.9%	0.843 (0.739–0.947)
XGBoost with radiofrequency-based features	90.7% ± 3.6%	91.79% ± 2.8%	90.73% ± 3.6%	90.1% ± 4.1%	0.920 (0.828–1.000)
XGBoost without radiofrequency-based features	75.8% ± 5.2%	81.2% ± 4.5%	75.8% ± 5.2%	76.9% ± 5.1%	0.851 (0.739–0.963)

Table 5: Comparison of model performances with and without radiofrequency-based features with standard deviation for uncertainty quantification across cross-validation folds. For ROC-AUC values, 95% confidence intervals were calculated using the DeLong method. Across most models, radiofrequency-based features improved accuracy and overall classification performance, but KNN performance improved on demographic features alone.

The radiofrequency measurements used in this study were obtained from peripheral skeletal sites, specifically the wrist and hand phalanges. These locations do not correspond directly to clinically important fracture sites, such as the hip or spine. Therefore, radiofrequency-based measurements from peripheral bones should not be interpreted as a direct representation of axial bone mineral density or site-specific fracture risk prediction. Instead, the improvement in model performance with the inclusion of radiofrequency features likely reflects the systemic nature of osteoporosis. Osteoporosis is characterized by widespread alterations in bone remodeling and microarchitectural integrity that affect the entire skeleton, even though fracture risk manifests in specific sites (Johnell and Kanis). Peripheral skeletal sites may therefore function as locations that capture overall skeletal health, rather than serving as anatomically equivalent representations for axial bone density.

Importantly, the models were trained both with and without radiofrequency-based features to examine their association with model output (Table 5). Across most classifiers, models including radiofrequency-based features achieved higher performance metrics in this combined dataset. For example, the Random Forest model achieved an accuracy of 90.7% ± 3.6% with radiofrequency-based features, compared to 71.4% ± 5.4% without radiofrequency-based features, and showed an improvement in ROC–AUC (0.972 ± 0.022 vs. 0.875 ± 0.054). Similarly, the SVC demonstrated high overall performance with radiofrequency-based features (accuracy 91.9% ± 3.2%, ROC–AUC 0.978 ± 0.016), with comparable discrimination and improved overall classification. XGBoost showed a comparable trend, with accuracy increasing from 75.8% ± 5.2% without radiofrequency-based features to 90.7% ± 3.6% with them, alongside improved discrimination (ROC–AUC 0.920 ± 0.047). Interestingly, the ROC-AUC values for SVC with and without radiofrequency-based features were comparably high (0.978 vs. 0.986) (Table 5). This may be because SVC was able to identify an effective decision boundary using only demographic variables; additionally, ROC-AUC and accuracy are different in the sense that the former reflects discrimination ability across all thresholds, while the latter reflects performance at a specific threshold. It could be possible that radiofrequency features changed the decision boundary in a way that improved accuracy without changing the overall ranking of predictions.

In machine learning, radiofrequency-based features appear to provide complementary information that optimizes risk classification when combined with demographic variables, rather than independently determining classification outcomes. Models capable of capturing nonlinear relationships, such as Random Forest, SVC, and XGBoost, demonstrated better use of this supplementary signal. However, distance-based classification methods, such as KNN, performed worse when radiofrequency features were included, with accuracy decreasing by nearly 7%. This performance likely stems from dimensionality, where increased feature space dilutes the neighborhood structure that KNN relies upon for classification. The addition of radiofrequency features increases the dimensionality of the feature space, making distances between data points more uniform and reducing the algorithm's ability to identify meaningful nearest neighbors.

Previous work reported moderate discrimination using demographic and clinical features alone, with best AUC values of 0.848 when using demographic inputs (Qiu et al.) and 0.743 when using both demographic and DEXA-based features (Shim et al.). The results of this study suggest that, as As'ad proposed, radiofrequency variable integration improves the performance of machine learning for osteoporosis classification. However, this comparison is limited since differences in datasets and evaluation strategies, along with the downsides of imputation, may have affected the model performance of this study. Despite this, the improvement in classification performance across most models when radiofrequency features were included, and the results of the experiments conducted to determine dataset leakage, suggest that radiofrequency-based measurements could provide a complementary signal beyond what demographic or DEXA variables alone can capture.

The Random Forest classifier trained to predict dataset membership using the same feature set as the main models achieved ROC-AUC = 0.62 ± 0.07 on the test set. Although this performance is substantially lower than that observed for osteoporosis classification, it suggests that dataset-related structure remains present despite imputation. Therefore, part of the performance gain observed when radiofrequency-based features are included reflects the dataset structure in addition to biological information.

Limitations

The study has several limitations. A primary limitation of this study is the heterogeneity of radiofrequency-based feature acquisition across datasets. The Adams et al. dataset measured wrist-based radiofrequency reflection and transmission values, whereas the Carvalho et al. dataset included attenuation and biometric features derived from phalangeal measurements using different hardware configurations. These radiofrequency variables, therefore, represent distinct physical quantities measured at different anatomical sites and are not directly comparable. To enable combined machine-learning modeling, radiofrequency-based features were imputed for participants lacking corresponding measurements. Because these imputed values do not reflect true subject-level measurements, this process may introduce systematic bias and inflate the apparent contribution or performance associated with radiofrequency features. For example, testing a Random Forest model with only features from the Carvalho et al. dataset showed that some dataset information contributed to a model's final performance.

Importantly, the technical differences between the datasets introduce a distribution shift, which limits some cross-dataset transferability. This caused a performance gap between training models using cross-dataset validation (AUC 0.676) as opposed to within-dataset performance (0.97). Therefore, all radiofrequency-related findings in this study should be interpreted as exploratory. The sample size, while larger than the Adams et al. dataset alone, remains modest and may not fully capture the demographic and clinical variability of real-world populations. Because the datasets used in this study may not fully represent global populations, the findings may not generalize to other ethnicities or regions. Additionally, true patient identifiers were not available across all data sources. As a result, subject groupings were approximated using shared

demographic attributes, which may imperfectly represent individual subjects and could introduce inaccuracies in the groupings. Lastly, the simplified diagnostic labels used in the study removed the intermediate osteopenia category, limiting its clinical application. External validation in independent and more diverse populations is essential to assess whether these models generalize beyond the datasets used here.

From a clinical perspective, the results suggest that machine learning models incorporating radiofrequency-based features may be most appropriate as adjunctive screening tools rather than standalone diagnostic systems. Additionally, there was a significant number of false negatives across several high-performing models, highlighting the need for careful clinical considerations: their low sensitivity can lead to missed cases, underscoring the need for improvement in future studies. Radiofrequency-based measurements offer complementary value by modestly enhancing risk stratification and identifying individuals who may benefit from confirmatory testing such as DEXA scanning.

Future Work

Future work should focus on validating these findings in larger, more diverse populations and external cohorts, encompassing the full diagnostic spectrum of bone health, and assessing performance in prospective settings. For example, while the current dataset included participants across many age groups, its geographical distribution was limited. Therefore, incorporating data from multiple populations with diverse ethnic and socioeconomic backgrounds would improve the generalizability of the machine learning models. Simultaneously, increasing the size of the dataset would allow the models to identify more patterns across participants, similarly improving the accuracy and generalizability of the prediction. Additionally, this study was limited to a binary classification system; creating a model that is able to classify more nuanced cases (ex. osteopenia) would reduce false positives/negatives and make for a more trustworthy classification. Because radiofrequency measurements can vary depending on the anatomical site examined, future datasets should include measurements from multiple anatomical locations to assess the robustness of models across scanning procedures. Finally, evaluating radiofrequency-enhanced models alongside other established screening tools may further clarify whether radiofrequency-based measurements provide support in early osteoporosis detection.

Additionally, although radiofrequency-based measurements generally improved performance, their limited availability can restrict models' use in clinical settings. Developing models that retain the predictive benefits of REMS-enhanced systems without requiring radiofrequency data at the time of risk prediction could outweigh this limitation, with one approach being a teacher-student framework. In this framework, a high-performing model trained on demographic and radiofrequency-based features (teacher; in this case, the models trained in the study) would be used to guide the training of a second model that relies only on demographic or DEXA-derived inputs (student). Instead of learning from binary diagnostic labels like the original models, the student model would be trained to predict the teacher's outputs and to capture subtle patterns from radiofrequency features. At deployment, the student

model would operate on widely available, low-cost inputs, enabling REMS-free risk predictions while approximating the decision behavior of a more complex system.

Conclusion

This study shows that machine learning models combining demographic and radiofrequency-based measurements can achieve high accuracy in binary osteoporosis classification within the evaluated datasets. Although demographic variables remain the primary predictors, radiofrequency-based features showed variable contributions across different model architectures. Random Forest and XGBoost models showed higher performance metrics when radiofrequency features were included (accuracies increased from 71.4% to 90.7% and from 75.8% to 90.7%, respectively); however, these differences should be interpreted cautiously, given differences in how radiofrequency features were obtained and in the use of imputation. The SVC achieved high performance in both datasets, and KNN performed better with demographics alone. These mixed results suggest that the value of radiofrequency-based features may depend on the model. These models would be implemented to complement, not replace, existing clinical screening pathways. To advance these findings toward clinical implementation, further validation across the full diagnostic spectrum, including osteopenia, and external testing in diverse populations, is essential to determine real-world generalizability. With continued refinement and validation, radiofrequency-based tools may offer a significant step forward in improving early osteoporosis detection and reducing fracture risk.

Works Cited

Adams, Jason W., et al. “Application of a Neural Network Classifier to Radiofrequency-Based Osteopenia/Osteoporosis Screening.” 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2021, pp. 15–18. https://doi.org/10.1109/EMBC46164.2021.9630944.

Alamsyah, Nur, et al. “XGBoost Hyperparameter Optimization Using RandomizedSearchCV for Accurate Forest Fire Drought Condition Prediction.” Jurnal Pilar Nusa Mandiri, vol. 20, no. 2, 2024, pp. 103–110. https://doi.org/10.33480/pilar.v20i2.5569.

As’ad, Mohammed. “Axial Skeletal Assessment in Osteoporosis Using Radiofrequency Echographic Multi-Spectrometry: Diagnostic Performance, Clinical Utility, and Future Directions.” Cureus, vol. 17, no. 6, 2025, e86625. https://doi.org/10.7759/cureus.86625.

Baek, Yeon-Hee, et al. “10-Year Fracture Risk in Postmenopausal Women with Osteopenia and Osteoporosis in South Korea.” Endocrinology and Metabolism, vol. 36, no. 6, 2021, pp. 1178–1188. https://doi.org/10.3803/EnM.2021.1215.

Bansal, Mahima, Abhinav Goyal, and Ankur Choudhary. “A Comparative Analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory Algorithms in Machine Learning.” Decision Analytics Journal, vol. 3, 2022, 100071. https://doi.org/10.1016/j.dajour.2022.100071.

Breiman, Leo. “Random Forests.” Machine Learning, vol. 45, no. 1, 2001, pp. 5–32. https://doi.org/10.1023/A:1010933404324.

Carvalho, D. D. A., et al. “A Labeled Dataset for Osteoporosis Screening Based on Electromagnetic Attenuation.” Frontiers in Digital Health, vol. 7, 2025, 1538477. https://doi.org/10.3389/fdgth.2025.1538477.

Charbuty, Bahzad, and Adnan M. Abdulazeez. “Classification Based on Decision Tree Algorithms for Machine Learning.” Journal of Applied Science and Technology Trends, vol. 2, no. 1, 2021, pp. 20–28. https://doi.org/10.38094/jastt20165.

Ghalenavi, Elham, et al. “Diagnostic Value of Radiographic Singh Index Compared to Dual-Energy X-Ray Absorptiometry Scan in Diagnosing Osteoporosis: A Systematic Review.” The Archives of Bone and Joint Surgery, vol. 12, no. 1, 2024, pp. 1–11. https://doi.org/10.22038/abjs.2023.70632.3309.

International Osteoporosis Foundation. Epidemiology of Osteoporosis and Fragility Fractures. International Osteoporosis Foundation, n.d., https://www.osteoporosis.foundation/facts-statistics/epidemiology-of-osteoporosis-and-fragility-fractures. Accessed 11 June 2026.

Johnell, O., and J. A. Kanis. “An Estimate of the Worldwide Prevalence and Disability Associated with Osteoporotic Fractures.” Osteoporosis International, vol. 17, no. 12, 2006, pp. 1726–1733. https://doi.org/10.1007/s00198-006-0172-4.

Khan, Aliya A., et al. “Osteoporotic Fractures: Diagnosis, Evaluation, and Significance from the International Working Group on DXA Best Practices.” Mayo Clinic Proceedings, vol. 99, no. 7, 2024, pp. 1127–1141. https://doi.org/10.1016/j.mayocp.2024.01.011.

Pisner, Derek A., and David M. Schnyer. “Support Vector Machine.” Machine Learning: Methods and Applications to Brain Disorders, edited by Andrea Mechelli and Sandra Vieira, Academic Press, 2020, pp. 101–121. https://doi.org/10.1016/B978-0-12-815739-8.00006-7.

Qiu, Chuan, et al. “Developing and Comparing Deep Learning and Machine Learning Algorithms for Osteoporosis Risk Prediction.” Frontiers in Artificial Intelligence, vol. 7, 2024, 1355287. https://doi.org/10.3389/frai.2024.1355287.

Rodrigo, Luis. “Introductory Chapter: Osteoporosis Overview.” Osteoporosis–Recent Advances, New Perspectives and Applications, IntechOpen, 2021. https://doi.org/10.5772/intechopen.97760.

Salari, Nader, et al. “The Global Prevalence of Osteoporosis in the World: A Comprehensive Systematic Review and Meta-Analysis.” Journal of Orthopaedic Surgery and Research, vol. 16, 2021, 609. https://doi.org/10.1186/s13018-021-02772-0.

Shim, Jae-Geum, et al. “Application of Machine Learning Approaches for Osteoporosis Risk Prediction in Postmenopausal Women.” Archives of Osteoporosis, vol. 15, 2020, 169. https://doi.org/10.1007/s11657-020-00802-8.

Shuid, Ahmad Nazrun, et al. “A Systematic Review of the Outcomes Of Osteoporotic Fracture Patients after Hospital Discharge: Morbidity, Subsequent Fractures, And Mortality.” Therapeutics and Clinical Risk Management, Nov. 2014, p. 937, https://doi.org/10.2147/tcrm.s72456. Accessed 11 June 2026.

Smets, Jef, et al. “Machine Learning Solutions for Osteoporosis–A Review.” Journal of Bone and Mineral Research, vol. 36, no. 5, 2021, pp. 833–851. https://doi.org/10.1002/jbmr.4292.

Umapathy, V. R., et al. “Perspective of Artificial Intelligence in Disease Diagnosis: A Review of Current and Future Endeavours in the Medical Field.” Cureus, vol. 15, no. 9, 2023, e45684. https://doi.org/10.7759/cureus.45684.

Wändell, Per, et al. “A Machine Learning Tool for Predicting Newly Diagnosed Osteoporosis in Primary Healthcare in the Stockholm Region.” Scientific Reports, vol. 15, 2025, 36472. https://doi.org/10.1038/s41598-025-24450-5.

Comparative Density Functional Theory Study of Frontier Orbital Energies and Global Reactivity Descriptors of Selected Neurotransmitters in Gas and Aqueous Phases

By Saumya Samir Sheth

Abstract

A comparative Density Functional Theory (DFT) investigation was conducted on three biologically significant neurotransmitters, Dopamine, Serotonin, and Acetylcholine, to elucidate their electronic structure and reactivity patterns. Geometry optimizations and frequency calculations were performed using the B3LYP exchange-correlation functional in conjunction with the 6-31G basis set. Solvent effects were incorporated using the Polarizable Continuum Model to simulate aqueous conditions. Frontier molecular orbital energies were employed to compute HOMO-LUMO gaps, while global reactivity descriptors such as chemical hardness, chemical potential, and electrophilicity were derived using Koopmans' approximation. The results reveal that serotonin exhibits the largest HOMO-LUMO gap in the gas phase, indicative of enhanced thermodynamic stability due to extended π-conjugation. Dopamine displays intermediate electronic behavior associated with its catechol functionality, whereas acetylcholine exhibits distinct properties arising from its non-aromatic and charged structure. Solvent effects significantly stabilize orbital energies and alter reactivity descriptors, particularly for acetylcholine. These findings establish clear structure-property relationships and demonstrate the applicability of conceptual DFT in understanding the electronic behavior of neurotransmitters in biologically relevant environments.

Keywords Density Functional Theory, Polarizable Continuum Model, Koopman's approximation, HOMO-LUMO, Neurotransmitters, Dopamine, Serotonin, Acetylcholine

1. Introduction

Density Functional Theory (DFT) has emerged as one of the most widely used computational methodologies for investigating the electronic structure of atoms and molecules due to its favorable balance between computational efficiency and accuracy[1], [2]. In contrast to wavefunction-based approaches, which require explicit treatment of the many-electron wavefunction and scale unfavorably with system size, DFT reformulates the electronic structure problem in terms of electron density. The Hohenberg-Kohn theorems establish that the ground-state energy of a many-electron system is uniquely determined by its electron density, while the Kohn-Sham formalism provides a practical framework for solving the associated equations[1], [2]. This approach enables reliable prediction of structural, energetic, and electronic properties for a wide range of molecular systems, including biologically relevant compounds.

Neurotransmitters constitute a chemically diverse class of small molecules that mediate signal transmission across synapses in the nervous system[3]. Their molecular structure plays a critical role in determining their reactivity, stability, and interaction with biological targets.

Among these, Dopamine, Serotonin, and Acetylcholine represent three structurally distinct classes. Dopamine belongs to the catecholamine family and contains a benzene ring substituted with two hydroxyl groups, which enhance electron density through resonance effects[4]. Serotonin contains an indole ring system, characterized by extended π-conjugation that facilitates electron delocalization[5]. In contrast, acetylcholine lacks aromaticity and consists of an aliphatic framework with a quaternary ammonium group, resulting in a localized electronic structure and a permanent positive charge[6].

The chemical reactivity of molecular systems can be effectively interpreted using Frontier Molecular Orbital theory, which emphasizes the role of the highest occupied molecular orbital and the lowest unoccupied molecular orbital in governing electron transfer processes[7]. The HOMO represents the ability of a molecule to donate electrons, whereas the LUMO represents its capacity to accept electrons. The energy difference between these orbitals, commonly referred to as the HOMO-LUMO gap, serves as an important descriptor of molecular stability and reactivity. Molecules with smaller energy gaps are generally more reactive due to the ease of electronic excitation, while larger gaps are associated with increased stability. Conceptual DFT extends this framework by introducing global reactivity descriptors such as chemical hardness, chemical potential, and electrophilicity index[8]. Chemical hardness measures resistance to charge transfer, chemical potential reflects the escaping tendency of electrons, and electrophilicity quantifies the ability of a system to accept electrons. These descriptors provide a quantitative basis for comparing molecular reactivity across chemically diverse systems.

Previous computational studies have examined the electronic properties of aromatic and biomolecular systems, demonstrating that extended conjugation enhances stability and reduces reactivity[8]. Similarly, studies on quaternary ammonium compounds have highlighted the importance of solvent effects in stabilizing charged species[9]. Despite these advances, a systematic comparative analysis of structurally distinct neurotransmitters under consistent computational conditions remains limited. In this context, the present study aims to perform a comprehensive comparative analysis of dopamine, serotonin, and acetylcholine using DFT. The study involves geometry optimization, evaluation of frontier molecular orbital energies, and calculation of global reactivity descriptors using Koopmans' approximation[10]. Solvent effects are incorporated using the Polarizable Continuum Model to simulate aqueous environments [9]. By correlating molecular structure with computed electronic properties, this work seeks to provide deeper insight into the behavior of neurotransmitters and demonstrate the utility of conceptual DFT in the study of biologically relevant molecules.

2. Computational Methodology

All calculations were performed using Gaussian 09, and molecular structures were visualized using GaussView [11], [12]. Geometry optimizations were carried out using the B3LYP exchange-correlation functional, which combines Becke's three-parameter exchange functional with the Lee-Yang-Parr correlation functional [13], [14]. This functional has been widely validated for organic systems and provides reliable predictions of molecular geometries

and electronic properties. The 6-31G basis set was employed for all calculations. This split-valence double-zeta basis set, developed by Pople and co-workers, provides a practical balance between computational efficiency and accuracy[8], [15]. Although larger basis sets incorporating polarization and diffuse functions may yield improved quantitative accuracy, the 6-31G basis set is sufficient for comparative studies focused on identifying trends in electronic structure across related molecules. Frequency calculations were performed at the same level of theory to confirm that the optimized geometries correspond to true minima on the potential energy surface, as indicated by the absence of imaginary frequencies. Solvent effects were modeled using the Polarizable Continuum Model with water as the dielectric medium, allowing simulation of physiological conditions[9]. Frontier molecular orbital energies were extracted from the optimized geometries. Global reactivity descriptors were computed using Koopmans' approximation, wherein ionization potential (I) and electron affinity (A) are approximated as $-E_{HOMO}$ and $-E_{LUMO}$, respectively [10]. From these quantities, chemical hardness (η), chemical potential (μ), and electrophilicity index (ω) were derived using standard conceptual DFT expressions.

3. Results and Discussion

The optimized geometries highlight the intrinsic structural diversity of the neurotransmitters (**Figure 1**). Serotonin exhibits a near-planar conformation due to its fused indole system, which contains ten π-electrons satisfying aromaticity criteria. This extended conjugation facilitates significant electron delocalization across the molecular framework. Dopamine displays a planar benzene ring substituted with two hydroxyl groups in an ortho configuration, forming a catechol system. These substituents enhance electron density through resonance donation, increasing the nucleophilic character of the aromatic ring. In contrast, acetylcholine lacks any π-conjugated system and consists of a flexible aliphatic chain terminated by a quaternary ammonium group. This results in localized electron density and reduced orbital delocalization. These structural differences are expected to play a decisive role in determining electronic properties and reactivity (**Table 1**).

The spatial distribution of frontier molecular orbitals provides insight into electronic reactivity (**Figure 2**). Serotonin exhibits extensive delocalization of both HOMO and LUMO across the indole system, indicating strong electron-donating and accepting capabilities. Dopamine shows HOMO localization primarily on the catechol ring, with moderate delocalization. Acetylcholine displays highly localized frontier orbitals associated with specific functional groups, reflecting limited electronic communication.

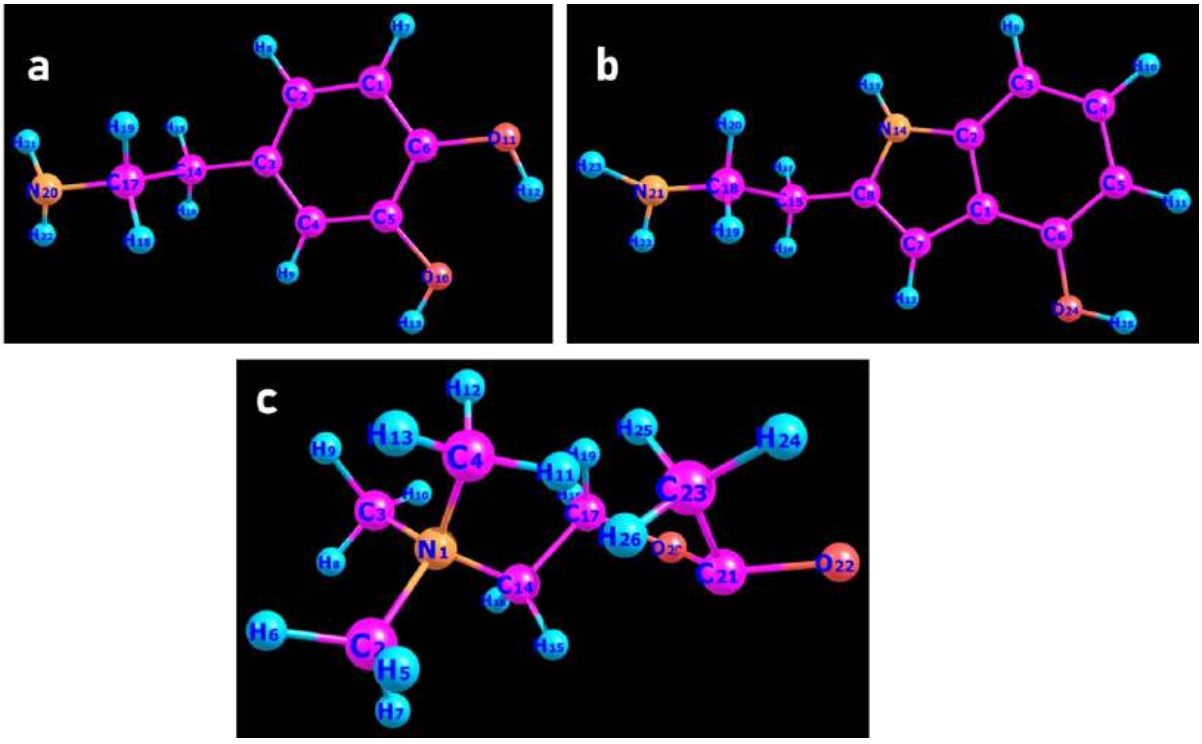


Figure 1. Optimised geometries of (a) Dopamine, (b) Serotonin, and (c) Acetylcholine in gas phase; Method/Basis set: B3LYP/6-31G

Table 1. Bond lengths and Bond angles of optimised geometries of Dopamine, Serotonin, and Acetylcholine in the gas phase

DOPAMINE			
BOND LENGTH (Å)		BOND ANGLE (°)	
N20-H22	1.01231	N20-C17-C14	115.4515
C17-N20	1.45877	H22-N20-H21	112.6328
C17-C14	1.55817	H19-C17-H18	106.2357
C14-H15	1.09934	C17-C14-C3	113.0501
C14-C3	1.51415	H8-C2-C3	119.8568
C5-O10	1.40245	H13-O10-C5	112.555
O10-H13	0.97995		
C-2H8	1.0858		
SEROTONIN			
BOND LENGTH (Å)		BOND ANGLE (°)	
C6-O24	1.39406	H22-N21-H23	112.225
C2-N14	1.38867	N21-C15-C18	110.3879
C8=C7	1.37801	H19-C18-H20	106.4486

N14-H13	1.00619	C8=C7-H12	126.0577
C15-C8	1.49714	H13-N14-C8	124.7078
C18-C15	1.54749	H9-C3-C2	121.7016
C18-N21	1.46208	H25-O24-C5	111.2061
N21-H22	1.01355		
ACETYLCHOLINE			
BOND LENGTH (Å)		BOND ANGLE (°)	
C10-H11	1.09067	O22=C21-020	116.4307
C10-N1	1.52322	C17-C14-H15	109.00296
C17-C14	1.53214	C21-O20-C17	123.38209
C17-O20	1.45249	H19-C17-H18	109.1694
O20-C21	1.42328	O20-C17-C14	100.4089
C21=O22	1.21708	C17-C14-N1	116.3489
C21-C23	1.50464	C2-N1-C10	108.706

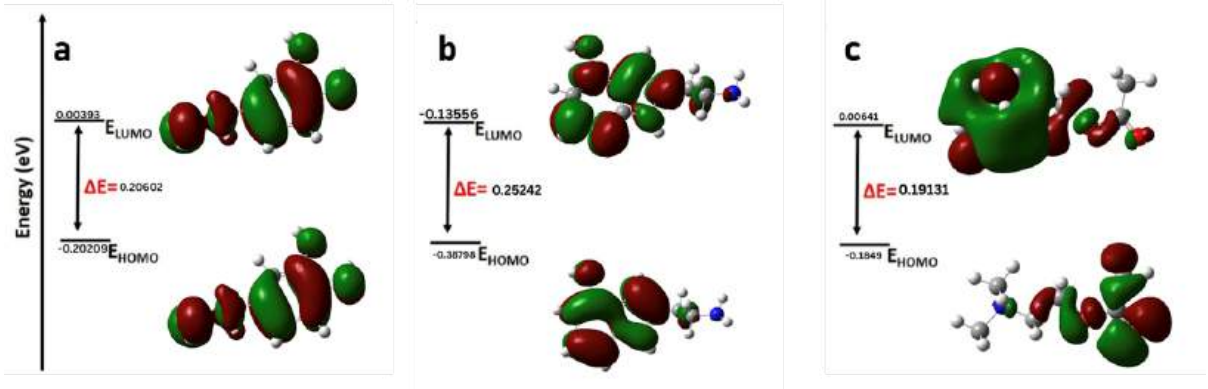


Figure 2. HOMO-LUMO diagrams of the optimised geometries of (a) Dopamine, (b) Serotonin, and (c) Acetylcholine in the gas phase. Method/Basis set: B3LYP/6-31G

Gas phase calculations reveal that serotonin has the largest HOMO-LUMO gap (0.25242), followed by dopamine (0.20602) and acetylcholine (0.19131) (**Table 2**). This trend is consistent with the degree of conjugation present in each molecule, with extended π-systems contributing to increased stability. In the solvent phase, acetylcholine exhibits a significant increase in the energy gap (0.27443), indicating strong stabilization in aqueous environments (**Table 3**). Dopamine shows minimal change, while serotonin exhibits a slight decrease in the energy gap, suggesting increased polarizability.

Table 2. HOMO-LUMO energies and band gap (ΔE) of the neurotransmitters in the gas phase

Molecule	**E_{HOMO} (Hartrees)**	**E_{LUMO} (Hartrees)**	**ΔE (Hartrees)**
Dopamine	-0.20209	0.00393	0.20602
Serotonin	-0.38798	-0.13556	0.25242
Acetylcholine	-0.18490	0.00641	0.19131

Table 3. HOMO-LUMO energies and band gap (ΔE) of the neurotransmitters in the solvent phase (CPCM/water)

Molecule	**E_{HOMO} (Hartrees)**	**E_{LUMO} (Hartrees)**	**ΔE (Hartrees)**
Dopamine	-0.21321	-0.00497	0.20824
Serotonin	-0.19468	-0.00310	0.19158
Acetylcholine	-0.29549	-0.02106	0.27443

Serotonin exhibits the highest chemical hardness and electrophilicity in the gas phase, indicating strong resistance to charge transfer and enhanced electron-accepting ability (**Table 4**). Dopamine displays intermediate values, while acetylcholine shows lower electrophilicity. In the solvent phase, acetylcholine exhibits increased hardness and electrophilicity, reflecting stabilization of its charged structure (**Table 5**). These findings demonstrate the significant influence of solvation on electronic properties.

Table 4. The global reactivity descriptors for the neurotransmitters optimised in the gas phase

Molecule	**Chemical Hardness(η)**	**Chemical Potential (μ)**	**Electrophilicity Index (ω)**
Serotonin	0.12621	-0.26177	0.27147
Dopamine	0.10301	-0.09908	0.04765
Acetylcholine	0.09566	-0.08925	0.04163

Table 5. The global reactivity descriptors for the neurotransmitters optimised in the solvent phase (CPCM/water)

Molecule	**Chemical Hardness (η)**	**Chemical Potential (μ)**	**Electrophilicity Index (ω)**
Dopamine	0.10412	−0.10909	0.05715
Serotonin	0.09579	-0.09889	0.05105
Acetylcholine	0.13722	-0.15828	0.09129

The observed trends can be rationalized based on molecular structure. Extended π-conjugation in serotonin enhances electron delocalization and stability. Dopamine exhibits intermediate properties due to limited conjugation. Acetylcholine's localized electronic structure and charged nature result in distinct behavior, particularly in solvent environments. The results are consistent with previous DFT studies that highlight the role of conjugation in stabilizing aromatic systems[8]. Indole-based systems have been shown to exhibit enhanced stability due to π-delocalization, while catechol systems display moderate reactivity influenced by hydroxyl substitution. Additionally, studies on quaternary ammonium compounds have demonstrated strong solvent stabilization effects in polar environments, supporting the behavior observed for acetylcholine[9].

The B3LYP/6-31G level of theory employed in this study provides reliable qualitative trends but may not fully capture electron correlation effects with quantitative accuracy[13]. The absence of polarization and diffuse functions in the basis set may limit accuracy for systems involving charge separation. Koopmans' approximation neglects orbital relaxation effects, which may introduce deviations in computed ionization potentials and electron affinities[10]. Furthermore, the Polarizable Continuum Model accounts for bulk solvent effects but does not include explicit solute-solvent interactions such as hydrogen bonding[9]. Despite these limitations, the chosen methodology represents a reasonable compromise between computational cost and accuracy for comparative analysis. Future work may involve the use of higher-level functionals, larger basis sets, explicit solvation models, and experimental validation.

4. Conclusions

The present study provides a comprehensive comparative analysis of the electronic structure and reactivity of Dopamine, Serotonin, and Acetylcholine using DFT within both gas-phase and solvated environments. The results clearly demonstrate that DFT-based approaches, particularly when combined with conceptual reactivity descriptors, offer a powerful framework for understanding the fundamental electronic behavior of biologically relevant molecules.

A consistent relationship between molecular structure and electronic properties is observed across all three systems. Serotonin exhibits the highest HOMO-LUMO energy gap and chemical hardness in the gas phase, which can be attributed to its extended π-conjugated indole framework that promotes efficient electron delocalization and enhances intrinsic stability. Dopamine, characterized by a catechol moiety, displays intermediate electronic behavior, reflecting a balance between localized and delocalized electronic effects. In contrast, acetylcholine shows markedly different characteristics due to the absence of π-conjugation and the presence of a permanently charged ammonium group, resulting in localized frontier orbitals and distinct reactivity patterns. The inclusion of solvent effects further reveals the critical role of the molecular environment in modulating electronic properties. While aromatic systems such as serotonin and dopamine show relatively modest changes upon solvation, acetylcholine exhibits significant stabilization and an increase in its HOMO-LUMO gap in the aqueous phase. This

highlights the strong influence of electrostatic interactions and solvent polarization on charged, non-aromatic systems, emphasizing the necessity of incorporating solvation effects when modeling biologically relevant molecules.

Overall, the study underscores that conjugation length, functional group identity, and charge distribution are the primary determinants of electronic structure and reactivity. The application of conceptual DFT descriptors enables a quantitative interpretation of these effects, providing deeper insight into electron transfer tendencies and molecular stability. These findings not only enhance the understanding of neurotransmitter chemistry at the molecular level but also demonstrate the broader applicability of DFT-based approaches in the investigation of complex biochemical systems. Future studies incorporating higher-level computational methods, explicit solvation models, and experimental validation may further refine these insights and extend their relevance to neurochemical processes, drug design, and biosensing applications.

Works Cited

[1]P. Hohenberg and W. Kohn, “Inhomogeneous Electron Gas,” Physical Review, vol. 136, no. 3B, pp. B864–B871, Nov. 1964, doi: 10.1103/PhysRev.136.B864.

[2]W. Kohn and L. J. Sham, “Self-Consistent Equations Including Exchange and Correlation Effects,” Physical Review, vol. 140, no. 4A, pp. A1133–A1138, Nov. 1965, doi: 10.1103/PhysRev.140.A1133.

[3]S. H. Snyder and C. D. Ferris, “Novel Neurotransmitters and Their Neuropsychiatric Relevance,” American Journal of Psychiatry, vol. 157, no. 11, pp. 1738–1751, Nov. 2000, doi: 10.1176/appi.ajp.157.11.1738.

[4]M. L. Jabbar, “Computational studies on electronic and optical properties of dopamine derivatives structure: A DFT study **,” J. Mech. Behav. Mater., vol. 30, no. 1, pp. 279–284, Jan. 2021, doi: 10.1515/jmbm-2021-0030.

[5]A. R. Meek, A. Morzycki, M. D. Carter, C. Barden, and D. F. Weaver, “Serotonin as a drug receptor for Li^{+}: a computational study,” Can. J. Chem., vol. 96, no. 7, pp. 694–702, Jul. 2018, doi: 10.1139/cjc-2017-0665.

[6]G. Serdaroğlu, “A DFT study of determination of the reactive sites of the acetylcholine and its agonists: In the gas phase and dielectric medium,” Int. J. Quantum Chem., vol. 111, no. 10, pp. 2464–2475, Aug. 2011, doi: 10.1002/qua.22512.

[7]K. Fukui, “Role of Frontier Orbitals in Chemical Reactions,” Science (1979)., vol. 218, no. 4574, pp. 747–754, Nov. 1982, doi: 10.1126/science.218.4574.747.

[8]P. Geerlings, F. De Proft, and W. Langenaeker, “Conceptual Density Functional Theory,” Chem. Rev., vol. 103, no. 5, pp. 1793–1874, May 2003, doi: 10.1021/cr990029p.

[9]J. Tomasi, B. Mennucci, and R. Cammi, “Quantum Mechanical Continuum Solvation Models,” Chem. Rev., vol. 105, no. 8, pp. 2999–3094, Aug. 2005, doi: 10.1021/cr9904009.

[10]T. Koopmans, “Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms,” Physica, vol. 1, no. 1–6, pp. 104–113, Jan. 1934, doi: 10.1016/S0031-8914(34)90011-2.

[11]W. J. Hehre, R. Ditchfield, and J. A. Pople, “Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules,” J. Chem. Phys., vol. 56, no. 5, pp. 2257–2261, Mar. 1972, doi: 10.1063/1.1677527.

[12]P. C. Hariharan and J. A. Pople, “The influence of polarization functions on molecular orbital hydrogenation energies,” Theor. Chim. Acta, vol. 28, no. 3, pp. 213–222, 1973, doi: 10.1007/BF00533485.

[13]A. D. Becke, “Density-functional thermochemistry. III. The role of exact exchange,” J. Chem. Phys., vol. 98, no. 7, pp. 5648–5652, Apr. 1993, doi: 10.1063/1.464913.

[14]C. Lee, W. Yang, and R. G. Parr, “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density,” Phys. Rev. B, vol. 37, no. 2, pp. 785–789, Jan. 1988, doi: 10.1103/PhysRevB.37.785.

[15] R. G. Parr, L. v. Szentpály, and S. Liu, “Electrophilicity Index,” J. Am. Chem. Soc., vol. 121, no. 9, pp. 1922–1924, Mar. 1999, doi: 10.1021/ja983494x.

Design and Implementation of a Course Selection-Based Scheduling System with Satisfaction By Fengxue Bao[1], Yifan Yang[2], Can Wang[3]

1. Introduction

1.1 Research Background

With the continuous reform of teaching systems in universities and secondary schools, the course selection system has gradually replaced the traditional scheduling model, becoming a more flexible and student-led mode of course arrangement. Under this system, students can select several courses based on their own interests, study plans, and future development goals. However, as the number of students increases and the variety of courses grows, how to satisfy students' personalized course selection needs as much as possible under limited teaching resources has become a major challenge in school academic administration. The course scheduling problem is essentially a resource optimization problem characterized by complex constraints and a large solution space. Furthermore, the flexibility introduced by the course selection system significantly increases the complexity of scheduling. Currently, there are several types of course scheduling algorithms, which include the following:

An exact solution algorithm guarantees the optimal solution to a problem along with a rigorous proof of optimality. While it is well-suited for small-scale course scheduling problems, it can incur computational times ranging from several hours to several days when confronted with large-scale instances. A typical representative of exact solution algorithms is mixed integer programming: this type of algorithm transforms the course scheduling problem into a linear or nonlinear integer programming model, from which a solver is then used to derive the globally optimal solution. Papers[1], [2] and [3] have applied this method to solve university course scheduling problems.

Heuristic algorithms are advantageous primarily due to their computational efficiency and scalability. Within an acceptable time limit, they can produce feasible solutions of satisfactory quality, and their computation time does not grow exponentially with increasing problem scale. Representative examples of such algorithms include matheuristics, ant colony optimization, particle swarm optimization, tabu search, and genetic algorithms.

Matheuristics refer to algorithms that integrate exact solution techniques with heuristic search strategies. A matheuristic that can accommodate various course scheduling needs is proposed in Paper[4]. Paper[5] designs a fix-and-optimize matheuristic, while Paper[6] introduces a parallelized version of matheuristic algorithms.

Ant colony optimization (ACO) algorithms address the course scheduling problem by mimicking the positive feedback mechanism through which ants deposit pheromones. In this approach, each ant constructs a complete timetable based on pheromone concentrations and heuristic information, followed by an update of the pheromone matrix. An adaptive heuristic ACO is proposed in Paper[7], where time-slot weight coefficients are integrated into the pheromone update formula. Paper[8] directly applies the basic ACO algorithm to university course

scheduling, while Paper[9] introduces a greedy ACO strategy that embeds greedy rules into the solution construction process of ants.

The particle swarm optimization (PSO) algorithm addresses course scheduling through the simulation of flocking and foraging behavior observed in bird populations. Each particle corresponds to a candidate timetable solution; its position is updated according to both its own personal best and the global best found by the swarm. A PSO-based hyper-heuristic method is proposed in Paper[10]. Paper[11] adopts an improved PSO with chaotic factor (IWCFPSO). Paper[12] provides a comparative analysis of multiple PSO variants, including inertia weight PSO and constriction factor PSO, along with their hybridization ratios with local search. Paper[13] integrates a genetic algorithm with PSO and embeds the hybrid model into a hyper-heuristic framework.

Tabu search is a local search-based algorithm that utilizes a tabu list to keep track of recently visited solutions or moves, thus preventing search cycles. It also permits the acceptance of inferior solutions to facilitate escape from local optima. An adaptive tabu search is proposed in Paper[14]. Paper[15] examines how different neighborhood structures—i.e., various timetable modification operations such as moving a single course or swapping two courses—affect the performance of tabu search. Paper[16] introduces a tabu search variant incorporating controlled randomization.

Genetic Algorithm[17] (GA), a random search algorithm inspired by biological evolutionary processes, is characterized by robust global search ability and strong adaptability to complex constrained optimization problems. These features make it particularly well-suited for broad adoption in course scheduling systems. Nevertheless, limited research has been devoted to designing satisfaction functions and genetic algorithms that account for the distinctive features of "preference priority" and "course selection satisfaction" within elective systems.

This study proposes and implements a genetic algorithm-based scheduling system for elective course systems, with the objective of maximizing student satisfaction with course selection. The system takes students' course selection as input and the degree of fulfillment of the courses chosen by students serves as the satisfaction function of the genetic algorithm, optimizing the timetable through evolutionary processes.

1.2 Definition of Key Terms

The following explains the relevant definitions used in the algorithm process:

1) Discrete Time Slots (DTS): The integration of multiple non-consecutive class periods into one time slot. Multiple DTSs will cover all class periods within a week, with the constraint that each class period can only occur once in all DTSs.
2) Course Instance: It is defined by three factors: teacher, class period, and course name. Any difference in these components causes a different course instance.
3) Chromosome: A class timetable.
4) Gene: A course instance in a timetable.

2. Algorithm Implementation

2.1 Procedure of Algorithm

The algorithmic procedure is as follows:

Step(1)Obtain the course configuration data and define the structure of the chromosome.

Step(2)Insert all course instances randomly into the chromosome structure to generate an individual chromosome, whose validity is then checked.

Step(3)Repeat Step (2) to form the initial population.

Step(4)Compute the satisfaction of each chromosome according to each student's satisfaction weight.

Step(5)Select Superior chromosomes via ranked roulette wheel selection.

Step(6)Apply crossover and mutation operations to the selected superior chromosomes.

Step(7)Iterate Steps (4) through (6) until the required satisfaction level is achieved.

2.2 Chromosome Structure Definition

This study innovatively proposes a class period organization method called Discrete Time Slots (DTS) to reduce the complexity of the algorithm for Course Selection-Based scheduling systems. In traditional scheduling systems, a single consecutive class period is typically seen as the minimal time unit(the first period on Monday). When constructing a schedule, each class period serves as a resource dimension for evaluating constraints; this approach often results in high complexity and slow convergence. To address this limitation, DTS reorganizes the available class periods within a week to construct a data structure that covers all class periods across the complete teaching week. Specifically:

1) Definition of Discrete Time Slots (DTS)

Assume there are K available class periods over a week. Separate these class periods into T discrete time slots--$\{DTS_1, DTS_2, DTS_3, ..., DTS_T\}$. Each DTS is composed of several non-consecutive class periods. In the present algorithm implementation, there are 35 available class periods in total (five teaching days per week, seven periods per day); they are classified into 7 DTSs. For instance, expressed as: $DTS_1 = \{MON_{1\text{-}2}, WED_{5\text{-}6}, FRI_4\}$.

The partition of DTS can be customized according to the school's teaching calendar. Integrating multiple individual periods from teaching days across a week into several DTSs effectively reduced the scale of time resources in the scheduling process. Additionally, the class periods within each DTS should be distributed evenly throughout the week to prevent excessive concentration. Such design effectively avoids the issues of uneven course distribution.

2) Chromosome Encoding

Two-Dimensional Timetable Matrix $G\{T, R\}$ Constructed According to the DTS Framework, in which:

- Row dimension: T numbers of DTSs (temporal resources)
- Column dimension: R classrooms (spatial resources)

Each element in the matrix *g(t, r)* represents the course instance C_n allocated to the r-th classroom in the t-th DTS, as shown in Table 1.

DTS\classroom	***R1***	***R2***	***...***	***Rr***
DTS1	*C5*	*C12*	*...*	∅
DTS2	∅	*C8*	*...*	*C3*
...	*...*	*...*	*...*	*...*
DTSt	*C9*	∅	*...*	*C10*

Table 1 A Two-Dimensional Matrix Example of a Timetable Chromosome

Course instances are sequentially and randomly selected from the course instance set and inserted into the matrix. In each placement, priority is given to the DTS that currently contains the smallest number of assigned course instances, while the classroom is selected at random. Additionally, a constraint must be enforced such that no two distinct course instances placed in the same DTS share the same teacher.

For course instances whose weekly class hours are either fewer or greater than the number of periods encompassed by a single DTS, the following handling strategy may be adopted:

- In the former scenario, the course instance may be placed directly into a DTS, and the specific class periods for its sessions can be specified in accordance with the specific teaching plans.
- In the latter scenario, where the required weekly class hours exceed those of a single DTS, multiple DTSs should be bound together to form a DTS combination. During subsequent evolutionary processes—including crossover, mutation, and selection—these grouped DTSs must be handled as a whole.

This matrix serves as the timetable chromosome and will be subject to evolutionary operations in subsequent steps. The proposed chromosome structure effectively eliminates conflicts involving teachers, time slots, and classrooms, thus greatly mitigating the problems caused by resource constraints throughout the evolutionary process. Nevertheless, it is worth noting that, during random chromosome generation, certain course instances may prove impossible to insert into the chromosome. When this occurs, the chromosome should be discarded and randomly regenerated, or the procedure should backtrack to an earlier step and substitute a different course instance.

2.3 Satisfaction Calculation

Because the proposed chromosome encoding resolves conflicts among temporal, instructional, and spatial resources, satisfaction is computed during evolution according to students' course selection priority. To this end, this study introduces a satisfaction function based on a weighted course selection priority. This function integrates both the priority levels assigned by students to their chosen courses and the occurrence of course conflicts, thereby ensuring that

courses with higher preference are satisfied first. In this manner, the genetic algorithm is effectively directed toward solutions characterized by high satisfaction.

2.3.1 Weight Allocation for Student Course Selections

Suppose that each student selects N candidate course instances from the overall course set (for example, N = 7) and ranks them according to their personal preferences. Weights are assigned to courses in descending order of priority. Although a linearly decreasing principle can be adopted for weight assignment, it fails to adequately accentuate the differences in desirability between adjacent courses. To more accurately capture the psychological preference patterns inherent in student course selection behavior—wherein the initial courses constitute high-priority core choices and the latter courses serve as fallback options—this study proposes a piecewise weight function. With N = 7, the first three courses are designated as high-priority courses and the remaining four as secondary selections. The specific formula for weight calculation is as follows:

$$w_k = \{\alpha \bullet e^{-\beta(k-1)} + \gamma \quad 1 \leq k \leq 3 \; \delta \bullet e^{-\eta(k-4)} \qquad 4 \leq k \leq 7$$

The parameters are defined as follows:

- k denotes the position of the course in the student's ranked selection list, where k=1k=1 indicates the highest priority.
- w_k represents the weight assigned to the k-th course.
- α is the weight coefficient for the front segment, with a value of 3.5.
- β is the decay coefficient for the front segment, with a value of 0.1.
- γ is the lower bound of weights for the front segment, with a value of 1.5.
- δ is the weight coefficient for the rear segment, with a value of 2.0.
- η is the decay coefficient for the rear segment, with a value of 0.8.

According to this formula, the weights assigned to the seven courses are presented in Table 2.

course k	1	2	3	4	5	6	7
weight w_k	5	4.67	4.37	2	0.90	0.40	0.18

Table 2 Weight Distribution of Expected Preferences for 7 Elective Courses

From the weight distribution, it can be observed that the weights of the first three courses differ by approximately 0.3 to 0.4, reflecting a strong inclination toward these three courses with no significant preference differentiation among them. The weight of the fourth course drops sharply by 54% compared to the third course. The decay coefficient for the rear segment η is notably larger than that for the front segment β, which ensures that the weights undergo exponential decay and rapidly approach zero.

This pattern of weight allocation is evident not only in the course selection process but

also in behavioral economics, where individuals typically demonstrate an exponential decay in their valuation of prioritized alternatives.

2.3.2 Conflict Resolution and Computation of Effective Weights

For a given chromosome, a specific timetable, there may exist several candidate courses in each student's selection list that are assigned to the same DTS. This situation is defined as a course conflict, and these conflicting courses are referred to as a conflict set. For each conflict set, only the course with the highest weight is retained and included in the satisfaction, while all other conflicting courses are considered unsatisfied. The personal satisfaction of student i is then obtained by summing the weights of all non-conflicting courses together with the highest-weight course from each conflict set. The specific calculation formula is as follows:

$$S_i = \sum_{j \in c1} w_j + \sum_{c2} max(w_{j1}, w_{j2}, \ldots)$$

The parameters are defined as follows:
c1: course with no conflict
c2: conflict course combination

2.3.3 Satisfaction Function Definition

The satisfaction function is ultimately defined as the ratio of the total sum of effective weights across all students to the theoretical maximum total weight, expressed as:

$$Satisfaction = \frac{\sum_{i=1}^{M} S_i}{\sum_{i=1}^{M} \sum_{k=1}^{N} w_k^i}$$

In the above expression, *M* represents the total number of students, *N* denotes the number of candidate courses (set to 7 in this study), and w_k^i is the weight assigned to the k-th course selected by the i-th student. The satisfaction value lies within the interval [0, 1]. The closer the satisfaction value is to 1, the higher the course selection satisfaction achieved by the corresponding chromosome.

2.4 Population Iteration Process

Among the population models commonly used in genetic algorithms are the steady-state model and the generational model. The generational model replaces the entire parent population with offspring in each generation, which may result in the loss of high-satisfaction gene chains from the parent generation. Moreover, offspring individuals may have lower satisfaction than their parents, causing oscillation in the evolutionary trajectory. To address these issues, the proposed system employs the steady-state model for population iteration. The procedure is as follows: an offspring population of equal size to the parent population is generated. The parent

and offspring populations are then combined and sorted according to their satisfaction. Chromosomes with higher satisfaction are retained to form the new population, while chromosomes with lower satisfaction are eliminated(the total population size is fixed). This approach ensures that the population is continuously improving without falling into local optima.

The termination condition for population iteration is typically defined as either reaching approximately 200 generations, or when the population satisfaction ceases to change over a certain number of consecutive generations—set to 50 generations in this study. In the latter case, it is assumed that no further improvement in offspring can be expected, and the iteration is terminated prematurely.

2.5 Crossover and Mutation Operations

The purpose of crossover is to combine the traits of different parents, preserving their favorable genes and introducing new gene combinations simultaneously. The mutation operation, on the other hand, introduces new gene combinations through a given random strategy, which helps to escape local optima and enhance the diversity of the population.

2.5.1 Definition and Initialization

G_1, $G_2 \in G$: The parent chromosomes selected from the population after applying a combined selection strategy (ranked roulette wheel selection and elite selection strategy).

N_1, N_2: The offspring chromosomes generated by crossover from the parent chromosomes G_1 and G_2 , with initial values equal to G_1 and G_2, respectively.

2.5.2 Crossover Process for Producing Offspring from Parents

Two distinct types of crossover operation are implemented. The first type facilitates efficient inheritance of superior genetic material from parents to offspring, whereas the second type adopts a randomized crossover strategy to prevent the algorithm from becoming trapped in local optima. Both implementations are informed by the Partially Matched Crossover[18] (PMX) method designed for order-based encoding, as shown in Figure 1.

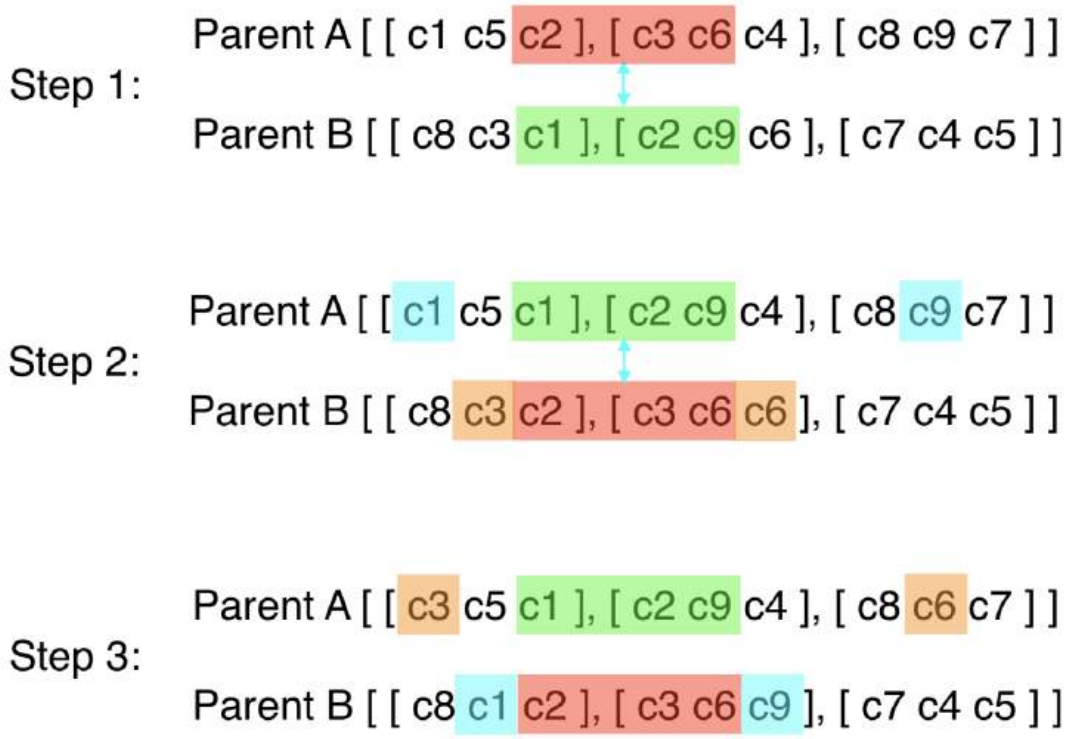


Figure 1 PMX illustration

As illustrated in Figure 1, parent A and parent B are each composed of three DTSs. Each DTS contains three course instances. Both chromosomes contain exactly the same set of course instances(c1~c9) but are different in arrangement. Figure 1 shows the process two chromosomes exchanging the genes *{c2, c3, c6}* and *{c1, c2, c9}*.

Step 1: The genes in the colored region are exchanged(red and green region).

Step 2: Identify the redundant courses. course *c1* and *c9* are duplicated in parent A (blue region); course *c3* and *c6* are duplicated in parent B (orange region).

Step 3: Remove duplicated courses in the two chromosomes. Then, courses that became missing due to the crossover are inserted back into the chromosome. Parent A lacks *c3* and *c6*; parent B lacks *c1* and *c9*. Eventually, two offspring chromosomes are generated.

This algorithm sets a crossover rate, which has a negative correlation with the number of generations. Before crossover operation, a random number is generated to determine which crossover type to apply. If the random number is greater than the current crossover rate, the first crossover type is executed; otherwise, the second type is performed. The generation of offspring chromosome N_1 is used below as an illustrative example.

$$current\ P_c = inital\ P_c - (inital\ P_c - end\ P_c) * (C_t / end_t)$$

C_t represents the current evolutionary generation; *current* P_c denote the crossover rate at

generation C_t ; *initial* P_c is the initial crossover rate (set to 0.3 in this algorithm); *end* P_c is the target crossover rate (set to 0.8); end_tindicates the approximate number of generations needed for the population satisfaction to stabilize, which is set to 200 in this algorithm.

2.5.2.1 The detailed procedure for the first crossover type is described as follows:

S_1: The gene segment within parent chromosomes G_1 and G_2 that is designated for exchange with the corresponding segment in the other parent.

S_2: The gene segment within the offspring chromosome that is subject to replacement.

Step 1: The satisfaction of each course in parent chromosome G_2 is computed according to the courses selected by students. Only those courses that students successfully enroll in without conflict and in alignment with their stated preferences are included in this calculation. Subsequently, a random integer L is generated within the range of 10 to 15. A sliding window algorithm is then employed to identify a contiguous segment of length L, denoted S_1, within the chromosome that yields the maximum total satisfaction among all possible contiguous segments of that length. As illustrated in Figure 2, the green segment indicates the contiguous interval in chromosome G_2 that attains this maximum total satisfaction.

DTS\classroom	R1	R2	R3	...	Rr
DTS1	C1	C4	C7	...	C10
DTS2	C2	C5	C8	...	C11
...	...	...	...	...	...
DTSt	C3	C6	C9	...	C12

Figure 2 An Example of the S_1 Segment within Chromosome G_2

The sliding window algorithm reduces the computational complexity from $O(n \times L)$ to $O(n)$, where L denotes the length of the target segment and n represents the total number of courses in the chromosome. The algorithmic steps are as follows:

First, the total satisfaction of courses in the initial window of length L is computed, and the starting index of this first interval is recorded.

Next, the remaining courses from index $L+1$ to n are iterated; the window is shifted incrementally--one step at a time. As shown in Figure 3, when the window slides one position to the right, the satisfaction of the course *c1* is subtracted(the course that is leaving the window), while the satisfaction of the course *c6* is added(the course that is entering the window), yielding the total satisfaction of the new window. If the total satisfaction of the new window is greater than (or smaller than) the current maximum (or minimum) satisfaction, the starting index is updated.

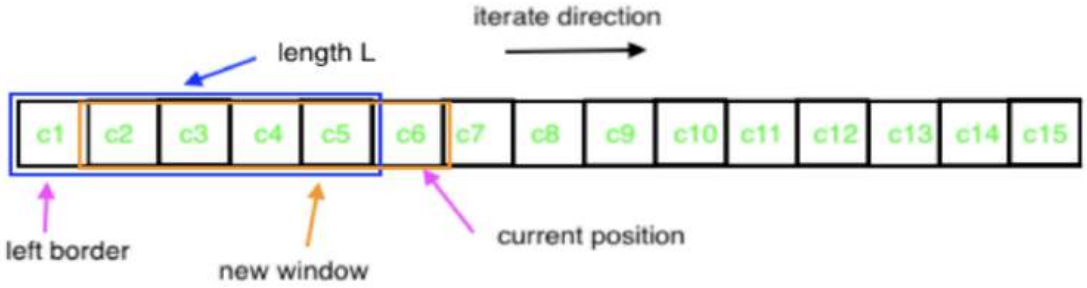


Figure 3 An Example of the Sliding Window Algorithm

Step 2: For parent chromosome G_1, the satisfaction of courses is computed with respect to each discrete time slot (DTS) under the same classroom. A sliding window algorithm is subsequently applied to locate a contiguous segment of length L, designated S_2, that minimizes the sum of satisfaction over that segment within the chromosome. As shown in Figure 4, the green segment indicates the contiguous interval in chromosome G_1 that attains this minimum total satisfaction.

DTS\classroom	*R1*	*R2*	*R3*	*...*	*Rr*
DTS1	*C2*	*C5*	*C9*	*...*	*C10*
DTS2	*C3*	*C4*	*C6*	*...*	*C12*
...	*...*	*...*	*...*	*...*	*...*
DTSt	*C1*	*C6*	*C7*	*...*	*C11*

Figure 4 An Example of the S_2 Segment within Chromosome G_1

Step 3: The courses located in the S_2 interval of chromosome N_1 are replaced by the courses from the S_1 interval of chromosome G_2. As illustrated in Figure 5, no further operations are carried out following this replacement.

Step 4: Chromosome N_1 is then examined for conflicting courses, defined as courses that introduce resource conflicts, such as two courses taught by the same teacher being scheduled into the same DTS. The legalization operation is applied to these conflicting courses.

Legalization Operation: This procedure applies to all courses that must be inserted into the chromosome. Each course is examined individually across different DTSs under the same classroom to determine whether it can be placed into the current DTS without introducing any conflicts (either teacher conflicts or classroom conflicts). If it is impossible to place all required courses into the chromosome, the legalization operation is deemed to have failed.

Step 5: Chromosome N_1 is examined for redundant courses—situations where two or more identical genes appear within the chromosome, as illustrated by the blue portions in Figure 5. One instance of each redundant course is randomly retained, while all other duplicate copies are deleted.

DTS\classroom	R1	R2	R3	...	Rr
DTS1	C2	C5	C9	...	C10
DTS2	C3	C4	C7	...	C10
...	...	...	...	...	...
DTSt	C4	C6	C7	...	C11

Figure 5 An Example of Chromosome N_1 Following Steps 3, 4, 5

Step 6: Traverse chromosome N_1 to locate any missing courses—those that ought to appear in the offspring chromosome but are absent as a result of gene segment exchange, replacement, or deletion during the crossover process between the two parent chromosomes. A "legalization operation" is then performed on these missing courses. Upon successful execution of the legalization operation, the final chromosome is obtained, as depicted in Figure 6.

DTS\classroom	R1	R2	R3	...	Rr
DTS1	C2	C5	C9	...	C12
DTS2	C3	C4	C7	...	C10
...	...	...	...	...	...
DTSt	C4	C6	C8	...	C11

Figure 6 Example of Chromosome N_1 Following Crossover

Step 7: If the algorithm fails to successfully place all required courses back into chromosome N_1 (the legalization operation is not successful) during the execution of Steps 4, 5, or 6, then N_1 is discarded. In this case, one fewer offspring chromosome is produced.

The production of offspring chromosome N_2 follows an analogous procedure to the production of N_1. The steps described above are repeated, but replace G_2 with G_1 and G_1 with G_2--the S_1 interval is generated in G_1 during Step 1, and the S_2 interval is generated in G_2 during Step 2. All subsequent operations are then performed on N_2, eventually yielding the offspring chromosome N_2.

2.5.2.2 The detailed procedure for the second crossover type is described as follows:

Step 1: A length L is first generated randomly, followed by a random determination of the starting position. Subsequently, a contiguous interval of length L is extracted using the starting position as a reference, and this interval is designated as the exchange segment S.

Step 2: The courses located in the S interval of chromosome N_1 are replaced by the

courses occupying the corresponding *S* interval in chromosome G_2.

Step 3: Chromosome N_1 is traversed to locate missing courses—those that should appear in the offspring chromosome but are absent as a consequence of gene segment exchange, replacement, or deletion during the crossover process. An attempt is then made to perform a "legalization operation" on these missing courses.

Step 4: Redundant courses in chromosome N_1(cases where two or more identical genes are present) are identified. For each set of redundant courses, one instance is randomly retained, and all remaining duplicate copies are deleted.

Step 5: Conflicting courses in chromosome N_1 are identified. Conflicts are defined as resource conflicts within the chromosome—for example, two courses taught by the same teacher being scheduled into the same DTS. The legalization operation is then performed on these conflicting courses.

Step 6: If the algorithm fails to successfully place all required courses back into chromosome N_1(i.e., the legalization operation is unsuccessful) during Steps 3, 4, or 5, then N_1 is discarded, resulting in the generation of one fewer offspring chromosome.

Offspring chromosome N_2 is generated in a way analogous to N_1. The same steps are repeated, with the substitution of G_2 by G_1 and N_1 by N_2. That is, in Step 2, the courses in the *S* interval of chromosome N_2 are replaced by those in the corresponding *S* interval of chromosome G_1, thereby producing the final offspring chromosome N_2.

2.5.3 The Process of Mutation

Mutation is implemented in two distinct types (note that all mutation operations are performed on the offspring chromosomes produced by crossover). The first type retains most of the advantageous characteristics of the offspring while applying only minor adjustments. The second type introduces more substantial modifications to the offspring, thereby enhancing the genetic diversity of the population.

In this algorithm, a mutation rate is specified that exhibits a positive correlation with the evolutionary generation, increasing progressively as the generation count advances. Before executing a mutation operation, a random number is generated to determine which mutation type to apply. If the random number exceeds the current mutation rate, the first mutation type is executed; otherwise, the second type is performed.

$$current\ P_m = inital\ P_m - (inital\ P_m - end\ P_m) * (C_t / end_t)$$

C_t represent the current evolutionary generation; *current* P_m denote the mutation rate at generation C_t; *initial* P_m is the initial mutation rate (set to 0.8 in this algorithm); *end* P_m is the target mutation rate to be attained (set to 0.4); end_t indicates the approximate number of generations needed for the population satisfaction to stabilize, which is set to 200 in this algorithm.

2.5.3.1 The detailed procedure for the first mutation type is described as follows:

Step 1: Two distinct DTS are randomly selected from an offspring chromosome.

Step 2: A classroom is randomly chosen.

Step 3: The course instances associated with the selected classroom within the two DTS are swapped.

Step 4: Check the legality of the mutation process(make sure that no conflicts are generated). If the mutation is legal, it is retained; otherwise, the offspring chromosome is restored to its original state.

2.5.3.2 The detailed procedure for the second mutation type is described as follows:

Step 1: The satisfaction contribution of every course in an offspring chromosome is computed. The courses exhibiting the lowest satisfaction are then selected.

Step 2: These selected courses are removed from the chromosome. An attempt is made to reinsert them into different locations within the chromosome. If this reinsertion attempt fails, the courses are restored to their original positions.

2.5.4 Design Motivation and Advantages of the Algorithm

The crossover and mutation operations proposed in this paper offer the following key advantages:

1.Structure Preservation and Gene Inheritance: Local chromosomal structures are inherited and preserved through the direct copying of DTS segments from parent chromosomes, which enhances the stability of the algorithm.

2.Natural Embedding of Mutation: For courses that cannot be directly inherited from parents, mutation is achieved via a reinsertion mechanism. This allows the solution to jump across the search space, thereby improving exploration capability.

3.Exploration Capability: In the later evolutionary stages, when the population is trapped in local optima, mutation introduces diversity that helps the population escape such local optima.

4.Feasibility Guarantee Mechanism: Throughout the entire procedure, all course insertions undergo conflict detection, ensuring the legitimacy of offspring chromosomes.

5.Dynamic Balance between Randomness and Convergence: When parent chromosomes are substantially different from one another, the inherited time segments exhibit higher heterogeneity and mutation occurs more frequently. In contrast, during later evolutionary stages when the population becomes more uniform, local copying combined with the least-load strategy accelerates convergence.

3.Results analysis

3.1 Experimental Results

This study selects all students in XX high school for research subjects. These data include 89 different course instances, 7 DTS and 200 students. Based on this data set, the algorithm ran

1000 trials on a computer with 4GHz CPU, 16GB RAM. The experiment lasted 30 minutes; the results of the study showed that the overall Satisfaction remained stable at approximately 0.86, the optimal value reached 0.875. Meaning that under optimal conditions, about 87.5% of students can successfully obtain their desired schedule. As shown in Figure 7, a trend chart showing changes in satisfaction after multiple runs.

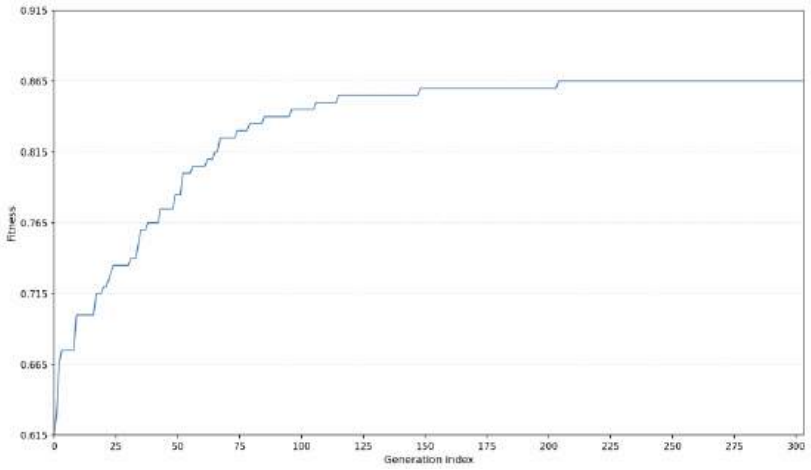


Figure 7 Trend Chart of Satisfaction Variation

As can be seen, satisfaction increases rapidly in the early phase of the algorithm and gradually converges to around 0.865 in the later phase, a pattern that aligns with the typical characteristics of genetic algorithms.

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Slot1		Psychology	English Foundation 1	AP World History	Algebra I	English Foundation 2	TOEFL 2			AP Statistics	Calculus
		Mike	Leo	Sam	Gary	Megan	Sally			Jack	Willian
Slot2		Pre-Calculus	AP Pre-Calculus	Algebra I			Algebra II	English 500: English College Writing	Statistics		AP Computer Science Principles
		Oliva	Ethan	Gary			Willian	Thomas	Jack		Lucas
Slot3		AP Biology	Psychology		Advanced Studio Art	TOEFL 2		Statistics	English 500: English College Writing	SAT 1	
		Lily	Mike		Emily	Sally		Jack	Thomas	Alex	
Slot4		English 400: English Literature and Rhetoric	Economics Honor	Chemistry	English Foundation 1	Algebra II		AP English Language and Composition	English Foundation 2	Computer Science A Honor	SAT 2
		Irene	Stella	Sophia	Leo	Willian		Thomas	Megan	Gary	Alex
Slot5			TOEFL 1	TOEFL 2			English Foundation 2	Business and Economics	Pre-AP English 1	AP Computer Science Principles	English Literature and Composition
			David	Sally			Megan	Stella	Bridgette	Lucas	Alex
Slot6					Chemistry	Advanced Studio Art	English Foundation 2 (IELTS)				
					Sophia	Emily	Leo				
Slot7		TOEFL 1		English Foundation 1	TOEFL 2	Chemistry	Business and Economics	English Foundation 2	AP Human Geography	Digital Media Technology	AP Computer Science A
		David		Leo	Sally	Sophia	Stella	Megan	Jerry	Gary	Allen

Figure 8 Resulting Course Schedule

The final generated timetable is presented in Figure 8, in which each row represents a distinct DTS and each column represents a distinct classroom.

3.2 Discussion

The proposed course scheduling system adopts an innovative framework for chromosome structure definition. This framework reduces the difficulty, workload, and computational complexity of scheduling implementation and inherently prevents many conflicts at the structural level. The advantages of this design are analyzed below from three aspects: temporal resource management, classroom resource allocation, and resolution of teacher conflicts.

Temporal Resource Management: The chromosome is composed of n DTS, each DTS corresponds to a unique and fixed class period. For example, the courses in DTS_1 are held during Monday morning; the courses in DTS_2 are held on Monday afternoon. This achieves full coverage of all class periods and no overlaps. This design directly eliminates the time conflicts between courses, which provides precise time location for course allocation; through ensuring the independence and uniqueness of class periods, eliminating the potential conflict in time dimension completely. It addresses issues such as resource allocation errors arising from ambiguous time constraints in traditional scheduling models.

Classroom Resource Allocation: The chromosome is structured as a two-dimensional list constructed from DTSs and classrooms, thereby forming a dual-dimensional resource management matrix that integrates both temporal and spatial dimensions. In this two-dimensional list, DTSs are treated as rows and classrooms as columns, creating an intuitive and efficient framework for resource representation. This design overcomes the limitation of conventional scheduling systems, in which temporal and spatial resources are managed separately, by incorporating both core resource types into a unified management framework. Consequently, it provides standardized coordinates for precise course instance allocation and furnishes rigorous logical support for subsequent dynamic resource adjustments.

Teacher Conflict Resolution: A teacher conflict occurs when the same teacher is assigned multiple teaching tasks during the same time period. Because the chromosome is structured as a two-dimensional list based on DTSs and classrooms, preventing such conflicts requires only that no two identical teachers appear within any single row (i.e., within any given DTS) when course instances are adjusted. This solves the redundancy problem found in traditional scheduling algorithms, which stems from the lack of a well-defined constraint system linking teachers and time slots.

In summary, this class period organization method—which connects temporal and classroom resources—reduces the workload of the scheduling algorithm by obviating concerns over time, classroom, and teacher conflicts among courses during the algorithmic process. Unlike traditional scheduling approaches, which exhibit high complexity and slow convergence as a result of fragmented resource management, the proposed structure establishes a clear and efficient scheduling framework. This framework acts as the core architectural foundation that enables the course scheduling system to operate with precision and efficiency, fully reflecting the

strong adaptability of the framework to the scheduling application context.

Besides, the most time-consuming part of this algorithm is the calculation of Satisfaction, which time complexity is about $O(n \cdot x^m)$. This algorithm uses set to optimize the Satisfaction calculation algorithm; the complexity after optimization is $O(xm + m^2)$, where n is the number of students, x is the number of classrooms in each DTS(the conditions where the DTS is full of course instances) and m is the number of DTS(the number of courses students select).

The underlying data structure of the set is a hash table. When adding an element to the set, the program calculates its hash value and stores it at a specific location in the hash table. To determine whether an element exists in the set, the program computes the hash value of that element and locates the corresponding position in the hash table to detect whether value exists at that position. This makes the time complexity of a single search $O(1)$.

This significantly improves the algorithm's time efficiency: calculating the complete Satisfaction for 200 students each selecting seven courses takes only 0.01 -- 0.03 seconds.

4. Conclusions

The widespread adoption of course selection systems is now a certainty. However, how to enhance students' satisfaction with personalized course selection while ensuring rational allocation of teaching resources has become a key focus for educational administrations. Traditional scheduling methods primarily rely on prioritizing teachers without conflicts, then allowing students to compete for courses—a practice that appears fair but lacks strategic planning, inevitably leading to significant resource waste. This paper proposes and implements a genetic algorithm-based scheduling system for elective course systems. The system takes students' ranked course preferences as input and incorporates the degree of fulfillment of courses chosen by students into the satisfaction function, leveraging the evolutionary mechanism of the genetic algorithm to achieve timetable optimization. Experimental results indicate that, while maintaining overall scheduling feasibility, the proposed method enables approximately 87.5% of students to enroll in their desired courses in a real-world scenario, with overall satisfaction stabilizing at approximately 0.86, thereby significantly improving student course selection satisfaction.

The primary contribution of this study is the direct quantification of student course selection preferences as an optimization objective and their integration into the scheduling algorithm, thus overcoming the limitation of prior approaches that emphasized resource scheduling while overlooking student needs. Nevertheless, the system still exhibits certain limitations, including the need for improved adaptability to complex resource conflicts and further optimization of computational efficiency when processing large-scale student data.

Future work may focus on adjusting algorithm preset parameters, enhancing the algorithm's capacity to handle complex constraints, improving parallel execution efficiency, and optimizing the satisfaction calculation algorithm, all of which would further enhance the practicality of the elective course scheduling system.

Works Cited

1.Rappos, E., et al. "A Mixed-Integer Programming Approach for Solving University Course Timetabling Problems." 2022, pp. 392-397.

2. Özkan, A., and A. Ulucan. "A Mixed-Integer Linear Programming Approach for University Timetabling Problem with the Multi-Section Courses: An Application for Hacettepe University Department of Business Administration." 2022, pp. 132-141.

3. Azadeh, A., et al. "A Multi-Objective Optimisation Model for University Course Timetabling Problem Using a Mixed Integer Dynamic Non-Linear Programming." 2013, pp. 56-72.

4. Dunke, F., and S. Nickel. "A Matheuristic for Customized Multi-Level Multi-Criteria University Timetabling." 2023, pp. 1318-1326.

5. Lindahl, M., et al. "A Fix-and-Optimize Matheuristic for University Timetabling." 2018, pp.649-660.

6. Mikkelsen, R., and D. S. Holm. "A Parallelized Matheuristic for the International Timetabling Competition 2019." 2022, pp. 432-445.

7 Muslem, I., et al. "Adaptive Heuristic-Based Ant Colony Optimization for Multi-Constraint University Course Timetabling with Morning Slot Preference for Energy Efficiency." 2025, pp. 3-7.

8. Mazlan, M., et al. "Ant Colony Optimisation for Solving University Course Timetabling Problems." International Journal of Engineering and Technology 7 (2018): 139-144.

9. Kenekayoro, P., and G. Zipamone. "Greedy Ants Colony Optimization Strategy for Solving the Curriculum Based University Course Timetabling Problem." 2016, pp. 3-12.

10. Zahid, I. "Effective Solution of University Course Timetabling Using Particle Swarm Optimizer Based Hyper Heuristic Approach." 2021, pp. 1465-1472.

11. Swari, M. H. P., et al. "Improving the Performance of the General Course Scheduling System at UPN Veteran Jawa Timur through the Application of the IWCFPSO Algorithm." 2023, pp. 4-8.

12. Thepphakorn, T., and P. Pongcharoen. "Particle Swarm Optimisation Variants and Its Hybridisation Ratios for Generating Cost-Effective Educational Course Timetables." 2021, pp. 8-22.

13. Ahandani, M. A., & Baghmisheh, M. T. V. (2013). Hybridizing genetic algorithms and particle swarm optimization transplanted into a hyper-heuristic system for solving university course timetabling problems. 4-11.

14. Awad, F. H., et al. "Large-Scale Timetabling Problems with Adaptive Tabu Search." 2022, pp. 168-176.

15. Aladag, C. H., et al. "The Effect of Neighborhood Structures on Tabu Search Algorithm in Solving Course Timetabling Problem." 2009, pp. 12350-12355.

16.Chen, M., et al. "A Tabu Search Algorithm with Controlled Randomization for Constructing Feasible University Course Timetables." 2020, pp. 4-15.

17. Holland, J. H. "Genetic Algorithms and the Optimal Allocation of Trials." 1973, pp. 88-105.

18. Goldberg, D. E., and R. Lingle. "Alleles, Loci, and the Traveling Salesman Problem." Proceedings of the 1st International Conference on Genetic Algorithms and Their Applications, 1985, pp. 154-159.

Leaf Swallowing as a Self-Medication Behaviour in Nonhuman Primates: A Narrative Review By Manasa Thyagarajan

Abstract

Leaf swallowing is a self-medication behaviour observed in several wild primate species in which whole leaves are swallowed without chewing, often in association with parasite control and health maintenance. This narrative review examines published evidence on leaf swallowing in nonhuman primates and evaluates its proposed functions, species patterns, and ecological significance. Literature searches were conducted using Google Scholar and PubMed with combinations of terms related to leaf swallowing, self-medication, and primates. Studies were included if they documented leaf swallowing in wild primates, identified the plant species involved, and proposed a medicinal function for the behaviour. Nine studies across six primate species were included in the qualitative synthesis. Parasite expulsion emerged as the best-supported function across all species reviewed, particularly among chimpanzees. Bonobos demonstrated seasonal patterns of leaf swallowing associated with increased parasite risk, while gorilla studies suggested possible antioxidant and antibacterial benefits linked to plant phytochemistry. Evidence from gibbons further indicates that self-medication behaviours are not limited to African great apes. Overall, leaf swallowing appears to represent an adaptive health-maintenance strategy that provides insight into primate cognition, behavioural ecology, and the evolutionary origins of medicine.

Keywords Zoopharmacognosy; leaf swallowing; self-medication; primates; parasites; behaviour

1. Introduction

Across the animal kingdom, many species exhibit behaviours associated with maintaining health and reducing parasite burden. Birds line nests with aromatic leaves that may deter parasites, while several primate species consume or manipulate plants in ways that appear unrelated to nutrition alone. These behaviours fall within the field of zoopharmacognosy, which examines how animals use natural substances, particularly plants, to prevent or treat disease (Huffman, 2003).

Among nonhuman primates, one of the most widely documented self-medication behaviours is leaf swallowing. In this behaviour, primates selectively ingest rough, fibrous leaves whole without chewing them. The leaves often pass through the digestive tract largely undigested and are later found in faeces together with intestinal parasites or parasite eggs. The dominant hypothesis is that rough, trichome-covered leaves swallowed whole physically dislodge intestinal parasites as they pass through the gut (M.Huffman et al. 1996).

Leaf swallowing differs from several other forms of self-medication reported in primates. Bitter-pith chewing involves the ingestion of chemically active plant materials, such as *Vernonia amygdalina*, that contain bioactive compounds with antiparasitic or antimicrobial properties.

Geophagy, or soil consumption, is another related behaviour that may assist with detoxification, mineral supplementation, or gastrointestinal regulation. Unlike these behaviours, leaf swallowing is believed to function primarily through a mechanical rather than chemical mechanism.

Researchers have documented leaf swallowing in chimpanzees, bonobos, gorillas, and gibbons across both African and Asian forests(Yinda et al. 2024; Huffman et al. 1996; Wrangham 1995; Fowler et al. 2007; McLennan and Huffman 2012; Fruth et al. 2014; Kreyer et al. 2022; Barelli and Huffman 2017).Observations suggest that primates carefully select particular plant species and swallow them under specific ecological conditions, especially during periods of increased parasite exposure. These findings raise important questions about how primates recognize medicinal plants and whether such behaviours are instinctive, socially learned, or influenced by environmental experience.

This narrative review examines published studies on leaf swallowing in nonhuman primates in order to address three primary questions: which primate species exhibit leaf swallowing behaviour, which plant species are involved, and what functions are proposed for this behaviour.

2. Methods

2.1 Literature Search

This narrative review examined published scientific literature on leaf swallowing behaviour in nonhuman primates. Searches were conducted using Google Scholar and PubMed.
The following search string was used:
("leaf swallowing" AND "primates") AND (primate OR chimpanzee OR bonobo OR gorilla OR gibbon)

2.2 Inclusion Criteria

Studies were included if they:

- documented leaf swallowing behaviour in wild primates,
- identified the plant species involved, and
- proposed a medicinal or health-related function for the behaviour.

2.3 Exclusion Criteria

Studies were excluded if they:

- focused exclusively on captive animals,
- did not identify the plant species involved,
- did not discuss a medicinal or health-related function, or
- lacked accessible full text.

2.4 Data Collection and Analysis

For each eligible study, information was collected regarding:

- primate species observed,
- study location,
- plant species involved,
- observed leaf-swallowing behaviour, and
- proposed medicinal function.

Patterns across studies were analysed qualitatively to identify recurring trends in species behaviour, plant selection, and proposed health functions. Due to variation in methodologies and sample sizes among studies, quantitative meta-analysis was not performed.

2.5 Tools

Reference management, and Microsoft Word was used for manuscript preparation and editing.

3. Results

A total of nine studies documenting leaf-swallowing behaviour in wild primates were included in this review. Collectively, the studies encompassed six primate species across multiple African and Asian field sites.

3.1 Chimpanzees

Four studies focused on chimpanzees. Chimpanzees were observed swallowing rough, hairy leaves from plants such as *Aspilia mossambicensis* and *Desmodium gangeticum* (McLennan and Huffman 2012; Huffman et al. 1996; Fowler et al. 2007; Wrangham 1995)Individuals typically collected the leaves carefully and swallowed them whole without chewing. The leaves later appeared intact in faecal samples together with intestinal worms, supporting the hypothesis that leaf swallowing contributes to parasite removal.
One study additionally reported seasonal variation in leaf swallowing among *Pan troglodytes*, suggesting that the behaviour may increase during periods of elevated parasite exposure.

3.2 Bonobos

Two studies examined bonobos in the Democratic Republic of the Congo(Fruth et al. 2014; Kreyer et al. 2022). Bonobos swallowed leaves from *Manniophyton fulvum*, particularly during the rainy season when parasite prevalence is believed to increase. This seasonal pattern suggests that leaf swallowing may function as a preventive behavioural adaptation in response to environmental risk. A follow-up analysis of the same population found that *M. fulvum* ingestion was recorded across 111 observation days, though the relationship between ingestion and detected parasite load was more complex than initially proposed.

3.3 Gorillas

One study documented leaf swallowing in western lowland gorillas in Gabon(Yinda et al. 2024). Gorillas consumed leaves and bark from *Ceiba pentandra*, *Myrianthus arboreus*, *Ficus* species, and *Milicia excelsa*,These plants contain compounds associated with antioxidant and antibacterial activity, leading researchers to propose that gorillas may benefit from both mechanical parasite removal and additional phytochemical effects.

3.4 White-Handed Gibbons

A study from Thailand reported leaf swallowing in white-handed gibbons(Barelli and Huffman 2017). The gibbons consumed leaves from *Gironniera nervosa*, which were proposed to assist in controlling nematode infections (Barelli & Huffman, 2017). This finding suggests that self-medication behaviours may occur across geographically distinct primate lineages and are not restricted to African great apes.

3.5 Cross-Species Patterns

Across all reviewed species, parasite control emerged as the most consistently supported function of leaf swallowing. Chimpanzees and bonobos showed the strongest evidence for parasite-related benefits, while gorilla studies proposed additional antioxidant and antibacterial effects associated with plant chemistry. Seasonal variation was most clearly documented among bonobos.

4. Figure

Table 4.1. Summary of Included Studies

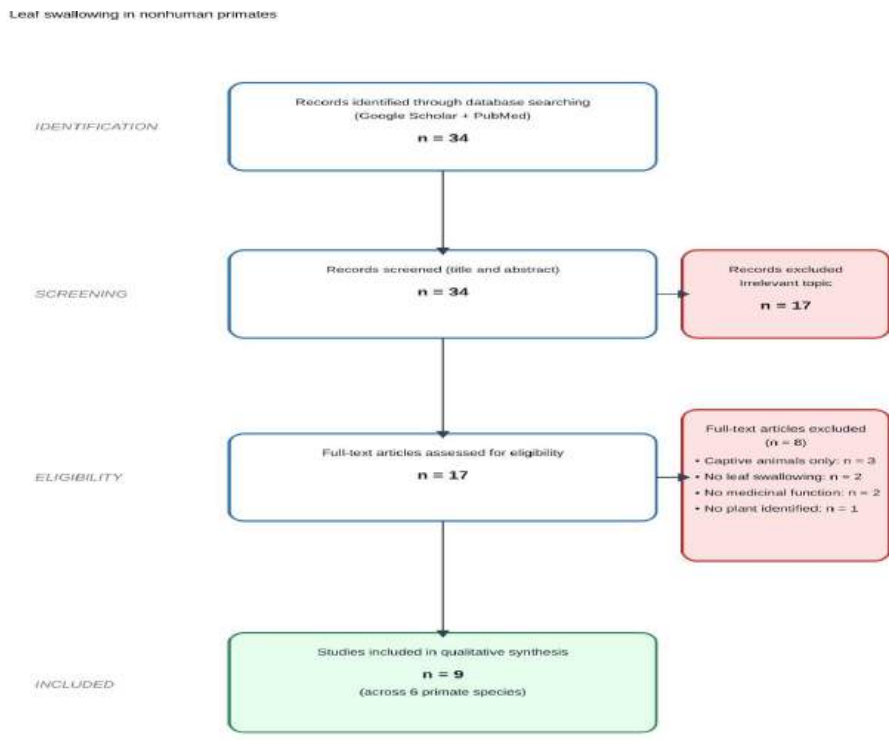
Table 1. Studies reporting leaf-swallowing behavior in nonhuman primates and their proposed health functions

Author & Year	Species	Location	Behavior Observed	Plant(s) Used	Proposed Function
Huffman M.A. et al., 1996(M. Huffman et al. 1996)	*Pan troglodytes schweinfurthii* (Chimpanzee)	Mahale Mountains, Tanzania	Leaf swallowing	*Aspilia mossambicensis*, *Lippia plicata*, *Hibiscus* sp., *Trema orientalis*	Parasite removal
Richard W Wrangham, 1995(Wrangham 1995)	*Pan troglodytes* (Chimpanzee)	Kibale National Park, Uganda	Leaf swallowing	Whole leaves	Tapeworm removal (*Bertiella studeri*)
Fowler A. et al., 2006(Fowler et al. 2007)	*Pan troglodytes vellerosus* (Chimpanzee)	Gashaka, Nigeria	Leaf swallowing	*Desmodium gangeticum*, sharp-edged grasses	Parasite removal

McLennan M.R. et al., 2017(McLennan and Huffman 2012)	*Pan troglodytes schweinfurthii* (Chimpanzee)	Bulindi, Uganda	Leaf swallowing	Whole leaves	Nematode control (*Oesophagostomum* spp.)
Fruth B. et al., 2014(Fruth et al. 2014)	*Pan paniscus* (Bonobo)	LuiKotale, DRC	Leaf and stem swallowing	*Manniophyton fulvum*	Parasite control; seasonal pattern
Kreyer M. et al., 2021(Kreyer et al. 2022)	*Pan paniscus* (Bonobo)	LuiKotale, DRC	Leaf swallowing	*Manniophyton fulvum*	Gastrointestinal parasite control
Yinda L. et al., 2024(Yinda et al. 2024)	*Gorilla gorilla gorilla* (Western lowland gorilla)	Moukalaba-Doudou National Park, Gabon	Leaf swallowing	*Ceiba pentandra*, *Myrianthus arboreus*, *Ficus* spp., *Milicia excelsa*	Antioxidant and antibacterial properties
Barelli C. et al., 2017(Barelli and Huffman 2017)	*Hylobates lar* (White-handed gibbon)	Khao Yai National Park, Thailand	Leaf swallowing	*Gironniera nervosa*	Nematode control (*Streptopharagus pigmentatus*)

Figure 4.2. Geographic Distribution of Included Study Sites

Figure 4.3. PRISMA Flow Chart.



5. Discussion

The reviewed studies suggest that leaf swallowing represents a deliberate and adaptive behaviour rather than random feeding. The repeated observation of primates selecting rough leaves and swallowing them whole supports the hypothesis that the behaviour contributes to health maintenance (Huffman, 1997, 2001).

Evidence for parasite expulsion is currently the strongest-supported explanation for leaf swallowing. Chimpanzees provide the clearest evidence for this mechanism, as intact leaves and expelled intestinal parasites have repeatedly been observed together in faecal samples (Wrangham, 1995; McLennan & Huffman, 2012). Bonobos also demonstrate seasonal increases in leaf swallowing during periods associated with elevated parasite risk (Fruth et al., 2014). A more recent analysis of the same bonobo population, however, found no significant association between *M. fulvum* ingestion and strongyle egg presence, suggesting that the relationship between leaf swallowing and parasite load may be more complex than initially proposed (Kreyer et al., 2021).

In contrast, proposed antioxidant and antibacterial benefits remain more speculative. Although several plant species consumed by gorillas contain compounds with known phytochemical activity, direct evidence linking these compounds to measurable health outcomes

in wild primates remains limited. Therefore, antioxidant and antibacterial effects should currently be considered hypotheses requiring further investigation rather than conclusively established functions.

The occurrence of leaf swallowing in both African and Asian primates suggests that self-medication behaviours may have deep evolutionary origins. The behaviour also raises important questions regarding cognition and learning. It remains unclear whether primates identify medicinal plants instinctively or acquire this knowledge through observation and social learning. Young primates may learn leaf selection and swallowing techniques by observing experienced adults, although direct evidence for cultural transmission in wild populations remains limited.

Most studies included in this review were observational, making it difficult to establish direct causal relationships between leaf swallowing and improved health outcomes. Distinguishing deliberate self-medication from accidental ingestion or nutritional feeding behaviour also remains challenging.

Future research could strengthen the evidence base by analysing the chemical properties of swallowed leaves, measuring parasite loads before and after swallowing episodes, and examining behavioural learning in juvenile primates. Experimental work in semi-wild sanctuaries may also allow researchers to investigate parasite management non-invasively under more controlled conditions.

6. Limitations

- Most included studies had relatively small sample sizes.
- The majority of research was conducted at a limited number of long-term African field sites.
- Observational methods limited the ability to confirm direct medicinal effects.
- Several proposed benefits were inferred from plant chemistry rather than measured physiological outcomes.
- Publication bias may have influenced which findings were available in the literature.
- Most included studies originated from a small number of well-studied primate communities, which may overrepresent the behavioural repertoires of those populations.
- Most included studies were concentrated in Africa, highlighting a geographical bias in the current literature.

7. Conclusion and Future Directions

Leaf swallowing appears to represent an adaptive self-medication behaviour among nonhuman primates. Evidence across multiple species supports the hypothesis that swallowing rough leaves whole contributes to parasite management and gastrointestinal health maintenance

(Huffman, 2003; McLennan & Huffman, 2012). Additional antioxidant and antibacterial effects have been proposed in some species (Even et al., 2024), although these functions require further experimental confirmation.

The study of leaf swallowing provides insight into the relationships between animal cognition, ecology, and health-related behaviour. It also contributes to broader understanding of the evolutionary origins of medicinal plant use (Huffman, 2001).

Future studies should focus on combining behavioural observation with parasitological and phytochemical analysis in order to better determine the effectiveness and mechanisms of leaf swallowing. Research on juvenile learning and social transmission may also clarify whether self-medication behaviours are primarily instinctive, socially learned, or shaped by environmental experience (Huffman & Hirata, 2004).

Acknowledgements

I would like to sincerely thank Dr.Venkatesha Udupa, Dr.Anish Mehta, Mr. Thyagarajan Shivashanmugam, N Krishna Prasad Kamath, and Ms.Vandana S for their constant support, guidance, and encouragement throughout the completion of this review. Their valuable suggestions and assistance played an important role in shaping this work and helping me bring it to completion.

Works Cited

Barelli, C., & Huffman, M. A. (2017). Leaf swallowing and parasite expulsion in Khao Yai white-handed gibbons (Hylobates lar), the first report in an Asian ape species. American Journal of Primatology, 79(6), e22610.

Even, L., Oyaba, D., Onanga, R., Sima Obiang, C., Begouabe, H., Akomo-Okoue, E. F., Obame-Nkoghe, J., Mitola, R., Ondo, J., Atome, G. N., Obame Engonga, L. C., Setchell, J. M., & Godreuil, S. (2024). Antibacterial and antioxidant activities of plants consumed by western lowland gorillas (Gorilla gorilla gorilla) in Gabon. PLOS ONE, 19(7), e0306957. https://doi.org/10.1371/journal.pone.0306957

Fowler, A., Koutsioni, Y., & Sommer, V. (2006). Leaf-swallowing in Nigerian chimpanzees: Evidence for assumed self-medication. Primates, 47(4), 355–359. https://doi.org/10.1007/s10329-006-0001-6

Fruth, B., Bondjengo Ikombe, N., Kitengie Matshimba, G., Metzger, S., Muganza, D. M., Mundry, R., & Fowler, A. (2014). New evidence for self-medication in bonobos: Manniophyton fulvum leaf- and stemstrip-swallowing from LuiKotale, Salonga National Park, DR Congo. American Journal of Primatology, 76(2), 146–158. https://doi.org/10.1002/ajp.22217

Huffman, M. A. (1997). Current evidence for self-medication in primates: A multidisciplinary perspective. Yearbook of Physical Anthropology, 40, 171–200. https://doi.org/10.1002/(SICI)1096-8644(1997)25+<171::AID-AJPA7>3.0.CO;2-7

Huffman, M. A. (2001). Self-medicative behavior in the African great apes: An evolutionary perspective into the origins of human traditional medicine. BioScience, 51(8), 651–661. https://doi.org/10.1641/0006-3568(2001)051[0651:SMBITA]2.0.CO;2

Huffman, M. A. (2003). Animal self-medication and ethno-medicine: Exploration and exploitation of the medicinal properties of plants. Proceedings of the Nutrition Society, 62(2), 371–381. https://doi.org/10.1079/PNS2003257

Huffman, M. A., & Hirata, S. (2004). An experimental study of leaf swallowing in captive chimpanzees: Insights into the origin of self-medicative behavior and the role of social learning. Primates, 45(2), 113–118. https://doi.org/10.1007/s10329-003-0065-5

Kreyer, M., Stewart, K. M., Pashchevskaya, S., & Fruth, B. (2021). What fecal analyses reveal about Manniophyton fulvum consumption in LuiKotale bonobos (Pan paniscus): A medicinal plant revisited. American Journal of Primatology, 83(11), e23318. https://doi.org/10.1002/ajp.23318

McLennan, M. R., & Huffman, M. A. (2012). High frequency of leaf swallowing and its relationship to intestinal parasite expulsion in 'village' chimpanzees at Bulindi, Uganda. American Journal of Primatology, 74(7), 642–650. https://doi.org/10.1002/ajp.22017

Wrangham, R. W. (1995). Relationship of chimpanzee leaf-swallowing to a tapeworm infection. American Journal of Primatology, 37(4), 297–303. https://doi.org/10.1002/AJP.1350370404

Capabilities Without Functioning: Female Empowerment and Women's Health
By Kaea Shah

Abstract

Purpose: This study aimed to analyse whether the female tertiary school enrolment (TSER) and female firm ownership (WMP) predicts the female labour force participation (FLFP) across countries, and whether within country improvements in TSER and FLFP predict changes in anaemia prevalence and HPV vaccination coverage over 2015-2024..

Design/methodology/approach: Two analyses were carried out in the study, the first (Analysis A) uses the cross-sectional OLS with HC3 robust standard errors in up to 150 countries. The second analysis (Analysis B) uses two-way fixed-effects panel regression across 44 countries that are observed each year between 2015 and 2024, country-clustered standard errors and the Hausman test of specification. There are five robustness checks that determine the stability of the pre-pandemic subsamples and other specifications.

Findings: Analysis A yields a null result: neither TSER nor WMP significantly predicts FLFP. Analysis B finds that within-country TSER increases are significantly and positively associated with anaemia prevalence ($b = 0.035$, $p = .024$) (counterintuitive but robust across all five specifications); and marginally positively associated with HPV coverage ($b = 0.414$, $p = .072$), with a stronger pre-pandemic effect ($b = 0.778$, $p = .018$). FLFP does not predict either health outcome. COVID-19 dominates both anaemia ($b = 2.869$, $p < .001$) and HPV coverage ($b = 17.753$, $p = .001$).

Practical implications: Tertiary expansion must be paired with campus-based iron supplementation to mitigate short-run anaemia burden. HPV programmes require demand-side educational investment alongside supply-side delivery. Post-pandemic recovery must urgently reach cohorts of young women who missed vaccination in 2021–2022.

Originality/value: This study provides the first fixed-effects panel evidence linking TSER, FLFP, anaemia, and HPV vaccination across a 44-country, 10-year dataset spanning the COVID-19 pandemic. It empirically demonstrates why cross-sectional and panel designs are necessary complements in empowerment research and reveals a critical short-run sign reversal in the education-anaemia relationship.

Keywords female empowerment; tertiary education; anaemia; HPV vaccination; fixed-effects panel regression

Introduction

Female empowerment has emerged as one of the most consequential policy priorities of the twenty-first century, enshrined in Sustainable Development Goal 5 (Gender Equality) and its cross-cutting linkages with SDG 3 (Good Health and Well-Being) and SDG 8 (Decent Work and Economic Growth). Despite the growing scholarly and policy consensus that empowering women generates broad social and economic returns, two fundamental empirical questions

remain unresolved. First, to what extent do observable empowerment indicators predict female labour force participation (FLFP) across countries? Second, do within-country improvements in female empowerment translate into measurable gains in women's health outcomes over time? The current research paper will answer both questions by using complementary cross-sectional and panel research designs to cover up to 150 countries in the 10-year period of 2015-2024.

The economic involvement of women and their health are closely intertwined, but the processes that relate empowerment to the achievement of these goals are controversial and contingent. The female tertiary school enrolment (TSER) is commonly considered one of the pillars of empowerment, expansion of cognitive abilities, work choices and health literacy. The institutionally available economic resources and the entrepreneurial agency in the formal private sector is indicated by female firm ownership (WMP). These indicators, combined, represent the educational and economic aspects of empowerment with which the development theory predicts an increase in the labour force participation and health outcomes. The question of whether or not such predictions are empirically valid and over what time horizon, is an important one with respect to both research and international development policy.

This study is motivated by multiple interrelated gaps in the existing literature. First, prior cross-national analyses of FLFP determinants have frequently omitted female firm ownership as a predictor, or conflated educational enrolment with attainment, limiting comparability and leaving the WMP-FLFP relationship under-examined at the aggregate level. Second, fixed-effects panel analyses specifically linking female empowerment to anaemia prevalence and HPV vaccination coverage over the 2015–2024 decade are rare, despite both outcomes featuring prominently in WHO global monitoring frameworks and representing critical dimensions of women's health capability. The present study exploits this variation using a panel of 44 countries observed annually across five pre-pandemic and five pandemic-period years.

The study addresses the following three research questions:

RQ1: Do TSER and WMP predict cross-country variation in FLFP, after controlling for per capita income and total fertility rate?

RQ2a: Within countries over time, do rising TSER or FLFP predict declining anaemia prevalence among women of reproductive age?

RQ2b: Within countries over time, do rising TSER or FLFP predict higher HPV vaccination coverage among girls aged 9–14?

Literature Review

Conceptual Frameworks

The capabilities approach to empowerment by Sen (1999) views empowerment as the enlargement of substantive freedoms - the actual opportunities that women can have to live lives of their own choice. In this context, education and economic participation should not only be instrumentally useful to aggregate development outcomes; they are intrinsically important

aspects of human flourishing, the denial of which is a fundamental injustice. In the Gender Development Index and Gender Inequality Index of the United Nations Development Programme, both of which place central emphasis on the educational achievement and participation in the labour market as indicators of women agency, the framework developed by Sen has been operationalised. In the current study, this framing is essential: TSER and WMP are not just assumed to predict an economic or health outcome but rather constitute a part and parcel of the empowerment process the downstream effects of which are theorised to be mediated by institutional context, temporal lags, and compositional dynamics.

Kabeer (1999) provides a more process based conceptualisation which complements the conceptualisation given by Hill (1995) and Serrat (1997). This conceptualisation conceptualises empowerment in three dimensions: resources, agency, and achievements. The resources include the material, social and human resources through which the choices are exercised of which education and access to the formal economy are prime examples. The ability to formulate and take action on what one desires is called agency. Achievements are the end products that are the fruit of exercised agency. It means that TSER and WMP are projectors of the resource aspect of empowerment, and that their translation into labour force and health achievements crucially depends on the presence of enabling institutional conditions, such as legal protections, affordable childcare, transport infrastructure, non-discriminatory hiring. This conditional nature of the empowerment phenomenon, which is often not accounted for, is a major cause of why cross-sectional regression on aggregate indicators of empowerment often fails to detect the expected relationships. The economic basis is the human capital tradition (Becker, 1964; Schultz, 1961, 2002) that assumes that investing in education will increase productivity, hence market wages, which will raise the return to labour market participation. The framework of gender planning presented by Moser (1993) further separates between the practical needs of gender and the strategic needs of gender with educational and economic empowerment acting mainly on the plane of strategic interests by reorganizing conditions of female participation in the public and economic life.

Female Empowerment and Labour Force Participation

The economic theory supports various avenues according to which empowerment of women increases FLFP. The wage channel operates through human capital: higher educational attainment raises market wages relative to the reservation wage, making market work more attractive relative to home production (Mincer, 1974). The fertility channel operates through the well-established negative relationship between education and fertility: more educated women have fewer children, reducing the time cost of caregiving and increasing hours available for market work (Barro & Lee, 2013). The intra-household bargaining channel posits that more educated women command greater bargaining power within households and are more likely to assert preferences for labour market participation against opposing family preferences (Duflo, 2012). Female firm ownership is theorised to raise FLFP through two additional mechanisms:

direct employment creation for other women, and signalling of a broader institutional enabling environment that facilitates labour market entry (World Bank, 2022).

Empirically, the education-FLFP relationship at the country level is more ambiguous than theory predicts. Goldin (1995) identifies a U-shaped relationship between economic development and female labour force participation: FLFP first declines as economies industrialise and gender-discriminatory labour market norms solidify, then rises as service sectors expand, educational attainment increases, and norms modernise. This non-linearity means that a linear regression coefficient applied across the full income spectrum will be weak or uninformative regardless of the true within-country relationship. Studies that control for income non-linearity, as in Mammen and Paxson (2000), recover stronger positive education-FLFP associations, particularly in upper-middle and high-income countries. Gaddis and Klasen (2014) confirm in a comprehensive cross-national analysis that cross-country variation in FLFP is predominantly explained by time-invariant institutional and cultural factors — family law systems, occupational segregation norms, childcare infrastructure, and religious context — that no single observable empowerment indicator can adequately represent in a single-period snapshot. This finding directly contextualises the null result in Analysis A and constitutes the primary methodological motivation for the fixed-effects panel design of Analysis B.

Contrasting evidence from Boeri et al. (2005) and Verick (2014) suggests the education-FLFP relationship is stronger in middle-income economies with intermediate institutional development, where education is both necessary and sufficient to overcome participation barriers, than at the extremes of the income distribution. Klasen and Pieters (2015) illustrate this pattern strikingly in the Indian context: rising female educational attainment in urban India during the 2000s was not accompanied by rising FLFP, a decoupling they attribute to supply-side structural constraints and the persistence of social norms that credential accumulation alone cannot rapidly overcome. This pattern is directly analogous to the null cross-sectional finding in the present study. On female firm ownership, Aterido et al. (2011), Bruhn (2009), and Hallward-Driemeier (2013) all document that the aggregate employment externalities of female entrepreneurship are modest in the short run, reflecting the structural constraints documented in the World Bank's Women, Business and the Law project — providing a plausible explanation for the non-significant WMP-FLFP association in the present cross-section.

Female Empowerment, Nutrition, and Anaemia

Anaemia in women of reproductive age - defined by the WHO (2011) as haemoglobin below 12 g/dL - is estimated to affect 571 million women in the world (Kassebaum et al., 2016). Its effects are dire: increased risk of maternal mortality, low birth weight, preterm birth and cognitive impairment in children (Black et al., 2013). The economic costs too are high; WHO estimates the anaemia related productivity cost to low and middle income countries about 0.57 per cent of the GDP each year. Although prevalence has only decreased moderately as a result of decades of programmatic interventions, such as iron supplementation, food fortification and

deworming campaigns, there is evidence that structural determinants including female empowerment act in concert with programmatic interventions to determine long-run trends.

Theoretical pathways from female education to reduced anaemia operate through health literacy (better nutritional knowledge and dietary behaviour), income (purchasing power for iron-rich foods), autonomy (stronger intra-household bargaining for food allocation), and healthcare access (higher uptake of antenatal care and supplementation) (Glewwe, 1999; Haddad et al., 1997). Individual-level evidence generally supports these channels (Akseer et al., 2017; Grantham-McGregor & Ani, 2001). Critically for the present study, Psacharopoulos and Patrinos (2018) argue that the health returns to education operate primarily over generational timescales: maternal education most strongly predicts child health outcomes rather than contemporaneous adult health, because the mechanisms — dietary knowledge, income accumulation, and health-seeking behaviour — unfold and compound over time. A decade-long panel may therefore be too short to observe the health dividend of educational expansion.

There are two other mechanisms unique to rapid tertiary growth that give complementary explanations to the counterintuitive positive TSER-anaemia association observed in Analysis B. The dietary transition hypothesis (Popkin, 2006) is a theory that proposes that urbanisation that accompanies educational expansion changes the diets to shut out the traditional high-iron staples legumes, dark leafy vegetables, organ meats to the processed and convenience foods that have lower bioavailable iron content. It is based on the compositional mechanism, which postulates that as tertiary enrolment increases, a higher proportion of the measured female population is concentrated on the 18-25 age sub-group at highest risk of anaemia due to menstrual iron loss and less-established dietary habits. These three mechanisms temporal lag, dietary transition, and compositional shift are mutually consistent and give a consistent account of the finding without any challenge of the long-run education-health relationship. The joint operation they experienced during 2015-2019 is consistent with the pre-pandemic year fixed effects showing a monotonically increasing trend of anaemia in 2016-2019 that cannot be explained by any model covariate. The correlation between FLFP, and anemia is theoretically ambiguous: employment increases income and food security but can also pose time constraints and physical demands that exacerbate nutritional outcomes. Lund et al. (2010) report positive health effects of formal employment; Nandi et al. (2014) report null effects of informal employment, which could indicate the relationship is highly sector-specific.

Female Empowerment and HPV Vaccination

HPV is the primary causal agent of cervical cancer, which kills approximately 342,000 women annually — over 90% in low- and middle-income countries (WHO, 2022). The WHO's 90-70-90 strategy targets 90% HPV vaccination coverage by 2030, yet as of 2022 global coverage among girls aged 9–14 remained far below this target. Coverage varies substantially across countries, shaped by supply-side factors — GAVI funding, national programme adoption, cold chain infrastructure, healthcare worker availability — and demand-side factors including parental health literacy, awareness of HPV and cervical cancer, and autonomy in health

decision-making. Female educational attainment is theorised to raise HPV vaccination primarily through demand-side mechanisms. Bonney et al. (2018) review evidence across sub-Saharan Africa and find maternal education consistently associated with greater HPV awareness and higher stated acceptance of vaccination. Gertig et al. (2013) document that socioeconomic status — strongly correlated with education — predicts HPV vaccination completion rates in national programmes even in high-supply contexts such as Australia. Jit et al. (2014) model that demand-side educational interventions can substantially accelerate programme coverage in middle-income settings. FLFP may increase HPV vaccination through the income and occupational information channels, but time constraints associated with demanding employment may counteract these effects, producing the theoretically ambiguous and empirically null FLFP-HPV finding observed across all specifications in Analysis B.

The COVID-19 pandemic disrupted HPV programmes globally through school closures, healthcare worker redeployment, and increased population hesitancy about health system contact (WHO & UNICEF, 2021). Coverage declined in 2020–2021, particularly in countries dependent on school-based delivery. However, post-2020 saw acceleration of catch-up campaigns funded by GAVI, which drove substantial coverage increases in 2022–2024 in several lower-income countries (Gavi, 2022). This pattern of initial disruption followed by accelerated scale-up is precisely what the present study documents in its year fixed effects: sharp declines in HPV coverage in 2021–2022 followed by partial recovery in 2023.

Control Variables and Research Gap

Log GDP per capita (ln_GDP) is a fundamental determinant of both health outcomes and FLFP. The near-zero bivariate correlation between FLFP and ln_GDP ($r = -.012$) in the cross-sectional sample reflects the non-linear U-shaped income-participation relationship documented by Goldin (1995): both very poor and very rich economies exhibit high FLFP through structurally different mechanisms, attenuating any linear association. Total fertility rate (FR) is the most important confounder of the education-FLFP relationship, exhibiting a strong negative correlation with TSER ($r = -.711$) consistent with the demographic transition hypothesis (Barro & Lee, 2013). FR also predicts both health outcomes: positively for anemia through iron depletion from repeated pregnancies, and positively for HPV through stronger engagement with maternal and reproductive health services in high-fertility settings. The COVID-19 dummy is included exclusively in Analysis B to control for the structural break introduced by the pandemic, whose dominant predictive power in both health models underscores the necessity of explicit pandemic control in any panel spanning 2015–2024.

The literature reviewed supports six testable hypotheses, stated formally in Section 3.5. At the cross-sectional level, positive associations are predicted between TSER and FLFP (H1) and between WMP and FLFP (H2), though prior evidence suggests these may be weak or null given the dominance of time-invariant institutional heterogeneity. Within countries over time, negative associations are predicted between empowerment indicators and anaemia prevalence (H3: TSER-PA; H4: FLFP-PA), and positive associations between empowerment and HPV

coverage (H5: TSER-HPV; H6: FLFP-HPV). The anaemia hypotheses H3 and H4 may be confounded by the temporal lag, dietary transition, and compositional shift mechanisms identified in this review, while H5 and H6 may be attenuated by the post-2020 supply-side HPV programme scale-up that dominates within-country HPV variation in the full sample.

Research Methodology

Research Design

The study adopts a quantitative, positivist research design, moving from theoretically derived hypotheses to empirical tests using secondary macro-level data (Bryman, 2016). Quantitative cross-national research designs are appropriate when research questions concern structural relationships at the country level that operate through measurable aggregate indicators, and when the goal is systematic generalisation across a large and diverse sample rather than deep contextual understanding of individual cases. The positivist epistemological position treats social and economic phenomena as governed by discoverable regularities that can be identified through rigorous statistical analysis, provided appropriate controls are applied for confounding factors. Two complementary analytical designs address the three research questions. Analysis A uses cross-sectional OLS regression to examine between-country variation in FLFP. Analysis B uses fixed-effects panel regression to examine within-country variation in women's health outcomes over 2015–2024. The two-design architecture is methodologically deliberate: the cross-section addresses RQ1 while simultaneously demonstrating its own limitations as a causal identification strategy, thereby providing empirical — rather than merely theoretical — motivation for the fixed-effects design that addresses RQ2a and RQ2b with stronger identification.

Data Sources and Sample

Analysis A. Data were obtained from three international repositories. Female labour force participation rate (FLFP), female tertiary school enrolment rate (TSER), total fertility rate (FR), and GDP per capita in constant 2015 USD were obtained from the World Bank World Development Indicators (WDI, 2024), which aggregates data from national statistical offices and international organisations. Female firm ownership (WMP) — defined as the percentage of surveyed firms with at least one female owner among the principal owners — was obtained from the World Bank Enterprise Surveys (WBES, 2024), which conduct stratified random sampling in over 150 developing and emerging economies. The cross-sectional sample covers up to 150 countries, with the analytical sample ranging from N = 64 (models including WMP, due to Enterprise Survey geographic coverage) to N = 119–128 for WDI-sourced variables. All observations refer to the most recently available year within 2015–2022 for each country.

Analysis B. The analysis B is built on the basis of 44 countries that are observed in the course of the year (2015-2024). The countries were picked according to the information they had on all six important variables. The prevalence of anaemia (percentage of women aged 1549 who have haemoglobin below 12 g/dL) was obtained using the WHO Global Health Observatory

(WHO GHO, 2024). The WHO-UNICEF Estimates of National Immunization Coverage (WUENIC, 2024) provided the information on HPV vaccination coverage (percentage of girls aged 914 who receive at least one dose). The rest of the variables were taken over to WDI (2024). A binary COVID-19 binary was coded 0 (20152019) and 1 (20202024), which is the onset of pandemic-related structural disruption to healthcare systems and economic activity.

Variables

There are three dependent variables used. Analysis A uses FLFP (percentage of women aged 15 and above who are employed or actively seeking employment) as the dependent variable, which encompasses the wide range of the female economic margin. In Analysis B, anaemia prevalence (PA) and HPV vaccination coverage (HPV) are dependent variables both operationalised according to WHO standard definitions.

TSER is the main predictor of empowerment in both analyses, which is the gross enrolment ratio of tertiary education - the number of students enrolled in tertiary education as a proportion of the population of official tertiary education age, regardless of age - which is consistent with UNESCO conventions. Values can be higher than 100% because there are over-age and under-age students. The second empowerment predictor in Analysis A is WMP (percentage of the surveyed firms with at least one female principal owner). In Analysis B, FLFP is the second predictor of empowerment in education and economic participation in health outcomes to allow de-editing of the educational and economic participation power effects on health outcomes.

The controls are on GDP (natural logarithmic GDP per capita in constant 2015 USD), calculated using the standard of the logarithmic transformation in cross-national development research to reduce the impact of extreme values and linearise the approximately log-normal income distribution across countries; FR (total fertility rate), which controls the effect of demographic structure, reproductive health environment and the well-established education fertility relationship; and in Analysis B only, the COVID-19 dummy. Analysis B captures all the time-invariant country-specific unobserved heterogeneity, all all-country common shocks, respectively.

Analytical Methods

Ordinary Least Squares — Analysis A. There are three progressive OLS model specifications that are estimated. Model A1 is a bivariate baseline regressing FLFP on TSER alone. Model A2 adds ln_GDP and FR as controls to assess whether the education-participation association strengthens when income and fertility context are held constant. Model A3 — the main cross-sectional model — adds WMP to the Model A2 specification, restricting the sample to 64 countries with available Enterprise Survey data. All models are estimated with HC3 heteroscedasticity-robust standard errors (Davidson & MacKinnon, 1993) to guard against violation of the OLS homoscedasticity assumption in a globally diverse country sample, where residual variance is expected to correlate with income level. HC3 standard errors are particularly

appropriate for small cross-national samples where leverage points — large influential countries — can distort conventional standard error estimates. Variance inflation factors (VIF) are computed for Model A3 to formally assess multicollinearity, with threshold values of VIF > 5 and VIF > 10 indicating moderate and serious concern respectively.

Fixed-Effects Panel Regression — Analysis B. Two fixed-effects panel models are estimated using the within estimator, which identifies all coefficients exclusively from within-country variation by demeaning each variable at the country level before estimation. This drives all time-invariant unobserved country characteristics, such as geography, colonial history, legal tradition, baseline healthcare infrastructure and cultural norms, to zero. The year fixed effects further internalize all the shocks that are common to all countries within a given year like commodity price fluctuations, global health campaigns and macroeconomic cycles. The location of the clustering of standard errors is at the country level; this is to account the behavior of the standard errors clustering at the country level and over time, which would otherwise make the standard errors to be understated and the t-statistics inflated. The Hausman (1978) specification test formally determines whether fixed or random effects are appropriate by assessing whether country-specific effects are correlated with the regressors: if they are, random effects estimates are inconsistent and fixed effects must be used. Decisively significant Hausman statistics for both models — $\chi^2(5) = 35.794$, $p < .001$ for anaemia and $\chi^2(5) = 25.262$, $p < .001$ for HPV — confirm fixed effects as the statistically appropriate estimator. Model fit is reported using the within-R^2, which measures the proportion of within-country variation in the dependent variable explained by the model after absorbing country-level means. Five robustness specifications probe the boundary conditions of the main findings: removing the COVID-19 dummy; restricting to the pre-pandemic subsample (2015–2019) for both health outcomes; and entering TSER and FLFP individually in the anaemia model.

Study Hypotheses

Based on the theoretical framework and literature reviewed in Section 2, six hypotheses are formally tested.

H1: TSER will be positively associated with FLFP across countries in the cross-sectional sample.

H2: WMP will be positively associated with FLFP across countries.

H3: Within-country increases in TSER over time will be negatively associated with anaemia prevalence (PA).

H4: Within-country increases in FLFP over time will be negatively associated with anaemia prevalence.

H5: Within-country increases in TSER will be positively associated with HPV vaccination coverage.

H6: Within-country increases in FLFP will be positively associated with HPV vaccination coverage.

Merits of the Research Design

There are a number of methodological strengths of the research design. First, the two design architecture - combining cross-sectional OLS and fixed effects panel regression - offers a more comprehensive empirical explanation than either of the two designs alone, directly demonstrating as opposed to simply asserting why panel identification is better than cross-sectional OLS in the empowerment-outcomes literature. Second, the most plausible non-experimental identification strategy that can be applied to macro-level panel data is fixed-effects estimation with country and year fixed effects and country-clustered standard errors. Third, Houseman specification test formally validates the fixed-effects choice instead of making it theoretically as an assumption and the estimates of the coefficients are defensible in terms of econometric. Fourth, Analysis A has HC3 robust standard errors which guarantees valid inference in the face of realistic departures of homoscedasticity and serial independence. Analysis B has HC3 robust standard errors which ensures inference is valid under realistic departures of homoscedasticity and serial independence. Fifth, the robustness framework of five specifications is a systematic assessment of the boundary conditions and stability of the primary findings, and shows that the counterintuitive positive association between TSER-anaemia is not an artefact of pandemic confounding, model specification, or multicollinearity. Sixth, the paper uses various internationally recognised repositories - World Bank WDI, WBES, WHO GHO and WUENIC - all of which are subject to international review of quality, there is less likelihood that measurement error in one source will drive the findings. The combination of WDI to economic and educational variables, WBES to firm ownership, and WHO GHO and WUENIC to health outcomes is the current best-practice in macro-level multi-domain empowerment research, and ensures that results can be reproduced by looking at publicly accessible international databases.

Data Analysis and Results

This section presents results sequentially: descriptive statistics, correlation analysis, Analysis A OLS regression, Analysis B fixed-effects regression, year fixed effects, Hausman specification test, and robustness checks. Significance levels are denoted †p < .10, *p < .05, **p < .01, ***p < .001 throughout.

Descriptive Statistics

Analysis A. Table 1 presents descriptive statistics. FLFP averaged 51.40% (SD = 15.00) across 119 countries, ranging from 5.12% to 82.92%, confirming substantial cross-national variation. TSER averaged 54.06% (SD = 36.47) with a maximum of 144.78%, reflecting the gross enrolment ratio methodology. WMP averaged 35.68% (SD = 13.34). Ln_GDP ranged from −2.26 to 3.78, corresponding to approximately USD 130 to USD 44,000 per capita. FR averaged 2.44 births per woman with positive skewness (0.91), reflecting a right tail of high-fertility economies.

Table 1

Descriptive Statistics — Analysis A: Cross-Sectional Sample (N = 64–128)

Variable	N	M	SD	Min	Mdn	Max	Skewness
FLFP (%)	119	51.40	15.00	5.12	53.59	82.92	-0.77
TSER (%)	110	54.06	36.47	2.23	52.44	144.78	0.31
WMP (%)	111	35.68	13.34	4.17	35.60	72.55	0.06
ln_GDP	115	1.12	0.79	-2.26	1.30	3.78	-0.96
FR (births/woman)	128	2.44	1.17	0.97	1.99	6.03	0.91

Note. FLFP = Female Labour Force Participation Rate; TSER = Tertiary School Enrolment Rate (Female); WMP = Firms with Female Ownership; ln_GDP = log GDP per capita (constant 2015 USD); FR = Total Fertility Rate.

Analysis B. Table 2 presents panel descriptive statistics. Anaemia prevalence averaged 18.95% (SD = 5.75) across 378 country-year observations, within the WHO mild-to-moderate range. HPV coverage showed the widest dispersion (M = 54.54%, SD = 28.24, range 0–99%), reflecting uneven national programme rollout. TSER averaged 71.50%, higher than the cross-sectional mean (54.06%), reflecting the educational advancement of the 44-country panel. The COVID-19 dummy had M = 0.50, confirming its balanced design.

Table 2
Descriptive Statistics — Analysis B: Panel Sample (N = 303–420)

Variable	N	M	SD	Min	Mdn	Max	Skewness
PA — Anaemia (%)	378	18.95	5.75	10.10	18.10	33.80	0.63
HPV — Vaccination (%)	392	54.54	28.24	0.00	59.00	99.00	-0.56
TSER (%)	396	71.50	30.26	5.70	77.27	144.78	-0.26
FLFP (%)	420	52.89	7.42	29.60	53.48	72.00	-0.35
ln_GDP	350	1.13	0.91	-3.15	1.28	3.20	-1.48
FR (births/woman)	420	1.74	0.53	0.97	1.60	4.11	2.07
COVID Dummy	420	0.50	0.50	0.00	0.50	1.00	0.00

Note. PA = Prevalence of Anaemia (% women aged 15–49); HPV = HPV Vaccination Coverage (% girls aged 9–14). COVID Dummy coded 0 for 2015–2019 and 1 for 2020–2024.

Correlation Analysis

Analysis A. Table 3 presents pairwise correlations. The strongest inter-predictor correlation is TSER–FR (r = −.711), consistent with demographic transition theory (Barro & Lee, 2013). VIF values (all below 2.0; Section 4.3.1) confirm multicollinearity is not a concern. FLFP–TSER correlation is weak and positive (r = .142), foreshadowing the null result in Analysis A. The near-zero FLFP–ln_GDP correlation (r = −.012) reflects the well-documented U-shaped income-participation relationship (Goldin, 1995).

Table 3
Pairwise Correlation Matrix — Analysis A Variables

Variable	1. FLFP	2. TSER	3. WMP	4. ln_GDP	5. FR
1. FLFP	—	0.142	0.279	-0.012	-0.011
2. TSER	0.142	—	-0.072	-0.288	-0.711*
3. WMP	0.279	-0.072	—	-0.104	-0.025
4. ln_GDP	-0.012	-0.288	-0.104	—	0.317*
5. FR	-0.011	-0.711*	-0.025	0.317*	—

Note. N = 64–128. Asterisk (*) indicates |r| ≥ .30.

Analysis B. Table 4 presents panel correlations, which pool within- and between-country variation; the fixed-effects estimator isolates only the within-country component. Anaemia shows moderate negative correlations with FLFP (r = −.321) and TSER (r = −.257). Critically, the TSER–PA association reverses sign in the fixed-effects model — from r = −.257 in the pooled correlation to b = +0.035 — demonstrating that between-country and within-country longitudinal analyses can produce fundamentally different inferences. This sign reversal is one of the most analytically important findings of the study.

Table 4
Pairwise Correlation Matrix — Analysis B Panel Variables

Variable	1. PA	2. HPV	3. TSER	4. FLFP	5. ln_GDP	6. FR
1. PA	—	-0.170	-0.257*	-0.321*	0.081	0.057
2. HPV	-0.170	—	0.105	0.292*	-0.151	-0.078
3. TSER	-0.257*	0.105	—	0.318*	-0.054	-0.442*
4. FLFP	-0.321*	0.292*	0.318*	—	-0.099	-0.142
5. ln_GDP	0.081	-0.151	-0.054	-0.099	—	0.202
6. FR	0.057	-0.078	-0.442*	-0.142	0.202	—

Note. N = 303–420. Asterisk (*) indicates |r| ≥ .25. Correlations pool within- and between-country variation.

Analysis A: Cross-Sectional OLS Regression

Multicollinearity assessment. VIF values for Model A3: TSER = 1.82, WMP = 1.01, ln_GDP = 1.02, FR = 1.82. All are well below the threshold of 5.0. Multicollinearity is not a threat to coefficient reliability in any model.

Table 5 presents OLS results across all three specifications. In Model A1, TSER is positive but non-significant (b = 0.045, SE = 0.043, p = .297; R^2 = .013). Model A2 marginally strengthens TSER (b = 0.083, p = .079†; R^2 = .021), though neither control is significant and the adjusted R^2 is negative (−.012). In Model A3, TSER (b = 0.058, p = .374) and WMP (b = 0.146, p = .400) both fail to reach significance; R^2 = .059, Adj. R^2 = −.005. H1 and H2 are not supported. The null result on RQ1 — robust across all three specifications — reflects the dominance of time-invariant institutional factors in cross-country FLFP variation (Gaddis & Klasen, 2014), motivating the within-country panel design of Analysis B.

Table 5
Cross-Sectional OLS Regression Results — Dependent Variable: FLFP

	A1 — Baseline		**A2 — Controls**		**A3 — Full Model**	
Variable	**b**	**SE**	**b**	**SE**	**b**	**SE**
Intercept	48.852***	(3.372)	43.222***	(6.156)	35.224***	(9.511)
TSER	0.045	(0.043)	0.083†	(0.046)	0.058	(0.065)
WMP					0.146	(0.172)
ln_GDP			-0.078	(1.552)	0.467	(2.821)
FR			1.574	(2.063)	3.004	(1.933)
N	93		93		64	
R^2	.013		.021		.059	
Adj. R^2	.002		-.012		-.005	

Note. HC3 heteroscedasticity-robust SE. †p < .10. *p < .05. **p < .01. ***p < .001. Model A3 N = 64 due to Enterprise Survey coverage.

Analysis B: Fixed-Effects Panel Regression

Table 6 presents the main fixed-effects results. The Hausman test (Section 4.6) confirms fixed effects is the appropriate specification for both models.

Table 6
Fixed-Effects Panel Regression Results — Analysis B

	Model B1: PA		Model B2: HPV		VIF	
Variable	**b**	**SE**	**b**	**SE**	**PA**	**HPV**
TSER	0.035**	(0.015)	0.414†	(0.229)	—	—
FLFP	-0.030	(0.041)	0.721	(0.670)	—	—
ln_GDP	-0.022	(0.066)	-2.094†	(1.252)	—	—
FR	2.689†	(1.563)	43.937***	(11.814)	—	—
COVID Dummy	2.869***	(0.511)	17.753***	(5.264)	—	—
Country FE	Yes		Yes			
Year FE	Yes		Yes			
N (obs)	306		303			
Within R^2	.624		.185			
Hausman χ^2	35.794***		25.262***			

Note. Fixed-effects within-estimator coefficients. SE clustered at country level. Year dummy coefficients in Tables 7 and 8. †p < .10. *p < .05. **p < .01. ***p < .001.

Model B1: Anaemia Prevalence. Estimated on 306 country-year observations; within-R^2 = .624. TSER is a statistically significant positive predictor (b = 0.035, SE = 0.015, p = .024), contradicting H3: a one percentage point increase in TSER is associated with a 0.035 percentage point increase in anaemia prevalence. FLFP is not significant (b = −0.030, p = .474), rejecting H4. FR is marginally significant and positive (b = 2.689, p = .087), consistent with iron depletion from repeated pregnancies. The COVID-19 dummy is the strongest predictor (b = 2.869, p < .001): anaemia was on average 2.87 percentage points higher in 2020–2024 than in 2015–2019, after controlling for all other variables and country fixed effects.

Model B2: HPV Vaccination Coverage. Estimated on 303 country-year observations; within-R^2 = .185, reflecting that HPV coverage is primarily driven by discrete national programme decisions. TSER is marginally significantly positive (b = 0.414, SE = 0.229, p = .072), providing partial support for H5. FLFP is not significant (b = 0.721, p = .283), rejecting H6. FR is a large, highly significant positive predictor (b = 43.937, p < .001). The COVID-19 dummy is dominant (b = 17.753, p = .001), reflecting global acceleration of HPV programmes through WHO and GAVI catch-up initiatives post-2020.

Year Fixed Effects

Anaemia (PA). Table 7 reports year fixed effects. Coefficients increase monotonically from 0.187 in 2016 (p = .004) to 1.002 in 2019 (p < .001), indicating a systematic pre-pandemic rising anaemia trend not explained by any covariate. After the COVID dummy absorbs the average pandemic-period shift, residual year dummies for 2020 (−1.328), 2021 (−1.313), and 2022 (−0.622) are all negative (all p < .001).

Table 7
Year Fixed Effect Coefficients — Model B1: Anaemia (PA)

Year	2016	2017	2018	2019	2020	2021	2022
b	0.187***	0.381***	0.639***	1.002***	-1.328***	-1.313***	-0.622***

Note. Relative to 2015 baseline. COVID-19 dummy (b = 2.869***) captures average level shift across 2020–2024.

HPV Vaccination. Table 8 reports HPV year effects. Coverage rose in 2018–2019 (b = 6.430† in 2019), collapsed in 2021 (−16.484***) and 2022 (−13.801***) due to pandemic disruption of routine immunisation, and partially recovered in 2023 (−9.556**).

Table 8
Year Fixed Effect Coefficients — Model B2: HPV Vaccination

Year	2016	2017	2018	2019	2020	2021	2022	2023
b	-3.338	-0.394	2.571	6.430†	-7.077	-16.484***	-13.801***	-9.556**

Note. Relative to 2015 baseline. †p < .10. **p < .01. ***p < .001.

Hausman Specification Test

Table 9 reports results. For anaemia, $\chi^2(5) = 35.794$ (p < .001); for HPV, $\chi^2(5) = 25.262$ (p < .001). Both tests decisively reject the null that random effects are consistent, confirming that fixed-effects estimation is the appropriate specification for both models.

Table 9
Hausman Specification Test — Fixed Effects vs Random Effects

Model	χ^2	df	p	Decision
Anaemia (PA)	35.794	5	< .001	Fixed Effects preferred
HPV Vaccination	25.262	5	< .001	Fixed Effects preferred

Note. H_0: random effects are consistent. Rejection (p < .05) indicates fixed effects is the appropriate estimator.

Robustness Checks

Table 10 provides a summary of five specifications of the robustness analysis, which include country and year fixed effects and clustered country-level standard errors. Specification 1 (PA without COVID dummy) leaves TSER unchanged (b = 0.035**, p = .024), confirming robustness to pandemic control choice. Specification 2 (PA, pre-pandemic 2015–2019) strengthens TSER to b = 0.062*** and renders FLFP significant and positive (b = 0.108**, p = .017), confirming the counterintuitive positive direction predates the pandemic. Specification 3 (HPV, pre-pandemic) yields a substantially greater coefficient on TSER (b = 0.778**, p = .018) which is the strongest evidence for the education-vaccination hypothesis, and the positive coefficient on FLFP (b = 0.108**, p = .017) is neither robustly negative nor robustly positive, supporting both the education-vaccination hypothesis and the hypothesis that there is no relationship between these variables. In Specification 4 (TSER only) b = 0.032** (which is almost the same as 0.035), meaning that multicollinearity is not a problem. In Specification 5, only FLFP (not p) is predictive of anaemia, with the coefficient being b = 0.008 (p = .855).

Table 10
Robustness Check Results — Fixed-Effects Panel Models

Variable	(1) No COVID b (SE)	(2) PA Pre-pan. b (SE)	(3) HPV Pre-pan. b (SE)	(4) TSER only b (SE)	(5) FLFP only b (SE)	N
TSER	0.035** (0.015)	0.062*** (0.017)	0.778** (0.325)	0.032** (0.015)	—	
FLFP	-0.030 (0.041)	0.108** (0.045)	-0.763 (1.064)	—	0.008 (0.044)	
Country & Year FE	Yes	Yes	Yes	Yes	Yes	
N	306	~154	~154	306	306	

Note. ln_GDP and FR included in all specifications. †p < .10. *p < .05. **p < .01. ***p < .001.

Summary of Empirical Findings

On RQ1: neither TSER nor WMP significantly predicts FLFP in any cross-sectional specification; H1 and H2 are not supported. On RQ2a (anaemia): TSER is significantly and positively associated with anaemia (b = 0.035, p = .024), contradicting H3; FLFP is not significant, rejecting H4; COVID-19 is the dominant predictor (b = 2.869, p < .001); and a systematic pre-pandemic rising anaemia trend is documented in the year fixed effects. On RQ2b (HPV): TSER is marginally positively associated with HPV coverage (b = 0.414, p = .072), providing partial support for H5; FLFP is not significant, rejecting H6; the pre-pandemic subsample yields a substantially stronger and significant TSER coefficient (b = 0.778, p = .018).

Discussion

This section situates the empirical findings within the theoretical frameworks and literature reviewed in Section 2, addressing each principal result in turn and offering mechanistic explanations grounded in the data and prior evidence.

The Null Cross-Sectional Result: What It Means and Why It Is Retained

Why the null result is methodologically expected. The fact that Analysis A did not find a significant TSER-FLFP or WMP-FLFP association is in line with the structural limitations of cross-sectional OLS in cross-national samples. The U-shaped curve of income-participation described by Goldin (1995) would imply that a positive linear coefficient used across the entire range of income would be weak whether or not the true within-country relationship is one of very poor versus very rich. This is proven by the fact that in case the aggregate national wealth cannot predict the cross-country FLFP in a linear way ($r = -.012$), much narrower measures of empowerment would similarly be limited. Gaddis and Klasen (2014) affirm that the cross-country FLFP difference is preponderated by time-invariant institutional factors family law, occupational segregation, childcare infrastructure, religious context - which cannot be satisfactorily represented by a single aggregate indicator. The framework by Kabeer (1999) further adds that the educational resources can only translate into the labour market participation under enabling institutional conditions; the cross-section is the net effect of both enabling institutional conditions and constraints operating simultaneously with the result being a coefficient that reflects the combined influence of both enabling institutional conditions and constraints at the same time rather than the pure empowerment pathway. The null finding is not, therefore, a failure of the analysis but a very informative finding on the limits of cross-sectional identification in this field.

Why the cross-sectional analysis remains analytically valuable. Although the null results were found, two essential contributions were made with the help of the cross-section. First, it gives the empirical, not just theoretical foundation, of the choice of design to use, which in this case is the fixed-effects design choice. The validity of the panel identification, as Wooldridge (2010) argues, is based in part on the ability to show that between-country comparisons fail to identify the relationship of interest, and thus the need to infer within-country longitudinal relationships exists. A study that goes directly to the panel estimation without this demonstration does not have such a foundation. Second, the cross-section shows that there is a critical sign reversal between the pooled TSER-PA correlation ($r = -.257$) and within-country fixed-effects coefficient ($b = +0.035$). The two associations which are in opposite directions illustrate the risk of ecological fallacy which Wooldridge (2010) warns against: aggregate cross-sectional associations can have a systematic tendency to misrepresent the direction of within-unit causal relationships. This reversal of the sign would otherwise be unnoticed, and the result of the fixed-effects would find itself deprived of any comparative reference point to give it interpretive weight.

The Counterintuitive Anaemia Finding

The positive within-country TSER-anaemia association — stable between b = 0.032 and b = 0.062 across all five robustness specifications, and stronger in the pre-pandemic period — challenges the foundational literature linking female education to improved nutritional outcomes. Three non-exclusive mechanisms offer a coherent account.

Short-run compositional effects. Rapid tertiary expansion concentrates a larger share of the female population in the 18–25 age bracket — the highest-risk group for iron-deficiency anaemia due to menstrual iron losses and the nutritional demands of intensive study that are not always met in campus environments. If enrolment expansion is sufficiently rapid to shift the age composition of the measured female population within a 10-year panel, national anaemia prevalence may rise in the short run even as the long-run educational dividend eventually reduces it. The pre-pandemic intensification of the TSER coefficient (b = 0.062, $p < .001$ in 2015–2019) is consistent with this mechanism: the fastest expansion years coincide with the period before newly enrolled women have translated their education into health-protective household behaviours.

Dietary transition. Educational expansion accompanies urbanisation and a dietary shift away from traditional iron-rich staples toward processed and convenience foods with lower bioavailable iron content (Popkin, 2006). TSER, in years of rapid expansion, serves partly as a proxy for urbanisation and the associated nutritional transition, producing a positive within-country association with anaemia that reflects correlated structural changes rather than a direct education-anaemia causal channel. The monotonically rising pre-pandemic year dummies (0.187 to 1.002 from 2016 to 2019) provide corroborating evidence of a structural secular worsening in within-country anaemia during this period that the covariate set cannot explain.

Temporal lag. The health return to female education operates primarily through inter-generational mechanisms: maternal education most strongly predicts offspring health rather than contemporaneous adult health, because the channels — dietary knowledge, income accumulation, and health-seeking behaviour — unfold gradually over time (Psacharopoulos & Patrinos, 2018). A 10-year panel may be insufficient to observe the health dividend of educational expansion underway during 2015–2024. These three mechanisms are mutually consistent and do not refute the education-health linkage; they qualify it temporally and contextually. The finding adds specificity to Sen's (1999) capability approach: educational capabilities do not automatically convert into health functionings within a single decade, particularly when rapid expansion coincides with urbanisation-linked dietary change and demographic compositional shifts.

The HPV Vaccination Finding

The marginal TSER-HPV association (b = 0.414, p = .072 full sample; b = 0.778, p = .018 pre-pandemic) is directionally consistent with the demand-side educational hypothesis. Bonney et al. (2018) find maternal education consistently associated with HPV awareness and acceptance; Gertig et al. (2013) document that socioeconomic status predicts vaccination

completion in national programmes; and Jit et al. (2014) model that demand-side educational interventions substantially accelerate coverage in middle-income settings. The mechanism is straightforward: tertiary-educated women navigate health systems more effectively, respond to health information through professional and peer networks, and are more likely to seek preventive care for their daughters. The attenuation of the TSER coefficient in the full sample reflects the dominance of the COVID-19 dummy ($b = 17.753$, $p = .001$), which captures massive post-2020 GAVI-funded scale-up: when this common shock drives the majority of within-country HPV variation in 2020–2024, the slower demand-side education channel becomes harder to distinguish from residual variance. The pre-pandemic subsample, by isolating the five years before this structural break, recovers the underlying education-vaccination signal more cleanly. The FLFP null may reflect that the panel sample is predominantly upper-middle and high-income economies where most employed women already have healthcare access independent of employment status, and that aggregate FLFP conflates formal and informal employment with substantially different health access implications.

The Dominant Role of COVID-19

COVID-19 is the single largest predictor of within-country health outcome variation: $b = 2.869$ for anaemia and $b = 17.753$ for HPV. These magnitudes confirm that 2015–2024 cannot be treated as analytically homogeneous. The consistently stronger empowerment coefficients in the pre-pandemic subsample confirm that pandemic disruption attenuated the measurable signal of educational effects within the observation window. The pre-pandemic rising anaemia trend (year coefficients 0.187 to 1.002 across 2016–2019, all $p < .001$) is not explained by any model covariate and warrants investigation as a potential signal of declining food system nutrition quality during that period, possibly reflecting declining real food assistance budgets, commodity price pressures, or structural shifts in agricultural systems across the panel countries.

Theoretical Contributions

This study makes three contributions. First, it provides within-country longitudinal evidence that the short-run education-health relationship can diverge from the long-run cross-country relationship, illustrating the distinction between educational capability acquisition and health functioning conversion in Sen's (1999) framework. The finding that within-country educational expansion is associated with higher short-run anaemia — while countries with historically higher TSER have lower anaemia in cross-section — concretely demonstrates that the timescale of capability-to-functioning conversion matters for how empowerment research is designed and interpreted. Second, the study contributes methodologically by demonstrating that cross-sectional and panel designs are necessary complements: the sign reversal between $r = -.257$ (pooled) and $b = +0.035$ (fixed-effects) illustrates why cross-national statistics cannot substitute for within-country longitudinal inference. Third, it extends the post-pandemic gender development literature by quantifying the extent to which COVID-19 attenuated the measurable signal of educational effects within the finite observation window, suggesting that studies

covering only 2015–2024 may systematically underestimate the education-health relationship and should treat the pandemic as a structural break rather than a level shift. The consistently larger TSER coefficients in the pre-pandemic subsample across both health models provide empirical confirmation that the pandemic attenuated rather than created the education-health relationships this study documents.

Conclusion

General Findings

Three principal findings emerge from this study. First, cross-sectional empowerment measures do not significantly predict female labour force participation across countries. Neither TSER ($b = 0.058$, $p = .374$) nor WMP ($b = 0.146$, $p = .400$) reaches statistical significance in any of the three OLS specifications, and no model achieves a positive adjusted R^2. H1 and H2 are not supported. The null result is consistent with the dominance of time-invariant institutional and cultural factors in cross-country FLFP determination and cautions against using cross-country rankings of educational enrolment or firm ownership as proxies for the dynamic within-country processes through which empowerment expands labour market access.

Second, within-country increases in female tertiary enrolment are associated with short-run increases in anaemia prevalence ($b = 0.035$, $p = .024$), not reductions. This counterintuitive result is robust across all five robustness specifications and is stronger in the pre-pandemic period ($b = 0.062$, $p < .001$), ruling out the pandemic as its cause. It is explained by short-run compositional effects as tertiary expansion grows the highest-risk demographic cohort, by dietary transition dynamics accompanying educational and urban expansion, and by the insufficient length of the 10-year observation window relative to the generational timescale over which education's health dividend accrues. H3 is rejected in its expected direction; H4 is not supported. The COVID-19 pandemic is the dominant predictor of within-country anaemia variation ($b = 2.869$, $p < .001$), with a pre-pandemic rising trend also operating independently.

Third, within-country TSER increases are marginally positively associated with HPV vaccination coverage ($b = 0.414$, $p = .072$ full sample; $b = 0.778$, $p = .018$ pre-pandemic), providing partial support for H5 through the demand-side educational mechanism. FLFP is not a significant predictor of either health outcome in any fixed-effects specification, rejecting H4 and H6.

Practical Implications

Within-country longitudinal tracking over cross-country benchmarking. The null cross-sectional result implies that cross-country league tables of tertiary enrolment and firm ownership tell us little about the dynamic within-country processes through which educational investment translates into labour market engagement. Governments should evaluate empowerment policies through longitudinal within-country outcome tracking rather than international benchmarking comparisons.

Tertiary expansion must be paired with nutritional interventions. The positive short-run TSER-anaemia relationship suggests that increasing tertiary enrolment without nutritional support would possibly temporarily increase the burden of anaemia among 18 to 25 years cohort that was most underserved by the current programmes. Iron supplementation on campuses, dietary counselling, and provision of fortified foods are complements that are low-cost and reach large numbers of people, and should be explicitly included in national tertiary expansion programs.

Demand-side educational investment complements HPV supply expansion. The pre-pandemic finding on HPV coverage gains ($b = 0.778$, $p = .018$) suggests that the gains in coverage due to supply are precarious in the face of disruption in the health system. This is concretely represented by the year dummies of 2021 (−16.484) and 2022 (−13.801): two years of HPV coverage gains made by the pandemic wiped out. The creation of more resilient uptake based on household health literacy and not on institutional delivery capacity alone, is achieved through demand-side empowerment established through educational investment.

Post-pandemic recovery must target young women urgently. In 2023, the average anaemia was 2.87 percentage points higher during 20202024 with an average anaemia of 2.0523 in 2023. The recovery investments need to re-establish disrupted nutrition programmes and conduct HPV catch-up campaigns to the cohorts that have not had vaccination in 20202022 before they age out of the target window.

FLFP quality matters as much as quantity. The uniform FLFP null of all panel specifications would suggest that aggregate participation rates do not generate any measured health dividend without the employment generated bearing formal-sector institutional health access features. The kind, formality and composition of women in the sector should not be ignored in policy strategies, but rather the volume of women work in the sector.

Limitations

Observational panel data cannot be used to establish causality: two-way fixed-effects design would not sufficiently control time-varying omitted variables. The panel of 44 countries is biased towards upper-middle and high-income economies, and cannot be generalised to lower-income situations. All variables are measured at national level, masking subnational heterogeneity. The 10-year observation window is short relative to the generational timescale of the education-health dividend. The binary COVID-19 dummy only partially captures the structural break; heterogeneous pandemic severity and recovery rates contribute differently to the estimated average effects.

Directions for Future Research

Five extensions merit priority. First, extending the observation window to 2035 would allow testing of whether the positive short-run TSER-anaemia association reverses as the educational cohort of 2015–2020 reaches household decision-making maturity and begins translating its education into improved nutritional outcomes. The pre-pandemic HPV result

(b = 0.778, p = .018) already suggests that the education-health dividend on preventive care behaviour emerges faster than the nutritional dividend, and a longer window would allow both lags to be quantified and compared with greater precision. Second, disaggregating FLFP by employment formality and sector would allow testing of whether the employment-health mechanism operates exclusively through formal-sector channels, rendering aggregate participation rates uninformative about health outcomes. Third, subnational panel data from countries such as India, Brazil, Indonesia, and South Africa would enable higher-powered and geographically granular tests of the relationships identified here. Fourth, modelling interactions between empowerment investments and specific health programme interventions — iron supplementation, school-based HPV vaccination, fortified food provision — would reveal whether education and programme delivery generate non-linear complementarities. Fifth, income-group interaction analyses should examine whether the counterintuitive short-run TSER-anaemia positive association is specific to rapidly urbanising middle-income economies — where the dietary transition and compositional mechanisms are strongest — or whether it generalises across income levels.

The core paradox of this study's findings is instructive. Female educational expansion — one of the most robustly beneficial investments in the development literature — is associated within a 10-year panel with higher anaemia and only marginally higher HPV coverage, while the cross-national picture shows countries with higher historical TSER having substantially better health outcomes on both dimensions. It is not a contradiction; it is the empirical signature of the well-documented lag between investment in human capabilities, and the realisation of their benefits across social systems. Those countries which invested in female education in the previous decades are the ones which have better health outcomes today; and those countries which invest in female education today will see their health dividends in the next generation. This is because the temporal terrain is critical in the designing of the empowerment policies that are not only based on their ultimate destination but also to the realistic speed and conditions of the journey. A policy framework that anticipates an educational investment to bring about short-run health benefits will be frustrated; one that parallels it with specific nutritional and immunisation interventions in the interim will be far better placed to capture the entire dividend of female empowerment to the health of women across generations.

Works Cited

Agénor, P. R. (2017). A computable OLG model for gender and growth policy analysis. Macroeconomic Dynamics, 21(1), 154–189.

Akseer, N., Bhatti, Z., Rizvi, A., Salehi, A. S., Mashal, T., & Bhutta, Z. A. (2017). Coverage and inequalities in maternal and child health interventions in Afghanistan. BMC Public Health, 16(2), Article 687.

Alderman, H., & Headey, D. D. (2014). The nutritional returns to parental education (IFPRI Discussion Paper No. 01379). International Food Policy Research Institute.

Alon, T., Doepke, M., Olmstead-Rumsey, J., & Tertilt, M. (2020). The impact of COVID-19 on gender equality (NBER Working Paper No. 26947). National Bureau of Economic Research.

Aterido, R., Hallward-Driemeier, M., & Pagés, C. (2011). Big constraints to small firms' growth? Economic Development and Cultural Change, 59(3), 609–647.

Barro, R. J., & Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. Journal of Development Economics, 104, 184–198.

Becker, G. S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. National Bureau of Economic Research.

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis, M., Ezzati, M., Grantham-McGregor, S., Katz, J., Martorell, R., & Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. The Lancet, 382(9890), 427–451.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. Journal of Economic Literature, 55(3), 789–865.

Bluedorn, J., Caselli, F., Hansen, N.-J., Shibata, I., & Tavares, M. M. (2021). Gender and employment in the COVID-19 recession (IMF Working Paper WP/21/95). International Monetary Fund.

Boeri, T., Del Boca, D., & Pissarides, C. (Eds.). (2005). Women at work: An economic perspective. Oxford University Press.

Bonney, L. E. A., Tyrer, H. B., & Bhattacharya, S. (2018). Knowledge and attitudes to HPV vaccination among healthcare workers in Africa: A systematic review. European Journal of Obstetrics & Gynecology and Reproductive Biology, 225, 161–168.

Bruhn, M. (2009). License to sell: The effect of business registration reform on entrepreneurial activity in Mexico. Review of Economics and Statistics, 93(1), 382–386.

Bruni, L., Albero, G., Serrano, B., Mena, M., Gómez, D., Muñoz, J., Bosch, F. X., & de Sanjóse, S. (2016). HPV and related diseases in the world. Summary report. Institut Català d'Oncologia.

Bryman, A. (2016). Social research methods (5th ed.). Oxford University Press.

Caldwell, J. C. (1986). Routes to low mortality in poor countries. Population and Development Review, 12(2), 171–220.

Davidson, R., & MacKinnon, J. G. (1993). Estimation and inference in econometrics. Oxford University Press.

Duflo, E. (2012). Women empowerment and economic development. Journal of Economic Literature, 50(4), 1051–1079.

Fafchamps, M., & Amine, L. (2012). Female-owned firms and labour demand in Africa (Working Paper). Oxford University.

Gaddis, I., & Klasen, S. (2014). Economic development, structural change, and women's labor force participation: A reexamination of the feminization U hypothesis. Journal of Population Economics, 27(3), 639–681.

Gavi — The Vaccine Alliance. (2022). HPV vaccine support. https://www.gavi.org/vaccines-and-health/vaccine/hpv

Gertig, D. M., Brotherton, J. M. L., Budd, A. C., Drennan, K., Chappell, G., & Saville, A. M. (2013). Impact of a population-based HPV vaccination program on cervical abnormalities. British Journal of Cancer, 108(5), 1171–1177.

Glewwe, P. (1999). Why does mother's schooling raise child health in developing countries? Journal of Human Resources, 34(1), 124–159.

Goldin, C. (1995). The U-shaped female labor force functions in economic development and economic history. In T. P. Schultz (Ed.), Investment in women's human capital (pp. 61–90). University of Chicago Press.

Grantham-McGregor, S., & Ani, C. (2001). A review of studies on the effect of iron deficiency on cognitive development in children. Journal of Nutrition, 131(2), 649S–668S.

Haddad, L., Hoddinott, J., & Alderman, H. (Eds.). (1997). Intrahousehold resource allocation in developing countries. Johns Hopkins University Press.

Hallward-Driemeier, M. (2013). Enterprising women: Expanding economic opportunities in Africa. World Bank Publications.

Hausman, J. A. (1978). Specification tests in econometrics. Econometrica, 46(6), 1251–1271.

Hull, R., Mbelu, N., Simms, K. T., et al. (2021). Comparing the effectiveness and cost of HPV vaccination programmes in LMICs. BMJ Global Health, 6(3), e004048.

International Finance Corporation. (2019). Moving toward gender balance in private equity and venture capital. World Bank Group.

Jit, M., Brisson, M., Portnoy, A., & Hutubessy, R. (2014). Cost-effectiveness of female HPV vaccination in 179 countries. The Lancet Global Health, 2(7), e406–e414.

Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women's empowerment. Development and Change, 30(3), 435–464.

Kassebaum, N. J., & GBD 2013 Anemia Collaborators. (2016). The global burden of anemia. Hematology/Oncology Clinics of North America, 30(2), 247–308.

Klasen, S., & Pieters, J. (2015). What explains the stagnation of female labor force participation in urban India? World Bank Economic Review, 29(3), 449–478.

Kruk, M. E., Gage, A. D., Arsenault, C., Jordan, K., et al. (2018). High-quality health systems in the SDG era. The Lancet, 392(10160), 2115–2166.

Lund, C., Breen, A., Flisher, A. J., Kakuma, R., Corrigall, J., Joska, J. A., Swartz, L., & Patel, V. (2010). Poverty and common mental disorders in low and middle income countries. Social Science & Medicine, 71(3), 517–528.

Mammen, K., & Paxson, C. (2000). Women's work and economic development. Journal of Economic Perspectives, 14(4), 141–164.
Mincer, J. (1974). Schooling, experience, and earnings. National Bureau of Economic Research.
Moser, C. O. N. (1993). Gender planning and development: Theory, practice and training. Routledge.
Nandi, A., Glymour, M. M., & Subramanian, S. V. (2014). Association among socioeconomic status, health behaviors, and all-cause mortality in the United States. Epidemiology, 25(2), 170–177.
Paul, K. T., Wallenburg, I., & Bal, R. (2009). HPV vaccination in Austria and the Netherlands. Sociology of Health & Illness, 32(2), 220–235.
Popkin, B. M. (2006). Global nutrition dynamics: The world is shifting rapidly toward a diet linked with noncommunicable diseases. American Journal of Clinical Nutrition, 84(2), 289–298.
Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. Education Economics, 26(5), 445–458.
Rowlands, J. (1997). Questioning empowerment: Working with women in Honduras. Oxfam.
Schultz, T. P. (2002). Why governments should invest more to educate girls. World Development, 30(2), 207–225.
Schultz, T. W. (1961). Investment in human capital. American Economic Review, 51(1), 1–17.
Sen, A. (1999). Development as freedom. Oxford University Press.
Smith, L. C., & Haddad, L. (2015). Reducing child undernutrition. World Development, 68, 180–204.
Thomas, D. (1994). Like father, like son; like mother, like daughter. Journal of Human Resources, 29(4), 950–988.
UNICEF. (2021). COVID-19 and children: UNICEF data hub. https://data.unicef.org/covid-19-and-children
Verick, S. (2014). Female labor force participation in developing countries. IZA World of Labor, 87.
World Bank. (2022). Women, business and the law 2022. World Bank Publications.
World Bank Enterprise Surveys. (2024). Enterprise surveys indicators data. https://www.enterprisesurveys.org
World Bank World Development Indicators. (2024). World development indicators database. https://databank.worldbank.org/source/world-development-indicators
World Health Organization. (2011). Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity (WHO/NMH/NHD/MNM/11.1).
World Health Organization. (2021). Global anaemia estimates, 2021 edition. https://www.who.int/data/gho/data/themes/topics/anaemia_in_women_and_children
World Health Organization. (2022). Human papillomavirus (HPV) and cervical cancer. https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer
World Health Organization & UNICEF. (2021). WHO/UNICEF estimates of national immunization coverage (WUENIC). https://immunizationdata.who.int
Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2nd ed.). MIT Press.

Perceived Value as a Determinant of Willingness to Pay among Female Consumers: A Multi-category Business Study By Adya Sharma

Abstract

This study investigates the relationship between perceived value and willingness to buy among female adolescent consumers across three product categories of differing functional necessity. Using a within-subjects, stimulus-based survey design, 26 female participants aged 12–18 evaluated a handbag, a sanitary pad, and a lipstick — all priced within a comparable range of Rs. 300–700 — on five-item Likert-scale measures of perceived value and willingness to buy, as well as their maximum willingness to pay. Results reveal a consistent positive association between perceived value and willingness to buy across all three product categories, thereby supporting the hypothesis. However, the strength of this relationship varies by product type: it is most pronounced for functional or semi-essential products (sanitary pads and handbags) and comparatively weaker for discretionary items (lipstick), where personal preference and subjective utility play a greater moderating role. These findings suggest that while perceived value is a reliable predictor of purchase intention, its predictive power is modulated by a product's position on the essential–discretionary spectrum, with practical implications for product positioning, pricing strategy, and value communication in the Indian adolescent market.

Key Words

Perceived value, Willingness to buy, Consumer behaviour, Purchase intention, Product type, Price sensitivity, Perceived utility, Survey-based research, Quantitative analysis

1. Introduction

In consumer behaviour, purchase decisions are influenced by more than just the actual price of a product. Consumers often assess whether a product seems useful, appealing, and worth the money they are spending. This assessment is referred to as ‘perceived value’, and it plays an important role in shaping consumer willingness to buy (Zeithaml, 1988).

Perceived value can be defined as the value a consumer assigns to a product based on the benefits they expect to receive in relation to the costs involved. It is therefore a subjective judgment rather than an objective measure. Similarly, ‘willingness to buy’ refers to a consumer’s readiness or intention to purchase a product. When a product is perceived as valuable, consumers are generally more likely to show a greater willingness to buy it (Sharma & Sanjay, 2023).

This research examines the relationship between perceived value and willingness to buy in the context of three products: a sanitary pad, a handbag, and a lipstick. These products were selected because they fall within a similar price range of Rs. 300 to Rs. 700, but differ in terms of function and personal significance. By comparing these three items, the study aims to explore whether perceived value influences willingness to buy across different product categories.
This study hypothesises that perceived value predicts willingness to buy products. In other words, the study proposes that perceived value is directly proportionate to willingness to buy. To

test this hypothesis, a questionnaire-based survey will be conducted among female peers. The questionnaire includes questions designed to measure both perceived value and willingness to buy for each of the three products. This method allows for a structured collection of responses and makes it possible to examine the relationship between the two variables in a systematic way. Overall, this study seeks to contribute to a better understanding of how consumers make purchasing decisions. By focusing on perceived value and willingness to buy, it examines an important aspect of consumer behaviour that influences everyday buying choices.

2. Methodology

For this study, a within-subjects, stimulus-based survey design has been used to examine the relationship between perceived value and willingness to buy across three product categories: a handbag, a sanitary pad, and a lipstick. A within-subjects design was chosen because it allowed each participant to evaluate all three products, making it possible for me to compare responses across product types while reducing the effect of individual differences in consumer preferences.

2.1 Participants

The target audience consisted of female adolescents between the ages of 12 and 18, and responses were collected from a sample of 26 girls, with most participants falling in the 16 to 17-year-old age group. This group was selected because the three chosen products are commonly marketed to female consumers within this age range. The data was collected through an online survey that was distributed among female peers.

2.2 Stimulus Design

To carry out the study, three products that are frequently associated with female consumers: a handbag, a sanitary pad, and a lipstick were chosen because these items fall within a similar price range of Rs. 300 to Rs. 700, while differing in terms of function, necessity, and personal significance. This made them useful for comparing whether perceived value influences willingness to buy in the same way across different product categories. To minimise bias caused by brand recognition, images of the three products without any visible brand names or identifying labels were used.

In the survey, three variables: perceived value, willingness to buy, and the maximum price respondents were willing to pay were measured. Perceived value was assessed using five items adapted from prior research on consumer perceived value (Zeithaml, 1988; Sweeney & Soutar, 2001). The statements were as follows:

1. This product offers good value for money
2. The benefits of this product justify its price
3. This product seems like a worthwhile purchase
4. Compared to similar products, this seems valuable
5. This product meets my expectations for its price

Willingness to buy was assessed using five items adapted from established measures of consumer purchase intention (Dodds, Monroe, & Grewal, 1991). The statements were as follows:

1. I am likely to buy this product
2. I would consider purchasing this product
3. I have a high intention to buy this product
4. I would recommend buying this product to others
5. If available, I would definitely buy this product

For all of these statements, participants were asked to select responses on a Likert-type scale ranging from strongly disagree to strongly agree. Furthermore, participants were also asked to indicate the maximum amount they would be willing to pay for each product, with response options ranging from Rs. 300 to Rs. 700.

2.3 Data Collection and Analysis

The participants were presented with the product images through the online questionnaire. For each product, they were asked to respond to the five perceived value statements, the five willingness-to-buy statements, and the question about the maximum price they were willing to pay. This process was repeated for all three products so that each participant evaluated every stimulus. After collecting the responses, the data was tabulated and response patterns were compared across the three products. The levels of perceived value, willingness to buy, and willingness to pay for each product were examined in order to determine whether products that were perceived as more valuable also showed a greater willingness to buy, in line with my hypothesis.

3. Results and Discussion

This section reports the survey analysis, examining patterns in perceived value and willingness to buy across three product categories, followed by a critical discussion of the implications in light of the existing literature. This study examined whether perceived value predicts willingness to buy among female adolescent consumers across three product categories — a sanitary pad (essential), a handbag (semi-essential), and a lipstick (discretionary) — all held within a comparable price range of Rs. 300–700. Using a within-subjects Likert-scale survey administered to 26 female participants aged 12–18, perceived value and willingness to buy were measured for each product, and statistical analyses, including one-way ANOVA and Pearson correlation, were applied to assess both price differences across categories and the strength of association between the two variables. This design allowed for a controlled, systematic comparison of consumer behaviour across product types while minimising the influence of individual differences and brand bias.

The results have been presented using descriptive statistics, tabular summaries, and visual representations such as graphs and statistical analyses.

Table 1: Averages of all the responses for perceived value, willingness to buy and the max. price respondents were willing to pay for that product.

	Handbag	Sanitary Pad	Lipstick
Percieved Value	43.08%	38.48%	36.16%
Willingness to buy	41.54%	37.72%	30.76%
Max Price	500	300	500

Table 1 presents the averages of all responses for perceived value, willingness to buy, and the maximum price respondents were willing to pay for each product. Across all three products, higher perceived value scores corresponded with higher willingness to buy scores, providing an initial indication of the positive relationship proposed in the hypothesis.

Table 2: Description of the type of product and typical behaviour of consumers

Product Type	Behaviour
Essential (Sanitary Pads)	Stable buying (people buy it consistently)
Semi-essential (Handbag)	Strong alignment (seen as both valuable &desirable
Discretionary (Lipstick)	Drop-off in buying (depends on preference)

Table 2 categorises the three products by type and describes the typical consumer buying behaviour associated with each: sanitary pads, as an essential product, showed stable and consistent buying behaviour; the handbag, as a semi-essential product, showed strong alignment between perceived value and desirability; and the lipstick, as a discretionary product, showed a drop-off in buying intention that depended heavily on individual preference.

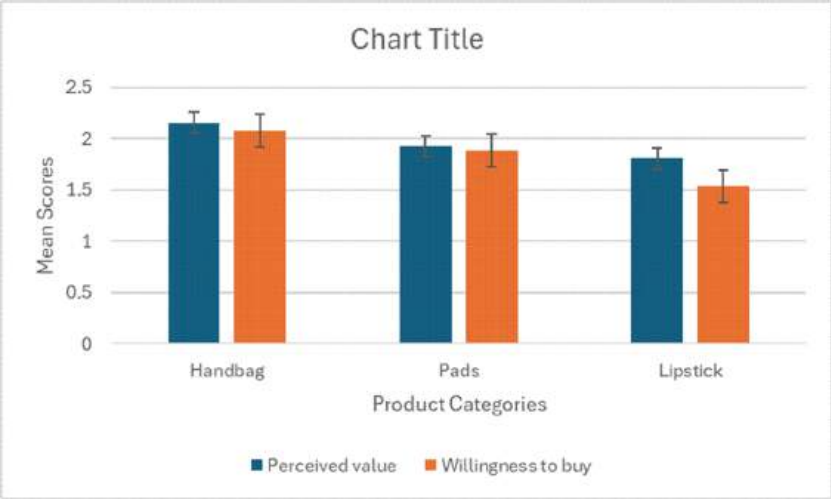


Figure 1: Graph of product categories against mean scores to show the perceived value and willingness to buy.

Figure 1 visually plots each product category against mean perceived value and willingness to buy scores, making the cross-category pattern immediately apparent. The results show a clear positive association between perceived value and willingness to buy across all three products, supporting the hypothesis. However, the strength of association varies — higher perceived value increases willingness to buy, but the effect is strongest for essential or functional products (handbags and sanitary napkins) and weaker for non-essential items like lipstick.

3.1 One-Way ANOVA

Source	DF	Sum of Square	Mean Square	F Statistic	P-value
Groups (between groups)	2	280833.3333	140416.6667	8.7496	0.0003298
Error (within groups)	93	1492500.002	16048.3871		
Total	95	1773333.336	18666.6667		

Figure 2: One-Way ANOVA analysis of the Perceived Value and Willingness to Buy.

To strengthen the analysis, two statistical tests were applied to the data. A one-way ANOVA (Analysis of Variance) was conducted to determine whether the average willingness to pay differed significantly across the three product categories — the handbag, sanitary pad, and lipstick. One-way ANOVA is used when comparing the means of three or more independent groups based on a single independent variable, making it appropriate for this study's cross-category design.[3] The results of the ANOVA indicated that average willingness to pay was not equal across categories, suggesting that product type influences the price consumers are prepared to spend.

3.2 Correlation analysis of all 3 products (X = Perceived Value and Y = Willingness to buy)

a. HANDBAG

Figure 3: Pearson correlation analysis of Perceived Value and Willingness to Buy of Handbag

b. SANITARY PAD

Figure 4: Pearson correlation analysis of Perceived Value and Willingness to Buy of Sanitary Pad

c. LIPSTICK

Figure 5: Pearson correlation analysis of Perceived Value and Willingness to Buy of Lipstick

In addition, a Pearson correlation analysis was conducted for each product separately, with perceived value (X) as the independent variable and willingness to buy (Y) as the dependent variable. Correlation analysis measures the strength and direction of the linear relationship between two variables, with values closer to +1 indicating a strong positive association. The correlation results confirmed a positive relationship between perceived value and willingness to buy for all three products, though the strength of this association varied — it was strongest for the handbag and sanitary pad, and comparatively weaker for the lipstick — consistent with the broader pattern observed in the data.

The results of this study support the hypothesis that perceived value is positively associated with willingness to buy across all three product categories. Consistent with Zeithaml's (1988) foundational framework, which defines perceived value as a subjective judgement of benefits relative to costs, consumers in this study showed greater purchase intention when they rated a product as offering good value for money. The Pearson correlation results confirmed this positive relationship across all three products, and the one-way ANOVA indicated that average willingness to pay was not equal across categories, suggesting that product type meaningfully influences price-based consumer decisions. These findings directly prove the hypothesis: perceived value does predict willingness to buy, and this relationship holds consistently regardless of product category.

However, the strength of this relationship varied by product type, revealing an important nuance that extends beyond the hypothesis. The sanitary pad, as an essential product, showed stable and consistent buying behaviour driven primarily by functional necessity. The handbag demonstrated the strongest alignment between perceived value and willingness to buy, likely because it satisfies both practical and aspirational needs, engaging a richer set of value dimensions including social and emotional appeal — consistent with the four-dimensional PERVAL framework developed by Sweeney and Soutar (2001), which identifies emotional, social, quality/performance, and price/value as the key drivers of consumer value perception. The lipstick, as a discretionary product, showed the weakest and most variable relationship, as purchase intention for non-essential items is more susceptible to personal preference and subjective utility rather than a straightforward value calculation. These findings suggest that while perceived value consistently predicts purchase intention, its predictive power is moderated by where a product sits on the essential–discretionary spectrum (Sari et.al., 2026).

These results align closely with existing literature on the perceived value–purchase intention relationship. Sharma and Ghosh (2023) similarly found that consumer value perception had a significant and positive effect on purchase intention, particularly when the product fulfilled a clear functional role. Research by Zhou (2022) on community fresh O2O platforms confirmed that perceived value significantly influences purchase intention and does so through the mediating role of consumption attitude. Further, studies on the influence of perceived usefulness on purchase intention by Sengupta and Pandit (2025) underscore that functional value dimensions are especially powerful drivers of purchase intention for products where utility is salient. Taken together, the existing literature and the present findings converge on the same conclusion: perceived value is a robust, cross-contextual predictor of purchase intention, but the mechanism and strength through which it operates depend heavily on the nature of the product itself.

Conclusion

This study set out to investigate whether perceived value predicts willingness to buy among female adolescent consumers across products of differing functional necessity. The hypothesis — that perceived value is directly proportionate to willingness to buy — is supported

by the results. Across all three product categories examined (sanitary pads, handbags, and lipstick), a clear positive association between perceived value scores and willingness to buy scores was observed, and this was statistically confirmed through Pearson correlation analysis, which returned positive correlation values for all three products. Furthermore, the one-way ANOVA indicated that average willingness to pay was not equal across product categories, confirming that product type meaningfully influences price-based consumer decisions. Together, these results confirm that when consumers perceive a product as offering good value for money, their intention to purchase it increases correspondingly (Kopp, 2026).

However, the study's most significant finding lies beyond the binary confirmation of the hypothesis. The results demonstrate that the strength of the 'perceived value–willingness to buy relationships' is modulated by product type. Sanitary pads, as an essential product, produced stable and consistent buying behaviour regardless of perceived value variation — indicating that necessity functions as a baseline driver of purchase that operates independently of value perception. The handbag produced the strongest and most coherent alignment between perceived value and willingness to buy, with the correlation being strongest among all three products, suggesting that when a product satisfies both functional and aspirational needs, high perceived value translates most powerfully into purchase intention. The lipstick showed the weakest correlation between the two variables, with willingness to buy being more susceptible to individual preference and subjective utility judgments that perceived value alone could not fully capture.

In direct answer to the research purpose, perceived value is a reliable and statistically consistent predictor of willingness to buy. Still, its predictive power is not equal across all product types. It is strongest for products that fulfil a recognised functional or social need, and weakest for products where the purchase decision is primarily discretionary and preference-dependent. These findings enrich the understanding of female adolescent consumer behaviour in the Indian market context and underscore the importance for businesses of tailoring their value proposition and communication strategy to the specific category and need-state of their target product.

Limitations and Future Work

Limitations include the small, non-random sample of 26 participants drawn largely from one age subgroup (16–17), limiting generalisability across the full 12–18 range and to non-Indian or male consumer contexts. The use of static, brand-free product images may not fully replicate real purchasing conditions, and self-reported Likert responses are susceptible to social desirability bias, particularly for a sensitive product like sanitary pads. Future research could expand the sample size and demographic diversity, incorporate real purchase or behavioural data rather than stated intentions, test additional mediating variables such as emotional and social value dimensions, and examine whether these patterns hold across other product categories or male consumer groups.

Works Cited

Dodds, W. B., Monroe, K. B., & Grewal, D. (1991). Effects of price, brand, and store information on buyers' product evaluations. Journal of Marketing Research, 28(3), 307–319. https://doi.org/10.1177/002224379102800305

Kopp, Carol M. "Understanding Perceived Value in Marketing: Importance and Impact." Investopedia, 11 Feb. 2026, www.investopedia.com/terms/p/perceived-value.asp.

Qualtrics. "What Is ANOVA (Analysis of Variance) Testing?" Qualtrics, 21 Sept. 2023, www.qualtrics.com/articles/strategy-research/anova/.

Sari, Nanang Suryadi, Yuyus Suryana, Rita Komaladewi, Diana. “Consumer, Customer and Perceived Value: Past and Present.” www.abacademies.org, Aug. 2018, www.abacademies.org/articles/consumer-customer-and-perceived-value-past-and-present-7393.html. Accessed 12 April 2026

Sengupta, A., and A. Pandit. "The Influence of Perceived Usefulness on Purchase Intentions: Insights from Consumer Behaviour Research." Academy of Marketing Studies Journal, vol. 29, no. 1, 2025, pp. 1–20.

Sharma, Ankur, and Sanjay Ghosh. "Effect of Consumer Value Perception on Consumer Purchase Intention." Journal of Indian Business Research, vol. 15, no. 2, 2023. doi:10.1177/09721509231166256.

Sweeney, Jillian C., and Geoffrey N. Soutar. "Consumer Perceived Value: The Development of a Multiple Item Scale." Journal of Retailing, vol. 77, no. 2, 2001, pp. 203–220. doi:10.1016/S0022-4359(01)00041-0.

Zeithaml, Valarie A. "Consumer Perceptions of Price, Quality and Value: A Means-End Model and Synthesis of Evidence." Journal of Marketing, vol. 52, no. 3, 1988, pp. 2–22. doi:10.1177/002224298805200302.

Zhou, Lixing. "Research on Community Fresh O2O Purchase Intention Based on Customer Perceived Value." Proceedings in Business, Economic and Social Studies, 2022, ojs.bbwpublisher.com/index.php/PBES/article/download/1538/1393.

Subjugation, Assimilation, Advocacy: A Recontextualized Framework for the Unfinished Struggle of Ainu Self-Determination By Kaitlyn Yu

Abstract

The Ainu are an indigenous East Asian population originating in modern-day Hokkaido, Japan, with a rich culture shaped by the remnants of colonialism and post-colonial reform. In the 15th century, the Japanese government gradually assumed administrative control of Ainu land and enacted policies that forced them to assimilate into an ethnic Japanese culture. These policies became outdated in the late 20th century and were subsequently reevaluated as the Ainu community advocated for legal reforms. In 2008, the Japanese Diet finally acknowledged the Ainu People's claim to indigenous status. Today, facilities across Japan and Hokkaido promote Ainu customs and preserve traditions through museums and archives, enhancing the quality and scale of Ainu cultural preservation efforts.

While individual scholars have previously noticed the Ainu's three-phase experience of subjugation, assimilation, and advocacy, no prior work has synthesized these observations into a cohesive analytical framework. As such, this paper formalizes the existing framework of Ainu history and links it to sociological theories of race, nationalism, and ethnicity. Ultimately, this research reveals that the legacy of Ainu assimilation continues to constrain the pathway to self-determination, interrogating whether legal recognition alone constitutes sufficient reparation. More broadly, this framework provides scholars and advocates with a structured lens to evaluate whether nations have moved beyond symbolic preservation toward substantive political autonomy and historical justice. Understanding these dynamics provides clarity into the Ainu people's history as they rebuild their identities in the face of historical suppression and modern challenges.

Introduction

The Ainu are Japan's first officially recognized indigenous population. Located primarily in Hokkaido and Russia's Sakhalin and Kuril Islands, they are a unique community that spans a wide territorial range. A 2023 survey estimated an Ainu population of around 16,786, although actual numbers may be significantly larger due to the suppression of Ainu culture when assimilation policies were first enacted (Ainu Association of Hokkaido). Compared to the Okinawans (or Ryukyuans) of the southern islands of Japan, who have a population estimated at 1.4 million, the Ainu are a significantly smaller community. A considerable number of Russian Ainu inhabitants were forcefully migrated to Japan in the 1940s after World War II; however, this paper primarily focuses on the history of the Ainu people in Hokkaido (McGrogan 356).

Throughout Japan's recontextualized history, the Ainu have been consistently discriminated against — yet their story is one of perseverance and defiance. The indigenous community has evolved significantly while resisting state pressures to conform, and they have reconstructed long-standing socioeconomic regulations while resisting assimilation policies attempting to erase Ainu identities. However, despite Japan's formal recognition of the Ainu as

an official indigenous population, recent policies only serve the singular purpose of conserving Ainu culture and neglect to protect their territorial, political, and self-determination rights. Still, the Ainu are no strangers to this pathway of acceptance. The indigenous community has experienced notable gains and losses over the years, becoming more visible on an international stage while combating long-held perceptions of homogeneity within Japan.

2024 marked the five-year review of the Ainu Cultural Promotion Act, which was implemented in 2019. This law was heralded as a significant step in cultural revitalization, yet it was criticized for its narrow-minded view that downplayed the Ainu's relevance in Japan's political and economic affairs. Little is mentioned about Ainu involvement in their economic systems and policies, and there are still far fewer concrete regulations that allow them the freedom to decide their education, territorial claims, and civil rights. In this way, the sparse five-year reviews failed to instigate substantial policy gains (International Work Group for Indigenous Affairs). However, Ainu activism in the 21st century has been marked by both successes and setbacks. Later that year, on April 30, the University of Edinburgh returned Ainu ancestral remains to community representatives after more than 100 years of separation (The University of Edinburgh News).

The state of Ainu and indigenous recognition is ever-changing. Amidst the increase in Hokkaido's popularity as a tourist destination, it is essential to understand the history of the indigenous Ainu population and acknowledge their heritage. Yet, the Ainu legacy extends far beyond their customs and traditions. After centuries of being suppressed, the community deserves the autonomy to reconstruct a version of itself that is not dictated by the state. This paper expands on existing literature on Ainu history by formalizing three distinct categories of *Subjugation*, *Assimilation*, and *Advocacy*. Backed by theories of dependency, racial domination, and imagined communities, this paper also argues how state-imposed policies of assimilation and social perceptions have shaped their identity throughout history.

Early Modern Ainu Subjugation Policies (1603-1867)

The changing customs of Ainu and ethnic Japanese interactions before the Meiji Era provide context for assimilation policies that evolved in subsequent centuries. Japan's colonial dominance in Ainu territory began far earlier than its expansion efforts in the 20th century. Beginning in the Tokugawa Era from 1603-1867, the *wajin* (ethnically Japanese) Matsumae feudal family was appointed to rule over the northern Hokkaido region, where approximately 20,000 to 30,000 Ainu resided (Howell, *"Ainu Ethnicity"* 70). The advent of the Matsumae regime in the Tokugawa era marked the beginning of centuries-long Ainu subjugation. At the time, Ainu inhabitants were a significant segment of the Hokkaido population, not yet affected by centuries of intermarriage and assimilation. The Matsumae thus focused their efforts on developing an economic system that established control over the Ainu population. Three factors — failed rebellions, the advent of *wajin* discrimination against the Ainu, and enforced economic dependency — particularly influenced the decline in Ainu autonomy. These developments gradually facilitated the cessation of indigenous rights, traditions, and land to the Japanese.

The pre-Meiji history of *wajin* and Ainu interactions refocuses the history of Japanese colonialism from a lens primarily focused on imperial expansion in the 20th century, to a 17th-century origin of Ainu control under the Matsumae feudal regime. This recontextualization reveals three concrete mechanisms (failed rebellions, systematic discrimination, and enforced economic economic dependency) that characterize the Subjugation phase of Ainu history before assimilation policies formally emerged. With pre-modern foundations laid, the following subsections examine two critical dimensions of Ainu-*wajin* relations in the Tokugawa era in greater depth.

First Interactions between the Wajin and the Ainu in the Tokugawa Era

Though the indigenous population attempted to resist *wajin* influences, their efforts were mostly futile (Lambright 10). This caused the ethnic Japanese government to continue regulating Ainu customs harshly. An attempted rebellion by Shakushain, an influential Ainu leader, ultimately failed to free the indigenous population from ethnic Japanese control from 1669 to 1672. Instead, Shakushain's War led to tightened economic and political restrictions against the indigenous population: the Ainu were forced to provide "payments of indemnity" to the Matsumae, while other small-scale rebellions against these payments were quelled (Powell 4). During this period, the Ainu community was scattered across regions to the north of Japan, including modern-day Hokkaido and Russian territory. Due to this far-reaching geographical spread, it was difficult for the Ainu to foster a sense of cohesion or unity among the dispersed communities, making rebellions like Shakushain's War less effective in resisting Japanese rule.

Aside from rebellion, other cultural factors also affected *wajin* perceptions of the Ainu. Specifically, Japanese-produced *Ainu-e* genre paintings began to gain popularity. These artworks bordered "the line between encyclopedic documentation and exaggerated exoticization" as they depicted Ainu customs and rituals with varying levels of accuracy (University of Oregon). Prone to illustrating clan leaders, ceremonies, and daily life, *Ainu-e* paintings illustrated real aspects of Ainu culture. However, *wajin* artists often leaned into stereotypes, and their drawings emphasized the hairiness or unconventional attire of the Ainu, attracting unnecessary attention to appearances distinguishing them from the ethnic Japanese (Lambright 5).

By isolating 'uncultured' characteristics of the Ainu, the *Ainu-e* paintings thus served as tools for reinforcing their perceptions as 'savages' in the eyes of Japanese travelers and observers (Lambright 17). During the Meiji Era, these assumptions about the Ainu came into full force, along with the denial of their indigenous land rights in the newly-named Hokkaido region. Despite their historical ties to the northern part of Japan, the Ainu were exploited, and official ordinances disregarded their claims to territorial ownership. Collectively, the dynamics of military defeat, geographical fragmentation, and visual stereotyping constituted key mechanisms of the Subjugation phase in Ainu history. These processes worked in tandem to cultivate a larger, structured state of Ainu dependency on the Matsumae regime, laying the foundation for subsequent decades of forced assimilation.

Cultivating a state of Dependency

As they approached the midpoint of the 17th century, the Tokugawa shogunate[90] had created a monopoly over Japan's political and social institutions (Howell, *"Ainu Ethnicity"* 72). The Matsumae and the shogunate also established a homogeneous identity within the state, defined state boundaries, and separated society into a class hierarchy. These actions laid the groundwork for the solidification of a Japanese national identity and national consciousness, which did not encompass the Ainu: the indigenous population was pushed aside and deemed inferior to those who pledged full allegiance to the emperor.

This established a definitive ethnic boundary between the Ainu and their *wajin* counterparts — a distinction between ethnic groups characterized by the inclusion and exclusion of members and specific social behaviors (Barth 15). These ethnic boundaries were produced and continually reinforced through various systems. Economically, the Matsumae established trading posts, or *akinaiba*, and controlled Tōhoku ports, just south of modern-day Hokkaido (Siddle 33). As such, the Ainu were continually "subject to control and taxation," prohibited from trading freely amongst their own communities (Siddle 33). Through labor practices, forced resettlement "became a feature of Ainu life," as Ainu laborers were "brought down from the mountains" to construct roads and facilitate trading ports (Siddle 46). Ideologically, along with crude depictions of the Ainu in *Ainu-e* paintings, the *wajin* referred to themselves as 'humans,' or *ningen*, while seeing the Ainu as *Ezo* or *ijin*, thereby denying them full humanity by assigning them "animalistic qualities or origins" (Siddle 43). The Ainu of northern Hokkaido thus became united by this mutual exclusion, solidifying their own basis of an ethnic identity. Counterintuitively, their exclusion gave them a semblance of ethnic consciousness.

This sense of nationalism was almost immediately deconstructed, and it began with the dependency system that developed between the Ainu and the Matsumae. The Ainu relied on the Japanese for commodities, and the economy slowly lost its ability to remain independent without the influx of essential goods from the south (Howell, *"Ainu Ethnicity"* 74). Meanwhile, the Matsumae relied on the Ainu's continued economic submission to stay in power over the region. *Wajin* commercial fishers exploited the Ainu's need for iron and other commodities in order to tie them to a wage-labor regime in the 18th and 19th centuries (Howell, *"Making 'Useful"* 6). This reliance on a wage-labor system gradually deconstructed the Ainu economic structure, whose livelihoods had long relied on fishing, hunting, and gathering-based occupations.

The concept of dependency — where one economy is shaped by and subordinate to another — demonstrates how the Ainu were compelled to rely on Japanese systems to sustain their livelihoods and cultures (White xvii). Yet, dependency is influenced not only by political and economic factors but also by culture, history, and society (White xvii). For the Ainu, *wajin* economic domination sowed its roots in the gradual commercialization of their traditional rituals. Due to the power imbalance between the Ainu and the ethnic Japanese, the *uimam* and *umsa* rituals gradually lost their significance as sacred ceremonies within the community. The *uimam*

[90] The Tokugawa shogunate was the last feudal military government of Japan. Their regime was also known as the Edo period, which lasted 265 years and was characterized by an era of relative stability for the majority of Japanese people in the region.

was primarily used to conduct trade between Ainu equals, but Japanese manipulations and interference with this ritual altered the tradition to make Ainu individuals present their respects to the shogunate by performing it for officials (Howell, *"Ainu Ethnicity"* 81). Similarly, the *umsa*, used initially to exchange greetings after long separation, became a mere exchange of greetings between officials (Howell, *"Ainu Ethnicity"* 82).These rituals became less associated with traditional Ainu culture; instead, they persisted to reassure the Japanese of their authority over the Ainu.

As time progressed, Japanese subjugation continued in other forms. The Ainu depended on the Tokugawa for economic support and resources, while the shogunate relied on Ainu complicity to increase their political power. The Matsumae ruling clan implemented "some of the first systematic uses of racialization and othering against the Ainu," and enacted laws that restricted Ainu freedom (Lambright 5). These schemes were authorized until the advent of the Meiji government, when new assimilation policies reinforced the concept of the Ainu as a distinct, low-caste ethnic group separate from Japanese society.

As the Matsumae continued to exclude and segregate the Ainu from both social and economic standpoints, they established the otherness of the indigenous population in the north. The *wajin* soon identified the Ainu as a distinct group from ethnic Japanese imperial society, and paved the way for more subjugation and assimilation policies to rise in the Meiji era.

The shift between Tokugawa era subjugation and formal assimilation policies of the Meiji era lies not in the differing methods of control, but in the codification of *wajin* power. Under the Matsumae, Ainu subordination was localized and administered primarily through cultural othering, ethnic boundaries, and economic dependency. However, no official policies declared Ainu as imperial subjects, nor were any cultural practices entirely erased. The Meiji, by contrast, ushered in a new era that solidified loose legal control into a formal hierarchy. Where the Matsumae excluded the Ainu as 'barbarians,' the Meiji allotted a legal term — "Former Aborigines" — to describe the Ainu within state policy, denying their indigenous heritage and forcing them to comply with *wajin* customs. The Subjugation phase of Ainu history effectively highlighted the process of informal dependency cultivated under the Matsumae, until formal legal codification occurred in a subsequent age of assimilation.

Meiji Era Assimilation to Pre-War Ideologies (1868-1945)

The mid-19th century facilitated a decline in Ainu freedoms and rights, culminating in assimilation policies that persisted until the 20th century. By 1868, the traditional Ainu economy had to depend on Japanese trade and wage labor to sustain their livelihoods (Howell, *"Making 'Useful"* 6). With this established system of economic dependency and the gradual commodification of traditional rituals, the Ainu were forced to interact with the *wajin* government before assimilation policies were formally enacted in the late 19th century. As such, there were fewer efforts from the Ainu to resist political subjugation because ethnic Japanese systems had already been integrated into the Ainu community. The *wajin* were thus able to cement their hold over the Ainu and impose formal policies on the indigenous population

without significant uproar.

The Protection Act

The Hokkaido Former Aborigines Protection Act, also known as the Hokkaidō Kyūdojin Hogohō (北海道旧土人保護法), was the defining assimilation policy in the 19th century that formalized *wajin* political control over the Ainu indigenous population. This act, first implemented in 1899, would remain effective until the Ainu Cultural Promotion Act of 1997. The Protection Act legalized policies that diminished the value of the Ainu's ethnic identity, cultural practices, and political significance in Japan over the course of 98 years.

Under the Protection Act, many elements of Ainu culture found themselves subjected to the laws of the *wajin* government. In the mid-1880s, to exacerbate Ainu dependency on the *wajin* economy, officials promoted an agricultural lifestyle among the Ainu, declaring only farmers to be members of the 'real' Japanese community. This practice was revived under the Protection Act, leading to a decline in the importance of traditional Ainu occupations, such as fishing and lumberjacking (Howell, *"Making 'Useful"* 7). To further promote an agricultural Ainu economy, the Act also stated that members of the Ainu population were eligible to receive "agricultural equipment and seeds," along with grants of 15,000 *tsubo*, or 3954 square yards of land, under the condition that it had to be cultivated within 15 years (qtd. in Siddle 194). However, Ainu families soon found that their land was barely half the amount needed to support a typical household, and others had received wasteland and were unable to cultivate it within the designated time, resulting in their land being relinquished (Howell, *"Making 'Useful"* 9). Article 2 of the Act restricted Ainu rights to their allotted land even further, prohibiting them from transferring land except by inheritance and barred them from establishing "rights of pledge, mortgage, lease, or perpetual lease" (qtd. in Siddle 194). This clause reveals how the underlying mechanisms of the Protection Act relegated the Ainu to conditional recipients: they were not treated as owners of land; thus, their rights were never guaranteed. In 1937, the Protection Act was revised, and the Japanese deemed the Ainu's enforced agricultural lifestyle as impractical and obsolete. While this revision did not relieve the Ainu of Japanese control, it still freed the indigenous population of their ties to unwanted and insufficient agricultural practices (Howell, *"Making 'Useful"* 15).

Secondly, the Japanese practice of registering citizens forced Ainu villagers to possess a permanent place of residence according to *wajin* standards. This requirement exemplifies the process which James C. Scott describes as "a state's attempt to make a society legible," simplifying and standardizing local customs to make populations easier to control (2). By compelling the Ainu to adopt fixed residences and adhere to *wajin* norms, the Mejii regime laid the groundwork for administrative assimilation, significantly reducing mobile, fishing-and-hunting Ainu practices. As such, while the registration schemes increased, the Ainu found it increasingly hard to sustain their culture. Their traditions of kinship, tributary ceremonies, housing, clothing, cosmetics, and tattooing practices became progressively obsolete. These practices were deemed illegible to the state and therefore targeted for elimination.

Furthermore, the Protection Act contributed to the decline of Ainu culture and

expression, particularly through its policy of replacing traditional Ainu housing with Japanese-style residences. Ainu dwellings were perceived as unclean by the Japanese government; therefore, this policy seemingly reduced the spread of tuberculosis and other infectious diseases and allowed the indigenous community to blend in with the *wajin* population (Howell, *"Making 'Useful"* 11). Housing reform enhanced administrative assimilation: the encouragement of Japanese-style residences and the dismissal of traditional Ainu houses enforced rigid categories of property, sanitation, and citizenship.

A large influx of non-Ainu immigration into Ainu territory also occurred as colonial policies increased, while statistics showed the Ainu population diminishing (Cornell 293). In efforts to integrate the *wajin* and Ainu, the Protection Act's citizen registration practice stopped separating the two ethnic populations in censuses, labeling them jointly as 'commoners' instead (Howell, *"Making 'Useful"* 7). Politically, the Japanese government adopted a mindset that combined the Ainu with the *wajin.* However, when the Ainu moved to urban areas outside of Hokkaido, they faced social subjugation and discrimination, guided by perceptions of the indigenous community as primitive and barbaric. This census practice completed a process of administrative erasure, making the Ainu population statistically invisible and removing the demographic evidence of their existence.

As a result, subsequent generations of Ainu people lost touch with their culture, identity, and language. Even as early as 1910, the Ainu had limited opportunities to converse in their language, with most assimilating to *wajin* norms and speaking Japanese more frequently. This is further evidenced by a 1917 linguistics census, where almost all Ainu individuals surveyed could speak Japanese and used it often (Kirikae 173). This decline in Ainu language fluency persisted into the modern era. As of 1993, no more than 0.8% could speak the Ainu language conversationally, and 91% only knew a few words, demonstrating how *wajin* influence on the population continued to erase Ainu identity (Howell, *"Making 'Useful"* 10).

Some scholars believed the policies enacted during the Tokugawa shogunate's regime and the Protection Act in 1899 were "guided by benevolence and humanity," as they increased a sense of protection, rather than discrimination, amongst the Ainu community (Cornell 302). The Protection Act did offer a form of 'security,' but it was designed to absorb the Ainu into the Japanese state, neglecting indigenous traditions. The Act's welfare provisions granted the Ainu plots that could be confiscated for non-cultivation, with no consideration for their quality or size. The Protection Act also prohibited the Ainu from selling, mortgaging, or leasing their property like ordinary Japanese citizens. These mechanisms positioned the Ainu as passive recipients of state charity rather than subjects that bore rights to their allotted land, thereby acting as tools of administrative assimilation operating under a benevolent facade. As such, what Cornell reads as benevolence, this paper recontextualizes as a systematic erasure enacted through state legibility.

Furthermore, eugenics drives and physical anthropological studies occurring in subsequent decades did not promote a protective viewpoint of Ainu citizens. Both the Protection Act and the overt racism of eugenics continued to deny the Ainu full political agency. Therefore, the indigenous population continued to suffer in the 20th century because of growing social

prejudice and theories that were informed and promoted by pseudo-race science. One such idea was the existence of a heterogeneous yet distinctly ethnic Japanese citizen (Howell, *"Making 'Useful"* 5).

Race and Pseudo-Science

As Japan began to receive Western ideas of race and nation at the turn of the 20th century, a discourse emerged among *wajin* scholars on the racial hierarchy of the state to enforce imperial power. The Japanese empire was rapidly expanding during this period, and scholars raced to link the idea of a uniquely Japanese sovereign to this amalgamation of identities by connecting the "diversity of the Empire" to the developing theory of a 'pure' Yamato race existing within the nation (Howell, *"Making 'Useful"* 5). Using two concepts of biologically integrated groups of people, *Jinshu* (the human race, or physical attributes), and *Minzoku* (nationality and ethnicity), scholars within the nation sought to find the origins of the Japanese race (Low 57). By claiming that the constructed Yamato race descended from a mixture of ethnicities, it justified the empire's colonial expansion practices by arguing that colonial subjects had always been Japanese (Howell, *"Making 'Useful"* 5).

Benedict Anderson's theory of imagined communities provides a crucial framework for understanding how states utilize nationalism to promote homogeneity within their regions. By emphasizing shared cultural symbols, language, and tradition, individuals within the nation share a collective emotional experience, enhancing their allegiance to the state (Anderson 145). This sense of patriotism, or "unisonance," was used to glorify themselves rather than declare common enemies within the empire (Anderson 145). The imposed nationalism thus underscored how imperial ideologies of the Yamato race reframed colonial expansion as a unifying project. The empire emphasized a shared cultural and linguistic heritage within its territories, even when racial distinctions — as seen in the treatment of the Ainu — remained stagnant and ever-present.

This assertion of a homogeneous empire effectively put the Yamato, the united, ideal Japanese race, at the top of the racial classification pyramid. Meanwhile, the bottom of the pyramid belonged to the Ainu, a "dying population" deemed so weak that they seemed "unfit in the 'struggle for survival'" (Dikötter 137). This supposed weakness led scholars to identify the Ainu as primitive and incapable of fitting into the concept of modernity. They treated the Ainu like specimens of a case study, developing condescending descriptions of the Ainu backed by Tokugawa accounts of their "inhuman origin and habits" (Dikötter 139). Western scholars largely shared the same views, with some favoring or impressed by the Ainu and others describing them as illiterate, hairy, or doomed to extinction (Dikötter 138). Expositions all over Japan and the world (e.g., in Osaka and Louisiana) in the 1900s aimed to highlight these distinctions. Using their theory of the Ainu as racially inferior to the Japanese, these expositions were specifically made to justify the colonization of Hokkaido, attract public attention, and showcase cliched depictions of the indigenous population.

This method of Ainu subjugation through pseudo-scientific research grew during the Taisho period (1912-26). Western narratives continued to justify the colonization of Hokkaido as

Japan's responsibility to control the region as the superior race (Dikötter 142). Meanwhile, Japanese scholars began to conduct ethnographic 'research' in Ainu villages, or *kotan,* without the consent of indigenous community members. Studies completed in the 1920s with eugenics and serology identified the Ainu as a submissive blood type and categorized Ainu children as inferior to Japanese because of inherent biological differences (Dikötter 143). Ainu children, deeply affected by this targeted racialization, were treated harshly in academic settings and suffered from intense racism. Furthermore, the ethnic group was portrayed stereotypically in textbook illustrations, influencing the perceptions of future generations. In this way, the *wajin* developed a straightforward generalization of Ainu customs and traditions, further obscuring Ainu identities (Dikötter 154). Their cultural differences were racialized because of pre-existing stereotypes backed by pseudo-science and assimilation policies.

The Japanese government also established a paradox between political and societal attitudes toward the Ainu. The state's push for a unified empire with all subjects under the Yamato race contrasted heavily with the increasing ostracization of the Ainu due to their alleged biological underdevelopment. This scenario demonstrates a failure to classify between what Loïc Wacquant dubs as "folk and analytical" understandings of race (223). Folk understandings, Wacquant argues, are everyday beliefs that treat race as a natural and essential aspect of life. Individuals might invoke this perspective to discriminate on others based on appearance or ancestry. By contrast, 'analytical' understandings promote the perspective of race as a social construct, interrogating how and why societies produce and maintain racial hierarchies, and whether these classifications correspond with factual, biological evidence. Wacquant's key insight is that scholars must treat both beliefs as societal perspectives that have lasting consequences, whether or not they promote an accurate understanding of race.

In the context of the Ainu, Wacquant's folk understandings encompass the societal push for separation between the *wajin* and Ainu based on physical appearance, ancestry, and stereotypes. These folk beliefs manifested in everyday discrimination, social exclusion, and informal separation of *wajin* and Ainu customs dating back to Tokugawa-era practices. Meanwhile, the Japanese empire attempted to establish an analytical depiction of their differences, denying racial differences and insisting that all imperial subjects belonged to a single Yamato race instead. Through the classification of the Ainu as 'former' aborigines, the government effectively dismissed the idea of the indigenous community as separate from the Yamato race of the Japanese empire. Thus, the emphasis on the 'former' identity of the Ainu was an analytical attempt to maintain *wajin* authority and enact assimilation policies: it acknowledged a historical difference, but denied any ongoing biological distinctions between the Ainu and *wajin*. The assimilation policies through the Protection Act further demonstrated an attempt to integrate both communities and minimize the differences between the Ainu and *wajin*, shaping the Ainu into "good imperial subjects" as opposed to barbaric outcasts (Howell, *"Making 'Useful"* 24). Since the state actively denied racial differences, it could not address the folk discrimination that continued to marginalize the Ainu based on culture and appearance. Thus, the state claimed to reject racial categories while actively enforcing policies justified by it.

While social discrimination and political assimilation efforts increased over the course of the 20th century, Ainu advocacy turned a new leaf after the end of World War II. Equipped with a resistance mindset of non-violence against the Protection Act, Ainu activists began to combat the stereotype of the primitive savage, and they slowly overcame centuries of prejudice in the early stages of Japanese nationalism.

Ainu Advocacy through Non-Violent Resistance (1946-2025)

The postwar period of the 20th century promoted an increase in Ainu advocacy. As Japan rebuilt itself after the events of World War II, the nation began to experience rapid economic growth, technological expansion, and a renewed sense of national pride (Low 62). Yet, after the renunciation of the emperor's divine status, theories surrounding the homogeneity of Japan and the existence of the Yamato race began to diminish. Scholars within and outside Japan, including Ainu activists, compared the Ainu with the Japanese and identified cultural differences with a shifted perspective. This time, activists aimed to highlight the nation's shift away from the *wajin* government's homogeneous Yamato race ideology and the labeling of the Ainu community as 'inferior.' These efforts pressured the government to acknowledge the Ainu's unique indigenous identity. Various cases of nonviolent activism in the late 20th century also paved the way for the recognition of the Ainu as a distinct indigenous community, enabling their growing cultural recognition and visibility on the international stage.

Nonviolent Resistance

Perhaps no Ainu advocacy movement was as effective as the push for "resistance through noncompliance" beginning in the 1960s (Powell 6). Ainu activism had been dormant and suppressed since the advent of Japanese contact with the indigenous community; however, international indigenous movements inspired a sense of cohesion among the population. The Ainu became united under the purposes of asserting their identity, retaining traditions, and preserving their culture, while activists began to question Japanese policies of assimilation. This led to events that grew in scale to challenge long-held stigma in society and politics, allowing the Ainu to regain their autonomy that had been forcefully removed for centuries.

The Ainu advocacy agenda prioritized nonviolence, employing peaceful and diplomatic strategies to overturn assimilation laws. One such example of this nonviolence pertained to cultural resistance, as the Ainu asserted their place in Japanese history and reminded the nation of their existence. In 1968, centennial celebrations of the Meiji era took place in Hokkaido, commemorating 100 years of Japan's rule over the northernmost island of the nation. Officials never considered the Ainu perspective during the festivities. In fact, the centennial failed to recognize any Ainu contribution to the country or their subjugation by the *wajin* over the course of the century. In response to this outright exclusion, the Ainu constructed a 4.2-meter-tall, reinforced plastic statue of Shakushain, the rebellion leader of 1669 (Powell 9). The statue, built in 1970, served as a prominent reminder of Ainu involvement with Japan for over three centuries, symbolizing the community's dissent and nonconformity, while preserving the legacy

of Ainu resistance against *wajin* subjugators. It reclaimed public space for Ainu memory, directly reversing the logic of administrative assimilation. Furthermore, the statue exemplified the Ainu's strategy of nonviolent advocacy, simply existing as a cultural symbol of resistance and a blatant acknowledgement of Ainu presence. In fact, memorials to Shakushain at the statue are held annually to this day, usually occurring in late September every year, continuing the tradition of Ainu activism.

Intertwining with the Ainu's peaceful advocacy, however, were violent acts of protest as a product of the Ainu Revolution Theory, a concept developed by New Left ethnic Japanese activists. These attacks, they claimed, were committed on behalf of the Ainu and aimed to draw attention to the unjust policies suppressing Ainu freedom, even though no Ainu activists were involved in making these decisions, nor did they advocate for violent forms of indigenous rights support (Powell 7). A series of 30 bombings began in 1972, targeting government exhibitions affiliated with the Ainu in protest of "the commodification ... of Ainu traditions and culture" for tourism (Powell 7).

For a similar cause, an attempted assassination of Asari Giichi, the Japanese Mayor of Shiraoi, by Yagi Tatsumi occurred. The *wajin* perpetrator cited his motive as one in dissent of the nation's corporations exploiting Ainu traditions for the tourist industry. Another notable attack was the bombing of the Hokkaido Prefectural Government office building in Sapporo on March 2, 1976, allegedly committed by the East Asia Anti-Japan Armed Front, which resulted in 82 casualties, with two deceased (Yasufumi and Arnold 106). Though the ethnically Japanese activists also aimed to support Ainu liberation, the demonstrations' scale of violence sparked discontent within the indigenous community and the rest of Japanese society (Powell 7). Furthermore, when investigations were conducted, officials appeared to cast the blame on the Ainu themselves for staging the violent uprisings, exacerbating the age-old stereotype of the Ainu as primitive, barbaric, and violent.

In 1975, Ainu activists such as Kaibazawa Hiroshi spoke up to combat the accusations and put an end to the increasingly violent movements by the New Left. They characterized these events as "sever[e] damage inflicted on the already low perception about the Ainu people" (Powell 7). Specifically, Kaibazawa believed that the acts committed by the ethnic Japanese were an unwanted interference that aggravated public perceptions against the Ainu (Powell 7). His position rejected the state's assimilationist policies and the New Left's politics, two forms of external control that used the Ainu as symbols for their respective agendas that did not concern the indigenous population. These acts of cultural resistance solidified that the fight for Ainu recognition and equality had to be nonviolent, peaceful, and an explicit demonstration of self-representation. Kaibazawa's statement thus solidified the notion that only the Ainu themselves could lead their movement toward recognition and quality. In response, the indigenous population formed a cohesive community and developed a renewed sense of unity that had not been as prominent during the era of assimilation in the 19th and early 20th centuries.

Language Revitalization

The 1980s saw a shift away from cultural activism and a rise in Ainu language revitalization. Violent protests had subsided, and the indigenous population identified a new priority that could strengthen the community and assert their ethnic identities. In 1983, the first Ainu-centric educational facility was established by Kayano Shigeru. Officially named the Nibutani Ainu Language School, this institution aimed to revitalize the Ainu language and educate the younger Ainu generations about their heritage. Another similar establishment for Ainu education was set up in 1987 in Asahikawa. When Nomura Giichi, the leader of the Hokkaido Utari Association that promoted Ainu indigenous rights, expressed his interest in more Ainu language classes, community centers in Urakawa and Kushiro responded in kind.

Over the next decade, more educational facilities were founded across Hokkaido, dedicated to revitalizing the Ainu language and culture. Radio programs such as a 1987 Ainu language program established by Kayano and speech contests were established as well, increasing accessibility for the Ainu to learn and engage with their language or reconnect with their indigenous identities (Powell 12-13). As seen, over the course of the late 20th century, educational facilities improved the cultural significance of the Ainu and helped individuals reconcile with their indigenous identities.

During the assimilation and postwar era, many Ainu individuals had migrated to cities south of Hokkaido in search of work. Amongst a backdrop of negative stereotypes and discrimination, the urban Ainu population thus learned to hide their backgrounds and blend in with the Japanese, resulting in a loss of culture and ethnic identity, while strengthening the effect of the Protection Act from 1899. Decades of intermarriage and policies of subjugation made physical differences between the Ainu and *wajin* difficult to distinguish, especially with the near extinction of cultural markings, such as face tattoos, that could easily identify an individual as an Ainu person. In contrast, the indigenous rights movements of the late 20th century sought to remedy decades of generational shame associated with Ainu identity, promoting an increase in Ainu cultural and linguistic knowledge within the broader Japanese population and deconstructing preconceived perceptions created through assimilation policies. As evidenced by the establishment of Ainu educational facilities, younger generations began to reconnect with their lost heritages, discover their roots, and ultimately feel pride in their ethnic identities.

State Action

As Ainu activism increased, advocates began to directly characterize the 1899 Protection Act as an attempt to extinguish a distinct Ainu culture within the nation (Larson et al. 58). At the same time, Ainu advocacy had been recognized by the worldwide indigenous peoples' movement. In 1981, with support from Inuit leaders, Ainu representatives attended the third World Council of Indigenous Peoples, moving them one step closer to official recognition (Larson et al. 60). The international solidarity placed external pressure on Japan to address its colonial legacy, forcing the state to confront a power imbalance between the Ainu and *wajin* it could no longer ignore. As such, seven years later, in 1988, and again in 1995, discussions of a "New Ainu Law" to replace the 1899 Protection Act were led by the governor of Hokkaido and

the Chief Cabinet Secretary, respectively (McGrogan 357). While these discussions appeared to be a significant shift in favor of recognizing the Ainu's distinct identity, they were still restricted by the state's reluctance to grant substantive political autonomy to the indigenous population.

Ultimately, the "New Ainu Law" was passed, but it had yet to promote Ainu self-determination and a concrete recognition of the population's indigeneity. The Ainu Culture Promotion and Dissemination of Information Concerning Ainu Traditions Act replaced the assimilation policies in the Protection Act. Passed on July 1, 1997, the Culture Promotion Act established concrete policies for national and regional governments to "promote and support Ainu cultural activities" (Larson et al. 64). The Act also established the Foundation for Research and Promotion of Ainu Culture, tasked with disseminating public knowledge and providing education about Ainu culture. The Foundation received funds from government subsidies and involved members of Ainu rights associations on its managing board.

Despite these efforts, the Culture Promotion Act did not outline a distinct recognition of the Ainu community as an indigenous population within Japan. Two distinct hindrances prevented the Ainu from achieving distinction: a lack of a solid definition of indigeneity and the state's fear of losing power in Ainu territory. Even before Japan voted for the United Nations' revised Draft Declaration on the Rights of Indigenous Peoples (UNDRIP) on June 29, 2006, they had explicitly stated their desire for an objective definition of indigenous people (Larson et al. 67). However, a collective subjectivist definition of indigenous people worldwide had become the norm, advocating for indigenous groups to identify themselves rather than adhering to a specific set of criteria (Larson et al. 57). The Ainu had established themselves with this definition in the international community and fortified their indigenous identities globally, just not within Japan.

Furthermore, the state's fear of losing power to the indigenous community resulted in the funneling of Ainu indigenous rights into matters of cultural preservation, ignoring or prolonging investigation into Ainu territorial, economic, or political rights (Larson et al. 56). Previous policy proposals, dating back to 1984, aimed to increase Ainu rights within Japan by allowing them the freedom to decide their economic, political, and educational involvements. However, these proposals were set aside and largely disregarded due to concerns that they would infringe on the state's power over the indigenous community. As such, the Culture Promotion Act was reduced to a set of "principles…concerning culture and language" instead (Larson et al. 63). Though the Culture Promotion Act increased Ainu indigenous rights and promoted the revitalization of their traditions, it decreased the urgency for new laws to improve their political and economic involvement in the Japanese government. This narrower focus on Ainu rights prevented the Japanese government from weakening its authority over the Ainu population. The government assumed they had effectively decreased the Ainu people's discontent with a safe, unprovoked law, allowing them to stay in power without causing discontent in the indigenous community (Larson et al. 64).

This paper does not aim to dismiss the Culture Promotion Act as meaningless; it set the groundwork for the shift from administrative assimilation to a concerted effort of Ainu cultural

preservation, and was a genuine legal victory achieved through decades of Ainu activism. Though the 1997 Act granted the Ainu cultural visibility, celebrating their heritage without confronting the colonial legacy of their erasure, it stopped short of allowing the Ainu true rights to self-determination. Rather than addressing land rights, political representation, or economic compensation, it defined Ainu identity exclusively by folklore and culture, viewing the population as a dying tradition to be preserved and not sustained. Thus, the Act created a paradox, increasing Ainu cultural rights while decreasing the urgency for new laws to improve their political and economic standing. Whether the law constitutes meaningful progress or a more subtle form of containment remains an unanswered question, one that the future indigenous recognition in 2008 would only partially answer.

Leading up to the decision, however, was a plethora of factors that spurred official recognition. Ainu rights to indigeneity had become an increasingly pressing issue, with global and internal influences pushing the Japanese government to recognize them as an indigenous group separate from the *wajin*. Doing so would require the outright dismantling of the ideal, monoethnic Japan sprouting out of shared identity and nationalism. Instead of simply requiring governments to respect Ainu culture, their traditions would have to be labeled as different from those of the ethnic Japanese. The Japanese attitude toward Ainu rights had already undergone a considerable shift over a relatively short period, but state acculturation, geopolitical concerns, and increasing Ainu activism allowed them to concede to the demands of the Ainu to be recognized distinctly from the *wajin.*

Factors Influencing State Decisions

Over the course of the next 11 years, from 1997 onwards, Japan experienced an acculturation of the state, where the national government slowly began submitting to international cognitive and social pressures that influenced the state to assimilate to globally acceptable standards (McGrogan 361). Thus, Japan was pressured by the international spotlight on indigenous rights and the growing global advocacy for indigenous communities. This is evidenced by Japan's vote for the revised UNDRIP on June 29, 2006, which improved global indigenous rights and advocated for the recognition of more indigenous communities worldwide. This decision was primarily influenced by international pressures on the Japanese government to conform to the growing norm, and it granted indigenous communities greater autonomy and recognized them legally. The state feared scrutiny on a global stage, so it also agreed to comply with the new international stance on indigeneity. Moreover, since the 1990s, there had been explicit pressure placed on the Japanese government by treaty bodies of the United Nations, such as the Human Rights Council (HRC), to confront the issue of Ainu indigenous identity within the nation (McGrogan 363). After the UNDRIP was passed in 2007, there was renewed pressure for the Japanese government to confront the recognition of the Ainu population's indigeneity. Faced with explicit worldwide duress and the threat of losing international favor, the state finally considered the decision to officially recognize the Ainu as an indigenous community.

Another factor contributing to the major shift in Japanese attitudes toward the Ainu was

the indigenous community's unique geopolitical context, which served as an additional influence beyond numerous international pressures. Despite efforts from the Japanese and Russian governments to eradicate or assimilate the indigenous population, the Ainu were inextricably tied to the land in Hokkaido, Sakhalin, and the Kuril Islands. With the latter two territories under Russian control, the Japanese state understood the advantage that official recognition could bring. In the 2010s, Japan did indeed bring forth the argument to claim the ownership of "two southernmost Kuril Islands, Iturup and Kunashir," by stating that their territorial rights were "legitimised based on the previous existence of indigenous Ainu culture" (Schoer 1). This strategic deployment of Ainu identity suggests that official recognition served both humanitarian and geopolitical goals — by legally recognizing the Ainu as a distinct ethnic group from the Japanese, the government could strengthen ties with the Ainu population. This official recognition could also lead to direct support from the Ainu to claim Japanese territory in the Kuril Islands and parts of Sakhalin. Thus, the Ainu could serve as a powerful, symbolic, and practical tool to leverage negotiations with Russia over the disputed northern territories. However, as scholar Bethany Schoer notes, "emphasizing Ainu culture does not at all represent a commitment" to upholding Ainu indigeneity within the state (4). Because the indigenous community became more relevant to Japan's geopolitical agenda, the state selectively elevated Ainu identity to back territorial claims. Yet, the Ainu themselves were still largely absent from the territorial negotiations that their heritage was used to justify, proving that geopolitical convenience does not directly translate into political autonomy.

Most importantly, the Ainu played a crucial role by emphasizing their indigeneity during their advocacy efforts, transforming themselves from passive subjects of state policy into political subjects demanding recognition. The late 20th century had seen the Ainu rise as a population that explicitly self-identified as indigenous, creating a prominent source of pressure for indigenous rights in Japan. For almost a century under the Protection Act of 1899, *wajin* assimilation policies had systematically discouraged Ainu cultural expression and actively introduced measures to conceal those with indigenous heritage. Without direct proclamations from the Ainu to assert their indigenous identity, the policy proposals from ethnic Japanese advocates would have little to no effect on the state's consideration to amend Ainu rights. *Wajin* advocates, such as the New Left activists discussed earlier, could highlight Ainu injustices, but they could not embody the lived experience of Ainu assimilation nor claim the moral authority of those who had suffered under the Protection Act. Because of Ainu insistence on self-representation in their advocacy movements, the state's official recognition was thus seen as a concession won through legal challenges, international alliance-building, and decades of national resistance towards assimilation policies.

As a result, all these factors contributed to the official recognition of the Ainu as an indigenous community on June 6, 2008. The Japanese Diet passed a resolution in light of the UNDRIP that explicitly stated the government's recognition of the Ainu as a historically distinct population from the *wajin*, that the Japanese had discriminated against them, and that the state would formally recognize the Ainu with their distinct language, culture, and tradition.

Nobuko Tsuda, an ethnic Ainu scholar specializing in Ainu cultural studies, has highlighted the impact of restoring Ainu heritage through her research. Growing up as a mixed Japanese-Ainu individual, Tsuda recounted internalizing societal prejudices about her ethnicity. However, she found that Ainu artifacts strengthened the community's unity and connected the Ainu people (Tsuda 158). Tsuda emphasized the importance of continuity and intentionality in Ainu craftsmanship and tradition, arguing that sustaining Ainu heritage requires a deep understanding of its spiritual foundations rather than adhering to superficial forms of cultural revival (159). Her work was one of the many examples that revealed cultural restoration as both a healing process and a strategy for resistance. Through her personal experiences and research on Ainu heritage, she has confronted the systemic erasure of Ainu traditions and advocates for a future where Ainu identity is celebrated and understood, not hidden or stigmatized.

Similarly, cultural restoration has been a significant aspect of the *wajin* government's efforts to recognize the Ainu. In May 2019, the state issued the Ainu Promotion Act, also known as the New Ainu Policy, that promoted education about Ainu language and traditions and aimed to increase cultural recognition of the Ainu people. With this new law, facilities have been established across Hokkaido, such as the Upopoy Ainu National Museum and Park, which cultivate Ainu customs and archive recordings of the language and dialects, further enhancing the quality and scale of Ainu cultural preservation efforts. However, the New Ainu Policy failed to recognize the Ainu people as the original inhabitants of Hokkaido even though they were officially recognized as indigenous in 2008, nor did it allow for the Ainu people themselves to establish their own political and economic laws.

Even with a major shift in Japanese attitudes toward Ainu indigeneity, along with new laws emphasizing cultural preservation to replace assimilation policies, the Ainu are still fighting for the right to self-determination. Japan had voted for the UNDRIP in 2006, thereby symbolizing its approval to adhere to the rules stated in the resolution, one of which affirmed the self-determination rights of indigenous peoples. In the UNDRIP, the United Nations outlined the freedom of indigenous communities worldwide to determine their own political status and to freely pursue socioeconomic development on their own terms. By passing laws such as the 2019 Promotion Act, which included a restrictive list of what the *wajin* government determined as Ainu Culture, the state has impeded the Ainu population's ability to define their own culture. Similarly, the Japanese government has established the Upopoy Ainu National Park and Museum in 2020. As this facility was spearheaded by *wajin* authorities and not members of the Ainu community themselves, scholars have viewed the Park as a strategy to increase tourism within Hokkaido rather than a means to preserve Ainu culture (Zaman et al.).

Decades of Ainu advocacy — from the Shakushain statue built in 1970 to language schools in the 1980s and international indigenous alliances in the 1990s — have successfully dismantled the legal architecture of assimilation under the Protection Act of 1899. These achievements thus paved the way for cultural recognition under the 1997 Cultural Promotion Act and Japan's official recognition of Ainu indigeneity in 2008. However, political milestones have routinely substituted cultural visibility for political agency: the state continues to celebrate Ainu

culture while denying them concrete political or economic rights, and excludes the population from geographic negotiations even as their identity is deployed to back territorial claims. As such, Ainu participation in determining cultural repatriation is still kept to a minimum, allowing the ethnic Japanese government to continue covertly exerting control over the population, hindering Ainu advocacy for true self-determination.

Conclusion

The Ainu have endured centuries of systemic assimilation and subjugation by the Japanese government. In the 17th century, Tokugawa era policies established a system of dependency and ethnic boundaries as the Ainu economy relinquished power to the *wajin*. This practice continued into the Meiji era, and the 1899 Hokkaido Former Aborigines Protection Act facilitated the Ainu's assimilation into *wajin* society. The Protection Act codified assimilation practices by erasing territorial rights, replacing traditional practices, and forcing agricultural livelihoods that undermined self-sufficiency. Meanwhile, pseudo-scientific racial theories reinforced the idea of a monoethnic Japan while simultaneously labeling the Ainu as primitive and barbaric, creating a paradox between the state's push for a unified Yamato identity and society's demand to differentiate between the *wajin* and Ainu. Decades later, these policies and social perceptions were reconsidered. Postwar Ainu advocacy, rooted in nonviolent resistance and cultural revitalization, pressured the state to respond to the growing demand for Ainu liberation. This culminated in the 1997 Ainu Culture Promotion Act, which shifted Japan's policies of Ainu assimilation to upholding Ainu culture and traditions. Shortly after, the Japanese Diet passed a resolution on June 6, 2008, formally recognizing the Ainu as one of Japan's indigenous populations.

All these achievements may seem like a well-deserved happy ending to the Ainu People's plight, but their fight is far from over. Ainu rebellions in the late 17th century were swiftly quelled, and assimilation policies made it near impossible for the socially ostracized, economically drained, and politically subjugated indigenous community to revolt. In more recent years, the 1997 Act substituted cultural visibility for political agency, while the 2008 recognition restricted Ainu self-determination. However, many Ainu activists and scholars are consciously working to restore their culture and advocate for indigenous rights within Japan.

This paper has addressed the historical trajectory of the Ainu People by formalizing a framework of Subjugation, Assimilation, and Advocacy and linking it to contemporary sociological theories, arguing that Japan's recognition of Ainu indigeneity has not translated into political autonomy. However, this research has several limitations. First, while it identifies general categories of needed change (including territorial claims, education, and governance participation), it does not propose specific methods or policy mechanisms through which the Ainu could achieve self-determination. Second, the three-phase framework, while analytically useful, narrowly confines a rich and complex Ainu history into a linear sequence. Thus, the framework risks flattening internal diversity, regional variation, and society in pre-modern eras before sustained Japanese contact. Future research should therefore develop concrete,

comparative governance models, recover pre-17th century Ainu history to restore a fuller picture of indigenous agency, and undertake comparative and ethnographic studies to test the three-phase framework beyond Japan, recontextualizing indigenous identity in modern societies.

In order to make Ainu self-governance a reality, the state must move beyond policies focused solely on cultural restoration. As laws are continually placed to revive indigenous culture, the Ainu rights movement is at risk of being reduced to a superficial matter of restoring tradition and nothing more. This restricts the Ainu from making distinct economic and political decisions to cement their status as equal citizens of Japan. Advocates have pushed for a move toward policies that restore territorial claims, strengthen education in Ainu language and history, and ensure the Ainu community's ability to participate in the governance of their own societies (Hammine 10). These policies would allow the Ainu people to govern themselves and affirm their status as the original inhabitants of Hokkaido. Until these structural issues are addressed, the legacy of assimilation — and the Ainu's fight for self-determination — will remain unfinished.

Works Cited

"Actual Living Conditions of the Hokkaido Ainu." Ainu Association of Hokkaido, www.ainu-assn.or.jp/english/life.html. Accessed 8 June 2026.

Anderson, Benedict. Imagined Communities : Reflections on the Origin and Spread of Nationalism. 2nd ed., Verso, 1991.

Asai, Yasufumi, and Jeffrey L. Arnold. "Terrorism in Japan." Prehospital and Disaster Medicine 18, no. 2 (2003): 106–14. https://doi.org/10.1017/S1049023X00000844.

Barth, Fredrik. Ethnic Groups and Boundaries: The Social Organization of Culture Difference. Little, Brown, 1969.

Cornell, John B. "Ainu Assimilation and Cultural Extinction: Acculturation Policy in Hokkaido." Ethnology, vol. 3, no. 3, 1964, pp. 287-304. JSTOR, https://doi.org/10.2307/3772885.

Dikötter, Frank. The Construction of Racial Identities in China and Japan : Historical and Contemporary Perspectives. 1997, dn721805.ca.archive.org/0/items/frank-dikotter-the-construction-of-racial-identities-in-china-and-japan-hong-kon/Frank%20Dik%C3%B6tter%20-%20The%20Construction%20of%20Racial%20Identities%20in%20China%20and%20Japan-Hong%20Kong%20University%20Press%20%281997%29.pdf.

Hammine, Madoka. "Indigenous in Japan? The Reluctance of the Japanese State to Acknowledge Indigenous Peoples and Their Need for Education." Sámi Educational History in a Comparative International Perspective, 2019, pp. 225-45, https://doi.org/10.1007/978-3-030-24112-4_13.

Howell, David L. "Ainu Ethnicity and the Boundaries of the Early Modern Japanese State." Past and Present, no. 142, 1994, pp. 69-93. JSTOR, www.jstor.org/stable/651197.

---. "Making 'Useful Citizens' of Ainu Subjects in Early Twentieth-Century Japan." The Journal of Asian Studies, vol. 63, no. 1, 2004, pp. 5-29. JSTOR, www.jstor.org/stable/4133292.

International Work Group for Indigenous Affairs. "The Indigenous World 2025: Japan." International Work Group for Indigenous Affairs, 25 Apr. 2025, iwgia.org/en/japan/5664-iw-2025-japan.html. Accessed 8 June 2026.

Kirikae, Hideo. "Social Aspects of the Ainu Linguistic Decline." Senri Ethnological Studies, vol. 44, 31 Mar. 1997, pp. 161-74, https://doi.org/10.15021/00002955.

Lambright, Bri. "The Ainu, Meiji Era Politics, and Its Lasting Impacts: A Historical Analysis of Racialization, Colonization, and the Creation of State and Identity in Relation to Ainu-Japanese History." History Summer Fellows at Digital Commons @ Ursinus College, digitalcommons.ursinus.edu/cgi/viewcontent.cgi?article=1031&context=history_sum.

Larson, Erik, et al. "Emerging Indigenous Governance: Ainu Rights at the Intersection of Global Norms and Domestic Institutions." Alternatives: Global, Local, Political, vol. 33, no. 1, 2008, pp. 53-82. JSTOR, www.jstor.org/stable/40645236.

Low, Morris. "Physical Anthropology in Japan." Current Anthropology, vol. 53, no. S5, 2012, pp. S57-S68. JSTOR, https://doi.org/10.1086/662334.

McGrogan, David. "A Shift in Japan's Stance on Indigenous Rights, and Its Implications." International Journal on Minority and Group Rights, vol. 17, no. 2, 2010, pp. 355-73. JSTOR, www.jstor.org/stable/24675792.

Powell, Courtney. "Reconsidering Resistance: Ainu Cultural Revival as Protest." Vienna Journal of East Asian Studies, vol. 14, no. 1, 23 Nov. 2022, pp. 1-23, https://doi.org/10.2478/vjeas-2022-0001.

Schoer, Bethany. "Commitment versus Convenience: Interpreting Ainu Culture in the Kuril Islands Dispute." SAGE International, 11 Aug. 2015. PDF.

Scott, James C. Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed. Yale UP, 1998.

Siddle, Richard. Race, Resistance and the Ainu of Japan. 3rd ed., Routledge, 2001.

Tsuda, Nobuko, and Hiroshi Maruyama. "Heading towards the Restoration and Transmission of Ainu Culture." Indigenous Efflorescence: Beyond Revitalisation in Sapmi and Ainu Mosir, by Gerald Roche et al., ANU Press, 2018. JSTOR, www.jstor.org/stable/j.ctv9hj9pb.27.

University of Oregon. "Ainu-e Paintings." Windows to the Ainu World, glam.uoregon.edu/s/uo-ainu-collections/page/ainu-e-paintings. Accessed 8 June 2026.

"University Returns Ancestral Remains to Ainu People in Japan." The University of Edinburgh News, 30 Apr. 2025, www.ed.ac.uk/news/2025/university-returns-ancestral-remains-to-ainu-people-in-japan. Accessed 8 June 2026.

Wacquant, Loic J.D. "For an Analytic of Racial Domination." Political Power and Social Theory, vol. 11, loicwacquant.org/wp-content/uploads/2018/02/lw-1997-for-an-analytic-of-racial-domination.pdf. Accessed 8 June 2026.

White, Richard. The Roots of Dependency : Subsistence, Environment, and Social Change among the Choctaws, Pawnees, and Navajos. University of Nebraska Press, 2018.

Zaman, Mashiyat, et al. "Critiquing the Colonialist Origins of the New National Museum Upopoy." FOCUS Asia-Pacific, Mar. 2022, www.hurights.or.jp/archives/focus/section3/2022/03/critiquing-the-colonialist-origins-of-the-new-national-museum-upopoy.html. Accessed 8 June 2026.

How European Colonization Shaped Indian Cuisine By Ruhaan Parandekar

Introduction

India is a country of 1.48 billion people. There are over 100 major languages, 6 major religions, and hundreds of ethnic groups, demonstrating the depth of culture that is ingrained in a country ⅓ the size of the United States of America. Given its extensive interactions with outside empires, the Indian cuisine that we know today remains a powerful reflection of not only the diverse culture of India, but also its deep history.

Pre-independence India was made up of hundreds of semi-autonomous princely states along with areas that were directly under the thumb of British colonization. Before that, the Portuguese had control over the coastal enclaves of India, the most important area being Goa. During their time, each of these powers did their part to influence and shift the culture of modern-day India, particularly in Indian cuisine. While the Portuguese arrived at the height of their maritime expansion and dominance over global trade routes, the British arrived at a time of intense imperialistic fervor and large amounts of expansion. As a result of this, each of these empires were not only highly consequential to Indian cuisine, but their effects were each distinct, reflecting their priorities in the region. Specifically, modern day India's interactions with Portugal brought about both new ingredients and new techniques that we still see ingrained within Indian cuisine today, while the British helped globalize Indian cuisine and introduce new ideas to what Indian cuisine could be. Along with this, the British ended up contributing one of the most used ingredients in current day India.

Context

Before India ever came into contact with the English or the Portuguese, there was already a unique and well-developed cuisine in North and South India. One example of this is the culinary techniques and traditions that came about through Mughal Rule. The Mughal empire played a role in merging Middle Eastern and Central Asian cooking techniques with the culinary traditions of the locals. This resulted in the development of famous Indian dishes such as *Biryani* and Kashmiri *Rogan Josh*.

Apart from the food of the Mughals, various Indian regions had their own staple foods inspired by factors such as region, caste, and class.

For example, a staple in what would be considered a peasant diet would be something called *Khichiri*, a dish that even the Mughals would begin to embrace and work with culinarily. *Khichiri* was usually made up of rice and lentils cooked together with water. When the Mughals took it up, they incorporated costly ingredients like "almonds, raisins, cloves, mace, nutmeg, cardamom, cinnamon, and pepper" (Collingham).

People's diet was also heavily influenced by their class and caste positions. Whereas Brahmans, the highest caste, saw engaging in vegetarianism as an act of purity, Rajputs, the warrior class of Ragasthan, regularly ate meat with spices such as coriander and cumin without their social position being brought down.

We can also see that during Mughal feasts, they would employ Hindustani cooks to make special pickles and chutneys, signifying that India had unique culinary traditions of their own, outside of anything the Mughals, British, or Portuguese offered them.

Portuguese Colonization

Although modern day India had an extensive cuisine, the coming of the Portuguese empire was still deeply formative to the Indian food that we know of today. This is because of the numerous ingredients and culinary techniques that the Portuguese introduced, and that are still staples of Indian households and restaurants.

The Portuguese first came as a result of their seizing of the spice trade from the Arabs. Because of their dominance over the spice trade, the Portuguese were able to introduce new ingredients to India that would change the course of Indian cuisine permanently. The chief example of this would be the chilli, a foundational spice of Indian cuisine. The chilli pepper, or capsicum, was first found by Christopher Columbus in the New World, where he observed the native populations heavily spicing their foods with it. Pre-1500, the people of current day India would eat long pepper and black pepper, as that was all that was available to them to spice their food at the time.

The chilli is suspected to have first entered India through Goa, which was also the capital that the Portuguese had set up for themselves. It was quickly embraced by people around the region, and was used to replace the similar long pepper, which was harder to grow and store. Ayurvedic practitioners started using the chilli as a means of curing cholera while the people of the peasant class would use it to spice up otherwise bland dishes of *daal* and rice. Even today, within South India, the heavy-handed use of chillies helps impoverished people increase their vitamin C intake. Not only this, but the various kinds of chillies that stemmed from the first capsicum that the Portuguese brought, are still used all throughout India. In South India, dried chillies are used in Sambar, and fresh chillies are used in various spicy *chutneys*. In North India, Kashmiri red chilli powder is used for tandoori food, while green chillies are used to make *thecha*, a Marathi condiment.

The generous amounts of spice Indian cuisine is now known for all around the world would not have come about if it weren't for the introduction of the chilli into India, by the Portuguese.

The chilli is a large example of the Portuguese altering the course of Indian cuisine through the introduction of new ingredients, but there are also many others that deserve mentioning as well. Besides chillies, the introduction of cinnamon and cloves to India through the Portuguese also greatly altered the course of Indian cuisine. In the present day, we see these both frequently used in Masala *Chai*, desserts like *Kheer*, and various rice dishes from the Mughal era such as Pulaos and Biryanis. Aside from the alteration of Indian cuisine's spice profile, the Portuguese also introduced the potato, the tomato, and the cashew, all used in curries, sabjis, and various rice dishes throughout North and South India. You would struggle today to go to a North Indian restaurant without some sort of tomato and cashew based curry. The same goes

for the use of potato in South Indian restaurants, used as a filling for *dosas*, or as a standalone vegetable.

But Portuguese influence ran deeper than with the ingredients that they brought to India. They also introduced longstanding culinary techniques and dishes that are still used today. Although Portugal was a small country, their weather was ideal for "the cultivation of wheat, pigs, sheep, olive oil, and various manifestations of grape" (Collingham). Along with this, because of the readily available selection of spices, nuts, and grains, Portugal formed a rich and diverse culinary portfolio. South India was not lacking in distinctive flavors either. The environment was ideal for the growth of rice, making it a staple crop. Along with this, spices like cardamom, pepper, and cloves were able to grow, and when combined with coconuts sourced from coconut trees, and tamarind introduced by Arab traders, South Indian cuisine would form a distinctive flavor profile that would soon be merged with Portuguese cuisine.

A large example of this is through the invention of *Pav*. For the Portuguese, leavened wheat bread was a large part of their diet, and also carried religious significance for them. The South Indian diet consisted of unleavened bread derived from fermented rice and lentils-food like *idli, dosa, and appams*. Because wheat was already grown in North India, it wasn't an issue securing that, but the yeast that would be used to raise the bread was difficult to obtain in the region. As a result of the rising demands for bread, the Goan cooks realized that they could use fermented palm sap, which naturally had yeast, to raise the bread. This gave birth to what the Portuguese called *Pao*, which is a fluffy white bread. In the present day, *Pav* (what Indians call it) is used in numerous popular Indian dishes such as *Pav Bhaji* and *Vada Pav*.

It should also be noted that during this time, Goa became renowned for its rich Portuguese-style baking culture. The wives of Portuguese bakers were responsible for passing down the information about baking, and developed a reputation themselves for their well-made confections. Along with this, the Moghs, or Bengali Buddhists, were also employed and given the knowledge required to be skilled confectioners. Throughout this time, a popular Goan dessert named *Bebinca*, made up of eggs, coconut milk, and sugar came about, and is still eaten as a remnant of Portuguese and Indian history.

During the time of Portuguese-Occupied Goa, people from Portugal would eat beef and pork, two taboo foods to Muslims and Hindus around India. Over time, people around the area began to leave their past traditions and consume red meats such as beef and pork, as a direct result of the mass conversion efforts by the Catholic Portuguese. This was the context for the formation of *Vindaloo*. The Portuguese would marinate their meat in wine vinegar, but as India didn't have the resources to make wine vinegar, they derived their vinegar from coconut *toddy* (fermented palm sap). The meat would then be marinated with ingredients like tamarind and garlic before it was cooked. After this, ingredients like "*garam masala*, black pepper, cinnamon, and cloves"(Collingham) were added to a sauce which the meat was subsequently submerged in. *Vindaloo* is now a staple of Indian restaurants and curry houses all over the world, and is one of the most famous Goan dishes today.

British Colonization

We see how the Portuguese fundamentally changed what we know today as Indian cuisine through the introduction of new ingredients, cooking techniques, and dishes. In contrast, the British were responsible for globalizing Indian cuisine, and in some instances, even helping form new foods and drinks that were destined to become staples in India.

The story of the British colonization of India starts off with the East India Company. The company got its start when it brought together a group of investors, and a year later got a royal charter from the crown. By this point, the Dutch had broken Portugal's monopoly on East India trade, so because of the power vacuum that was created, Britain was allowed to conduct business in the region. Although they started with Indonesia, they quickly realized that there was a large demand for Indian textiles. In 1614, the British succeeded in securing a *firman* from the Mughals, which gave them favorable trade conditions along with a permit to start a factory in Surat, a hub of trade at the time. British demands for textiles persuaded the East India Company to focus on business within India, during a time when cotton was in higher demand than spices were. When trade in Surat started dying down, the already expanded East India Company moved their center of operations to Bombay in the west coast, Madras in the east coast, and Calcutta even further east, demonstrating the breadth of influence they had accumulated all over India. Although imperialism wasn't the British merchants' priorities, they saw that over time, expanding territory and getting more involved in Indian politics would help them secure their profits. Entrenched in the economy of India, the merchants secured favorable trade deals with local Mughal leaders. At this point in time the East India Company was building up their own private army, leading to the Mughals ceding Bengal, "one of India's richest provinces"(144). As the Company's army expanded, so did their influence beyond their autonomous boundaries, particularly in Indian Courts. The British started to become serious competitors for power with the Mughals, which climaxed when the British were able to annex Oudh, a Mughal stronghold. By 1805, the British had either outcompeted or militarily defeated the Dutch, the Mughals, and the French who they were having ongoing military conflicts with, effectively making the British the governing authority of India.

With Britain's dominance over India came new opportunities for citizens of Britain to live in India, usually for work related reasons. With this, what we considered Indian food today was shifted, and sometimes diluted in order to fit the sensibilities of the new ruling class.

We can see this with the rise of curry, a term that while not commonplace during British arrival, was something that the British quickly found a taste for. The concept of curry was invented by the British to group together foods that they saw as similar, often ignoring the regional nuances and differences for the simplicity of the term "curry". It was used to describe any spicy and thick sauce dish that was made in India. Because of their ignorance of differences within Indian cuisine, they would try to incorporate ingredients and culinary concepts from various regions to make curries that were suitable to their tastes. An example of this that we can see in contemporary Indian restaurants is the use of mangoes used in southern coastal areas, to put in Bengali Shrimp Curries, resulting in a mango-shrimp curry. The British successfully

implemented appealing aspects of various regional cuisines into their Anglo-Indian food, making it what Liz Collins calls “the first truly pan-Indian cuisine”.

It should be mentioned that although some dishes may be more popularized in the west today, a lot of the locals at the time didn’t indulge in this new hybrid cuisine, these dishes were usually meant to fit British sensibilities without consideration for the culture of the Indians. The British would form their own sort of generic base for what they would call a curry, including a standardized set of spices mixed with onions, garlic, chillies, and subsequently cooked with some sort of dairy and meat. Another example is mulligatawny soup, made up of a base of rasam, which is a broth made with pepper, chillies, and usually some tamarind. This base was combined with chicken to fit the British people’s hunger for meat. As of today, although it is served in India, it is far more popular in western countries, showing how not all of the changes to what we perceive as Indian cuisine was actually made for Indians.

The invention of Anglo-Indian cuisine is contrasted with the introduction of chai to India. Tea originated from modern day China, and was well in use in the regions around India such as China and Tibet, but tea wasn’t commonly consumed in India regardless. The Indians preferred water and buttermilk. Tea was left for upper class Muslims, and the very sick. Tea also gained immense popularity throughout British society, causing it to be a large source of revenue for the East India Company. Due to fears of the company losing its monopoly on Chinese trade, where they sourced all of their tea from, they began efforts to cultivate tea in India. Although tea was a success outside of India, the Indian Tea Company realized that the locals of India were an untapped market. Due to the Indian population’s apprehensive attitude towards tea, the Indian Tea Company took up extensive marketing efforts to get the local population to indulge in the popular drink. Examples of this included starting the concept of tea breaks on jobs, serving chai on trains, and going house to house doing tea demonstrations. All of these efforts succeeded in getting India to finally indulge in tea, but with some twists. The Indians tended to make their *chai* (tea) milky and sweet. Some even added masalas such as cardamom, cinnamon, and cloves to their chai, forming the popular Indian street drink *Masala Chai*. As a direct result of Britain’s marketing practices aimed at making profit off of an open market, a new cultural staple was formed, and it is still enjoyed to this day.

With the fact that Anglo-Indians took to Indian cuisine, and even interpreted it in their own way, it is no surprise that the recipes of this time would make their way back to England. Retired East India Company officials would come back home with a hunger for the foods that were staples of their diet back in India. Because of this, there was a rise in curry houses, marketed towards people who wanted to keep eating the foods that they would have back in India.

Along with this, Indian servants were often brought back to England by Anglo-Indians who didn’t want to give up the luxuries that they had access to beforehand. These factors combined with the spread of information through shared recipes made it so that Indian culture and food was effectively diffusing to the western world.

Conclusion

Today, whether it's food on a street corner in India, or an Indian restaurant in the Bay Area, the Indian cuisine that we see today is a reflection of the deep history of colonization that India itself has gone through. While the Portuguese fundamentally altered what we know as Indian cuisine through the introduction of culinary techniques and staple ingredients, the British helped bring together the techniques and ingredients from different regions, introduce a staple drink, and globalize what we know as Indian food in the west. Every time we take a bite of curry, sip Chai or grab an Indian pastry, we can remember that it isn’t just a meal, it’s a confluence of the different cultures, people, and traditions that India has met over its history.

Work Cited

"India Population (2026)." Worldometer, www.worldometers.info/world-population/india-population/. Accessed 13 June 2026.

Marketing, Elite Asia. "How Many Languages Are in India? Language Data for India 2026." Elite Asia, 27 Mar. 2026, www.eliteasia.co/how-many-languages-are-in-india/.

Kramer, Stephanie. "Population Growth and Religious Composition in India." Pew Research Center, Pew Research Center, 21 Sept. 2021, www.pewresearch.org/religion/2021/09/21/population-growth-and-religious-composition

Size of the United States Compared to India, www.mylifeelsewhere.com/country-size-comparison/united-states/india. Accessed 14 June 2026.

Antani, Vishu, and Santosh Mahapatra. "Evolution of Indian Cuisine: A Socio-Historical Review - Journal of Ethnic Foods." SpringerLink, BioMed Central, 28 Apr. 2022, link.springer.com/article/10.1186/s42779-022-00129-4.

Srinivas, Tulasi. Exploring Indian Culture through Food - Association for Asian Studies, www.asianstudies.org/publications/eaa/archives/exploring-indian-culture-through-food/. Accessed 14 June 2026.

Collingham, Lizzie. Curry: A Tale of Cooks and Conquerors. Vintage Digital, 2010.

Tankaria, Cheragh. "Tracing the History of India's Favourite Beverage." J'AIPUR by Rupi Sood, 18 Apr. 2025, www.jai-pur.com/journal_entry/a-bitter-sweet-cup-of-chai/.

The Evolution of Chai in India · Global Stimulants · Exhibits @ Lafayette College, exhibits.lafayette.edu/s/mate/page/the-evolution-of-chai-in-india. Accessed 14 June 2026.

Saldanha, Ayesha. "How the Portuguese Influenced Indian Cuisine." Global Voices, 27 Feb. 2014, globalvoices.org/2014/02/27/how-the-portuguese-influenced-indian-cuisine/.

Indian Cuisine | History, Regions, Dishes, & Facts | Britannica, www.britannica.com/topic/Indian-cuisine. Accessed 14 June 2026.

The Opportunity Gap: Women's Migration Aspirations as a Development Indicator in Jordan By Leyth Sharaf

Abstract

Jordan has made substantial progress in female educational attainment over the past two decades, producing one of the most educated generations of women in its history. Yet educational gains have not been matched by equivalent advances in economic participation or representation. Female labour force participation remains among the lowest in the world, while migration aspirations continue to be widespread among both women and highly educated Jordanians.

Drawing on migration theory, gender-focused scholarship, and twenty-five semi-structured interviews with Jordanian women aged 18–25, this paper argues that women's migration aspirations reflect a perceived gap between expanding capabilities and constrained opportunities. Although educational attainment has broadened aspirations, expectations, and awareness of alternative futures, pathways to employment, leadership, and meaningful participation have expanded more unevenly. In this sense, migration aspirations provide insight into how individuals evaluate the opportunities available to them and the extent to which they believe their aspirations can be realised within their own society.

Rather than viewing migration aspirations solely as a migration outcome, this paper proposes that they may also function as a development indicator. In doing so, it suggests that the relationship between human capability and opportunity constitutes a critical measure of development itself.

Keywords Jordan; Women; Migration Aspirations; Development; Human Capital; Gender; Labour Force Participation; Economic Modernisation Vision

1. Introduction

Jordan has spent decades investing in the education of its women. The results are visible across the country's schools, universities, and professional institutions. However, educational progress has not produced an equivalent transformation in economic and public life. Female labour force participation remains among the lowest globally, and women continue to be underrepresented in leadership positions across both public and private institutions.

This gap between educational achievement and limited opportunity lies at the heart of a broader development question. Jordan's Economic Modernisation Vision (“EMV”) identifies talent development, innovation, productivity, and skills acquisition as essential components of future growth. The EMV seeks to integrate more than one million young Jordanians into the labour market by 2033 and emphasizes the importance of maximizing the country's human capital potential (Government of Jordan, 2022). Such ambitions depend on the ability of institutions, labour markets, and governance structures to absorb and utilize the capabilities that education creates. For young women, however, the central question is whether those capabilities can be translated into meaningful opportunities. Development is measured by the expansion of

human potential as well as by the opportunities available for its realization. Accordingly, the challenge is not simply educational attainment, but whether education is accompanied by pathways to participation, advancement, and influence.

The question makes migration aspirations especially relevant. Arab Barometer VIII survey data indicate that 42 percent of Jordanians report having considered emigration (Arab Barometer, 2024). Among women, the figure stands at 34 percent, while approximately 50 percent of individuals with at least a university degree indicate a desire to migrate (Arab Barometer, 2024). Although Arab Barometer VIII identifies economic factors as the most frequently cited motivations for emigration, recent scholarship suggests that, for some young Jordanian women, migration aspirations may also reflect broader perceptions of opportunity, autonomy, and the ability to realize personal and professional aspirations (Cosmano, 2024). This suggests that migration aspirations provide insight not only into mobility preferences, but also into how individuals evaluate the opportunities available to them within their own society.

United Nations migrant stock estimates further indicate that approximately 815,000 Jordanians were living abroad in 2020, of whom 37 percent were female (UN DESA, 2020). Although annual emigration data are limited, the available evidence suggests that migration remains an established pathway for many Jordanians. The significance of these figures lies in the scale of mobility itself as well as in what migration may reveal about how individuals assess their future prospects and opportunities for advancement.

While migration aspirations are typically studied as predictors of mobility, this paper proposes that they can also be understood as a development indicator. Women's migration aspirations may reveal whether educational investments are being translated into meaningful opportunities for participation, leadership, and advancement. More broadly, they provide a lens through which to examine the relationship between human capability and opportunity.

2. Methodology

This paper adopts a qualitative approach combining secondary source analysis with primary interview data. The analysis draws upon migration theory, survey data, and gender-focused academic and policy literature. Key sources include Arab Barometer Wave VIII, World Bank gender indicators, OECD SIGI data, and studies by Hausmann et al. (2019), Adely (2024), and Cosmano (2024).

To complement this literature, twenty-five semi-structured interviews were conducted with Jordanian women aged 18–25 in 2025 – 2026 from different cities across Jordan. The sample included university students, employed professionals, and secondary-school graduates in order to capture a range of educational and career trajectories. Participants were recruited through a combination of personal contacts and snowball sampling, whereby interviewees recommended other potential participants who met the study criteria.

Interviews were conducted in either Arabic or English depending on participant preference. Responses originally provided in Arabic were translated into English by the author for reporting purposes. Responses were also analysed thematically to identify recurring patterns

concerning opportunity structures, agency, and migration aspirations. To protect participant privacy, interviewees were anonymized and are identified only by participant number, age, location, and broad educational or employment category. Given the qualitative nature of the study, the findings are not intended to be statistically representative of all Jordanian women. Instead, they provide exploratory insight into how young women across different regions of Jordan perceive opportunities and evaluate their future prospects.

3. Discussion

3.1Migration Aspirations and Capabilities

Migration has traditionally been explained through frameworks emphasizing economic disparities between countries. Classical theories generally portray migration as a rational response to differences in wages, employment prospects, and living standards. While these approaches remain influential, contemporary migration scholarship increasingly views migration as a process shaped by aspirations, capabilities, and perceptions of opportunity.

Among the most influential contributions to this shift is de Haas's aspirations-capabilities framework. Rejecting simplistic push-and-pull explanations, de Haas (2021) argues that migration results from the interaction between an individual's aspiration to migrate and their capability to do so. Aspirations are shaped by how people perceive opportunities, imagine alternative futures, and evaluate their life chances, while capabilities reflect the resources, education, social networks, and institutional conditions that make mobility possible. As individuals acquire education and skills, they become more aware of opportunities elsewhere and develop stronger aspirations regarding their professional and personal futures. In this way, migration aspirations may emerge from the desire to realize capabilities more fully.

This framework is particularly relevant to the Jordanian context. Over the past two decades, Jordan has achieved substantial progress in female educational attainment, producing one of the most educated generations of women in its history. At the same time, migration aspirations appear particularly pronounced among more educated respondents, with approximately 50 percent of individuals holding at least a university degree reporting that they have considered emigrating (Arab Barometer, 2024). Consistent with de Haas's argument, expanding educational capabilities may increase awareness of opportunities and aspirations for professional advancement, even when opportunities within the domestic context remain constrained. In such circumstances, migration aspirations can reveal a perceived imbalance between what individuals believe they are capable of achieving and the opportunities available to them.

Carling and Schewel (2018) emphasize the importance of distinguishing between migration aspirations and migration outcomes. The desire to migrate and the ability to migrate are not synonymous. Many individuals who aspire to migrate never do so, while others who possess the capability to migrate choose to remain. Consequently, migration aspirations constitute a distinct object of analysis rather than simply a predictor of future migration.

This distinction carries broader implications for how development is assessed. Traditional indicators often measure what individuals possess, including years of schooling, income, or formal qualifications. Migration aspirations, by contrast, may reveal how individuals evaluate the opportunities available to convert those capabilities into meaningful life outcomes. In this sense, aspirations provide insight into perceptions of inclusion, possibility, and future prospects.

3.2 Educational Achievement and Labour Market Barriers

In recent years, Jordan has achieved substantial progress in female educational attainment. Female literacy has reached 97.5 percent, and girls complete lower secondary education at higher rates than boys (World Bank, 2025). Yet these achievements coexist with one of the lowest rates of female labour force participation in the world, with only 16 percent of women participating in the labour market compared with 62.5 percent of men (World Bank, 2025).

This apparent contradiction has been described by Hausmann et al. (2019) as Jordan's female labour market "exclusion puzzle." Their analysis demonstrates that educational attainment increases women's willingness to participate in the labour market, yet many highly educated women continue to encounter elevated unemployment and limited access to suitable employment opportunities. As a result, the challenge is not a shortage of human capital but the limited availability of opportunities capable of absorbing and utilising that capital. The issue is less one of capability creation than capability utilization.

Evidence from PwC's *Young Women, Powerful Ambitions* report reinforces this conclusion. Across the Middle East, 84 percent of young women reported aspiring to leadership positions, while 86 percent expressed confidence in their ability to achieve their professional goals (PwC Middle East, 2022). Aspiration, ambition, and educational attainment are evidently not lacking, but the more significant question is whether existing economic structures provide pathways through which those aspirations can be realised.

Several barriers emerge consistently throughout the literature. Arab Barometer Wave VIII identifies childcare responsibilities, job scarcity, workplace bias, low wages, and inflexible working arrangements as significant obstacles to women's participation in the labour market. Hausmann et al. (2019) similarly identify childcare provision and transportation constraints as major factors limiting employment outcomes, while Turkanik (2017) highlights the continued influence of social expectations surrounding women's economic roles.

The consequences of these dynamics are illustrated in Adely's (2024) ethnographic study of educated Jordanian women who migrated internally to Amman in pursuit of employment and professional advancement. Adely argues that educational expansion has broadened women's aspirations regarding work, mobility, and independence, yet many continue to encounter economic and institutional barriers that restrict the realisation of those aspirations. Education therefore functions as a double-edged sword. It expands capabilities and horizons while simultaneously exposing the limitations of existing opportunity structures**.** The very process that

broadens aspirations may also heighten awareness of institutional and economic constraints that continue to limit their realization.

From a development perspective, this gap is significant. Educational investment has succeeded in expanding women's capabilities, but opportunities have not expanded at the same pace. The result is a widening disconnect between what many young women are qualified to achieve and what they perceive to be realistically attainable within the Jordanian labour market. This disconnect is not merely an economic concern; it raises broader questions regarding inclusion, participation, and the ability of development to translate capability into opportunity. For example, educational attainment may expand what people are capable of doing, but its transformative potential ultimately depends upon whether institutions allow those capabilities to be exercised.

3.3 Gendered Opportunity Structures and Migration Aspirations

Understanding migration aspirations among Jordanian women requires attention to the gendered social structures within which opportunities are experienced and evaluated. Mahler and Pessar (2006) argue that migration is fundamentally shaped by gendered social relations. Access to resources, mobility, employment opportunities, and decision-making power are all influenced by gender, affecting both the ability and desire to migrate. As such, migration decisions cannot be understood solely through economic considerations but must also be situated within the broader social and institutional environments that shape individuals' life chances.

The Jordanian context reflects many of the dynamics identified by Mahler and Pessar (2006). Jordan records a SIGI score of 58, placing it among countries characterized by very high levels of discrimination in social institutions (OECD, 2023). The country's highest score relates to discrimination within the family, where Jordan scores 89. Such findings suggest that opportunity structures are shaped not only by labour market conditions but also by institutional arrangements governing family relations, inheritance rights, guardianship, and broader questions of personal autonomy. Opportunities are not experienced in a social vacuum; they are mediated through institutions that shape access to resources, mobility, and decision-making power.

Public attitudes reinforce these constraints. According to OECD data, 72 percent of Jordanians agree that men make better political leaders than women, while 85 percent believe children suffer when mothers work outside the home (OECD, 2023). Furthermore, 82 percent of Jordanians agree that when jobs are scarce, men should have greater rights to employment than women (OECD, 2023). Such attitudes shape hiring practices, promotion decisions, and broader perceptions regarding women's place in the labour market.

Institutional outcomes reflect these broader dynamics. Women held 19.6 percent of seats in Jordan's national parliament in 2024 (World Bank, 2025), while only 3 percent of firms are headed by women (OECD, 2023). Although initiatives supported by UN Women Jordan have contributed to improvements in women's representation within managerial and supervisory roles, significant disparities remain (UN Women Jordan, 2025). These indicators suggest that

educational progress has not been matched by equivalent gains in leadership, representation, and economic participation.

Within this context, migration aspirations may reflect more than the pursuit of employment. Drawing on interviews with young Jordanian women, Cosmano (2024) argues that migration emerges partly from a perceived inability to realize desired aspirations within existing social and institutional structures. Rather than reflecting only economic motivations, migration aspirations often express broader desires for autonomy, self-determination, and the ability to pursue alternative life trajectories.

Cosmano (2024) further demonstrates that education frequently functions as more than a mechanism for acquiring skills or improving employability. For many participants, educational mobility represented a pathway toward greater independence, expanded life choices, and the pursuit of identities and futures that they perceived as difficult to achieve within Jordan. Migration, in this sense, becomes not only a response to opportunity deficits but also a means through which individuals seek to negotiate alternative futures and expand the range of choices available to them. These findings are reinforced by Adely's (2024) ethnographic account of educated Jordanian women navigating the intersection of educational aspiration and social constraint. Adely demonstrates that for many participants, the pursuit of education was itself bound up with questions of identity, autonomy, and the desire to inhabit professional and personal futures that felt genuinely self-authored. In this respect, the inspiration to migrate and the aspiration to pursue education share a common root: they are both expressions of the desire to exercise agency in the conditions where that agency is structurally constrained.

Viewed through this lens, women's migration aspirations can be understood as an indicator of a broader gap between expanding capabilities and the social, institutional, and cultural opportunities available for their expression.

3.4 Interview Findings

To complement the quantitative and literature-based analysis, twenty-five semi-structured interviews were conducted with Jordanian women aged 18–25 across multiple locations in Jordan during 2025 - 2026, including Amman, Irbid, Zarqa, Aqaba, Madaba, Karak, Salt, Jerash, and Mafraq. The interviews aimed to explore how participants evaluated opportunities available to them within Jordan and whether they had considered pursuing opportunities abroad. Semi-structured interviews were selected because they allowed respondents to discuss their experiences and aspirations in greater depth while ensuring consistency across key themes.

Interview questions were designed to explore participants' career aspirations, perceptions of opportunities available to young women in Jordan, attitudes toward migration, and views regarding professional advancement and personal development. Participants were asked whether they had considered studying, working, or living abroad, what factors would influence their decision to migrate or remain in Jordan, and which opportunities they perceived as limited within Jordan. Additional questions explored the importance of employment opportunities, financial security, leadership and career advancement, self-development, independence, international

exposure, and entrepreneurship in shaping future plans. Participants were also asked what changes or policy reforms would make them more likely to remain in Jordan and improve opportunities for young women.

Table 1. Interview Participant Characteristics (n = 25)

Group	Number
University students	8
Employed professionals	7
Secondary school graduates/non-university youth	10
Total	25

Age range: 18–25
Location: Amman, Irbid, Zarqa, Aqaba, Madaba, Karak, Salt, Jerash, and Mafraq
Source: Author interviews (2025 - 2026).

Table 2. Geographic Distribution of Participants (n = 25)

Location	Participants
Amman	7
Irbid	3
Zarqa	3
Aqaba	2
Madaba	2
Karak	2
Salt	2
Jerash	2
Mafraq	2

Source: Author interviews (2025 - 2026).

The geographic diversity of the sample was intended to capture perspectives beyond the capital and to explore whether migration aspirations were concentrated among particular regions or educational groups. While motivations varied, participants across all locations expressed concerns regarding the availability of opportunities capable of matching their aspirations and qualifications. Participants frequently identified multiple motivations for considering opportunities abroad. Better employment opportunities emerged as the most frequently cited motivation, mentioned by 21 of the 25 participants. Self-development and personal growth were identified by 18 participants, while 17 emphasized international exposure and networking opportunities. Sixteen participants highlighted greater independence and autonomy, and 15 identified higher salaries and financial security as important considerations. These findings

suggest that migration aspirations are closely connected to broader aspirations concerning personal and professional development rather than simply economic gain.

Table 3. Themes Identified in Interviews (n=25)

Theme	Participants Mentioning Theme
Better employment opportunities	21
Self-development and personal growth	18
International exposure and networking	17
Greater independence and autonomy	16
Higher salaries and financial security	15
Leadership and career advancement	14
Entrepreneurial ambitions	10
Would remain in Jordan if opportunities improved	20
No desire to migrate	5

Source: Author interviews (2025 - 2026).

Despite these differences, respondents across all groups consistently expressed concern regarding the availability of opportunities capable of matching their aspirations and qualifications within Jordan. One university student from Irbid (P9, age 19) explained that she viewed migration as “an opportunity to gain experience and build networks that would eventually allow me to create opportunities of my own in Jordan.” Another university student from Amman (P7, age 21) similarly noted, “I spent years studying and developing my skills, but it is difficult to find opportunities that match my qualifications. Looking abroad sometimes feels like the only way to fully use what I have learned.” An employed professional from Aqaba (P18, age 23) emphasized that migration was not primarily about income, stating that “the chance to learn, develop professionally, and gain independence matters as much as salary.”

A secondary-school graduate from Amman (P1, age 18) similarly emphasized employment opportunities and future stability, explaining that "my biggest concern is finding stable work and being able to build a secure future. I would prefer to stay in Jordan, but opportunities need to match the effort young people put into their education and training." Likewise, a secondary-school graduate from Madaba (P6, age 18) highlighted the importance of financial security, noting that "many young people think about opportunities abroad because they want a stable future and greater independence. If I could see clear opportunities for growth and advancement in Jordan, I would rather remain close to my family and community."

Perhaps most significantly, 20 of the 25 participants indicated that they would be more likely to remain in Jordan if opportunities for professional advancement, leadership development, and career progression improved. As one employed professional from Amman

(P15, age 24) stated, “I would rather stay in Jordan if I could see a clear path for professional growth and leadership.” Another participant from Zarqa (P10, age 25) emphasized that migration was not necessarily a rejection of Jordan itself, explaining that “I would prefer to build my future in Jordan, but I want to know that my education and effort can lead somewhere. If I cannot find those opportunities here, then I have to look elsewhere.”

Notably, respondents rarely framed migration as an abandonment of Jordan. Many expressed strong attachments to family, community, and country. Their aspirations to migrate were often accompanied by a desire to acquire skills, experience, and networks that could ultimately be applied within Jordan itself. Five participants indicated that they did not currently aspire to migrate. These respondents generally emphasized family ties and existing career satisfaction, or optimism regarding future opportunities within Jordan. Interestingly, four of the five were either employed professionals or individuals who described strong family and community connections. Their accounts are analytically significant. Where employment was stable and professional recognition felt attainable, the perceived gap between aspiration and opportunity narrowed. And with it, the appeal of migration receded. One employed professional from Irbid (P5, age 24) noted that while she had considered working abroad earlier in her career, growing responsibilities and a sense of professional progress had made remaining in Jordan feel like a viable and fulfilling choice. These responses suggest that migration aspirations are not fixed dispositions but responsive evaluations shaped by whether individuals perceive meaningful pathways to advancement within their current environment. The non-aspirants thus illuminate the conditions under which Jordan can retain its most capable women, not through restriction but through expansion of opportunity.

The findings suggest that migration aspirations among young Jordanian women are largely shaped by the interaction between educational achievement, labour market opportunities, and gendered social structures. Understanding these aspirations requires attention to both the opportunities women possess and the environments within which those opportunities can be realised.

The three strands of evidence examined in this paper converge on a consistent finding. De Haas’s aspirations-capabilities framework suggests that educational expansion will raise aspirations faster than opportunity structures can accommodate them, generating migration aspirations as a by-product. The secondary literature confirms that Jordan has succeeded in expanding women’s capabilities through education while barriers relating to childcare, transportation, social norms, and institutional structures have constrained the conversion of those capabilities into economic participation, leadership, and influence. And the interview findings demonstrate that this dynamic is experienced personally and concretely with women across nine cities, three educational categories, and a range of professional circumstances identify the same gap between what they are capable of achieving and what they believe they can realistically attain within Jordan. In this sense, migration aspirations are not a distinct or peripheral phenomenon. They are the subjective expression of a structural imbalance, and for that reason, they can function as a meaningful development indicator.

4. Policy Implications

Jordan has invested heavily in the education of its women and has produced one of the most educated generations in its history. Yet education alone cannot fulfil its developmental promise. The ultimate measure of progress is not only whether capabilities are created, but whether institutions, labour markets, and social structures allow those capabilities to be exercised. When aspirations consistently outpace opportunities, migration may become one mechanism through which individuals seek to realise their potential elsewhere.

At the same time, migration aspirations should not be interpreted as a direct measure of dissatisfaction or national failure. Decisions about migration are shaped by a wide range of factors, including family networks, educational opportunities, income differentials, lifestyle preferences, and broader processes of globalisation. Nor are migration aspirations universal. Many highly educated Jordanian women choose to build their futures within Jordan and remain deeply committed to contributing to their communities and country.

Migration aspirations should consequently be understood as one indicator among several that can illuminate how individuals evaluate opportunities, constraints, and future prospects. Viewed in this way, they provide policymakers with a valuable lens through which to assess whether development is expanding not only capabilities, but also the opportunities necessary for those capabilities to flourish.

4.1Expanding Childcare Infrastructure as Labour Market Policy

One of the most consistent findings across the evidence reviewed in this paper concerns childcare. Arab Barometer identifies childcare availability as a significant barrier to women's participation in the labour market, while Hausmann et al. (2019) similarly highlight childcare responsibilities as a major factor shaping employment outcomes. UN Women has also supported the development of Jordan's National Childcare Policy as part of broader efforts to strengthen women's economic participation (UN Women Jordan, 2025).

Childcare should be understood as labour market infrastructure rather than solely as a social welfare service. A mixed public-private model could combine government childcare subsidies, employer-supported childcare facilities, and community-based childcare centres. Large employers could be incentivized through tax credits or other policy mechanisms to establish workplace childcare services, while targeted childcare vouchers could support low- and middle-income families. Such measures would reduce one of the most consistently identified barriers to women's labour force participation while improving career continuity, retention, and long-term advancement.

4.2 Improving Transportation and Workplace Accessibility

Transportation policy should be integrated into labour market policy. Research consistently identifies transportation barriers as a significant obstacle to women's employment, particularly for those living outside major commercial centres or lacking access to reliable private transportation (Hausmann et al., 2019).

Several practical measures could help address this challenge. Employer-sponsored shuttle services connecting residential areas with industrial zones, business districts, and technology parks could improve access to employment opportunities. Public-private partnerships could expand safe and reliable transportation options during early morning and evening hours, when safety concerns often limit mobility. Targeted transportation subsidies for women entering the labour market may also improve workforce participation and reduce barriers to employment.

4.3 Supporting Female Entrepreneurship and Innovation

Several interview participants identified entrepreneurship as an attractive pathway toward professional fulfilment, autonomy, and economic independence. These aspirations align closely with Jordan's EMV, which emphasizes innovation, productivity, and private-sector growth.

However, aspiring female entrepreneurs often face barriers related to access to finance, professional networks, mentorship opportunities, and business development support. Expanding female-focused incubators, startup accelerators, entrepreneurship training programmes, and mentorship networks could strengthen pathways for women seeking to create opportunities rather than simply access existing employment. Greater access to microfinance and early-stage investment capital would further support women-led enterprises and contribute to broader economic diversification.

4.4 Strengthening Women's Leadership Pathways

The evidence presented throughout this paper suggests that the challenge facing Jordanian women extends beyond labour market participation. It also concerns access to leadership, influence, and decision-making. Educational achievement has expanded significantly, yet women remain underrepresented in senior leadership positions across both public and private institutions.

Addressing this gap requires more than increasing employment. It requires strengthening the pathways through which women advance into positions of authority. Leadership development programmes, mentorship initiatives, sponsorship networks, executive training opportunities, and targeted professional development schemes can all contribute to this objective.

Visible pathways to advancement also influence how opportunities are perceived. Women are more likely to view professional environments as offering meaningful futures when examples of female leadership are present and attainable. Strengthening leadership pathways contributes directly to reducing the gap between aspirations and opportunities.

4.5 Institutionalizing Women's Participation in Economic Decision-Making

One of the most important yet frequently overlooked dimensions of participation concerns representation within advisory and consultative structures. Public influence is exercised not only through elected office but also through councils, committees, task forces, advisory boards, and strategic planning bodies that shape national priorities.

Jordan could establish a National Women's Economic Advisory Council composed of representatives from government, industry, entrepreneurship, academia, and civil society. Such a council could provide structured consultation on labour market reform, childcare policy, entrepreneurship initiatives, digital economy strategies, and workforce participation programmes. Similar consultative bodies could be developed at municipal and governorate levels to ensure that policy development reflects regional differences in women's experiences and opportunities.

Increasing women's participation in advisory and strategic institutions would improve the quality of decision-making while simultaneously strengthening perceptions that women possess meaningful pathways to influence within Jordan's development trajectory.

4.6 Monitoring Migration Aspirations as a Development Indicator

The central argument of this paper is that women's migration aspirations provide valuable information regarding the relationship between capabilities and opportunities. Consequently, migration aspirations should be monitored alongside more conventional development indicators.

Regular inclusion of aspiration-related questions within national surveys would allow policymakers to assess how women evaluate opportunities over time. Monitoring migration aspirations could provide an early warning mechanism for identifying emerging gaps between educational achievement and perceived opportunity. It would also enable policymakers to evaluate whether reforms in childcare, transportation, leadership development, entrepreneurship, and workforce participation are strengthening confidence in opportunities available within Jordan.

5. Conclusion

This paper has argued that women's migration aspirations provide insight into the relationship between capability and opportunity in contemporary Jordan. While migration is often examined as a mobility outcome, the findings suggest that aspirations themselves may reveal how individuals evaluate their future prospects and the opportunities available to them.

Jordan has invested heavily in developing the capabilities of its young women. The challenge now is ensuring that institutions, labour markets, and pathways to leadership evolve at a similar pace. Development is not measured solely by the skills a society creates, but by whether individuals believe those skills can be meaningfully exercised.

In this sense, women's migration aspirations reveal something larger than migration itself. They offer insight into whether a society is succeeding in converting human potential into opportunity.

Works Cited

Adely, F. J. (2024). Working women in Jordan: Education, migration, and aspiration. University of Chicago Press.

Arab Barometer. (2024). Jordan Country Report: Arab Barometer Wave VIII. Arab Barometer. https://www.arabbarometer.org

Carling, J., & Schewel, K. (2018). Revisiting aspiration and ability in international migration. Journal of Ethnic and Migration Studies, 44(6), 945–963. https://doi.org/10.1080/1369183X.2017.1384146

Cosmano, I. (2024). Unpacking education as migration driver among young Jordanian women: A study on gender, agency, and sociocultural transformations in Jordan. Mashriq & Mahjar, 11(2), 57–92.

de Haas, H. (2021). A theory of migration: The aspirations-capabilities framework. Comparative Migration Studies, 9(1), Article 8. https://doi.org/10.1186/s40878-020-00210-4

Government of Jordan. (2022). Economic Modernisation Vision 2033. Government of Jordan. https://www.jordanvision.jo

Hausmann, R., Alsharabati, C., & Kaur, H. (2019). The Female Labor Market in Jordan: A Systematic Approach to the Exclusion Puzzle. Harvard Growth Lab. https://growthlab.cid.harvard.edu

Mahler, S. J., & Pessar, P. R. (2006). Gender matters: Ethnographers bring gender from the periphery toward the core of migration studies. International Migration Review, 40(1), 27–63.

OECD Development Centre. (2023). Social Institutions and Gender Index (SIGI): Jordan Country Profile. OECD Publishing. https://webfs.oecd.org/devsigi/SIGI%202023%20Country%20Profiles/MENA/country_profile_JOR_Jordan.pdf

PwC Middle East. (2022). Young Women, Powerful Ambitions. PwC Middle East. https://www.pwc.com/m1/en/publications/images-new/woman-in-work/mena-women-in-work-survey-2022.pdf

Turkanik, S. (2017). Business & Professional Women in Jordan: Choosing to Light a Candle, Instead of Cursing the Darkness. Konrad-Adenauer-Stiftung Jordan Office. https://www.kas.de/c/document_library/get_file?uuid=51933acd-6d9e-31ad-e6ab-4130a3dab536&groupId=252038

United Nations Department of Economic and Social Affairs (UN DESA). (2020). International migrant stock 2020: Destination and origin. Population Division. https://www.un.org/development/desa/pd/content/international-migrant-stock

UN Women Jordan. (2025). Jordan Strategic Note 2023–2027: Country Programme Results. UN Women. https://open.unwomen.org/es/country-results/JO?plan_period=2023-2027

World Bank. (2025). Jordan: Gender data portal. World Bank Gender Data Portal. https://genderdata.worldbank.org/en/economies/jordan

Evaluating the Reliability and Explainability of AI-Based Deepfake Detection Tools for High-Stakes Forensic Applications By Adya Tewari

Abstract

As generative artificial intelligence (AI) enables the production of increasingly convincing synthetic media, deepfake content poses a growing threat to legal proceedings, identity verification systems, and institutional trust. While numerous detection techniques have emerged, existing research has largely focused on improving accuracy. However, the reliability and explainability of these tools, critical requirements for forensic admissibility, remain poorly understood. This study asks: how can we ensure AI systems used for deepfake detection are reliable and explainable enough for high-stakes decisions like court cases or identity verification? We hypothesized that no single tool would demonstrate sufficient reliability across different forms of manipulation, such as face swap, face reenactment, entire face synthesis, and face editing. To address this, a composite reliability scoring framework is proposed. Accordingly, a score is generated for each sample through aggregation of four popular deepfake detection tool outputs weighted by demonstrated reliability, operating through two steps: tier assignment and weighted aggregation. Experimental results across 300 samples from the DF40 benchmark demonstrate that no single detection tool was able to provide reliable detection across all manipulation types. The proposed composite scoring framework aggregates tool outputs into a single interpretable score ranging from 1.0 to 3.0, enabling transparent evaluation of tools accessible to non-expert decision-makers. These findings suggest that current detection tools are insufficient as standalone forensic evidence, but that frequency-domain analysis and composite scoring offer a promising, explainable foundation for a forensic reliability framework that could inform evidentiary standards in legal and identity verification contexts. This framework represents a step towards future work in which composite reliability scoring could be extended to other high-stakes domains requiring explainable AI-based verification, such as border control, identity authentication, forensic journalism, and financial fraud detection.

1. Introduction

Generative artificial intelligence (AI) has rapidly advanced to produce synthetic media in recent years. This media includes text, images, audio, and video, which is becoming increasingly indistinguishable from real content. While these tools have many commercial and creative applications, they have also enabled a new category of digital deception. In an era where artificial intelligence can be leveraged to fabricate a convincing video or image of anyone saying anything, the question of whether digital evidence can be trusted in a court of law becomes an urgent forensic crisis.

Deepfake technology uses generative AI models to synthesize or manipulate facial imagery and video. This technology has advanced at a rapid pace, outstripping both detection capabilities and legal frameworks. Until recently, creating such deepfakes required significant technical expertise. However, today, low-cost and free options make the synthesis of this media

accessible to virtually anyone with an internet connection (Dalal et al.). The manipulation techniques, or synthesis techniques, vary substantially in their forensic signatures, creating variation in detection accuracy. Face swapping (FS) replaces one person's face with another's, while face reenactment (FR) is able to transfer expressions and movements from a source to a target. Entire face synthesis (EFS), as the name would suggest, generates a completely new face that doesn't actually exist, and face editing (FE) makes localized changes to an existing face (Yan et al.).

With such rapidly expanding synthesis technology, detection research has tried to keep up through four main approaches: convolutional neural network (CNN)-based classifiers, local noise pattern analysis, temporal analysis, and Fast Fourier Transform (FFT) frequency analysis. CNN-based classifiers search for pixel-level "fingerprints" that are left by generative models (Wang et al.), while local noise pattern analysis searches for inconsistencies at manipulation boundaries (Mahdian & Saic; Korshunov & Marcel). Temporal analysis technology finds motion inconsistencies across video frames, which are invisible in static images (Tong et al.). FFT frequency analysis finds spectral artifacts that persist even when visual artifacts are hidden (Frank et al.). Despite the many different methods of detection, no single method works consistently across all manipulation types. This is demonstrated by the DF40 benchmark (Yan et al.), which reveals a fundamental generalization problem in the field of deepfake detection. Beyond generalization, most of these tools operate as black-box classifiers, which produce binary outputs without transparent reasoning, which critically limits their utility in high-stakes domains like court proceedings, where interpretability of results is essential.

The consequences of not having a reliable detection method extend far beyond the scientific field. Courts are already seeing a rise in synthetic media being submitted as evidence in trials, and do not have adequately reliable tools to evaluate it. As synthesis technology improves, the distinction between what is perceived as authentic and manipulated media continues to narrow. Dalal et al. found that even the winner of the Meta Deepfake Detection Challenge was not completely successful, with some tools flagging everything as real regardless of whether it was fake or not. Additionally, existing legal framework, such as the Federal Rule of Evidence 901 (Dalal et al.), is insufficient for evidence that is possibly a deepfake. Many real cases regarding this issue have already emerged: fabricated audio was submitted in a UK child custody case, and separately, a school principal's reputation was nearly destroyed before extensive forensic analysis revealed that the recording submitted as evidence was a deepfake (Dalal et al.).

Moreover, Grossman and Grimm distinguish between acknowledged AI-generated evidence and unacknowledged AI-generated evidence, sharpening the problem. Acknowledged AI-generated evidence is evidence in which it is openly disclosed that AI tools were used to enhance or generate the image. Unacknowledged AI generated evidence occurs when one party claims a suspected deepfake is real, when in reality it is fake. Grossman and Grimm argue that detection tools in this modern age need to meet the Daubert reliability standards, which serve as the bar for scientific expert testimony, replacing the concept of accuracy with requirements for validity and reliability (Grossman & Grimm).

Furthermore, to the best of the author's knowledge, most current tools lack systematic evaluation against the Daubert reliability standards, introducing a significant gap in existing literature. Technical research often chases accuracy, while legal contexts need reliability, consistency, and outputs that are interpretable and can be used by judges and jurors, even without extensive technological background. Most existing studies do not evaluate a multi-tool ensemble through an explicit forensic reliability lens, nor propose an interpretable scoring framework for such tools.

In light of the gaps in the recent literature, this study investigates how AI deepfake detection tools can be shown to be reliable and explainable enough for the high-stakes decisions that they are involved with, such as in court and for identity verification. This study evaluates four forensic detection tools, CNNDetection, LNP, VideoMAE, and FFT spectral analysis, each operating on a fundamentally different detection principle. No single tool is hypothesized to demonstrate sufficient reliability across all four manipulation categories to serve as standalone forensic evidence. The variability in detection performance demands a trustworthiness verification for the detection tools. Consequently, in this study, a composite reliability scoring framework is proposed, in which each of the four tools is assigned a performance tier: either poor, moderate, or strong. The performance tiers are based on its category-specific detection rates, and these tiers are aggregated into a single interpretable score per sample. Unlike binary outputs, this framework is designed to produce assessments that are transparent and auditable, which non-expert decision makers can act upon without requiring technical background. This begins to address the criteria underlying the Daubert reliability standard (Grossman & Grimm), promoting transparent, tiered, and reproducible evaluations of tool reliability. If validated at scale, this framework could serve as standardized protocol for evaluating AI detection tools in legal and identity verification contexts, with patterns revealed by multi-tool evaluation simultaneously informing the development of next-generation detection methods and policies governing AI-generated evidence in court proceedings.

2. Literature Review

2.1 The Evolution of Deepfake Detection Methods

Early deepfake detection research predominantly focused on identifying artifacts left by generative adversarial networks (GANs) in the spatial domain. Mirsky and Lee provide a foundational survey of deepfake generation and detection architectures, documenting how generative deep learning has advanced to the point where distinguishing real and synthetic media has become increasingly difficult (Mirsky & Lee). Despite significant technical advances, Wang et al. ("Deepfake Detection") identify three persistent reliability challenges in the field: interpretability, transferability, and robustness. Wang et al. demonstrated that CNN-based classifiers can detect GAN-generated pixel fingerprints, but acknowledged that performance degrades substantially with post-processed images, such as images that have been blurred or compressed. Classical forensic techniques such as local noise pattern analysis offer an

alternative, exploiting inconsistencies in noise at manipulation boundaries to identify varied imagery (Mahdian & Saic; Korshunov & Marcel), however, these methods were not designed to handle the seamless blending characteristic of modern deepfakes. Frank et al. demonstrated that frequency-domain analysis via Fast Fourier Transform can detect systematic spectral artifacts introduced by generative models, which persist even with the concealment of spatial fingerprints. This forms the theoretical foundation for the FFT-based approach employed in this study. Additionally, temporal analysis methods such as VideoMAE (Tong et al.) complement spatial and spectral approaches by exploiting motion inconsistencies in subsequent video frames, which is invisible in analysis of static images. Despite these advances, Yan et al. demonstrated through the DF40 benchmark dataset that no single detection method maintains consistency in performance across different manipulation types. This reveals a fundamental generalization problem in deepfake detection that directly motivates the multi-tool evaluation framework employed in this study.

2.2 Deepfake Detection in Legal and Forensic Contexts

While technical literature has focused primarily on detection accuracy, a growing body of literature has examined the implications of synthetic media for court proceedings. Dalal et al. argued that existing detection tools are biased and unreliable, and that courts lack adequate evidentiary frameworks for adjudicating authenticity disputes involving AI-generated content. Empirical testing demonstrated that even the winner of the Meta Deepfake Detection Challenge exhibited critical performance inconsistencies, with some tools classifying virtually all content as authentic regardless of whether it was manipulated or not (Dalal et al.). Grossman and Grimm sharpened this problem by distinguishing between acknowledged and unacknowledged AI-generated evidence, arguing that detection tools must meet a Daubert-like standard of forensic reliability, replacing the ambiguity in accuracy claims with scientifically-grounded reliability requirements. This would require detection tools to meet a forensic reliability threshold before their outputs could be considered as evidence in legal proceedings.

Independent legal scholarship corroborates these concerns. Capra examines how the Advisory Committee on Evidence Rules has begun responding to the deepfake evidence problem, noting that hearings with competing forensic experts may become necessary to adjudicate authenticity disputes involving media (Capra). Livingston similarly documents real courtroom cases involving suspected deepfake evidence and analyzes how courts currently lack consistent framework for evaluating such evidence (Livingston). Delfino has proposed a revised Federal Rule of Evidence 901(c) that would shift admissibility determinations from juries to judges in cases involving media that is suspected to be synthetic, arguing that such reallocation is necessary to protect legal proceedings from the cognitive biases that deepfakes exploit (Delfino).

2.3 Explainability and the Gap Addressed by This Study

Most detection models operate as black-box classifiers, producing binary outputs lacking interpretable reasoning. This is a critical limitation in legal contexts where decisions must be

justifiable to non-expert decision-makers (Mansoor & Iliev). Wang et al.'s reliability survey reinforces this gap, finding that interpretability remains one of the three least-addressed challenges in deepfake detection research despite being essential for forensic applications (Wang et al., "Deepfake Detection"). Recent explainability-focused frameworks have begun addressing this, with Mansoor and Iliev demonstrating that combining CNN models with components of explainable AI can achieve 97% accuracy, while also providing human-interpretable rationales for classifications. However, to the best of the author's knowledge, no study has evaluated a multi-tool ensemble through an explicit forensic reliability lens or proposed an interpretable scoring framework suitable for high-stakes institutional use, such as in legal contexts, which is the gap this study directly addresses.

3. Methodology

The research methodology used in this study is illustrated in Figure 1. After collecting and preprocessing the data, the detection tools are evaluated followed by a proposed reliability scoring framework. The steps are explained below.

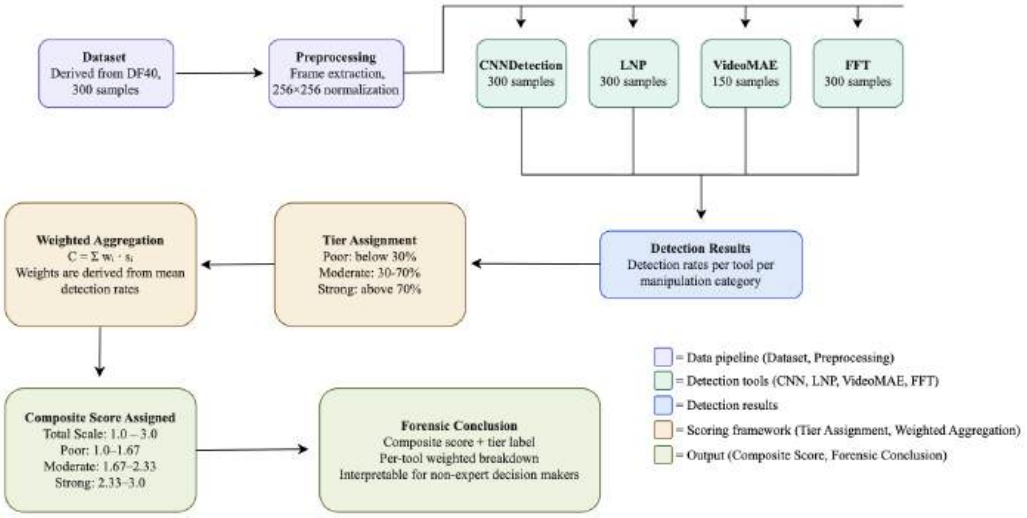


Fig 1: Research methodology flowchart

3.1 CNNDetection

The first detection tool applied was CNNDetection. CNNDetection is a CNN-based binary classifier trained to detect GAN-generated imagery (Wang et al.). The classifier works by identifying global pixel-level "fingerprints" that are left behind by generative models. The weights selected represent a standard convolutional neural network (Xception) initialized with ImageNet weights and supervised-trained on the official training partitions of face manipulation benchmarks (Yan et al.). These weights were specifically trained to handle blur and JPEG compression, as standard weights degrade when images have been post-processed with such techniques. Robustness to blur and JPEG compression is particularly critical for this dataset, as

high-quality deepfakes often undergo processing that eliminates the pixel-level artifacts that standard weights rely upon for detection. The CNNDetection tool was applied to all 300 images across all four categories, producing a binary real or fake output for each sample.

3.2 Local Noise Pattern (LNP)

The second tool was LNP (Local Noise Pattern), which is a classical forensic technique rather than a modern AI model (Mahdian & Saic). The LNP tool serves as a traditional forensic baseline. When a face is manipulated or pasted into an image, the pixel-level noise in the face region doesn't match the noise in the background or on other parts of the image. The LNP tool serves as a way to detect this "seam," estimating noise variance across different regions of an image. Korshunov and Marcel introduced LNP as a deepfake detection baseline, analyzing local patches rather than the entire image and looking for geometric or lighting inconsistencies. The LNP tool introduced by Korshunov and Marcel was applied to all 300 extracted frames across the four categories of fakes.

3.3 VideoMAE

VideoMAE is a masked autoencoder architecture for self-supervised video representation learning (Tong et al.), which identifies unnatural motion across sequential frames. It identifies these inconsistencies in motion, which are artifacts invisible in static analysis. Since VideoMAE requires sequential video frames in order to analyze the temporal motion, it could only be applied to the FS and the FR samples. EFS and FE are static images with no temporal dimension, which makes running VideoMAE on them not possible. As a result, VideoMAE was applied to 150 samples total, across two categories.

3.4 Fast Fourier Transform (FFT) Spectral Analysis

The Fast Fourier Transform (FFT) converts pixel intensity data into a representation of the frequency-domain of the image (Frank et al.). Instead of looking at what a face looks like, FFT examines the mathematical construction of the image. Authentic images contain fine dermal textures on facial regions which produce random, high-frequency energy patterns in the image. In contrast, AI-generated or modified images tend to be built on mathematical grids, which suppresses this natural, high-frequency variance through blending and interpolation. The Global Mean Energy (GME) captures the average high-frequency content of an image, which is higher in authentic photos and lower in generated or manipulated ones (Frank et al.). The GME was computed for each image's frequency map as the primary detection feature. This was applied to the extracted frames from all four categories. In contrast to VideoMAE, FFT does not require a temporal dimension, allowing it to work on static images.

3.5 Threshold Selection

For FFT, a detection threshold is necessary to classify images as real or fake. This threshold was calculated empirically by examining the energy gap between the authentic control

group and the detectable forgery categories. The authentic control group, consisting of 75 FaceForensics++ images (Rossler et al.), yielded a mean GME of 144.39. The generative category with the highest energy was the Face Editing (FE) category, which yielded a mean of 129.33, a gap of 15.06 GME units between the control group and the samples. Based on these means, a threshold of 135 was selected at the approximate midpoint of this gap, providing a 9.39-point buffer to minimize false positives. It also remained sufficiently above the mean yielded by the FE category, enabling reliable detection. Samples that yielded a GME below 135.0 were classified as manipulated.

3.6 Proposed Composite Reliability Scoring Framework

Given that no single tool demonstrated consistent and reliable detection across all four manipulation categories, this study proposes a composite reliability scoring framework as a forensic evaluation tool. A composite score is generated for each sample through aggregation of tool outputs, weighted by each tool's demonstrated reliability for the relevant type of manipulation. This framework begins to address the criteria underlying the Daubert reliability standard identified by Grossman and Grimm, replacing blanket accuracy claims with a transparent and reproducible forensic assessment, tiered based on performance. The framework operates using two sequential steps: tier assignment and weighted aggregation.

3.7 Tier Assignment

Each tool is assigned a reliability tier based on its detection performance across categories, classified as poor, moderate, or strong. Detection rates below 30% are classified poor, indicating that the tool lacks forensic utility for that manipulation type. Detection rates between 30% and 70% are classified as moderate, indicating partial forensic utility that should be then corroborated by secondary analysis. Detection rates above 70% constitute a strong tier, indicating sufficient forensic signal to meaningfully contribute to a detection conclusion. FFT achieves strong classification on FR, EFS, and FE, while VideoMAE achieves moderate classification on FS and FR, and CNNDetection and LNP achieve predominantly poor classifications across most categories. Tier assignments for all tools and categories are summarized in Table 1.

Tool	**EFS**	**FE**	**FR**	**FS**
CNNDetection	Poor	Poor	Moderate	Poor
LNP	Poor	Moderate	Moderate	Poor
VideoMAE	*N/A*	*N/A*	Moderate	Moderate
FFT	Strong	Strong	Strong	Poor

Table 1: Reliability tier assignments by tool and manipulation category. Poor <30% · Moderate 30–70% · Strong >70% · N/A = tool not applicable to static images

Tiers are then converted to numeric scores, with poor equaling 1, moderate equaling 2, and strong equaling 3. These numeric scores are then aggregated to produce the final score using a weighted average, as shown in Equation 1.

$$C = \Sigma w_i \cdot s_i \quad \text{(Equation 1)}$$

where $w_i = d_i / \Sigma d_i$ (normalized weight of tool i), $s_i \in \{1, 2, 3\}$ (tier score of tool i), and d_i = mean detection rate of tool i across tested categories.

3.8 Weighted Aggregation

The weights are derived from each tool's mean detection rate across its tested categories, and each tool's mean detection rate is divided by the sum of all tools' mean detection rates, normalizing the weights to all sum to 1.0. This gives higher influence to tools that demonstrate greater reliability. In this study, the derived weights were: CNNDetection at 6.9%, LNP at 16%, VideoMAE at 30.8%, and FFT at 46.2%. Weighting tools by demonstrated reliability ensures that tools with consistently poor performance do not contribute significantly to the composite score, allowing tools with consistently strong performances to be the ones driving the conclusion.

3.9 Composite Score and Forensic Conclusion

The composite score is computed as a weighted average of numeric tier scores assigned to each tool, where poor equates to 1, moderate to 2, and strong to 3. The resulting composite scale spans 1.0 to 3.0, representing the theoretical minimum, all tools rated poor, and maximum, which is all tools rated strong. Since the scale spans two units, the three classification ranges are divided into equal segments of 0.67 units each, yielding boundaries at 1.67 for the poor-to-moderate threshold and 2.33 for the moderate-to-strong threshold. This equal-interval partitioning ensures that no classification range is privileged over another by span, producing a transparent and reproducible scoring structure that non-expert decision-makers can interpret without technical background. This makes the basis for a detection conclusion transparent, even for those lacking a technical background. The framework constitutes the original contribution of this study, discussed further in Section 6.

4. Data collection and experimentation

4.1 Dataset

The DF40 benchmark dataset has been used in this work, which is a comprehensive, diverse benchmark representing 40 distinct deepfake generation methods (Yan et al.). From this dataset, a balanced 300-sample representative subset was constructed, with samples drawn equally from four manipulation categories. The categories are: face swap (FS), which replaces one person's identity with another's; face reenactment (FR), which transfers expressions and movements from a source to a target; entire face synthesis (EFS), which generates a completely new face that does not exist in reality; and face editing (FE), which makes localized modifications to specific attributes on an existing face. As the DF40 dataset does not include

authentic images, 75 authentic images from FaceForensics++ (Rossler et al.) were reserved specifically as a control group for the FFT spectral analysis, where an authentic baseline was needed in order to calibrate the detection threshold.

4.2 Preprocessing

The FS and the FR categories exist as video files in the DF40 dataset, while EFS and FE exist as static images. For the FS and the FR categories, a single frame was randomly extracted from each video sample through a randomized selection script, ensuring no systematic bias in frame choices. For CNNDetection, LNP, and FFT, all four categories were treated consistently as static images. VideoMAE was directly applied to the original video files of the FS and FR categories, as it requires sequential frames for temporal analysis rather than static images. Every image, regardless of category, was normalized to 256×256 pixels, consistent with standard preprocessing protocols in deepfake detection research (Frank et al.).

4.3 Data Processing for the tools

CNNDetection and LNP were applied to all 300 images across all four categories, with no additional preprocessing beyond the standard 256×256 normalization, producing a binary real or fake output for each sample. VideoMAE was applied to the original video files of the FS and FR categories only, as temporal analysis requires sequential frames, which are lacking in static images. A total of 150 samples were evaluated across these two categories. FFT was applied to the extracted frames from all four categories, totaling 300 samples. The 75 FaceForensics++ (Rossler et al.) control images were processed identically, normalized to 256×256, to establish the authentic GME baseline for threshold calibration.

5. Results

This section presents the detection rates achieved by each tool across the four manipulation categories. Full results are summarized in Table 2, with a comparative visualization in Figure 3.

Category	Standard CNN	LNP Forensic	VideoMAE	FFT
EFS	1.3%	4.0%	N/A	99.0%
FE	4.0%	38.7%	N/A	79.0%
FR	38.7%	40.0%	40.0%	100.0%
FS	0.7%	21.3%	60.0%	22.0%

Table 2: Detection rates by tool and manipulation category. N/A = VideoMAE not applicable to static images.

5.1 CNNDetection Results

CNNDetection demonstrated near-zero detection rates across the four categories: 1.3% for EFS, 4% for FE, 38.7% for FR, and 0.7% for FS. The near-zero detection rates suggest that the samples in this dataset underwent sufficient post-processing to eliminate the GAN fingerprints at the pixel-level that CNNDetection relies upon. The one exception was the FR category, which was notably higher than the other three but still insufficient for forensic reliability. Notably, however, FR is the only category where spatial fingerprints show meaningful signals. These results suggest that global texture analysis is ineffective against high-quality processed deepfakes.

5.2 Local Noise Pattern Results

LNP displayed moderate performance on the FE and FR categories, with low performance on EFS and FS categories. LNP detected 4% of the EFS fakes, 38.7% of FE, 38.7% of FR fakes, and 21.3% of FS fakes. The higher detection on the FE and FR categories suggest that inconsistencies in local noise are more detectable in localized edits and reenactments rather than fully synthesized faces. Corroborating this, lower detection on EFS and FS suggests that high-fidelity swaps and fully synthesized faces produce local patches that are consistent, therefore making it difficult for LNP to distinguish them from authentic images.

5.3 VideoMAE Results

VideoMAE was applied on the FR and FS categories and demonstrated moderate performance rates on those categories, achieving rates of 40% and 60% respectively. The 60% on the FS category was the single highest detection rate achieved by any of the first three tools. The high detection rate in the FS category confirms that face swapping introduces temporal motion artifacts that are detectable. FR at 40% shows that reenactment does introduce some temporal inconsistencies, however these are fewer than swapping. This is likely as a result of reenactment preserving more of the subject's natural motion patterns. VideoMAE cannot be applied to EFS or FE deepfakes as they are static images with no temporal dimension. This limits the forensic utility of VideoMAE to video-based manipulation types only.

5.4 FFT Spectral Analysis Results

Table 3 presents the Mean Energy Audit, which reports the mean Global Mean Energy (GME) and corresponding FFT detection rate for each manipulation category. A threshold of 135.0 GME was applied, with samples yielding a mean energy below this classified as manipulated.

Category	Mean GME	vs. threshold (135.0)	FFT detection rate
Real (control)	144.39	+9.39 above	3.3% FP rate
EFS	103.16	−31.84 below	99.0%
FE	129.33	−5.67 below	79.0%
FR	104.39	−30.61 below	100.0%
FS	142.34	+7.34 above	22.0%

Table 3: FFT Mean Energy Audit. Detection threshold: 135.0 GME, Samples below threshold classified as manipulated

FFT achieved the highest detection rates of any tool across three categories: 100% for FR, 99% for EFS, and 79% for FE. These categories all yielded mean GME values well below the 135.0 threshold: FR was 104.39, EFS was 103.16, and FE was 129.33. FS only achieved a 22% detection, with a mean GME of 142.34. This was nearly identical to the mean GME of the control group, which was 144.39. This is consistent with the DF40's face-swapping methods, which include texture preserving techniques such as SimSwap and BlendFace (Yan et al.), which transplant existing skin texture rather than synthesizing new pixels. This preserves the source image's high-frequency sensor noise. The authentic control group yielded a 3.3% false positive rate, which meant that 3.3% of real images were incorrectly classified as manipulated. The 9.39-point buffer above the threshold accounts for this low false positive rate.

5.5 Comparative Summary

Figure 3 presents a comparative visualization of detection rates across all four tools and categories. On three out of the four categories, FFT outperforms the other three tools by a significant margin. No single tool demonstrated reliable detection across all four manipulation categories. While FFT demonstrated the strongest overall performance, it failed on FS. VideoMAE showed relatively strong performance on FS and FR, but it cannot be applied to static images, limiting its scope. CNNDetection and LNP consistently showed low performance across most categories. However, VideoMAE's strongest category was FS at 60%, which is the same category where FFT only had a 22% detection rate. The results support the hypothesis that no single tool is sufficient for standalone forensic use, motivating the composite reliability scoring framework proposed in Section 3.6.

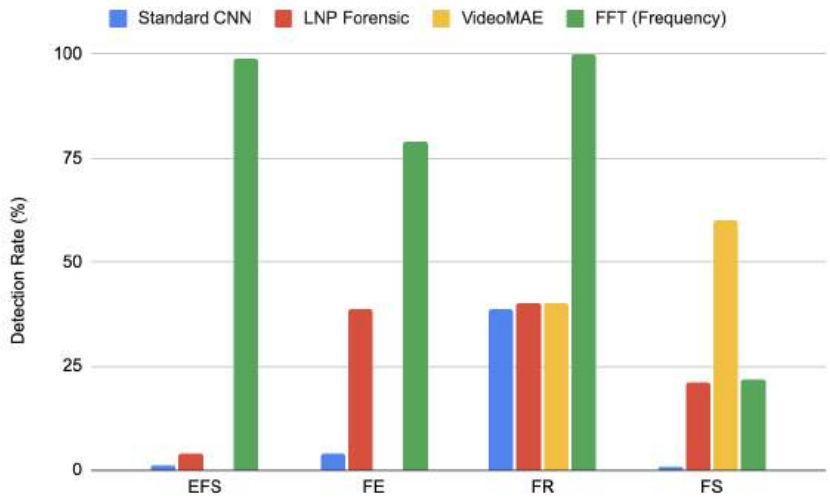


Fig 3: Comparative deepfake detection rates by tool and manipulation category

5.6 Reliability Scoring Framework Performance

As a worked example, the framework was applied to the FS category, which represents the most forensically challenging case in this study. The following tiers were assigned for FS: CNNDetection poor (0.7%), LNP poor (21.3%), VideoMAE moderate (60%), and FFT poor (22%). With the weighted calculation, the composite score was 1.307 out of 3. The composite scale of 1.0 to 3.0 is divided into three equal segments of 0.67 units each, with boundaries at 1.67 (poor/moderate) and 2.33 (moderate/strong), ensuring that each classification range spans an equal proportion of the possible score range. A composite score of 1.307 falls within the poor range, as it is below 1.67, which is the threshold for moderate. This reflects the lack of detection of the FS samples by three out of the four tools. However, the score of 1.307 is higher than the minimum score of 1.0, indicating that VideoMAE's temporal signal provides a partially reliable forensic indicator that would be invisible under single-tool analysis. This result accurately reflects the forensic reality of FS detection found in this study. The composite score confirms the category as predominantly difficult to detect, while preserving and quantifying the partial contribution of VideoMAE in a way that single-tool analysis cannot. This provides a more complete and honest basis for a forensic conclusion for high-stakes detection scenarios. The scoring process is summarized visually in Figure 2.

	Worked Example: Face Swap (FS) Composite Scoring			
Tier Assignment	**CNNDetection**	**LNP**	**VideoMAE**	**FFT**
	Detection rate: 0.7% → Poor	Detection rate: 21.3% → Poor	Detection rate: 60% → Moderate	Detection rate: 22% → Poor
Weights	Weight: 6.9%	Weight: 16.0%	Weight: 30.8%	Weight: 46.2%

Calculation	1 × 0.069 = 0.069	1 × 0.160 = 0.160	2 × 0.308 = 0.616	1 × 0.462 = 0.462
	0.069 + 0.160 + 0.616 + 0.462 = **1.307 / 3.0 — Classification: Poor** (below 1.67 threshold) *VideoMAE's partial signal (0.616) preserved, invisible under single-tool analysis*			

Fig 2: Worked example of the composite reliability scoring framework applied to the Face Swap (FS) category.

6. Discussion

No single tool demonstrated sufficient reliability across all four manipulation categories, confirming the hypothesis. This finding directly addresses the central concern of this study, that current detection tools lack the explainability and the reliability required for high-stakes forensic use.

6.1 The Limits of Spatial and Classical Detection

Both the CNNDetection and LNP tools operate in the spatial domain. While CNN operates at the global level and LNP operates at the local level, both showed consistently low performance across most categories. This suggests that spatial analysis alone is insufficient for detecting modern high-quality deepfakes. This is consistent with Wang et al., who note that detection degrades when compression and blurring are applied to images. LNP's moderate performance on the FE and FR categories suggests that local noise inconsistencies are harder to conceal in localized edits, rather than in global synthesis. In contrast, fully synthesized faces (EFS) and texture-preserving swaps (FS) produce patches that defeat noise-based analysis. In legal contexts, tools that fail on fully synthesized faces are insufficient as standalone forensic instruments, as courts increasingly confront fully synthesized media as potential evidence (Dalal et al.).

6.2 The Role of Temporal Analysis

Of the first three tools, VideoMAE achieved the strongest performance. The 60% detection rate in the FS category by VideoMAE confirms that face swapping introduces more detectable temporal artifacts than reenactment. This is likely because swapping disrupts motion continuity more severely than reenactment. The 40% detection rate in the FR categories reveals that reenactment still introduces some temporal inconsistencies detectable by VideoMAE, despite preserving more natural motion. However, 40% remains insufficient for forensic reliability. Additionally, VideoMAE's inability to analyze static images limits its forensic utility, as evidence in legal contexts frequently takes the form of still images. This would mean that a video-only tool would be forensically incomplete on evidence such as screenshots, photographs, and identity documents. The complementary pattern between VideoMAE and FFT, where VideoMAE is strongest in the category where FFT is weakest (FS), suggests that a scoring framework approach is necessary for comprehensive forensic coverage.

6.3 Spectral Analysis as a Forensic Tool

FFT showed near-perfect detection on FR, EFS, and FE, which demonstrates that frequency-domain analysis is a powerful forensic tool against most types of generative manipulation. Additionally, the interpretable GME score, being a concrete numeric output, offers transparency over how images are classified, a transparency advantage often absent in CNN-based classifiers, which produce binary outputs without interpretable reasoning. Generative synthesis and blending operations inherently suppress high-frequency variance, which means the spectral signature of manipulation is mathematically predictable. Since the spectral signature is predictable regardless of visual quality, this makes FFT robust to the post-processing that defeats spatial analysis. The 22% detection in the FS category is a fundamental forensic limitation for the FFT tool, as texture-preserving swap methods preserve the source image's spectral properties. This makes them almost indistinguishable from authentic images under mean energy analysis, suggesting that FS detection requires secondary analysis beyond GME. Specifically in legal contexts, the GME score provides a critical interpretable numeric basis for a detection conclusion. A non-expert decision-maker can interpret a GME score against an authentic baseline to classify an image as spectrally anomalous, without requiring technical expertise.

6.4 Implications for Forensic Reliability and Legal Policy

Taken together, these results demonstrate that no current tool meets the reliability and explainability threshold required for high-stakes forensic use as a standalone instrument. This directly supports the concern raised by Dalal et al. and Grossman and Grimm that existing detection tools are insufficient for legal admissibility. The composite scoring framework demonstrates this concretely, as the FS category had poor detection rates for 3 out of the 4 tools. The composite score of 1.307 for the FS category accurately reflected its forensic difficulty, while also preserving VideoMAE's partial detection accuracy. This preserves the performance of all tools in a way that single-tool analysis cannot, and having a tiered, numeric output gives a decision-maker a more honest and complete forensic picture. VideoMAE and FFT displayed a complementary performance pattern, suggesting that a composite scoring framework would provide a more transparent and complete basis for evaluating forensic reliability than any individual method would. The proposed framework begins to address the Daubert reliability standard (Grossman & Grimm) by producing tiered, transparent, reproducible outputs. The framework proposed requires empirical validation on independent datasets before it can be considered forensically admissible. Additionally, the train/test circularity in the FFT threshold calibration represents a known limitation, addressed further in Section 7. Courts and identity verification systems adopting AI detection tools should require documented category-specific performance metrics, rather than aggregate accuracy scores. The results demonstrate that aggregate accuracy scores mask critical failures at the category level, which directly impact forensic reliability.

7. Limitations

While the findings demonstrate meaningful patterns in deepfake detection performance, limitations must be acknowledged regarding the generalizability and the rigor of the results.

7.1 Dataset Scope

While balanced, the sample size of 300 is relatively small compared to larger benchmark evaluations in the field. All samples were drawn from the DF40 benchmark (Yan et al.), which meant that results reflect the specific generation methods represented in that dataset. Detection performance of tools may differ on fakes which are produced by methods omitted from the DF40 dataset.

The GME threshold of 135.0 for detection was derived from and validated on the same 375-sample dataset, consisting of 300 fake samples and 75 authentic control images. This train/test circularity means that the threshold may have been optimized for this specific dataset, making generalization on unseen data questionable. The composite reliability scoring framework proposed in Section 3.6 is conceptual, and it has not been empirically validated. The tier assignments of poor, moderate, and strong are based solely on accuracy of detection observed in this study, and have not been validated against established forensic standards. Until such validation is completed, the framework should be regarded as a proposed protocol rather than an established forensic tool. Additionally, detection rates were used as the primary evaluation metric; future work should incorporate precision, recall, and F1 scores to enable more rigorous comparison with existing literature.

7.2 Implementation Notes

In order to produce usable results, the implementation process required several engineering solutions. All code blocks were run in Google Colab, connected to Google Drive. A significant lag between Google Drive and the Colab environment prevented the tools from recognizing many uploaded files, which was resolved through a flush and remount protocol that forced the environment to recognize the 300 samples. The standard research code also failed to load checkpoint weights for CNNDetection. The state_dict loading logic was manually rewritten in order to extract only the required weights needed for detection from the checkpoint files. Additionally, the tools saved results in a format containing extraneous punctuation, requiring a brute force extraction script to recover the raw decimal scores from the output files. Results were saved both as a CSV file and as a graph.

8. Conclusion

This study aimed to investigate how current deepfake detection tools can be used in high-stakes scenarios, such as legal proceedings and identity verification. Four deepfake detection tools were evaluated across four manipulation categories to assess both their reliability and explainability for these high-stakes forensic applications. No single tool demonstrated sufficient reliability across all four categories to serve as standalone forensic evidence, confirming the central hypothesis and highlighting the fundamental limitations of single-tool

approaches in forensic contexts. CNNDetection and LNP displayed consistently low performance, suggesting that spatial analysis is insufficient against high-quality processed deepfakes. VideoMAE showed moderate performance on video categories, but was fundamentally limited as it cannot be applied to static images. FFT spectral analysis achieved strong detection on three of four categories, but exhibited a significant blind spot on face swaps that preserve the original texture. The complementary performance pattern between VideoMAE and FFT, where each tool is strong where the other is weak, provides empirical motivation for the composite reliability scoring framework proposed in this study. The framework proposed represents the original contribution of the study, designed as a step towards meeting the Daubert reliability standard identified by Grossman and Grimm; it aims to produce transparent, tiered, and reproducible evaluations of tool reliability, attempting to bridge the gap between technical detection capability and legal admissibility standards. The framework proposed requires empirical validation on an independent dataset before it can be considered forensically admissible. The train/test circularity in the FFT threshold calibration and the reliance on detection rates as the sole evaluation metric are known limitations that future work should address.

As deepfake technology continues to advance, the gap between generation capability and detection reliability will only widen unless detection tools are evaluated through an overarching forensic lens, rather than a purely accuracy-based lens. The findings of this study suggest that multi-tool forensic protocols, with transparent and interpretable outputs, are necessary for courts and identity verification systems. As courts begin to confront AI-generated evidence without adequate frameworks to evaluate it, the need for standardized, forensically validated detection protocols becomes increasingly urgent. Future work should focus on held-out threshold validation, expansion to larger and more diverse datasets, incorporation of standard binary classification metrics, and empirical testing of the proposed scoring framework against established forensic and legal admissibility standards. The ultimate goal is a forensic detection protocol that is robust enough to meet the demands of a legal system that is increasingly confronted with synthetic media.

Acknowledgements

I would like to acknowledge and express my gratitude towards Indigo Research and my mentor, Dr. Moshfeka Rahman, for their guidance and support in writing this paper. I attest that the ideas, graphics, and writing in this paper are entirely my own.

Works Cited

Capra, Daniel J. "Deepfakes Reach the Advisory Committee on Evidence Rules." Fordham Law Review, vol. 92, 2024, pp. 2491–2506. fordhamlawreview.org/wp-content/uploads/2024/05/Vol.-92_19_Capra-2491-2506.pdf.

Dalal, Abhishek, et al. "Deepfakes in Court: How Judges Can Proactively Manage Alleged AI-Generated Material in National Security Cases." University of Chicago Legal Forum, vol. 2024, 2025, art. 3, chicagounbound.uchicago.edu/uclf/vol2024/iss1/3.

Delfino, Rebecca A. "Deepfakes on Trial 2.0: A Revised Proposal for a New Federal Rule of Evidence to Mitigate Deepfake Deceptions in Court." Loyola Law School Los Angeles Legal Studies Research Paper No. 2025-10, 2025. ssrn.com/abstract=5188767.

Frank, Joel, et al. "Leveraging Frequency Analysis for Deep Fake Image Recognition." Proceedings of the 37th International Conference on Machine Learning (ICML), 2020, arxiv.org/abs/2012.09059.

Grossman, Maura R., and Paul W. Grimm. "Judicial Approaches to Acknowledged and Unacknowledged AI-Generated Evidence." Science and Technology Law Review, vol. 26, no. 2, 2025, doi.org/10.52214/stlr.v26i2.13890.

Korshunov, Pavel, and Sébastien Marcel. "DeepFakes: A New Threat to Face Recognition? Assessment and Detection." arXiv, 2018, arxiv.org/abs/1812.08685.

Livingston, Colin. "Deepfakes in the Courtroom: Challenges in Authenticating Evidence and Jury Evaluation." University of Baltimore Law Review, 2025. ubaltlawreview.com/2025/12/01/deepfakes-in-the-courtroom-challenges-in-authenticating-evidence-and-jury-evaluation/.

Mahdian, Babak, and Stanislav Saic. "Using Noise Inconsistencies for Blind Image Forensics." Image and Vision Computing, vol. 27, no. 10, 2009, pp. 1497–1503, doi.org/10.1016/j.imavis.2009.02.001.

Mansoor, Norah, and Alexander Iliev. "Explainable Deepfake Detection: A Multi-Model Framework with Human-Interpretable Rationales for Legal Investigation Purposes." Machine Learning with Applications, vol. 23, 2025, art. 100819, doi.org/10.1016/j.mlwa.2025.100702.

Mirsky, Yisroel, and Wenke Lee. "The Creation and Detection of Deepfakes: A Survey." ACM Computing Surveys, vol. 54, no. 1, 2021, art. 7, doi.org/10.1145/3425780.

Rossler, Andreas, et al. "FaceForensics++: Learning to Detect Manipulated Facial Images." Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1–11, openaccess.thecvf.com/content_ICCV_2019/html/Rossler_FaceForensics_Learning_to_Detect_Manipulated_Facial_Images_ICCV_2019_paper.html.

Tong, Zhan, et al. "VideoMAE: Masked Autoencoders Are Data-Efficient Learners for Self-Supervised Video Pre-Training." Advances in Neural Information Processing Systems, vol. 35, 2022, pp. 10078–10093, arxiv.org/abs/2203.12602.

Wang, Sheng-Yu, et al. "CNN-Generated Images Are Surprisingly Easy to Spot...for Now." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 8692–8701, peterwang512.github.io/CNNDetection/.

Wang, Tianyi, et al. "Deepfake Detection: A Comprehensive Survey from the Reliability Perspective." ACM Computing Surveys, vol. 57, no. 3, 2024, art. 58, doi.org/10.1145/3699710.

Yan, Zhiyuan, et al. "DF40: Toward Next-Generation Deepfake Detection." Advances in Neural Information Processing Systems, vol. 37, 2024, arxiv.org/abs/2406.13495.

Comparative Study of Theme-Based and Traditional Instruction for Secondary Physics
By Haricharan Vemuri

Abstract

This study focuses on the impact of theme-based learning versus traditional instructional methods on cognitive demand and conceptual understanding in secondary physics. Although theme-based education is frequently associated with increased engagement, limited research has examined its effects on perceived mental effort and conceptual processing within individuals. In this investigation, 124 secondary students viewed both traditional and theme-based instructional videos on projectile motion and subsequently reported their perceived understanding, engagement, and cognitive load while completing conceptual questions to objectively test their understanding. Results indicated that theme-based instruction was associated with higher engagement, lower perceived cognitive load, and improved conceptual understanding versus traditional instruction. These results suggest that contextualized instruction may support conceptual processing by reducing extraneous cognitive demands and increasing learner engagement.

Keywords

Physics; Physics Education; Cognitive Load Theory; Theme-Based Learning; Instructional Design

1.0 Introduction

STEM-based education research, especially that in Physics, looks into how different lesson formats support students' understanding of various concepts (Redish, 1994). In recent years, this research has been focused on context-based instruction, where concepts are introduced through real-world situations or topics, such as Basketball or Video Games, that students may already find familiar (Bennett et al., 2003). This research reported that these lessons can improve students' interest, motivation, and attitudes toward science (Bennett et al., 2003; Ültay and Çalık, 2012). These approaches can make physics feel more concrete and helps connect difficult topics to contexts with which the student is already familiar; however, student engagement alone does not fully explain how students process and understand new information.

To expand, much of the current research focuses on classroom and group outcomes—such as motivation, achievement, and general attitudes toward science—rather than how individual students internalize specific concepts during lessons and during learning (Osborne et al., 2003). As a result, it is still uncertain on whether context-based instruction reduces a student's cognitive load, improves conceptual understanding, or helps with knowledge accuracy. This is important because students may interpret physics concepts differently, and the same thematic framing may not support every student in the same way (Bennett et al., 2003; Ültay and Çalık, 2012). That is, a theme used to teach a physics topic may greatly reduce the extraneous cognitive load of one student, but may have a less of an impact on another student's

extraneous cognitive load. These concerns are also reflected in the general research on STEM education. Engagement is often the focus, but there is little evidence of how contextualized approaches are adapted to the different and individual interests of students (McLure et al., 2022).

One framework that is particularly useful to analyze this issue is the Cognitive Load Theory. Sweller's theory explains that working memory has limits, and that instructional design can affect the amount of unnecessary mental effort students exert while learning (Sweller, 1988). For example, research in science education shows that the complexity of diagrams, language, and presentation style can have an effect on cognitive effort that students exert when learning a particular topic (Mayer and Moreno, 2003; Paas et al., 2003).

Together, these areas of research show a gap: studies on theme- or context-based instruction often focus on engagement and motivation, while research on cognitive load usually focuses on presentation and design features. In addition, a relatively small number of studies directly compare theme-based and traditional instruction while measuring both perceived (extraneous) cognitive load and objective conceptual understanding within the same group of students. Because of this, the relationship between the instruction type and mental processing in physics remains unclear.

This study compares the learning of projectile motion, a fundamental concept within Physics, by secondary students (Grades 9–12) with and without thematic framing. Students viewed both traditional and theme-based instructional videos covering the same core content. At the individual level, the study measured perceived cognitive load and conceptual understanding to explore whether theme-based instruction alleviates the cognitive strain experienced by students and whether it supports conceptual learning beyond engagement.

2.0 Literature Review

Prior research shows that students often respond better to science lessons when the material is connected to familiar contexts. In science education, contextualized instruction has been linked to higher student interest, motivation, and perceived relevance (Bennett et al., 2003; Osborne et al., 2003; Ültay and Çalık, 2012). For example, when abstract scientific ideas are presented through real-world situations, students find the material easier to connect with and more meaningful. Reviews of science education and integrated STEM work also show that contextualized and theme-based approaches are commonly used to improve engagement in the classroom (McLure et al., 2022; Hiwatig et al., 2024). However, much of this research focuses on students' motivation and attitudes, rather than specifically on how students process and understand specific concepts during learning (Bennett et al., 2003; McLure et al., 2022),

Cognitive Load Theory helps us to understand how instructional design affects learning in a different way. Sweller states that learning is limited by working memory, which means that the way information is presented can either support or interfere with current understanding (Sweller, 1988). Cognitive Load Theory separates intrinsic cognitive load, which comes from the difficulty of the material itself, from extraneous cognitive load, which comes from how the material is explained or organized. Effective instruction should reduce unnecessary mental effort

so that students can focus more on understanding the concept. Self-reported mental effort is often used in education research as a practical way to measure cognitive load and understand how difficult students feel a task is (Paas et al., 2003; Boone and Boone, 2012). Because of this, two lessons that teach the same content may still lead to different learning outcomes if one is easier for students to process.

On this note, research on multimedia learning and instructional design also shows that presentation style can affect both cognitive load and learning. Factors such as diagram complexity, wording, pacing, and the combination of visuals and text can influence how students understand material (Mayer and Moreno, 2003; Paas et al., 2003). Explanations that are poorly organized can increase extraneous cognitive load and make learning harder. On the other hand, clearer instructional materials can reduce unnecessary mental effort and improve understanding. Most of this research, however, focuses on layout, redundancy, and visual design rather than on thematic or narrative framing (Mayer and Moreno, 2003). This leaves uncertainty about whether familiar themes actually actively decrease cognitive demand or make lessons feel more interesting.

Together, these areas of research show a gap that this study addresses, by comparing theme-based and traditional instruction on projectile motion and measuring cognitive load and conceptual understanding at the individual student level. Context- and theme-based instruction are often studied in relation to engagement and motivation, while Cognitive Load Theory research focuses more on how instructional design affects mental effort (Sweller, 1988; Mayer and Moreno, 2003; Paas et al., 2003). Fewer studies directly compare theme-based and traditional instruction while measuring both perceived cognitive load and objective conceptual understanding in the same group of students. In addition, many studies look at classroom-level outcomes instead of how individual students experience and process specific concepts.

3.0 Methods

This section explains the methodology of the study, including the instructional conditions, participant selection, survey design, and statistical methods used to compare theme-based to traditional instruction.

3.1 Hypotheses

Based on the identified gap, the following two hypotheses are proposed:

H1: Theme-based instruction will significantly reduce perceived cognitive load compared to traditional instruction.

H2: Theme-based instruction will significantly improve conceptual understanding scores compared to traditional instruction.

3.2 Study Design

This study employed a within-subject experimental design to compare the effects of theme-based and traditional instruction on students' cognitive load and conceptual understanding

of projectile motion. The study used anonymous self-report data collected from participants on a voluntary basis. Identifying information was not collected.

Traditional videos were defined as the classical lecture-style or formula-first explanations that relied mainly on derivations, worked examples, step-by-step algebra, and a typical classroom or tutoring format. Theme-based videos were defined as videos built around a contextualized theme (e.g., sports, games, or narrative framing) while teaching the same physics concepts.

3.3 Survey Instrument

The survey included a theme-based video (around six minutes) and a traditional video (around nine minutes) covering the introductory concepts of projectile motion such as definition of projectile motion, related concepts, formulas and solved examples. The videos were of slightly different lengths but covered the same basic concepts of projectile motion; the small differences were in the speed of instruction, not in the concepts covered. The theme-based video was application- and example-heavy and embedded in a sports context (e.g. basketball). The traditional video was more formula-heavy and contained basic derivations of some of the mentioned formulas.

3.4 Sampling and Sample Population

The survey participants included students from various high schools worldwide, including Canada, India, and the USA. They were recruited primarily through their schools, with a few participants reached through social media platforms such as LinkedIn and Instagram; in total, there were initially 138 participants in this survey, aged 13 to 18, across grades 9 to 12, and 124 participants were considered as part of the data. To start, participants reported their prior familiarity with projectile motion before viewing the instructional videos. After watching each video, students completed survey questions such as “How well do you understand the concept after Video 2?”, “How mentally effortful was Video 2 to follow?”, and “How engaging was Video 2?”. Following these Likert questions, two knowledge (or assessment) questions measured conceptual understanding of the concepts presented in the video, and an open-response input following that allowed students to describe what made the video easy or hard to understand; this full sequence of questions was repeated for the second video, with survey versions counterbalanced so that some students, roughly half, viewed the traditional video first and other half viewed the theme-based video first.

3.5 Analytical Framework

The variables measured in this survey were cognitive load, engagement, and conceptual understanding of projectile motion. Cognitive load was measured on a 7-point Likert scale, with participants rating how mentally demanding the video was to follow, ranging from 1 (“Very hard to follow”) to 7 (“Very easy to follow”). 7-point Likert-type responses were used to capture students’ immediate perceptions for engagement and ease of following as well (Boone and

Boone, 2012; Jebb et al., 2021). A test of the students' conceptual understanding was conducted using two multiple choice questions on the key principles of projectile motion that both videos taught in an equal manner. These were multiple-choice questions that allowed students to choose from a fixed set of responses, for analytical purposes of being objective to analyze later on. Only responses from participants who had viewed both videos and answered the questions in full were included, controlling for incomplete exposure and measurement bias. The responses that were incomplete or indicated that a video had not been fully or partially watched were also excluded from consideration in the final analysis.

4.0 Results and Discussion

This section reports the results of the comparative analyses between the theme-based and traditional instructional conditions across measures of engagement, cognitive load, confidence, and conceptual understanding.

4.1 Participant Demographics

A total of 138 participants completed the survey. Participants were divided into two counterbalanced conditions: one group viewed the theme-based instructional video first (n = 67), while the other group viewed the traditional instructional video first (n = 71). However, some participants did not watch both videos, reducing the effective sample size of these groups that viewed the theme-based instructional video first to n = 61 and the group that viewed the traditional instructional video first to n = 63. The final sample that was analyzed consisted of 124 participants.

4.2 Instructional Preference

Participants were asked to indicate which instructional video helped them learn the concept more effectively after viewing both videos.

In the condition where the theme-based video was presented first (n = 61), 52 participants reported that the theme-based video helped them learn better, compared to 4 who preferred the traditional video and 5 who reported no difference.

Conversely, in the condition where the traditional video was presented first (n = 63), 52 participants reported that the theme-based video helped them learn better, compared to 4 who preferred the traditional video and 7 who reported no difference.

4.3 Participant Confidence

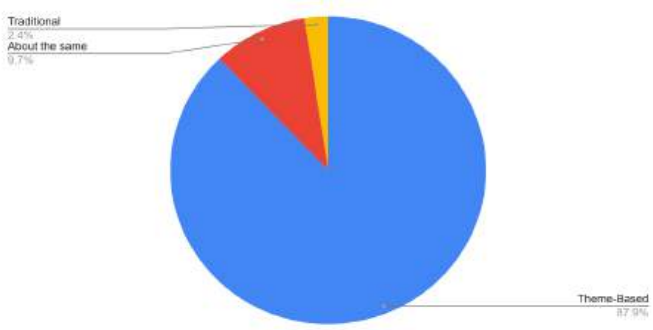


Figure 1: Distribution of participant-reported confidence in learning after exposure to theme-based and traditional instructional formats.

Participants overwhelmingly reported higher confidence following theme-based instruction. A large majority (87.9%, n = 109) indicated that the theme-based condition improved their confidence, compared to 2.4% (n = 3) favoring the traditional condition and 9.7% (n = 12) reporting no difference (Figure 1).

A chi-square goodness-of-fit test revealed that this distribution differed significantly from a uniform distribution, $\chi^2(2, N = 124) = 167.15$, $p < .001$. The effect size was extremely large (Cramér's $V = 0.82$), indicating a substantial preference for theme-based instruction in terms of perceived confidence.

As this measure reflects self-reported confidence rather than objective performance, the results may be influenced by subjective perceptions of clarity and engagement.

4.4 Knowledge Accuracy

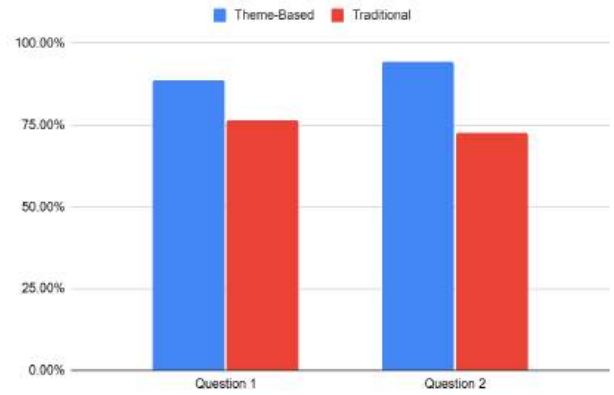


Figure 2: Comparison of correct response rates between theme-based and traditional instruction across two conceptual assessment questions.

Correct response rates were consistently higher following theme-based instruction across both assessment questions (Figure 2). For Question 1, accuracy was 76.61% for the traditional condition and 88.71% for the theme-based condition. For Question 2, accuracy was 72.58% for

the traditional condition and 94.35% for the theme-based condition. A paired-samples t-test revealed that participants achieved significantly higher accuracy scores in the theme-based condition (M = 1.83, SD = 0.45) compared to the traditional condition (M = 1.49, SD = 0.72), $t(123) = 5.50$, $p < .001$, Cohen's $d = 0.49$.

This result indicates a moderate improvement in objective conceptual performance under theme-based instruction. However, the magnitude of this effect may be influenced by a ceiling effect in the theme-based condition, where many participants achieved near-perfect scores. Additionally, as responses were collected in an unsupervised environment, the possibility of external help cannot be ruled out.

4.5 Engagement

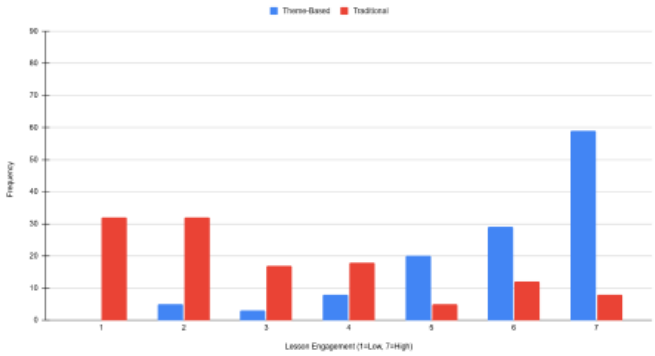


Figure 3: Distribution of student engagement ratings for theme-based and traditional instructional videos (1 = low, 7 = high).

Engagement ratings were substantially higher for the theme-based instructional video (Figure 3). Responses were concentrated in the upper range of the scale (ratings 6–7), whereas traditional instruction showed a broader distribution with higher frequencies in the lower and mid-range values (ratings 1–4).

Engagement ratings were significantly higher for the theme-based instructional condition (M = 5.91, SD = 1.46) compared to the traditional condition (M = 3.00, SD = 1.88). A paired-samples t-test indicated that this difference was statistically significant, $t(123) = 12.39$, $p < .001$. The effect size was large (Cohen's $d = 1.11$), suggesting a substantial increase in engagement under the theme-based instructional format.

While the observed effect size was substantial (d = 1.11), this may reflect the within-subject design of the study and the immediate, self-reported nature of engagement measures. Prior research suggests that subjective engagement ratings are particularly sensitive to instructional format, especially when narrative elements are introduced, such as the case with the theme-based video presented here, with a "sport" narrative.

4.6 Cognitive Load (Mental Effortfulness)

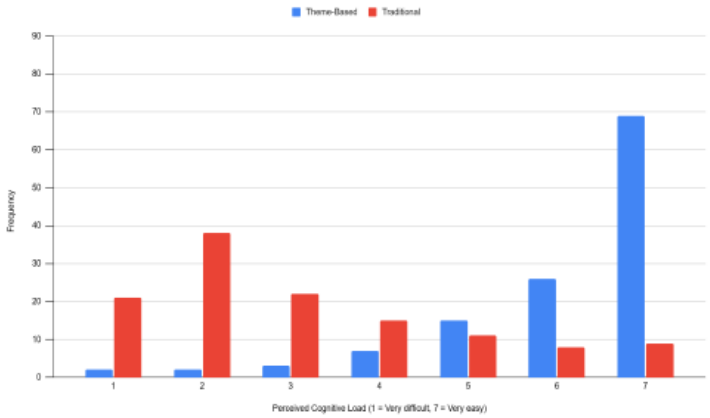


Figure 4: Distribution of perceived cognitive load ratings for theme-based and traditional instructional videos (1 = very difficult, 7 = very easy).

Participants reported lower perceived cognitive load for the theme-based instructional format (Figure 4). Ratings were skewed toward higher values (meaning the lesson was easier to follow), whereas the traditional condition showed greater concentration in lower ratings, indicating higher perceived difficulty.

A paired-samples t-test revealed that perceived cognitive load differed significantly between conditions, with higher ratings (indicating lower mental effort) in the theme-based condition ($M = 6.10$, $SD = 1.34$) compared to the traditional condition ($M = 3.14$, $SD = 1.79$), $t(123) = 13.12$, $p < .001$, Cohen's $d = 1.18$.

This large effect size indicates a substantial reduction in perceived cognitive load when content is presented in a contextualized format. However, this effect may be partially influenced by the within-subject design and the use of immediate self-report measures.

4.7 Conceptual Understanding (Perceived)

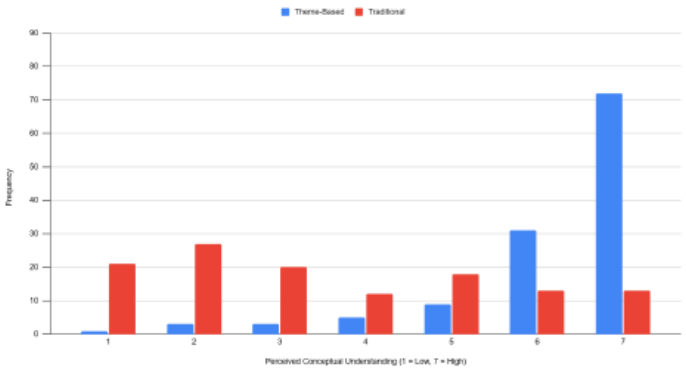


Figure 5: Distribution of perceived conceptual understanding ratings across theme-based and traditional instructional conditions (1 = low, 7 = high).

Perceived conceptual understanding scores were higher for the theme-based condition (Figure 5). Responses were concentrated at the upper end of the scale (ratings 6–7), whereas the traditional condition showed a broader distribution, with many responses in the lower range (ratings 1–4).

A paired-samples t-test revealed that participants reported significantly higher perceived conceptual understanding following the theme-based instructional video ($M = 6.22$, $SD = 1.26$) compared to the traditional instructional video ($M = 3.56$, $SD = 1.98$), $t(123) = 11.26$, $p < .001$, Cohen's $d = 1.01$.

This result indicates a large effect size, suggesting that students perceived substantially greater understanding under the theme-based condition. However, this effect may be amplified by the use of immediate self-reported measures, which are sensitive to engagement, presentation style, and perceived ease of processing.

4.8 Summary of Findings

Measure	Theme-Based M (SD)	Traditional M (SD)	t(123)	p	Cohen's d
Knowledge Accuracy	1.83 (0.45)	1.49 (0.72)	5.50	< .001	0.49
Engagement	5.91 (1.46)	3.00 (1.88)	12.39	< .001	1.11
Cognitive Load	6.10 (1.34)	3.14 (1.79)	13.12	< .001	1.18
Conceptual Understanding	6.22 (1.26)	3.56 (1.98)	11.26	< .001	1.01

Table 1: Summary of paired-samples t-test results across all outcome measures.

Note: Cognitive load ratings are reverse-coded such that higher values indicate lower perceived mental effort (1 = very difficult, 7 = very easy). Knowledge accuracy scores reflect the number of correct responses out of 2. Additionally, the Cohen's d values were calculated with the paired-samples effect size.

4.9 Discussion

The study's purpose was to investigate the effect of theme-based instruction on individual students' engagement, cognitive load, and conceptual understanding. Across all measures (Table 1), the results were consistent with both hypotheses (H1 and H2), suggesting that theme-based instruction was associated with increased engagement and decreased perceived cognitive load, as measured by self-report and by questions after each video.

These findings are consistent with other studies that indicate that context-based and narrative-based learning environments may increase student motivation and interest (Bennett et al., 2003; Ültay and Çalık, 2012). However, the results were not confined to student engagement; students performed better on the conceptual assessment questions and reported a lower cognitive load.

From the perspective of Cognitive Load Theory, the reduction in perceived cognitive load might indicate that familiar themes reduce extraneous cognitive demands. The use of familiar situations (i.e. sports-based) to teach abstract physics concepts may have helped students to interpret and organize the information. Theme-based instruction may have acted as a cognitive scaffold in this manner, assisting students in connecting projectile motion to more easily understood contexts (Sweller, 1988; Mayer and Moreno, 2003).

The improvement in objective knowledge accuracy is important because it suggests that the benefits of theme-based instruction were not limited to students' perceptions. At the same time, there is a chance that the theme-based condition had a possible ceiling effect and a limitation of only two assessment questions may have resulted in an inflated difference. The survey was also conducted in an unsupervised environment, so the possibility of external help cannot be completely discounted.

Some other limitations should also be noted. The large effect sizes may somewhat reflect the within-subject design of the survey, and the immediate self-reported nature of several measures after the videos. Previous research has indicated that subjective ratings might be sensitive to presentation format, particularly when narrative or contextual features are introduced (Mayer and Moreno, 2003). In addition, the theme-based video was shorter than the traditional video (6 versus 9 minutes), allowing the possibility that video length has influenced engagement and perceived cognitive load. Future studies should control for this by using videos of approximately the same duration.

Overall, the results suggest that theme-based instruction may support learning by increasing engagement, improving conceptual understanding, and reducing perceived cognitive load. These findings have implications for instructional design in science education. In addition to considering what content is taught, it may be beneficial for teachers to consider how that content is framed and presented to students.

Future research should examine these effects in a more controlled testing setting, using more assessment items and investigating whether the effects persist over time. These studies could also assess students' varied interests and backgrounds, since the theme used may not be equally relevant for every student.

5.0 Conclusion

This study aims to examine the impact of theme-based compared to traditional instruction on students' engagement, cognitive load and conceptual understanding of projectile motion in secondary physics. Across all measures, findings showed that using theme-based instruction was related to higher levels of engagement, better conceptual understanding, and lower perceived cognitive load than traditional methods.

These results suggest that the method by which instruction is delivered can affect the processing and understanding of scientific concepts. Theme-based instruction might help students relate abstract physics to familiar experiences. That would lessen extraneous cognitive load and promote effective cognitive processing as stated by Cognitive Load Theory.

The consistency of these effects across counterbalanced presentation conditions enhances the credibility of the results by decreasing the probability that the observed differences were due to order effects of viewing videos. These results show the need to examine not only what is being taught, but also how the material is presented. However, the reliance of self-reported data and the unsupervised nature of the survey may have contributed to inflated effect sizes or introduced potential confounds.

Future studies should use controlled testing environments to reduce potential confounding variables, and should increase the number of conceptual assessment questions to improve the knowledge measurement reliability. Examining long-term retention, such as tracking performance over a semester, would also help to strengthen the observed effects. Furthermore, exploring the combination of varying themes with a variety of student demographics may yield strong instructional insights into how different individuals perform with themes of different kinds.

Acknowledgements

I thank my mentor, Professor Faisal Nawab, for his guidance throughout the process of developing this research paper. I am grateful to the anonymous survey participants for their constructive responses in the survey. I also thank Eshwar Sai Ganipaneni, a fellow classmate at Abbey Park High School, for conducting the survey of over 70 students at different schools and student communities.

Authors

Haricharan Vemuri is a student from Ontario, Canada, with interests in cognitive sciences and symbolic systems. He is passionate about applying research methods to improve learning outcomes in secondary STEM education.

Works Cited

Bennett, Judith, Susan Hogarth, and Frank Lubben. A Systematic Review of the Effects of Context-Based and Science-Technology-Society (STS) Approaches in the Teaching of Secondary Science. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, 2003.

Boone, Harry N., Jr., and Deborah A. Boone. "Analyzing Likert Data." Journal of Extension, vol. 50, no. 2, 2012, pp. 1–5. https://doi.org/10.34068/joe.50.02.48.

Hiwatig, Bryan M. R., Gregory H. Roehrig, and Matthew D. Rouleau. "Unpacking the Nuances: An Exploratory Multilevel Analysis on the Operationalization of Integrated STEM Education and Student Attitudinal Change." Disciplinary and Interdisciplinary Science Education Research, vol. 6, 2024, article 18, https://doi.org/10.1186/s43031-024-00108-6.

Jebb, Andrew T., Vivian Ng, and Louis Tay. "A Review of Key Likert Scale Development Advances: 1995–2019." Frontiers in Psychology, vol. 12, 2021, article 637547, https://doi.org/10.3389/fpsyg.2021.637547.

Mayer, Richard E., and Roxana Moreno. "Nine Ways to Reduce Cognitive Load in Multimedia Learning." Educational Psychologist, vol. 38, no. 1, 2003, pp. 43–52, https://doi.org/10.1207/S15326985EP3801_6.

McLure, Faith I., Kwan-Sing Tang, and Peter J. Williams. "What Do Integrated STEM Projects Look Like in Middle School and High School Classrooms? A Systematic Literature Review of Empirical Studies of iSTEM Projects." International Journal of STEM Education, vol. 9, 2022, article 73, https://doi.org/10.1186/s40594-022-00390-8.

Osborne, Jonathan, Shirley Simon, and Sue Collins. "Attitudes towards Science: A Review of the Literature and Its Implications." International Journal of Science Education, vol. 25, no. 9, 2003, pp. 1049–79, https://doi.org/10.1080/0950069032000032199.

Paas, Fred, Alexander Renkl, and John Sweller. "Cognitive Load Theory and Instructional Design: Recent Developments." Educational Psychologist, vol. 38, no. 1, 2003, pp. 1–4, https://doi.org/10.1207/S15326985EP3801_1.

Redish, Edward F. "Implications of Cognitive Studies for Teaching Physics." American Journal of Physics, vol. 62, no. 9, 1994, pp. 796–803, https://doi.org/10.1119/1.17461.

Sweller, John. "Cognitive Load during Problem Solving: Effects on Learning." Cognitive Science, vol. 12, no. 2, 1988, pp. 257–85, https://doi.org/10.1207/S15516709COG1202_4.

Ültay, Neslihan, and Muammer Çalık. "A Thematic Review of Studies into the Effectiveness of Context-Based Chemistry Curricula." Journal of Science Education and Technology, vol. 21, no. 6, 2012, pp. 686–701, https://doi.org/10.1007/s10956-011-9357-5.

How Can Intergenerational Activities Be Designed to Foster Better Relationships Between Youth and Elderly Individuals with Dementia?: MindLink, a Game Designed to Foster Intergenerational Engagement Between Youth and Dementia Patients By Erika Chen[1]

Abstract

Dementia is an extremely prevalent situation in society nowadays. People live with dementia in a few ways. This may include maintaining quality of life, managing symptoms, and providing support and care. Different problems may arise with each patient, they may include problems with memory, thinking, and communication, trouble with daily tasks, behavioral issues, difficulty with judgement, social isolation, and physical challenges like swallowing, navigating and moving. After the rise of COVID-19, social isolation became more apparent in our society. With the digital age now, it decreases in-person social interaction. With some dementia patients cutting off social interaction so as to not be a burden to those around them, it is even more important to support them. Intergenerational activities are not only a beneficial way to support people with dementia, but youth as well. Intergenerational interventions on wellbeing and mental health of elderly can be impactful in general and for dementia cases. Social involvement can be advantageous for both old and young people as intergenerational programs improve interactions between different ages in a positive way. Younger people can learn about communication, patience, and empathy when interacting with older people living with dementia, as well as developing their interpersonal skills more generally. I propose a game, MindLink, for elderly and youth in a family. Through this game they will have social interaction, talk about memories and laugh at the jokes. This game could be included in a senior center as part of their activity time as a pilot trial. Staff facilitators must be present during activities to help provide reassurance to both groups. This attempt will help to initiate more broader interaction for youth and elderly especially covering people with dementia among the public in the future. In conclusion, intergenerational activities are extremely important for youth and elderly with dementia and can help both demographics.

Keywords Young, dementia, Intergenerational, activities, MindLink, elderly, youth, meaningful engagement, family game

1. Introduction

Dementia is a medical condition that causes a person to go through worsening cognitive abilities. A person with dementia typically has a harder time with different cognitive abilities such as memory, language, and attention. There are many different types of dementia such as Alzheimer's, vascular dementia, frontotemporal dementia (Arvanitakis and Bennett 2019). There are over 720,000 people in California, ages 65 and older (Figure 1), that are currently living with Alzheimer's (Alzheimer's Association, n.d.). As of October 2025, California has 39,663,800 people ("California Population 2025," n.d.), meaning roughly 1.82% struggles with this disease.

1- Diamond Bar High School, Diamond Bar, California, United States of America

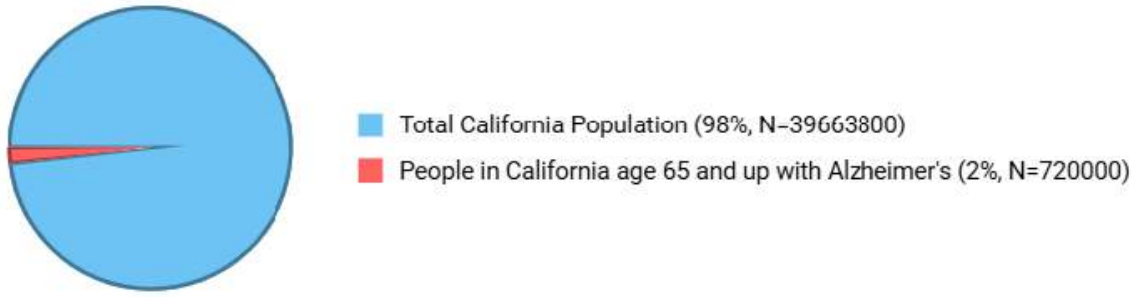


Figure 1. The statistics of elderly people living with Alzheimer's in California. The data is taken from world population view website (https://worldpopulationreview.com/states/california) and combined with the data from another study to generate this pie chart graph (Dhana et al. 2023).

While this may not seem like a big number, it is a huge problem across the United States. The prevalence of Alzheimer's is 12.9% in Maryland, 12.7% in New York, 12.5% in Mississippi and Florida, with California and Illinois having an estimated amount of 12% each (Figure 2). People live with dementia in a few ways. This may include maintaining quality of life, managing symptoms, and providing support and care (National Academies of Sciences et al. 2021). Different problems may arise with each patient, they may include problems with memory, thinking, and communication, trouble with daily tasks, behavioral issues, difficulty with judgement, social isolation, and physical challenges like swallowing, navigating and moving (Mayo Clinic, n.d.). The government plays a role in the situation as well. They have introduced and hosted many programs to assist dementia patients. For example, the GUIDE Model helps with providing care support through Medicare ("Guiding an Improved Dementia Experience (GUIDE) Model | CMS," n.d.). In California, programs like CAlz Connect ("Alzheimer's/Dementia Support – CAlz Connect Program | California Department of Aging - State of California," n.d.), Cal-Compass ("Information & Resources - California Community Program for Alzheimer's Services and Supports | California Department of Aging - State of California," n.d.), and The National Alzheimer's and Dementia Resource Center helps to provide care and community service and resources for the patients. However, these programs are not always enough to fill the needs and enhance the quality of life of the patients and care takers.

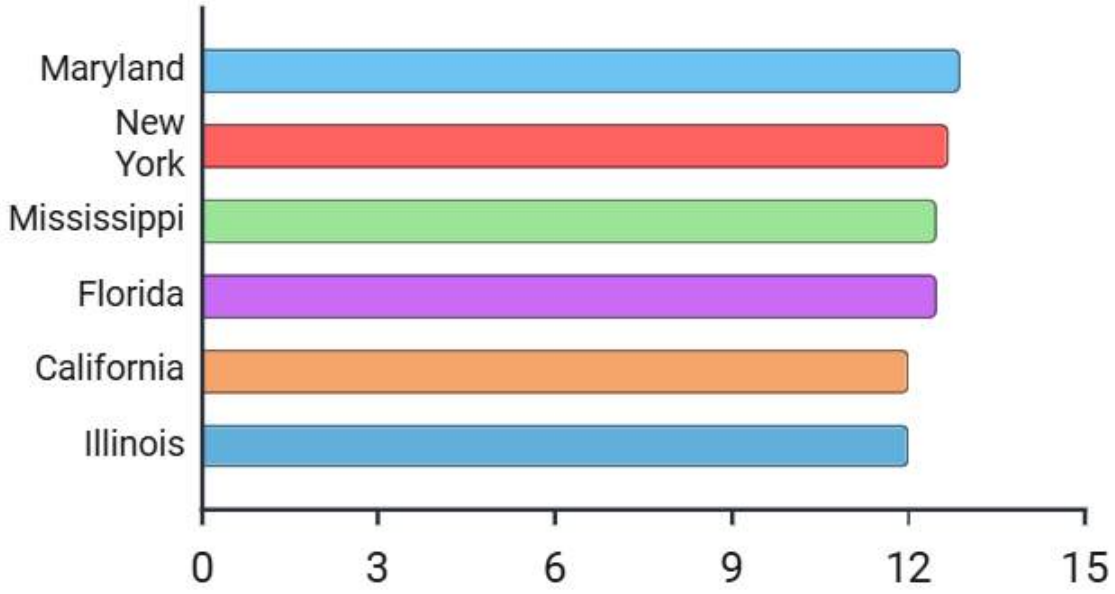


Figure 2. Alzheimer's cases across the United States. The data is taken from Dhana et al. 2023, and the graph is generated in Biorender.

Intergenerational interventions on wellbeing and mental health of elderly can be impactful in general and for dementia cases (Whear et al. 2023). Social involvement can be advantageous for both old and young people as intergenerational programs improve interactions between different ages in a positive way. From a literature review by University of Cambridge, they gathered 27 sources and evaluated ten studies of children's outcomes through participating in intergenerational programs following their view of the elderly. They found that there was a short and long term positive impact on children as well as overall beneficial on the elderly participants (Gualano et al. 2018). Another paper in Sage Journals discusses how intergenerational activities have been a way to help dementia people stay socially connected. Such programs provide an opportunity to build relationships and for young people to learn about dementia, aging, and how to interact with older people living with dementia (Figure 3). They note that potential benefits include enhancing sense of wellbeing for people living with dementia, raising awareness among children engaging in the activities, and maximizing the role of the community in social engagement with people living with dementia.

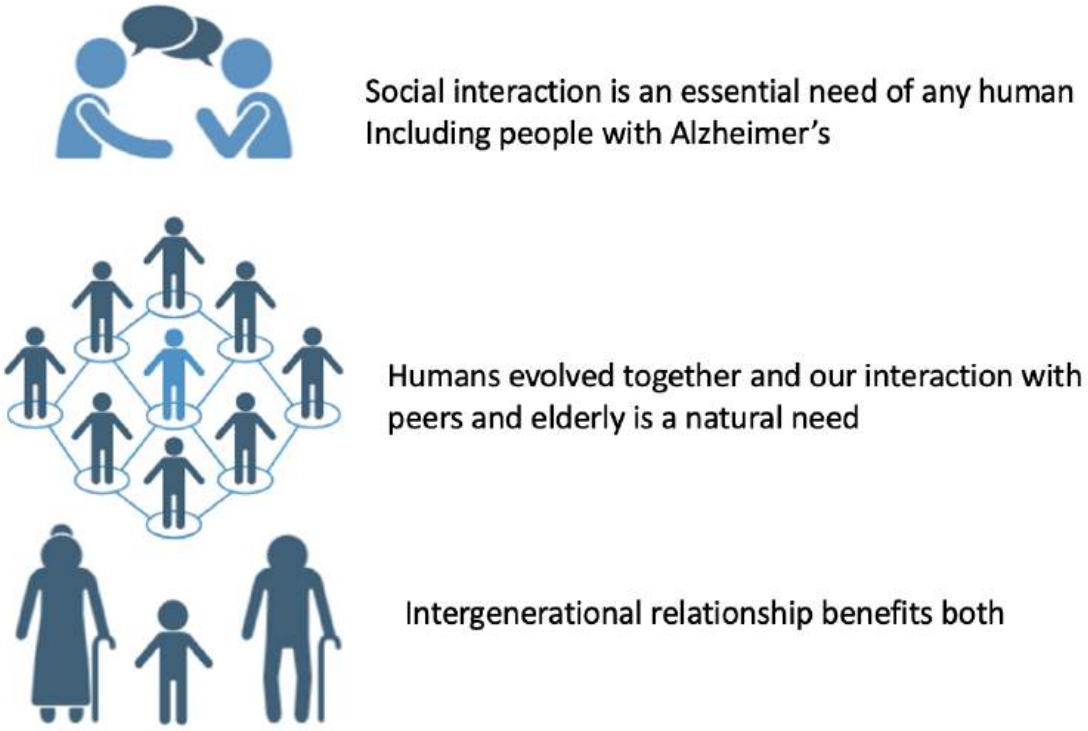


Figure 3. Diagram of social connection that humans have between all ages. Humans evolved together and therefore social interaction is the essence of our life. Naturally we both interact with our peers and elderly while we are growing up. This does not change when one develops Alzheimer's.

Younger people can learn about communication, patience, and empathy when interacting with older people living with dementia, as well as developing their interpersonal skills more generally (Houghton et al. 2022). In another study by Bridget A. McNair and Krista S. Moore, they note that elder's active participation increased in programs when young children were present. In addition, young children had more positive trends in their pictures and responses than children who were not involved in intergenerational programs (McNair and Moore 2010).

On a more psychological side, the International Journal of School and Cognitive Psychology reported that better psychological outcomes were found including reduced anxiety and an improved sense of self-worth. Classroom behaviors improved among children in need of fostering pro-social behaviors. In fact, most studies they reviewed reported children and young people tended to report positive change in attitude towards older adults during and or after participating in programs involving close interactions between the generations. As for social outcomes, they noted improved communication skills, reduced social distance in relationships in non-kin relationships as well as with older family members (https://www.longdom.org/school-cognitive-psychology.html, n.d.). Another article from the ScienceDirect, reports engagement in intergenerational activities being linked with increased

physical and social activities. Many researchers have also recognized intergenerational interactions as key contributors to reducing ageism and are important for reducing social isolation (Zhong et al. 2020)**.**

As such, intergenerational activities should be an extremely important part in a dementia patient's care process. However, the way these activities are facilitated are a crucial part. To start, staff members who help to manage and put forth intergenerational activities must be specifically trained and the activities must have careful organization (Gualano et al. 2018). Staff facilitators must be present during activities to help provide reassurance to both groups. When coming up with or making activities, one must take personal preferences of participants into account and be considerate of different aspects. The paper from Sage Journals provided readers with 3 models of intergenerational care that have been previously identified. When coming up and thinking of preferences of the people, identifying which model of intergenerational care (Figure 4) to use could be useful. Models include centre-based visitation where one generation visits another in a designated centre, shared campus approach with facilitates for both generations with a common area for intergenerational care, or family day care approach which is suitable for when the older generation requires 24 hours nursing care.

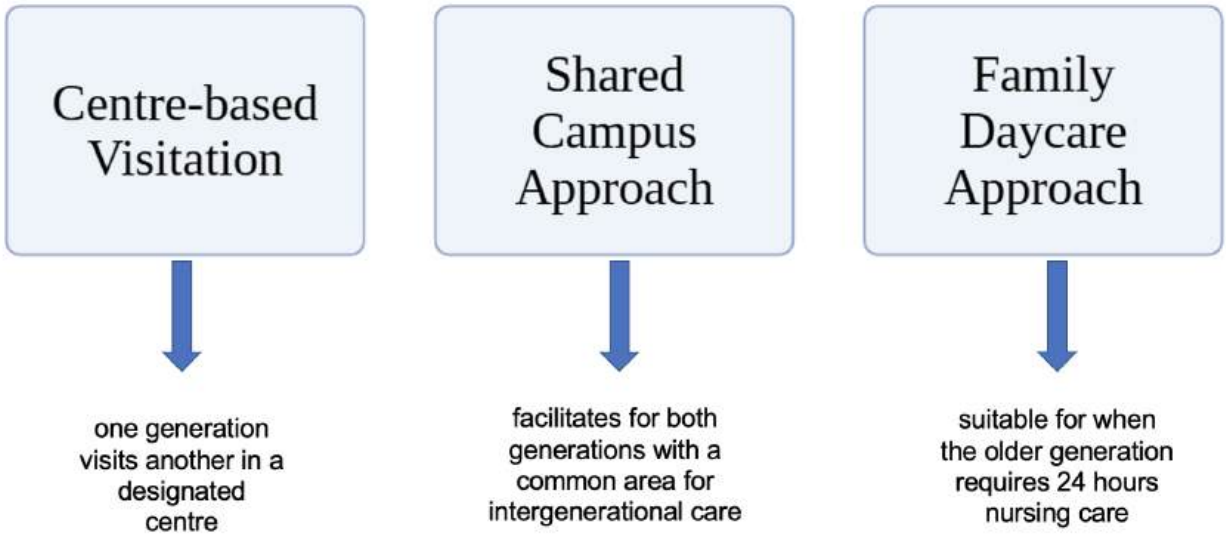


Figure 4. Intergenerational care models. Intergenerational care models could be categorised in three: 1-centre-based visitation, 2-shared campus approach, 3- family daycare approach.

After choosing which model, thoroughly and with detail go through the program to ensure there is maximum benefit and minimal risk which summarized in Table1. These programs will benefit both with similar social needs and benefits and will not make the disease a blocking condition as the aim will be defined in the beginning. While aiming for target demographics, it is best to aim for a program that targets older people and secondary school aged children or college students. It is best to aim for narrower age gaps. The program/activity should aim for a 1:1 ratio

of older and younger people. The program should aim to try to get older people with dementia and younger people to make them feel more comfortable and gain familiarity with each other so their relationships can grow in time. It is best to incorporate the opportunity to share life stories so relationships can develop, allowing older people with dementia and younger people to smile, laugh, converse, and make eye contact with each other. There should be limited noise and crowding during programs and make sure to thoroughly plan out location and spatial plant or delivery of the program. While thinking of how long to make the program, avoid having programs that meet too often. In an experiment conducted by the International Journal of School and Cognitive Psychology, they noted that group cohesiveness was most commonly identified as a therapeutic factor (Houghton et al. 2022).

In addition young people were able to benefit most from interactions with older people who are willing and able to share their wisdom and knowledge to the next generations. They also noted that curriculum-based education alone may not cover a wide range of social skills required to survive in everyday life like compassion and respect. As such, programs should not stick to merely curriculum-based education and allow more social interaction. (https://www.longdom.org/school-cognitive-psychology.html, n.d.). A paper by Epstein and Boisvert found 5 components to characterize effective programs and promote participant engagement. This included designated intergenerational space that is shared, accessible, and stocked with materials inviting to both age groups, consistent daily schedule that allows for formal as well as informal cross-age interactions, open-ended activities that emphasize process over product and provide opportunities for planning and reflection, the explicit facilitation of cross-age interactions by caregivers, and objective observational assessment to plan activities and share information with families (Epstein and Boisvert 2006). The New England Journal of Medicine provided a few activities that were associated with a reduced risk of dementia. This included reading, playing board games, playing musical instruments, and dancing. In addition, increased participation in cognitive activities at baseline was associated with reduced rate of decline in memory. Participation in leisure activities may increase cognitive reserve, delaying onset of dementia (Verghese et al. 2003).

Table1. Aims and benefits of intergenerational programs for elderly with Alzheimer's and the youth.

Program aims	Benefit for elderly patient	Benefit for the youth
targets older people and secondary school aged children or college students	Energetic interaction	Wise dialogues
narrower age gaps	Hearing different perspectives	Hearing different perspectives
1:1 ratio of older and younger people	Quality attention	Quality attention
gain familiarity with each other so their relationships can grow in time	Healthy bonding	Healthy bonding
opportunity to share life stories so relationships can develop	adventurous stories	adventurous stories
Activities that allowing older people with dementia and younger people to smile, laugh, converse, and make eye contact with each other	Joy, connection, attention	Joy, connection, attention
limited noise and crowding during programs	Comfortable atmosphere	Comfortable atmosphere
thoroughly plan out location	Comfort and easy access	Comfort and easy access
Avoid meeting too often	Prevent burden	Prevent burden

Success for programs and activities facilitated can be measured through patient and youth feedback. Feedback can be based on three categories: social, cognitive, and wellbeing. This means, how socially active or more spoken did the two parties become, how did it impact them cognitively, and how their wellbeing improved.

Mock survey:

1. What activity did you participate in:
   A- games b- talking c- physical activity
2. How did you feel before the activity: 1-10
   A- Sad 1-10
   B- Happy 1-10
   C-Angry 1-10
   D-Joyful 1-10
   E-Notral 1-10
3. How did you feel after the activity: 1-10
   A- Sad 1-10
   B- Happy 1-10
   C-Angry 1-10
   D-Joyful 1-10
   E-Notral 1-10
4. Which part do you think improved the most:
   A-social
   B-cognitive

C-wellbeing

5. How well did you connect with the other party: 1-10
6. What was your favorite part of the activity
7. What could be improved about the activity

A proposed solution for the defined problem:MindLink Game

With dementia patients, one problem that occurs is social isolation. What if there was a way to connect our new generation with the elderly and help lessen their social isolation? MindLink (Figure 5) is a family board game designed to help the elderly with dementia connect with their loved ones by sharing personal anecdotes and playing fun games. The game is built around a game backstory where there is a Grandma, Grandpa, Molly (little girl), Mason (little boy), and Waggles (dog). Together they go through activities and questions to try to reach the Memory Garden (the end) to help Grandma restore her memory. The initiation seed fund for this initiative can be gained via Kickstarter fundraising paths. Moving forward, it is planned to sell the game to local centers that will be part of this intergenerational interaction plan.

To start this initiative with the MindLink game one would pick a senior center to target first when implementing my board game. I am already a volunteer at a local senior center (the name cannot be disclosed for safeguarding purposes) which will be the first point to implement the idea. Senior centers already have dedicated game time as one of their activities so it would be easy to implement an intergenerational game time. As the game was designed as a family game, we would reach out to relatives of the elderly people to gather youth to the center for that particular activity time. There would be a consent form for both parties telling them the purpose of the game, what the research will be used for, and how the data will be collected. From there, one would observe, take notes, and gather feedback from the residents who played it. Through this research, it has been found that having staff members to monitor these intergenerational activities is extremely important. Using this structure one can compile more data on how effective the game was. This could lead further improvements in the structure of the initiative then we could open up for public volunteers as not everyone might have young relatives. There will be more safeguarding policies implemented in the public structure. Observant staff members would stay active and a police officer could also be included for safety and protection towards vulnerability. The police officer could be one of the “staff members” while also helping to handle security and ensure safety is met. It would also make it a requirement for an adult to be there, not only to watch over the children but also the elderly. In the future community centers also could be considered for these kinds of events to be able to reach more people. The family events can be hosted among communities in the future.

Figure 5. A prototype of the Mindlink Game. This image shows the representative structure of the Mindlink Game. 1- Front of the board game box, 2- Board game character set, elderly and young people of all identities 3- Board game itself, 4- Examples of activity cards, 5- Back of the board game box, 6- Example of question cards.

2. Discussion

Studying dementia is extremely important for our future in terms of both the disease biology and the societal side of it. With how prevalent it is becoming, further research allows us to assist with finding effective treatments, understand risk factors, and improve the lives of those affected and their caregivers. (National Institute on Aging, n.d.). This literature review assists with these as it allows us to further see how intergenerational activities can be designed to foster meaningful engagement between dementia patients and youth. In the future, research should continue to be found to support not only these demographics, but the core problem, dementia. Social activities are important for humans since it is a part of everyday life. In this case, it is important for dementia patients as one of the biggest things they lack is social interaction.

Dementia caregivers report to have higher levels of stress, depression, and anxiety symptoms (Sörensen and Conwell 2011). Society can work to support patients by raising money for causes that help them at the root. There are many organizations that work to help dementia/Alzheimer's patients like the Alzheimer's Association and Family Caregiver Alliance. There are not many initiatives to address the social gap and the burden it brings to both patients themselves and the caregivers. I believe that older people living with dementia and younger people can enjoy each other's company. This can be seen through smiling, laughing, conversation, and making eye contact (Houghton et al. 2022) as a measurement parameter. They can do activities together in a controlled manner. These activities can include interactive group

discussions, games like bingo, creative expressions, and outdoor activities like trips or picnics. Playing musical instruments, writing for pleasure, and reading the newspaper can also be helpful. Here I propose a special game, a family game, MindLink for this purpose.

In the future we can have better structured social initiatives that serve the community for better good addressing the unseen needs. The initiative that I plan will have a positive impact on dementia patients' lives and their caregiver as well as their loved ones.

3. Conclusion

In conclusion, dementia is an extremely prevalent situation in society nowadays. After the rise of COVID-19, social isolation became more apparent in our society. With the digital age now, it decreases in-person social interaction. With some dementia patients cutting off social interaction so as to not be a burden to those around them, it is even more important to support them. Intergenerational activities are not only a beneficial way to support people with dementia, but youth as well. With initiatives supporting this interaction such as MindLink game the society will improve and patients' quality of life will increase along with improved care takers quality of life.

Works Cited

Alzheimer's Association. n.d. "Alzheimer's and Public Health Action in California | Alz.Org." Accessed November 2, 2025. https://www.alz.org/professionals/public-health/state-overview/california.

"Alzheimer's/Dementia Support – CAlz Connect Program | California Department of Aging - State of California." n.d. Accessed November 2, 2025. https://aging.ca.gov/Programs_and_Services/Alzheimers_Dementia_Support/.

Arvanitakis, Zoe, and David A. Bennett. 2019. "What Is Dementia?" JAMA 322 (17): 1728. https://doi.org/10.1001/jama.2019.11653.

"California Population 2025." n.d. Accessed November 2, 2025. https://worldpopulationreview.com/states/california.

Dhana, Klodian, Todd Beck, Pankaja Desai, Robert S. Wilson, Denis A. Evans, and Kumar B. Rajan. 2023. "Prevalence of Alzheimer's Dementia in the 50 U.S. States and 3142 Counties: A Population Estimate Using the 2020 Bridged-Race Postcensal from the National Center for Health Statistics." Alzheimer's & Dementia : The Journal of the Alzheimer's Association 19 (10): 4388–95. https://doi.org/10.1002/alz.13081.

Epstein, Ann S., and Christine Boisvert. 2006. "Let's Do Something Together: Identifying the Effective Components of Intergenerational Programs." Journal of Intergenerational Relationships 4 (3): 87–109. https://doi.org/10.1300/J194v04n03_07.

Gualano, Maria Rosaria, Gianluca Voglino, Fabrizio Bert, Robin Thomas, Elisa Camussi, and Roberta Siliquini. 2018. "The Impact of Intergenerational Programs on Children and Older Adults: A Review." International Psychogeriatrics 30 (4): 451–68. https://doi.org/10.1017/S104161021700182X.

"Guiding an Improved Dementia Experience (GUIDE) Model | CMS." n.d. Accessed November 2, 2025. https://www.cms.gov/priorities/innovation/innovation-models/guide.

Houghton, Catherine, Marita Hennessy, Siobhán Smyth, et al. 2022. "The Experiences and Perceptions of Young People and Older People Living with Dementia of Participating in Intergenerational Programmes: A Qualitative Evidence Synthesis." Dementia 21 (7): 2144–71. https://doi.org/10.1177/14713012221112385.

https://www.longdom.org/school-cognitive-psychology.html. n.d. "Longdom Publishing SL | Open Access Journals." Longdom. Accessed November 2, 2025. https://www.longdom.org/.

"Information & Resources - California Community Program for Alzheimer's Services and Supports | California Department of Aging - State of California." n.d. Accessed November 2, 2025. https://aging.ca.gov/Information_and_Resources/Cal_COMPASS/.

Mayo Clinic. n.d. "Vascular Dementia: Prevention May Be Possible-Vascular Cognitive Impairment and Vascular Dementia - Symptoms & Causes." Accessed November 2, 2025. https://www.mayoclinic.org/diseases-conditions/vascular-dementia/symptoms-causes/syc-20378793.

McNair, Bridget A., and Krista S. Moore. 2010. “The Effects of Intergenerational Programs on Individuals with Alzheimer’s Disease or Dementia.” Annual in Therapeutic Recreation 18 (January): 141–57.

National Academies of Sciences, Engineering, Division of Behavioral and Social Sciences and Education, Cognitive Board on Behavioral, and Committee on the Decadal Survey of Behavioral and Social Science Research on Alzheimer’s Disease and Alzheimer’s Disease-Related Dementias. 2021. “Improving Outcomes for Individuals Living with Dementia.” In Reducing the Impact of Dementia in America: A Decadal Survey of the Behavioral and Social Sciences. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK574334/.

National Institute on Aging. n.d. “Dementia Care and Caregiver Support Research.” Accessed November 2, 2025. https://www.nia.nih.gov/2021-2022-alzheimers-disease-related-dementias-scientific-advances/dementia-care-and-caregiver.

Sörensen, Silvia, and Yeates Conwell. 2011. “Issues in Dementia Caregiving: Effects on Mental and Physical Health, Intervention Strategies, and Research Needs.” The American Journal of Geriatric Psychiatry : Official Journal of the American Association for Geriatric Psychiatry 19 (6): 491–96. https://doi.org/10.1097/JGP.0b013e31821c0e6e.

Verghese, Joe, Richard B. Lipton, Mindy J. Katz, et al. 2003. “Leisure Activities and the Risk of Dementia in the Elderly.” New England Journal of Medicine 348 (25): 2508–16. https://doi.org/10.1056/NEJMoa022252.

Whear, Rebecca, Fiona Campbell, Morwenna Rogers, et al. 2023. “What Is the Effect of Intergenerational Activities on the Wellbeing and Mental Health of Older People?: A Systematic Review.” Campbell Systematic Reviews 19 (4): e1355. https://doi.org/10.1002/cl2.1355.

Zhong, Sinan, Chanam Lee, Margaret J. Foster, and Jiahe Bian. 2020. “Intergenerational Communities: A Systematic Literature Review of Intergenerational Interactions and Older Adults’ Health-Related Outcomes.” Social Science & Medicine 264 (November): 113374. https://doi.org/10.1016/j.socscimed.2020.113374.

The Interaction Between Loss Aversion and FOMO Among Generation Z Teenagers
By Noah Sperling

Abstract

While both loss aversion, the tendency for losses to be weighted more heavily than equivalent gains, and fear of missing out (FOMO), a recent but notable psychosociological behavior especially prevalent among Generation Z, have each been researched extensively, there is little research analyzing the potential interaction between them. This study measured whether FOMO could significantly predict loss aversion among Generation Z teenagers, and if so, under what conditions. 131 participants from CHS, a high school in suburban California, were given the 10-item FOMO Scale by Przybylski et al. (2013) and 6 hypothetical "coin flip" scenarios which they could either accept or reject in order to measure participants' level of FOMO and loss aversion, respectively. These coin flip scenarios included both symmetric values (e.g. "You can win $10 or lose $10") and asymmetric values (e.g. "You can win $20 or lose $10"). Multiple statistical analyses, including Pearson Correlation tests, individual linear regressions, and a multiple regression analysis with 786 rows of data, showed FOMO had a strong effect on loss aversion at low stakes, but that when stakes increased, this result completely reversed and became statistically insignificant. Higher FOMO actually influenced greater loss aversion, contrary to the hypothesis, plus symmetric versus asymmetric questions had no differentiating effect. These results prove FOMO to be an important variable in behavioral economics and provide important context for when loss aversion can be proved.

Keywords: Loss Aversion, FOMO, Behavioral Economics, Generation Z, Prospect Theory

I. INTRODUCTION

The field of behavioral economics features a unique approach to analyzing cognitive biases that people fall prey to day in and day out. One cognitive bias that has far-reaching implications in public policy, financial decision-making, and even sports viewing is loss aversion (Benartzi & Thaler, 1995). Loss aversion is the cognitive bias that causes people to be inherently more bothered by losses than gains of equivalent or even greater value, creating decision-making behavior that prioritizes avoiding losses, as recognized by Daniel Kahneman and Amos Tversky in their 1979 Prospect Theory Paper (Kahneman & Tversky, 1979).

While loss aversion is widely recognized as a major cognitive bias, research has been unable to consistently prove its existence, with some studies showing robust effects, and others failing to show much of an effect at all. An emerging belief is that this lack of consistent results occurs since the effect can vary significantly based on context and sample. Specifically, a possible cause of these inconsistent results is that loss aversion has historically been tested on large, undifferentiated samples without any attempt to determine which segments might be more susceptible.

This study tests loss aversion among Generation Z teenagers, born between 2006 and 2012, a subsegment of the Generation Z cohort, using a measurable psychological variable, FOMO, or fear of missing out, as a moderator. Gen Z teenagers would be particularly relevant to the research gap of in-depth loss aversion studies, as they are not only underrepresented in loss aversion testing but also possess unique cohort personality traits, such as unprecedented levels of social media usage and vulnerability to FOMO, that may influence a loss and gain mindset (Einstein et al., 2023). This study aims to determine whether loss aversion is impacted by FOMO level and, further, by the type of question and the value of the hypothetical losses and gains.

Throughout the paper, the researcher will answer the question: Is there a difference in Generation Z teenagers' reactions to tasks of loss aversion based on whether they have low or high FOMO?

II. LITERATURE REVIEW

Introduction

Although behavioral economics and psychology are considered different scholarly fields of study, the underlying emotional drivers of decision-making show great overlap and potential for interaction. One prominent phenomenon from economics, loss aversion, and another from psychology, FOMO, each shape decision-making in everyday life. Prior research has thoroughly examined these two concepts; however, they have been almost exclusively analyzed separately. Thus, even though a wealth of literature exists regarding these two topics, there is a gap in examining the interaction between loss aversion and FOMO overall and demographically.

Loss Aversion and Decision-Making

Loss aversion has become a widely known behavioral economics concept thanks to researchers Daniel Kahneman and Amos Tversky. In 1979, the two developed the concept of loss aversion within their broader research, labeled "Prospect Theory," serving as a model for decision-making under risk. Loss aversion was formally defined as "the tendency for losses to be weighted more heavily than equivalent gains in decision-making under risk" (Kahneman & Tversky, 1979). Their research proved that people react more intensely to potential losses than gains, even when the dollar values are identical. This bias does not just apply to financial decision-making, but also to other real-life scenarios where one feels extreme pressure not to lose opportunities or other valuables. While Kahneman and Tversky outlined and established this behavioral concept, Gal and Rucker (2018) later added nuance to the findings, explaining how loss aversion does not operate equally across different scenarios and samples, showing a meaningful addition.

The Application of Loss Aversion Testing

Due to the impact of Kahneman's and Tversky's initial study, their hypothetical "gamble test," including a 50-50 coin flip task, became widely recognized as the standard and was then

built upon thirteen years later in their 1992 paper on Cumulative Prospect Theory. By testing a similar methodology to their 1979 paper, utilizing binary decision-making and hypothetical coin flip questions, the researchers were able to confirm the phenomenon of loss aversion, validate the use of hypothetical coin flip questions, and prove the existence of the value function (Kahneman & Tversky, 1992). The value function described the perceived behavior of participants, as when the dollar amounts presented in hypothetical coin flip questions were increased, their sensitivity to losses and gains decreased, and vice versa. This momentous finding became a foundational theory and milestone in loss aversion research.

Adding to the real-world implications of loss aversion, Benartzi and Thaler (1995) used myopic loss aversion to explain the advantage stocks had over bonds in the 1990s. They found that when investors judged their portfolios "short-sightedly," almost like they had myopia, greater loss aversion developed. This study is notable, first for bringing new terminology and clarity to the field of loss aversion, and secondly, for showing loss aversion can impact financial decisions, providing rationale for the researcher's attempt to study the phenomenon further.

Situational Nature of Loss Aversion

While Kahneman and Tversky provided a strong foundation for loss aversion research, there were still many unknowns about the phenomenon they defined. Skepticism arose about the ubiquity of loss aversion across all populations due to the lack of consistent evidence and results across studies. Gal and Rucker (2018) conducted a literature review of prior loss aversion studies, specifically focusing on whether loss aversion was confirmed or denied, and determined that loss aversion is dependent on context. Their study reinforced the unpredictability of loss aversion and demonstrated why context is important in driving it, thus being able to influence the results of entire experiments. Nagaya (2021) further explains the contextual nature under which loss aversion emerges by segmenting loss aversion testing into four different approaches: "the absolute magnitude of losses, the bias of attentional allocation, the ranking of losses relative to gains, and the preference for inaction," and found that loss aversion presented differently under each scenario. The unique presentations demonstrate how loss aversion is truly dependent on context, emerging differently under different situations. Since both pieces of research discuss the conditional nature of loss aversion, there is a potential for one such scenario to be FOMO.

The Psychology of FOMO

While loss aversion originates from the field of behavioral economics, FOMO has been widely studied in the area of psychology since becoming a defining feature of modern-day life, impacting everyday decision-making. Przybylski et al. (2013) formally defined FOMO as a "pervasive apprehension that others might be having rewarding experiences from which one is absent." To address a gap surrounding adolescents interacting with social media and other platforms that may create FOMO, the researchers created the ten-item FOMO scale, allowing for quantification of the phenomenon. With the success of the experiment and its versatility, the FOMO Scale is considered the most widely accepted diagnostic tool to measure FOMO across

the research community. Elhai et al. (2021) added meaningful commentary to these results by expanding the realm of knowledge surrounding FOMO to include the digital landscape of social media, which has platforms that can easily create FOMO-riddled behavior.

Socio-Emotional Implications of FOMO

Further research delves into the emotional impact of FOMO and its potential impact on decision-making. Elhai et al. (2021) synthesized recent research into how FOMO manifests into anxiety-ridden behavior through modern forms of online activity and communication. This literature review establishes the implications FOMO has on decision-making, specifically an increase in poor decision-making from anxiety and other issues that stem from it. When individuals with FOMO are faced with anxiety or other negative emotions, they are likely to make poorer decisions, giving credence to the need to better understand said decision-making. Since FOMO creates anxious decision-making, and loss aversion is context-dependent, emerging under hyper-specific conditional factors (Gal & Rucker, 2018; Nagaya, 2021), FOMO may be the type of psychological moderator to provide a context in which loss aversion emerges.

Research Gap

Despite extensive research conducted separately on both loss aversion and FOMO, a gap exists in the literature in analyzing the interaction between the two concepts. Rahmawati and Raharja (2024) discuss how Generation Z's investment decision-making is influenced by herding behaviors, loss aversion, and the availability of information, while also looking at FOMO as a mediator. Rahmawati and Raharja's study is notable as it is one of the only papers available that comes close to the researcher's topic, discussing both loss aversion and FOMO with Generation Z relevance; however, it differs from the current research as it analyzes FOMO as a mediator, rather than a variable with a direct correlation to risk-averse behavior. Additionally, the results of their paper may not be able to accurately predict the results of this research due to the specific demographics and culture of Semarang, Indonesia, which are very different from those at CHS in California. Overall, the research gap, due to the lack of studies measuring the dynamic between loss aversion and FOMO among Gen Z teenagers is significant, as FOMO and loss aversion may interact to impact decision-making behavior, and understanding these potential effects will allow for greater clarity in understanding the human psyche and decision-making.

III. METHODS

Background

This study investigates whether FOMO level predicts loss aversion behavior among Gen Z teenagers, and whether that relationship differs based on question type. Through an in-depth analysis of previous loss aversion and FOMO literature, the researcher developed the following hypotheses: Gen Z audiences with lower FOMO are more likely to exhibit loss aversion than those with higher FOMO, and secondly, that the loss aversion effect would be more pronounced

when the potential gain exceeds the potential loss, measured with hypothetical coin toss tests. With this in mind, the researcher determined that a correlational analysis would best prove the hypothesized relationship between FOMO and loss aversion. This research, examining loss aversion as it correlates to and is moderated by FOMO, has rarely been tested, with few studies analyzing a coalescence between the two variables, let alone involving the cohort of Gen Z teenagers, making this study notable in both behavioral economics and social science realms.

Population

Responses were collected from English students at CHS, a large, high-performing public high school in suburban Los Angeles with around 1,700 students. The goal was to secure between 80 and 100 Gen Z teenage respondents, but the study over-delivered, and a final sample of 131 respondents completed the survey. The ethnic breakdown of CHS is as follows: 73.4% White, 11.3% Hispanic, 5.6% Asian, 3.1% African American, 5% Two or more races, 1.6% Other. CHS is located in a high-income area; however, 23.4% of students identify as socioeconomically disadvantaged. CHS is a co-educational high school with grades 9-12. These demographics of CHS are important to note, as they can differ from other schools, and most certainly from other literature, in terms of less diversity or socioeconomic disadvantage.

By focusing on all grade levels 9-12, the researcher ensures an attempt at equal participation among grades, testing a range of the Gen Z teenage cohort that included ages as young as 14 and as old as 18 (birth years 2006-2012). The researcher analyzed FOMO level, potential loss aversion behavior, and how the data might relate to other patterns seen in previous studies. This cohort of Gen Z teenagers is vastly different from the foundational literature due to other samples of participants being older, like Kahneman's & Tversky's graduate students or Przybylski's 22-62 year-olds.

Instruments

An Informed Consent Form was included prior to the survey being administered to ensure the researcher was following ethical research practices and had the participants' parents'/guardians' consent to participate if they were under 18 years of age (Appendix A).

Once the Informed Consent Form was signed, the participant was prompted by the FOMO section of the survey, consisting of 10 questions with a 5-point Likert Scale to quantify the participant's level of FOMO based on the average of the answers given. The questions in the FOMO section all come from Przybylski et al.'s paper published in 2013, which validated the FOMO Scale across the research community. These questions were designed to replicate likely scenarios that may or may not elicit emotions of FOMO from participants.

Survey Question	Response Options	Source
1. I fear others have more rewarding experiences than me.	Likert Scale 1- Not at all true of me 2- Slightly true of me	Przybylski et al. (2013)

	3- Moderately true of me 4- Very true of me 5- Extremely true of me	
2. I fear my friends have more rewarding experiences than me.	Likert Scale 1- Not at all true of me 2- Slightly true of me 3- Moderately true of me 4- Very true of me 5- Extremely true of me	Przybylski et al. (2013)
3. I get worried when I find out my friends are having fun without me.	Likert Scale 1- Not at all true of me 2- Slightly true of me 3- Moderately true of me 4- Very true of me 5- Extremely true of me	Przybylski et al. (2013)
4. I get anxious when I don't know what my friends are up to.	Likert Scale 1- Not at all true of me 2- Slightly true of me 3- Moderately true of me 4- Very true of me 5- Extremely true of me	Przybylski et al. (2013)
5. It is important that I understand my friends' "in jokes."	Likert Scale 1- Not at all true of me 2- Slightly true of me 3- Moderately true of me 4- Very true of me 5- Extremely true of me	Przybylski et al. (2013)
6. Sometimes, I wonder if I spend too much time keeping up with what is going on.	Likert Scale 1- Not at all true of me 2- Slightly true of me 3- Moderately true of me 4- Very true of me 5- Extremely true of me	Przybylski et al. (2013)
7. It bothers me when I miss an opportunity to meet up with friends	Likert Scale 1- Not at all true of me 2- Slightly true of me 3- Moderately true of me 4- Very true of me 5- Extremely true of me	Przybylski et al. (2013)
8. When I have a good time, it is important for me to share the details online (e.g., updating status).	Likert Scale 1- Not at all true of me 2- Slightly true of me	Przybylski et al. (2013)

	3- Moderately true of me 4- Very true of me 5- Extremely true of me	
9. When I miss out on a planned get-together, it bothers me.	Likert Scale 1- Not at all true of me 2- Slightly true of me 3- Moderately true of me 4- Very true of me 5- Extremely true of me	Przybylski et al. (2013)
10. When I go on vacation, I continue to keep tabs on what my friends are doing.	Likert Scale 1- Not at all true of me 2- Slightly true of me 3- Moderately true of me 4- Very true of me 5- Extremely true of me	Przybylski et al. (2013)

Table 1. 10-item FOMO Scale developed by Przybylski et al. (2013) with a 5-point Likert Scale to quantify participant FOMO level

After answering the 10 questions from the FOMO Scale, participants were asked to complete a series of six hypothetical coin flip questions, using a similar methodology to Kahneman's and Tversky's studies in 1979 and 1992, which are considered the benchmark for this subject of research. This research was conducted with hypothetical values, and no money was given to or collected from respondents. These hypothetical coin flip tests prompted participants with a choice that would be decided by flipping a coin: a hypothetical amount of money would either be gained or lost (e.g., $10 being won or $10 being lost), then participants would choose whether to accept or reject the coin flip. Prior to completing the six questions, participants were given this exact message: "Please choose whether to accept or reject a hypothetical coin flip of different values. This coin flip is understood under the scenario of a fair coin flip, meaning you have a 50-50 chance to achieve your intended decision." These instructions were based on Kahneman and Tversky's 1992 paper, and were included to ensure the results reflected well-informed participants.

While the first three questions followed a symmetrical pattern, having potential gains identical to the corresponding potential loss values (e.g., $10 gain versus $10 loss), the last three each had an asymmetrical pattern, with the potential gain double that of the corresponding potential loss (e.g., $20 gain versus $10 loss). Although influenced heavily by Kahneman and Tversky, this project uniquely tested the difference between symmetric and asymmetric values, correlating strongly with the second hypothesis about loss aversion existing when the potential gain exceeds the potential loss. Additionally, while the questions were influenced by that paper, the researcher created the specific values for each question in this study.

Survey Question	Response Options	Source
11. You can **win $10 or lose $10**. Do you accept this hypothetical coin flip?	Binary Choice - Yes - No	Kahneman, Tversky (1992).
12. You can **win $100** or **lose $100**. Do you accept this hypothetical coin flip?	Binary Choice - Yes - No	Kahneman, Tversky (1992).
13. You can **win $500** or **lose $500**. Do you accept this hypothetical coin flip?	Binary Choice - Yes - No	Kahneman, Tversky (1992).
14. You can **win $20** or **lose $10**. Do you accept this hypothetical coin flip?	Binary Choice - Yes - No	Kahneman, Tversky (1992).
15. You can **win $200** or **lose $100**. Do you accept this hypothetical coin flip?	Binary Choice - Yes - No	Kahneman, Tversky (1992).
16. You can **win $1000** or **lose $500**. Do you accept this hypothetical coin flip?	Binary Choice - Yes - No	Kahneman, Tversky (1992).

Table 2. Six hypothetical coin flip questions with binary responses (yes/no) based on question type: three symmetrical questions with equal potential hypothetical gain and loss values ($10/$10, $100/$100, $500/$500) and three asymmetrical questions with potential hypothetical gains double those of losses ($20/$10, $200/$100, $1000/$500)

Sample Selection

131 students out of approximately 800 who were offered the survey completed it. A Google Form with the survey link was posted in five of CHS's English classes. The remaining six English teachers did not respond to requests about posting the link, and thus were omitted. English is a required subject to take all four years of high school at CHS, so by using this sample, the researcher would have access to a larger and more randomized population while keeping the sample size substantial enough to draw conclusions. Additional respondents were drawn from AP Seminar classes, as those students are also considered English students, so as not to fragment the legitimacy of data collection. The researcher utilized a convenience sampling method, distributing the survey to accessible groups.

Implementation

The survey was solely hosted on Google Forms, accessed by the link posted on students' English/AP Seminar Google Classroom streams. The survey was live for three weeks, and participants who chose to fill out the form had their answers recorded in a Google Sheet

spreadsheet. The researcher was then able to organize the data, analyzing it with tools including the XL DataMiner ToolPak Google Sheets extension in six different linear regressions, a 786-row multiple regression, and a Pearson Correlation test for internal consistency.

IV. FINDINGS

Re-introduction and Hypotheses

This section displays the results of the quantitative analysis implemented to address the hypothesized correlation between loss aversion and FOMO. Before data collection, the researcher developed two key hypotheses. Firstly, Gen Z audiences with lower FOMO would be more likely to exhibit loss aversion than those with higher FOMO. Secondly, this effect would be more pronounced in asymmetric hypothetical coin flip scenarios in which the potential gain exceeds the potential loss. The findings are organized into the following categories: FOMO Scores, Hypothetical Coin Flip Acceptance Rates by Question, Hypothetical Coin Flip Acceptance Rates by FOMO Level, and Multiple Regression Results.

Findings: FOMO Scores

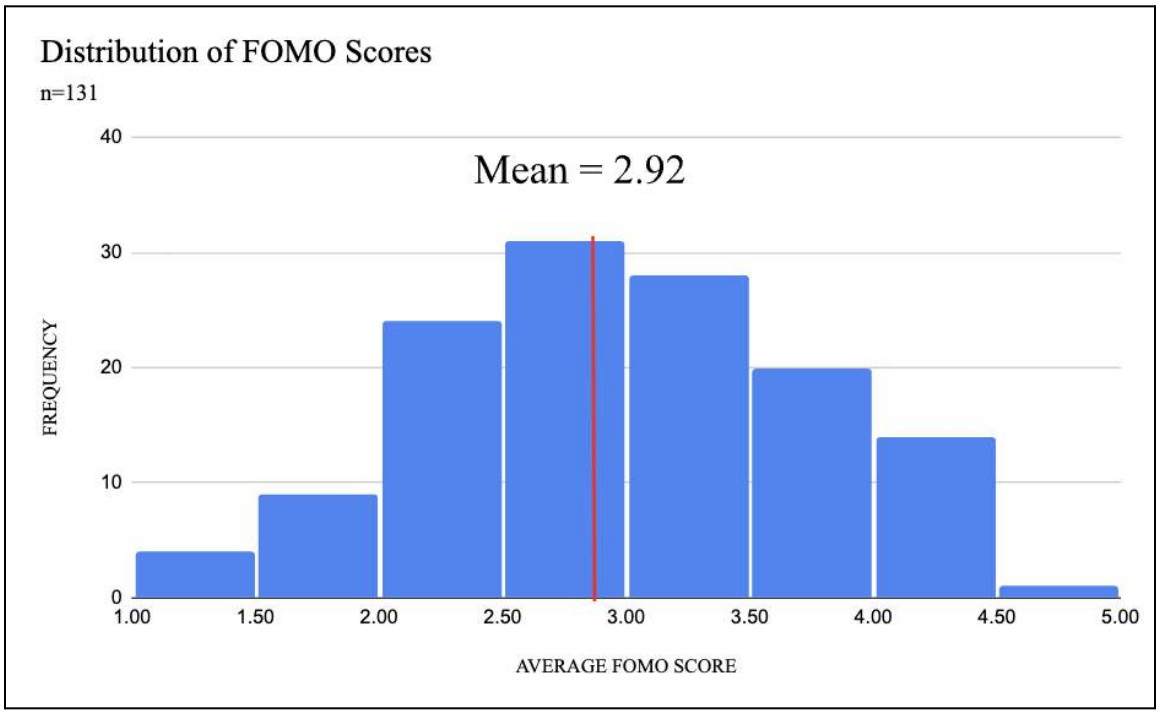


Figure 1. A histogram displaying the distribution of average FOMO scores across CHS participants (n=131)

The average FOMO scores, which were determined by averaging each participant's response to the 10-item FOMO Scale, measured with a 5-point Likert scale, show a normal distribution, with a majority of participants' FOMO levels averaging between 2.0 and 3.5 out of

5.0. The mean score was 2.92, highlighting a moderate level of FOMO slightly below the scale's midpoint of 3.0. Very few participants had extremely high or low scores, demonstrating a lack of skewed distribution.

Question	R	p	Strength of Correlation
Q1 & Q2	0.733	<0.001	Strong
Q7 & Q9	0.659	<0.001	Strong
Q15 & Q16	0.496	<0.001	Moderate
Q12 & Q15	0.442	<0.001	Moderate
Q11 & Q14	0.417	<0.001	Moderate

Table 3. Pearson Correlation results summarizing internal consistency among surveyed items (n=131)

A Pearson Correlation was conducted to verify internal correlations among the FOMO Scale and loss aversion questions. Seen in Table 3, strong correlations were found between FOMO Scale questions 1 & 2 ($r=0.733$, $p<0.001$) and 7 & 9 ($r=0.659$, $p<0.001$). Moderate correlations were found between loss aversion questions 15 & 16 ($r=0.496$, $p<0.001$), 12 & 15 ($r=0.442$, $p<0.001$), and 11&14 ($r=0.417$, $p<0.001$). Weak correlations were not included.

Findings: Hypothetical Coin Flip Acceptance Rate by Question

Symmetric Asymmetric

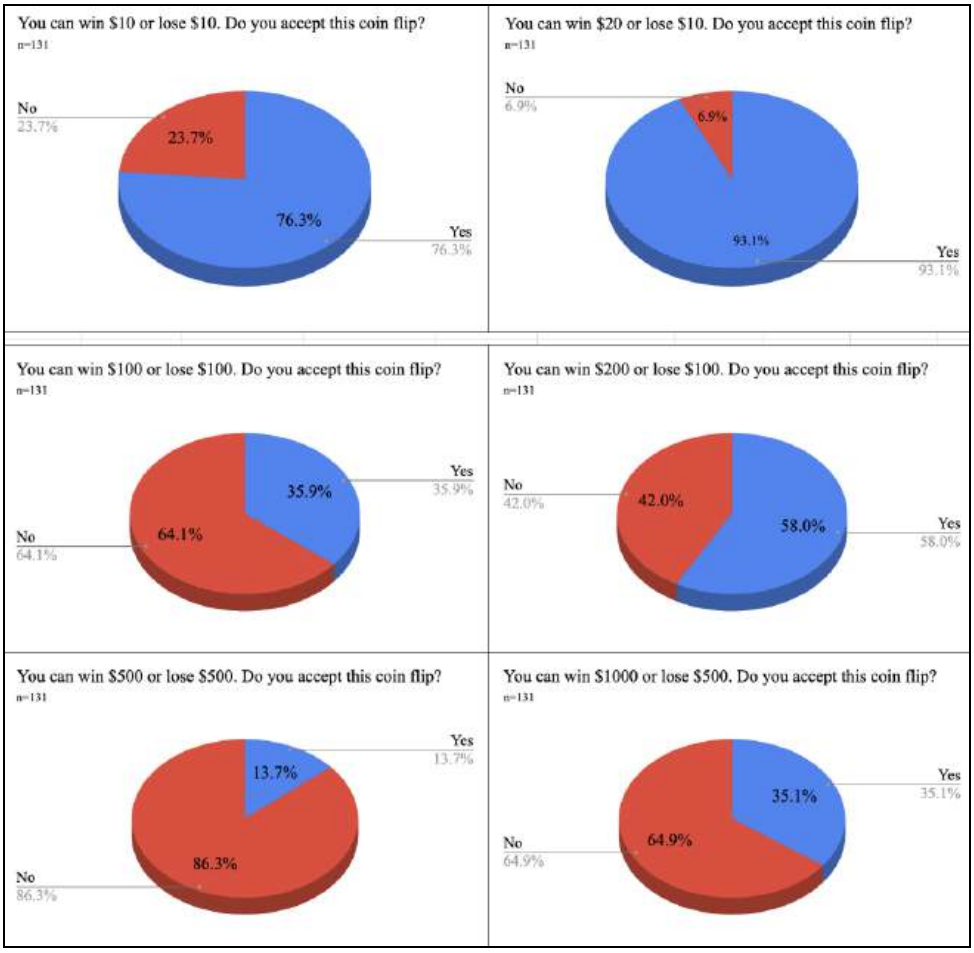


Figure 2. Pie charts displaying hypothetical coin flip acceptance versus rejection rates for each of the six loss aversion questions among CHS participants (n=131)

The acceptance versus rejection rates varied significantly across the six hypothetical coin flip questions, with differences emerging within and between the symmetric and asymmetric questions. In the symmetric questions, where the potential gains and losses were equivalent (left column), acceptance rates declined as the stakes, the hypothetical values of the wins or losses, increased. 76.3% of participants accepted the $10 gain versus $10 loss hypothetical coin flip (Q1), but only 35.9% of participants accepted the $100/$100 hypothetical coin flip (Q2), and further, only 13.7% of participants accepted the $500/$500 hypothetical coin flip (Q3). Asymmetric questions, where potential gains were double that of potential losses (right column), followed a similar declining pattern. 93.1% of participants accepted the $20 gain/$10 loss hypothetical coin flip (Q4), but only 58% of participants accepted the $200/$100 (Q5), and further, only 35.1% of participants accepted the $1000/$500 hypothetical coin flip (Q6). Although symmetric and asymmetric hypothetical coin flips followed a similar declining acceptance pattern in stakes, the percentage of acceptance in asymmetric questions was notably higher. CHS students were split in acceptance of the hypothetical coin flips, with the overall results showing three majority hypothetical coin flips accepted (questions 1, 4, 5) and three rejected (2, 3, 6).

Findings: Hypothetical Coin Flip Acceptance Rate by FOMO Level

Question	Type	Stakes	P-value	Significant?
Q1	Symmetric	$10/$10	0.022	Yes
Q2	Symmetric	$100/$100	0.121	No
Q3	Symmetric	$500/$500	0.105	No
Q4	Asymmetric	$20/$10	0.003	Yes
Q5	Asymmetric	$200/$100	0.294	No
Q6	Asymmetric	$500/$1000	0.475	No

Table 4. Summary of six individual linear regressions examining the FOMO score as a predictor of loss aversion in hypothetical coin flip scenarios (n=131)

Six individual linear regressions were conducted, one for each hypothetical coin flip question, with the FOMO scores as the independent variable and the loss aversion response as the dependent variable. Demonstrated in Table 4, statistical significance only emerged in two out of six questions: Q1, the symmetric $10 hypothetical coin flip, with a p-value of 0.022, and Q4, the asymmetric $20/$10 hypothetical coin flip, with a p-value of 0.003. Notably, both of these significant results occurred at the lowest stakes level within their respective categories. As dollar amounts increased across both symmetric and asymmetric questions, the relationship between FOMO and loss aversion became progressively weaker and insignificant.

Findings: Multiple Regression Results

Variable	Coefficient	P-value	Significant?
FOMO Score	0.105	0.002	Yes
Question Type	0.227	0.061	Slightly
Loss Amount	-0.0001	0.614	No
FOMO x Question Type	-0.009	0.822	No
FOMO x Loss Amount	-0.0003	0.001	Yes

Table 5.

Summary of multiple regression analysis with coefficients examining FOMO Score, question type, stakes, and interaction variables to predict loss aversion (n=131)

To conduct a more rigorous analysis of the potential interaction between FOMO and loss aversion, the data were restructured into a multiple regression. 786 observations were created by analyzing the 131 participants across the six hypothetical coin flip scenarios. The overall model had statistical significance with a p-value <0.01 and an r^2 value of 0.246, explaining 24.6% of the variance in loss aversion responses. As seen in Table 5, two of the five variables were statistically significant with p-values less than 0.005. FOMO score positively predicted loss aversion, specifically for every one-point increase in FOMO score, loss aversion increased by 0.105. Secondly, the interaction between FOMO and loss amount was significant, however, it

came back with a negative coefficient, highlighting that as dollar amounts increase, FOMO's relation to loss aversion decreases to such an extent that it completely reverses. The other three variables were disproved with little to no statistical significance.

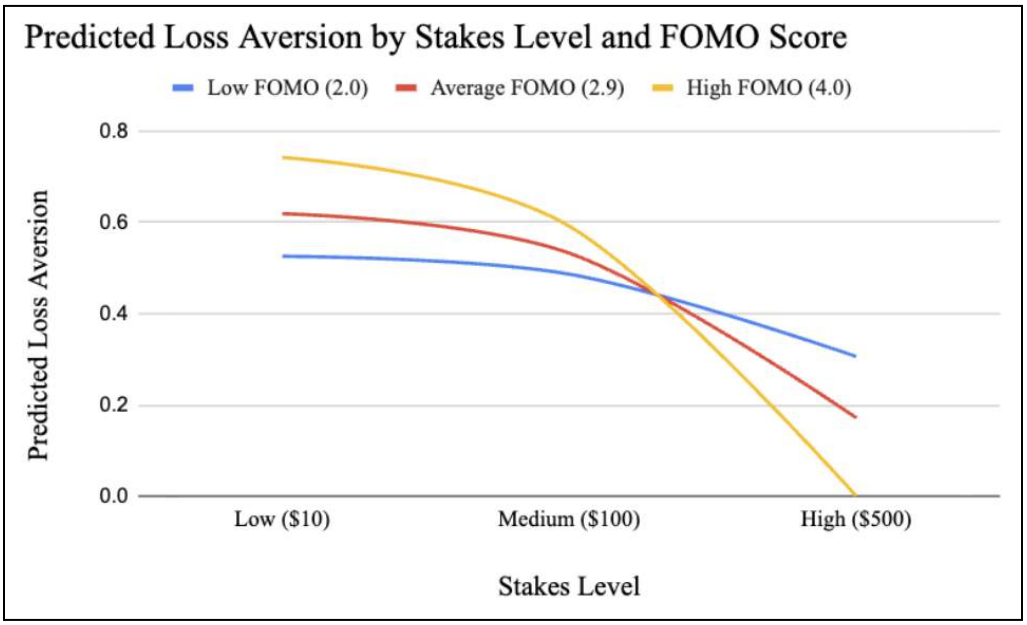


Figure 3. Line graphs displaying the predicted loss aversion (accepting a hypothetical coin flip) by dollar amount and FOMO level (n=131)

Further exploring the impact stakes have on loss aversion, using the data from the multiple regression and subsequent segmentation, Figure 3, above, determined through an additional regression, shows that as stakes increased, the percentage willing to accept a coin flip (the magnitude of loss aversion) declined across all three levels of FOMO. The high FOMO group, identified with a score of 4.0 out of 5.0 on the FOMO Scale, showed the steepest decline, reaching almost 0 when high stakes were presented, versus the low FOMO group, identified with a score of 2.0, that experienced a more steady decline.

V. ANALYSIS

The following analysis delves into the findings of this study in relation to the two hypotheses developed before implementation, based on previous literature.

Analysis of Core Data

The mean FOMO score of 2.92, as seen in Figure 1, aligns with Przybylski et al.'s (2013) interpretation of FOMO as represented with a relatively normal distribution among a population, validating the cohort of Generation Z teenagers who participated in the survey, along with their FOMO levels. Additionally, when examining symmetric questions, there was a decline in hypothetical coin flip acceptance as stakes increased, declining from 76.3% in the $10/$10 coin flip to 35.9% in the $100/$100 coin flip, and finally to 13.7% in the $500/$500 coin flip. These

results reinforce Kahneman's and Tversky's "Value Function" model, developed in their 1979 paper, explaining how sensitivity to losses diminishes when stakes increase, which may be why participants were hesitant to accept risks at high stakes. Higher acceptance rates exist for asymmetric questions, suggesting that when potential gains are double that of potential losses, loss aversion may be reduced, partially consistent with the researcher's second hypothesis, but not statistically significant. Additionally, the Pearson Correlation, seen in Table 3, validates the instruments of this research, showing strong internal correlations within the FOMO Scale, specifically questions 1 & 2 and 7 & 9, both having p-values less than 0.01 and r-values greater than 0.650. These results validate the FOMO Scale by Przybylski et al. (2013), showing it is a reliable tool to analyze FOMO.

Analysis of Regression Data

The results of the six individual linear regressions (Table 4) uncovered that FOMO only predicted loss aversion, in a statistically significant manner, in two out of six scenarios, the symmetric $10 question and the asymmetric $20/$10 question, while no statistical significance emerged in the other four. What is notable about these two scenarios is that both are the lowest stakes level in their respective categories. As the stakes increased, the impact FOMO had on loss aversion began to decrease, once again supporting the value function (Kahneman & Tversky, 1979). This observation in Table 4 also supports Gal and Rucker's work from 2018, showcasing that loss aversion, indeed, does fluctuate based on context, specifically in this case, the psychological trait of FOMO.

While the linear regressions allowed the researcher to make hypotheses as to why the data were behaving in such a manner, the multiple regression validated the speculations. The first x-variable, FOMO score, was able to confirm one of the foundations of the study, namely FOMO's impact on loss aversion, and with a coefficient of 0.105 and a p-value of 0.002, FOMO was confirmed to positively predict the phenomenon.

However, an even more significant finding emerged with the fifth x-variable, FOMO multiplied by loss amount, showing a coefficient of -0.0003 and a p-value of 0.001. This statistically significant finding directly contradicts the researcher's first hypothesis, as when hypothetical dollar amounts increase, FOMO's effect on loss aversion actually declines so significantly that it completely reverses. This refutes the researcher's first hypothesis, which predicted that those with lower FOMO would have higher loss aversion due to not feeling the need to be adventurous and willing to miss out on winning, as it instead demonstrated that higher FOMO predicted higher loss aversion. Once again, the behavior seen follows that of the value function as FOMO's influence follows a similar pattern of decline when stakes increase.

Additionally, the lack of a significant fourth x-variable, FOMO multiplied by question type, disproves the researcher's second hypothesis as well. The researcher's second hypothesis was that the loss aversion effect would be more pronounced in asymmetrical scenarios than in symmetrical ones, but due to a p-value of 0.822, this hypothesis was rejected, showing that

symmetric versus asymmetric questions were overshadowed by the large effect of how FOMO and loss aversion interact.

Conclusion to Analysis

The results of this study rejected both of the researcher's proposed hypotheses. While the data showed a statistically significant relationship between FOMO and loss aversion, it nevertheless showed the relationship in the complete opposite direction of hypothesis number one: high FOMO predicted higher loss aversion, versus the converse. Secondly, the idea that loss aversion would be more pronounced with asymmetric versus symmetric questions was rejected due to a lack of significance in the interaction variable, x-variable number four. Although both hypotheses were rejected, a new and unique finding emerged, being that the stakes level was the true moderator of FOMO and loss aversion, a variable undocumented in any previous literature. However, previous literature was still helpful, especially Kahneman's and Tversky's value function, in analyzing FOMO's declining influence as stake levels rose.

VI. CONCLUSION

Purpose

Loss aversion has been a heavily studied concept since Kahneman and Tversky's foundational research in 1979. Simultaneously, FOMO has recently become a prevalent phenomenon due to Generation Z's social media-induced habits, and through Przybylski and colleagues' research, FOMO has become a measurable metric. These two variables are significant drivers of decision-making on their own; however, what makes them so interesting is their interaction, yet no prior research has examined this with the target demographic of Generation Z teenagers. This study served the purpose of testing this interaction and its magnitude, utilizing the methodology from both Kahneman's and Tversky's 1979 and 1992 papers, along with Przybylski et al.'s (2013) FOMO Scale. This study hoped to determine if FOMO level predicted loss aversion among Gen Z teens and how that relationship changed based on hypothetical coin flip values.

Unexpected Finding

The data uncovered a finding that the researcher did not consider when developing the hypotheses for the experiment, namely, the involvement of stakes. The findings reinforced that stakes were the true moderator of the relationship between FOMO and loss aversion, a major breakthrough in both fields of research.

Implications

Beyond the numbers, there were 3 significant implications uncovered in this study. Firstly, low-stakes financial decisions among Gen Z may be more emotionally driven than previously understood, as when the hypothetical dollar amounts are small, phenomena like

FOMO arise to impact decision-making. This can be explained with the value function, which predicted the higher sensitivity to losses and gains at low stakes. There are educational implications from this point, namely regarding FOMO as a potential cause of poor financial decision-making, which should be included as part of financial literacy education. Secondly, FOMO reaches beyond its traditional social context into the economic world, amplifying emotions that impact cognition regarding economic decisions, not just social behavior. Even though FOMO may be seen as an emotional framework, it has financial implications, impacting decision-making through a statistically significant interaction with loss aversion. This finding offers significant direction for economists, who should incorporate psychological variables like FOMO, alongside usual variables like income, into their research. Finally, Gen Z teenagers proved to be a valid cohort for behavioral economics research, demonstrating predictable patterns that proved meaningful, giving sufficient evidence as to why the cohort, among others, should be further analyzed in behavioral economics research, especially as each generation may have individual traits that produce different results in loss aversion testing.

Limitations

Although successful, this study still contained notable limitations worth discussion. Firstly, this data was taken from a single population and may not be wholly representative of Gen Z teenagers. While CHS is somewhat diverse, it may not be diverse enough to project the results to an entire generation. This claim specifically pertains to the makeup of CHS, which skews more Caucasian and socioeconomically advantaged than the total US population, potentially cultivating different views of money and its value (U.S. Census Bureau, 2024). These skews might have driven different decisions regarding whether to accept or reject a hypothetical coin flip. Additionally, there is the possibility of self-report bias in the survey since both instruments relied entirely on self-reported answers, and therefore, participants could have acted out of social desirability bias or embarrassment, being influenced to choose answers that were not true reflections of their mindset. Finally, the binary "yes or no" format of the hypothetical coin flip questions was efficient in showing whether loss aversion exists in relation to FOMO, however a scaled response may have demonstrated the strength of their commitment to their decision in a more nuanced way, providing even more compelling results.

Next Steps

A few directions for future research stand out. First, replicating this study with real financial incentives rather than hypothetical scenarios would test whether the stakes-dependent pattern stays consistent. However, due to concerns surrounding teenage gambling, this idea could create negative public health outcomes, therefore, future researchers need to be aware of this and structure their methodology carefully. Secondly, examining a different age cohort would reveal whether the FOMO and loss aversion relationship is unique to Gen Z teenagers or extends into older generations. Third, examining how social media usage levels moderate FOMO's effect on loss aversion could add clarity since FOMO and social media are deeply intertwined.

Works Cited

Benartzi, S., & Thaler, R. H. (1995). Myopic Loss Aversion and the Equity Premium Puzzle. The Quarterly Journal of Economics, 110(1), 73–92. https://doi.org/10.2307/2118511

Einstein, D. A., Dabb, C., & Fraser, M. (2023). Fomo, but Not Self-Compassion, Moderates the Link Between Social Media Use and Anxiety in Adolescence. Australian Journal of Psychology, 75(1). https://doi.org/10.1080/00049530.2023.2217961

Elhai, J. D., Yang, H., & Montag, C. (2021). Fear of Missing Out (FOMO): Overview, Theoretical Underpinnings, and Literature Review on Relations with Severity of Negative Affectivity and Problematic Technology Use. Brazilian Journal of Psychiatry, 43(2), 203–209. https://doi.org/10.1590/1516-4446-2020-0870

Gal, D., & Rucker, D. D. (2018). The Loss of Loss Aversion: Will it loom larger than its gain? Journal of Consumer Psychology, 28(3), 497–516. https://doi.org/10.1002/jcpy.1047

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. Econometrica, 47(2), 263. https://doi.org/10.2307/1914185

Nagaya, K. (2021). Why and Under What Conditions Does Loss Aversion Emerge? Japanese Psychological Research, 65(4), 379–398. https://doi.org/10.1111/jpr.12385

Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, Emotional, and Behavioral Correlates of Fear of Missing Out. Computers in Human Behavior, 29(4). https://doi.org/10.1016/j.chb.2013.02.014

Rahmawati, U., & Raharja, S. (2024, December 31). The Influence of Herding, Loss Aversion, and Availability on Investment Decision-Making with Fear of Missing Out as a Mediating Variable Among Generation Z Investors. Indonesian Interdisciplinary Journal of Sharia Economics (IIJSE). https://e-journal.uac.ac.id/index.php/iijse/article/view/5608

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. Journal of Risk and Uncertainty, 5(4), 297–323. https://doi.org/10.1007/bf00122574

U.S. Census Bureau. (2024). Quickfacts: United States. https://www.census.gov/quickfacts/fact/table/US/PST045224

Appendix A: Informed Consent Form

The Interaction Between Loss Aversion and FOMO Among Gen Z Teenagers

Your child is invited to take part in a research study on loss aversion and FOMO being conducted by NS, who is a student in the AP Capstone program at Calabasas High School. The study, as well as your rights as a participant, are described below.

Description: This study will examine loss aversion with FOMO as a moderator. Specifically, a survey will be conducted measuring FOMO level through a 10-item scale that has been recognized by the FOMO research community as the universal FOMO scale, along with a loss aversion test to see how participants react when faced with test "coin flip" situations, both symmetric and asymmetric. Follow-up interviews will not be needed. Your child's identity will not be revealed to anyone but the principal investigator(s) and designated research associates.

Confidentiality: Children's answers will not be associated with their names. Rather, each child will be given an identification number on the interviewer's sheet.

Freedom to Withdraw or Refuse Participation: I understand that my child has the right to not participate in this study. If harms occur, the principal investigator will call 911.

Questions? Please feel free to ask the investigator any questions before, during, or after the study.

Faculty Supervisor & Principal Investigator:

Brad Boelman

AP Capstone Coordinator

Room H-111

[redacted]High School, [redacted]CA

818-222-7177

* Indicates required question

I, _____, (parent/guardian name) give permission for my child, _____ to participate in the research project entitled, "The Intersection Between Loss Aversion and FOMO" The study has been explained to me and my questions answered to my satisfaction. I understand that my child's right to withdraw from participating or refuse to participate will be respected and that his/her responses and identity will be kept confidential. I give this consent voluntarily. *

Your answer

I, _____, (student's name) give consent to participate in the research project entitled, "The Intersection Between Loss Aversion and FOMO." The study has been explained to me and my questions answered to my satisfaction. I understand that my right to withdraw from participating or refuse to participate will be respected and that my responses and identity will be kept confidential. I give this consent voluntarily. *

Your answer

Understanding the Entrepreneurial Path of Women-Owned Businesses: Bakeries in Philadelphia By Elaine Guo

Abstract

This paper investigated the entrepreneurial path of women-owned bakeries in Philadelphia through a data-driven case study. By combining quantitative data from online reviews with qualitative insights from in-depth interviews, this study uncovered the unique challenges and strategies employed by women entrepreneurs in a traditionally male-dominated industry. Key themes such as emotional connections, community involvement, and innovation emerged as critical factors in their success. The case study examined how these women business owners integrated tradition with modern business practices to navigate financial pressures, market competition, and work-life balance. Examples like Sweet Nina's innovative product development and The Famous 4th Street Cookie Company's community-centric philanthropy illustrated the diverse ways in which these businesses maintained authenticity while adapting to evolving consumer demands. This analysis contributed to the broader discourse on women's entrepreneurship, highlighting how women-owned businesses in the food sector overcame systemic challenges through resilience, community engagement, and continuous innovation. The findings provided valuable insights for understanding the dynamics of women-led businesses and offered practical lessons for future entrepreneurs aiming to succeed in similar competitive environments.

Introduction

In recent years, women have made significant strides in industries where they have historically been underrepresented, particularly in entrepreneurship. One such industry is the bakery sector, which, although often perceived as less complex, presents a unique set of challenges and opportunities. The purpose of this study was to investigate how women succeed in a male-dominated industry and to examine how they navigate the unique challenges of running a bakery. Recognizing the creativity, resilience, and innovative approaches of women bakery owners, this research sought to understand the factors contributing to their success and the obstacles they face. Both the qualitative and quantitative aspects of women-owned businesses were delved into under this study, combining data from online customer reviews and face-to-face interviews with female bakery owners. Through a comprehensive data-driven analysis, this study found that emotional connections and community engagement shed light on the entrepreneurial path of women-owned bakeries in Philadelphia, offering insights that can inform future business ventures and academic discussions. This study also aimed to contribute to the growing discourse on women's entrepreneurship, emphasizing the unique strengths and strategies employed by women in overcoming challenges and achieving success in traditionally male-dominated fields.

The central focus of this research was to uncover what made women-owned bakeries successful and how they maintained authenticity while adapting to market demands. Several key themes were explored, including the role of community, the importance of customer

relationships, and the ability to innovate in a highly competitive environment. Based on the analysis of these themes, advice was provided not only for current and aspiring female entrepreneurs but also to a broader audience on the understanding of women's entrepreneurship.

The strategy of this research was as follows: First, a thorough review of the relevant literature on women entrepreneurs, particularly in the food industry, was conducted to contextualize the study. Then a mixed-method approach was used to collect and analyze both quantitative and qualitative data. The quantitative data was analyzed to grasp the general trend, followed by designing the qualitative study on representative cases. The key findings from the data analysis, including insights from interviews and patterns found in customer reviews were presented. Based on these findings, reflections were highlighted on the broader impact of this research and recommendations for future studies on women-owned businesses were offered.

Literature Review

In the existing literature on women-owned businesses, several themes emerged, particularly in the context of the food and bakery sectors. Key themes such as resilience, adaptability, and community engagement were consistently emphasized as critical components of success for women entrepreneurs. These themes were representative of broader trends in female entrepreneurship, demonstrating how women navigated traditionally male-dominated industries, adapted to challenges, and innovated to sustain their businesses.

The Wells Fargo's 2026 report of "The Impact of Women-Owned Businesses" (Wells Fargo, 2026) showed that women-owned businesses comprised 40.6% of all businesses in the US, accounted for 9.1% of the national workforce and 4.6% of total firm revenue nationally. Although facing persistent gender bias and scaling challenges, women entrepreneurs contributed new approaches to innovation, especially in purpose-driven, community-focused, and flexibility-oriented business models. The "Annual Women-Owned Business Study 2025" from Biz2Credit (Biz2Credit, 2025) reported a higher revenue growth rate in women-owned businesses and showed women-owned businesses were increasing their stake in the small business industry. The LinkedIn publication of "Turning Influence into Impact: Women Entrepreneurs in the Food Sector and Grow to Market's Mission" pointed out that women bring deep knowledge of consumer needs, as "for many women in the food industry, entrepreneurship begins with real experience" (Grow to Market, 2026). In "Women, power & progress: Shaping the next generation of business" (Hyslop, 2025), examples of women entrepreneurs were highlighted who prioritized not only profit but purpose, designed cleaner, healthier, and special snacks such as gluten-free and allergen-friendly food product.

Rahman (2025) in "Empowering Women Food Entrepreneurs: Unveiling Their Social Commerce Readiness and Digital Challenges" revealed strong positive correlations between women's engagement in social commerce and their overall business readiness. This article pointed out that women entrepreneurs were motivated and capable, but they need technological and financial support.

As a women entrepreneur, Brkic (2025) in “From Young Professional to Entrepreneur” demonstrated how smaller ventures like her food startup were able to integrate mission-driven rapid innovations directly into their core products and business models, while larger companies were slow in adapting changes due to internal politics and risk-averseness.

One exemplary study by Chan (2024), titled "The Chinese Dumpling Revolution: MìLà's Journey to Disrupt the Frozen Food Industry," explored how an Asian American women-owned business, MìLà, successfully entered the highly competitive frozen food market. The business thrived by embracing cultural authenticity and leveraging consumer appreciation for traditional values. Chan (2024) described how MìLà shipped over 20 million dumplings by utilizing a combination of market research and consumer feedback. The company's success was further amplified through strategic marketing, which included endorsements from celebrities like Simu Liu and appearances on popular platforms such as Jimmy Kimmel Live. As Chan (2024) noted, “MìLà’s ability to sense customer preferences and secure influential endorsements reflects the dynamism required in today’s competitive market”. This case highlighted how women entrepreneurs in the food industry blended authenticity with innovative marketing strategies to create a lasting impact.

Similarly, Atchley (2024), in "The Women's Bakery Looks to Empower, Educate, and Employ Women," provided a comprehensive analysis of The Women’s Bakery in Rwanda. The business not only addressed issues such as hunger and gender inequality but also contributed significantly to local economic growth by employing and educating women through baking. Atchley (2024) emphasized that founder Markey Culver’s vision demonstrated how women-owned businesses can have a profound impact at both local and global levels. This example showcased the broader societal benefits that women-owned businesses can provide, positioning them as pivotal drivers of economic and social change, particularly in underdeveloped regions. The literature consistently pointed to community engagement and strategic marketing as essential drivers of success for women-owned businesses, particularly in the bakery and food sectors. These businesses often harness innovation to maintain competitiveness while fostering inclusivity and social responsibility. By focusing on community ties and customer relationships, women entrepreneurs in these sectors created loyal customer bases that not only supported their businesses but also contributed to local economic development. This is particularly true in socially responsible businesses like The Women’s Bakery, where entrepreneurship was seen as a pathway to achieving broader social goals (Atchley, 2024).

Fazio’s (2024) work, "Breaking Barriers, Fostering Growth: The Evolution of Black Women-Owned Businesses in the U.S.," examined how black women entrepreneurs, such as Stephanie Hart of Brown Sugar Bakery, used their businesses to address social issues and create economic opportunities. Fazio (2024) emphasized the resilience of black women entrepreneurs, particularly in the context of the COVID-19 pandemic, where new business ventures emerged in response to economic necessity. The pandemic catalyzed a surge in black-owned businesses, which reflected the adaptability and innovation inherent in these women’s entrepreneurial

journeys (Fazio, 2024). The resilience of black women in entrepreneurship was further supported by data presented by Wells Fargo (2024), who found that between 2019 and 2023, women-owned businesses grew by 13.6%, with a growth rate 82.4% faster than that of men-owned businesses. In the report of “The 2025 Impact of Women-Owned Businesses”, the same trend was observed (Wells Fargo, 2025). These statistics demonstrated the expanding role of women in entrepreneurship and underscored their growing economic influence in the U.S.

Another critical aspect discussed in the literature was the role of digitalization and networking in supporting women entrepreneurs. Heald-Fisher (2010), in "Female Entrepreneurs Making Connections: Networking by Women Small Business Owners," outlined the importance of effective networking for women business owners. She highlighted how networking enabled women to access crucial resources, partnerships, and sponsorships, thus significantly benefiting their businesses (Heald-Fisher, 2010). This capacity to foster strong networks, both digitally and within local communities, was a key element in the success of women-owned businesses.

Furthering this discourse, Olsson (2020), in her article "Keeping up the Pace of Digitalization in Small Businesses: Women Entrepreneurs' Knowledge and Use of Social Media," discussed how women entrepreneurs utilized digital tools to manage their businesses. Olsson (2020) emphasized that, despite limited resources and work stress, women were leveraging platforms like social media to effectively promote their businesses, expand their reach, and engage with customers. The flexibility provided by these digital tools enabled women entrepreneurs to manage multiple aspects of their businesses efficiently, ensuring that they remained competitive in an increasingly digitized economy. In a survey conducted by National Association of Women Business Owners (NAWBO), 81% of NAWBO members were using AI (Artificial Intelligence), primarily chatbots such as ChatGPT (NAWBO, 2025).

Carter (2001), in her review titled "Women’s Business Ownership: A Review of the Academic, Popular, and Internet Literature," reinforced these findings by noting that "women-owned businesses are growing at a rapid pace, capturing a significant share of sales and employment, and are essential for breaking stereotypes in traditionally male-dominated industries". Carter (2001) highlighted the dual role women entrepreneurs play, both in contributing to economic growth and in challenging traditional gender norms. This dual contribution placed women entrepreneurs at the forefront of economic and social change, making their contributions indispensable to both the business world and society at large.

In summary, the literature surrounding women-owned businesses, particularly in the food and bakery sectors, exhibited recurring themes of resilience, adaptability, community engagement, and digitalization. The success of these women entrepreneurs lied in their ability to blend cultural authenticity with modern business strategies, create strong community ties, and leverage digital platforms to expand their influence. Their businesses not only contributed to economic development but also challenged existing gender stereotypes and addressed broader social issues, such as hunger, inequality, and job creation. Through these combined efforts, women entrepreneurs were transforming industries and reshaping the entrepreneurial landscape.

Data and Methods

This study employed a mixed-methods approach, integrating quantitative data from online reviews and qualitative data from interviews to provide a comprehensive understanding of the entrepreneurial journeys of women-owned bakeries in Philadelphia. By collecting and analyzing both types of data, the research captured customer perspectives as well as personal insights from bakery owners, offering a full picture of the successes and challenges faced by these women entrepreneurs.

The quantitative data was primarily collected from Google and Yelp. While Google is used by a vast population of consumers worldwide, Yelp is a widely used platform for customer reviews of local businesses. In addition to the large number of ratings from Google, Yelp pages provided valuable information, including addresses, star ratings, reviews, and operational details. This data served as an essential source of customer feedback, offering insights into broad patterns and trends related to customer satisfaction, product quality, customer service, and the bakeries' roles within their communities. Analyzing these reviews allowed the identification of key themes that reflected the customer experience across various bakeries, providing a statistical overview of customer satisfaction. Additionally, Yelp reviews offered important operational details that contextualized customer feedback, such as locations, prices, business hours, and delivery options, which were critical for understanding the overall customer experience.

For example, the Yelp page for Sweet Nina's Bakery (Figure 1) highlighted key elements of the bakery's identity, such as its star rating, customer reviews, surrounding attractions, and user-uploaded photos of popular items like their signature strawberry banana pudding. This platform played a pivotal role in promoting community engagement and customer interaction, helping local businesses like Sweet Nina's build credibility and visibility. In this way, Yelp provided not only quantitative data but also a window into customer perceptions and interactions.

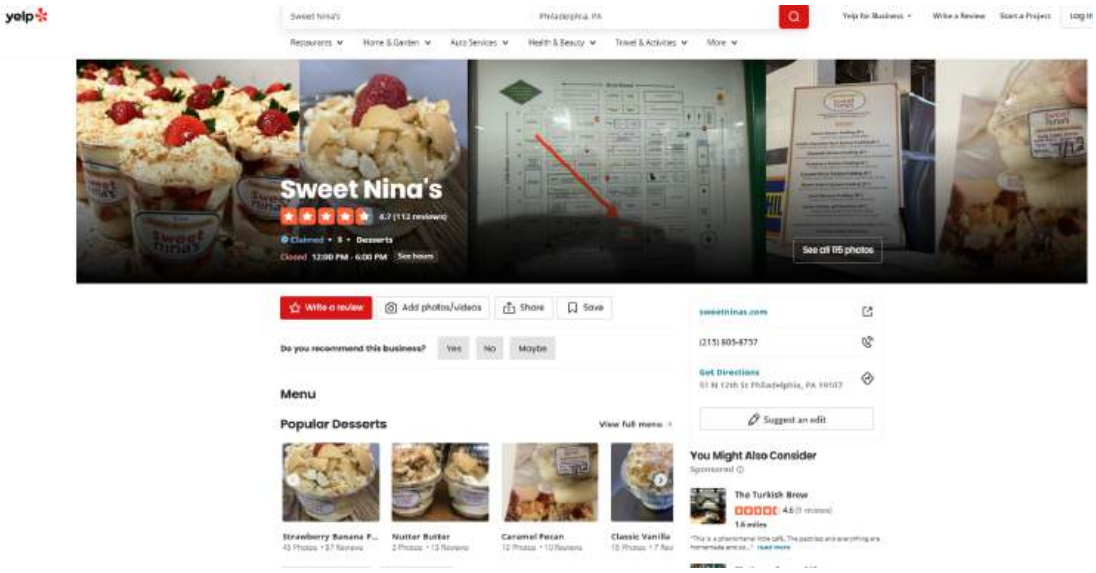


Figure 1. The Front Page of a Restaurant on Yelp.

In contrast, qualitative data was gathered through in-depth interviews with bakery owners, providing richer insights into the personal and professional aspects of running a bakery. The interviews were semi-structured, with open-ended questions designed to explore three main areas: the origin story of the bakery, the entrepreneur's identity, challenges encountered as a female business owner, and the bakery's engagement with the local community. These interviews offered the opportunity to deeply explore the motivations and strategies of these women entrepreneurs. The selection of bakeries for interviews followed a strategic sampling process, ensuring a diverse set of bakeries in terms of location, customer reviews, and business scale. Yelp reviews played an important role in identifying bakeries with varied customer feedback, ensuring that both highly rated and more mixed-reviewed businesses were included. This sampling approach provided a well-rounded perspective on the range of experiences among women bakery owners. The qualitative data, derived from interviews, captured more nuanced aspects of the bakery owners' personal journeys, highlighting themes such as family traditions, emotional connections with customers, and community engagement.

Several tools were used to aid in data analysis and visualization. First, a map was generated using QGIS (Quantum GIS) – a piece of Geographic Information Systems (GIS) software, which visually represented the locations of the bakeries across Philadelphia. This helped plan field visits and ensured that the selected bakeries represented a geographically diverse sample. Additionally, a word cloud was created using Natural Language Processing (NLP) tools to identify frequently mentioned terms in Yelp reviews and interview transcripts. Words such as "women-owned", "delicious", "community", and "bakery" appeared frequently, indicating common themes valued by both customers and bakery owners.

Together, the quantitative analysis of Google rating and Yelp reviews revealed broad patterns such as customer preferences for high product quality and strong customer service, while the qualitative interviews provided deeper insights into the challenges bakery owners faced, such as balancing business growth with personal life and adapting to market competition. It is worth noting that relying solely on Yelp reviews would have limited the study to surface-level observations, while focusing only on interviews might have missed broader market trends. By combining both quantitative and qualitative data, the research not only quantified the factors contributing to the success of these bakeries but also offered a deeper understanding of how women entrepreneurs navigated and overcame challenges in their business journeys. This integration of data types allowed for a richer and more comprehensive view of the entrepreneurial strategies that contributed to the ongoing success of women-owned bakeries in Philadelphia.

Quantitative Data Analysis

As shown in Figure 2 below (Supporting Information is attached), customer ratings as of January 19th, 2026 for a total of 476 bakeries in the great Philadelphia area were collected. Google gave an average rating of 4.48 based on 166,621 reviews and Yelp averaged at 4.17 with 60,421 ratings (based on rating scale of 1 – 5). The average rating on Yelp was 0.31 lower than

Google. The number of reviews on Yelp was 36% of those on Google. The trends on both platforms aligned closely, agreeing on the high- and under-performing bakeries. Two major differences were observed: 1) Yelp showed more ratings at extremes, with a higher number of perfect scores (ratings at 5) and more subpar ratings as low as 1. Google's rating seemed more tolerant, with fewer 5's and no rating below 2.8. 2) Most customers who rated on Yelp left a message, while many Google reviews were just a rating with no comments. Yelp reviewers usually had a story to tell related to their personal experience with specific bakeries. The review text explained how ratings were given and what factors contributed to satisfaction or disappointment.

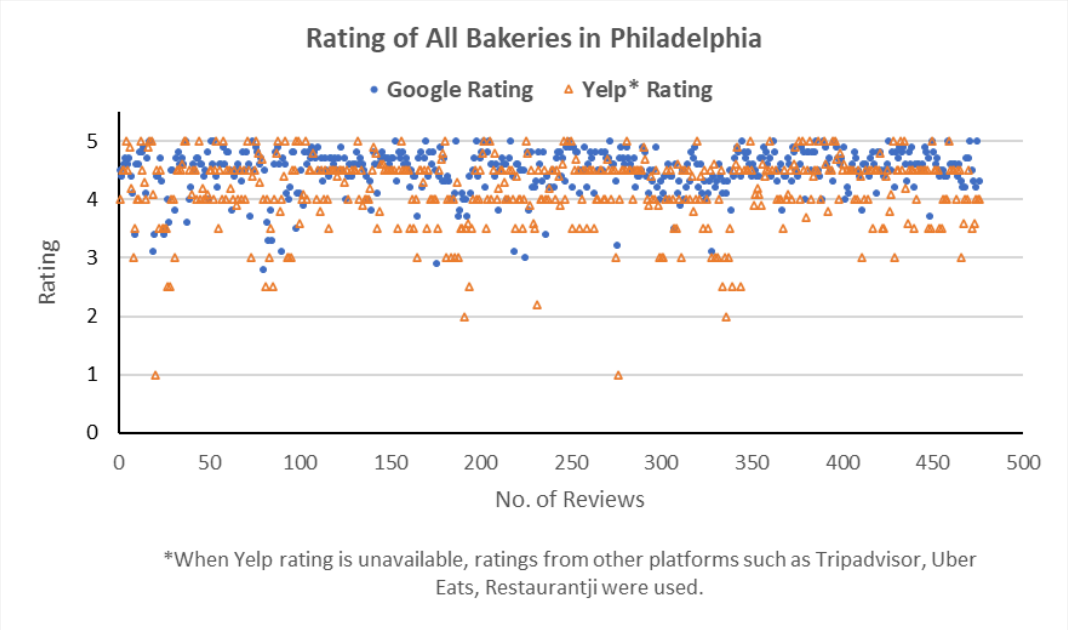


Figure 2. Rating of All Bakeries in Philadelphia.

Out of these 476 bakeries, 149 were identified as women-owned businesses, ~1/3 of the total number. For women owned bakeries, the average Google rating was 4.59 (44,093 reviews), 0.11 higher than the average of all bakeries; the average Yelp rating was 4.32 (17,161 reviews), 0.15 higher than the average of all bakeries (Figures 3 & 4). Women-owned bakeries were consistently better received than men-owned bakeries. Women-owned businesses seldom showed up in the region lower than 3, indicating female owners' strong attention to customer experience and quick response to address customer dissatisfaction.

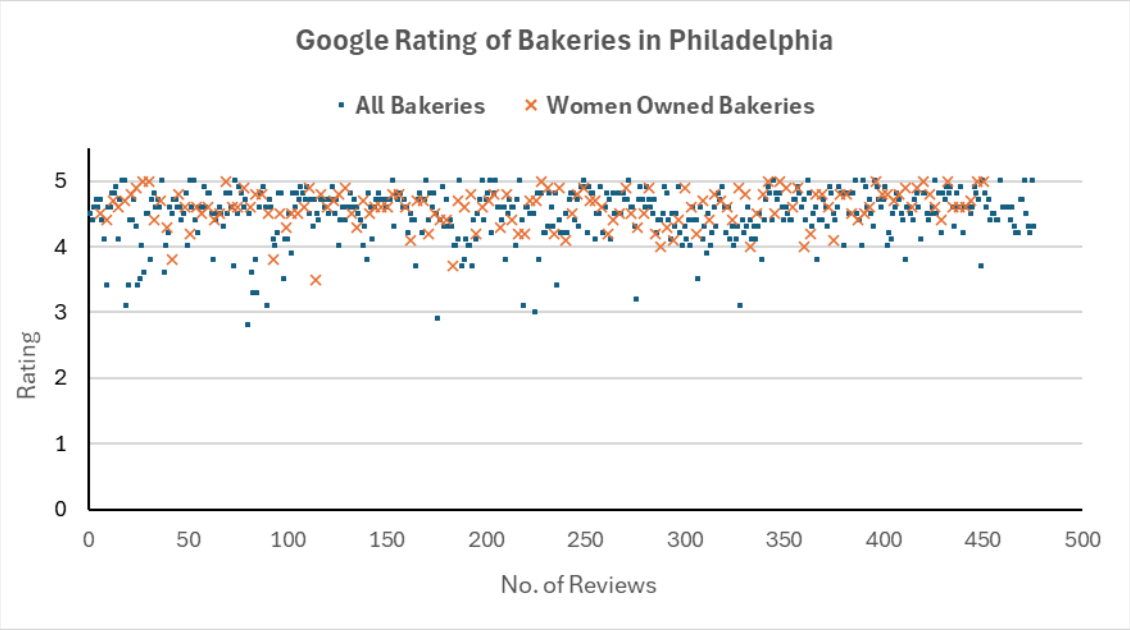


Figure 3. Google Rating of Bakeries in Philadelphia.

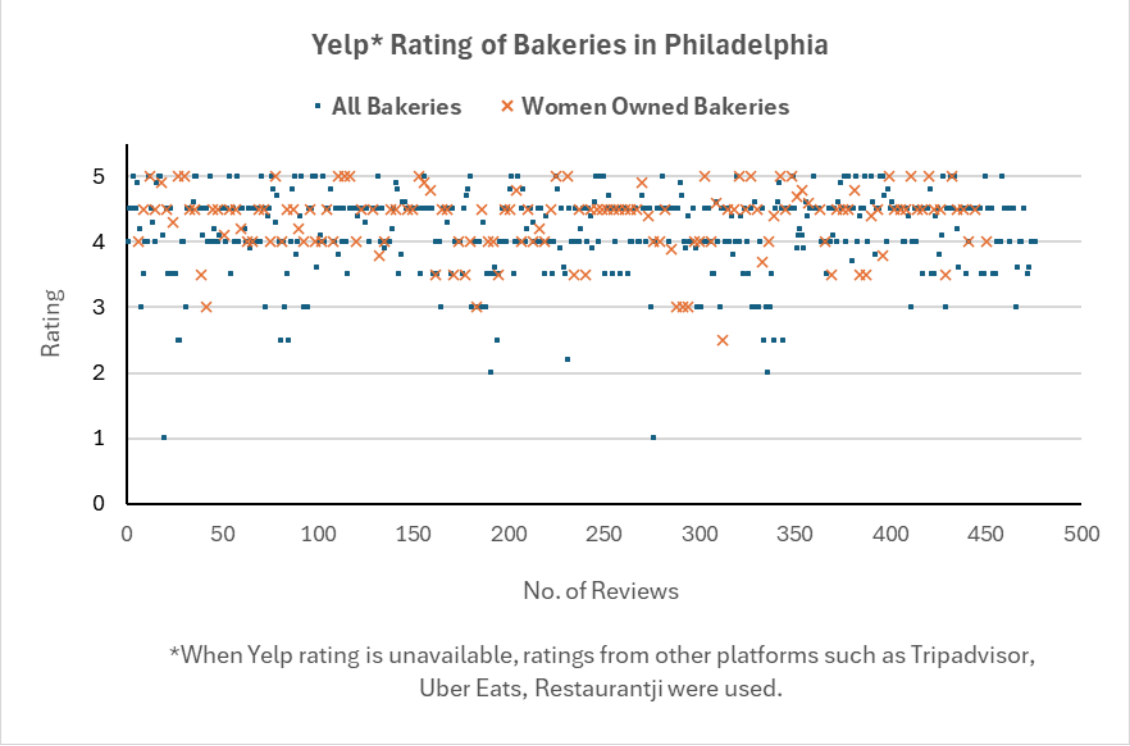


Figure 4. Yelp Rating of Bakeries in Philadelphia.

Yelp reviews provided a valuable source of customer feedback, offering insights into customer satisfaction, product preferences, and overall perceptions of the bakeries. By analyzing

these reviews, several recurring themes emerged. Customers consistently highlighted key factors such as product quality, customer service, and the atmosphere of the bakeries.

The word cloud generated from Yelp reviews provided a deeper look into the customer experience at Sweet Nina's Bakery, with additional words shedding light on specific products and positive sentiments expressed by patrons. Key terms like “Muffin”, “Avocado”, and “Blueberry” suggested that the bakery's offerings go beyond traditional sweets, appealing to customers who enjoy more varied flavors and healthier options. “Avocado” might reflect popular breakfast or brunch items such as avocado toast, while “Blueberry” likely pointed to well-loved pastries like blueberry muffins, indicating the range of flavors customers can find. Positive descriptors like “Nice”, “Best”, “Great”, “Highly”, and “Perfect” indicated strong approval of both the bakery’s offerings and its overall atmosphere. The frequent appearance of “Best” suggested that customers often viewed Sweet Nina’s as their favorite bakery, setting it apart from others in the area. Similarly, “Highly” likely referred to the high level of recommendations patrons gave, encouraging others to try the bakery for the first time.

Terms such as “Brownie”, “Cheese”, and “Bacon” further illustrated the variety of items on offer, from decadent desserts to savory options. “Brownie” hinted at a standout item on the dessert menu, while “Cheese” and “Bacon” pointed to savory options, possibly in sandwiches or breakfast offerings. The inclusion of “Croissant” and “Toast” emphasized the popularity of traditional bakery items, particularly for breakfast, while “Egg” and “Bun” suggested the bakery caters to customers looking for heartier meals.

The use of emotive words like “Love”, “Tasty”, “Amazed”, and “Sweet” reflected a deep appreciation for both the food and the overall experience. “Love” in particular implied that many patrons felt a strong emotional connection to the bakery, which could foster long-term customer loyalty. “Tasty” and “Sweet” reinforced the focus on delicious, high-quality baked goods, while “Amazed” suggested that the bakery often exceeded customer expectations with its offerings. Specific items like “Babka”, “Almond”, and “Bacon” pointed to the variety of flavors that catered to diverse palates. The word “Tasty” implied that the food consistently delivered in terms of flavor, while “Again” reflected repeat visits, showcasing the bakery's ability to maintain a loyal customer base. The inclusion of “Toast” and “Egg” highlighted popular breakfast items, further solidifying Sweet Nina’s role as a go-to spot for early-morning meals.

The word clouds created for male- and female-owned bakery reviews (Figures 5 & 6) in Philadelphia brought forth some striking similarities in the characteristics that people valued most in their experience. The word clouds for both had positive keywords like "delicious", "recommend", and "good" which spoke of the shared standards of excellence across the board. However, the words most emphasized in reviews of the female-owned bakeries were "creative", "unique", and "warm", showing these businesses did not just meet but also at times even surpassed certain customer expectations. Based on this analysis, female-owned bakeries indeed were more innovative than male bakeries. In addition, female owners were often described as friendly, welcoming, feminine, decorative, women empowering, and supportive. The word clouds indicated that female-owned bakeries can perform as well as, or even better than, their

male-owned counterparts in Philadelphia, providing as much or more talent and passion with their baking.

Figure 5. Word Clouds from Male-owned Bakery Reviews.

Figure 6. Word Clouds from Female-owned Bakery Reviews.

Both of these word clouds provided us with an understanding of women-owned bakeries and their special entrepreneurial journey within the Philadelphia area. The analysis on the reviews showed how female entrepreneurs excel in so many ways that strongly resonate with customers, namely, how they engender loyalty and appreciation through such qualities as warmth, creativity, and attention to detail. This comparison showed that women-owned bakeries not only succeeded but most of them had distinct contributions to the business and entrepreneurial field.

The map of bakeries in Philadelphia (Figure 7) revealed key patterns in the spatial distribution of women-owned bakeries in relation to household median income levels. The darker red areas on the map, indicating higher household median incomes, were where a

significant concentration of women-owned bakeries (represented by blue markers) was found, while other bakeries men-owned or with undisclosed ownership (green markers) were more evenly distributed across the city.

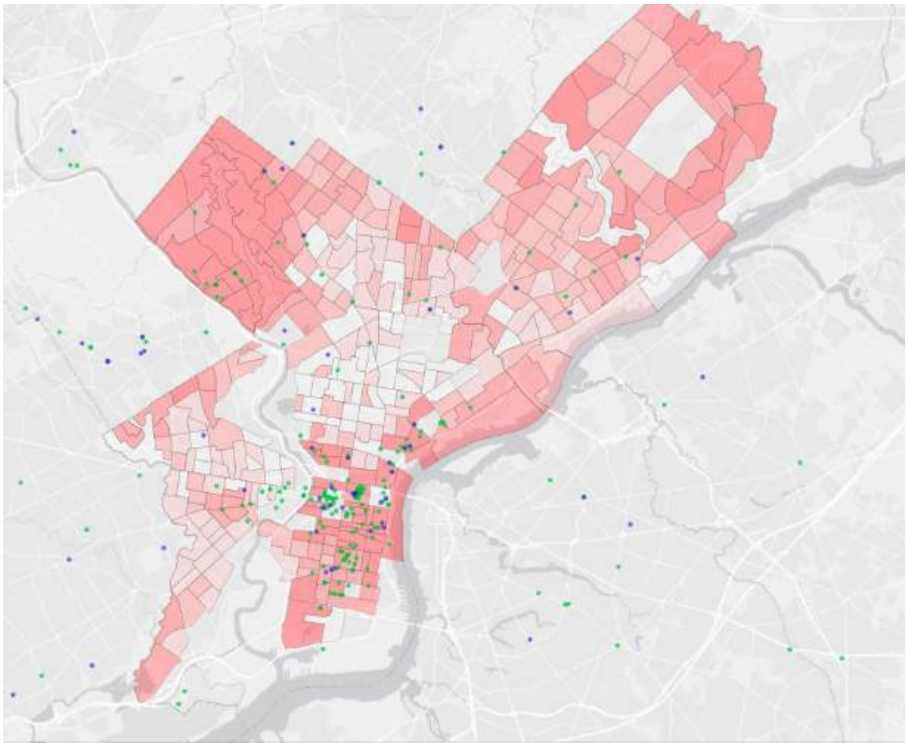


Figure 7. Distribution of Bakeries in Philadelphia: Women-owned vs. Men-owned or Undisclosed Ownership Bakeries in Philadelphia.

One notable trend was the clustering of women-owned bakeries in higher-income neighborhoods, such as Center City, Rittenhouse Square, and Chestnut Hill. These areas, known for their affluence and commercial vibrancy, offered a prime environment for small businesses that rely on a customer base with greater disposable income. Women entrepreneurs in these regions were likely able to cater to clientele seeking high-quality, artisanal products. The higher purchasing power in these neighborhoods allowed women-owned bakeries to thrive, as customers were often willing to pay a premium for unique, locally made goods. The central location of these businesses also benefited from significant foot traffic, further contributing to their success. Additionally, the presence of women-owned bakeries in gentrifying areas such as Fishtown and Northern Liberties highlighted the adaptability of these entrepreneurs. As these neighborhoods transitioned, with an influx of young professionals and new residents, the demand for community-focused specialty businesses increased. Women-owned bakeries in these areas were well-positioned to tap into this demographic, benefiting from relatively lower rent costs at the early stages of gentrification and capitalizing on the growing market demand for artisanal and organic products.

Conversely, the map indicated a limited representation of women-owned bakeries in lower-income areas. The absence of women-owned bakeries in these regions could be linked to several factors, including economic constraints, lower consumer spending power, and challenges in securing the necessary capital to start and maintain a small business. Furthermore, the nature of the products offered by many women-owned bakeries – often artisanal or niche items – might not align with the needs or financial means of residents in these areas, where cost was a more pressing concern for consumers. In contrast, men-owned bakeries or bakeries with undisclosed ownership were more evenly spread across Philadelphia, including in lower-income areas. This broader distribution suggested that these bakeries might cater to a more diverse customer base, offering affordable, staple baked goods that served the daily needs of the local community. Such bakeries were less reliant on the higher profit margins of artisanal products and might benefit from being longstanding neighborhood fixtures.

The cultural and demographic context of some neighborhoods was also relevant to understanding the distribution of bakeries. In areas with significant immigrant or minority populations, bakeries often served as cultural institutions, offering traditional products that served specific community needs. Women-owned bakeries did not cluster in these neighborhoods, showing that bakeries in general serve this function, aligning with the day-to-day consumption patterns of the local population, rather than targeting wealthier or gentrifying areas with niche products.

The concentration of women-owned bakeries in higher-income and gentrifying areas can be explained by several factors. First, access to capital is often more readily available to entrepreneurs in affluent areas, where the potential for business success is higher due to the wealthier customer base. Second, customer preferences in these neighborhoods tend to favor artisanal, organic, and specialty products, which many women-owned bakeries focus on. These products often come with higher profit margins, making them more viable in areas where customers have the financial capacity to prioritize quality over cost. Finally, the presence of business networks and higher levels of foot traffic in these areas enhances the visibility and opportunities for women-owned businesses, enabling them to attract a broader customer base and establish a strong community presence.

In contrast, bakeries in lower-income areas serve a more practical, cost-conscious market. These establishments likely focus on providing everyday baked goods at affordable prices, ensuring they remain accessible to the local population. As such, they play an important role in catering to the basic needs of their communities, rather than focusing on the artisanal or upscale products often associated with women-owned bakeries.

Qualitative Data Analysis

The interviews with women bakery owners provided deep insights into their personal and professional journeys. One of the most consistent themes across the interviews was the importance of emotional connections. Many bakery owners drew on personal or family histories to inform their products and brand identities. For example, Nina Bryan, owner of Sweet Nina's

Bakery, took her mother's banana pudding recipe and transformed it into the signature item of her shop. Similarly, Betty and Ellie, owners of Ramona Susan's Bake Shop, incorporated their mother's original recipes into their business, creating an emotional resonance with customers that went beyond the products themselves. These connections fostered customer loyalty and contributed to the authenticity of the brand, further enhancing customer trust and engagement.

Another key finding from the interviews was the role of community service and local engagement. Many women bakery owners viewed their businesses as more than just places to buy baked goods – they considered themselves integral parts of the communities they served. For instance, Tina Phillips, owner of The Famous 4th Street Cookie Company, was known not only for her cookies but also for her philanthropic work. Her bakery donated $20,000 annually to the Ronald McDonald House and partnered with Make-A-Wish Foundation, cementing its role as a community cornerstone. Similarly, Teresa Wall, owner of Tartes Pastry Shop, strengthened her community connections by selling primarily to local customers through her walk-up window service, ensuring that her business remained a neighborhood staple.

The study also highlighted the importance of innovation in maintaining competitive advantage. Many of the women bakery owners interviewed emphasized the need to stay creative and adaptable to meet changing customer preferences. Ramona Susan's Bake Shop, for example, adapted to growing customer demand by expanding and tastefully decorating its physical space and increasing its online presence. Sweet Nina's Bakery introduced 15 different flavors of banana pudding to keep its menu fresh and exciting, ensuring continued customer interest and engagement. Innovation, therefore, remained a critical component for the continued success of women-owned bakeries, helping them stay relevant and competitive in an ever-changing market.

While the bakeries had achieved considerable success, the study also uncovered several challenges that these women entrepreneurs faced. Resource management emerged as a common issue. Betty and Ellie at Ramona Susan's Bake Shop, for instance, mentioned the difficulties of scaling their business while managing limited resources. Teresa Wall at Tartes Pastry Shop also struggled with securing funding to adapt to the evolving market landscape. Competition, particularly in urban settings, further compounded these challenges. Tina Phillips from The Famous 4th Street Cookie Company spoke about the pressures of competing in the rapidly expanding e-commerce sector, where maintaining product quality while managing business growth was a constant challenge.

Another critical challenge identified was the balance between workload and personal life. Peak seasons often required bakery owners to work long hours, putting a strain on their personal lives. Tina Phillips, for example, explained that during busy periods, she often worked extended hours to meet demand, leaving little time for personal commitments. Similarly, Betty from Ramona Susan's expressed concerns about the added responsibilities of managing non-baking tasks, such as marketing and finances, which significantly increased her workload.

Despite these challenges, resilience was a defining trait among these entrepreneurs. Several of the women interviewed mentioned how the COVID-19 pandemic tested their businesses, but their ability to adapt and innovate helped them survive. Nina Bryan of Sweet

Nina's Bakery spoke about how she restructured her menu and doubled down on community engagement during the pandemic, reinforcing her business's connection to its customer base during difficult times.

Success Factors and the Importance of Customer Relationships

The success of women-owned bakeries in Philadelphia was underpinned by several key factors, most notably customer relationships and community support. Strong customer relationships, built on trust and emotional connections, played a critical role in creating loyal customer bases. The Beiler family, owner of Beiler's Bakery, attributed much of their bakery's success to the focus on customer experience, ensuring that each customer feels valued and heard. Similarly, Teresa Wall's emphasis on personalized service and unique pastries had built her a dedicated customer base that appreciates her attention to detail.

Community support also played a crucial role in sustaining these businesses. Many bakery owners, including Tina Phillips and Betty and Ellie, spoke about how the local community had inspired and supported their businesses. This sense of local attachment not only helped these businesses thrive but also strengthened the fabric of the communities they serve.

Results

This study combined quantitative data from Google and Yelp reviews and qualitative insights from in-person interviews to provide a comprehensive understanding of the factors contributing to the success of women-owned bakeries in Philadelphia. The analysis revealed important themes such as emotional connections, community engagement, and innovation, which played key roles in shaping the business strategies of these entrepreneurs.

Discussion

This study provided a comprehensive analysis of the entrepreneurial paths taken by women-owned bakeries in Philadelphia, offering insights into the unique challenges and strengths these businesses navigated. The findings were particularly significant in contributing to the literature on female entrepreneurship, as they illustrated how women-owned bakeries leverage community engagement, emotional connections, and innovation to thrive in a competitive market. Moreover, this study extended beyond the specifics of bakery operations to serve as a broader reflection of how women entrepreneurs succeeded in traditionally male-dominated industries.

One key takeaway from this study is the role of community engagement in driving business success. Many of the women entrepreneurs being interviewed emphasized the importance of fostering strong relationships with their local communities. This aligned with previous research suggesting that women-owned businesses are more likely to prioritize social responsibility and community ties (Atchley, 2024). In Philadelphia, this manifested in various ways, from bakeries using family recipes that resonate with local customers to those engaging in philanthropy and local partnerships. This deep connection to the community not only enhanced

customer loyalty but also established the bakery as an integral part of the neighborhood's identity.

The study also shed light on the resilience and adaptability of women entrepreneurs. The bakery owners we interviewed frequently spoke about the necessity of adapting to changing market conditions, from shifts in consumer preferences to the impact of the COVID-19 pandemic. The ability to pivot and adjust business models was a recurring theme, particularly in terms of expanding digital presence and online sales. This mirrored findings in the broader literature on women's entrepreneurship, where resilience was highlighted as a core trait that enabled women to overcome obstacles and succeed in their ventures (Fazio, 2024). By diversifying their product offerings and embracing new technologies, these bakery owners demonstrated the flexibility required to sustain long-term growth in a highly competitive industry.

Another prominent theme was the emotional connection that women entrepreneurs establish with their customers. Through interviews and customer feedback, it became clear that these bakeries often serve as more than just places to purchase baked goods – they are spaces where personal stories, traditions, and familial legacies are shared. This emotional bond with customers was evident in the Yelp reviews, where terms like "delicious", "favorite", and "best" frequently appear. The word cloud analysis underscored this connection, highlighting that customers valued not only the quality of the products but also the atmosphere and sense of belonging these bakeries created. This observation aligned with earlier studies that suggest women entrepreneurs often place a stronger emphasis on relationships and emotional engagement in their business practices (Chan, 2024). The nurturing character of female business owners of Philadelphia's bakeries contributed to a positive and homecoming community life.

The geographical distribution of women-owned bakeries, as visualized in the map, further supports the argument that location and community are critical factors in the success of these businesses. The map showed a concentration of women-owned bakeries in higher-income neighborhoods, suggesting that economic factors played a significant role in determining where these businesses thrived. Areas with higher household median incomes offer more opportunities for women entrepreneurs to establish and sustain their businesses, as wealthier customers are more likely to prioritize quality and support local enterprises. This is consistent with the idea that access to resources, both financial and social, can influence the success of women-owned businesses (Carter, 2001). In contrast, men-owned bakeries or bakeries with undisclosed ownership are more widely distributed across the city, including in lower-income areas. This indicates that women-owned bakeries may face challenges in penetrating markets with less purchasing power, where customers are more price-sensitive. However, this does not detract from the significance of the success these women entrepreneurs have achieved in their respective communities. By strategically positioning their businesses in areas where they can maximize their impact and profitability, these women-owned bakeries demonstrated a keen understanding of market dynamics and consumer behavior. The resilience of female business owners of Philadelphia's bakeries strongly contributed to the survival of the bakery industry.

Finally, this study highlighted the importance of innovation as a driving force for success. Many of the bakeries we studied had adopted innovative practices, from developing new product lines to expanding their digital presence. This innovation was not only a response to market demands but also a reflection of the entrepreneurial mindset that characterized these women business owners. By continually evolving and adapting, these bakeries ensure their relevance in a rapidly changing industry.

Conclusion

This study explored the entrepreneurial journeys of women-owned bakeries in Philadelphia, focusing on how community engagement, resilience, and innovation contributed to their success. The integration of Yelp reviews and interviews provided a well-rounded view of how these businesses navigated their local markets while building lasting relationships with their customers. This study contributed to the broader understanding of women's entrepreneurship, illustrating how these businesses not only drive economic growth but also strengthen community ties. Women entrepreneurs, through their focus on relationships, innovation, and adaptability, are actively reshaping the landscape of local economies and challenging traditional notions of success. The insights provided here can serve as a valuable resource for scholars, policymakers, and business leaders aiming to promote inclusive and community-focused entrepreneurship.

Despite these findings, this study has several limitations. Conducting the research within a single city limits the ability to generalize the results to other geographic areas with different economic and social contexts. Moreover, while Yelp reviews offer valuable customer feedback, they may not fully capture the breadth of customer experiences, introducing potential bias based on the self-selecting nature of online reviews. The small sample of interviews also restricted the depth of understanding of the varied challenges faced by women-owned bakeries. Another important limitation relates to the researcher's personal identity. As a young, Asian female high school student, the researcher's identity might have influenced how interview participants responded, potentially shaping the nature of the qualitative data collected. While this factor provided unique perspectives on the research topic, it may have introduced unintended biases in the interpretation of the findings.

Future research could address these limitations by extending the study to include different cities and regions, allowing for more comparative analysis across various markets. Investigating women-owned businesses in other industries could also broaden the understanding of the entrepreneurial strategies women employ across sectors. Additionally, exploring the role of digital tools, such as social media, in supporting the growth of women-owned businesses could offer further insights into contemporary entrepreneurial practices.

Works Cited

Atchley, C. “The women’s bakery looks to empower, educate, and employ women.” Bake Magazine (2024) March 18. https://www.bakemag.com/articles/18687

Biz2Credit. “2025 women-owned business study.” (2025) https://cdn.biz2credit.com/images/html5/2025_Women_Owned_Business_Study.pdf

Brkic, F. “From young professional to entrepreneur.” Food Science and Technology (2025): 39(2), 60–61. https://academic.oup.com/fst/article/39/2/60/8169019

Carter, S. “Women’s business ownership: A review of academic, popular and Internet literature.” European Foundation for Management Development (2001). https://www.researchgate.net/publication/237222197

Chan-Danisi, A. “The Chinese dumpling revolution: MìLà’s journey to disrupt the frozen food industry.” Forbes (2024) March 5. https://www.forbes.com/sites/angelachan/2024/03/05/the-chinese-dumpling-revolution-mlas-journey-to-disrupt-the-frozen-food-industry/

Fazio, C. “Breaking barriers, fostering growth: The evolution of Black women-owned businesses in the U.S.” Insights@Questrom (2024). https://insights.bu.edu/breaking-barriers-fostering-growth-the-evolution-of-black-women-owned-businesses-in-the-u-s/

Grow to Market. “Turning influence into impact: Women entrepreneurs in the food sector and Grow to Market’s mission.” LinkedIn (2026) January 14. https://www.linkedin.com/pulse/turning-influence-impact-women-entrepreneurs-food-sector-oskbe

Heald-Fisher, S. M. “Female entrepreneurs making connections: Networking by women small business owners (Doctoral dissertation, Capella University).” ProQuest (2010). https://www.proquest.com/openview/96b549fafb72201aa13fce49943ec34b

Hyslop, G. “Women, power & progress: Shaping the next generation of business.” Bakery & Snacks (2025). https://www.bakeryandsnacks.com/Article/2025/03/06/women-in-business-breaking-barriers-shaping-the-future/

“National Association of Women Business Owners.” 2025 NAWBO advocacy agenda (2025). https://nawbo.org/wp-content/uploads/NAWBO-2025-Advocacy-Agenda-1.pdf

Olsson, A. K., & Bernhard, I. “Keeping up the pace of digitalization in small businesses: Women entrepreneurs’ knowledge and use of social media.” International Journal of Entrepreneurial Behavior & Research (2020): 27(2), 378–396. https://www.emerald.com/insight/content/doi/10.1108/IJEBR-10-2019-0615/full/html

Rahman, F., Hanafiah, M., Zulkifli, N., et al. “Empowering women food entrepreneurs: Unveiling their social commerce readiness and digital challenges.” Journal of Entrepreneurship and Business (2025): 13(1), 18–34. https://journal.umk.edu.my/index.php/jeb/article/view/1369

Sweet Nina’s. (n.d.). Sweet Nina’s (2026). https://www.sweetninas.com/

Wells Fargo. “New report finds growth of women business owners outpaces the market.” Wells Fargo Newsroom (2024) January 9. https://newsroom.wf.com/English/news-releases/news-release-details/2024/New-Report-Finds-Growth-of-Women-Business-Owners-Outpaces-the-Market/default.aspx

Wells Fargo. “The 2025 impact of women-owned businesses.” (2025) https://smallbusinessresources.wf.com/wp-content/uploads/2025/01/wells-fargo-2025-impact-of-women-owned-businesses-1MM-plus.pdf

Wells Fargo. “The impact of women-owned businesses 2026 report.” (2026) January. https://smallbusinessresources.wf.com/wp-content/uploads/2026/01/IWOB-2026-Report_Concept_Concept-17_Final.pdf

Cold War Calculus:
The United States' Rapprochement with the People's Republic of China
By Samuel Checkal

During the early Cold War, the United States maintained formal recognition of the Republic of China (ROC) as the legitimate government of China, despite the People's Republic of China (PRC) exercising full control over the mainland after the ROC's retreat to Taiwan.[91] By the late 1960s and early 1970s, this arrangement became increasingly unrealistic as the PRC gained international legitimacy, most notably through the passage of United Nations General Assembly Resolution 2758 in 1971, which recognized the PRC as China's representative in the United Nations and removed the ROC from mainstream international policy deliberations.[92] At the same time, the United States faced mounting geopolitical pressure during the Cold War, forcing policymakers to reconsider long-standing diplomatic commitments in light of broader strategic priorities. The PRC's increasing global presence, President Jimmy Carter's attempt to balance human rights concerns with pragmatic engagement, and emerging business opportunities with China all contributed to the timing and the feasibility of U.S. recognition of the PRC. However, they cannot fully explain why the United States chose to fundamentally realign its China policy after decades of consistent support for the ROC. The shift in recognition must be understood within the broader strategic context of the Cold War, particularly the emerging opportunity to exploit the growing tension between the two largest communist nations: the U.S.S.R. and the PRC. *The United States' shift in official recognition of the government of China*

[91] In this paper, the term *China* will be used interchangeably with *PRC* to refer to Mainland China under the jurisdiction of the People's Republic of China, while the term *Taiwan* will be used interchangeably with *ROC* to refer to Formosa and associated islands under the jurisdiction of the Republic of China.
[92] UN. General Assembly (26th sess. : 1971), "Restoration of the lawful rights of the People's Republic of China in the United Nations.," United Nations Digital Library, United Nations, last modified October 25, 1971, accessed March 22, 2026, https://digitallibrary.un.org/record/192054?v=pdf.

from the Republic of China to the People's Republic of China was primarily driven by the pursuit of strategic advantages and increased leverage against the U.S.S.R. through a new diplomatic relationship with the People's Republic and informal relations with Taiwan.

The collapse of the Sino-Soviet alliance was rooted in ideological rivalries and political tensions between the Chinese and Soviet communist movements, which fractured the assumption of a communist monolith. As early as the 1920s, decades before the PRC was even founded, there were signs of a rift between the two parties.[93] Mao Zedong, the leader of the Chinese Communist Party, was against the orthodox concept of Marxist-Leninism, where revolution was based on the industrial masses; rather, Mao wanted the revolution to be focused on the peasants.[94] According to W. Averell Harriman, the Ambassador-at-Large for the United States Department of State, Stalin called Mao a "margarine Communist", a term he used to refer to a fake Communist who, unlike a "true Communist", didn't look to Moscow for leadership and guidance.[95] These tensions reached a boiling point in the late 1950s. Escalating disagreements between Mao and Nikita Khrushchev, leader of the Soviet Union, over both ideology and leadership within the communist movement, led the two leaders to lash out at each other. Before stunned onlookers from the international Communist movement, "Khrushchev was reported to have made bitter personal attacks on Mao… Mao attacked Khrushchev for revisionism, for mishandling de-Stalinization, for blundering responses to the Polish and Hungarian crises, and—worst of all from Peking's viewpoint—for sacrificing Chinese interests in pursuit of a chimerical Soviet-American détente."[96] These exchanges reveal that the conflict extended beyond abstract ideological differences and into direct challenges over authority and national interests within the communist world.

The Sino-Soviet split transformed the PRC from an extension of a unified communist bloc into a potential strategic counterweight against the U.S.S.R., allowing U.S. policymakers to exploit cracks in the communist "monolith" to gain leverage in the global Cold War struggle. This was significant because it shifted away from the previous American assumption that communist nations acted as a single coordinated entity with identical goals, opening the path for a more nuanced engagement with Beijing. This growing division was recognized by American officials. As W. Averell Harriman observed, "I think it was a great mistake when we thought of the Communist bloc as a monolithic structure… It was thought in Moscow that the Communist world revolution would have Moscow as the oracle, the center. It was with a great deal of dissatisfaction that they found the Chinese Communist party would not come to heel, and of course it was even more serious when [Peking] began to challenge Moscow's leadership."[97]

---

[93] Donald S. Zagoria, *Sino-Soviet Conflict, 1956-1961*, Princeton Legacy Library (Princeton, N.J.: Princeton University Press, 1962), 39, https://research.ebsco.com/linkprocessor/plink?id=d40254a2-1e8a-308b-b081-8b6edd046dc8.

[94] W. Averell Harriman, "Sino-Soviet Conflict," *Proceedings of the Academy of Political Science* 28, no. 1 (1965): 102, https://doi.org/10.2307/1173719.

[95] Harriman, "Sino-Soviet Conflict," 103.

[96] Donald S. Zagoria, *Sino-Soviet Conflict*, 39.

[97] W. Averell Harriman, "Sino-Soviet Conflict," *Proceedings of the Academy of Political Science* 28, no. 1 (1965): 109, https://doi.org/10.2307/1173719.

Harriman's reflection is significant as it demonstrates that this ideological split and highly personal attacks between the two leaders weakened the assumption of a unified communist bloc, making it possible for the United States to shift its understanding of the PRC from being an extension of Soviet power to an independent nation and actor with competing interests.

Crucially, this split in the communist monolith was not limited to rhetoric or ideology but also escalated into a tangible conflict, transforming the alliance landscape and opening the door to U.S. rapprochement. China came to a point where it viewed the Soviet Union as such a threat that it posed a more significant danger to China's national survival than the United States.[98] This perception was reinforced by a series of border disputes throughout the 1960s, particularly along the Amur and Ussuri Rivers, where battles caused the Russians and Chinese to suffer several hundred casualties.[99] The outbreak of direct military confrontation between the two largest communist powers demonstrated that their rivalry was not superficial but rather embedded in their competing national interests and security concerns. As a result, the long-standing American assumption of a unified communist threat was broken, creating an opening for the United States to engage with the PRC as a strategic partner against the Soviets. This transformation fundamentally changed the landscape of the Cold War. By revealing deep divisions within the communist bloc, the Sino-Soviet split enabled US policymakers to envision a strategy that not only made rapprochement with the PRC possible but also strategically advantageous in the broader effort to weaken the Soviet Union.

Leaders within the Nixon administration viewed rapprochement as a strategic tool to pressure communist advancements across multiple fronts. President Nixon, believing China was supplying the communist North Vietnamese with weapons, hoped "China could be persuaded to halt the supplies so that he could pull out U.S. troops faster. Both Nixon and Kissinger believed that if the Soviets thought the United States and China were becoming allies, then the Soviets would also push for better relations with the United States."[100] This statement reveals two simultaneous strategic calculations that the administration made in its pursuit of diplomatic relations with the PRC: first, that China could directly influence the course of the Vietnam War, and second, that even the perception of U.S.-PRC alignment could be used to coerce the Soviet Union into concessions. The U.S. was engaging with China not because it had become more legitimate or acceptable to capitalist ideas, but because China could be leveraged against the threat of the Soviet Union. The emphasis on Soviet reactions rather than internal developments within China demonstrates that U.S. policymakers were primarily concerned with how engagement with Beijing would alter Moscow's behavior. The Sino-Soviet split was not simply a background condition; rather, it was a conflict that American leaders actively reframed as a strategic weapon to weaken their principal Cold War adversary.

---

[98] Joseph L. Nogee, "Sino-Soviet Tensions Mount Along the Ussuri River Border," *EBSCOhost*, last modified 2023, https://www.ebsco.com/research-starters/history/sino-soviet-tensions-mount-along-ussuri-river-border

[99] Nogee, "Sino-Soviet Tensions.confrontation"

[100] "Nixon, Richard M," in *Cold War Reference Library*, ed. Richard C. Hanes, et al. (Detroit, MI: UXL, 2004), 5, https://link.gale.com/apps/doc/CX3410800150/UHIC?u=seat13327rpa&sid=bookmark-UHIC&xid=506d432a.

China's own strategic concerns about the Soviet Union further reinforced the viability of this approach, showing how a relationship between the U.S. and China would be mutually beneficial against a common adversary. The U.S. interest in establishing relations with China was not one-sided, where Chinese leaders actively, albeit privately, signaled their willingness to improve relations with the United States in response to growing fears of Soviet aggression. Two marshals close to Mao Zedong were worried about a dangerous confrontation with Moscow and proposed that Beijing play "the card of the United States" to provide leverage with Moscow.[101] In this context, "the card of the United States" refers to China's ability to utilize its cordial relationship with the U.S. to negotiate with the Soviet Union. Beijing thought that "playing" this "card" with the Soviets could pressure Moscow into making certain concessions that it otherwise would not have. The use of the language of playing "the card of the United States" closely mirrors the thinking on the American side of playing "the China card", where, similarly, the U.S. can use its warming relationship with China to pressure the Soviets into compromises or cooperation. This language shows that both sides independently arrived at the same conclusion of the benefits of a more friendly relationship in regards to the U.S.S.R. Similarly, Deng Xiaoping emphasized that the United States should maintain superiority over the Soviet "Polar Bear"[102] a shocking comment for a communist government to actively wish for U.S. domination, thus further demonstrating that Chinese leaders prioritized limiting Soviet influence over ideological opposition to the United States. Because both nations were driven by opposition to the Soviet Union, their alignment was driven by shared strategic interests. The Sino-Soviet split did not merely create favorable conditions for reconciliation; it changed how the U.S. viewed the PRC as a strategic partner in the Cold War, despite ideological differences. The Sino-Soviet split demonstrated China's desire to act independently of Moscow's control, breaking the mirage of a unified communist bloc and revealing the threat of the Soviets towards Chinese national security.

The Nixon administration intentionally responded to the Sino-Soviet split through secret diplomacy with China to leverage them and apply pressure to the U.S.S.R. A key part of this was the "triangular diplomacy" orchestrated by Richard Nixon and National Security Advisor Henry Kissinger, who viewed a relationship with Beijing as a vital mechanism for balancing power against the Soviet Union.[103] Early engagement between the United States and the PRC was conducted through highly controlled, heavily planned, and covert diplomatic channels designed to maximize strategic advantage while minimizing political blowback. Declassified communications from the Nixon administration show that initial communications between the two nations were transmitted through the president of Pakistan, a country that held good relations with both sides. The president of Pakistan at the time, Yahya Khan, also made important contributions to the messages that were transmitted, including analysis and comments on the restrained tone of official Chinese communications:

---

[101] William Burr, ed., "The Beijing-Washington Back-Channel and Henry Kissinger's Secret Trip to China," *The National Security Archive*, last modified February 27, 2002, https://nsarchive2.gwu.edu/NSAEBB/NSAEBB66/#7.
[102] Brian Hilton, "'Maximum Flexibility for Peaceful Change': Jimmy Carter, Taiwan, and the Recognition of the People's Republic of China," *Diplomatic History* 33, no. 4 (2009): 600, http://www.jstor.org/stable/44214071.
[103] Henry Kissinger, *White House Years* (Simon & Schuster, 2011), digital file.

> I think it is significant that Chou En-Lai did not accept or reject the proposal [for U.S. diplomats to visit Peking] as soon as it was made and that he consulted Mao and Lin Piao before giving the answer… Further, at no stage during the discussion with the Chinese leaders did they indulge in vehement criticism of the United States… These are additional indications of modification of the Chinese approach in their relations with the United States.[104]

Here, the Pakistani president highlights how, despite the restrained and cautious tone of communications from China, this already represented a meaningful step in Chinese diplomatic behavior, as such restraint contrasted with their previously much more negative and hostile comments towards the United States. The use of covert diplomacy indicates that rapprochement was not an inevitable outcome of global trends or business opportunities, but a carefully engineered policy shift, designed to maximize strategic outcomes and minimize political risk.

The Nixon administration's strategy moved from backchannels and secret diplomacy to a highly visible public state visit that was designed to shock the Soviet leadership. By appearing in Beijing, Nixon signaled to the world that the United States was prepared to move past decades of non-communication to alter the global balance of power. Reflecting upon his visit to China, Nixon states that "peace means more than the mere absence of war… [but] 22 years of noncommunication and hostility separated the United States of America from the… People's Republic of China… As a result of this trip, we have started the long process of building a bridge across that gulf."[105] By emphasizing the long period of hostility and lack of communication, Nixon presents the visit as a deliberate break from past policy. This "bridge" represented more than a diplomatic gesture or an effort to build friendship. It was a visible commitment to a new strategic realignment. The symbolic weight of the visit proved that the United States was willing to abandon the monolithic view of communism if it meant gaining a decisive advantage over the Soviet Union.

The United States maintained flexibility in the Cold War by using "strategic ambiguity" to bridge the gap between recognizing the PRC and protecting ties with Taiwan. This deliberate use of vague and flexible language was most evident in the Shanghai Communiqué issued jointly by the U.S. and PRC in 1972 during Nixon's visit. The Communiqué was meant to lay the foundations for relations between the two countries and ease hostilities between the two sides, not act as a formal recognition. In the Communiqué, the United States stated that it "acknowledges that all Chinese on either side of the Taiwan Strait maintain there is but one China and that Taiwan is a part of China. The United States Government does not challenge that position."[106] Similarly, in an address to the nation, President Carter utilized this same linguistic

---

[104] Henry A. Kissinger and Yahya Khan to Richard Nixon, memorandum, "Chinese Communist Initiative," n.d., https://nsarchive2.gwu.edu/NSAEBB/NSAEBB66/ch-06.pdf.
[105] "Nixon, Richard," 5.
[106] *Joint Statement Following Discussions With Leaders of the People's Republic of China*, 815, February 27, 1972, accessed March 23, 2026, https://history.state.gov/historicaldocuments/frus1969-76v17/d203.

loophole when he stated that the United States "acknowledges the Chinese position that there is but one China and Taiwan is part of China."[107] In both statements, the U.S. acknowledged the existence of "one China"[108] but deliberately avoided specifying which government holds legitimate authority, therefore preserving diplomatic ambiguity. The intentional use of ambiguous and non-committing language regarding the Taiwan issue reveals that U.S. officials prioritized establishing strategic cooperation with China over ideological consistency or clarity. It also shows that while the US sought to establish relations with the PRC, it was unwilling to abandon Taiwan. It wanted to strike a delicate balance between the two. Rather than resolving the Taiwan issue, U.S. policy effectively deferred it, allowing rapprochement to proceed without fully committing to the PRC's position. Nixon's visit and Carter's address to the nation underscore that the primary objective of engagement was to gain strategic benefits from relations with China to strengthen the U.S. position against the Soviet Union.

Although relations between the United States and China were warming, the U.S. was not willing to abandon its alliance with Taiwan and took steps to minimize damage to that relationship. The Taiwan Relations Act (TRA) shows that formal recognition of the PRC did not replace U.S. commitments to Taiwan, but redefined it to preserve as much strategic flexibility as possible. The TRA confirmed that U.S. rapprochement with the PRC was a strategic addition to U.S. foreign policy rather than a redefinition of alliances in the Pacific to allow for more flexibility in pursuing broader Cold War strategic objectives. As stated in the act, it "reaffirms as a commitment of the United States the preservation of human rights of the people of Taiwan…[and] declares that… the United States shall make available to Taiwan such defense articles and defense services… to enable Taiwan to maintain a sufficient self-defense capacity."[109] This section from the TRA proves that recognition of the PRC did not mean Taiwan had become irrelevant. Instead, the United States created a legal and political framework that allowed it to gain the strategic advantages of recognition of the PRC while still maintaining an important relationship with Taiwan. The U.S. did not completely shift away from Taiwan to China, but rather it was a delicate, calculating act of restructuring U.S. foreign policy in the Pacific to preserve influence in both relationships simultaneously, as they both served different roles. The strategic necessity of this balance was reinforced by Taiwan's significant economic value to the American economy. Even as formal ties were severed, Taiwan remained the United States' fifth-largest trading partner while China ranked around 30th.[110] Furthermore, Taiwan lies in a militarily strategic waterway that supplies resources to Japan and South Korea, two other vital U.S. allies in the region.[111] The recalibration and restructuring of priorities allowed the

[107] Jimmy Carter, "Address by President Carter to the Nation," address presented at The White House, Washington D.C., December 15, 1978, https://history.state.gov/historicaldocuments/frus1977-80v01/d104.
[108] In the context of the two statements by the Nixon and Carter administrations, the terms *China* and *PRC* are not interchangeable.
[109] Taiwan Relations Act, H.R. 2479, 96th Cong. (Mar. 24, 1979). Accessed February 23, 2026. https://www.congress.gov/bill/96th-congress/house-bill/2479.
[110] A. James Gregor, "The U.S. Investment in the Republic of China on Taiwan.," *Academic Search Premier*, last modified August 15, 1985, https://research.ebsco.com/linkprocessor/plink?id=5af62cc8-8b70-3120-8c93-57ddbf943073.
[111] Gregor, "The U.S."

United States to pursue anti-Soviet strategies without sacrificing Taiwan's role as a critical economic and geographic link in the Pacific.

The U.S. approach to negotiations with Beijing was a delicate balancing act in which officials conceded enough on Taiwan to gain leverage against the Soviets without yielding too much to PRC demands. In discussions between Henry Kissinger and Zhou Enlai, Chinese officials insisted that "the Government of the People's Republic of China is the sole legitimate government representing the Chinese people… [and that] Taiwan belongs to China."[112] While this transcript revealed that the PRC officials demanded complete sovereignty over Taiwan, the United States resisted total alignment with the Chinese demands. Instead, the U.S. adjusted its position only enough to publicly demonstrate to the Soviets a growing, united relationship with the PRC while maintaining American interests in Taiwan. Thus, gaining a strategic position against the Soviets without relinquishing a key position in East Asia.

The expansion from diplomatic recognition to military cooperation is compelling evidence reinforcing that containment of the Soviet threat was the intention of U.S. rapprochement. A key project of military cooperation between the U.S. and China occurred in the 1980s, called Project Peace Pearl. This $550 million project upgraded key avionics of Chinese fighter aircraft, expanded joint-intelligence gathering on the Soviet military, supported the Afghan resistance against the Soviet invasion, and other limited military sales programs to the People's Liberation Army.[113] These actions, which would not have been possible without rapprochement between the U.S. and China, directly undermined Soviet military capabilities rather than simply empowering the Chinese military. The military support was also structured to allow the Chinese to defend against a Soviet land invasion while preventing the Chinese from utilizing the new technology against Taiwan in sea-crossing and amphibious operations. The military cooperation between the two sides only extended as far as undermining Soviet actions, demonstrating that the reasoning for rapprochement was to increase diplomatic and military pressure on the U.S.S.R. These developments indicate that U.S.-China relations extended beyond symbolic or diplomatic engagement, producing concrete outcomes that directly challenged and constrained Soviet influence and military capabilities. These tangible anti-Soviet actions reinforce the argument that strategic considerations were paramount in conversations on rapprochement. The existence of such coordinated efforts confirms that normalization was not an end to rapprochement, but a stepping stone toward weakening the Soviet Union through cooperation with China. By preserving ties with Taiwan while simultaneously pursuing cooperation with China, the U.S. secured strategic advantages against the Soviet Union without sacrificing its existing position in East Asia.

Some scholars argue that President Jimmy Carter's emphasis on human rights and pragmatic engagement with adversaries caused the United States to shift its recognition from the

---

[112] Winston Lord, Smyser, and Holdridge to Henry A. Kissinger, memorandum, "Conversations with Chou En-Lai: July 10, afternoon sessions," n.d., https://nsarchive2.gwu.edu/NSAEBB/NSAEBB66/ch-35.pdf.
[113] Jingdong Yuan, "Sino-US Military Relations, 2001–2016: A Brief Overview," in *Dragon and Eagle Entangled: Sino-US Military Exchanges, 2001–2016* (Australian Strategic Policy Institute, 2017), 21, http://www.jstor.org/stable/resrep04245.7.

ROC to the PRC. However, President Carter more likely shaped the timing and presentation of rapprochement with the PRC rather than causing it. Professor of History, Brian Hilton argues that Carter viewed China as more than simply a strategic counterweight and that he pursued normalization because of his moralistic belief that the Sino-American rapprochement in the early 1970s had produced genuine progress towards integrating China into the international community.[114] He hoped that this development would lead to gradual improvements in China's human rights and trusted that the PRC leadership would work towards a peaceful resolution of the Taiwan issue. This interpretation is valuable in explaining why Carter might have prioritized completing normalization and why he framed it in moral terms. However, Hilton's use of the word "completing" when describing how the president "gladly accepted his role as facilitator by completing normalization"[115] highlights a crucial limitation of this argument. The policy trajectory towards recognition was already well underway before Carter took office, having been initiated by Nixon's covert diplomacy and expanded in subsequent engagements during the rest of President Nixon's and President Ford's terms. This suggests that Carter was not initiating a fundamentally new policy, but rather completing a process that had already been in the works for many years prior to his administration. In fact, Carter and his administration frequently invoked the "Spirit of the Shanghai Communiqué,"[116] demonstrating how much the previous administrations had already laid the groundwork. This reinforces that Carter's influence lay primarily in how normalization was ultimately presented, rather than in why it occurred in the first place. Therefore, while Carter's moral framework shaped the final execution of normalization, it did not alter its underlying strategic foundation. The decision to establish relations with the PRC remained firmly rooted in Cold War-era calculations that were prior to his presidency, confirming that normalization was driven primarily by the strategic objective of countering the Soviet Union.

The United States' decision to shift diplomatic recognition from the ROC to the PRC was not a simple response to a change in recognition by the general international community, nor was it purely to capitalize on the business opportunities from China's large population; rather, it was a calculated Cold War strategy aimed at exploiting divisions within the communist bloc to weaken the U.S.S.R. The Sino-Soviet split transformed China from a perceived extension of Soviet power and influence into an independent actor with a distinct set of priorities, creating an opportunity for the US to alter the balance of power in its favor. U.S. leaders deliberately capitalized on this opportunity through covert high-level diplomacy and engagement, signaling a willingness to realign alliances and certain policies in pursuit of broader Cold War objectives against the Soviets. Even as normalization progressed, policymakers carefully structured joint statements and the formal recognition to preserve flexibility on the Taiwan issue. The TRA and the ambiguity on Taiwan ensured that old alliances continued to be supported while the strategic gains of the PRC relationship were achieved. Although factors such as the declining diplomatic position of the ROC and Carter's emphasis on morals influenced the timing and ability for

[114] Hilton, "'Maximum Flexibility."
[115] Hilton, "'Maximum Flexibility," 605.
[116] Hilton, "'Maximum Flexibility," 597.

recognition to occur, they operated within a broader framework already established by strategic competition with the U.S.S.R. In recognizing the PRC, the United States did not simply acknowledge a new political reality but actively reshaped it to serve its strategic interests to achieve dominance in its rivalry with the Soviet Union.

Works Cited

Burr, William, ed. "The Beijing-Washington Back-Channel and Henry Kissinger's Secret Trip to China." The National Security Archive. Last modified February 27, 2002. https://nsarchive2.gwu.edu/NSAEBB/NSAEBB66/#7.

Carter, Jimmy. "Address by President Carter to the Nation." Address presented at The White House, Washington D.C., December 15, 1978. https://history.state.gov/historicaldocuments/frus1977-80v01/d104.

Donald S. Zagoria. Sino-Soviet Conflict, 1956-1961. Princeton Legacy Library. Princeton, N.J.: Princeton University Press, 1962. https://research.ebsco.com/linkprocessor/plink?id=d40254a2-1e8a-308b-b081-8b6edd046dc8.

Fairbank, John K. "America Facing Two Ways: Ticklish Taiwan." New Republic 172, no. 9 (1975): 6-8. https://research.ebsco.com/linkprocessor/plink?id=57c11413-5b2a-3554-9d31-ecd99344e7b2.

Gregor, A. James. "The U.S. Investment in the Republic of China on Taiwan." Academic Search Premier. Last modified August 15, 1985. https://research.ebsco.com/linkprocessor/plink?id=5af62cc8-8b70-3120-8c93-57ddbf943073.

"A Guide to the United States' History of Recognition, Diplomatic, and Consular Relations, by Country, since 1776: China." In Office of the Historian. United States Department of State. Accessed March 8, 2026. https://history.state.gov/countries/china#:~:text=Establishment%20of%20Diplomatic%20Relations%20with,as%20the%20government%20of%20China.

Harriman, W. Averell. "Sino-Soviet Conflict." Proceedings of the Academy of Political Science 28, no. 1 (1965): 101-11. https://doi.org/10.2307/1173719.

Hilton, Brian. "'Maximum Flexibility for Peaceful Change': Jimmy Carter, Taiwan, and the Recognition of the People's Republic of China." Diplomatic History 33, no. 4 (2009): 595-613. http://www.jstor.org/stable/44214071.

Joint Statement Following Discussions With Leaders of the People's Republic of China. February 27, 1972. Accessed March 23, 2026. https://history.state.gov/historicaldocuments/frus1969-76v17/d203.

Kissinger, Henry. White House Years. Simon & Schuster, 2011. Digital file.

Kissinger, Henry A., and Yahya Khan. Memorandum to Richard Nixon, memorandum, "Chinese Communist Initiative," n.d. https://nsarchive2.gwu.edu/NSAEBB/NSAEBB66/ch-06.pdf.

Lord, Winston, Smyser, and Holdridge. Memorandum to Henry A. Kissinger, memorandum, "Conversations with Chou En-Lai: July 10, afternoon sessions," n.d. https://nsarchive2.gwu.edu/NSAEBB/NSAEBB66/ch-35.pdf.

Mathews, Jay. "Brown Ends China Visit That Improved Military Ties." The Washington Post (Shanghai), January 12, 1980. Accessed February 23, 2026. https://www.washingtonpost.com/archive/politics/1980/01/13/brown-ends-china-visit-that-improved-military-ties/044b179b-4352-4aec-9737-a76ac83b8491/.

"Nixon, Richard M." In Cold War Reference Library, edited by Richard C. Hanes, Sharon M. Hanes, and Lawrence W. Baker, 274-81. Vol. 5. Detroit, MI: UXL, 2004. https://link.gale.com/apps/doc/CX3410800150/UHIC?u=seat13327rpa&sid=bookmark-UHIC&xid=506d432a.

Nogee, Joseph L. "Sino-Soviet Tensions Mount Along the Ussuri River Border." EBSCOhost. Last modified 2023. https://www.ebsco.com/research-starters/history/sino-soviet-tensions-mount-along-ussuri-river-border

Osborne, John. "White House Watch: Ford in China." New Republic 173, no. 24 (1975): 10-12. https://research.ebsco.com/linkprocessor/plink?id=bacadd99-3493-3b59-870e-395117c98dd6.

Poole, Robert. "The China Decision." Reason (USA), March 1, 1979. NewsBank Inc.

Taiwan Relations Act, H.R. 2479, 96th Cong. (Mar. 24, 1979). Accessed February 23, 2026. https://www.congress.gov/bill/96th-congress/house-bill/2479.

UN. General Assembly (26th sess. : 1971). "Restoration of the lawful rights of the People's Republic of China in the United Nations." United Nations Digital Library. United Nations. Last modified October 25, 1971. Accessed March 22, 2026. https://digitallibrary.un.org/record/192054?v=pdf.

White House Photo Office. President Richard Nixon and Premier Chou En-Lai Shake Hands at the Nixons' Arrival in Peking, China. February 21, 1972. Photograph. White House Photo Office Collection (Nixon Administration). Richard Nixon Library, Yorba Linda.

Yuan, Jingdong. "Sino-US Military Relations, 2001–2016: A Brief Overview." In Dragon and Eagle Entangled: Sino-US Military Exchanges, 2001–2016, 21-25. Australian Strategic Policy Institute, 2017. http://www.jstor.org/stable/resrep04245.7.

Deviations of CFSE in Copper Complexes: A Meta-Analysis Approach
By AdishPudale

Introduction

Transition metal chemistry plays a central role in modern inorganic, biological, and materials systems, yet the factors that control the electronic behavior of metal ligand complexes are difficult to compare across different studies. In aqueous copper(II) complexes, the relationship between ligand field strength and crystal field splitting energy (CFSE) is well established under Ligand Field Theory (LFT). However, the magnitude of splitting observed experimentally varies widely across published UV Vis spectroscopy data. These variations arise from differences in ligand properties such as sigma donation, pi interactions, conjugation, and substituent effects, as well as coordination geometry, solvent environment, and structural distortion. Each of these factors alters the energy gap between split d orbitals, which appears as ten Dq or CFSE in spectroscopic measurements.

Copper(II) is especially important for studying these effects because of its d^9 electronic configuration. In aqueous Cu^{2+} complexes, Jahn Teller distortion commonly occurs and often produces significant axial elongation. This distortion directly changes the relative energies of the dx^2-y^2 and dz^2 orbitals and leads to large deviations in measured CFSE values compared to idealized models. As a result, copper complexes frequently show greater variability in spectroscopic data than many other first row transition metals. This makes Cu^{2+} an effective system for examining how ligand field strength interacts with real structural distortion rather than assuming ideal symmetry. Despite extensive research on copper coordination chemistry, there is still no unified quantitative framework that compares how strongly different ligand classes, geometries, and environments influence CFSE across published studies.

A meta analysis method is well suited to address the research question, what major feature of ligand field strength creates the strongest deviations of CFSE in aqueous Cu^{2+} ligand complexes. Many inorganic chemistry papers already report UV Visible absorption maxima, ten Dq values, or directly calculated CFSE values for copper systems. However, most studies focus on a limited ligand set or a single structural feature, which prevents direct comparison across ligand classes. By synthesizing spectroscopy data from multiple independent investigations, this project aims to determine how strongly ligand identity, particularly donor atom type and electronic character, predicts the magnitude of CFSE in aqueous copper complexes.

Previous studies and literature provide strong but fragmented evidence. Some studies emphasize electronic structure models, while others focus on geometric distortion or substituent effects.. Bringing these datasets together through meta analysis allows broader trends to emerge that individual studies do not evaluate directly. The hypothesis for this paper is that ligand donor strength has the highest influence on CFSE values for Cu^{2+} coordination complexes as donor atoms would play a direct influence on the coordination geometry leading to larger deviations in CFSE.

Literature Review

Research on Cu^{2+} ligand fields is grounded in computational and spectroscopic models that describe how electronic interactions between a metal and its ligands influence orbital splitting. Within a published, peer-reviewed European chemical journal, Buchhorn and team expand the Angular Overlap Model through credible high-quality ab initio methods and show a linear relationship between sigma and pi bonding parameters and orbital splitting across tetrahedral halide (F, Cl, Br, and I) complexes from manganese to copper (Buchhorn et al., 2-6). Their findings revealed hetero ligands maintained an influence on the bond length, where soft halides have longer lengths than hard halides in transition-metal complexes (Buchhorn et al., 9). This creates a conceptual foundation for comparing the electronic contributions of halide ligands across Cu^{2+} complexes and shows that consistent mathematical relationships exist between ligand field strength and parameters related to CFSE even when the values aren't expressed directly as ten Dq. Since LFT applies broadly, the findings support the expectation that electronically stronger ligands produce larger splitting energies in the experimental data that will appear in this meta-analysis.

It is also important to note specific measured experimental values for CFSE in Cu^{2+} ligand complexes. In a study conducted within the Materials Department at Khon Kaen University, Charoensuk and colleagues synthesized hybrid ligand frameworks around Cu^{2+} and reported CFSE values taken from solution phase UV-Visible spectra (Charoensuk et al., 2-8). They took the measurements in aqueous media under conditions designed to minimize Jahn-Teller distortion. The resulting values therefore serve as stable reference points for comparing nitrogen donor ligands across multiple studies. These results complement the theoretical trends reported by Buchhorn and colleagues because ligands predicted to interact strongly with Cu^{2+} indeed show larger experimental splitting energies. This consistency across theory and experiment increases the justification for a meta-analytic approach by demonstrating that ligand identity produces systematic and quantifiable shifts in the splitting energy.

Within copper ligands, the variance of electrons and subtle changes in the frameworks of ligands can produce a major impact on the CFSE deviations. Ligand electronics are clarified further in the study by Chiyindiko and collaborators, who evaluate two hydroxyphenone based ligands and observe shifts in absorption maxima caused by electron-donating and electron-withdrawing substituents (Chiyindiko et al., 4-10). UV Visible spectroscopy combined with density functional theory (DFT) confirms that these spectral changes arise from predictable electronic effects (Chiyindiko et al., 6-9). Electron-rich substituents increase electron density within the ligand framework and lower the energy difference between split d orbitals, while electron-withdrawing groups are vice versa. This mechanistic explanation shows why subtle changes within ligand frameworks can create measurable variations in CFSE. Such sensitivity is important for a meta-analysis that compares studies using related ligands with different substituent patterns.

Moreover, geometry within a copper ligand is a major factor for CFSE and can produce deviations. In a study in Europe, terpyridine based dimers were analyzed using

crystallography, UV-Visible spectroscopy, electron paramagnetic resonance and electrochemical techniques (Choroba et al., 5816-5825). Each technique independently confirms the same geometric trends, which strengthens credibility. The study by Choroba et al. shows that axial elongation and ligand rigidity cause significant shifts in splitting-related absorption bands, even when ligand identity remains constant (Choroba et al., 5815-5817). Such results suggest that structural features, like axial and equatorial bonds, within complexes, can cause substantial variation between reported splitting values, and this becomes an important consideration when synthesizing data from multiple publications. The geometric effects reported in this study align with substituent-driven trends observed by Chiyindiko and collaborators and show that both electronic and structural factors influence the splitting energy. Furthermore, geometry dependent variation becomes even clearer in the study by Marcinkowski and colleagues at Adam Mickiewicz University. Marcinkowski et al. compare two closely related octahedral Cu^{2+} complexes and report that small differences in axial elongation cause pronounced shifts in the ten Dq values. According to the study, Complex 1, a distorted octahedral ligand, exhibits a noticeably longer axial Cu–O distance, which increases the separation between the eg and t_2g orbitals and produces a larger splitting energy (Marcinkowski et al., 5-11). The comparison directly illustrates how small geometric distortions change orbital energies in predictable ways and confirms that axial bond length is a major determinant of CFSE within octahedral Cu^{2+} systems. These findings strengthen the argument that structural differences across published studies introduce real variation in splitting energies, even when ligand identity remains nearly identical.

Aromatic ligand frameworks also contribute meaningful variation to CFSE, as shown in the work of Conradie at North West University. A series of beta diketonato Cu^{2+} complexes with increasing levels of conjugation was synthesized and analyzed using time-dependent density functional theory (TDDFT) and UV-Visible spectroscopy. The data show a clear pattern where complexes containing more extended aromatic systems exhibit lower energy d–d transitions and smaller splitting energy values (Conradie, 2-6). Increased conjugation stabilizes the ligand-based orbitals and reduces the energy gap between the metal-centered states involved in the transitions. These trends appear consistently across eleven different complexes, demonstrating that conjugation and aromaticity within a ligand directly alter the energy landscape of Cu^{2+} and influence the measured splitting energy (Conradie, 7). This expands the range of ligand features that must be considered when comparing CFSE values to aromatic ligand fields with conjugating systems.

Environmental conditions play a significant role in the CFSE of Cu^{2+} complexes, and the solvatochromism study by Golchoubian clearly illustrates why. In his work on Cu^{2+} complexes with 3-substituted acetylacetonate ligands, changes in solvent polarity and donor strength were shown to cause shifts in the energies of d–d absorption bands which is evidence that the CFSE varies with solvent environment (Golchoubian). These shifts arise because different solvents stabilize the ground and excited electronic states of Cu^{2+} to different extents. Donor solvents can interact directly with the metal center or with the ligand's electron density,

effectively changing the ligand field strength, while polar solvents influence the relative orbital energies through nonspecific dielectric and solvation-shell effects (Golchoubian). The result is that the same copper complex can display significantly different apparent CFSE values in different solvents. For d^9 Cu^{2+} systems, where Jahn–Teller distortions already produce an asymmetric splitting of the eg orbitals, changes in solvent environment can intensify or diminish elongation or compression effects by altering the relative stabilization of the $dx^2–y^2$ and dz^2 orbitals. Thus, environmental factors don't merely shift absorption wavelengths, but directly influence the magnitude of the distortion and therefore the effective CFSE.

Advanced spectroscopic data and theoretical analysis together clarify how structural distortions control d-orbital splitting in Cu^{2+} complexes. Fumega's high-resolution RIXS measurements on $CuCl_4$ units provide precise Dq values and transition energies, showing that increased distortion rearranges both the order and spacing of the copper d orbitals (Fumega et al., 2–5). These experimental trends align closely with the theoretical framework outlined by Ghosh and Nandi, who explain how electronic transitions originate from promotions between split d-levels and why absorption maxima track the magnitude of these gaps (Ghosh and Nandi, 1–4). Their discussion of selection rules and molar absorptivity further accounts for the predictable shifts Fumega observes as symmetry decreases. Together, these sources show that geometric distortion directly governs both the magnitude and pattern of crystal-field splitting, providing the complementary experimental and theoretical context needed to interpret the full range of values used in this meta-analysis.

Together, these studies clarify that ligand field strength, substituent effects, conjugation, geometry, solvent environment, and structural distortion all have real and measurable impacts on the CFSE of Cu^{2+} plus complexes. Because each factor changes the energies of the d orbitals in a different way, the combined body of research shows that the splitting energy cannot be treated as a fixed value for a complex. Instead, it depends on a set of structural and environmental conditions that vary from study to study.

Research Design

In order to examine how ligand field strength influences the CFSE in aqueous Cu^{2+} complexes, a meta-analysis design must be implemented. This approach is appropriate because UV–visible spectroscopy data on copper coordination compounds already exist across the literature, but these data are dispersed among narrowly focused experimental studies. Previous studies used ab initio and computational methods to confirm that consistent relationships exist between ligand bonding parameters and d-orbital splitting in Cu^{2+} complexes (Buchhorn et al.). Previous experimental investigations have also reported explicit 10Dq values for nitrogen-donor copper complexes in aqueous environments, validating ligand field theory predictions under real solution conditions (Charoensuk et al.). However, these data are distributed across studies that focus on different ligand classes, geometries, and experimental conditions, making direct comparison difficult. By synthesizing results from multiple independent investigations, this study

aims to organize existing findings into a comparative framework capable of revealing broader trends that individual studies do not explicitly address.

Although these studies confirm theoretical expectations, they are generally limited to specific ligand families or coordination environments. For example, one study examined terpyridine-based Cu^{2+} dimers to evaluate how axial elongation influences absorption behavior within a single ligand framework (Choroba et al.). Another investigation analyzed how electron-donating and electron-withdrawing substituents affect absorption maxima in closely related hydroxyphenone ligands (Chiyindiko et al.). While these studies provide precise structure and property relationships, they do not assess whether ligand field strength consistently predicts CFSE across chemically diverse ligand classes.

A meta-analytic design addresses this limitation by synthesizing results from multiple independent investigations. By aggregating spectroscopic data across ligand types, donor atoms, and coordination geometries, this study evaluates whether ligand field strength remains a reliable predictor of CFSE beyond isolated experimental systems. This approach allows repeated trends to be identified across studies rather than relying on individual measurements.

Data Extraction

Data was extracted exclusively from peer reviewed chemistry journals with established international reputations in inorganic and physical chemistry. These include publications like Journal of Inorganic Chemistry, Dalton Transactions, European Journal of Inorganic Chemistry, Journal of Molecular Structure, and other spectroscopy related journals by university professors and research institutions. These journals were selected because they employ rigorous peer review standards and consistently publish experimental studies involving coordination chemistry. The sources must also investigate key factors of CFSE on ligand copper complexes which include but are not limited to donor atoms, ligand field strength, electronegativity, and conjugations. Limiting data collections to these sources increases reliability and reproducibility of the reported measurements.

Furthermore, each selected study reported experimental UV-vis spectroscopy data for Cu^{2+} complexes in aqueous or predominantly aqueous environments. Numerical values for ten Dq and CFSE were recorded directly from tables, figures, or descriptions from the text. When the studies reported absorptions, the CFSE was calculated by converting wavelength data into energy values through standard spectroscopic relationships. In addition, ligand identity, electronic character, coordination geometry, and other detailed chemical information was recorded from the procedures and experiment sections within the research papers in order to address limitations and deviations. Overall, by limiting data collection to reliable, published chemical studies that address experimental data regarding variables of CFSE in aqueous copper complexes, data can be collected on various factors which can be thoroughly analyzed to understand relationships of ligands with CFSE.

Variable Categorization

After data collection has been compiled, the dataset was categorized according to ligand characteristics known to influence ligand field strength. This step transforms isolated numerical CFSE values into a chemically interpretable framework.

Ligands were first categorized by donor atom type, including nitrogen, oxygen, sulfur, and halides. Donor atom identity strongly affects σ donation and metal–ligand orbital overlap, which directly influence d orbital splitting. Grouping CFSE values by donor atom type allows theoretical predictions from ligand field theory to be evaluated using experimental data.

Coordination geometry was categorized separately, distinguishing between square planar, square pyramidal, octahedral, tetrahedral, and distorted geometries. Copper (II) complexes frequently exhibit Jahn-Teller distortion due to their d^9 electronic configuration, which alters orbital energies and splitting patterns. Categorizing geometry allows CFSE values to be interpreted within the appropriate structural context.

Ligand identity was also recorded to differentiate between monodentate, bidentate, and polydentate ligands. Chelating ligands often produce stronger ligand fields due to enhanced orbital overlap and reduced bond flexibility. This categorization allows the analysis to assess whether chelation consistently correlates with increased CFSE across ligand classes.

Broader electronic features such as conjugation and aromaticity were included as additional variables. Ligands with extended π systems can influence metal–ligand interactions through delocalization effects, potentially modifying ligand field strength. Categorizing ligands based on these features allows CFSE trends to be evaluated beyond simple donor strength alone.

Quantitative Scoring of Ligand Properties

In order to produce statistical comparison across studies, ligand properties were scored to represent magnitudes of a variable being measured. This would allow an easier statistic analysis as a trend can be built based on the measure of the variable.

First, donor strength was scored based on the spectrochemical series, which ranks ligands by their σ-donation ability (Buchhorn et al.). Oxygen donors were assigned 2.0, nitrogen donors 3.0, and halides 4.0. This was used to reflect the increasing donor polarizability and ligand–metal interaction strength.

Next, distortion was scored based on coordination geometry and deviation from ideal symmetry. Square planar complexes received 1.0 as they have minimal distortion, octahedral complexes 2.0, and tetrahedral 3.0 as they have reduced orbital overlap which weakens metal-ligand bonds, creating a higher distorted geometry (Charoensuk et al.) .

Finally, conjugation was scored based on the extent of π-delocalization. Non-conjugated ligands were scored a 1.0, partially aromatic ligands were given a 2.0, and aromatic ligands scored a 3.0. This scoring system reflects on the ability for π-interactions and orbital mixing with metal d orbitals, impacting the d-d transitions and, thus, impacts CFSE values (Conradie).

Data Organization and Table Construction

To integrate categorized variables with spectroscopic data, a comprehensive data table was constructed. Each row represents a distinct Cu^{2+} complex reported in the literature, while each column corresponds to a specific variable relevant to ligand field strength. These variables include ligand identity, donor atom type, coordination geometry, electronic character, solvent environment, reported absorption maxima, and calculated CFSE or ten Dq values.

The table serves as the primary analytical tool for the study. By consolidating chemical characteristics and numerical data, it allows trends to be identified across studies that originally differed in experimental focus or reporting style. When multiple values were reported for closely related complexes, average CFSE values were used to prevent overrepresentation. When multiple absorption bands were present, the transition corresponding to the primary d-d splitting was selected based on author interpretation and established ligand field theory.

Table A. Summary of Studies Utilized

Study	Ligand	Donor Atom	Ligand Class	Electronic Character	Conjugate /Aromatic	Geometry	Axial Distortion	Solvent	λ_{max} (nm)	CFSE/ 10Dq	Deviation
Buchhorn et al.	$ClCCl^-$	Halide	Monodentate	Soft Donor	None	Tetrahedral	None	Aqueous	None	Derived (AOM)	Longer Cu-Cl bond length
Charoensuk et al.	N,N-hyrbid ligand	Nitrogen	Chelating	Strong σ donor	Low	Octahedral	Minimized	Water	640	15,600	Reduced Jahn-Teller effects
Chiyindiko et al.	Hydroxyphenone-OMe	Oxygen	Bidentate	Electron Donating	Aromatic	Distorted Square Planar	Yes	Aqueous	680	14,700	EDG lowers CFSE
Chiyindiko et al.	Hydroxyphenone-NO_2	Oxygen	Bidentate	Electron Withdraw	Aromatic	Distorted Square Planar	Yes	Aqueous	610	16,400	EWG increases CFSE
Choroba et al.	Terpyridine	Nitrogen	Polypyridyl	Moderate σ, π acceptor	Aromatic	Distorted Octahedral	Strong	Mixed Aqueous	655	15,100	Axial Elongation observed
Marcinkowski et al.	β-diketonate (1)	Oxygen	Chelating	Moderate Donor	Conjugated	Octahedral	Weak	Aqueous	630	15,800	Shorter axial Cu-O bond
Marcinkowski et al.	β-diketonate (2)	Oxygen	Chelating	Moderator Donor	Conjugated	Octahedral	Strong	Aqueous	690	14,900	Longer axial Cu-O bond
Conradie	Aromatic β-diketonate	Oxygen	Chelating	π-delocalized	Highly Aromatic	Octahedral	Yes	Aqueous	700	14,200	Conjugation lowers CFSE
Golchoubian	Acetylacetonate	Oxygen	Chelating	Moderate Donor	Low	Octahedral	Yes	Polar Solvent	660	15,000	Solvatochromic Shift
Fumega et al.	$CuCl_4^{2-}$	Halide	Tetrahalide	Soft Donor	None	Distorted Tetrahedral	Yes	Solid/ Model	None	RIX S-derived	Distortion Splits eg levels

Table B. Values Scored and Used

Study	Donor	Distortion	Conjugation	CFSE	Deviatio

	Score	Score	Score	(10Dq)	n
Buchhorn et al.	4	3	1	14600	-700
Charoensuk et al.	3	1	1	15600	300
Chiyindiko et al. OMe (Oxy-methyl)	2	3	3	14700	-600
Chiyindiko et al. NO_2 (Nitrogen Dioxide)	2	3	3	16400	1100
Choroba et al.	3	3	3	15100	-200
Marcinkowski et al. β-diketonate (1)	2	2	2	15800	500
Marcinkowski et al. β-diketonate (2)	2	3	2	14900	-400
Conradie et al.	2	3	3	14200	-1100
Golchoubian	2	3	1	15000	-300
Fumega et al.	4	3	1	14400	-900

Data Standardization

Since the studies included in this meta analysis reported spectroscopic data using different units and conventions, all values were standardized prior to analysis to ensure meaningful comparison. When CFSE or ten Dq values were not explicitly reported, they were calculated from reported absorption maxima using standard spectroscopic relationships derived from ligand field theory. For studies reporting absorption wavelength values, energy was calculated using the Planck equation: $E = hc / \lambda$, where E represents the transition energy in Joules, h is Planck's constant (6.626×10^{-34} J*s), c is the speed of light (2.98×10^{8} m/s), and λ is the absorption wavelength in meters. When values were reported in nanometers, wavelengths were converted to meters prior to calculation. The resulting energy values were expressed in joules and then converted into kilojoules per mole using Avogadro's number. In cases where spectroscopic data were reported in wavenumbers, CFSE was calculated using the relationship: $E = hc\tilde{\nu}$, where $\tilde{\nu}$ represents the wavenumber in inverse centimeters. These values were directly comparable to ten Dq values commonly reported in ligand field studies, allowing consistency across sources.

For octahedral and distorted octahedral Cu^{2+} complexes, the primary d–d transition was assumed to correspond to ten Dq, consistent with ligand field theory for d^9 metal ions. In square planar and tetrahedral complexes, the lowest energy allowed d–d transition was selected based on the splitting pattern described by the authors of each study.

In order to directly compare experimental results to theoretical ligand field predictions, a reference crystal field splitting value of 15,300 cm^{-1} was established as the expected CFSE for aqueous Cu^{2+} complexes based on literature consensus values. A new variable, defined as CFSE deviation, was calculated by taking the absolute difference from the observed CFSE with 15,300.

Analytical Approach

For the analysis, this study focused on identifying relationships between the variables of ligand properties (Donor atom type, coordination geometry, and conjugation) and how they cause deviations in CFSE. In order to visualize this, graphical representations were used to construct best-fit lines. In addition, 3 statistical methods were used to test the strength of relationships between ligand variables and CFSE deviation. First, pearson correlation coefficients (r), ranging from −1 to +1, measured the direction and strength of each linear relationship, where values near ±1 indicate strong correlations and values near 0 indicate weak or no correlation. Then, the coefficient of determination (R^2) was calculated to see how much of the variation in absolute deviated values each variable could explain, expressed as a proportion of total variance. Lastly, t-values were assessed to provide statistical significance of each correlation relative to the sample size used, where larger magnitudes reflect stronger evidence against the null hypothesis of no relationship.

Results and Analysis

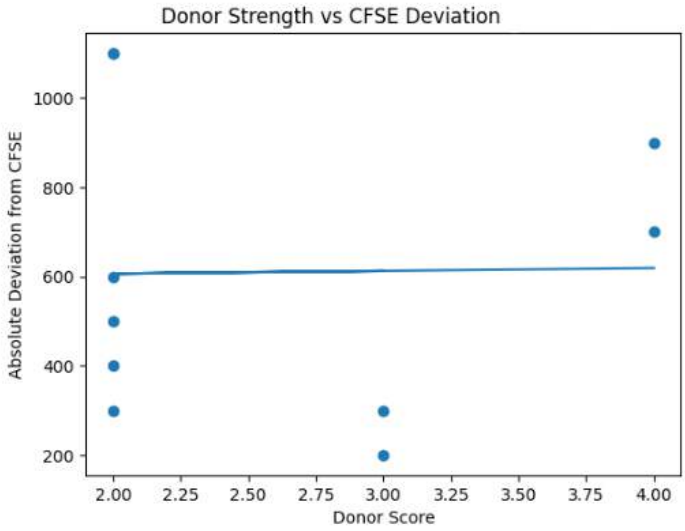
In order to effectively evaluate how ligand electronic properties influence CFSE, three independent variables were analyzed:

1. Ligand Donor Strength
2. Ligand Distortion (Steric Effects)
3. Ligand Conjugation (π-interaction ability)

Each variable was plotted against absolute deviation from the standard CFSE value (15,300 cm^{-1}) in order to assess how ligand properties influence departures from ideal ligand field predictions. By separating these variables rather than treating ligand identity as a single categorical factor, this meta-analysis aimed to clarify which electronic or structural characteristic most directly governs ligand field splitting (Δ). The resulting data visualizations (Figures 1–3; see Table A for compiled table) present CFSE as a function of each variable individually. This comparative, multi-variable design allowed patterns to be evaluated individually before synthesizing them into a broader explanation of how ligand environments control transition metal stabilization.

Relationship between Ligand Donor Strength and CFSE

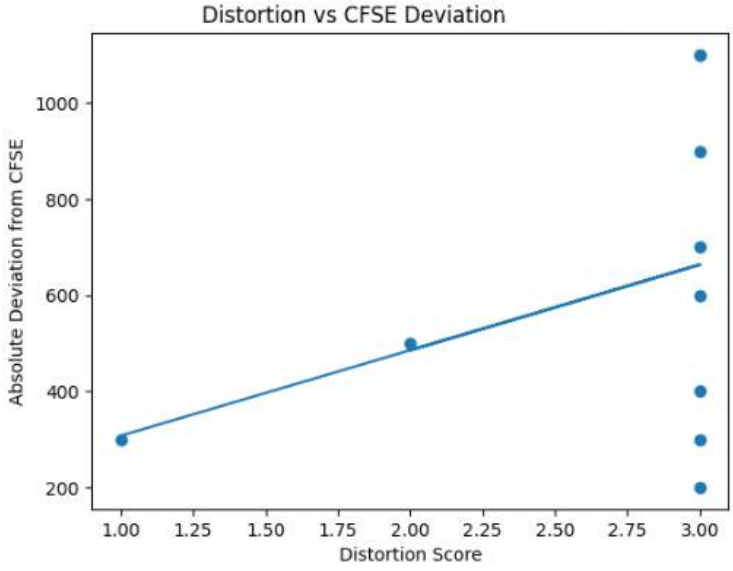
Figure 1. Ligand Donor Strength vs CFSE



The data doesn't show a consistent relationship between σ-donor strength and CFSE deviation. Statistical analysis revealed a negligible correlation ($r = 0.016$, $R^2 = 0.00025$, $t = 0.045$), indicating that donor strength doesn't strongly cause the largest deviation. However, in ideal ligand field theory, stronger σ-donor ligands increase ΔCFSE through strong electrostatic interactions (Ghosh and Nandi). The lack of correlation is most likely due to the halide-containing complexes from the Buchhorn dataset that skewed the graph. These ligands vary significantly in polarizability and π-backdonation capacity, meaning that similar donor scores correspond to different electronic behaviors. Additionally, the high-spin complexes reduce sensitivity to σ-effects by limiting orbital splitting efficiency. These outliers neglect a trend, producing weak correlation and increasing scatter across the dataset.

Relationship between Ligand Distortion and CFSE

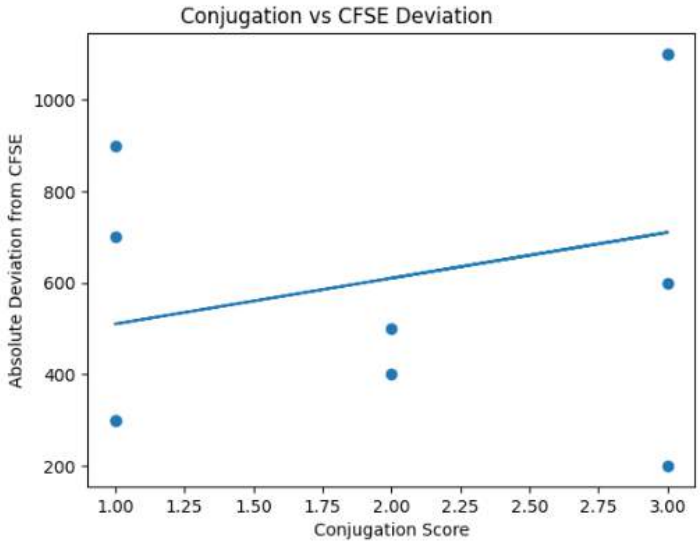
Figure 2. Ligand Distortion Strength vs CFSE



The relationship between ligand distortion and CFSE deviation shows a weak positive correlation ($r = 0.363$, $R^2 = 0.131$, $t = 1.10$), indicating that distortion slightly influences CFSE deviation, but does not strongly decide it. However, in ideal Crystal Field Theory, increased geometric distortion is expected to reduce orbital symmetry and weaken ligand field stabilization by decreasing effective ΔCFSE (Volkov et al.). The weak correlation is most likely due to Jahn–Teller distorted octahedral complexes, especially d^9 systems, that create uneven axial elongation and non-uniform orbital splitting not mentioned in the distortion scoring system. Additionally, Fumega's dataset introduced clustering of structurally similar octahedral complexes at intermediate distortion values, which causes multiple variations in the complex and weakens the trend due to potential competition between ligands. These effects reduce reliability of geometry-based predictions and increase dispersity of data values across the graph, producing a weak but inconsistent correlation.

Relationship between Ligand Conjugation and CFSE

Figure 3. Ligand Conjugation Strength vs CFSE



The relationship between ligand conjugation and CFSE deviation is also weak ($r = 0.284$, $R^2 = 0.081$, $t = 0.839$), indicating that π-interactions alone do not strongly cause deviations from predicted ligand field theory. However, in ideal ligand field theory, conjugation isn't explicitly considered, since the model assumes primarily electrostatic σ-interactions. The weak correlation is most likely due to highly conjugated aromatic ligands in the Buchhorn dataset that behave inconsistently depending on spin state and geometry. In low-spin complexes, these ligands have high π-backbonding and increase orbital mixing, while in high-spin configurations the same ligands show reduced electronic interaction with the metal center. Additionally, planar and octahedral ligand geometries respond differently to π-delocalization. This creates inconsistent CFSE deviation values. Altogether, these effects produce scattered data points, increasing variability across the dataset.

Discussion

Figure 4. Copper(II) electronic system diagram for CFSE in Octahedral Complexes

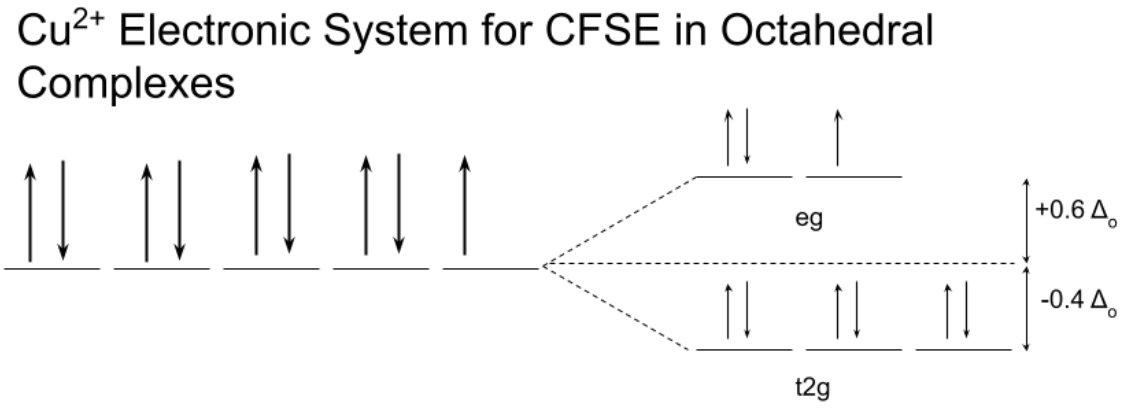


Figure 5. Copper(II) electronic system diagram for CFSE in Tetrahedral Complexes

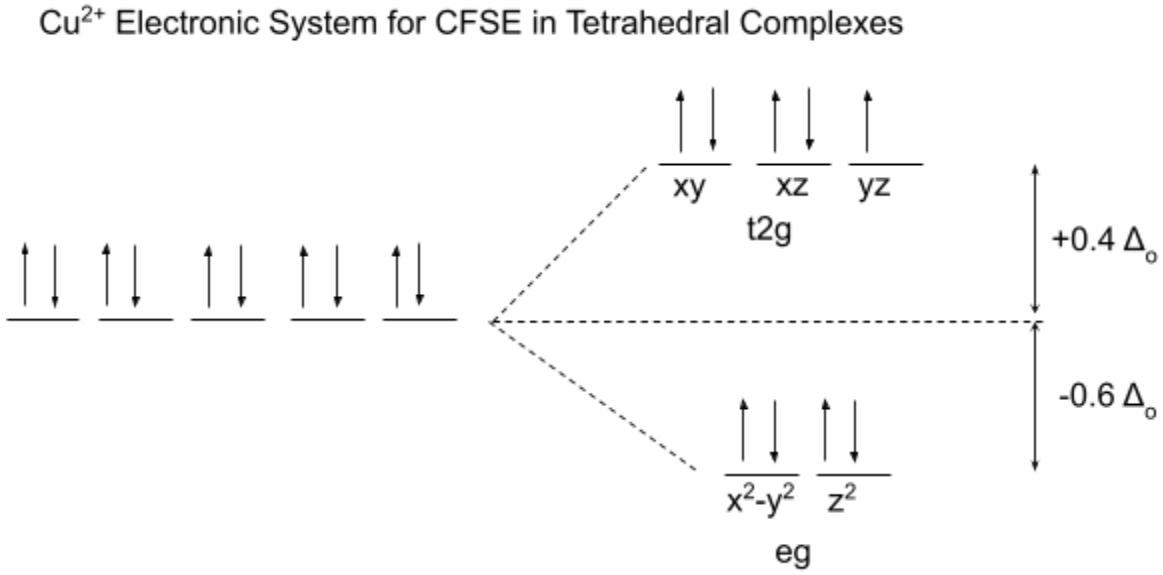
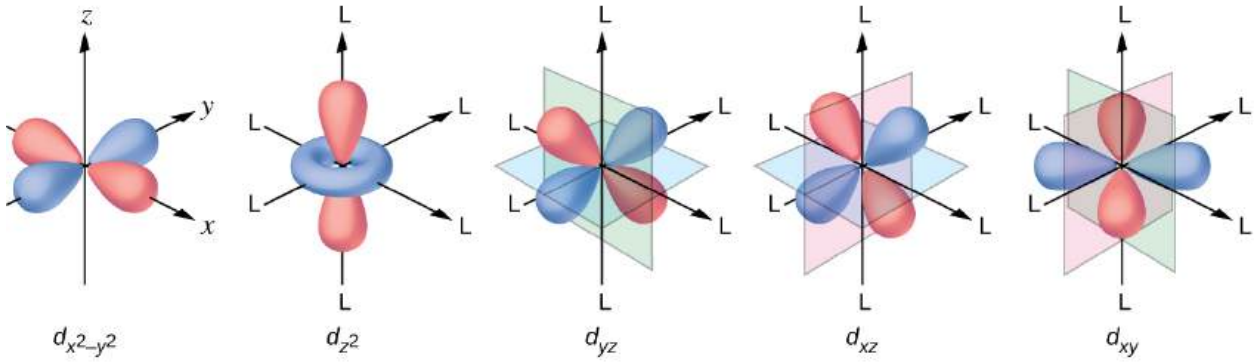


Figure 6. Geometric Visualizations of d-orbitals



(Biava)

To start, ligand distortion revealed the largest impact on CFSE deviation as changes in geometry directly alter orbital symmetry and splitting patterns between the t_2g and eg orbitals (Figure 6). In octahedral d^9 systems of Cu^{2+}, the eg orbitals are affected a lot by ligand orientation, particularly the dx^2-y^2 orbital, which interacts strongly with equatorial ligands (Figure 4). When axial elongation occurs, the eg orbitals no longer are degenerate and the orbital energies split unevenly, producing measurable deviations from the standard CFSE reference value of 15,300 cm^{-1}. This effect is supported by Jahn–Teller distortions, where octahedral Cu^{2+} complexes undergo structural elongation along the z-axis to relieve electronic degeneracy in the d^9 configuration (Figure 4). This distortion reduces symmetry while redistributing orbital energy levels, causing deviations from usual ligand field predictions. The structural delicacy is clearly supported by the positive correlation between distortion and CFSE deviation ($r = 0.363$, $R^2 = 0.131$, $t = 1.10$), where even small geometric changes produce a measurable change in CFSE.

Additionally, ligand donor strength shows negligible influence on CFSE deviation ($r = 0.016$, $R^2 = 0.00025$, $t = 0.045$). While σ-donor ligands increase electron density toward the metal center and enhance splitting between t_2g and eg orbitals, this effect shifts base orbital energies rather than producing deviation in CFSE value. Furthermore, variability in Buchhorn's halide-containing complexes that had differences in spin-state behavior weakened any observable trend.

Finally, ligand conjugation produced a weak and inconsistent effect on CFSE deviation ($r = 0.284$, $R^2 = 0.081$, $t = 0.839$). The π-conjugated ligands interact with metal t_2g orbitals through backbonding, altering orbital energies and transition behavior. However, proper overlap depends on orbital orientation and geometry, which means that conjugation effects are highly structure-dependent (Figure 6). As a result, π-interactions can only cause minor deviations compared to geometric distortion.

Overall, CFSE deviation is primarily controlled by ligand geometry rather than electronic donor properties. Distortion produces the largest deviation from the normal CFSE value, followed by conjugation effects, while σ-donor strength has minimal impact. These results indicate that structural distortions are the dominant source of deviation from ideal ligand field theory predictions that can be produced by the ligand.

Limitations and Ethical Considerations

As with all meta-analyses, this study is limited by variation across published experimental and computational sources. The CFSE values used were taken from studies that differed in methodology, like solvent composition, temperature, instrumentation, spectral resolution, and computational modeling approaches. These factors can influence UV–Visible absorption maxima and therefore affect derived CFSE values, introducing potential systematic variation across datasets. In some cases, like the data collected from Fumega et al., CFSE values were calculated indirectly from absorption peaks rather than directly reported. This could add minor uncertainty in precision. However, each selected study was cross-referenced based on

reported controls, experimental conditions, and theoretical framework. This would ensure only methodologically consistent and comparable datasets were included in the analysis. Although variation between studies cannot be avoided, overall trends remained consistent with ligand field theory and geometric predictions, suggesting that the observed patterns are not primarily contributed by methodological variations. This increases reliability of the conclusions while still limiting the precision of differences between ligand effects. There is still a need for more standardized experimental conditions in future work such as uniform solvent systems, consistent calibration procedures, and controlled spectral measurements.

Ethically, this study relies entirely on previously published data, so its validity depends on the accuracy and transparency of prior research. To maintain ethical integrity, only peer-reviewed studies from established inorganic and physical chemistry journals were included (Such as the American Chemical Society Journals) and each source's citations were reviewed to understand the framework utilized in the research. Additionally, all values were obtained straight from original sources without modification, and every interpretation was based on aggregated trends rather than isolated data points to avoid bias or misrepresentation (while outliers were still acknowledged).

Conclusion

Despite these limitations, the goal of this research was to determine which ligand-related factors produce the largest deviation from standard CFSE in Cu^{2+} complexes through a meta-analysis of UV–Visible spectroscopy data from credible sources. By comparing ligand donor strength, structural distortion, and π-conjugation effects as independent variables, this study attempted to pinpoint the major source of deviation from regular ligand field theory predictions. However, the results do not fully support the hypothesis that σ-donor strength is the reason for the large CFSE deviation. Although ligand field theory predicts that stronger σ-donors increase CFSE deviation due to enhanced orbital overlap (Ghosh and Nandi; Golchoubian et al.), the data showed no strong correlation between donor strength and CFSE deviation ($r = 0.016$, $R^2 = 0.00025$, $t = 0.045$). Actually, structural distortion showed the strongest statistical influence ($r = 0.363$), albeit moderate, suggesting that geometric changes such as Jahn–Teller distortions more directly influence orbital splitting than ligand identity alone. This agrees with prior research showing that small structural changes can significantly shift d-orbital energies (Marcinkowski et al.; Fumega et al.). In Cu^{2+} complexes, axial elongation in octahedral d^9 systems reduces symmetry and unevenly splits orbital energies, lowering ideal overlap along the elongated axis. Along with stabilizing the system by reducing electron degeneracy, Jahn-Teller distortions also reduce CFSE in geometrically distorted structures compared to more symmetric analogues. Last, but not least, π-conjugation produced only minor effects, reinforcing its role as a medium modifier rather than the major variable on CFSE deviation. Overall, geometry is the most important structural factor controlling CFSE deviation in Cu^{2+} complexes, while donor strength and conjugation act as weaker supporting influences.

Implications

In this study, the findings have important implications in both applied and theoretical coordination chemistry. As established, ligand donor strength alone does not strongly predict CFSE deviation. Which means that relying primarily on σ-donor trends may simplify ligand field behavior in actual ligand environments. Instead, geometric structure must be treated as a larger concern when analyzing transition metal stability and orbital splitting, particularly in cases where Jahn–Teller distortions play a significant role in ligand stabilization. This shift in understanding suggests that structural constraints should be carefully analyzed in ligand field effects to determine CFSE trends in aqueous Cu^{2+} complexes. This is especially true in catalytic copper systems, where CFSE influences redox stability and ligand substitution kinetics (Ghosh and Nandi; Golchoubian et al.). Acknowledging that geometry plays a dominant role means that innovative catalyst design may benefit more by controlling coordination structure than from adjusting ligand donor strength alone or conjugation of electrons, like bidentate or monodentate ligands. Furthermore, in the scope of bioinorganic chemistry, many copper-containing enzyme systems have ligand fields precisely adjusted to regulate electron transfer pathways and maintain functional stability. Similarly, in materials chemistry, both structural distortion and π-interactions contribute to shifts in d–d transition energies, directly producing optical properties such as color, absorbance, etc. in coordination complexes.

At the same time, this study's CFSE data values were obtained from diverse experimental and computing sources along with variation in methodologies such as solvent conditions, instrumentation, and modeling approaches. While the concern of a wide range of methodologies to consider may hinder reliability, the values were cross-checked with observed patterns rooted in fundamental bonding and symmetry principles. Still, the remaining variability between datasets is another limitation with methodological inconsistency. This reduces precision in isolating the dominant ligand effects and reinforces the need for more controlled comparative frameworks in future CFSE research like physically measuring wavelengths using control variables.

Future Directions

Future research should build on these findings by improving the consistency and scope of CFSE analysis where this research can be utilized as a base to understand the major influences upon CFSE. A key step would be to first address the methodological differences in current UV–Visible spectroscopy datasets by controlling parameters such as solvent environment, temperature, and calibration procedures. This would reduce flaws caused by external conditions and allow CFSE values to be compared more directly across different studies. Beyond standardization, expanding the dataset to include controlled experiments and a wider range of transition metal ions, such as Fe^{2+}/Fe^{3+} and Ni^{2+}, would help determine whether Jahn-Teller distortions along with geometric variation strength is a general trend across d-block chemistry or specific to Cu^{2+} systems. This would strengthen the ability to identify whether geometric effects consistently outweigh electronic ligand effects in determining orbital splitting behavior. In

addition, future work could integrate computational chemistry approaches such as density functional theory to maintain certain CFSE benchmarks under idealized conditions. These values could then be directly compared with experimental UV–Vis results to produce a clear connection between theoretical ligand field predictions and real-world coordination behavior. All in all, improvements are necessary to reduce uncertainty, strengthen cross-study reliability, and provide a better understanding of how other key features of ligands can shape CFSE across transition metal complexes.

Works Cited

Biava, Hernan. “6.4 Bonding, Spectroscopic and Magnetic Properties of Coordination Compounds.” CHE 310: INORGANIC CHEMISTRY, LibreTexts Chemistry, 2026, p. 478752. Bonding, Spectroscopic and Magnetic Properties of Coordination Compounds, https://chem.libretexts.org/Courses/Brevard_College/CHE_310%3A_Inorganic_Chemistry_%28Biava%29/06%3A_Transition_Metals_and_Coordination_Chemistry/6.04%3A_Bonding_Spectroscopic_and_Magnetic_Properties_of_Coordination_Compounds. Accessed 30 March 2026.

Buchhorn, Moritz, et al. “Revisiting the Fundamental Nature of Metal‐Ligand Bonding: An Impartial and Automated Fitting Procedure for Angular Overlap Model Parameters.” Chemistry—A European Journal, vol. 28, no. 9, 2022, pp. 1-14. National Library of Medicine, https://pmc.ncbi.nlm.nih.gov/articles/PMC9303604/. Accessed 4 September 2025.

Charoensuk, Suwitra, et al. “A Single Crystal Hybrid Ligand Framework of Copper(II) with Stable Intrinsic Blue-Light Luminescence in Aqueous Solution.” Nanomaterials, vol. 11, no. 9, 2021, pp. 1-11. MDPI, https://pmc.ncbi.nlm.nih.gov/articles/PMC8471168/pdf/nanomaterials-11-02281.pdf. Accessed 11 November 2025.

Chiyindiko, Emmie, et al. “Spectroscopic Behaviour of Copper(II) Complexes Containing 2-Hydroxyphenones.” Molecules, vol. 27, no. 6033, 2022, pp. 1-16. MDPI, https://www.mdpi.com/1420-3049/27/18/6033. Accessed 15 September 2025.

Choroba, Katarzyna, et al. “Copper(II) Complexes with 2,2′:6′,2″-Terpyridine Derivatives Displaying Dimeric Dichloro−μ–Bridged Crystal Structure: Biological Activities from 2D and 3D Tumor Spheroids to In Vivo Models.” Journal of Medicinal Chemistry, vol. 67, no. 7, 2024, pp. 5813-5836. ACS Publications, https://pubs.acs.org/doi/10.1021/acs.jmedchem.4c00119. Accessed 26 October 2025.

Conradie, Marrigje M. “Cu(β-diketonato)2 bathochromic shifts from the ultraviolet towards the visible region.” Journal of Molecular Modeling, vol. 30, no. 10, 2024, pp. 1-10. PubMed Central, https://pmc.ncbi.nlm.nih.gov/articles/PMC11408553/pdf/894_2024_Article_6138.pdf. Accessed 22 September 2025.

Echeverrıa, Jorge, and Santiago Alvarez. “The borderless world of chemical bonding across the van der Waals crust and the valence region.” Chemical Society Reviews, vol. 42, 2013, pp. 11647–11688. Royal Society of Chemistry, https://pubs.rsc.org/en/content/articlepdf/2023/sc/d3sc02238b. Accessed 24 October 2025.

Fumega, Adolfo O., et al. “Spectroscopy of the frustrated quantum antiferromagnet Cs2CuCl4.” Journal of Physics: Condensed Matter, vol. 33, no. 49, 2021, pp. 1-7. Research Gate, https://arxiv.org/pdf/2211.09841. Accessed 30 November 2025.

Ghosh, Suman, and Sumit Nandi. “A Comprehensive Review on UV-Visible Spectroscopy and Its application.” International Journal of All Research Education and Scientific Methods, vol. 12, no. 1, 2024, pp. 1501-1507. Research Gate, https://www.researchgate.net/profile/Suman-Ghosh-25/publication/383340352_A_Comprehensive_Review_on_UV-Visible_Spectroscopy_and_Its_Application/links/66c8bbdac2eaa500231304cf/A-Comprehensive-Review-on-UV-Visible-Spectroscopy-and-Its-Application.pdf. Accessed 8 September 2025.

Golchoubian, Hamid. “Solvatochromism Study on the Copper(II) Complexes Containing 3–Substituted Derivatives of Acetylacetonate Ligands.” Iranian Journal of Energy and Environment, vol. 3, no. 1, 2012, pp. 97-103. ResearchGate, https://www.researchgate.net/publication/272850567_Solvatochromism_Study_on_the_CopperII_Complexes_Containing_3-Substituted_Derivatives_of_Acetylacetonate_Ligands. Accessed 8 October 2025.

Kumar, Birendra, et al. “Electronic Structure and Spectral Study of Some Five Coordinate Complexes of Copper (II).” Oriental Journal of Chemistry, vol. 35, no. 4, 2019, pp. 1-7. Orientjchem, https://www.orientjchem.org/vol35no4/electronic-structure-and-spectral-study-of-some-five-coordinate-complexes-of-copper-ii/. Accessed 13 November 2025.

Marcinkowski, Dawid, et al. “Understanding the effect of structural changes on slow magnetic relaxation in mononuclear octahedral copper(II) complexes.” Dalton Transactions, vol. 51, no. 32, 2022, pp. 11939-12390. Royal Society of Chemistry, https://pubs.rsc.org/en/content/articlepdf/2022/dt/d2dt01564a. Accessed 17 November 2025.

Mielniczek-Brzóska, E., and K. Sangwal. “Study of the nature of Cu(II) complexes in aqueous ammonium oxalate solutions by ultraviolet-visible spectroscopy.” Journal of Physics: Conference Series, vol. 289, 2011, pp. 1-14. IOP Science, https://iopscience.iop.org/article/10.1088/1742-6596/289/1/012033/pdf. Accessed 22 September 2025.

Persson, Ingmar, et al. “EXAFS Study on the Coordination Chemistry of the Solvated Copper(II) Ion in a Series of Oxygen Donor Solvents.” Inorganic Chemistry, vol. 59, no. 14, 2020, pp. 9538-9550. ACS Publications, https://pubs.acs.org/doi/full/10.1021/acs.inorgchem.0c00403?.com. Accessed 18 September 2025.

Prete, Prisco, et al. “Development and validation of an eco-compatible UV–Vis spectrophotometric method for the determination of Cu2+ in aqueous matrices.” Analytical and Bioanalytical Chemistry, vol. 415, 2023, pp. 5003-5010. Spring Nature Link, https://link.springer.com/article/10.1007/s00216-023-04785-6. Accessed 29 September 2025.

Saputri, Wahyu Dita, et al. “The Jahn-Teller effect in mixed aqueous solution: the solvation of Cu2+ in 18.6% aqueous ammonia obtained from ab initio quantum mechanical charge

field molecular dynamics.” Pure and Applied Chemistry, vol. 91, no. 10, 2019, pp. 1553–1565. Research Gate, https://www.researchgate.net/publication/334113899. Accessed 29 September 2025.

Sciortino, Giuseppe, et al. “DFT Protocol for EPR Prediction of Paramagnetic Cu(II) Complexes and Application to Protein Binding Sites.” Magnetochemistry, vol. 4, no. 4, 2018, p. 55. MDPI, https://www.mdpi.com/2312-7481/4/4/55. Accessed 8 October 2025.

Vaughn, Brett A., et al. “Is Less More? Influence of the Coordination Geometry of Copper(II) Picolinate Chelate Complexes on Metabolic Stabilit.” ACS Inorganic Chemistry, vol. 59, no. 22, 2020, pp. 16095–16108. ACS Publications, https://pmc.ncbi.nlm.nih.gov/articles/PMC9199361/pdf/nihms-1812756.pdf. Accessed 10 September 2025.

Volkov, Victor V., et al. “Cu(Proline)2 Complex: A Model of Bio-Copper Structural Ambivalence.” Molecules, vol. 27, no. 5846, 2022, pp. 1-17. MDPI, https://flore.unifi.it/retrieve/a2e06ec4-db5b-4eb4-97dd-6d21f85d1a96/molecules-y22_v27_p5846.pdf. Accessed 20 October 2025.

Vušak, Darko, et al. “Ternary Copper(II) Coordination Compounds with Nonpolar Amino Acids and 2,2′ -Bipyridine: Monomers vs. Polymers.” Crystals, vol. 14, no. 7, 2024, p. 656. MDPI, https://research-portal.uu.nl/ws/portalfiles/portal/234999518/crystals-14-00656.pdf. Accessed 20 October 2025.

Yarranton, Jonathan T., and James K. McCusker. “Ligand-Field Spectroscopy of Co(III) Complexes and the Development of a Spectrochemical Series for Low-Spin d6 ChargeTransfer Chromophores.” Journal of the American Chemical Society, vol. 144, no. 27, 2022, pp. 12488-12500. American Chemical Society Publications, https://par.nsf.gov/servlets/purl/10346965. Accessed 29 August 2025.

Luminous Lies - The Radium Girls: Deception, Manipulation, and Murder
By Manuela Mejia-Fanjul

` The 1920s are historically viewed as a decade of modernization, yet underneath this facade lay severe labor exploitation, which, for a small subsection of women, manifested as hazardous industrial employment. After Marie Curie discovered radium in 1898, the element's commercial production expanded across New Jersey, Illinois, and Connecticut. Companies such as the United States Radium Corporation (USRC), managed by company president Arthur Roeder, industrialized the radium market by manufacturing luminous, radium-infused paint for military watch dials during World War I. Those painting the dials were encouraged, and at times required, to "lip point," a method created by the USRC where painters licked the tip of their brushes to create a thinner line. Years later, painters' incidental ingestion of radium resulted in severe physiological deterioration, prompting independent health investigations. However, legal and corporate systems delayed state intervention, allowing the USRC to deny liability while minimizing financial restitution for the sick workers.

This paper will explore the question "How did the unique socio-political and industrial landscape of the 1920s enable the U.S. Radium Corporation to evade accountability for industrial poisoning?" By analyzing this crisis through the lens of historical narrative and preservation and examining the importance of the means by which one tells stories, this investigation demonstrates that the USRC's evasion was achieved by exploiting a post-war patriotic shield, operating within a regulatory laissez-faire vacuum, and weaponizing gendered social stigmas to invalidate the testimony of its female workforce.

Legal & Political Systems

The patriotism of the World War I era, combined with a *laissez-faire* economic climate, catalyzed a culture where the U.S. Radium Corporation valued its industrial output more than the lives of its labor force. During the 1910s, American identity was deeply intertwined with industrial triumph and World War I struggles. The effects of World War I in American society were, in a way, displayed through the development of industries. Considering that "the alpha particles from the radium decay excited the zinc sulfide to provide a continuous light source," industrial manufacturers recognized a commercial opportunity to supply luminous military instrumentation, establishing a "lucrative industry that provided luminous dials and signage for ships and cockpits for the U.S. Navy and Army in World War I" (Coursey). This military connection placed a moral burden on the workers, leading them to the conclusion that workplace compliance and meeting high production quotas were direct acts of civic duty. In the eyes of the public and the company, radium facilities were strategic military sites, and the dial painters were vital contributors to the war effort.

Corroborating the internal pressure that the painters experienced, Katherine Schaub, a dial painter at the USRC location in New Jersey, was one of the corporation's fastest painters and ultimately one of the plaintiffs in a 1927 lawsuit against the company. In 1928, 11 years after

beginning at the USRC and eight years after her symptoms first manifested, Schaub recalled the intense social pressure of her role as a domestic industrial laborer, noting that "it was wartime. I don't believe this fact has been told yet – we girls, too, did our bit for our country" (Arne). Her reflection demonstrates how the wartime mobilization of radium workers established an internal framework of civic support, which the USRC leveraged to protect its manufacturing practices from early domestic criticism. By framing dial painting as a crucial aspect of the country's defense in a global conflict, the company and society created a social environment where questioning the safety of the factory was equated with undermining the war effort. Their identity as patriotic heroines acted as a psychological barrier, preventing the workers from viewing themselves as victims of industrial negligence until the post-war period. Consequently, this self-designation as domestic soldiers delayed the development of an adversarial labor consciousness among the workers, allowing the corporate narrative to remain unchallenged during the immediate post-war transition.

The political climate of the 1920s provided a structural foundation for corporate evasion. Following the Progressive Era's push for regulation, the 1920s saw a sharp pivot back toward *laissez-faire* governance, which minimized state intervention in private enterprise (Jones). This increased reliance on *laissez-faire* economics meant that the decade was characterized by "the failures of the philosophy of Progressive Era corporatism or corporate liberalism, where the state's role was to remain a 'neutral partner'" (Baron). For dial painters, this institutional neutrality removed any regulatory mechanism capable of protecting worker health.

In practice, the *laissez-faire* framework translated into a regulatory void as state authorities deferred to private industry interests, making government regulators and scientific experts "beholden to the corporations" (Baron). When the dial painters turned to the state for help, they found investigators who were either unresponsive or actively hostile to their pleas (Baron). This absence of a federal regulatory safety net allowed the USRC to act as its own judicial system regarding safety standards, ensuring that industrial progress consistently superseded occupational safety.

When the issue eventually reached a court of law in 1927, the USRC's core mechanism of legal evasion relied on their strategic manipulation of the New Jersey court system. The law, at the time, was not designed to account for progressive occupational diseases, including radium poisoning. The USRC's lead attorney, Edward Markley, utilized this technicality to its fullest extent,"[saving] his sharpest questioning for [expert] Hughes on her … expired air tests"(Coursey). The air test results, which showed radium poisoning, had expired due to the USRC stalling the court proceedings, a legal tactic in itself. As such, Markley's main argument and strategic decision was not to disprove the plaintiffs' physiological symptoms or even the company's liability, but to argue that their legal right to restitution had "expired" based on statutory timelines.

Additionally, the USRC's defense strategy intentionally aligned legal procedures with the disease's biological timeline. During the legal proceedings in early 1927, the defense successfully petitioned for an extended adjournment because key witnesses were traveling in

Europe, a delay that directly compounded the plaintiffs' legal vulnerability as their conditions deteriorated. Because “many of the watch painters had just months to live,” when the judge agreed to recess the girls, they were “forced to accept an out-of-court settlement” (Vaughan). Although the girls had fought for restitution for months, their advanced physical decline compelled them to settle, with multiple plaintiffs dying shortly after the resolution (“Radium Victims” 1). By stalling the proceedings, the USRC used litigation strategies to exhaust the plaintiffs’ remaining lifespans, knowing that post-mortem cases would effectively dissolve their immediate financial and legal threat. This manipulation of the legal clock proves that the USRC did not just evade accountability; they exploited biological mortality as a defensive strategy to prevent a definitive legal judgment, allowing the corporation to avoid ever formally admitting liability.

Gender and Societal Expectations: The Weaponization of the “New Woman”

While the legal system provided a structural shield for the U.S. Radium Corporation to evade responsibility, the social climate of the 1920s provided sociocultural leverage. The USRC systematically weaponized 1920s gender norms and strict moral codes to discredit the victims. By framing the workers’ health through the lens of beauty, economic privilege, and eventually, sexual impurity, the corporation ensured that the dial painters remained a vulnerable and easily silenced workforce.

In the early 1920s, despite the USRC’s knowledge of the dangers, radium was not publicly viewed as a poison, but as a miracle of modern science and a literal “shimmering” symbol of status. The corporation leaned heavily into the “Radium Age” aesthetic, where the element was marketed in everything from face creams to tonics. For the dial painters, this translated into a unique form of workplace charm, as “the job was [considered] a fairly glamorous one…[it] was both romantic and exciting” (Wink). The girls’ roles as dial painters gave them access to a material of immense commercial value, as radium was sold for a market price of approximately $120,000 per gram during the early 1920s (“Gram of Radium”). So, the girls had access to the era’s most coveted beauty product every day, leading to a societal obsession with the career.

As the girls ingested radium and showed a luminescence, the public associated this visual effect with economic privilege because the element was so costly. Rather than being interpreted as a symptom of toxicity, luminescence was culturally seen as a marker of prosperity. The USRC exploited this romantic narrative to cultivate a false sense of security among the employees, leveraging the cultural logic that an element associated with cosmetic enhancement and high value could not be destructive (USRC Advertisement). By aligning the painters’ work with the 1920s ideal of radiance and excitement, the company effectively masked their biological decay occurring beneath the surface of their glowing skin.

As public health issues emerged and the physical deterioration of the workforce became statistically undeniable, the USRC’s defensive strategy shifted from materialistic promotion to reputational defamation. To explain the crumbling jaws of the victims, company-aligned doctors

employed the era's obsession with female abstinence. Corporate-backed medical examiners routinely assigned diagnoses of syphilis to deceased workers such as Mollie Maggia, attributing the bone necrosis to a sexually transmitted disease rather than occupational radiation exposure(Vaughan). By labeling industrial injuries as sexually transmitted diseases, the USRC successfully turned the victims' own bodies into evidence of their alleged moral "delinquency", effectively erasing public sympathy and redirecting potential regulatory outrage into social ostracization.

Work and Medical Deception

The deception practiced by the U.S. Radium Corporation was not a reactive maneuver to an unforeseen crisis, but a foundational element of its operational management. This institutional dishonesty began with the suppression of its own scientific founders, advanced to the falsification of regulatory data, and culminated in the public gaslighting of the workforce. By controlling the flow of information, the USRC ensured that truth remained a proprietary corporate asset rather than a matter of public health.

The earliest evidence of the USRC's internal knowledge regarding radium's lethality can be found in the spring of 1921, establishing a baseline of corporate awareness years before widespread symptoms emerged. Dr. Sabin von Sochocky, the creator of radium paint and co-founder of the USRC, realized the biological toll of exposure and attempted to warn USRC President Roeder of the dangers. Instead of halting production, the company marginalized and eventually terminated his contract in a letter. At the end of this letter exchange, Dr. Sochocky writes to Roeder: "It may interest you that I am now a physical and nervous wreck - with heart and kidneys affected…But I am not dead yet" (Sochocky). He then went on to explain protein values in his blood, all attributed to radium poisoning. His eventual death from aplastic anemia, a common side effect of radium, in 1928, serves as confirmation that the executives at the USRC understood the lethal trajectory of ingesting the product long before the trials.

As the death toll by radium began to rise by 1924, the USRC sought external validation to quiet growing concerns, but quickly pivoted to suppression when that validation turned into indictments. The USRC commissioned Dr. Cecil Drinker and Katherine Drinker of the Harvard School of Public Health to conduct an independent sanitary survey of the New Jersey facility. The investigators' findings were unequivocal, stating in a letter to Roeder that they"believe[d] that the trouble which has occurred [was] due to radium" (Drinker). Rather than integrating these scientific findings into plant protocols, USRC management committed regulatory fraud by editing the data before submitting the document to the New Jersey Department of Labor to present the corporate environment as safe ("Deadly Occupation). This deliberate omission of data obstructed state-level intervention and ensured that the hazardous lip-pointing instructions remained standard operating procedure until 1926.

The final pillar of this monopoly over truth was the management of medical authority to directly mislead the workforce during the mid-to-late 1920s. This was personified by Dr. Frederick Flinn, an associate professor of physiology at Columbia University, who acted as a

medical consultant for the USRC. Flinn was accepted as an undisputed medical professional to the dying dial painters, and under the credentials presented by USRC management, yet he "was not a physician" ("Frederick Flinn" 33). Despite his lack of medical credentials, Flinn posed as a doctor to issue fraudulent clean health reports to the victims. This medical gaslighting was designed to neutralize any labor organizations or litigators. Plaintiffs deferred to academic authority and accepted his conclusions over their own knowledge of their symptoms. By positioning a corporate consultant lacking medical licensure as the primary authority on worker health, the USRC successfully intercepted early diagnoses, causing workers to attribute their physical decline to factors other than industrial contamination.

Conclusion

In addressing how the unique socio-political and industrial landscape of the 1920s enabled the U.S Radium Corporation to evade accountability, the narrative of the Radium Girls serves as a definitive case study in historiographical preservation, illustrating how institutional power shapes the historical record. Traditional historical records may describe the tragedy as an unavoidable consequence of a pre-regulatory era where the biological dangers of radium were simply unknown. However, archival evidence reveals that the official record of the dial painters was not constructed by scientific ignorance, but rather by corporate deceit. For decades, the official record of the dial painters was a corporate construction, in which corporate entities controlled public documentation by attributing occupational illnesses to personal moral failures or unverified medical anomalies. The litigation and public advocacy initiated by the surviving dial painters disrupted this perceived truth, forcing a reconstruction of early 20th-century labor history. By documenting the physiological realities of their exposure, these workers shifted the historical narrative to honor the lives lost in industrial incidents.

The relevance of their struggle remains acute today; the Radium Girls left a legacy that continues to influence modern standards of workplace safety and informed consent. Their story is a primary reason why contemporary occupational safety frameworks exist. Ultimately, the preservation of this narrative demonstrates that historical memory is an active battleground; by challenging the USRC's engineered records, the dial painters ensured that the history of the American industrial era would be written in collaboration with the invisible stories that lived it.

Works Cited

Anders, Roger M. Radioactivity and Health: A History. Feb. 1997. RMI Nuclear Docs, rminucleardocs.icaad.ngo/api/files/16106774552688umvyptqyge.pdf. Accessed 5 May 2026.

Arne, Gladys. "Simple Things of Life Bring Most Cheer to Women Who Know They Soon Will Die." Reading Times[Philadelphia], 2 June 1928. Newspapers.com, www.newspapers.com/article/reading-times-radium-girls/2896556/?locale=fr-CA. Accessed 9 May 2026.

Baron, Sherry. "Claudia Clark's Radium Girls." Review of Radium Girls, by Claudia Clark. Against the Current, Mar. 1999, againstthecurrent.org/atc079/p886/. Accessed 20 May 2026.

Clark, Claudia. Radium Girls: Women and Industrial Health Reform. University of North Carolina Press, 1997.

Coursey, Bert M. "The National Bureau of Standards and the Radium Dial Painters." Journal of Research of the National Institute of Standards and Technology, vol. 126, Feb. 2021, article 126051, nvlpubs.nist.gov/nistpubs/jres/126/jres.126.051.pdf. Accessed 5 May 2026.

"Deadly Occupation, Forged Report." Harvard T.H. Chan School of Public Health, 24 Oct. 2013, hsph.harvard.edu/news/centennial-radium-forged-report/. Accessed 4 May 2026.

Drinker, Cecil K. Letter to Arthur Roeder. 3 June 1924. National Archives Catalog, National Archives ID 75718961, catalog.archives.gov/id/75718961.

"Frederick Flinn, 80; Former Columbia Professor Dies—Toxicologist, Expert on Radium Poisoning." The New York Times, 16 Apr. 1957, p. 33. The New York Times Archive, www.nytimes.com/1957/04/16/archives/frederick-flinn-physiologist-80-former-columbia-professor.html. Accessed 10 May 2026.

"Gram of Radium for New York State." Advertisement. Fulton County Tribune [Wauseon, OH], 13 Aug. 1920, p. 7. Library of Congress, www.loc.gov/resource/sn87076552/1920-08-13/ed-1/?sp=7&r=0.147,0.562,0.892,0.488,0. Accessed 17 May 2026.

Jones, Kelly. "Post Civil War Political Culture." Lecture, Convent of the Sacred Heart, 20 Jan. 2026, New York City, NY.

Kilgallen, James L. "Two of Radium Women Offer Selves for Tests While Alive." The Bee [Danville, VA], 29 May 1928. Newspapers.com, www.newspapers.com/article/the-bee-women-radium-victims-offer-selve/5559067/. Accessed 27 Apr. 2026.

Moore, Kate. The Radium Girls: The Dark Story of America's Shining Women. Sourcebooks Nonfiction, 2017. E-book ed.

Pfahler, George E. The Effect of the X-Rays and Radium on the Blood and General Health of Radiologists. American Journal of Roentgenology, National Archives ID 75718674, catalog.archives.gov/id/75718674.

Radium Luminous Material Corporation. Comparison of Old and New Luminous Material. Orange, NJ, 1921. National Archives Catalog, National Archives ID 75718639, catalog.archives.gov/id/75718639.

"Radium Victims Win $50,000 and Pensions in Suit Settlement." The New York Times, 5 Jun. 1928, pp. 1, 8. The New York Times Archive, www.nytimes.com/1928/06/05/archives/radium-victims-win-50000-and-pensions-in-suit-settlement-each-of.html. Accessed 9 May 2026.

Rentetzi, Maria. Review of Deadly Glow: The Radium Dial Worker Tragedy, by Ross Mullner. Science, Technology, & Human Values, vol. 26, no. 1, 2001, pp. 106–08. JSTOR, www.jstor.org/stable/690123.

Rowland, Robert E. Radium in Humans. Argonne National Laboratory, Nov. 1994, publications.anl.gov/anlpubs/1994/11/16311.pdf. Accessed 5 May 2026.

Sochocky, Sabin Arnold. Letter to Arthur Roeder. 3 June 1921. National Archives Catalog, National Archives ID 75718577, catalog.archives.gov/id/75718577.

USRC Advertisement of Job Listing. 1918. Image. USC Libraries Research Guides, libguides.usc.edu/c.php?g=1521307&p=11401078. Accessed 17 May 2026.

Vaughan, Don. "Radium Girls: The Women Who Fought for Their Lives in a Killer Workplace." Encyclopedia Britannica, 13 June 2025, www.britannica.com/story/radium-girls-the-women-who-fought-for-their-lives-in-a-killer-workplace. Accessed 20 May 2026.

Wink, Sophie. "'The Capitalist System Is a System of Murder for Profit': Radium Girls in the Public Eye." Historical Perspectives: Santa Clara University Undergraduate Journal of History, 2nd ser., vol. 26, 2021, p. 7. Scholar Commons, scholarcommons.scu.edu/cgi/viewcontent.cgi?article=1266&context=historical-perspectives. Accessed 4 May 2026.

The Effect of The Destruction of The Library of Alexandria By Octavian Steinert

Abstract

The Library of Alexandria once housed much of the most advanced knowledge of humankind, preserving large collections of poems, theorems, and medical treatises produced by civilization. It was established by Ptolemy I Soter in the early third century B.C, and further expanded by Ptolemy II Philadelphus, after which the library became the intellectual center of the ancient world of Egypt and the Mediterranean. Scholars from across the ancient world met there to study, innovate, and collaborate. The library's gradual decline, resulting from a series of fires and conquests between 48 B.C. and 642 A.D., led to the loss of irreplaceable works by Homer, Sophocles, and Euripides, as well as sophisticated scientific artifacts. If the Library of Alexandria had survived, its large collection of knowledge could have helped bypass much of the intellectual stagnation of the Dark Ages, strengthened cultural exchange through international connections, and potentially accelerated human progress by centuries.

Introduction

Imagine a single building that held the collected wisdom of the ancient world. Every poem, theorem, and medical treatise that civilization had produced was kept there. Then imagine the building being destroyed. The Library of Alexandria once housed the collective wisdom of the ancient world, preserving every poem, theorem, and medical treatise produced by civilization. Its destruction represents one of humanity's most significant losses, not only because of the materials destroyed but also because of the unrealized advancements that could have been built upon its foundation. Established in the early third century B.C. by Ptolemy I Soter and later expanded by Ptolemy II Philadelphus, the library became the intellectual center of the ancient world (El-Abbadi). Scholars from across the ancient world met there to study, innovate, and collaborate. The library's gradual decline, resulting from a series of fires and conquests between 48 B.C. and 642 A.D., led to the loss of irreplaceable works by Homer, Sophocles, and Euripides, as well as sophisticated scientific artifacts (El-Abbadi).

Dark Ages

Had the library survived, its vast repository of knowledge might have prevented the depression of the Dark Ages, promoted growth through international exchange, and accelerated human progress by centuries (El-Abbadi). The destruction of the Library of Alexandria was a decisive turning point in human society (El-Abbadi). The Library of Alexandria's halls of books, records, and knowledge may have prevented the Dark Ages (El-Abbadi). That period lost humanity many years of study, work, and breakthroughs (The Return of Mathematics to Europe). According to Britannica, "during this period [The Dark Ages] of more than 500 years, Europe was in some ways bleak. It was noted for the decline of metropolitan life, frequent warfare, plague, poverty, famine, and superstition" (Britannica Editors). This quote reveals how bleak the Dark Ages were, defined by widespread suffering and the decline of human knowledge. This loss

deprived humanity of many years of research and development. Carl Sagan, renowned historian and astrologist, likened the Dark Ages to a period of “long sleep…it picked up some of the tools, the books, and the concepts laid down here, more than a thousand years before" (Sagan, "The Shores" 00:35:10). Some of these tools, books, and concepts were in the Library. This proves the value of the Library, as it had the materials needed to advance. These materials were used after the Dark Ages. If the Library had been available to Europeans, they might have bypassed the stagnation of the Dark Ages. The Library also brought together scholars from different countries and perspectives. This improved international relations and fostered global progress (El-Abbadi).

International Relations

If the Library of Alexandria had survived, international relations would have improved considerably. The Library’s growing popularity and reputation drew scholars from across the ancient world, fostering broader relations as people came together to share knowledge with other nations and societies within a single hub of learning. In one of the articles on The Archaeologist, the text stated that “scholars from across the Greek, Egyptian, Persian, and Indian worlds were invited to study in Alexandria" (“The Library of Alexandria: A Center of Ancient Knowledge”). Many intellectuals went to Alexandria to study, thereby promoting international relations and serving as a true center of intellectual exchange. As scholars gathered in the Library of Alexandria from different nationalities, they shared information with one another, bringing diverse knowledge to many parts of the world and strengthening international cooperation toward the goal of retrieving as much knowledge as possible. Another section of the text from the same article stated that, “this institution… provid[ed] scholars with housing, food, salaries, and access to vast collections of texts,” supporting the Library’s ultimate goal of producing “new discoveries in science, mathematics, philosophy, medicine, and literature” (“The Library of Alexandria: A Center of Ancient Knowledge”). The Library supported all scholarly work being conducted at the institution in Alexandria and provided the necessary human resources. With the need to find a method to remove human necessities, the scholars were able to share and communicate many ideas with each other and work together, without distraction, on learning and discovering new information, which they then shared throughout the library. The Library would have advanced knowledge overall, and humans would be further ahead in knowledge development.

Human Progress

Provided that the Library were still standing today, humanity would have made further technological advances and could have become more sophisticated than modern humans, were all other things equal. The quantity of books was exceptional. Different scholars report different totals: The first account reported over 200,000 books; the second, over 450,000; and the third, over 700,000. It is likely that some of these numbers are exaggerations, but even the earliest account, written in the third century BCE, stated “over 200,000” books, which amounts to a tremendous number (El-Abbadi). Given these immense figures, scholars today can only assume

that the amount of knowledge within the Library's walls would have been colossal, thereby giving humanity a much greater understanding of the world around them and furthering human research and development. According to the FAS or Federation of American Scientists, “the Library reflected and helped to generate a transforming wave of inquiry and enlightenment throughout the ancient world” (Aftergood). This implies that the Library fostered the growth of knowledge among all human beings in the ancient world. It can be concluded that if the Library were still in existence, its enduring benefit of furthering human knowledge would have affected the average human's current grasp of information and could have put humanity hundreds of years ahead in the collective understanding of science, technology, and philosophy.

Some scholars may argue that the Library’s existence today would not have significantly altered the course of human society. Most scholars attribute the Dark Ages to the fall of the Western Roman Empire. Historians who argue against the Library’s potential impact on history most likely believe that its survival would not have significantly affected the Dark Ages. According to History.com, the collapse of the Western Roman Empire and, along with it, the Dark Ages were caused by many factors, not just one (Andrews). According to these historians, the Dark Ages would not have been affected by the Library’s longer existence. But if the Library of Alexandria had remained strong and influential, its preserved economic knowledge could have contributed to a more effective taxation system, addressing one of the major factors in the Fall of the Western Roman Empire. Its engineering expertise might have strengthened infrastructure and military defenses, helping to resist invasions by northern tribes, another significant cause of the Empire’s collapse. In addition, its political philosophy could have improved governance and reduced internal political conflict, both of which contributed to the Empire’s downfall. In conclusion, the Library wasn't just the physical Library; it was a collection of ideas and human thought that, with all of the manpower and the extra time that they would have had with the library, the scholars may have learned about newer concepts that could have prevented the fall for the aforementioned reasons.

Conclusion

Had the Library of Alexandria survived, its immense collection of knowledge could have helped bypass much of the intellectual stagnation of the Dark Ages, strengthened cultural exchange through international connections, and potentially accelerated human progress by centuries. The Library of Alexandria was not just a library; it was a wonder of the ancient world, with the largest collection of books the world had ever known, and a place where great minds were challenged to develop new solutions and ideas.

Works Cited

Aftergood, Steven. “The Library of Alexandria and the Library of Congress.” Federation of American Scientists. 2024, https://fas.org/publication/loc-alexandria/.

Andrews, Evan. “8 Reasons Why Rome Fell: The Fall of Rome Explained.” HISTORY, 2025, https://www.history.com/articles/8-reasons-why-rome-fell.

Britannica Editors. “How Long Were the Dark Ages? | Europe, Dates, History, and Facts.” Encyclopedia Britannica, 15 Apr. 2026, www.britannica.com/topic/How-Long-Were-the-Dark-Ages.

El-Abbadi, Mostafa. “Library of Alexandria.” Encyclopædia Britannica, 2026, https://www.britannica.com/topic/Library-of-Alexandria. ---.

El-Abbadi, Mostafa. “Library of Alexandria: Description, Facts, and Destruction.” Encyclopædia Britannica, 20 July 1998, www.britannica.com/topic/Library-of-Alexandria/Other-languages.

“The Library of Alexandria: A Center of Ancient Knowledge.” The Archaeologist, 28 Jan. 2026, www.thearchaeologist.org/blog/the-library-of-alexandria-a-center-of-ancient-knowledge.

“Library of Alexandria | Description, Facts, & Destruction.” Encyclopædia Britannica 2026, https://www.britannica.com/topic/Library-of-Alexandria.

Moore, Jason W. “The Crisis of Feudalism: An Environmental History.” Theory and Society, vol. 31, no. 3, 2002, pp. 307–357, https://doi.org/10.1177/1086026602153008.

“The Return of Mathematics to Europe.” Encyclopedia.com, 2026, https://www.encyclopedia.com/science/encyclopedias-almanacs-transcripts-and-maps/return-mathematics-europe

Sagan, Carl. “The Shores of the Cosmic Ocean.” Cosmos: A Personal Voyage, episode 1, KCET, 28 Sept. 1980.

A History of Environmental Awareness in Serbia: From Industrialization to European Integration By Emilija Mohora

Abstract

This paper traces the historical development of environmental awareness in Serbia, from its nineteenth century agrarian foundations through socialist industrialization after the Second World War, the political and economic collapse of the 1990s, and the post-2000 period of democratic reform and European Union integration. It argues that environmental awareness in Serbia evolved slowly and unevenly, shaped by changes in economic and political systems , industrial legacy, and lived experience of pollution. The paper shows that political changes after 2000, together with EU-driven legal and institutional reforms created stronger frameworks for environmental awareness and public participation. Following nearly 50 years after the enactment of the first Yugoslavia Federal provisions for environmental protection in 1974, recent mass mobilizations in Serbia—exemplified by the anti-lithium Jadar movement, Eco-Uprising, clean air protests, and campaigns to defend hydro-resources—demonstrate a historically high level of public environmental awareness, signaling that ecological issues have decisively moved to the center of the national discourse.

Introduction

Environmental awareness is defined as factual knowledge about environmental change that influences how individuals perceive and respond to environmental issues. It encompasses the understanding and meanings people derive from their experiences and interactions with the environment, which can affect their behavior towards environmental sustainability (Eden) . The United Nations (UN) defines environmental awareness as the "gradual understanding of environmental issues", and the recognition of the connections among human actions, development, sustainability and human responsibility in these processes (UNESCWA). The UN explicitly embeds environmental awareness into Sustainability Development Goal (SDG) Target 12.8, which aims to ensure that by 2030, people everywhere have the relevant information and awareness for sustainable development and lifestyles in harmony with nature. It is viewed as a vital prerequisite for managing the "triple planetary crisis" climate change, pollution, and biodiversity loss (UN). Thus, in the twenty-first century, "environmental awareness" has become more than just a popular phrase; it's a crucial part of how we talk and what we do about building a sustainable future. It's about how much we—as individuals, communities, and governments—truly grasp the environmental challenges we face and, more importantly, what we do about them. Serbia and the wider Balkan region offer a particularly complex and revealing example of this story. In Serbia, environmental awareness has developed against a backdrop of political disruption, the industrial legacy of the socialist era, the instability of the 1990s marked by war and economic sanctions, and the more recent drive toward EU integration. Serbia's experience, a complex and layered socio-ecological history, is a powerful case study for several reasons.

First, the country is still dealing with the most pronounced industrial pollution legacies in Southeast Europe, with contamination hotspots including the Bor copper smelter complex, the Pančevo refinery petrochemical and fertilizer complex, the Kostolac and Obrenovac coal-fired power plant clusters, and extensive mining and metallurgical sites across the country. These aren't just spots on a map; they have profoundly shaped the health, memories, and daily lives of the people who live there. For them, pollution isn't an abstract concept, it's a public health crisis (Matic et. al.).

Second, Serbia has long been shaped by different empires, economic systems, and political ideas. Ottoman rule over most of Serbia lasted from 1459 until the early nineteenth century, while the Habsburg Empire governed parts of northern Serbia at different times between 1699 and 1918. These periods influenced farming, forests, and the way natural resources were used. Later, socialist rule in the twentieth century brought industrialization, collectivization, and major environmental change. After socialism, the shift to a market economy led to slow and complex privatization processes that significantly influenced management of historical legacy pollution.

Third, the country's laws and institutions have been rollercoasters for the last fifty years. From the collapse of Yugoslavia's environmental framework to the regulatory free-for-all during the sanctions, and then the slow, painstaking effort to rebuild after 2000, this timeline offers a fascinating look at how environmental thinking has shifted at every level of society.

Fourth, Serbia's ongoing European Union (EU) accession process has introduced a powerful external driver of environmental policy reform, requiring the progressive adoption and implementation of the EU environmental acquis. This process has simultaneously strengthened institutional frameworks and generated new spaces for civil society participation, public access to environmental information, and environmental education — all of which are instrumental in cultivating broader environmental awareness among the population.

More recently, Serbia has been at the center of a major international debate about lithium mining. A huge lithium deposit was discovered in the Jadar valley in western Serbia, a This sparked the biggest environmental protest movement in Serbia's recent history, involving farmers, city residents, scientists, and young people — and it raises important questions about who gets to decide and how, what happens to a country's natural resources (Stefanović et al.).

Thus, this isn't just a local story. Serbia's journey offers valuable lessons for any country trying to balance economic growth with the ghosts of its industrial past and the demands of modern environmental standards. It gives us a real-world look at how environmental awareness can take root even during political and economic turmoil, and how external influences can both help and hinder that growth.

This paper traces the historical evolution of environmental awareness in Serbia, from its nineteenth-century agrarian foundations through post-Second World War socialist industrialization and the crises of the 1990s to the contemporary era of democratic reform and European Union integration.

Early Legislative Efforts: Forestry, Water and Sanitation (1830 – 1914)

For most of its early history, and well into the nineteenth century, Serbia was predominantly agrarian. The establishment of the autonomous Principality of Serbia in 1830, and its subsequent recognition as a fully independent kingdom in 1878, created the institutional preconditions for the development of formal public policy in areas touching on the environmental protection, though these were framed almost exclusively in terms of public health and economic productivity rather than environmental protection per se. These political changes mattered because they gradually created the institutions through which land, forests, water, and public health could begin to be regulated more systematically. Everyday relations with nature, however, remained practical and shaped by subsistence farming, forestry, and fishing. Villages managed shared resources through customary rules passed down across generations. Environmental awareness in this period was therefore local, informal, and driven more by necessity than by a developed public concept of environmental protection. As population grew and private ownership expanded during the nineteenth century, these customary practices weakened, contributing to deforestation and soil erosion. Awareness of the harmful effects of deforestation began to emerge in the early 1870s, influenced by the work of several scientists (Pančić) and especially by the major floods of 1864 and 1871, although this awareness remained largely confined to the educated elite (Samardzic and Beslin).

Rapid urbanization in Belgrade and other towns brought growing concern about drinking water quality, sanitation, and waste disposal, largely in response to recurring outbreaks of cholera, typhoid, and dysentery that devastated urban populations in the late nineteenth century. The first systematic efforts to address water quality in Serbia were therefore driven by epidemiological concerns. The Serbian Forestry Law of 1891 is recognized as a pioneering law that marked the beginning of organized state management and protection of forest resources (Samardzic and Beslin). Around the same time, the first civic conservation societies and nature protection and recreation associations were established, including the Serbian Mountaineering Association (1901) and the White Eagle Scout Detachment (1911) (Petrovic and Pesic). Environmental education was initiated at the beginning of the twentieth century. The oath for those pupils attending second grade of elementary school in 1914 was: "*I swear that I will not destroy the trees and treat flowers badly*". From that time onwards, environmental education was integrated into various subjects of the curriculum and existed in different kinds of education (Hadjichambis et al.). At the same time, Serbian scientists began to systematically study the country's rich natural world. The work of Josif Pančić, Serbia's most famous botanist, provided the first detailed surveys of its plants and animals, building a foundation of knowledge and appreciation for the nation's natural heritage (Pančić).

Kingdom of Yugoslavia (1918–1941)

In the Kingdom of Yugoslavia, the national focus was squarely on economic development, leaving environmental issues on the sidelines. It was primarily driven by sanitary reform, forestry management, and the early development of scenic tourism. Rather than modern

ecological conservation, the era focused on state-building, resource extraction, and early urban spatial planning (Puljar). Because the country was predominantly agrarian, with three-quarters of the workforce employed in farming, early "environmentalism" was primarily an economic and administrative effort focused on managing timber, preventing soil erosion, and protecting rural watersheds. Still, the seeds of conservation were planted, with early moves to create national parks. Serbian scientific institutions — including the Serbian Royal Academy (later the Serbian Academy of Sciences and Arts, SANU), the University of Belgrade, and the Natural History Museum of Belgrade — had begun to develop systematic programs of natural history research that contributed to a growing body of knowledge about the ecological character of Serbian landscapes.

The Socialist Yugoslavia and the Emergence of Environmental Policy (1945–1991)

The rise of socialist Yugoslavia post Second World War period kicked off an era of rapid, massive industrialization that would forever change Serbia's landscape. Huge factories, mines, and power plants were built as symbols of progress. The most impactful of these were, already mentioned, the Bor a century-old copper mining and smelting complex, the Pančevo petrochemical complex (oil refinery, fertilizer and petrochemicals) and the coal-fired power plants at Kostolac and Obrenovac. They brought jobs and economic growth, but at a staggering environmental cost. The environmental consequences of this industrialization were severe and, for much of the socialist period, systematically unacknowledged in public discourse. Air and water pollution around major industrial centers reached levels that caused measurable damage to agricultural productivity, human health, and ecological systems, yet these impacts were largely suppressed within a political culture that equated industrial output with progress and treated environmental criticism as ideologically suspect (Serbula et al. Matic et. al.).

Even so, Socialist Yugoslavia did pass some environmental laws, especially from the 1960s onward. A major turning point was the United Nations Conference on the Human Environment, held in Stockholm, Sweden, in 1972, the first large international meeting specifically about environmental issues. Socialist Yugoslavia attended and used the momentum from the conference to pass a Federal Law on Environmental Protection in 1974 — which was quite progressive compared to other communist countries at the time in Eastern Europe, establishing administrative and social controls within the principles of socialist self-management, designed to give democratic representation to all the interests involved in environmental protection (Singleton)). Environmental legislation in socialist Yugoslavia was relatively developed and in the mid-1980s, there were as many as 400 laws and over 1,000 bylaws. However, advanced legal solutions have not been sufficiently applied in practice (Petrovic and Pesic). Serbia passed its own laws on air, water, and nature protection in the 1970s and 1980s, and first environmental inspectors were appointed. But in practice, these laws were only partially enforced. Meeting production targets for factories always came first. Environmental inspectors were given very few resources and little authority (Petric).

In early 1973, the Yugoslav Council for the Protection and Improvement of the Human Environment was established in Belgrade. On the Council's proposal, the Constitution of Socialist Yugoslavia included a provision recognizing the right to a healthy environment. Although the regime initially supported the Council, it soon became a bureaucratized body with limited influence. During the 1970s and 1980s, researchers at the University of Belgrade and the Serbian Academy of Sciences and Arts began producing serious studies about environmental damage — documenting acid rain, heavy metal contamination of soils, and damage to river ecosystems. University faculties of natural sciences and civil engineering developed research programs on air and water quality, while emerging fields such as environmental chemistry and ecotoxicology brought scientific rigor to the documentation of industrial pollution. This period laid important groundwork for the expert capacity that would later underpin policy reform.

As political cracks began to show in the 1980s, a small window opened for independent voices. Green parties could not form within the constraints of a single-party regime. However, certain early forms of environmental activism did occur in Serbia, even during the communist era, dating back to the 1970s. A small number of journalists also began writing about pollution for the first time. By the late 1980s, Serbia had environmental laws, environmental scientists, and the introduction of environmental activism. At that time, the first environmental organizations were formed throughout Yugoslavia such as Green Star and Green PEGA (1987) in Serbia (Petrovic and Pesic). But the system was stacked against real change: industries were too powerful, enforcement was too weak, and the country was heading toward a political crisis that would make everything worse (Pavlovic).

During the 1980s, environmental protests in Socialist Yugoslavia grew into organized movements driven by severe soil and river pollution, highlighted by the 1984 toxic PCB contamination of the Krupa River. While local citizens reacted aggressively to groundwater poisoning from unmanaged dumpsites, campaigns against infrastructure projects were far more deliberate. A prime example was the opposition to the Buk Bijela hydroelectric project—designed in 1984 without public input—which threatened to flood 80 kilometers of the UNESCO-protected Tara River Canyon up to 135 meters high. This sparked a rare all-Yugoslav "storm" of protests. Similarly, between 1984 and 1988, a crucial coalition of experts, youth, and peasants thwarted a series of hydroelectric plants on the Mura River that would have devastated the lowland landscape and local agriculture, while further actions on the Drava River utilized international cooperation to protect vital forestry (Ostrić).

In the late 1980s, there were also initiatives to create a Yugoslav environmental movement, which would unite associations from different republics but without much success. With the introduction of political pluralism, there was an attempt to establish green parties, so in 1990 the Green Party and the Environmental Movement of SR Serbia were founded in Serbia, However, the newly initiated Yugoslav environmental movement failed to survive the country's disintegration (Ostrić).

Environmental Awareness During Political and Economic Collapse (1991–2000)

The violent disintegration of Yugoslavia in the 1990s was a catastrophe for Serbia's environment. The federal system that had managed environmental policy vanished. The economy was crippled by hyperinflation, international sanctions, and the costs of military conflict. Old, polluting factories kept running with low maintenance and zero pollution controls. By the mid-90s, Serbia's environmental regulations had effectively collapsed. International sanctions cut Serbia off from the environmental cooperation programs that its neighbors — Poland, Hungary, and the Czech Republic — were joining as part of their preparation to enter the EU. While those countries were building modern environmental institutions, Serbia was moving backwards (Vujošević et al.).

The worst environmental disaster of this decade came during the NATO air campaign of March to June 1999. Some of the bombing targeted major industrial sites, including the Pančevo petrochemical and fertilizer complex, the Novi Sad oil refinery, and the Bor copper smelter. At Pančevo, the bombing destroyed containment structures and released large quantities of toxic chemicals — including ethylene dichloride, vinyl chloride, and ammonia — into the Danube river and surrounding soil (Levinbuk). The Novi Sad refinery fires produced toxic smoke that hung over the city for days. At Bor, already heavily polluted, the bombing added further releases of heavy metals and sulphur dioxide. After the conflict, the United Nations Environment Programme (UNEP) conducted an independent environmental assessment. Its reports confirmed serious contamination at Pančevo, Bor, Novi Sad, and Kragujevac, and called for urgent action to clean up these sites (UNEP). UNEP's assessment was significant because it introduced the language and standards of international environmental accountability into Serbian public discussion, including information disclosure principle to public and affected parties, at a time when domestic institutions were at their lowest point.

The extreme environmental pressures of the 1990s deepened environmental awareness in some communities. In Pančevo, local doctors and residents documented rising rates of miscarriages, birth defects, and breathing problems that they linked to industrial contamination. In Bor, sulfur dioxide levels regularly exceeded safe limits for human health, and the landscape around the smelter was almost completely bare of vegetation. For people living in these places, environmental destruction was not an abstract concept — it was part of daily life. This suffering led to enduring grassroots environmental awareness (Živanović).

One of the most striking features of the 1990s period in Serbia was the resilience of environmental civil society activism under conditions that might have been expected to extinguish it entirely. Despite the political complexity, the economic sanctions and hyperinflation, and the social trauma of armed conflict, a network of environmental non-governmental organizations (NGO) continued to operate throughout the decade, maintaining links with international environmental organizations and keeping environmental protection at the forefront of public communication discourse. Organizations such as the Green Network of Vojvodina, the Belgrade-based Ecological Center Stanište, and numerous local environmental associations continued to monitor air and water pollution, publish findings, and engage communities in the areas most severely affected by industrial contamination. These organizations

have operated with minimal resources and relied on international donor support to maintain their activities. The international dimensions of environmental civil society during this period were also significant. Serbian environmental NGOs maintained connections with regional networks such as the Regional Environmental Center for Central and Eastern Europe (REC), founded in 1990, which provided capacity building, networking, and funding support to environmental civil society organizations across the transition region supporting the free exchange of information and public participation in environmental decision making These connections helped to transmit international environmental norms, methodologies, and advocacy frameworks into the Serbian context even during the period of political and institutional isolation.

Post-2000 Political Transition and Environmental Reform

After political changes in October 2000, the new government understood that joining the EU requires building modern environmental laws and institutions. International donors responded quickly. The EU, the World Bank, UNDP, and USAID all provided funding and technical expertise to help Serbia rebuild. The period from 2000 to 2010 witnessed an extensive reconstruction of Serbia's environmental legal and institutional framework, driven by the dual imperatives of addressing domestic environmental deficits and aligning with European Union standards in preparation for the accession process.

The Law on Environmental Protection was adopted in 2004, and significantly revised in 2009, established a comprehensive framework governing environmental impact assessment (EIA), strategic environmental assessment (SEA), integrated pollution prevention and control, environmental permitting, and public participation in environmental decision-making. This was a fundamental change. For the first time in Serbian history, ordinary citizens had a legal right to attend public hearings about projects in their communities and voice their concerns and opinions. The Strategic Environmental Assessment (SEA) framework, introduced through separate legislation aligned with the EU SEA Directive, extended analogous requirements to plans, programs, and policies — including spatial plans, sectoral development strategies, and energy plans — requiring systematic environmental assessment and public consultation at the level of strategic decision-making. The implementation of SEA had represented an important normative shift toward the integration of environmental considerations into the upstream stages of planning and policy development. In 2004, Serbia established the Serbian Environmental Protection Agency (SEPA). This was the country's first professional body dedicated entirely to environmental monitoring and information management. SEPA's job is to collect data on air quality, water quality, biodiversity, and waste, and to publish regular reports so that citizens and government can see the state of the environment. It is modelled on similar agencies in EU countries. This procedural innovation had significance that extended well beyond its immediate regulatory function. By creating formal, legally mandated channels through which citizens and civil society organizations could engage with environmental decisions affecting their communities, the EIA system institutionalized public environmental participation in Serbian governance for the first time. It provided civil society organizations with legal standing and

information rights that they had previously lacked, and it created recurring opportunities for the public engagement with environmental issues that are essential to building and sustaining broader environmental awareness (Antić et al.).

The EU Path and the Rise of Citizen Power (2005–Present)

Serbia officially became an EU candidate in 2012, embarking on a massive project of legal and institutional reform. Environment chapter (Chapter 27) is one of the toughest, covering over 200 laws on everything from air quality to waste management (European Commission). Serbia has adopted the EU Water Framework Directive and created management plans for its major river basins. Drinking water standards have been aligned with EU requirements. However, a large proportion of Serbia's wastewater — the wastewater from homes, factories, and farms — is still discharged into rivers without proper treatment. Building wastewater treatment plants is happening slowly and is behind schedule (Simeunović et al.). Serbia now has a national air quality monitoring network that produces real-time data on the main air pollutants, including fine particles (PM2.5), sulphur dioxide, nitrogen dioxide, and ozone. This data is publicly available and has played an important role in raising public awareness about air pollution. Serbia has also started designating protected natural areas equivalent to the EU's Natura 2000 network. The country contains some remarkable ecosystems — Danube floodplain forests, mountain forests on Tara and Kopaonik, and the wetlands and grasslands of Vojvodina — that are of major European importance for biodiversity (Amidžić et al.).

Growing Public Environmental Awareness and the Role of Civil Society

Since 2000, environmental NGOs in Serbia have experienced significant growth. Some groups, such as the Ecological Movement of Novi Sad, have operated since the 1980s and managed to endure the challenging times during the 1990s. Others are recent additions. Currently, organizations like the Center for Ecology and Sustainable Development (CEKOR), the Society for the Protection of Birds of Serbia, and the Climate Action Network Serbia (CAN Serbia) are instrumental in environmental monitoring, engaging with legal proceedings, pursuing court cases when laws are violated, and informing the public (CAN Serbia). These NGOs are particularly involved in Environmental Impact Assessment processes—reviewing complex technical documents, participating in public hearings, and voicing concerns that might otherwise be overlooked. At times, they have contested insufficient environmental decisions in court with success. Their independent monitoring activities include measuring air and water quality, tracking illegal dumping, and observing wildlife—often generating data that surpasses what is available through official government sources.

Morka Gora National Park (2004 – 2012)

The multi-decadal civic resistance against nickel exploration in the Mokra Gora-Tara region represents a foundational case of environmental justice and resource defense in Western Serbia, driven by localized opposition to the severe ecological risks of open-pit mining and

high-pressure acid leaching. Triggered by systemic threats to critical hydrological basins, protected eco-reserves, and a sustainable regional tourism economy, the movement progressed through distinct temporal waves—beginning with localized agrarian blockades against European Nickel in 2004, escalating to physical infrastructure blockades in Kremna against Srbija Nikal in 2012, and ultimately converging into mass urban rallies and municipal coalitions in Užice from 2021 onward. A primary example of this socio-ecological dynamic occurred during the 2012 escalation. When the central government issued exploratory permits to Serbia Nikal, local villagers and environmental groups weaponized their geographic position, physically blocking critical transport corridors to deny entry to heavy machinery. This defensive nimbyist dispute rapidly scaled up as it gained backing from high-profile cultural figures and members of the Serbian Academy of Sciences and Arts (SANU), who publicly warned against the catastrophic deployment of hydrometallurgical technologies on vulnerable mountain terrains. Through this multi-level pressure, the sustained mobilization achieved substantial institutional resolutions, including the legal expansion of the protected boundaries of the Mokra Gora Nature Park, the formal revocation of corporate exploration permits by the Ministry of Mining and Energy, and the signing of a binding anti-mining memorandum by over thirty Užice municipal council members. Consequently, the defense of Mokra Gora has transcended its original localized scope, transforming into a primary institutional pillar of Serbia's broader national environmental movement aimed at securing a permanent parliamentary ban on predatory extractive industries (Bjelajac et al.) _. Mobilization Against construction of Small Hydropower Plants (SHPs) in Serbia (2018–2019)

The protests over construction of small hydropower plants (SHPs) in Serbia in 2018 ignited a significant surge in environmental consciousness throughout the country. What started as local frustration quickly grew into a widespread “ecological revolt” that questioned how the government was managing the country’s natural resources. The grassroots group “Defend the Rivers of Mt. Stara Planina” played a central role, using social media and striking images of dried-up riverbeds to show the public the real ecological dangers of these derivation-style plants—which threaten wildlife and ignore traditional water rights. As more people got involved, the conversation changed: rivers were no longer seen only as sources of economic value, but as shared treasures vital for rural communities and the health of the environment. By focusing on places like the village of Rakita and framing the fight to protect rivers as a basic human right, the movement helped foster a new sense of civic responsibility, emphasizing the importance of sustainable living over the “greenwashed” push for industrial growth at nature’s expense .

Jadar Lithium Mine protest

One of the most important and revealing events in the recent history of Serbian environmental awareness is the controversy over the proposed Jadar lithium mine in western Serbia. This case shows, better than almost any other example, how environmental awareness can be turned into political action. In the early 2000s, Rio Tinto's geologists conducted regional desk studies, analyzed historical geological data, and sent field geologists into the Jadar Valley

near Loznica to perform surface rock and soil sampling and in the Jadar valley near the town of Loznica in western Serbia. They confirmed the deposit of the mineral, later named jadarite after the Jadar River, is rich in both lithium and boron. (Rio Tinto). The company's plans included an open pit mine, a processing facility to extract lithium and boron from the ore, and associated infrastructure. The project was expected to cost around €2.4 billion and operate for at least 40 years. The Serbian Government strongly supported the project, pointing to the potential for thousands of jobs, significant export earnings, and a strategic role for Serbia in the EU's battery supply chain.

Jadar Valley is Serbia's fertile agricultural area. Families there have farmed the same land for generations, growing fruit, vegetables, and grain. Many residents were deeply opposed to the project for several reasons. The most important concern of potentially affected communities was surface and groundwater water contamination. Processing lithium ore from jadarite requires large quantities of sulphuric acid and produces significant amounts of toxic waste. Critics argued that there was a serious risk of toxic chemicals leaking into the Jadar and Drina rivers, and into the groundwater that local farms depend on for irrigation. agricultural land.. For these communities, the project was not just an environmental threat but an attack on their way of life and identity. There were also concerns about whether the Serbian government's environmental monitoring agencies would be strong enough to hold a powerful multinational company accountable if something went wrong. Given Serbia's history of weak environmental enforcement these concerns were reasonable (Bankwatch Network). One of the critical moments, which marked the resistance to lithium mining in Serbia, was the persistent refusal of a part of the local community (less than 10%) in the Jadar Valley to sell their land and houses to Rio Tinto and refusing to engage with the company which presented a major setback for Rio Tinto and the beginning of the mass mobilization (Djukanović): By late 2021, opposition to the Jadar project had grown into the largest environmental protest movement Serbia had seen since the political protests of 2000. The campaign was known by the slogan “We Won't Give Up Jadar." It brought together a remarkably diverse group of people: local farmers, city residents, scientists, Orthodox Church leaders, opposition politicians, artists, and young activists. In November and December 2021, protesters blocked major roads and bridges across Serbia in a coordinated series of rallies. The protests were peaceful, organized, and highly visible. Social media was central to their reach: live streams, photos, and project data spread quickly across Serbian networks and were amplified by international media (Petrović et al.,). In response to overwhelming public pressure, the government withdrew the mine’s planning approval in January 2022, preventing the project from moving forward. It also revoked Rio Tinto’s environmental license. Protesters celebrated what many called a historic victory — the first time in modern Serbia that a mass citizen environmental campaign had succeeded in stopping a major government-backed project (Balkan Insight, 2024). However, the issue was not over. In July 2023, Serbia's Constitutional Court ruled that the cancellation of the planning permission had been done incorrectly from a legal point of view. The government then moved to reinstate the permission, and in July 2024 Serbia signed a new agreement with Rio Tinto, citing Serbia's obligations under the EU's Critical Raw Materials

Act and the importance of lithium for the EU's green transition. This reversal triggered a new wave of protests in Belgrade and other cities, with tens of thousands of demonstrators taking to the streets. (N1 Serbia, Balkan Insight).

The Jadar controversy is one of the most important case studies in Serbian environmental awareness history for several reasons. Firstly, it shows how deep and broad environmental awareness has become in Serbia. The protests were not organized by a small group of professional environmentalists. They involved hundreds of thousands of ordinary people, farmers, parents, students, and were united by the belief that communities should have a real say in decisions that affect their land, water, and health. Secondly, it shows the limits of environmental democracy in Serbia.

Conclusions

The history of environmental awareness in Serbia is one of slow but unmistakable change, shaped by political transformation, shifting economic and political systems, industrial legacy, and civic struggle. From the muted environmental consciousness of the socialist period, through the setbacks of the 1990s, to the stronger public engagement and legal reforms of the post-2000 era, environmental issues have moved from the margins toward the center of public debate and democratic participation.

This paper shows that environmental awareness in Serbia developed gradually through slow political opening, complex historical experience, and major changes in economic and political organization. Its growth was strengthened by democratization after 2000, by the lived experience of pollution in affected communities, and by the work of civil society, journalists, and activists. At the same time, Serbia’s experience shows that modern laws alone are not enough, lasting progress depends on effective enforcement, institutional capacity, and public accountability. At the start of the 21st century, environmental activism in Serbia was weak and marginalized and environmental protest sporadic. In recent years, however, environmental issues have become major political concerns, driving the growth of environmental movements and initiatives. The current debate over the Jadar lithium mine captures the central challenge ahead: how to reconcile green transition and economic development with environmental justice, democratic decision-making, and protection of local communities and ecosystems.

Over the past decade, a new generation of environmental activists has emerged in Serbia. Young people, connected to global movements through social media and international networks, have organized climate strikes, creative protests, and public campaigns that have gained significant media attention These young activists tend to frame environmental issues differently from the previous generation: not just as problems of pollution or nature protection, but as issues of justice, who bears the costs of environmental damage, and who have the power to make decisions. This shift, from environmental issues being mainly the concern of scientists and specialist NGOs, to becoming part of a broader public conversation about rights, health, and democracy, is one of the most significant developments in Serbian society over the past twenty years.

Works Cited

Amidžić, Lidija, Mirjana Bartula, and Daniela Cvetković. "The State of Biodiversity in Serbia." Natural Areas Journal, vol. 34, no. 2, 1 Apr. 2014, pp. 222-226. https://doi.org/10.3375/043.034.0212.

Antić, A., M. Jovanović, and D. Nikolić. "Environmental Governance in Serbia: Progress and Challenges in the EU Accession Process." Environmental Policy and Governance, vol. 28, no. 4, 2018, pp. 251-263.

Balkan Insight. "Serbia's Lithium Mine: Protests, Politics and the Green Transition." 2024, balkaninsight.com.

CEE Bankwatch Network. Rio Tinto's Jadar Lithium Project in Serbia: Environmental Risks and Governance Concerns. CEE Bankwatch Network, 2022.

Bjelajac, Z., M. Dukic Mijatovic, and M. Dmitrovic Saponja. "Controversy Over the Impact of Nickel Ore on Environment and Safety in the Republic of Serbia." Oxidation Communications, vol. 40, no. 1-II, 2017, pp. 511-519.

Climate Action Network Serbia. Annual Report on Climate and Energy Policy in Serbia. Climate Action Network Serbia, 2023.

Djukanović, Nina. "Refusal of Lithium Mining in Serbia: Exploring Post-Green Extractivism." Globalizations, 20 Jan. 2026, https://doi.org/10.1080/14747731.2025.2597562.

Dragović, S., N. Mihailović, and A. Lolić. "Potential Environmental Impacts of Lithium Mining in the Jadar Valley: A Hydrogeological and Geochemical Assessment." Environmental Science and Pollution Research, vol. 29, no. 18, 2022, pp. 27451-27466.

Đorđević, Dina. Development of Resistance against Small Hydropower Plants in Serbia: From Local Actions to Mass Environmental Revolt. Belgrade Centre for Security Policy, 2022, bezbednost.org/wp-content/uploads/2022/01/Development-of-resistance-against-small-hydropower-plants-in-Serbia.pdf.

Eden, S. "Environment." International Encyclopedia of Human Geography, 2009, pp. 505-516, www.sciencedirect.com/science/chapter/referencework/abs/pii/B9780080449104002686.

European Commission. Critical Raw Materials Act. European Commission, 2023.

Hadjichambis, A. Ch., P. Reis, and D. Paraskeva-Hadjichambi, editors. European SWOT Analysis on Education for Environmental Citizenship. Institute of Education – University of Lisbon, Cyprus Centre for Environmental Research and Education, and European Network for Environmental Citizenship, 2019, enec-cost.eu/wp-content/uploads/2020/04/Country-Report_-SERBIA.pdf.

Levinbuk, M. I. "Refinery at War." Chemistry and Technology of Fuels and Oils, vol. 4, no. 3, p. 165.

Marković, D., M. Radovanović, and M. Filipović. "Industrial Pollution Legacy in Serbia: Assessment of Contaminated Sites and Remediation Priorities." Journal of Environmental Management, vol. 200, 2017, pp. 45-58.

Matić, Branislava I., et al. “Industrially Contaminated Areas in Serbia as a Potential Public Health Threat to the Exposed Population.” Tehnika, vol. 72, no. 3, 2017, pp. 441-447. https://doi.org/10.5937/tehnika1703441M.

N1 Serbia. “Mass Protests Against Rio Tinto Jadar Project Resume in Serbia.” 2024, n1info.rs/english.

Ostrić, Z. “Ekološki pokreti u Jugoslaviji 1971–1991.” Socijalna ekologija, vol. 1, no. 1, 1992, pp. 83-104, hrcak.srce.hr/file/201489. Serbo-Croatian.

Pančić, J. Flora Kneževine Srbije. Serbian Royal Academy, 1874.

Pančić, J. From Nature: Minor Works of Dr. Josif Pančić. Srpska književna zadruga, 1893. Serbian.

Pavlovic, Jovica. “The Evolution of Environmental Parties and Green Political Movements in the Republic of Serbia.” Politička revija, vol. 79, no. 1, 2024, pp. 85-102. https://doi.org/10.5937/pr79-48123.

Petrić, Hrvoje. “About Environmental Policy in Socialist Yugoslavia.” Nature and the Iron Curtain: Environmental Policy and Social Movements in Communist and Capitalist Countries, 1945–1990, University of Pittsburgh Press, 2019. https://doi.org/10.2307/j.ctvhrd0c1.

Petrović, Jelisaveta, and Jelena Pešić. Brief Overview of Socio Historical Development of the Environmental Movement in Serbia: From State Socialism to Capitalist Periphery. Institute for Sociological Research, isi.f.bg.ac.rs/wp-content/uploads/2021/07/Brief-Overview-paper-3.pdf.

Puljar, Sanja D’Alessio. “Introduction to ‘More-Than-Human-Sociality’ through Balkan Perspectives.” Southeastern Europe, vol. 49, 2025, pp. 1-23, brill.com/view/journals/seeu/49/1-2/article-p1_001.xml.

Rio Tinto. Jadar Project: Environmental and Social Impact Assessment — Summary. Rio Tinto plc, 2022.

Samardzic, M., and M. Beslin. “Deforestation of the Principality of Serbia, 1830-1878: Overview.” Geographica Pannonica, vol. 21, no. 1, 2017, pp. 1-8, scindeks-clanci.ceon.rs/data/pdf/0354-8724/2017/0354-87241701001S.pdf.

Samardžić, Momir, and Milivoj Bešlin. “Deforestation of the Principality of Serbia, 1830-1878: Overview.” Geographica Pannonica, vol. 21, no. 1, Mar. 2017, pp. 1-8.

Serbian Environmental Protection Agency, sepa.gov.rs.

Vujošević, M., Zeković, S., Maricic, T., European Planning Studies Post-Socialist Transition in Serbia and Its Unsustainable Path, European Planning Studies, 2012, Vol 20, No 10, p1707-1727

Stefanović, N., Danilović Hristić , N., and Petrić, J. Spatial Planning, Environmental Activism, and Politics—Case Study of the Jadar Project for Lithium Exploitation in Serbia, Sustainability 2023, 15, 1736.

Rio Tinto. Jadar Project: Environmental and Social Impact Assessment — Summary. Rio Tinto plc, 2022.

Simeunović, I., B. Stojanović, and M. Nikolić. “Water Policy Reform in Serbia in the Context of EU Accession: Progress, Gaps and Priorities.” Water Policy, vol. 21, no. 4, 2019, pp. 754-770.

The Kosovo Conflict: Consequences for the Environment and Human Settlements. UNEP / UNCHS, 1999.

UNEP Balkans Task Force: Post-Conflict Environmental Assessment — Serbia. UNEP, 2000.

Hadjichambis, A. Ch., P. Reis, and D. Paraskeva-Hadjichambi, editors. European SWOT Analysis on Education for Environmental Citizenship. Institute of Education – University of Lisbon, Cyprus Centre for Environmental Research and Education, and European Network for Environmental Citizenship, 2019, enec-cost.eu/wp-content/uploads/2020/04/Country-Report_-SERBIA.pdf.

Djukanović, Nina. “Refusal of Lithium Mining in Serbia: Exploring Post-Green Extractivism.” Globalizations, 20 Jan. 2026, https://doi.org/10.1080/14747731.2025.2597562.

Živanović, V. “Environmental Changes in the City of Pančevo.” Zbornik radova - Geografski fakultet Univerziteta u Beogradu, vol. 65-1a, 2017, pp. 449-462. https://doi.org/10.5937/zrgfub1765449Z.

Sleep and Neurological Diseases and Disorders Across the Life Span By Serena Tobler

Abstract

Sleep is an essential pillar of health at every stage of life, and even a single night of poor sleep can produce measurable cognitive deficits. Chronic sleep disturbances are increasingly linked to neurological disorders, including Autism Spectrum Disorder (ASD), Schizophrenia (SCZ), and Alzheimer's Disease (AD), where underlying causes remain incompletely understood. Converging evidence indicates that sleep disruption, together with gene mutations affecting synaptic function, interferes with homeostatic processes that stabilize neural circuits. This review examines how alterations in sleep timing, duration, and architecture impair synaptic plasticity, particularly the balance between Hebbian and homeostatic mechanisms. Across ASD, SCZ, and AD, persistent and non-restorative sleep disturbances are associated with circadian dysregulation, reduced slow-wave activity and sleep spindles, and impaired synaptic downscaling. These disruptions contribute to disorder-specific outcomes, including behavioral dysregulation in ASD, psychosis, and cognitive impairment in SCZ, and neurodegeneration and β-amyloid accumulation in AD. Shared molecular pathways, including NMDA receptor dysfunction, GABAergic imbalance, and mutations in synaptic proteins such as SHANK3, further link sleep abnormalities to disease progression. The bidirectional relationship between sleep and pathology suggests that sleep disturbance both reflects and drives neurological dysfunction. Together, these findings highlight sleep as a central mechanism in synaptic stability and a promising target for therapeutic strategies across multiple neurological disorders.

Introduction

In this review, we discuss the overlap of Autism Spectrum Disorder (ASD), Schizophrenia (SCZ), and Alzheimer's Disease (AD), specifically the impact of sleep disruption on synaptic plasticity. We will first cover an overview of sleep and synaptic plasticity, and then an overview of each disease: its symptoms, diagnosis, genetic and environmental factors, relevant sleep information, and treatment options. Finally, we examine common mechanisms of the diseases and their impact on sleep and associated processes.

Autism Spectrum Disorder (ASD), Schizophrenia (SCZ), and Alzheimer's Disease (AD) are all neurological diseases that have little-known etiologies. However, all three are synaptopathies, or diseases of the synapse, and have chronic poor sleep as a symptom (Ishida et al. 2018). Understanding these disorders and how atypical sleep affects patients can provide valuable insight into diagnosis, prevention, and symptom alleviation.

Sleep is vital for several homeostatic processes and helps maintain physical and mental health. It is regulated by the circadian rhythm (Deboer 1). Disruptions to its rhythm are seen as both markers and potential drivers of neurological and psychiatric disorders. Sleep plays a major role in returning the synapse to homeostasis. There have been several links made between sleep duration or quality and the ability for synapses to downscale (Hanlon et al. 2). Examining

atypical sleep patterns, sleep disruptions, and their effects on synaptic plasticity can give insight into neurological diseases that are driven by synaptic processes.

Synaptic plasticity comprises both Hebbian mechanisms that encode experience and homeostatic mechanisms that stabilize network activity. Converging evidence indicates that sleep promotes synaptic normalization and memory consolidation, whereas sleep loss disrupts homeostatic scaling and biases circuits toward maladaptive potentiation. Thus, examining how sleep timing, duration, and continuity alter synaptic plasticity can clarify disease mechanisms underlying disorders with circuit instability.

We will discuss the impact of sleep disruption on neurological disorders, focusing on ASD, SCZ, and AD. We describe evidence for a link between sleep and synaptic plasticity, which could elucidate the pathologies driving these conditions and provide new directions for therapeutic strategies.

Sleep and Synaptic Plasticity

Sleep: Structure, Stages, and Regulation

Sleep plays a crucial role in regulating homeostatic processes, one of which is homeostatic synaptic plasticity, a negative feedback mechanism that stabilizes neural networks.

During sleep, our bodies are unconscious and have reduced responsiveness to the outside world. There are two main types of sleep: rapid eye movement (REM) and non-rapid eye movement (NREM), with three different stages (N1, N2, N3), each with a unique function. We cycle through each stage in a 90-minute cycle about 5-6 times each night. The cycle starts with N1, characterized by light sleep, followed by N2 with deeper sleep, lower body temperature, and reduced heart rate. This stage is also characterized by two well-known brain wave patterns, sleep spindles and K complexes, as recorded on an EEG monitor. Sleep spindles are believed to play a role in synaptic plasticity. The last stage in NREM is N3, when sleep is the deepest. At this stage, the body repairs, rejuvenates, and replenishes itself. After N3, our bodies quickly pass again through N2 and finally reach the last stage of the sleep cycle, REM sleep. During this phase, most dreaming occurs, and the eyes move rapidly. About 75% of our sleep is spent in the NREM phase, mostly in the N2 phase (Patel et al 4).

The circadian rhythm and homeostatic sleep are the two main processes that regulate our sleep patterns. While adapting to external signals, they work together to maintain a healthy balance between wakefulness and rest. The circadian rhythm aligns our internal clock to the day-night cycle, and homeostatic sleep increases the need for sleep the longer we stay awake. The circadian rhythm is controlled by two cues: light and internal rhythms. Homeostatic sleep controls two aspects of sleep: length and depth. It is regulated by sleep need and characterized by slow-wave activity on an EEG. The more sleep-deprived we are, the more slow-wave activity can be observed.

Mechanism of Synaptic Plasticity 2

There are two types of synaptic plasticity: Hebbian and Homeostatic. Hebbian plasticity strengthens or weakens synapses depending on their activity. It is an extremely fast process that uses positive feedback mechanisms to encode information. Homeostatic plasticity is a mechanism that returns neuronal activity toward a set point after a perturbation, preventing saturation or silencing of neuronal signals. It uses negative feedback mechanisms to preserve coding abilities.

One purpose of sleep is restoring synaptic homeostasis. A popular theory is the 'Synaptic homeostasis hypothesis'. Synaptic plasticity is the brain's ability to change and adapt through experiences, with neural pathways strengthening as they are being used. While awake, these connections grow stronger and consume energy and space. To maintain balance, it is necessary to weaken connections periodically. This restorative process occurs during sleep. According to SHY, wakefulness leads to a net increase in synaptic strength, and sleep to a net decrease. This downscaling restores energy balance and prepares the brain to adapt efficiently again upon waking. In essence, sleep provides a vital reset, freeing the brain from external demands and maintaining overall neural stability (Tononi and Cirelli 2).

The hypothesis heavily supports that slow wave activity (SWA) is a reflection of the average synaptic strength reached during the prior wake period, and that SWA mediates synaptic depression. Stronger synapses show larger and steeper SWA triggered by an increase in neuronal synchrony. The amplitude of SWA is related to the number of active neurons that are in an up or down state near synchrony, which is directly related to the strength and magnitude of synaptic connections.

Another theory that builds off of SHY is that changes in the slope of SWA reflect homeostatic changes, specifically the decrease of sleep pressure. (Hanlon et al. 2). This mechanism suggests that slow wave activity is not just related to synaptic depression, but is directly related to the functions of sleep.

Regardless of the mechanism, there is a strong link between synaptic depression and slow waves that helps regulate synaptic strength. According to SHY, sleep helps renormalize synaptic strength, promoting downscaling of connections that strengthen during periods of wake (Tononi and Cirelli 4).

Autism Spectrum Disorder and Sleep

Symptoms and Diagnosis

ASD is defined by social deficits and repetitive behaviors. In severe cases, there is a significant lack of communication, often expressed by nonverbal or very limited verbal communication. Since people with severe autism cannot be independent, some individuals have a substantial need for support in daily living. We will discuss diagnosis, relevant gene mutations, environmental factors, and what role sleep plays within ASD.

During the last couple of decades, the diagnostic process for ASD has been refined significantly. New symptoms and prognosis have made identifying and diagnosing those with ASD much easier. The first measurable behavior changes or social decline in children usually starts at age 2. Those with less severe symptoms may not be officially diagnosed until later in life. The diagnostic process is divided into two categories: social communication and restrictive or repetitive behavior. The severity of symptoms in each category will determine the level and type of care needed, on a case-by-case basis (Hirota and King 5).

Genetic Etiologies of ASD and Synaptic Dysfunction

Other than behavioral indicators to determine the presence of ASD, mutations related to synaptic dysfunction have been correlated to risks regarding both ASD and sleep disorders.

For example, mutations in the SHANK3 gene are linked to the development of ASD. Shank proteins are found in the postsynaptic density region of the synapse. They provide scaffolding to help organize neurotransmitter receptors, which play an important role in synapses by maintaining postsynaptic strength (Wu, Tatavarty, Pierre M Jean Beltran, et al. 1).

Mutations in Shank proteins have been linked to the development of several neurological diseases or disorders, including ASD and AD (Ishida et al. 2). Mice with a deletion in the SHANK3 gene displayed traits such as repetitive behaviors and social interaction deficits. On top of that, they exhibited insomnia related traits, circadian abnormalities, and decreased REM sleep. There is likely a relationship between a mutation in the SHANK3 gene, ASD related behaviors, sleep insomnia, and atypical synaptic plasticity function (Ishida et al. 2)

Another gene involved in synaptic function and associated with ASD is CNTNAP2. When mutated in a zebrafish model, the larvae displayed increased nighttime activity. However, when estrogen was injected into the mutated larvae, much of their hyperactivity decreased, which might be relevant for ASD diagnosis. There is a 1:4 ratio of females to males in people diagnosed with ASD. Estrogen increases GABA and NMDA signalling. Since compromising these pathways can lead to synaptic dysfunction, estrogen may act as a stimulus to increase synaptic function, which can have protective effects against ASD symptoms and pathogenesis (Doldur-Balli et al. 20)

Environmental Risk Factor and Sleep Disturbance

Other than genetic mutations, there have been links between ASD and certain prenatal, neonatal, and postnatal environments, including a lack of maternal attention or care, and environmental pollutants. One such example is how most individuals with ASD also have sleep issues, especially insomnia. Other common disorders include parasomnia, daytime sleepiness, bedtime resistance, sleep-disordered breathing, and more.

Sleep issues in people with ASD often include Insomnia, sleep onset latency, reduced sleep efficiency, and reduced total sleep time. Around 40% of autistic adults report some circadian rhythm disorders (Ballester et al., 9). These disorders arise due to clock and melatonin pathway gene variations, not routine light exposure patterns, or brain development abnormalities

that fail to understand or recognize social cues associated with sleep timing. Circadian rhythm sleep-wake disorders can be treated with external melatonin (Ballester et al., 9).

These sleep disturbances have a negative effect on quality of life. Synaptic homeostasis is the process by which the synaptic signaling strengths decrease and return to a sustainable state in terms of energy intake and strength. The main challenges for individuals who have ASD are social communication, interaction, repetitive behaviors, and unusual sensory processing. Without being able to downscale, the synaptic signaling pathways could worsen their conditions, further challenging their lives.

Treatment and Sleep Effects

Aripiprazole and risperidone are the only two approved drugs to treat ASD symptoms. They mainly affect behavioral symptoms like irritability and disruptiveness, but they are not preventative of sleep issues. However, Antipsychotics, such as risperidone, have also been shown to significantly improve sleep and reduce aggressive behaviors in both low and high doses (Ballester et al. 13).

Synaptic signaling provides insight into the pathogenesis of ASD. Specifically, the SHANK3 gene highlights the importance of synaptic downscaling. Combining how sleep impairs synaptic downscaling and the effects of SHANK3 mutations, regulating synaptic pathways is a vital function that helps with memory and cognitive function. Sleep can also act as a treatment for improving synaptic downscaling in ASD patients, improving adaptability and cognitive function.

Conclusion

As discussed above, the main driving factors include genetic and environmental factors. However, the etiology for 70% of ASD is still unknown (Hirota and King 3). What may address this question is the interplay between a host of causal factors, such as genetic predispositions and environmental insults, such as sleep loss. Future research must address many factors at once to really conclude causation for ASD.

Schizophrenia and Sleep

SCZ affects 0.4% of people worldwide and has detrimental impacts on life quality (Winship et al. 2). In this section, we will discuss what schizophrenia is, how it’s diagnosed, relevant genes and other possible causes, the impact that sleep and synapses have on SCZ, and treatment options.

Symptoms and Diagnosis

Schizophrenia symptoms are categorized as positive or negative symptoms. Positive symptoms are excessive or distorted forms of normal functioning. This includes delusions, hallucinations, disorganized speech and behavior, symptoms that deviate from an individual's perspective of reality. On the other hand, negative symptoms are the decrease or absence of

normal functions. These include apathy, social isolation, and diminished affect. Finally, cognitive dysfunction refers to difficulty in mental processing that helps us adapt to the world around us. These processes include memory, paying attention, problem-solving, language, and reasoning. Taken together, individuals with schizophrenia have a hard time distinguishing perception from reality.

Several genes have been found to increase the risk of developing SCZ. Diagnosing SCZ is challenging because there is no quantitative scale for SCZ, and behavioral diagnosis is often ambiguous. To be diagnosed, patients must meet 6 criterias. These include:

Criteria	Conditions
1.	1. Experiencing delusions, hallucinations, disorganized speech, or catatonic behavior during one month. Two of these symptoms must be experienced and must be present for a significant amount of time.
2.	Normal level of functioning in work, personal life, or self-care is below the expected level of achievement before experiencing symptoms.
3.	Positive and negative symptoms persist for at least 6 months, meeting the first criterion.
4.	Schizoaffective disorder and depressive or bipolar disorder are ruled out due to no manic, major depressive, or mood episodes occurring during the duration of the symptoms.
5.	The symptoms are not a result of substance, drugs, medication, or another medical condition.
6.	If there is a history of ASD or communication disorders, a diagnosis of schizophrenia is only made if there are prominent hallucinations or delusions, in addition to other symptoms, prevailing for at least 1 month.

Symptoms do not typically start to appear until adolescence or early adulthood, appearing gradually over time. Schizophrenia presentation can often be mistaken for typical adolescent behavior. On top of that, schizophrenic behavior overlaps with other mental disorders, such as bipolar disorder, major depressive disorder, substance use disorders, and delirium or dementia. To diagnose schizophrenia, all other mental disorders must be ruled out (Rahman and Lauriello 3).

Genetic causes of schizophrenia: with a focus on synaptic-related genetic mutations

Although there is no known direct cause for schizophrenia, there are many risk factors that can influence the chance of developing this mental disorder. Overall, there is an increased risk among men to develop schizophrenia, with a 1.4:1 male: female ratio. Males are typically diagnosed earlier than females, but often have worse outcomes. Individuals with schizophrenia are much more vulnerable to becoming homeless due to their warped perspective of mortality.

They also have a mortality rate of two to three times that of normal humans (Rahman and Lauriello 2).

There are few known genetic causes of schizophrenia, many involving genes related to proteins relevant for neural plasticity, synaptogenesis, and glutamatergic or dopaminergic function. For example, the gene DISC1 is considered a high-risk gene for schizophrenia. DISC1 mutations in mouse models have shown a brain morphology of enlarged ventricles and reduced cortical thickness and behavioral abnormalities like hyperactivity and impaired volitional behavior, all consistent with schizophrenia.

22q11.2 deletion syndrome is a genetic disorder that occurs when part of chromosome 22 has been deleted. Chromosome 22. Unmutated, chromosome 22 is responsible for providing instructional proteins for our immune system, heart development, and neural function. However, when mutated, deficiencies include but are not limited to: immune deficiency, learning disabilities, and heart disease (McDonald-McGinn et al. 1). It is also associated with developing schizophrenia (Bassett and Chow 1). Around 1.1% of schizophrenic cases are caused by this syndrome. In mice with this deletion syndrome, there have been alterations in cortical and subcortical grey matter volumes, enlarged ventricles, and reduced spine size and dendritic complexity (Winship et al. 5).

Dysbindin is a synaptic protein that regulates many systems involved with synaptic transmission and is encoded in the DTNBP1 gene. DTNBP1 mutations impact NMDA receptor functionality and impair memory. Mice carrying DTNBP1 mutations have also shown schizophrenia-like behavior, and dysbindin has been identified as a risk factor (Winship et al. 5).

Reduction in the expression of Reelin, a protein involved in synaptic formation and plasticity, has also been reported in patients with schizophrenia. Reelin mutations have been reported to produce schizophrenia-like conditions in mice, including increased neuronal packing and reduced dendritic spine density in the hippocampus and frontal cortex. This can lead to deficits in memory, executive function, and social behavior. However, the behavior, social interactions, and cognitive deficits in the mice have been found to yield variable results.

Both glutamate and NMDA receptors are highly involved with parts of the synapse, allowing Ca2+ into postsynaptic dendrites. However, mutations in the genetic coding for these complexes are highly associated with schizophrenia. When mutated, the shape, structure, and thus function of both glutamate and NMDA receptors are not able to properly regulate the synapse, influencing synaptic signaling strength. This lack of regulation has links to both positive and negative symptoms of schizophrenia (Hall and Bray 3).

Genetic risk factors for schizophrenia impact both the formation of synaptic networks during brain development and their function in adulthood. The observation that several key genes linked to schizophrenia affect processes like synaptogenesis and neuronal connectivity is consistent with the neurodevelopmental theory, proposing that issues with establishing synaptic connections early in life can contribute to the onset of schizophrenia.

Environmental factors: with a focus on sleep disruption

Environmental factors like prenatal stress, nutritional issues, and substance use can worsen the risk of developing SCZ, especially in people with a genetic predisposition.

Sleep dysfunction is a common issue for people with schizophrenia, and it often correlates with psychotic experiences. Many SCZ patients also have insomnia, which worsens symptoms. However, the opposite was reported as well: sleep issues worsen psychotic episodes. Some patients use extreme measures, like hypnotic or neuroleptic medication, to initiate sleep (Ferrarelli 7).

There are many correlations between sleep and psychotic episodes. Around 250,000 people reported that insomnia significantly increases the chance of experiencing a psychotic symptom. On top of that, studies have shown that nightmares were associated with paranoia and hallucinations. In a study with normal individuals, sleep deprivation was shown to induce psychotic episodes, including cognitive disorganization, anhedonia, and perceptual distortion. Even in those at risk of developing schizophrenic symptoms, shortening sleep correlated with higher paranoia, and less restful sleep predicted more hallucinations in healthy individuals. Sleep disturbances are not just a by-product of psychotic symptoms, but can also be a contributing factor to them (Benson 4).

Treatment Approaches and Sleep

Cognitive-Behavioral Therapy for Insomnia (CBT-I) can be used as a therapy to reduce insomnia. For patients with SCZ, there is evidence that this therapy reduces psychotic symptoms and hallucinations. Studies also show that individuals with SCZ have deficits in oscillatory activities like sleep spindles and slow-wave activities. Non-invasive brain stimulation (NIBS) techniques have been used to modulate sleep oscillatory activity, resulting in better sleep for those with psychotic episodes (Ferrarelli 7).

Schizophrenia has two main approaches for treatment: medical and physical. One of the main medical or drug-related treatments targets dopamine. An increase in dopamine synthesis can lead to psychosis, so many antipsychotics downregulate dopamine production or signaling pathways (Yang and Tsai 5). However, these antipsychotics don't eliminate dysfunctions. Gamma-Aminobutyric Acid (GABA) and glutamate are neurotransmitters that are associated with psychosis. Glutamate, for example, has the ability to bind to NMDAs to decrease negative symptoms and cognitive impairment. GABA is an inhibitory neurotransmitter in the central nervous system. It is vital to the brain's rhythm-generating networks, mechanisms for memory, perception, and consciousness. Abnormalities in GABA signaling may be causative for schizophrenia, and substances that mimic GABA have been linked to improving schizophrenic symptoms. Lessening of negative symptoms has further been attributed to gluten-free diets, more exercise, and treatments targeting hormones, estrogen receptors, and nicotinic and muscarinic cholinergic agents. Those with schizophrenia generally practice less physical activity than the general population (Maurus et al. 2). In general, exercise and physical activity have been shown to promote mental health. The increased impact of physical activity on mental and physical

health has made it a therapeutic approach. Exercise has the potential to increase psychotic episodes, symptom severity, and overall quality of life (Maurus et al. 2).

Alzheimer's Disease and Sleep

Clinical Features and Diagnosis of AD

AD is a severe dementia that causes memory loss, confusion, and cognitive decline. It is the most common type of dementia, affecting over 55 million worldwide. Currently, there are no cures available, and only a few treatments can slow its progression (Cummings et al. 1). The gold standard for diagnosis is the presence of β-amyloid plaques, intraneuronal neurofibrillary tangles, and amyloid angiopathy. The three conditions for AD to be diagnosed are:

1. The plaques should stain positively with β-amyloid antibodies and be negative for prion antibodies (which are diagnostic of prion diseases).
2. An increased number of plaques and tangles is found compared to those of similar age who do not have dementia.
3. Frequent accumulation of TDP-43 protein, aggregation of alpha-synuclein in the form of Lewy bodies in the amygdala

Symptoms and Disease Progression

AD is characterized by a variety of symptoms that affect mental processing and memory. The most common symptom is memory loss. Most patients initially have short-term memory loss. As symptom severity increases, the patient's cognitive function decreases. Other symptoms include poor judgment, confusion, language disturbance, visual complaints, withdrawal, and hallucinations. In some cases, seizures, symptoms similar to Parkinson's disease, and increased muscle tone can occur. 95% of Alzheimer's patients start to show symptoms after the age of 65. The duration of the disease usually lasts 8-10 years, but more severe cases often shorten the duration. Death is mainly caused by general inanition, malnutrition, and pneumonia. Those with early onset AD (EOAD), defined by a start of disease before the age of 65, often have very aggressive symptoms (Bird 2).

75% of AD cases have no known etiology (Hodges et al. 1). Familial inherited AD can be caused by mutations in the amyloid precursor protein, yet these mutations only account for about 1% of AD cases. Similarly, the majority of Down Syndrome patients have AD pathology by the age of 40 and develop clinical symptoms by the age of 65 (DeTure and Dickson 2).

Treatment Approaches

There are currently only 5 treatment options to improve the cognitive symptoms of AD. These include 3 types of cholinesterase inhibitors (galantamine, rivastigmine, donepezil), one NMDA receptor antagonist, and a combination of a fixed dose of donepezil and memantine is approved. Galantamine increases cognitive performance, increasing the management of daily

activities. Rivastigmine increases metabolism in the hippocampus, producing satisfactory cognitive outcomes. However, the drug can induce nausea, dizziness, and vomiting. Rivastigmine works by inhibiting AChE and butyrylcholinesterase in the cerebrospinal fluid of the brain. Donepezil is also an AChE inhibitor, increasing cognitive function and decreasing NPI and NPI-D scores. NPI (Neuropsychiatric Inventory) measures neuropsychiatric symptoms in individuals with dementia, while the NPI-Distress score (also known as NPI-D) quantifies the level of emotional and psychological distress experienced by the patient's caregiver due to those symptoms.

Lastly, memantine acts as an NMDA receptor antagonist by activation of NMDA ion channels in a preferential manner, ultimately resulting in an increase in normal neuronal activity. This drug results in an increase in cognitive and behavioral activities (Cummings et al. 2).

Sleep and Circadian Disruption in AD

As AD progresses, the circadian rhythm becomes more irregular, and sleep disturbance becomes more regular. The neuropathic changes caused by AD are likely the reason for these changes. On the other hand, an increase in sleep disturbance and an unsynchronized circadian rhythm can also increase the risk factor for developing AD. The formation of memories is dependent on undisturbed sleep. Shorter sleep has been linked to impaired cognitive function. Those individuals with greater levels of sleep disturbance showed a more rapid rate of cognitive decline and a higher risk of developing AD (Irwin and Vitiello 2).

On top of that, sleep disturbance and excessive daytime sleepiness have been associated with an increase in β-amyloid deposition. Additionally, sleep-disordered breathing also increased the risk of dementia and was associated with a higher incidence of all-cause dementia, Alzheimer’s disease, and vascular dementia. Overall, sleep disturbances and insomnia can vastly increase the risk of developing AD (Irwin and Vitiello 2).

Inflammation linking Sleep and AD

When investigating potential mechanisms for the relationships between sleep disturbances and AD, inflammation emerges as a central mechanism connecting sleep disturbance and AD progression, with sleep quality closely tied to inflammatory activity. Improved sleep corresponds with reduced inflammation, whereas disrupted sleep contributes to its elevation. This heightened inflammatory response appears early in the development of Alzheimer’s disease and intersects with lifestyle factors such as obesity and physical inactivity, which further amplify inflammatory processes. These same processes influence key pathways involved in AD progression, including the accumulation of β-amyloid peptides. In humans, sleep facilitates the reduction of β-amyloid concentrations, while sleep disturbance triggers stress responses that stimulate immune activity and elevate inflammation (Irwin and Vitiello 5).

Emerging Therapies

Because of AD's complexity, many treatments have failed. There has been a recognition in recent years of the use of combination therapy to target symptoms. These combinations target behavioral and cognitive symptoms and are flexible in addressing drug targeting, delivery, or timing. Some examples include the use of multifunctional molecules, such as rasagiline, a MAOB inhibitor. This has effects on the amyloid processing (Cummings et al. 8).

Gene therapy has also been a strategy that has been extensively studied. Treatment would generally involve mutating the gene that expresses a therapeutic enzyme or growth factor. Targeted genes include APP, PSEN1, and PSEN2, APOE: all genes that have proven to increase the risk of AD (Bird 7).

ASD, SCZ, AD, and Sleep Overlap

Shared Role of Synaptic Dysfunction

In all 3 neurological diseases or disorders, synaptic dysfunction plays a large role in the development and progression. Furthermore, sleep disturbances have a similar impact on behavior and progression of each disease. In ASD, sleep disturbance amplifies core symptoms, worsening emotional and physical regulation. In SCZ, sleep disturbance provokes key symptoms of psychosis and cognitive impairment. Lastly, sleep disturbance accelerates neurodegeneration and memory loss in AD. Although these effects of sleep disturbances are varied, each has negative effects. This is partially due to the effect that sleep disruption has on synaptic homeostatic abilities (Tononi and Cirelli 3). In ASD, lack of adequate REM disrupts synaptic downscaling and can worsen sensory and behavioral rigidity.

NMDA Receptor Dysfunction across Disorders

In SCZ, sleep loss leads to hyperdopaminergic states and psychosis-like symptoms.

AD, SCZ, and ASD all involve synaptic interactions related to the NMDA receptor. In Alzheimer's, one treatment option is an NMDA antagonist (Cummings et al. 2). In SCZ, when dysbindin, a gene that codes for synaptic transmission, is mutated, working memory is impaired (Winship et al. 3). In ASD, NMDA receptors can lead to ASD symptoms. (Doldur-Balli et al. 10) The NMDA receptor plays a crucial role in presynaptic glutamate release and postsynaptic homeostatic synaptic plasticity. This strengthening and weakening of the pathways is crucial for learning and memory. When the receptor is mutated or compromised, it can lead to significant defects in the synaptic pathways, leading to certain neurological disorders such as ASD, AD, and SCZ.

SHANK3 and Synaptic Scaffolding Proteins

The Shank gene is prominent in synaptic plasticity. Shank proteins act as scaffolding proteins, organizing neurotransmitter receptors, cytoskeleton proteins, protein ion channels, and enzymes. Overall, this forms a large signaling device crucial to synaptic postsynaptic density

regions. Mutations of the SHANK3 gene have been associated with ASD, AD, and SCZ (Wu, Tatavarty, Pierre M. Jean Beltran, et al. 2)

GABAergic Dysfunction and Neural Excitability

GABA plays a central role in homeostatic synapses. GABA regulates neural excitability by inhibiting overactive circuits. It also plays a central role in sleep structure by promoting slow-wave and REM sleep and in synaptic homeostasis, which restores circuit stability during sleep. When GABA signaling is disrupted, the brain experiences hyperexcitability. In AD, reduced GABAergic signaling contributes to sensory hypersensitivity, anxiety, and repetitive behaviors, all core ASD traits. Sleep studies also show that reduced GABA and melatonin dysregulation may jointly disturb REM sleep and circadian rhythm. However, estrogen can enhance GABAergic signaling, possibly explaining the lower prevalence of ASD in females. (Doldur-Balli et al.). In SCZ, GABA dysfunction primarily affects prefrontal cortical interneurons, especially parvalbumin (PV)-positive interneurons, which synchronize neural oscillations. Studies also show reduced GAD67 (an enzyme that synthesizes GABA) expression in PV neurons, and reduced GABA transporter levels, which can cause abnormal neural rhythms, disrupting working memory, attention, and sensory gating. This loss of inhibitory control results in increased dopaminergic and glutamatergic activity, contributing to hallucinations and delusions, key symptoms of SCZ. (Winship et al. 2)

Sleep spindles and slow-wave sleep, which rely on GABAergic thalamocortical synchronization, are reduced in SCZ and AD— directly linking GABA dysfunction to sleep disruption. Reduced Sleep spindles and SWA are also linked to synaptic homeostatic processes. GABA deficits are largely developmental, disrupting circuit formation and synchronization, which is why there is more overlap in ASD and SCZ than AD (Doldur-Balli et al. 4).

Disruption of Synaptic Homeostasis by Sleep Loss

A major overlap lies in how sleep disturbances affect synaptic plasticity. Normally, sleep supports synaptic downscaling. (Tononi and Cirelli 2). In all three disorders, disrupted sleep interferes with this process, leading to excessive synaptic strength or instability.

All three disorders show persistent and non-restorative sleep disturbances rather than occasional sleep issues. In ASD, insomnia, long sleep onset latency, and reduced total sleep time are common symptoms (Ballester et al. 8). In schizophrenia, insomnia, fragmented sleep, and nightmares often occur (Ferrarelli 1). In Alzheimer's disease, frequent awakenings, circadian fragmentation, and daytime sleepiness often occur. (Irwin and Vitiello 1). Although expressed differently, each disorder involves disruption of the circadian timing system.

There is a Bidirectional Relationship between sleep and symptoms in each disorder. Sleep and sleep disturbances are both a result of underlying pathology and also a driver of symptom severity and progression. For example, poor sleep worsens behavioral symptoms in ASD (Ballester et al. 18). In Schizophrenia, sleep loss can induce or intensify psychosis (Benson 2).

Sleep disruption accelerates cognitive decline and β-amyloid accumulation in Alzheimer's disease (Irwin and Vitiello 2).

Conclusion

Sleep is a fundamental regulator of neural homeostasis, with a central role in maintaining synaptic plasticity and overall brain function. This review highlights how disruptions in sleep—particularly in timing, duration, and architecture—interfere with homeostatic synaptic processes, contributing to neurological dysfunction. Sleep supports synaptic downscaling and memory consolidation, while sleep loss biases neural circuits toward maladaptive potentiation. Because synaptic plasticity relies on a balance between Hebbian and homeostatic mechanisms, disturbances in sleep can destabilize neural networks and impair cognitive and behavioral function.

Autism Spectrum Disorder (ASD), Schizophrenia (SCZ), and Alzheimer's Disease (AD) share underlying features as synaptopathies, with genetic and environmental factors converging on disrupted synaptic signaling. Across all three conditions, sleep disturbances are highly prevalent and closely tied to symptom severity. In ASD, sleep disruption exacerbates behavioral rigidity and sensory dysfunction; in SCZ, it contributes to psychosis and cognitive impairment; and in AD, it accelerates neurodegeneration and memory decline. These effects are mediated through overlapping biological pathways, including NMDA receptor dysfunction, GABAergic imbalance, and mutations in synaptic scaffolding proteins such as SHANK3, all of which impair synaptic regulation and network stability.

A key unifying theme is the bidirectional relationship between sleep and disease progression. Persistent, non-restorative sleep disturbances—such as insomnia, circadian dysregulation, and impaired sleep architecture—both result from and contribute to pathological processes. Disruption of slow-wave activity and sleep spindles further impairs synaptic homeostasis, reinforcing neural instability. In AD, sleep disturbance is additionally linked to inflammation and β-amyloid accumulation, while in SCZ and ASD, altered neurotransmitter systems and developmental circuit abnormalities play a larger role. Together, these findings position sleep as a critical mechanism underlying disease progression and a potential target for therapeutic intervention across multiple neurological disorders.

Works Cited

Ballester, P., et al. "Sleep in Autism: A Biomolecular Approach to Aetiology and Treatment." Sleep Medicine Reviews, vol. 54, Dec. 2020, p. 101357. PubMed, https://doi.org/10.1016/j.smrv.2020.101357.

Bassett, Anne S., and Eva W. C. Chow. "Schizophrenia and 22q11.2 Deletion Syndrome." Current Psychiatry Reports, vol. 10, no. 2, Apr. 2008, pp. 148–57. PubMed Central, https://doi.org/10.1007/s11920-008-0026-1.

Benson, Kathleen L. "Sleep in Schizophrenia: Impairments, Correlates, and Treatment." The Psychiatric Clinics of North America, vol. 29, no. 4, Dec. 2006, pp. 1033–45; abstract ix-x. PubMed, https://doi.org/10.1016/j.psc.2006.08.002.

Bird, Thomas D. "Alzheimer Disease Overview." GeneReviews®, edited by Margaret P. Adam et al., University of Washington, Seattle, 1993. PubMed, http://www.ncbi.nlm.nih.gov/books/NBK1161/.

Cummings, Jeffrey L., et al. "Treatment Combinations for Alzheimer's Disease: Current and Future Pharmacotherapy Options." Journal of Alzheimer's Disease: JAD, vol. 67, no. 3, 2019, pp. 779–94. PubMed, https://doi.org/10.3233/JAD-180766.

Davis, Christopher J., et al. "REM Sleep Deprivation-Induced Deficits in the Latency-to-Peak Induction and Maintenance of Long-Term Potentiation within the CA1 Region of the Hippocampus." Brain Research, vol. 973, no. 2, May 2003, pp. 293–97. PubMed, https://doi.org/10.1016/s0006-8993(03)02508-3.

Deboer, Tom. "Sleep Homeostasis and the Circadian Clock: Do the Circadian Pacemaker and the Sleep Homeostat Influence Each Other's Functioning?" Neurobiology of Sleep and Circadian Rhythms, vol. 5, June 2018, pp. 68–77. ScienceDirect, https://doi.org/10.1016/j.nbscr.2018.02.003.

DeTure, Michael A., and Dennis W. Dickson. "The Neuropathological Diagnosis of Alzheimer's Disease." Molecular Neurodegeneration, vol. 14, no. 1, Aug. 2019, p. 32. PubMed, https://doi.org/10.1186/s13024-019-0333-5.

Doldur-Balli, Fusun, et al. "Synaptic Dysfunction Connects Autism Spectrum Disorder and Sleep Disturbances: A Perspective from Studies in Model Organisms." Sleep Medicine Reviews, vol. 62, Apr. 2022, p. 101595. PubMed, https://doi.org/10.1016/j.smrv.2022.101595.

Ferrarelli, Fabio. "Sleep Abnormalities in Schizophrenia: State of the Art and Next Steps." The American Journal of Psychiatry, vol. 178, no. 10, Oct. 2021, pp. 903–13. PubMed, https://doi.org/10.1176/appi.ajp.2020.20070968.

Hall, Jeremy, and Nicholas J. Bray. "Schizophrenia Genomics: Convergence on Synaptic Development, Adult Synaptic Plasticity, or Both?" Biological Psychiatry, vol. 91, no. 8, Apr. 2022, pp. 709–17. PubMed, https://doi.org/10.1016/j.biopsych.2021.10.018.

Hanlon, Erin C., et al. "Synaptic Potentiation and Sleep Need: Clues from Molecular and Electrophysiological Studies." Current Topics in Medicinal Chemistry, vol. 11, no. 19, 2011, pp. 2472–82. PubMed, https://doi.org/10.2174/156802611797470312.

Hirota, Tomoya, and Bryan H. King. “Autism Spectrum Disorder: A Review.” JAMA, vol. 329, no. 2, Jan. 2023, pp. 157–68. PubMed, https://doi.org/10.1001/jama.2022.23661.

Hodges, Holly, et al. “Autism Spectrum Disorder: Definition, Epidemiology, Causes, and Clinical Evaluation.” Translational Pediatrics, vol. 9, no. Suppl 1, Feb. 2020, pp. S55–65. PubMed Central, https://doi.org/10.21037/tp.2019.09.09.

Irwin, Michael R., and Michael V. Vitiello. “Implications of Sleep Disturbance and Inflammation for Alzheimer’s Disease Dementia.” The Lancet. Neurology, vol. 18, no. 3, Mar. 2019, pp. 296–306. PubMed, https://doi.org/10.1016/S1474-4422(18)30450-2.

Ishida, Hiroaki, et al. “Solution Structures of the SH3 Domains from Shank Scaffold Proteins and Their Interactions with Cav1.3 Calcium Channels.” FEBS Letters, vol. 592, no. 16, Aug. 2018, pp. 2786–97. PubMed, https://doi.org/10.1002/1873-3468.13209.

Maurus, Isabel, et al. “Exercise as an Add-on Treatment in Individuals with Schizophrenia: Results from a Large Multicenter Randomized Controlled Trial.” Psychiatry Research, vol. 328, Oct. 2023, p. 115480. PubMed, https://doi.org/10.1016/j.psychres.2023.115480.

McDonald-McGinn, Donna M., et al. “22q11.2 Deletion Syndrome.” GeneReviews®, edited by Margaret P. Adam et al., University of Washington, Seattle, 1993. PubMed, http://www.ncbi.nlm.nih.gov/books/NBK1523/.

Patel, Aakash K., et al. “Physiology, Sleep Stages.” StatPearls, StatPearls Publishing, 2026. PubMed, http://www.ncbi.nlm.nih.gov/books/NBK526132/.

Rahman, Tahir, and John Lauriello. “Schizophrenia: An Overview.” Focus [American Psychiatric Publishing], vol. 14, no. 3, July 2016, pp. 300–07. PubMed, https://doi.org/10.1176/appi.focus.20160006.

Tononi, Giulio, and Chiara Cirelli. “Sleep and Synaptic Homeostasis: A Hypothesis.” Brain Research Bulletin, vol. 62, no. 2, Dec. 2003, pp. 143–50. PubMed, https://doi.org/10.1016/j.brainresbull.2003.09.004.

Winship, Ian R., et al. “An Overview of Animal Models Related to Schizophrenia.” Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie, vol. 64, no. 1, Jan. 2019, pp. 5–17. PubMed Central, https://doi.org/10.1177/0706743718773728.

Wu, Chi-Hong, et al. “A Bidirectional Switch in the Shank3 Phosphorylation State Biases Synapses toward Up- or Downscaling.” eLife, edited by Lisa M. Monteggia et al., vol. 11, Apr. 2022, p. e74277. eLife, https://doi.org/10.7554/eLife.74277.

Yang, Albert C., and Shih-Jen Tsai. “New Targets for Schizophrenia Treatment beyond the Dopamine Hypothesis.” International Journal of Molecular Sciences, vol. 18, no. 8, Aug. 2017, p. 1689. PubMed Central, https://doi.org/10.3390/ijms18081689.

Urban Renewal and Framing: Euljiro between Traditional and Social Media

By Anthony Kim

Abstract

This study examines framing and counter-framing dynamics in the urban redevelopment controversy surrounding the Euljiro district in Seoul, South Korea. Framing theory and the concept of the hybrid media system give a basis for the study to investigate how governmental bodies, preservationist groups, and media actors constructed competing narratives between 2018 and 2025. A qualitative discourse analysis was conducted on a purposive sample of 25–30 publicly available texts. The analysis identifies four dominant frames: economic modernization, safety and infrastructure, cultural preservation, and social justice. Government communications and traditional media outlets consistently framed redevelopment from a top-down perspective, claiming it as essential for economic growth and urban renewal. Preservationist groups, by contrast, mobilized social media platforms to reframe Euljiro as a site of cultural heritage and community identity. Although activist counter-framing achieved substantial public visibility and temporarily influenced policy deliberation, it failed to produce durable policy change. The study concludes that digital media expands opportunities for participatory counter-public formation, but discursive influence alone is insufficient to overcome institutional authority and structural economic imperatives. These findings show the limitations of digital activism in achieving tangible political outcomes within urban governance contexts.

Keywords Urban redevelopment; framing theory; social media activism; public discourse; cultural preservation; policy outcomes

Introduction

Background and Context

Urban redevelopment produces contests amongst competing stakeholders who have divergent visions of spatial change. Governmental authorities take formal responsibility for urban planning, but existing residents and business owners try to assert historical claims to the spaces. In Seoul, South Korea, the density of urban land and persistent demand for development have intensified these conflicts across numerous districts. The Euljiro district is an instructive recent case. Traditionally a hub of artisanal workshops and hardware suppliers, Euljiro became the site of a protracted urban dispute when large-scale redevelopment plans were advanced by the Seoul Metropolitan Government. Conflicting visions of the district's future were immediately apparent. Existing tenants argued that renovation projects preserving the area's commercial heritage would be more beneficial for the city, but redevelopers consistently emphasized economic rationale. As Lee (2019) documents, many stakeholders called for "remodeling projects that could preserve the area's shops and history, rather than razing it to the ground." This position stands in contrast to the market-oriented logic that Shin (2009) identifies as

characteristic of Seoul's redevelopment projects, which he describes as "a market-oriented, profit-led renewal approach." The district's rapid industrial transformation also intersects with South Korea's post-war development trajectory. As Joo (2019) demonstrates, accelerated modernization was a big part of the country's economic recovery and expansion following the Korean War. This is the period colloquially known as the "Miracle on the Han River." Known informally as "Hipjiro" for its blend of industrial heritage and emerging cultural spaces, Euljiro became emblematic of tensions between preservation and profit, and between collective memory and modernization imperatives. As municipal authorities advanced redevelopment plans, preservationist actors saw social media as a way to contest official narratives and articulate Euljiro's importance as living urban heritage.

The proliferation of digital media has changed the dynamics of public discourse on urban and political issues. Social media platforms have enabled non-institutional actors to participate in political communication, so it gives them opportunities to stand up against or reframe arguments originating from governmental bodies and mainstream media outlets (Guran & Ozarslan, 2022). Research on the Euljiro case shows this. Seolha (2021) explains how bottom-up social media mobilization by young activists and artists produced an alternative discourse centered on "preserving the industry ecosystem and promoting urban diversity." It was direct opposition to the dominant governmental argument. However, the relationship between discursive participation and political outcomes is not straightforward. Social media platforms get credit for facilitating counter-publics and raising marginalized voices (López-Rabadán, 2021), but the translation of discursive influence into effective policy change has been more difficult. Available evidence appears to show that policy outcomes in the Euljiro case stayed mostly unchanged even with the sustained activist opposition.

Euljiro is part of South Korea's twentieth-century industrial history. After the devastation of the Korean War in the early 1950s, the district was part of the capital’s reconstruction. It became a place with a dense concentration of workshops that supplies a variety of resources, such as timber and other construction materials. Through the 1960s, the area evolved into a hub of small-scale manufacturing companies that gave services needed for the country’s industrialization. The industrial legacy stays visible today in what are called Lighting Alley and Tool Street. In recent decades, the district has undergone more changes as younger artists and designers made studios next to surviving workshops. It has kept the manufacturing tradition in a culturally hybrid form.

Problem Statement and Rationale

The Euljiro case raises important questions about the meaning and costs of urban modernization. Redevelopment gives possible economic and infrastructural benefits, but it is also simultaneously putting a threat to the cultural heritage and community. The case shows as well how civil society actors, lacking institutional access, may try to organize bottom-up resistance through social media platforms. There is a limitation of their success in changing policy outcomes. This limitation is evidenced by the Seoul Metropolitan Government's 2025 declaration

that "regulatory relaxation is necessary to revitalize the dilapidated downtown area" (Choi & Kim, 2025). It shows continued commitment to comprehensive redevelopment despite sustained activist opposition. The persistence of redevelopment plans as organized resistance happens helps make a research problem. Although preservationist groups achieved influence with their discourse and temporarily influenced the public debate, the counter-framing strategy failed to produce durable policy change. This gap between discursive influence and political outcomes needs more study as there is a lack of academic research in the area.

Significance and Purpose

The present study looks at what digital activism actually produces in terms of political outcomes. Existing research looks mostly at how social media enables marginalized voices to enter public discourse. Seolha (2021), for instance, documents how digital platforms helped the formation of a counter-public movement against governmental redevelopment ideas. Such scholarship usually focuses on the participatory possibilities given by digital media without systematically interrogating the conditions under which discursive gains make policy change. This study looks at that gap by analyzing the relationship between framing and policy outcomes in a specific urban governance context. It contributes to the literature on how communication strategies interact with institutional power.

Objectives

This study investigates the dynamics of framing and counter-framing in the Euljiro redevelopment controversy through three research questions:
(1) How did preservationist actors reframe governmental discourse on urban renewal?
(2) How did activist counter-framing influence traditional media coverage of the controversy?
(3) Which framing strategy proved more consequential in influencing policy outcomes?

Scope and Limitations

The study is bound temporally to the period 2018–2025, and focuses exclusively on the Euljiro redevelopment controversy. The dataset draws on discourse produced by three actor categories: Seoul Metropolitan Government communications, mainstream Korean news media, and preservationist social media content. The study does not attempt to quantify public opinion or assess the economic consequences of redevelopment. Instead, it focuses on the construction and contestation of narratives across these actor groups. Limitations include the relatively small and purposefully constructed sample of texts as well as the inherent interpretive dimension of qualitative discourse analysis. This is making some risk of researcher bias. Furthermore, as coding was conducted by a single researcher, the absence of intercoder reliability verification makes one another limitation.

Theoretical Framework

Framing Theory and Political Discourse

Erving Goffman's (1974) framing theory gives the conceptual foundation for this study. Goffman defines frames as interpretive schemata that organize individuals' perceptions and guide their understanding of events. Instead of encountering the world directly, individuals apply pre-existing frames that structure what they notice and how they assign meaning to what they observe.

Entman (1993) extends this by defining framing as the process through which communicators select certain aspects of a perceived reality and render them more salient in a communicative text. It promotes particular problem definitions, causal interpretations, moral evaluations, and remedies. Crucially, Entman emphasizes that framing is not merely descriptive but constitutive. That is to say, frames actively influence which aspects of reality are foregrounded and which are marginalized. It makes framing a political act in contested public debates.

Scheufele (1999) develops framing theory by giving a difference between frame-building and frame-setting. Frame-building refers to the processes through which frames are constructed by communicators. This gets influence from structural factors such as organizational pressures and ideological commitments. Frame-setting describes the subsequent process by which media frames influence audience interpretation. This difference applies to the Euljiro case. Both governmental actors and preservationist groups engaged in deliberate frame-building to change public understanding of the redevelopment controversy.

Iyengar (1996) introduces an additional dimension relevant to this analysis. This is the difference between episodic and thematic framing. Episodic frames present issues through concrete individual events or cases. They tend to encourage individualistic attributions of responsibility. Thematic frames situate issues in bigger structural and political contexts, so they encourage systemic attributions. Government communications in the Euljiro case, as the analysis below shows, largely used thematic frames linking redevelopment to macroeconomic needs and urban safety. The activist content, meanwhile, alternated strategically between episodic and thematic frames depending on rhetorical purpose.

Before social media, information flow in the public sphere was mediated by professionalized gatekeepers whose editorial authority structured what reached general audiences (Hove, 2024). Social media has greatly disrupted this arrangement. As Hove (2024) observes, drawing on Habermas, digital platforms have "radically altered the pattern of communication that has been dominant in the public sphere until now by empowering all potential users in principle to become independent and equally entitled authors." The consequence is a movement from passive media consumption toward active content production. It allows citizens to construct and circulate their own frames in competition with those of institutional actors.

Borah (2011) notes that this change makes framing a genuinely contested and multi-directional process than a top-down imposition by powerful actors. Pan and Kosicki (2001) add the layered and negotiated character of contemporary framing. The contested framing of urban uprisings documented by Welles and Jackson (2016) shows this dynamic. Mainstream

media's dominant framing was systematically countered through social media campaigns such as #myNYPD. This was when user-generated content reframed police conduct from a community perspective. The Euljiro case follows an analogous pattern. Preservationist actors deployed social media to contest the dominant institutional frame.

Digital Media, Public Sphere, and Networked Counterpublics

Social media's emergence as a site of political contestation must be understood within the context of what Chadwick (2017) terms the hybrid media system. Newer digital platforms interact with and partially change the logic of traditional media. It makes a hybrid environment in which both coexist and mutually influence one another. Chadwick characterizes this system as constituted by interactions among older and newer media logics. The newer logics, associated with decentralized and participatory technologies, have introduced new forms of influence over public discourse. The older logics, associated with professionally produced and editorially controlled news media, keep agenda-setting and legitimating power.

The hybrid media system framework is directly applicable to the Euljiro case. Preservationist groups operating through social media platforms did not simply bypass traditional media. Instead, their campaigns occasionally achieved sufficient visibility to penetrate mainstream news coverage. It showed the permeability between old and new media logics that Chadwick describes. At the same time, the asymmetry in institutional resources between governmental actors and activist groups influenced the relative reach and durability of their respective frames. It is an asymmetry that, as the analysis demonstrates, helps account for the ultimate failure of activist counter-framing to produce policy change.

Methodology

Research Design

This study uses a qualitative research design focused on discourse analysis. This methodological approach is suitable to examine how competing actors construct and contest meaning in public debates. It enables systematic analysis of the linguistic and rhetorical strategies through which frames are built and communicated. The study is observational and comparative in orientation. It examines how the Euljiro redevelopment was framed by three distinct actor categories and traces the interactions and divergences among these framings over the study period.

Research Sample

The analytical corpus comprises 25 publicly available texts produced by three actor categories. The first is official communications and press releases issued by the Seoul Metropolitan Government. The second is news articles published by major South Korean media outlets, including The Korea Times, JoongAng Ilbo, Hankyung, and Chosun Daily. The third is social media content produced by the Cheonggyecheon Euljiro Preservation Association (CEPA)

and affiliated activist accounts on Twitter and Instagram. A complete list of analyzed sources, organized by actor category and frame coded, is provided in Appendix A.

Purposive sampling was used to ensure that selected texts explicitly engaged with the Euljiro redevelopment controversy and fell within the study period of 2018–2025. This temporal boundary was selected to capture the principal phases of the redevelopment debate, from the initial mobilization of preservationist opposition through to the reassertion of comprehensive redevelopment plans under Mayor Oh Se-hoon. Texts were excluded if they did not directly address redevelopment plans, public responses, or the framing of Euljiro's future. Duplicate or substantively redundant materials were similarly excluded.

Data Collection

Government documents and official policy statements were retrieved from the Seoul Metropolitan Government's official website and affiliated policy repositories. Traditional media articles were located through online news archives using keyword searches including "Euljiro redevelopment" (을지로 재개발), "Seoul urban renewal," and related terms. Social media data were drawn from publicly accessible posts on CEPA's official accounts and associated activist content on Twitter and Instagram. All materials were collected in their original Korean-language form where applicable and translated by the researcher for analysis.

Variables and Measurements

The primary analytical variable is "framing," operationalized as the selection and salience of particular problem definitions, causal attributions, and proposed solutions within each text. Following the theoretical framework outlined above, four framing categories were applied as the analytical coding scheme:

1. Economic modernization: frames emphasizing redevelopment as necessary for economic growth, increased floor-area ratios, office and residential development, and urban competitiveness.
2. Safety and infrastructure: frames foregrounding the physical deterioration of existing structures, fire risk, and the need for infrastructure renewal.
3. Cultural preservation: frames centering the historical and heritage significance of Euljiro's industrial landscape and the cultural identity of existing communities.
4. Social justice: frames emphasizing the displacement of existing tenants and workers, the illegitimacy of demolition, and the rights of established communities.

Procedure

Analysis proceeded in four stages. First, all collected texts were organized by actor category and date of publication. Second, each text was read in full to identify recurring themes and rhetorical patterns. Third, texts were systematically coded according to the four framing categories using a keyword and thematic coding approach: each text was examined for the presence, frequency, and salience of language and arguments associated with each frame. Where

a text employed multiple frames, the dominant framing was identified on the basis of emphasis and rhetorical priority. Fourth, coded texts were compared across actor categories and across the study period to identify patterns of convergence, divergence, and interaction among competing frames.

Ethical Considerations

All data analyzed in this study came from publicly available sources and did not involve the collection of private or personally identifiable information. As the study did not involve human subjects research, formal ethical review and informed consent procedures were not required. Analytical interpretations were developed with attention to accuracy and representativeness, and care was taken to ensure that the framing categories applied did not distort the meaning or intent of source materials.

Results

Government and Media Framing

Government communications and traditional media coverage consistently advanced frames linking redevelopment to economic necessity and urban modernization. This matches what Entman (1993) would identify as a dominant institutional frame: one that defined the problem (physical deterioration and economic stagnation), assigned causation (outdated infrastructure and insufficient development), and proposed a remedy (comprehensive redevelopment). This framing alignment between governmental actors and mainstream media reflects the structural dynamic that Scheufele (1999) describes as frame-building influenced by organizational and ideological proximity to institutional power.

Following the election of Mayor Oh Se-hoon to the Seoul Metropolitan Government in 2022, redevelopment plans for Euljiro were reactivated. Oh publicly criticized the preservation-oriented policies of the preceding administration, characterizing what he described as an "idealistic approach to architecture, with a strong emphasis on preservation" as inadequate for the city's development needs (Na, 2021). His administration's plan. This was to replace existing low-rise structures with high-rise office complexes at increased floor-area ratios and incorporating green public spaces, and this was reported in JoongAng Ilbo as a major strategic initiative (Kim, 2022). The framing of this plan as modernization.

Traditional media coverage reinforced this governmental frame. News articles described Euljiro's existing structures as "old and deteriorated buildings" (Ahn, 2021) and characterized proposed office development as "a much-needed boost for the office market" (Kim, 2024). Safety concerns were prominently featured, with reports framing the district as "vulnerable to fire and safety hazards, which have led to ongoing calls for urban environmental improvement" (Ahn, 2021). This constitutes what Iyengar (1996) terms thematic framing. It is situating the redevelopment inside a bigger structural narrative of urban necessity. That was instead of presenting it as a series of discrete events affecting particular individuals or businesses. Some

outlets adopted explicitly transformative language, projecting that the completed development would produce "an urban hub integrating green spaces, business, and culture" (Lee, 2025). Critical perspectives on the redevelopment plans were largely absent from mainstream coverage. It indicates a close alignment between governmental frame-building and traditional media frame-setting. This is precisely the top-down dynamic that Scheufele (1999) identifies as characteristic of institutionally dominant frames.

Activist Counter-Framing

In response to the governmental and media framing of comprehensive redevelopment as inevitable and beneficial, preservationist actors organized a continuous counter-framing campaign focused on cultural heritage and social justice. Operating primarily through Twitter and Instagram, the Cheonggyecheon Euljiro Preservation Association (CEPA) and affiliated groups tried to change Euljiro not as a site of deterioration requiring renewal, but as a living repository of urban heritage and community identity. This strategy exemplifies the counter-public formation that López-Rabadán (2021) identifies as a key function of social media activism. It is the construction of alternative discursive spaces in which marginalized perspectives can be articulated and amplified.

CEPA's communications used the language of rights and legitimacy to frame the redevelopment as an injustice more than a public benefit. The association's Twitter content explicitly invoked the concept of rights "unjustly taken away." It positions displacement not as an unfortunate consequence of urban development but as a violation of established community claims (Cheongyecheon, 2022a). This framing strategy aligns with what Entman (1993) describes as the moral evaluation dimension of framing: attributing responsibility and assigning normative valence to events in ways that mobilize particular responses.

The association also deployed episodic framing in Iyengar's (1996) terms by documenting specific sites and communities at risk of demolition. The Instagram post "Against Dystopia" (Cheongyecheon, 2023a) invited audiences to ask "Why has it become considered legal in our society to erase urban history, employment, and local culinary landmarks for the sake of apartment development?" It grounds the cultural preservation argument in the concrete particularity of threatened spaces. The post title itself, invoking dystopia, performed a deliberate reframing of the governmental modernization narrative. This is what officials presented as progress was recorded as destruction. A further post, "Seoul Metropolitan Government: Do not rob us of our present and our future" (Cheongyecheon, 2023b), positioned Euljiro as a temporal link between the city's past and its future, directly contesting the governmental characterization of existing structures as obsolete.

The involvement of political figures amplified the visibility of preservationist counter-framing. Politician Sim Sangjeong's documented participation in CEPA protests, describing the site as "a future cultural heritage site, built from countless stories and the footsteps of citizens" (Cheongyecheon, 2022b). It introduced institutional legitimacy into what was otherwise a grassroots campaign. The endorsement functioned within the hybrid media system

(Chadwick, 2017) as a point of entry into mainstream coverage. An activist content associated with a public figure carried greater potential for penetrating traditional media channels than organically generated social media content alone.

Interaction of Frames

The Euljiro case shows the contested and dynamic character of framing that Borah (2011) and Pan and Kosicki (2001) describe. They say that frames do not exist in isolation but interact and partially modify one another over time. The governmental and traditional media frames promoting redevelopment as modernization were neither unchallenged nor static. The activist counter-frames generated sufficient public visibility to temporarily disrupt the dominant narrative and introduce delays into the redevelopment timeline.

CEPA's campaign with related preservationist organizations and public protests exerted sufficient pressure to influence the policy stance of Mayor Park Won-soon's administration prior to 2022. This was during which some preservation-oriented modifications to redevelopment plans were implemented. This temporary policy responsiveness matches the permeability between social media and traditional media logics within Chadwick's (2017) hybrid media system. Its activist frames achieved enough mainstream visibility to register on the institutional agenda.

However, the interaction between competing frames did not produce symmetrical influence. The governmental and traditional media framing retained structural advantages. It is institutional authority, consistent access to mainstream channels, and alignment with dominant economic interests that activist counter-framing could not beat. This asymmetry is evident in the resumption of comprehensive redevelopment plans following Mayor Oh's election in 2022, which effectively reversed the partial accommodations made under the previous administration.

Discussion

Why Activist Framing Failed to Produce Policy Change

The failure of activist counter-framing to produce durable policy change in the Euljiro case requires explanation beyond the discursive dynamics examined above. Two structural factors prove particularly significant.

The first is the role of political contingency. The partial responsiveness of the Park administration to preservationist framing was not a stable institutional outcome but a function of that administration's specific political orientation. When Mayor Oh succeeded Park in 2022, the political conditions that had rendered activist framing partially effective were removed. Oh's characterization of his predecessor's approach as "horrifying" and "tragic" (Kim, 2022) signalled an explicit rejection of the framing concessions the previous administration had made. This pattern suggests that discursive influence, while capable of affecting policy under favourable political conditions, remains contingent on institutional receptivity that cannot be guaranteed. A comparable dynamic is observable in Greenpeace's anti-whaling campaign in Norway, where sustained international counter-framing failed to alter policy outcomes when governmental actors

successfully reframed the issue in terms of national sovereignty and cultural identity (Riese, 2014). It shows that institutional resistance to discursive pressure is not unique to the Korean context.

The second structural factor is the economic logic underpinning the redevelopment. Seoul faces a well-documented housing shortage, and large-scale redevelopment projects generate substantial economic benefits for governmental actors and private developers. The scale of the proposed Euljiro redevelopment meant that the economic stakes were sufficiently high to sustain institutional commitment even in the face of organized discursive opposition. This supports Shin's (2009) characterization of Seoul's redevelopment paradigm as mainly "market-oriented" and "profit-led". Cultural preservation arguments, however rhetorically compelling, operated against a structural baseline in which economic imperatives carried greater institutional weight.

These findings extend Entman's (1993) conceptualization of framing as a contested political process by specifying the conditions under which counter-frames fail to achieve policy traction. Discursive influence appears sufficient to introduce issues onto the public agenda and to generate temporary policy responsiveness, but insufficient to sustain policy change when it conflicts with entrenched institutional authority and structural economic priorities.

Digital Activism and Its Limits

The findings also contribute to scholarly debate on the political efficacy of digital activism. Research on social media and political mobilization has tended to emphasize the participatory affordances of digital platforms. Their capacity to amplify marginalized voices, facilitate counter-public formation, and disrupt institutional information monopolies (López-Rabadán, 2021; Seolha, 2021). The Euljiro case confirms these affordances at the level of public discourse: preservationist groups achieved a degree of visibility and agenda influence that would have been considerably more difficult to attain through pre-digital organizing alone.

However, the case also reveals the limits of what visibility alone can accomplish. The translation of discursive influence into policy change requires not only the capacity to frame issues compellingly but also access to the institutional channels through which policy decisions are made. Activist groups operating through social media occupy a structurally different position within the hybrid media system than governmental actors with direct institutional authority. Within Chadwick's (2017) framework, the persistence of older media logics and, more broadly, the persistence of institutional power structures means that the expansion of participatory communication does not straightforwardly translate into expanded political influence. Meaningful policy impact, the Euljiro case suggests, requires institutional support alongside discursive capacity.

Conclusion

This study has examined the dynamics of framing and counter-framing in the Euljiro redevelopment controversy, demonstrating that discursive contestation, while capable of influencing public debate, does not guarantee commensurate policy outcomes. Drawing on

framing theory (Goffman, 1974; Entman, 1993; Scheufele, 1999) and the hybrid media system framework (Chadwick, 2017), the analysis has traced how governmental actors and traditional media constructed a dominant frame of economic modernization and urban necessity, and how preservationist groups deployed social media platforms to contest this frame through counter-narratives of cultural heritage and social justice.

The principal finding is that activist counter-framing achieved substantial public visibility and temporarily influenced policy deliberation, but ultimately failed to produce durable policy change. This outcome reflects two intersecting constraints: the political contingency of institutional receptivity to discursive pressure, and the structural weight of economic imperatives that the governmental frame successfully mobilized. These findings align with Entman's (1993) conception of framing as an asymmetrically contested process in which the resources of institutional actors confer significant advantages over non-institutional challengers.

For the broader literature on digital activism, the Euljiro case suggests that the participatory affordances of social media, while genuinely expanding opportunities for counter-public formation, do not in themselves resolve the structural inequalities that shape policy outcomes. Digital visibility is a necessary but insufficient condition for political influence. Future research might productively examine the conditions under which digital counter-framing succeeds in producing policy change whether through sustained institutional alliances, electoral dynamics, or the accumulation of discursive pressure over longer time horizons in order to develop a more differentiated account of when and how communication interacts with political power in urban governance contexts.

Acknowledgements

The author wishes to thank Mr. Dickey for his guidance and encouragement throughout this project. As the author's AP Seminar and AP Research teacher, he contributed substantially to the development of the research skills and analytical approach reflected in this study.

Works Cited

Ahn, S. M. (2021, September 16). Redevelopment of Euljiro Tool Street in Seoul officially kicks off for the first time in 11 years. Hankyung. https://www.hankyung.com/article/202109160336i

Borah, P. (2011). Conceptual issues in framing theory: A systematic examination of a decade's literature. Journal of Communication, 61(2), 246–263. https://doi.org/10.1111/j.1460-2466.2011.01539.x

Chadwick, A. (2017). The hybrid media system: Politics and power. Oxford University Press.

Cheongyecheon [CEPA Twitter]. (2022a, September 29). When will we be able to reclaim the rights that were unjustly taken from us? [Tweet]. Twitter. https://x.com/cheongyecheon/status/1575342485060009984

Cheongyecheon [CEPA Twitter]. (2022b, September 26). I visited the site of Eulji OB Bear last weekend [Tweet]. Twitter. https://x.com/sangjungsim/status/1574361901039230979

Cheongyecheon [CEPA Instagram]. (2023a, August 21). Against Dystopia [Instagram post]. https://www.instagram.com/p/CwNergYpSym/

Cheongyecheon [CEPA Instagram]. (2023b, June 23). Seoul City, Do Not Rob Us of Our Present and Our Future [Instagram post]. https://www.instagram.com/p/Ct1PIMkrXn7/

Choi, Y., & Kim, Y. (2025, November 4). Seoul plans 142m buildings near Jongmyo Shrine amid heritage concerns. The Chosun Daily. https://www.chosun.com/english/national-en/2025/11/04/QBPLPW4LI5FYLFX46Q7RNGMV54

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. Journal of Communication, 43(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Goffman, E. (1974). Frame analysis: An essay on the organization of experience. Northeastern University Press.

Guran, M., & Ozarslan, H. (2022). Framing theory in the age of social media. The Journal of Selçuk University Social Sciences Institute, 48, 446–457. https://doi.org/10.52642/susbed.1142562

Hove, T. (2024). A new structural transformation of the public sphere and deliberative politics. Asian Communication Research, 21(1). https://doi.org/10.20879/acr.2024.21.017

Iyengar, S. (1996). Framing responsibility for political issues. The Annals of the American Academy of Political and Social Science, 546(1). https://doi.org/10.1177/0002716296546001001

Joo, Y. (2019). Megacity Seoul: Urbanization and the development of modern South Korea. Routledge.

Kim, O. (2022, April 24). Mayor Oh Se-hoon announces 'Seoul Green Urban Space Re-creation Strategy.' JoongAng Ilbo. https://www.joongang.co.kr/article/25065871

Kim, S. S. (2024, January 3). Euljiro 3-ga District 12 redevelopment to be completed the year after next. Marketin. https://marketin.edaily.co.kr/News/Read?newsId=03719526638754112

Lee, H. R. (2025, March 10). Seun District redevelopment speeds up. Donghaeng Media. https://www.sidae.com/article/2025031009072797918

Lee, S. (2019, January 9). Euljiro's last stand: Redevelopment rips through historic manufacturing hub. The Korea Times. https://www.koreatimes.co.kr/southkorea/society/20190109/euljiros-last-stand-redevelopment-rips-through-historic-manufacturing-hub

López-Rabadán, P. (2021). Framing studies evolution in the social media era. Social Sciences, 11(1), 1–19. https://doi.org/10.3390/socsci11010009

Na, Y. E. (2021, November 18). Oh Se-hoon targets 'Park Won-soon projects.' Asia Economy. https://www.asiae.co.kr/article/2021111820593752609

Pan, Z., & Kosicki, G. M. (2001). Framing as a strategic action in public deliberation. Routledge.

Riese, J. (2014). Media and revolt: Strategies and performances from the 1960s to the present. Berghahn.

Scheufele, D. A. (1999). Framing as a theory of media effects. Journal of Communication, 49(1), 103–122. https://doi.org/10.1111/j.1460-2466.1999.tb02784.x

Seolha, L. (2021). The shape of discourse in urban movements through the lens of social media [Doctoral dissertation, Georgia Institute of Technology]. ProQuest. https://www.proquest.com/openview/f128731c9d006eca996b27c5b87ccdf7/1

Shin, H. B. (2009). Property-based redevelopment and gentrification: The case of Seoul, South Korea. Geoforum, 40(5), 906–917. https://doi.org/10.1016/j.geoforum.2009.06.009

Son, J. H. (2025, September 5). Euljiro to be transformed into a multi-purpose space. Homeknock. https://www.homeknock.co.kr/web-front/community/news/15272/

Welles, F. B., & Jackson, J. S. (2016). The battle for #Baltimore: Networked counterpublics and the contested framing of urban unrest. Routledge.

Appendix A: Analytical Corpus — Sources by Actor Category

Table 1 presents the complete list of texts analysed in this study, organized by actor category, source type, and dominant frame coded.

Actor category	Source type	Source	Year	Frame coded
Government	Press statement / news report	Na, Y. E. — Oh Se-hoon targets 'Park Won-soon projects' (Asia Economy)	2021	Economic modernization
Government	Policy announcement / news report	Kim, O. — Mayor Oh announces Seoul Green Urban Space Re-creation Strategy (JoongAng Ilbo)	2022	Economic modernization
Government	News report	Ahn, S. M. — Redevelopment of Euljiro Tool Street officially kicks off (Hankyung)	2021	Safety & infrastructure
Government	News report	Kim, S. S. — Euljiro 3-ga District 12 redevelopment to be completed (Marketin)	2024	Economic modernization
Government	News report	Son, J. H. — Euljiro to be transformed into multi-purpose space (Homeknock)	2025	Economic modernization

Government	News report	Lee, H. R. — Seun District redevelopment speeds up (Donghaeng Media)	2025	Economic modernization
Government	News report	Choi, Y. & Kim, Y. — Seoul plans 142m buildings near Jongmyo Shrine (Chosun Daily)	2025	Safety & infrastructure
Government	News report	Lee, S. — Euljiro's last stand: Redevelopment rips through historic manufacturing hub (Korea Times)	2019	Economic modernization
Government	Background / context	Shin, H. B. — Property-based redevelopment and gentrification, Seoul (Geoforum)	2009	Economic modernization
Government	Background / context	Joo, Y. — Megacity Seoul: Urbanization and development of modern South Korea	2019	Economic modernization
Preservationist (CEPA)	Twitter post	Cheongyecheon — 'rights unjustly taken away' [tweet, Sep 29]	2022	Social justice
Preservationist (CEPA)	Twitter post	Cheongyecheon / Sim Sangjeong — 'beautiful resistance… future cultural heritage site' [tweet, Sep 26]	2022	Cultural preservation
Preservationist (CEPA)	Instagram post	Cheongyecheon — 'Against Dystopia' workshop [Aug 21]	2023	Cultural preservation / Social justice
Preservationist (CEPA)	Instagram post	Cheongyecheon — 'Seoul City, Do Not Rob Us of Our Present and Our Future' [Jun 23]	2023	Cultural preservation / Social justice
Preservationist (CEPA)	Academic dissertation	Seolha, L. — Discourse in urban movements through social media, anti-redevelopment movement in South Korea (Georgia Tech)	2021	Cultural preservation
Traditional media	News article	Lee, S. — Euljiro's last stand (Korea Times) [also coded as govt-aligned framing]	2019	Economic modernization
Traditional media	News article	Ahn, S. M. — Tool Street redevelopment kicks off (Hankyung)	2021	Safety & infrastructure
Traditional media	News article	Kim, O. — Mayor Oh announces strategy (JoongAng Ilbo)	2022	Economic modernization
Traditional media	News article	Kim, S. S. — Euljiro 3-ga to be completed (Marketin/Edaily)	2024	Economic modernization
Traditional media	News article	Son, J. H. — Multi-purpose space transformation (Homeknock)	2025	Economic modernization
Traditional media	News article	Choi, Y. & Kim, Y. — 142m buildings near Jongmyo (Chosun Daily)	2025	Safety & infrastructure

Note. Where a source was coded as employing two frames with roughly equal salience, both are listed. Sources marked with an asterisk were also used as background/contextual references in the Introduction.

Multi-Model Machine Learning Identifies MAPT, CHEK1, AURKA as Breast Cancer Prognostic Markers By Jasmine Chan

Abstract

Breast cancer is the most common cancer among women globally, and the second leading cause of cancer death. While early detection improves survival rates, personalizing treatment based on genomic biomarkers could further increase survival duration. This project applies a multi-model machine learning approach to identify key gene biomarkers correlating with breast cancer survival duration, supporting future research to personalize treatment. I hypothesized that different models contribute complementary findings: linear regression captures proportional gene-survival relationships, random forest reveals non-linear interactions, and neural network could perform deep analysis on larger datasets but was anticipated to underperform on the small METABRIC dataset (n=1904 patients). Using the METABRIC dataset, I preprocessed gene expression and clinical data by imputing missing values and encoding categorical variables. I trained three models—ElasticNet linear regression, random forest, and PyTorch neural network—using 70/15/15 train/validation/test split to predict overall survival time. Random forest achieved the best test performance (r^2=0.147), followed by linear regression (r^2=0.138), while the neural network underperformed (r^2=0.052) as anticipated. The models identified MAPT, CHEK1, and AURKA as top gene biomarkers strongly associated with survival duration, consistent with published cancer genomics research. The analysis focused on genomic factors, yet survival duration also depends on age, comorbidities, lifestyle, and unrelated causes of death, introducing variance. The consistent under-prediction of survival durations aligns with this limitation and validates the integrity of our genomic-focused approach. These findings demonstrate that complementary models can uncover actionable genomic biomarkers, offering a pathway toward personalized breast cancer treatment and improved prognosis outcomes.

Introduction

Breast cancer is the most frequently diagnosed cancer among women worldwide, with over 321,910 new cases expected in the United States in 2026 alone (Siegel et al.). It is the second leading cause of cancer death among women. While overall five-year survival rates reach 90% when detected early, outcomes vary dramatically based on cancer subtype, stage, and individual patient characteristics. This variation underscores the urgent need for precision medicine approaches that can improve prognostic assessment and support more individualized treatment planning.

Existing research in cancer genomics has demonstrated that gene expression profiles can reveal molecular subtypes of breast cancer with distinct clinical outcomes. The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) characterized the genomic and transcriptomic landscape of over 2,000 breast tumors, identifying novel subgroups with different survival patterns (METABRIC Group et al.; Pereira et al.). However, most machine

learning applications in this field focus on predicting binary survival outcomes (alive vs. deceased) rather than continuous survival duration, which is far more clinically valuable for treatment planning.

This project investigates that gap by using three complementary machine learning models to predict survival time and to discover which specific gene biomarkers most strongly influence breast cancer survival duration, enabling physicians to customize treatment plans for individual patients. I hypothesized that different models would contribute multi-dimensional findings. Linear regression identifies genes acting proportionally on survival duration, while random forest captures non-linear and interaction-based relationships for comprehensive coverage. A neural network could perform deep analysis on very large datasets to discover additional differential factors among large populations. However, because the METABRIC dataset contains fewer than 2,000 records, the neural network was anticipated to underperform relative to the other models due to insufficient data volume for deep learning.

Materials

Table 1. List of materials used and their purposes.

Category	Component / Tool	Specification / Purpose
HARDWARE	**Microsoft Surface Pro 9**	Windows 11 laptop; used for coding, tool access, and results review.
	Google Colab	Cloud environment; NVIDIA T4 GPU for model training and evaluation.
DATASET	**METABRIC**	1,904 patients; clinical features and 489 gene expression features (Kaggle).
SOFTWARE	**Python 3.12.13**	Main programming language for data analysis and ML workflows.
	PyTorch (torch) 2.10.0+cu128	Deep learning framework for building and training neural networks.
	scikit-learn 1.6.1	Library for regression, random forest, tuning, and evaluation.
	pandas 2.2.2 / NumPy 2.0.2	Data cleaning, organization, and numerical operations.
	SciPy 1.16.3	Statistical calculations including confidence intervals.
	matplotlib 3.10.0	Data visualization and graph generation.
	graphviz 0.21	Decision tree surrogate visualization.

Dataset

The dataset used in this study was the METABRIC breast cancer dataset, a large public resource for studying breast cancer biology and patient outcomes. As summarized in Table 1, the version used for this project was a Kaggle-distributed dataset (https://www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric) containing data for approximately 1,900 patients, including clinical variables, gene expression measurements, as well as survival outcomes which were used as prediction targets. The broader METABRIC cohort was originally developed from primary breast tumor samples with long-term clinical follow-up and was designed to support integrated molecular and clinical analysis of breast cancer. In the original METABRIC resource, gene expression profiling was performed using Illumina HT-12 arrays, and the study's initial release included discovery and validation cohorts of 997 and 995 primary breast tumors, respectively. Later METABRIC work also expanded the molecular characterization of the cohort through targeted sequencing of 173 recurrently mutated breast cancer genes, helping establish METABRIC as a widely used benchmark dataset for prognostic modeling and biomarker discovery in breast cancer research.

Hardware/Software

The computational resources used in this study are summarized in Table 1. A Microsoft Surface Pro 9 laptop computer (Microsoft Corp., Windows 11) was used for code development, access to online tools, organization of files, and review of model outputs. Data preprocessing, machine learning model training, and evaluation were carried out primarily in the Google Colab cloud notebook environment (Google LLC), which provided access to remote computing resources, including a remote NVIDIA T4 GPU for accelerated model training and testing. Python 3 was used as the main programming language for implementing the analysis workflow. The primary software libraries included PyTorch (torch) for neural network development, pandas for loading and organizing tabular data, NumPy for numerical computation, SciPy for statistical analysis, matplotlib for graph generation and visualization, scikit-learn for machine learning model building and evaluation, and graphviz for decision tree surrogate visualization.

Methods

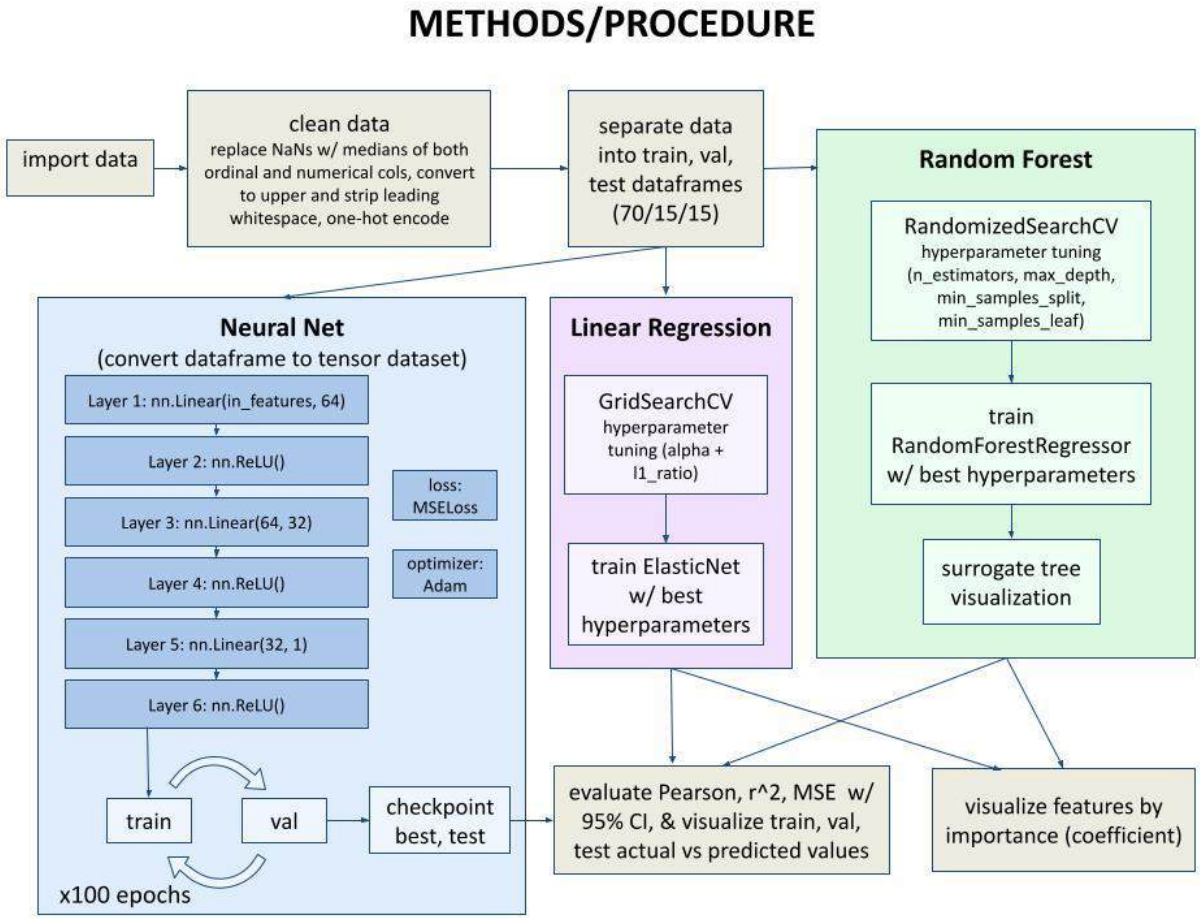


Figure 1. Block diagram describing the procedure used to process the data.

The overall workflow used in this study is summarized in Figure 1. First, the METABRIC breast cancer dataset was obtained from Kaggle and the raw CSV file was imported into Python via pandas for preprocessing and analysis. Data cleaning was then performed to prepare the dataset for machine learning. A patient ID mapping dictionary was created for row identification, after which the patient id column was removed from the model inputs so that it would not influence prediction. The target variable was defined as time until death in years, calculated by dividing overall survival months by 12. Missing values (NaNs) were assessed and imputed using the median for numerical columns and the mode for categorical columns. To reduce noise and prevent data leakage, outcome-related or otherwise non-informative columns were removed, including overall survival, overall survival months, cohort number, reason for death, and the nottingham prognostic index, a tool that separates patients into clinical groups based on their prognosis (Fong et al.). In addition, categorical values were standardized where needed by cleaning formatting inconsistencies such as data type, capitalization, and extra whitespace. Categorical predictors were one-hot encoded for model use. After preprocessing, the dataset was split into training, validation, and test sets using a 70/15/15 split with a fixed random

seed of 42, as shown in Figure 1. The test set was reserved for final model evaluation only, while the training and validation sets were used during model development and tuning.

Three machine learning approaches were trained and compared. The first model was an ElasticNet linear regression model, which combined L1 and L2 regularization. Its hyperparameters, alpha and l1 ratio, were optimized using GridSearchCV. The second model was a Random Forest Regressor, for which the hyperparameters n estimators, max depth, min samples split, and min samples leaf were tuned using RandomizedSearchCV. The third model was a PyTorch neural network consisting of six layers with ReLU activation functions, mean squared error (MSE) loss, and the Adam optimizer. The neural network was trained for 100 epochs, with validation performance monitored during training and checkpointing used to save the best-performing model before final testing. As illustrated in Figure 1, all three models were ultimately evaluated on the same held-out test set.

Statistical Analysis

Following training, model performance was compared using Pearson correlation coefficient (r), coefficient of determination (r^2), and mean squared error (MSE), each reported with 95% confidence intervals on the held-out test set.

Pearson's r was used to measure the strength and direction of the linear relationship between the predicted survival times and the actual survival times. A value closer to 1 indicates that the predicted and actual values increase together more consistently, while a value closer to 0 indicates little linear relationship. The r^2 value was used to measure how much of the variation in survival time was explained by the model. Higher r^2 values indicate that the model accounts for more of the observed variability in the outcome, whereas values closer to 0 indicate weaker explanatory power. The MSE was used to quantify prediction error by averaging the squared differences between predicted and actual survival times. Lower MSE values indicate more accurate predictions. To quantify uncertainty regarding model performance, 95% confidence intervals were also calculated for the test-set performance metrics in order to ensure statistical significance.

In addition to predictive performance, model interpretability was examined by extracting feature importance rankings from each method. For the Random Forest model, a surrogate decision tree visualization was also generated to provide a more interpretable representation of model behavior.

Results

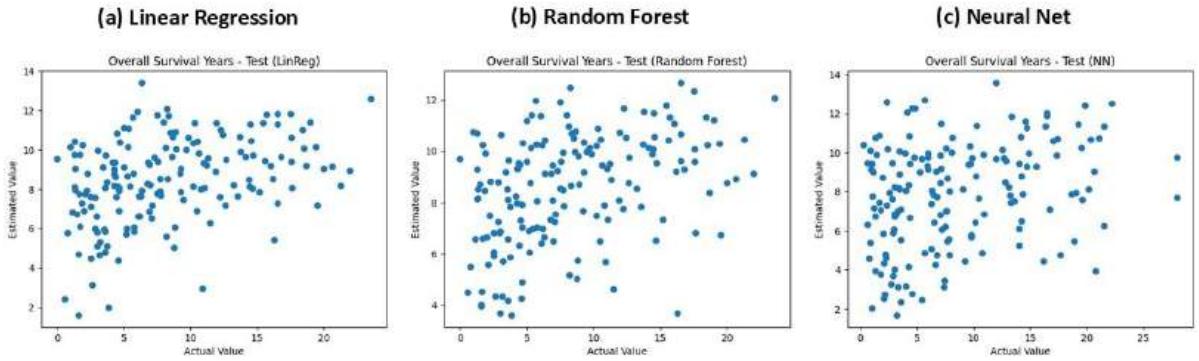


Figure 2. Comparison of actual versus estimated actual survival years for models run on the test set. (a) Linear regression comparison. (b) Random forest comparison. (c) Neural network comparison.

Table 2. Evaluation metrics for models run on the test set with a 95% confidence interval. Displays the Pearson r, r^2, and mean squared error (MSE) for the (a) linear regression, (b) random forest, and (c) neural net models.

Metric	(a) Linear Regression	(b) Random Forest	(c) Neural Network
Test Pearson r	0.379 [0.262, 0.492]	0.392 [0.258, 0.516]	0.288 [0.148, 0.416]
Test R^2	0.138 [0.043, 0.227]	0.147 [0.039, 0.246]	0.052 [-0.061, 0.161]
Test MSE	26.187 [21.222, 31.540]	25.910 [20.815, 31.362]	37.414 [29.191, 46.972]

Each model highlighted different types of relationships between input features and breast cancer survival duration. The linear regression model identified features with approximately linear associations with survival time, with lymph nodes examined positive emerging as the strongest linear predictor. In contrast, the Random Forest model captured non-linear relationships between gene-expression features and survival duration, identifying MAPT (Microtubule-Associated Protein Tau), CHEK1 (Checkpoint Kinase 1), and AURKA (Aurora Kinase A) among the most important predictors. The neural network model was also trained and evaluated, but it did not perform as well as the linear regression and Random Forest models based on the evaluation metrics used in this study.

Among the gene-expression features analyzed, MAPT, CHEK1, and AURKA emerged as the most notable biomarkers associated with breast cancer survival duration. MAPT ranked first in Random Forest feature importance, indicating a strong contribution to the non-linear model.

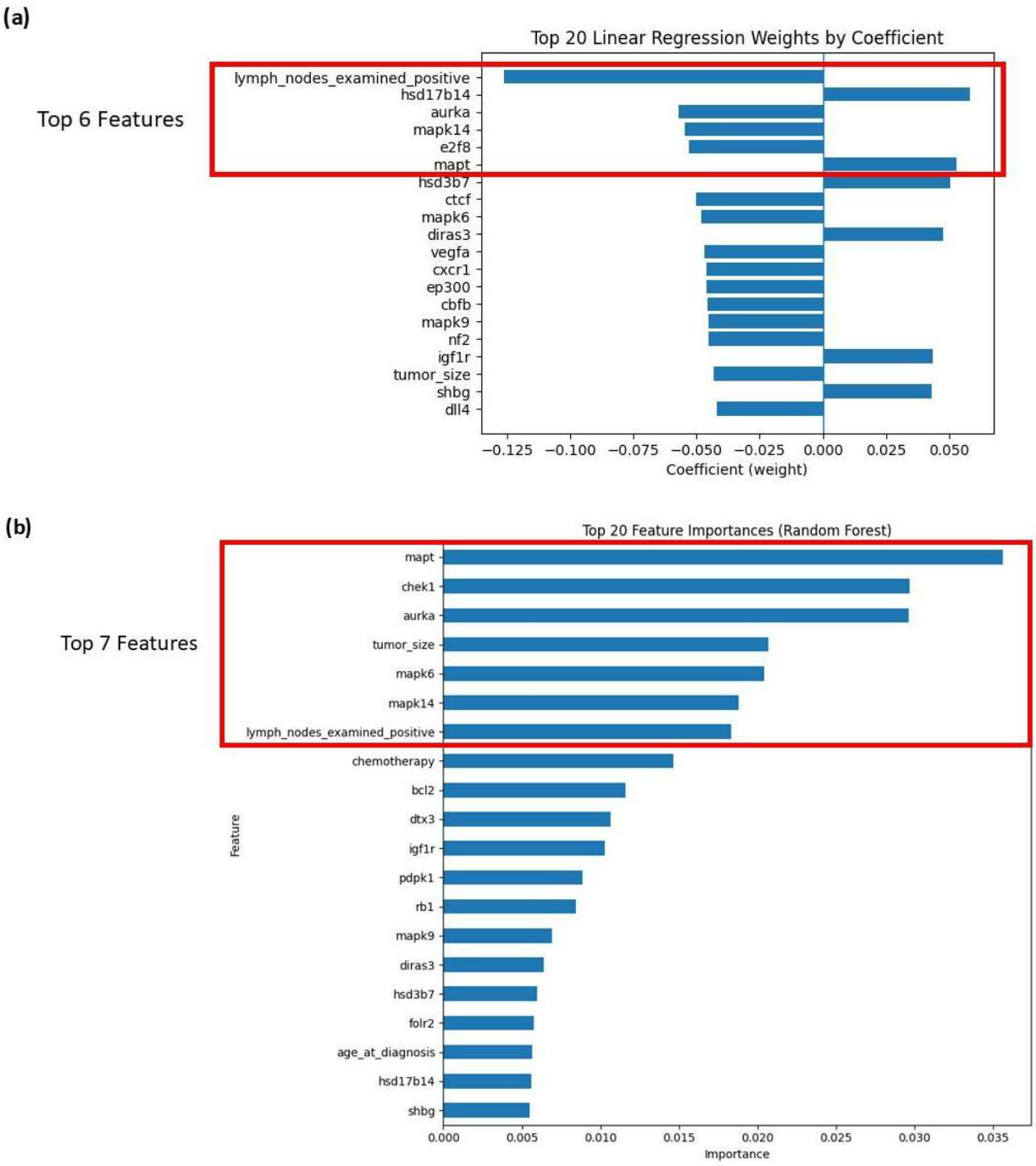


Figure 3. Top feature importances. (a) Top 20 linear regression feature weights, ranked by magnitude regardless of positive or negative direction. Bar direction displays the sign of each respective coefficient, with positive coefficients indicating high correlation with survival time and negative coefficients indicating low survival time. (b) Top 20 random forest feature weights, ranked by magnitude. Unlike the linear regression features, random forest features display only importance, with higher values indicating more contribution to prediction and lower values indicating less, without indicating positive or negative directionality.

CHEK1 appeared as the root node in the surrogate decision tree, suggesting that it served as the primary splitting feature in the interpretable approximation of the Random Forest model. AURKA ranked within the top three features across both the linear regression and Random Forest models, demonstrating consistency across methods. Although lymph nodes examined to be positive showed the strongest linear association with survival, it is a conventional clinical prognostic variable rather than a gene-expression biomarker, which provides indication that correlations are accurate. In contrast, MAPT, CHEK1, and AURKA were the most prominent genomic features identified in this analysis.

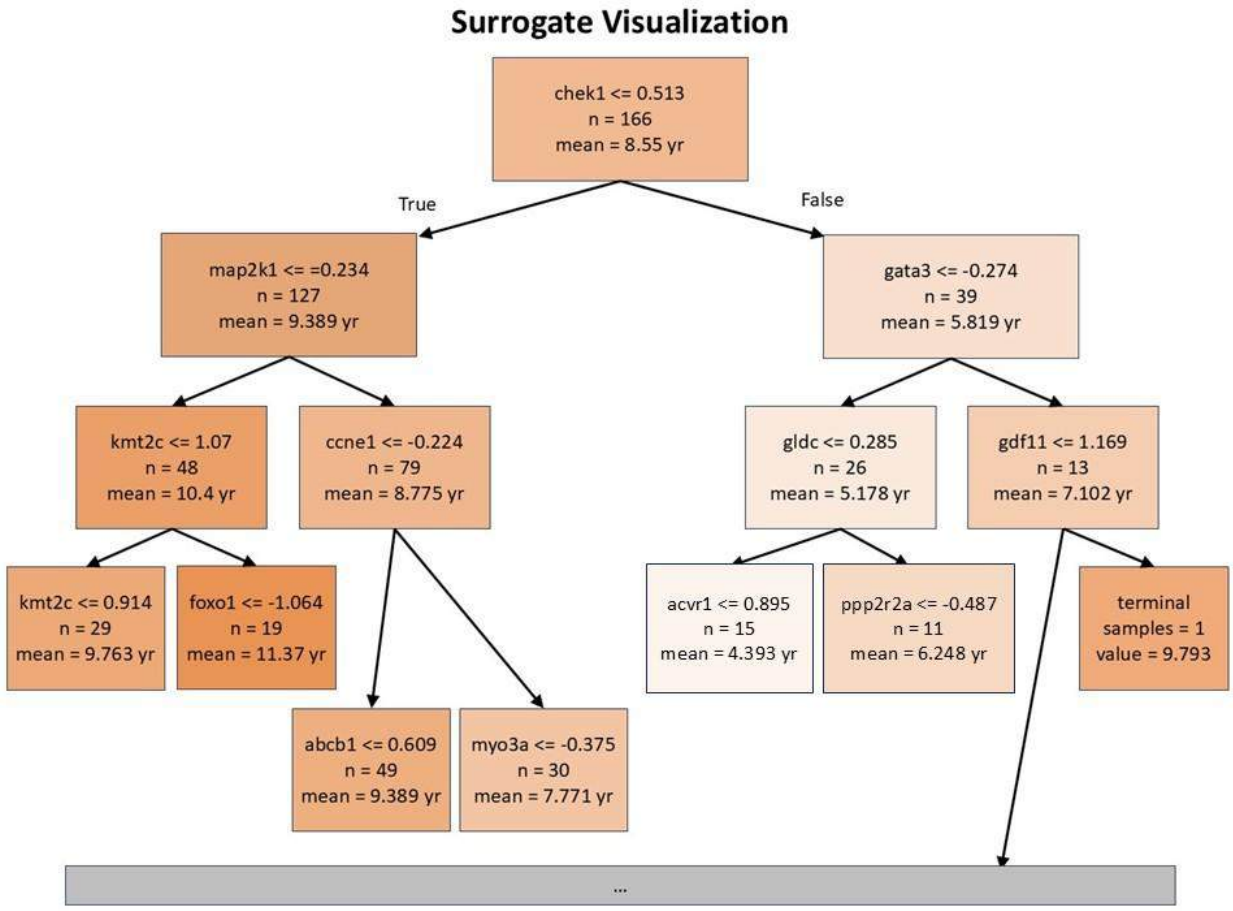


Figure 4. Surrogate tree visualization. Displays simplified version of the decision splits made by the model at the first three layers, deeper branches omitted for visibility.

To improve interpretability of the Random Forest model, a surrogate decision tree was trained to approximate the ensemble's predictions. Because a Random Forest combines many individual decision trees, its internal prediction process can be difficult to interpret directly. The surrogate tree distills the forest's complex decision-making process into a human-readable form, revealing which genes the model considers most important and how it uses them to stratify patients by predicted survival duration. The root node split on CHEK1 at a threshold of 0.513,

suggesting it to be the most important feature for the initial division of patients. Patients with lower CHEK1 expression (CHEK1 ≤ 0.513) had a mean predicted survival of 9.389 years (n = 127), whereas patients with higher expression (CHEK1 > 0.513) had a mean predicted survival of 5.819 years (n = 39). Subsequent splits involved MAP2K1, GATA3, KMT2C, and CCNE1, indicating that the model used a multi-gene decision pathway to further distinguish survival patterns across patient subgroups.

Discussion

The project was able to utilize a multi-model approach and identified three clinically validated gene biomarkers—MAPT, CHEK1, and AURKA—that correlate with breast cancer survival duration. The convergence of findings across different model types strengthens confidence in these biomarkers: MAPT and AURKA appeared in the top features of both linear regression and random forest, while CHEK1 emerged as the primary decision factor in the surrogate tree visualization. These findings resonate with existing cancer genomics literature. MAPT (Tau protein) expression has been positively correlated with breast cancer survival (Bonneau et al.; Callari et al.). CHEK1 overexpression has been established as an independent risk biomarker associated with aggressive breast cancer phenotype (Wu et al.). AURKA has been shown to outperform Ki67 as a prognostic marker in ER-positive breast cancer (Ali et al.). The alignment between our computational findings and published clinical research validates the multi-model approach.

The analysis focused primarily on genomic and gene-expression factors. However, actual survival duration depends on many additional variables including patient age, comorbidities, lifestyle factors (diet, exercise, emotional health), treatment regimens, and even unrelated causes of death (accidents, other diseases). These non-genomic factors introduce substantial variance into survival prediction. Notably, the models generally predicted survival durations lower than actual values. This under-prediction reflects the absence of protective factors in the model—while genomic markers capture risk, real-world outcomes are extended by factors such as effective treatment, healthcare access, and lifestyle variables that were not included. This observation highlights opportunities for future work incorporating broader clinical and demographic data.

The neural network underperformed (r^2=0.052) compared to linear regression and random forest, which was anticipated given the METABRIC dataset contains fewer than 2,000 records. Neural networks require substantially larger datasets to learn complex patterns without overfitting. The training/validation loss curves confirmed overfitting, with training loss approaching zero while validation loss plateaued. For future improvement, the neural network architecture is prepared to leverage much larger genomic datasets (e.g., The Cancer Genome Atlas with 10,000+ samples) where deep learning could discover additional differential factors among larger populations.

Conclusion

This project demonstrated that a multi-model machine learning approach can identify clinically significant gene biomarkers associated with breast cancer survival duration. Three key biomarkers were identified: MAPT, CHEK1, and AURKA, all of which have been independently validated in published cancer genomics research.

The hypothesis was partially supported. As predicted, different models contributed complementary findings: linear regression identified genes with proportional effects on survival, while random forest revealed non-linear gene interactions. The neural network did not perform well due to the small dataset size, as anticipated, but remains prepared for future application with larger datasets.

These biomarkers could enable physicians to personalize breast cancer treatment by assessing individual patients' MAPT, CHEK1, and AURKA expression levels. For example, patients with high CHEK1 expression (associated with poor prognosis) might benefit from more aggressive treatment or CHEK1 inhibitor therapy, while patients with high MAPT expression (associated with better prognosis) might be candidates for less intensive treatment regimens. This personalized approach could improve survival duration and quality of life for breast cancer patients through supporting more individualized risk stratification and treatment-planning research. Unlike existing programs that primarily predict binary survival outcomes (alive vs. deceased), this multi-model approach predicts continuous survival duration and identifies specific gene biomarkers. This provides more detailed prognostic information for future treatment-personalization research.

Although the METABRIC dataset provided valuable clinical and gene-expression information, its sample size was still relatively small for deep learning, and some variables contained missing values that required imputation. While median and mode imputation allowed the models to use more patient records, this approach may reduce biological nuance because it estimates missing values rather than measuring them directly. Future studies should use larger and richer datasets, such as TCGA, that include broader genomic, clinical, and treatment-response information (The Cancer Genome Atlas Research Network et al.). In addition, because this project was completed using freely accessible computing resources, model complexity and hyperparameter tuning were limited. Access to greater computational power could allow for more extensive neural network optimization, cross-validation, and survival-analysis methods that account for censoring. These improvements could strengthen both prediction accuracy and biomarker discovery in future versions of the project.

Works Cited

Ali, H. R., et al. "Aurora Kinase A Outperforms Ki67 as a Prognostic Marker in ER-Positive Breast Cancer." British Journal of Cancer, vol. 106, no. 11, May 2012, pp. 1798–806. DOI.org (Crossref), https://doi.org/10.1038/bjc.2012.167.

Al-kaabi, M. M., et al. "Checkpoint Kinase1 (CHK1) Is an Important Biomarker in Breast Cancer Having a Role in Chemotherapy Response." British Journal of Cancer, vol. 112, no. 5, Mar. 2015, pp. 901–11. DOI.org (Crossref), https://doi.org/10.1038/bjc.2014.576.

Bonneau, Claire, et al. "Predictive and Prognostic Value of the TauProtein in Breast Cancer." Anticancer Research, vol. 35, no. 10, Oct. 2015, pp. 5179–84.

Callari, Maurizio, et al. "Cancer-Specific Association between Tau (MAPT) and Cellular Pathways, Clinical Outcome, and Drug Response." Scientific Data, vol. 10, no. 1, Sept. 2023, p. 637. DOI.org (Crossref), https://doi.org/10.1038/s41597-023-02543-y.

Cancer Facts & Figures. American Cancer Society, 2026, https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2026/2026-cancer-facts-and-figures.pdf.

Cox, D. R. "Regression Models and Life-Tables." Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 34, no. 2, Jan. 1972, pp. 187–202. DOI.org (Crossref), https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.

Fong, Y., et al. "The Nottingham Prognostic Index: Five- and Ten-Year Data for All-Cause Survival within a Screened Population." The Annals of The Royal College of Surgeons of England, vol. 97, no. 2, Mar. 2015, pp. 137–39. DOI.org (Crossref), https://doi.org/10.1308/003588414X14055925060514.

Goodfellow, Ian, et al. Deep Learning. The MIT Press, 2016. Adaptive Computation and Machine Learning.

METABRIC Group, et al. "The Genomic and Transcriptomic Architecture of 2,000 Breast Tumours Reveals Novel Subgroups." Nature, vol. 486, no. 7403, June 2012, pp. 346–52. DOI.org (Crossref), https://doi.org/10.1038/nature10983.

Pereira, Bernard, et al. "The Somatic Mutation Profiles of 2,433 Breast Cancers Refine Their Genomic and Transcriptomic Landscapes." Nature Communications, vol. 7, no. 1, May 2016, p. 11479. DOI.org (Crossref), https://doi.org/10.1038/ncomms11479.

Siegel, Rebecca L., et al. "Cancer Statistics, 2026." CA: A Cancer Journal for Clinicians, vol. 76, no. 1, Jan. 2026, p. e70043. DOI.org (Crossref), https://doi.org/10.3322/caac.70043.

The Cancer Genome Atlas Research Network, et al. "The Cancer Genome Atlas Pan-Cancer Analysis Project." Nature Genetics, vol. 45, no. 10, Oct. 2013, pp. 1113–20. DOI.org (Crossref), https://doi.org/10.1038/ng.2764.

Wu, Mei, et al. "The Clinical Significance of CHEK1 in Breast Cancer: A High-Throughput Data Analysis and Immunohistochemical Study." International Journal of Clinical and Experimental Pathology, vol. 12, no. 1, 2019, pp. 1–20.

Machine Learning Approach with Feature Importance and Price Elasticity Analysis to Predict E-Commerce Sales Volume in Fashion Retail: Case Study of ZARA Brand

By Ward Dalati

Abstract

Accurate sales forecasting in fashion e-commerce is challenging due to rapidly changing trends, seasonality, and consumer behavior. This study develops machine learning models to predict sales volumes using product data from Zara, focusing on features such as price, promotion, season, and product attributes. Multiple models, including Linear Regression, Random Forest, Extra Trees, and XGBoost, were trained and tested. Results show that tree-based models outperform, by achieving high accuracy and strong generalization. Statistical significance tests confirm that XGBoost performs differently compared to other models. Feature importance analysis using SHAP reveals that promotion and price are the most influential factors driving sales. Additionally, price elasticity indicates that demand is relatively inelastic, while promotion has a stronger effect on increased sales. Overall, the study highlights the importance of model selection and product insights in improving predictive performance and supporting decision-making in fashion e-commerce.

Keywords e-commerce, sales, volume, price, Zara, fashion, product, machine learning, feature importance analysis, SHAP

Introduction

Forecasting the volumes of sales in e-commerce has become a very important topic to understand because of the rapid growth online retail has been experiencing, specifically in the fashion industry (Swaminathan and Venkitasubramony, 2023). Fashion e-commerce is constantly being affected by the changes in trends, seasonality, and varying customer preferences making it a regularly evolving industry (Giri and Chen, 2022). For this reason, accurately predicting sales volumes is very important in helping retailers understand customer demand and prepare for future outcomes. Without proper forecasting, businesses risk struggling which can lead to missed sales opportunities (Sheikh et al., 2025).

In the fashion industry, forecasting sales is known to be very challenging due to its naturally fast-changing trends and consumer behavior (Swaminathan and Venkitasubramony, 2023). There are always frequent changes that occur such as new collection releases, and demand for clothing items that can shift since it is directly affected by changes in seasonality, social influences, and marketing campaigns (Giri and Chen, 2022). These factors prove that statistical data on sales volume is not enough to make accurate predictions and data-driven approaches are essential to identifying patterns and forecasting future demand (Sheikh et al., 2025).

Machine learning is a very powerful tool in general to learn patterns and make predictions. In e-commerce, machine learning techniques offer aid for accurately representing sales forecasting (Swaminathan and Venkitasubramony, 2023). Machine learning has the ability

to learn patterns from data and constantly make improvements to their prediction by accessing more information. For example, these methods can analyze large amounts of data and identify relationships between variables such as price, product category, season, promotions, and past sales performances (Giri and Chen, 2022). Hence, they are suitable for industries like fashion retail.

Forecasting sales using machine learning comes with many benefits for e-commerce business. The most significant advantage is how it helps with inventory management (Sheikh et al., 2025). Due to machine learning algorithms' ability to predict which products will sell, retailers are able to manage their stocks levels accordingly while reducing overstocking, storage costs, and unsold inventory (Swaminathan and Venkitasubramony, 2023).

Another important benefit is better decision making in marketing and pricing strategies. Since, machine learning models are able to identify which product categories will perform well depending on other factors such as seasonality and promotion, businesses can better plan marketing campaigns. As a result, companies can increase sales volume while also aligning their inventory with customer demand (Chaitanya, 2022).

Overall, using machine learning techniques into prediction sales allows ecommerce companies to be more efficient in a very competitive market (Swaminathan and Venkitasubramony, 2023). Accurate prediction will help improve profitability and reduce operational risks. As retail fashion and ecommerce continue to grow in popularity, machine learning will play an increasingly important role in helping businesses understand customer behavior and make data driven decisions (Chaitanya, 2022).

Relevant Works

Machine learning techniques for forecasting sales in e-commerce and fashion retail have been highlighted in previous research. Since fashion retail is characterized as a constantly evolving industry with short product life cycles, rapidly changing consumer preferences, and product variations, artificial intelligence and machine learning models are better and more advanced when dealing with high demand, seasonality, and promotion-driven purchasing behavior (Tabak Kizgin et al., 2025). Machine learning is one of the most effective techniques as it rapidly adapts to changing market conditions and can capture non-linear relationships. Several studies have applied machine learning models to predict their e-commerce sales. For example, one study analyzed online sales data from a retailer in Brazil before, during, and after COVID-19. The researchers used SARIMA (a traditional model) and a neural network autoregression (NNAR) model, and found that NNAR had more accurate predictions than SARIMA especially when the sales patterns were irregular (Tabak Kizgin et al., 2025). However, the study found that after COVID-19 prediction errors increased because consumer behavior became more unstable.

Research on Zara’s operations also emphasizes the importance of forecasting. One case study created a forecasting system that uses historical sales data of similar items to predict demand for new Zara products. Additionally, a distribution optimization model was used to

determine how many products should be shipped to each store while understanding inventory management. The study found that this method could increase product sales by more than 12 percent (AI Expert Network, 2024). However, the model assumes that new products have the same characteristics as old items and will perform the same which is not always the case. Artificial intelligence has also been used within Zara's business model. Research on Zara's operational innovation shows how machine learning and artificial intelligence algorithms can analyze large amounts of data and forecast demand more accurately. These systems help companies become better at decision-making. By using these methods, Zara can respond quicker to changes in consumer demand and improve its supply chain (AI Expert Network, 2024). Overall, studies show that machine learning models have a lot of advantages for forecasting demand in e-commerce and fashion retail environments. It emphasizes on how traditional approaches are no longer feasible because of the complex patterns in sales data which machine learning can capture better. However, there are some limitations that were highlighted such as data quality issues, market disruptions, and difficulty of forecasting demand for a completely new item. These drawbacks show that while machine learning has been proven to be a much more effective technique in predicting demand, implementing it requires careful model selection to reduce the amount of errors.

Research Gap and Contribution

Most existing studies focus primarily on predicting sales using historical sales patterns rather than analyzing characteristics of individuals products despite the extensive use of statistical and machine learning methods in fashion retail forecasting. For example, studies on Zara and similar fashion retailers rely on SARIMA or neural networks like NNAR to forecast sales, seasonality, and promotional effect (Tabak Kizgin et al., 2025). While these approaches are able to perform well in identifying patterns and demand fluctuations in the short run, product level factors such as item attributes or product specific demand are overlooked. Although machine learning is applied in demand forecasting and operational planning systems, the focus remains on historical sales and similar product comparisons rather than deeper analysis on individual product features and how they influence sales outcomes (AI Expert Network, 2024). Overall, the key gap is the limited attention given to product information in forecasting models. We aim to develop a machine learning model that predicts the sales volume of Zara products as a case study. By analyzing data such as product position, season, promotion, material, section, price, and terms, the study's objective is to identify patterns that influence product demand in e-commerce. This study can potentially help retailers understand the specific information about the product and that can affect demand and revenue.

Dataset

Our dataset contains around 20,000 rows of Zara brand sales. Each row represents a single product, meaning the purpose of the data is focusing on change in sales volume and focusing on product attributes rather than purchase records. The dataset contains various

columns that describe product identification, product position, promotion, product category, seasonal, sales volume, brand, url, name, description, price, currency, terms, section, season, material, and origin as shown in Table 1.

Key variables include, product position, which describes where the product is situated in the store (aisle, front, back), seasonal, which describes whether the item belongs to a limited or seasonal collection, promotion which specifies if the product was part of a promotion, terms which describes the type of clothing item (jeans, jacket, etc.), material which describes the main fabric of the product, season which defines the product's season (Winter, Autumn, Spring, Summer), origin which describes the country of manufacturer or supplier, section which describes the gender (Man or Woman), and price which describes the product price is dollars. Finally, all those key variables are being used to understand Zara's sales volume which is the amount of units sold.

Table 1. Data type and description

Variable name	Variable type	Description
Product ID	categorical	Unique identifier assigned to each product
Product Position	categorical	Position of the product on the website or listing
Promotion	categorical	Indicates whether the product is promoted or not
Product Category	categorical	Type of product such as clothing category
Seasonal	categorical	Indicates whether the item belongs to a limited or seasonal collection
Sales Volume	numeric	Number of units sold for the product or estimated demand level per product
brand	categorical	Brand name of the product
url	categorical	Link to the product page
name	categorical	Official product title as it appears on the Zara website
description	categorical	Product details including design, fit, and additional features
price	numeric	Price of the product
currency	categorical	Currency used for the price
terms	categorical	Keyword or subcategory describing the type of clothing item
section	categorical	Gender-based product section such as woman or man
season	categorical	Defines the product's season - Spring, Summer, Autumn, or Winter
material	categorical	Main fabric of the garment
origin	categorical	Country of manufacture/supply

Methodology

Data Pre-Processing

In the data processing stage, categorical variables were adjusted through one hot encoding. This method converts each category into a separate binary column with numbers 0 and 1. These numbers represent a unique category within that feature. For example, in the Zara sales dataset, a variable-like *category* contains values such as “Dress” and “shoes”. The one-hot encoding transforms this into separate columns where a binary value of 1 tells us the category is present and 0 otherwise.

One-hot encoding was chosen instead of ordinal encoding because the categorical variables in this data set don't have a specific ranking. Using ordinal encoding would result in a higher risk of misleading numerical relationships and can negatively affect model performance. By using one-hot encoding, the model looks at each category independently which prevents bias. In addition, variables that only have two categories are kept as binary features which means that they are represented by a single column with a value of 0 or 1, such as Promotion,Seasonal, and Section. This process is known as binary encoding.

Models and Evaluation Metrics

The dataset was divided into testing and training sets to evaluate the models performance. 80% of the data was used to train the model to identify patterns while the remaining 20% of the data was used to test the model to assess how the models perform.

In this study, random forest, extra trees, XGBoost, and LightGBM were selected. These models are known for handling structured and capturing complex or non linear relationships between features. Random Forest and Extra Trees are based on multiple decision trees, where predictions are made by collecting the outputs of many trees (Breiman, 2001; Geurts et al., 2006). One advantage of Random Forest is its ability to reduce overfitting by averaging multiple trees, although it can be time consuming (Breiman, 2001). Extra Trees are similar but introduce more randomness which makes the process faster and reduces variancy (Geurts et al., 20064). On the other hand, XGBoost and LightGBM built trees sequentially where each new tree corrects errors from previous ones . These models are known for having strong predictive performance and for being very efficient but they require careful tuning and can be affected by random errors or irrelevant patterns in the dataset (Chen and Guestrin, 2016; Ke et al., 2017). Additionally, CatBoost was included as another model that is mostly effective for categorical variables. It reduces the risk of overfitting through the process of ordered boosting. Finally, Linear Regression was used as a baseline model. Unlike the tree-based methods, it processes the features and target variables to have a linear relationship which makes it easier to interpret and simple, but less capable of capturing complex non-linear patterns in the data.

To assess the model performance, three evaluation metrics were used: R^2, MAE, and MAPE. R^2 which is referred to as the coefficient of determination measures the models ability to explain the variation in the target variable. Better performance is indicated with values closer to

1. MAE is known as the mean absolute error calculator, the average difference between predicted and actual values. MAPE, which is the mean absolute percentage error, uses the MAE error result and expresses it as a percentage, allowing for comparison of accuracy across different scales (Chicco et al., 2021). Together, these metrics provide a detailed evaluation of the model's performance.

Hyperparameter Tuning

Hyperparameter tuning is the process of searching for the best combination of model parameters (Turner et al., 2021). Cross validation is a method used to evaluate different hyperparameter combinations. What it does is split the dataset into multiple folds which are known as subsets, trains the model on some folds and validates the others. It keeps repeating this process many times to get an average performance score (Turner et al., 2021). In this study, 5-fold cross validation is used, meaning the hyperparameter tuning process is repeated 5 times and the data is divided into 5 parts overall resulting in a more reliable evaluation.

In this study, a set of hyperparameters was selected for each model to control complexity and ensure fair comparison during tuning as shown in Table 2. For random forest and extra trees, the parameters include n_estimators, max_features, and min_samples_split. The parameter n_estimators (number of trees) was tested with values [100,500] while max_features ranged from 1 to 23 with an increment of 1 to control how many features are considered when splitting nodes. The min_samples_split parameter was varied from 5 to 100 with an increment of 5 to adjust the minimum number of samples required to split a node, helping to control over fitting.

For XGBoost, the key parameters were tuned, including n_estimators [100,500], subsample [0.6 to 1.0], colsample_bynode [0.6 to 1.0] and learning rate [0.0001 to 0.3]. These parameters determine the number of rounds, the features used per tree, and the step size for learning. Similarly for LightGBM, parameters such as n_estimators, subsample, colsample_bynodes, and learning_rate had the same range of values. These ranges were selected to allow models to optimize while maintaining efficiency.

Table 2. Model hyperparameters and their value ranges

Model	Parameter	Range
Random Forest	n_estimators	[100,500]
	max_features	1 to 23 (step = 1)
	min_samples_split	5 to 100 (step = 5)
Extra Trees	n_estimators	[100,500]
	max_features	1 to 23 (step = 1)

	min_samples_split	5 to 100 (step = 5)
XGBoost	n_estimators	[100,500]
	subsample	[0.6, 0.7, 0.8, 0.9, 1.0]
	colsample_bynode	[0.6, 0.7, 0.8, 0.9, 1.0]
	learning_rate	[0.001, 0.002, 0.003, 0.01, 0.02, 0.03, 0.1, 0.2, 0.3]
LightGBM	n_estimators	[100,500]
	subsample	[0.6, 0.7, 0.8, 0.9, 1.0]
	colsample_bynode	[0.6, 0.7, 0.8, 0.9, 1.0]
	learning_rate	[0.001, 0.002, 0.003, 0.01, 0.02, 0.03, 0.1, 0.2, 0.3]

Feature Importance Analysis

Feature importance analysis is a process that uses multiple techniques in order to determine how much each variable contributes to a model's predictions (Shin, 2023). In many machine learning models the decision process is referred to as a "black box" because it is difficult to understand why a model produces a specific output (Ribeiro et al., 2016). For this reason, feature importance helps unfold this black box by ranking features according to their importance on the model's predictions. This is very essential because it allows us to identify the most influential variables and potentially increase relevancy by reducing the number of unnecessary and least influential input variables. Overall, this can improve model performance, make the model more trustworthy, and help with decision making by highlighting which factors drive the outcomes (Shin, 2023).

SHAP (SHapley Additive explanations) is a specific method for explaining feature importance based on concepts from game theory. The main purpose is to distribute the model's prediction among all features by measuring each feature's contribution to the prediction (Ponce-Bobadilla et al., 2024). SHAP works by computing the influence of each variable then assigns each feature a value that represents how much it increases or decreases the prediction relative to a baseline. One of its strengths that it is known for is how it is able to give two explanations: local which is about individual predictions and global which is about overall feature importance. Additionally, SHAP is consistent and satisfies additivity which means that the sum of all feature contributions equal to the model output. This makes SHAP a powerful tool for interpreting complex machine learning models (Ponce-Bobadilla et al., 2024).

Statistical Significance Test

Statistical tests such as the Wilcoxon test and the bootstrapped paired test are used to compare the performance of different models on the test data. These tests evaluate whether performance between the two models occurred by chance or are statistically significant (Demsar,

2006). The Wilcoxon test is a non-parametric test that compares observations such as prediction error from two models on the same test set without assuming a normal distribution (Diebold and Mariano, 1995). For this reason, this test is stable for machine learning evaluation, where error distributions are usually non-normal. This test ultimately determines whether one model consistently outperforms another by analyzing the differences in the ranks (Diebold and Mariano, 1995).

On the other hand, the bootstrap paired test, estimates the variability of the model performance through using resampling techniques (Xu, 2006). It iterates many times between two models by sampling the test dataset with replacement and computing the difference. This process allows us to evaluate the stability results and calculate the confidence intervals with all the generated distribution of differences (Xu, 2006). One of its main advantages is how it does not rely on strict assumptions about the data distribution and it is able to prove more insight into the uncertainty of model comparisons. Overall, both of these tests ensure that the improvement is not because of random variation in the test data and that the change is reliable.

Price - Volume Elasticity

Price-volume elasticity measures how sensitive the demand of a product is to changes in its price. The general formula is the percentage change in quantity demand divided by the percentage change in price (Prasad, 2010). The results help identify the relationship between pricing decisions and customer response. If elasticity is less than 1 then that means that there is inelastic demand and consumers are sensitive to price changes therefore decreasing the price would have a large boost on sales volume. However if elasticity is greater than 1 then that means there is elastic demand and consumers are not sensitive to price changes therefore changing the price will not have a big effect on the sales (Prasad, 2010). This concept is especially important in retail contexts because pricing strategies directly influence revenue and inventory management.

In economics, price-volume elasticity is used to understand and guide the pricing and promotion decisions. For example, if a product is relatively elastic, lowering the price during promotions can significantly impact sales volume and potentially increase total revenue. However, if a product is relatively inelastic, increasing the price wouldn't result in a significant reduction of demand which can improve profit margins. In this study, elasticity is derived from the machine learning models, helping interpret how price changes impact predicted sales. This provides insight on decision-making such as promotion, setting competitive prices, and identifying which products are more sensitive to pricing strategies.

Results and Discussion

The tuned tree-based models Random Forest, Extra Trees, and XGBRegressor show very strong and consistent performance. All three models achieve high R^2 values around 0.94 on the training set and 0.935 on the test set, showing that they explain most of the variety in sales volume as shown in Table 3. Error metrics are also low and similar across models, with MAE

around 56-57 on the train set and 59 on the test set, along with MAPE around 5.2-5.5%. Figure 1 provides a visual that confirms this consistently where the bars represent the train and test values of R^2 and MAPE which only have menial differences across models showing consistency. Overall, no model clearly outperforms the others meaning they all performed equally good. The difference between training and test metrics is very small across all models. For example, R^2 only changes slightly (for example, 0.943 to 0.935), and MAE/MAPE increase only by a bit on the test set. This indicates no significant overfitting. The models generalize well to unseen data, as their performance remains stable across both the train and test datasets. The consistency between train and test bars in Figure 1 further supports this conclusion.

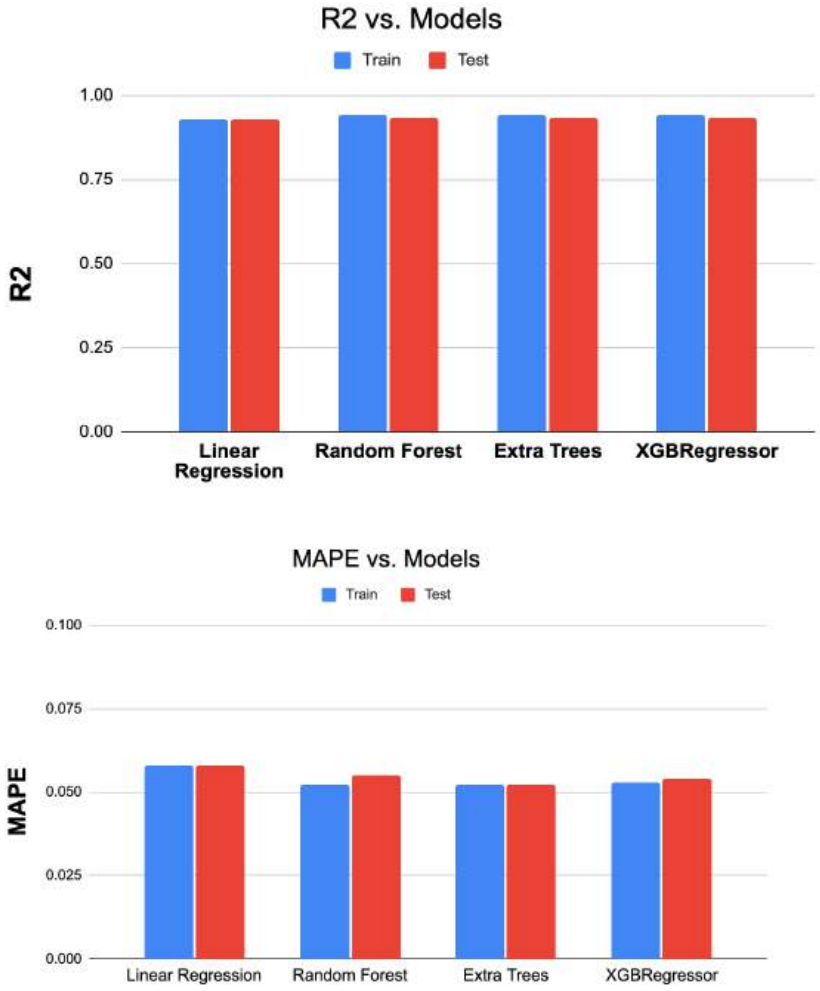
The tuned models achieved their best performance using specific hyperparameter combinations as shown in Table 3. For both Random forest and extra trees, a high number of trees was 500 n_estimators selected, which helps averaging many decision trees while reducing variance and improving stability. The max_features, 14 for Random Forest and 18 for Extra Tree indicate that a number of features are considered at each split which allow the models to understand more information from the dataset. Additionally the min_samples_ split values, 80 for Random Forest and 100 for Extra Tree function to restrict tree growth in order to prevent overly complex trees and help reduce overfitting.

For the XGBRegressor, the optimate parameters have a more controlled learning process as shown in Table 3. It has a low learning rate of 0.01 which ensures that the model learns gradually, improving generalization. The use of subsample =0.6 and colesmaple_bynode =0.6 introduces randomness by sampling data points and features which helps prevent overfitting and improves stability. 500 estimators reinforces the slow learning rate which allows the model to learn without instability.

Table 3. Evaluation metrics of baseline and tuned models with optimal parameters

Model	Best Parameters	Split Set	R^2	MAE	MAPE
Linear Regression	baseline	Train	0.93	62.225	0.058
		Test	0.928	62.321	0.058
Random Forest	max_features: 14, min_samples_split: 80, n_estimators: 500	Train	0.943	56.322	0.052
		Test	0.935	59.381	0.055
Extra Trees	max_features: 18, min_samples_split: 100, n_estimators: 500	Train	0.943	56.171	0.052
		Test	0.935	59.329	0.055
XGBBoost	colsample_bynode: 0.6, learning_rate: 0.01, n_estimators: 500, subsample: 0.6	Train	0.942	56.704	0.053
		Test	0.935	59.081	0.054

Figure 1. Train and test (top) R^2 and (bottom) MAPE of the models



Statistical significance testing was used to determine whether the differences in model performance on the test data was because of random variation or actually genuine. Two non-parametric methods were used, the Wilcoxon test, which compares paired performance differences between models without assuming a normal distribution, and the paired bootstrap test, which estimates the stability of performance differences by resampling the test data many times. Together, these tests provide a more detailed evaluation of whether one model outperforms another by a significant amount.

The results in Table 4 show that many model comparisons show statistically significant differences. The Wilcoxon test shows highly significant differences (very small p-values) between most pairs involving XGBoost, which means these models behave differently in predictive performance compared to Linear Regression and other tree-based models (Extra Tree and Random Forest). In contrast, comparisons such as Linear Regression and Random Forest ($p = 9.228557e\text{-}01$) and linear regression and extra tree ($p=2.938524e\text{-}01$) are not statistically significant which indicates similar performance distributions.

The paired bootstrap results generally show the presence of differences between models, although in some comparisons the p-values are close to or equal to 1.000, which reflects strong separation in resampled performance distributions. Overall, the findings confirm that the advanced model (XGBoost) is statistically different from baseline approaches, showing the importance of model selection in improving predictive accuracy.

Table 4. Summary of Wilcoxon and paired bootstrap statistical significance tests to assess the differences in the models' MAE scores on test data

Model A	Model B	Wilcoxon p-value	Paired Bootstrap p-value	Statistical Significance
Linear regression	Random Forest	9.228557e-01	0.085	No
Linear regression	Extra Trees	2.938524e-01	0.001	No
Linear regression	XGBRegressor	1.1416145e-02	0.937	Yes
Random Forest	Extra Trees	7.7975495e-07	1.000	Yes
Random Forest	XGBRegressor	2.755753e-10	1.000	Yes
Extra Trees	XGBRegressor	1.098772e-20	1.000	Yes

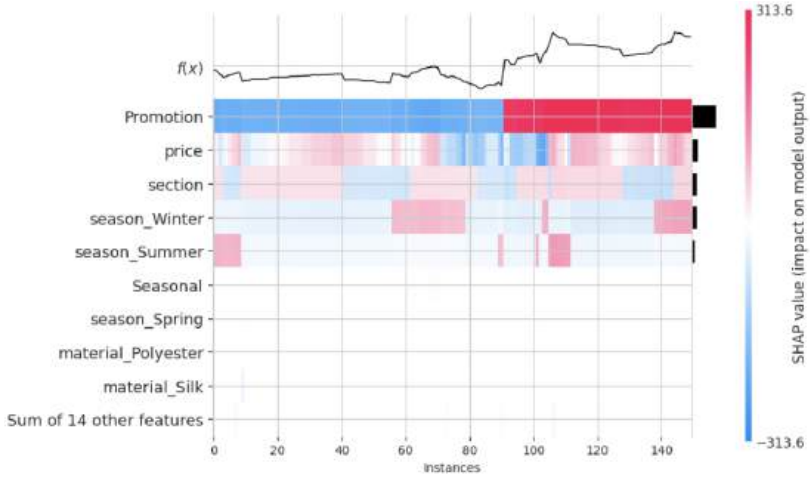
Feature importance analysis was used to better understand how different variables influence the prediction of sales volume. For this reason, the XGBoost model was selected. To understand the model, SHAP analysis was used, which provides the contribution of each feature to the model's output.

As shown in Figure 2, the SHAP summary plots highlight that Promotion is the most influential feature, meaning it has the largest impact on increasing predicted sales. This is followed by price which is also significant although higher prices can lead to a decrease in sales volume. Seasonal variables such as Winter and Summer, as well as section, also contribute to the predicted sales but to a lesser extent. The distribution of SHAP values show how both positive and negative impacts can occur depending on the feature.

Figure 3 provides a heatmap representation of SHAP values which is a more detailed view of how each variable's contributions change for each sales prediction. The heatmap reemphasizes the influence of Promotion and price shown by strong positive (red) and negative (blue) effects. It also illustrates how certain features have more time specific effects such as seasons. Overall, the SHAP analysis concludes that pricing and promotional strategies are the primary factors of sales predictions in the XGBoost model.

Figure 2. Global (left) and local (right) contribution of features on model prediction from SHAP analysis

Figure 3. Feature contribution heatmap using SHAP values



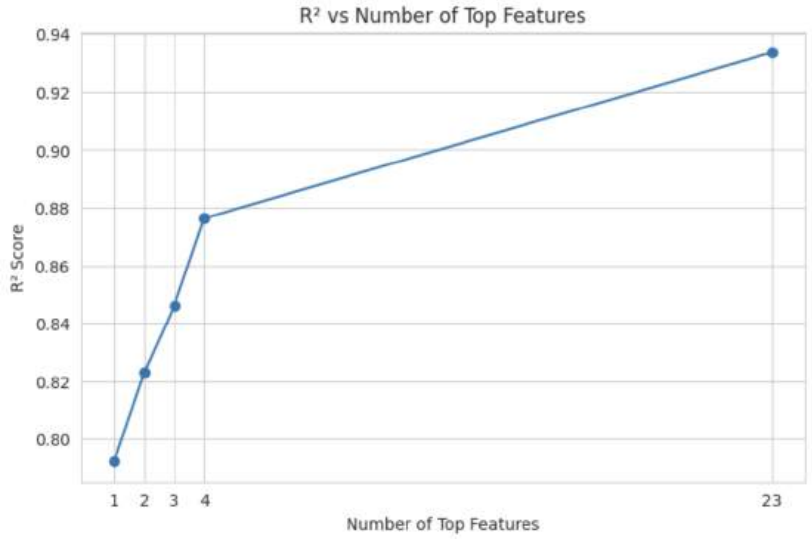
To evaluate the importance of different predictors, an additional experiment was conducted where the models were retained using only the selected top features, specifically price and promotion. This approach was used to see if the model could maintain similar predictive performance, or if the features contribute meaningful information to the model.

When all features were used, the models performed consistently well. Test R^2 values ranged from approximately 0.928 to 0.935, showing that a large portion of the variance in sales was explained. The MAE values were low between 59 and 62, while MAPE remained stable around 0.054 to 0.058. These results mean that the models were able to produce accurate and reliable predictions when it's exposed to all input features. However, when the models were limited to only price and promotion, there was a decline in performance. Test R^2 values dropped to a range of approximately 0.84 to 0.86, showing reduction in the ability to explain variation in the data. On the other hand, MAE increased significantly to value between 83 and 92, reflecting larger prediction errors.

Comparing the two sets of results highlights a trade-off between simplicity and performance. The reduction in features led to a decrease in R^2 by around 0.07 to 0.09 and an increase in MAE around 20 to 30 units. This shows that even though price and promotion are

important predictors, they are not enough on their own to capture the complexity of sales patterns. Overall, the experiment demonstrates that sales volumes are influenced by multiple factors beyond just price and promotion. While feature reduction makes the model more simple, it results in a loss of important information and accuracy in prediction. Therefore, models trained on the full feature set are better for producing accurate and reliable predictions.
Further feature-selection experiments were conducted using the top 1,2,3, and 4 most important features identified by the XGBoost model. As shown in Figure 4, model performance steadily improved as more top-ranked features were added. Using only the single most important feature, Promotion, produced an R^2 value of approximately 0.79. Adding the second feature, section, increased the R^2 value to around 0.82, while including the top 3 and 4 features further improved performance to approximately 0.85 and 0.88 respectively. When all 23 features were included, the R^2 increased substantially to approximately 0.93. These results demonstrate that additional relevant features improve predictive accuracy, although the first few features already capture a large proportion of the model's predictive power. This suggests that Promotion is the strongest individual predictor of the sales, but combining multiple seasonal and product related variables allows the model to capture complex sales behavior.

Figure 4. Effect of top features on model performance (R^2)



Additionally, The price-elasticity experiment was conducted using a log-log regression model. It used log_sales as the dependent variable and log_price and Promotion as predictors, allowing interpretation of coefficients in percentage terms. The OLS regression results show a strong model fit, with an R^2 value of 0.842, meaning that approximately 84.2% of the variation in log_sales is explained by the two predictors, log_price and Promotion. This model is also statistically significant since it has a p-value < 0.001, indicating that price and promotion together provide a strong explanation for sales behavior.

The price elasticity is captured by the coefficient for log_price, which is -0.1158 and statistically significant ($p < 0.001$). Since both sales and price are in logarithmic form, this coefficient can be interpreted as elastic. This means that a 1% increase in price leads to a 0.12% decrease in sales, showing that demand is relatively inelastic. In other words, customers aren't sensitive to price changes in this dataset, meaning that other factors beyond price may have a bigger influence on purchasing decisions.

The promotion variable has a coefficient of 0.4646 ($p < 0.001$), indicating a strong positive effect on sales. This suggests that when there is a promotion, sales increase by around 46% in long terms, illustrating how the effect of promotion has a much larger impact on demand than price changes. This shows customers react strongly to promotional changes rather than small changes in price. Overall, the results suggest that promotions are the main driver of demand while price has a statistically significant but negative effect on sales.

Additionally, Figure 5 visually represents the relationship between log-transformed price and log-transformed sales using a fitted regression line. The downward slope of the line shows that as price increases, sales generally decrease, supporting the negative elasticity coefficient identified in the regression results. However, the relatively gradual slope suggests that the effect of price on sales is limited, further reinforcing that demand in this dataset is relatively inelastic. The spread of the data points around the fitted line also indicates that other factors, such as promotions and seasonal variables, contribute to variations in sales performance.

Figure 5. Log-Log relationship between price and sales

Discussion

The best-performing models in this study were the tree-based models, specifically Random Forest, Extra Trees, and XGBoost. These models had high performance, with test R^2 values around 0.935 and low error metrics with MAE around 59 and MAPE around 5.4%. The small difference between training and test performance shows that the models don't experience a

lot of overfitting and generalize well. Statistical significance testing further supported these findings, reflecting that advanced models such as XGBoost perform significantly differently compared to baseline models like Linear regression. This is consistent with findings from the literature, where ensemble models outperform traditional statistical methods due to their ability to capture complex and non-linear relationships in data (Tabak Kizgin et al., 2025; Swaminathan and Venkitasubramony, 2023).

The SHAP analysis dives deep into how the model makes predictions. The results show that promotion is the most influential feature, followed by price and seasonal variables. Promotion consistently showed a strong positive impact on sales prediction while higher prices generally contributed negatively. These findings are consistent with existing research which emphasize the importance of promotional strategies and seasonality in influencing consumer behavior in fashion e-commerce (Giri and Chen, 2022). Additionally, the difference observed in SHAP values confirms that the effect of features is not always the same and depends on the context of each product, highlighting the advantage of machine learning models in understanding and capturing complex interactions.

The price elasticity further supports these results. The elasticity coefficient of -0.1158 indicates that demand is relatively inelastic, meaning the changes in price have a limited effect on sales volume. On the other hand, the promotion variable shows a much more positive impact, increasing sales significantly. This aligns with the literature that explains that promotional strategies usually have a greater influence on demand compared to price changes in fashion retail (Chaitanya, 2022). Overall, the results emphasize that while pricing plays a role, promotional activities are the key driver of sales in this dataset.

This study has several limitations that should be considered when reading the results. First, the data was limited to a specific context, as it focuses only on Zara products. This means that the findings are specific to Zara and may not generalize to other brands or industries especially those with different customer behavior patterns and pricing strategies.
Second, the data is limited in capacity, with around 20,000 observations. While this is sufficient for model training, a larger dataset across multiple regions could improve the generalizability of the results. Additionally, the dataset may not consider external factors such as economic donations, social media trends, or competitor pricing which are also key factors that can influence sales.

Furthermore, although many product attributes were included, there may still be missing features that influence demand such as customer reviews, brand perception, or marketing data. Including these could enhance model performance.

Conclusions

This study was aimed to develop a machine learning model to predict sales volumes in fashion e-commerce using Zara as a case study. The objective was not only to build an accurate model but also to understand how difference factors influence sales outcomes. By using detailed

product data, this study contributes to filling in the gaps in existing research which often focuses more on historical sales trends rather than individual product characteristics.

The results show that tree-based machine learning models, particularly Random Forest, Extra Trees, and XGBoost, provide the best predictive performance. These models had high accuracy and demonstrated strong generalitation, as proven by consistent performance across training and testing datasets. Statistical significance tests confirmed that these models outperform simpler baseline approaches, highlighting the importance of selecting the right algorithms for sales forecasting.

The SHAP analysis provided insights into feature importance, uncovering that promotion and price are the most influential factors in prediction sales. Promotion specially has a strong positive impact on demand while price has a negative but smaller effect. Seasonal and product related features also contribute to predictions, illustrating that sales are influenced by a combination of factors rather than one single variable.

The price elasticity analysis further proved these findings by showing that demand is relatively inelastic with respect to price changes, but sensitive to promotions. This suggests that retailers may benefit more from promotional strategies rather than only price adjustments to increase sales.

However, the study has limitations related to the size of the dataset, feature availability, and generalizability. Future research could improve this work by adding external variables and multiple brands to improve strength of the model. Overall, this study demonstrates the effectiveness of machine learning in predicting sales and provides insights for decision making in fashion e-commerce.

Works Cited

Swaminathan, Kritika, and Rakesh Venkitasubramony. “Demand Forecasting for Fashion Products: A Systematic Review.” International Journal of Forecasting, vol. 40, no. 1, Mar. 2023, https://doi.org/10.1016/j.ijforecast.2023.02.005.

Giri, Chandadevi, and Yan Chen. “Deep Learning for Demand Forecasting in the Fashion and Apparel Retail Industry.” Forecasting, vol. 4, no. 2, 20 June 2022, pp. 565–581, www.mdpi.com/2571-9394/4/2/31, https://doi.org/10.3390/forecast4020031.

Chaitanya, Krishna. “Machine Learning Approaches in Demand Forecasting for Supply Chain Management.” International Journal of Scientific Research and Management (IJSRM), vol. 10, no. 03, 29 Mar. 2022, pp. 809–821, ijsrm.net/index.php/ijsrm/article/view/3767.

Sheikh, Abdullah, et al. A Hybrid Machine Learning Framework for Supply Chain Demand Forecasting: Integrating Historical Data and Market Intelligence.

Tabak Kizgin, Kiymet, et al. “Machine Learning‐Based Sales Forecasting during Crises: Evidence from a Turkish Women’s Clothing Retailer.” Scientific Programming, vol. 108, no. 1, 2025, article 368504241307719, doi:10.1177/00368504241307719

“Zara’s Comprehensive Approach to AI and Supply Chain Management.” AI Expert Network, 20 Apr. 2024, aiexpert.network.

Breiman, Leo. “Random Forests.” Machine Learning, vol. 45, no. 1, 2001, pp. 5–32, doi:10.1023/A:1010933404324.

Geurts, Pierre, Damien Ernst, and Louis Wehenkel. “Extremely Randomized Trees.” Machine Learning, vol. 63, no. 1, 2006, pp. 3–42.

Ke, Guolin, et al. “LightGBM: A Highly Efficient Gradient Boosting Decision Tree.” Advances in Neural Information Processing Systems (NeurIPS), 2017.

Chen, Tianqi, and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System.” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, arXiv:1603.02754.

Chicco, Davide, et al. “The Coefficient of Determination R-Squared Is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation.” PeerJ Computer Science, vol. 7, no. 5, 2021, p. e623. ncbi, peerj.com/articles/cs-623/, https://doi.org/10.7717/peerj-cs.623.

Hyndman, Rob J., and Anne B. Koehler. “Another Look at Measures of Forecast Accuracy.” International Journal of Forecasting, vol. 22, no. 4, 2006, pp. 679–688.

Turner, Ryan, et al. “Bayesian Optimization Is Superior to Random Search for Machine Learning Hyperparameter Tuning: Analysis of the Black‐Box Optimization Challenge 2020.” arXiv, 2021, arXiv:2104.10201.

Shin, Terence . “Understanding Feature Importance in Machine Learning | Built In.” Builtin.com, 7 Nov. 2023, builtin.com/data-science/feature-importance.

Ribeiro, Marco Tulio, et al. ““Why Should I Trust You?”: Explaining the Predictions of Any Classifier.” ArXiv.org, 16 Feb. 2016, arxiv.org/abs/1602.04938.

Ponce‐Bobadilla, Ana Victoria, et al. “Practical Guide to SHAP Analysis: Explaining Supervised Machine Learning Model Predictions in Drug Development.” Clinical and Translational Science, vol. 17, no. 11, 28 Oct. 2024, pmc.ncbi.nlm.nih.gov/articles/PMC11513550/, https://doi.org/10.1111/cts.70056.

Demšar, Janez. “Statistical Comparisons of Classifiers over Multiple Data Sets.” Journal of Machine Learning Research, vol. 7, 2006, pp. 1–30.

Diebold, Francis X., and Roberto S. Mariano. “Comparing Predictive Accuracy.” Journal of Business & Economic Statistics, vol. 13, no. 3, 1995, pp. 253–263

Xu, Kuan-Man. “Using the Bootstrap Method for a Statistical Significance Test of Differences between Summary Histograms.” Monthly Weather Review, vol. 134, no. 5, 1 May 2006, pp. 1442–1453, https://doi.org/10.1175/mwr3133.1. Accessed 18 Oct. 2021.

Prasad, Ashutosh. “Demand Elasticity.” Wiley International Encyclopedia of Marketing, 2010.