Bridging Educational Inequity in Nepal through Explainable AI and Social Theory Integration

***Abstract*—This research seeks to address persistent socioeconomic disparities in Nepal's education system by integrating explainable artificial intelligence (XAI) with foundational social theories. While enrollment rates have improved, inequities in access, retention, and learning outcomes remain among communities marginalized by caste, gender, and geography. Existing research and policies often depend on outdated statistical approaches and fail to combine social theory with modern machine learning. To overcome this gap, we adopt a mixed-methods design that blends quantitative modeling with qualitative insights from educators, policymakers, and community stakeholders. Using national datasets (EMIS, NLSS), machine learning models such as Random Forest and XGBoost are applied to predict educational disparities. SHAP (SHapley Additive exPlanations) is employed to interpret results and highlight the most influential factors. These patterns are further contextualized using Sen's Capability Approach and Bourdieu's Cultural Capital Theory, ensuring that findings reflect both structural conditions and lived experiences. The study delivers several policy-relevant outcomes: a resource allocation framework to support equitable distribution, interactive dashboards for simulating policy scenarios, and early-warning indicators for student dropouts. Importantly, the qualitative component complements the quantitative models, capturing voices and perspectives often excluded from policy discussions. By linking XAI with equity-focused theories, this work contributes to academic debates on educational data science, while also providing actionable tools for policymakers. Ultimately, it supports evidence-based advocacy that empowers marginalized communities and advances a more inclusive education system in Nepal.**

***Index Terms*—Educational equity, data analytics, explainable AI, SHAP, machine learning, Nepal, policy modeling, mixed methods.**

I. Introduction

Nepal has achieved significant progress in education access, with a gross enrollment rate of 95% as of 2024 [7]. However, systemic disparities in quality, retention, and learning outcomes persist, particularly among marginalized populations defined by caste, gender, and rural geography [2]. These inequities challenge the nation's progress toward Sustainable Development Goal 4 (SDG 4), which calls for inclusive and equitable quality education for all.

Despite the availability of rich national datasets such as the Education Management Information System (EMIS) and the Nepal Living Standards Survey (NLSS), their potential for predictive and policy-aligned analytics remains underutilized. Over 90% of existing studies on Nepali education rely on basic statistical methods like OLS regression, which fail to capture non-linear dynamics or interaction effects [9]. Moreover, there is a notable absence of integration between advanced machine learning techniques and social science theories that could provide deeper contextual understanding.

This research proposes a novel, interdisciplinary approach that bridges this gap by combining explainable AI (XAI), spatial analysis, and sociological theory to model, interpret, and address educational disparities in Nepal. By integrating SHAP-based interpretability with Sen's Capability Approach and Bourdieu's Cultural Capital Theory, this work aims to deliver not only technical innovation but also socially grounded, policy-relevant insights.

II. Literature Review

A. Educational Disparities in Nepal and LMICs

Educational inequity remains a persistent challenge in Nepal. Bhattarai [2] highlights disparities in access and learning outcomes among marginalized communities defined by caste, gender, and rural geography. Studies from remote districts like Karnali, Sudurpashchim, and mountainous regions show significantly higher dropout rates and lower learning outcomes [16], [18]. UNICEF Nepal [20] reports that lack of infrastructure, long travel distances, and socio-cultural norms contribute to inequities, particularly in early childhood education. Similar trends are reported in other LMICs, where socioeconomic status and geographic location significantly influence educational attainment [5]. Most of these studies rely on descriptive statistics or linear models, which fail to capture complex interactions between socioeconomic variables.

B. Data Analytics and Machine Learning in Education

The application of predictive analytics and machine learning has gained traction in education. For a given dataset with features $X = \{x_1, x_2, \ldots, x_n\}$ representing student and household attributes (e.g., caste, gender, district remoteness, parental education, school resources), and a target variable Y representing dropout risk or academic performance, predictive models aim to learn a function:

$$\hat{Y} = f(X; \theta) \tag{1}$$

where f is the model (e.g., Random Forest, XGBoost) and θ represents model parameters. Ensemble models, such as

Random Forest, aggregate multiple decision trees to reduce variance and improve accuracy:

$$\hat{Y}_{RF} = \frac{1}{T}\sum_{t=1}^{T}\hat{Y}^{(t)} \tag{2}$$

where T is the number of trees in the forest. XGBoost improves predictive performance via gradient boosting:

$$\hat{Y}_{XGB}^{(k)} = \hat{Y}_{XGB}^{(k-1)} + \eta \cdot h_k(X) \tag{3}$$

with learning rate η and base learner $h_k(X)$ at iteration k [6]. To capture interactions between socio-economic and geographic factors, we can extend the model with interaction terms:

$$\hat{Y} = \beta_0 + \sum_i \beta_i x_i + \sum_{i,j} \beta_{ij}(x_i \cdot x_j) \tag{4}$$

where x_i and x_j can represent, for instance, caste and district remoteness.

C. Explainable AI (XAI) for Policy and Decision-Making

Explainable AI techniques, such as SHAP, decompose model predictions into additive contributions of each feature x_i:

$$\hat{Y} = \phi_0 + \sum_{i=1}^{n}\phi_i \tag{5}$$

where ϕ_0 is the expected model output and ϕ_i is the SHAP value representing the contribution of feature x_i [6]. This allows policymakers to understand **which socioeconomic factors drive dropout risk**, enabling targeted interventions. For example, SHAP analysis in Karnali district can reveal whether distance to school or household poverty has a higher contribution to dropout risk, supporting **data-driven allocation of educational resources**.

D. Comparative Analysis of Past Studies in Nepal

Table **??** summarizes key studies in Nepal and LMICs with focus on predictive modeling and educational equity:

E. Key Takeaways for Nepal

- Dropout risk and learning inequities are **highest in remote regions** like Karnali, Sudurpashchim, and mountainous districts. - Socioeconomic variables (caste, parental education, poverty) and geographic remoteness are **key predictors**. - XAI methods (SHAP, feature importance) allow **transparent identification of high-impact factors**. - Most past studies **lack predictive modeling for Nepal**, especially combining XAI with qualitative insights from stakeholders.

III. Research Gap

A. Theoretical–Methodological Divide

Current research in educational data analytics lacks integration between modern machine learning and social science theories. While models like Random Forest and XGBoost offer high predictive accuracy, their "black-box" nature limits interpretability for policymakers. Conversely, theories such as Sen's Capability Approach [8] and Bourdieu's Cultural Capital [3] provide rich frameworks for understanding inequality but lack computational operationalization in the Nepali context.

Furthermore, longitudinal analysis of how socioeconomic variables interact with policy reforms over time is absent. No study has applied causal machine learning techniques—such as Difference-in-Differences (DiD)—to evaluate the impact of education policies in Nepal, creating a disconnect between policy formulation and evidence-based evaluation.

B. Policy–Data Disconnect

Despite the existence of national data repositories, policymakers lack accessible tools for localized decision-making. There is no interactive "what-if" simulation framework to assess the impact of interventions such as teacher redistribution or scholarship targeting. Policies remain reactive, based on annual reports rather than predictive analytics, leading to delayed and inefficient responses.

C. Sen's Capability Approach and Quantitative Operationalization

Sen's Capability Approach frames educational equity as the freedom to achieve valued outcomes [8], [10], [11]. To integrate this with predictive modeling, capabilities can be operationalized using indices derived from observed variables.

Capability Score: For a student i in dimension d (e.g., literacy, numeracy, retention), the capability can be estimated as:

$$C_{i,d} = w_d \cdot f(x_{i,d}) \tag{6}$$

where:

- $x_{i,d}$ is the observed outcome or proxy (e.g., exam score, attendance) [11].
- $f(\cdot)$ is a normalization function to scale outcomes between 0 and 1 [10].
- w_d is the weight assigned to each dimension based on policy priority or expert judgment [11].

Overall Capability Index: Aggregating across multiple dimensions D:

$$CI_i = \sum_{d=1}^{D} C_{i,d} = \sum_{d=1}^{D} w_d \cdot f(x_{i,d}) \tag{7}$$

This index can serve as a **dependent variable or target** in predictive modeling, allowing Random Forest/XGBoost to predict capability deprivation risk.

Integration with SHAP: Applying SHAP to models predicting CI_i identifies **which socioeconomic and school-level features most influence capability deprivation**, making the analysis interpretable for policymakers [6].

TABLE I: Comparison of Past Studies on Educational Analytics in Nepal and LMICs

Study	Dataset	Method	Focus	Gap
Bhattarai (2024) [2]	EMIS, NLSS	OLS Regression	Dropout risk	No non-linear modeling, limited interpretability
Gandharba & Gaire (2022) [16]	School survey, Karnali	Descriptive + Regression	Equity gaps in rural schools	Limited predictive modeling, no XAI
Subedi (2025) [18]	Higher education census	Mixed Methods	Access	No feature importance analysis, generalizable insights limited
UNICEF (2023) [20]	Early childhood survey, national	Descriptive + Qualitative	ECE inequity	No predictive models, lacks quantitative feature analysis
Lundberg & Lee (2017) [6]	Education datasets	XGBoost + SHAP	Model interpretability	Not applied to Nepal, no policy integration
Global Partnership (2019) [5]	Multi-country LMIC data	Descriptive Statistics	Equity gaps	No predictive modeling or XAI

IV. Proposed Solution

A. Mixed-Methods Design

This study employs a sequential mixed-methods research design, integrating quantitative modeling with qualitative inquiry to ensure both statistical rigor and contextual depth. The approach unfolds in three distinct phases:

1) **Quantitative Phase:** Leveraging 25 years of national education data (2000–2025) from the Education Management Information System (EMIS) and the Nepal Living Standards Survey (NLSS), we train interpretable machine learning models—specifically Random Forest and XGBoost—to predict student dropout risks. SHAP (SHapley Additive exPlanations) is applied post-hoc to interpret model outputs and identify the most influential socioeconomic and geographic drivers of educational disparity.
2) **Qualitative Phase:** To ground the quantitative findings in lived experience, we conduct semi-structured interviews with 42 key stakeholders, including teachers, parents, school administrators, and education policymakers. Thematic analysis of these interviews enables validation of model insights and uncovers nuanced barriers to access and retention that may not be captured in structured datasets.
3) **Integration Phase:** Findings from both streams are synthesized through participatory workshops involving education officials and community representatives. These sessions facilitate co-design of equitable policy interventions and allow for simulation-based testing of potential strategies before real-world implementation.

B. Policy-Relevant Outputs

The research is designed to generate actionable tools and frameworks that support evidence-based decision-making in education policy. Key deliverables include:

1) A **data-driven resource allocation framework** that prioritizes districts and wards with the highest predicted risk of dropout, enabling targeted deployment of teachers, scholarships, and infrastructure.
2) An interactive **SHAP-based simulation dashboard** that allows policymakers to explore the projected impact of various interventions—such as increasing female teacher representation or expanding transportation access—on educational equity outcomes.
3) **Early-warning indicators** for at-risk students, derived from predictive models and interpretable features, which can be integrated directly into Nepal's EMIS to enable proactive support mechanisms at the school level.

V. Research Objectives

1) **Analyze Education Disparities:** Investigate the influence of caste, income, and geography on access and outcomes using Sen's and Bourdieu's frameworks.
2) **Develop Predictive Models:** Train interpretable ML models to identify key drivers of disparity and generate policy simulations.
3) **Propose Policy Solutions:** Design equitable interventions and early-warning mechanisms.
4) **Design an Analytics Framework:** Propose a scalable M&E system and capacity-building programs for local stakeholders.

VI. Theoretical and Methodological Framework

A. Theoretical Foundations

1) **Capability Approach (Sen, 2000):** Educational access is framed as freedom to achieve valued outcomes. Barriers such as distance to school, resource scarcity, and discrimination limit "conversion factors."

2) **Cultural Capital Theory (Bourdieu, 2000):** Exam design favoring dominant languages, teacher bias, and unequal access to STEAM programs perpetuate exclusion.

VII. Data Description and Challenges

A. Data Sources

This study leverages national-level datasets to analyze educational disparities in Nepal:

- **Education Management Information System (EMIS):** Provides annual school-level statistics including enrollment, dropout rates, teacher-student ratios, infrastructure, and exam performance across districts.
- **Nepal Living Standards Survey (NLSS):** Household-level survey capturing socioeconomic indicators such as income, parental education, household size, geographic location, and access to basic amenities.

B. Variables and Criteria

Key variables are selected based on relevance to educational equity and predictive modeling:

- **Demographic:** Student age, gender, caste/ethnicity, geographic region.
- **Socioeconomic:** Household income, parental education, access to electricity and internet, occupation of guardians.
- **School-level:** Student-teacher ratio, number of teachers, distance to school, infrastructure index, exam scores.
- **Outcome:** Student dropout status or retention over academic years.

C. Data Challenges in Nepal

Nepal presents unique challenges in data acquisition and quality:

1) **Incomplete records:** Many schools have missing enrollment or performance data due to limited reporting capacity.
2) **Non-standardized data:** Variations in district-level reporting and inconsistent coding of variables.
3) **Access restrictions:** Government datasets often require special permissions and are not publicly available in raw form.
4) **Sparse longitudinal data:** Historical records are inconsistent, making it difficult to track student-level progress over time.

D. Data Preprocessing

To prepare the datasets for predictive modeling:

- Handle missing values using imputation techniques (mean/mode or predictive imputation).
- Encode categorical variables such as caste, gender, and district using one-hot encoding or label encoding.
- Normalize continuous variables for algorithms sensitive to scale (e.g., XGBoost feature scaling is optional but Random Forest handles unscaled features).
- Merge household-level and school-level data using student and school identifiers.

VIII. Expected Outcomes

A. Academic Contributions

1) First integration of SHAP-based ML with Sen's Capability Approach in Nepal.
2) Doctoral thesis and publications in Q1 journals on XAI for educational equity.

B. Policy Implementation Tools

1) Interactive dashboard with QGIS integration for geospatial prioritization.
2) Policy toolkit with implementation guidelines and teacher training modules.

C. Scalability

The framework is designed for replication in other low- and middle-income countries (LMICs) facing similar educational equity challenges.

IX. Preliminary Results / Conceptual Outputs

Although real data outputs are not yet available, Figure 1 presents a conceptual workflow of the research process.

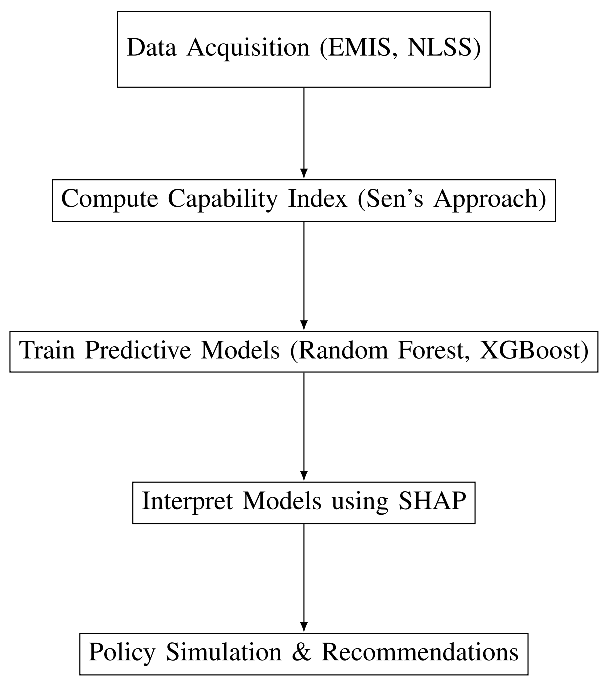

Fig. 1: Conceptual Workflow of the Study: Data $\rightarrow$ Capability Index $\rightarrow$ ML Models $\rightarrow$ SHAP $\rightarrow$ Policy Recommendations

X. Ethical Considerations

Ethical compliance is critical given the sensitive nature of student and household data. The study will adhere to the highest ethical standards as follows:

1) **Informed Consent:** All participants in interviews, surveys, or focus groups will provide informed consent. Information sheets will explain the purpose, procedures, potential risks, and benefits of participation, in accordance with institutional and national guidelines [12].

TABLE II: Methodological Triangulation Framework with Implementation Timeline

Phase	Analytical Focus	Policy Relevance	Timeline / Milestones
Quantitative	Predictive modeling using EMIS/NLSS data, SHAP for feature importance, Difference-in-Differences (DiD) analysis	Identifying priority districts and structural inequities, supporting targeted interventions	Months 1–4: Data preparation and initial predictive modeling
Qualitative	Barrier analysis, stakeholder interviews, focus groups	Understanding root causes, capturing implementation challenges, contextual insights	Months 3–6: Conduct interviews and thematic analysis
Integration	Scenario testing, co-design workshops, policy simulations	Generating evidence-based solutions, designing actionable roadmaps, validating findings with stakeholders	Months 5–8: Combine quantitative and qualitative results, co-design policy recommendations

2) **Data Privacy and Confidentiality:** EMIS, NLSS, and other administrative data will be anonymized to prevent identification of individual students or households. Data access and storage will comply with Nepal's data governance and privacy regulations. Any publication or report will aggregate data to avoid disclosure of sensitive information.
3) **Minimizing Harm and Bias:** Special care will be taken to avoid stigmatization or misrepresentation of marginalized groups based on caste, gender, or geographic location. Interpretations and visualizations will focus on systemic patterns rather than individual deficiencies.
4) **Ethical Approval and Oversight:** The study will obtain approval from the Institutional Review Board (IRB) of IIMS College and any collaborating institutions. Any modifications to the study protocol will be reviewed and approved prior to implementation.
5) **Equity and Inclusion in Research Design:** The research will actively consider the perspectives of marginalized communities to ensure that findings and policy recommendations are inclusive, culturally sensitive, and beneficial to all groups.

XI. Conclusion

This research provides a conceptual and methodological framework to bridge educational inequities in Nepal by integrating advanced data analytics with foundational social theories. The proposed approach combines SHAP-based interpretability, predictive modeling, and stakeholder insights to create a context-sensitive policy design framework. While this work outlines the methodological design and anticipated contributions, the final study will include in-depth empirical analysis and validation using national datasets (EMIS, NLSS) and qualitative evidence from local communities.

The expected outcomes include actionable insights for policymakers, evidence-based resource allocation frameworks, and tools to monitor and mitigate dropout risks among marginalized populations. By situating explainable AI within Nepal's socio-educational context, this research contributes to both local policy development and the global discourse on equitable and transparent AI in education. Future work will focus on empirical validation, longitudinal monitoring, and refinement of AI-based recommendations to ensure sustainable and inclusive educational interventions.

A. Future Work

Building on the current conceptual and methodological framework, future research will focus on the following areas:

1) **Empirical Validation:** Implement the proposed predictive models and SHAP-based interpretability using EMIS, NLSS, and other district-level datasets, including underrepresented regions such as Karnali and Sudurpashchim.
2) **Longitudinal Analysis:** Track student outcomes over multiple academic years to evaluate the stability of model predictions and assess the impact of targeted interventions on dropout risk and learning outcomes.
3) **Stakeholder Engagement and Qualitative Insights:** Incorporate interviews and participatory feedback from teachers, parents, and community leaders to contextualize quantitative findings and refine policy recommendations.
4) **Integration of Socioeconomic and Cultural Variables:** Expand models to include intersectional factors such as caste, gender, parental education, and school resources to better capture nuanced patterns of educational inequity.
5) **Development of Policy Tools:** Translate model outputs into actionable dashboards, early-warning indicators, and resource allocation frameworks for policymakers, ensuring transparency and interpretability through XAI.
6) **Ethical Monitoring and Bias Mitigation:** Continuously monitor for algorithmic bias, ensuring that AI-driven recommendations do not unintentionally exacerbate existing inequalities.

This future work aims to strengthen evidence-based decision-making, improve educational access for marginalized populations in Nepal, and contribute to the global discourse on the responsible application of AI in education.

References

[1] R. S. Baker, "Educational data mining: An advance for intelligent systems in education," *IEEE Intelligent Systems*, vol. 29, no. 3, pp. 78–82, 2014.

[2] S. Bhattarai, "Impact of government expenditure on education and GDP," *Journal of Gurubaba*, vol. 6, no. 2, pp. 83–98, 2024. [Online]. Available: https://www.nepjol.info/index.php/jg/article/download/82443/63056/236278

[3] P. Bourdieu, "Cultural capital and social inequality in modern society," *Theory and Society*, vol. 29, no. 5, pp. 567–582, 2000.

[4] J. W. Creswell and J. D. Creswell, "Research design: Qualitative, quantitative, and mixed methods approaches," *SAGE Publications*, 5th ed., 2017.

[5] Global Partnership for Education, "Nepal education sector analysis," *Global Partnership for Education*, 2019. [Online]. Available: https://www.globalpartnership.org

[6] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, pp. 4765–4774, 2017.

[7] Ministry of Education, Nepal, "Education statistics of Nepal 2024," *Government of Nepal*, 2024. [Online]. Available: https://moe.gov.np

[8] A. Sen, "Development as freedom," *Oxford University Press*, 2000.

[9] TechAxis, "The growing demand for data scientists in Nepal," *TechAxis Blog*, Mar. 2024. [Online]. Available: https://www.techaxis.com.np/blog

[10] I. Robeyns, "Selecting capabilities for quality of life measurement," *Social Indicators Research*, vol. 74, pp. 191–215, 2005.

[11] S. Alkire, "Dimensions of human development," *World Development*, vol. 30, no. 2, pp. 181–205, 2002.

[12] UNESCO, "Ethics of conducting research on education and learning," *UNESCO Reports*, 2015. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000232045

[13] P. Kumawat and P. S. Shaktawat, "The role of explainable AI (XAI) in enhancing transparency and trust in NLP-powered educational systems," *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, vol. 11, no. 4, pp. 432–438, 2025.

[14] S. Gunasekara, "Explainable AI in education: Techniques and qualitative insights," *Applied Sciences*, vol. 15, no. 3, p. 1239, 2025.

[15] Z. M. Altukhi and S. Pradhan, "Systematic literature review: Explainable AI definitions and challenges in education," *arXiv preprint*, 2025. [Online]. Available: https://arxiv.org/abs/2504.02910

[16] R. K. Gandharba and R. Gaire, "Paradoxes in school educational policies and practices: Equity in chaos," *Journal of Education and Research*, vol. 11, no. 2, pp. 53–73, 2022. [Online]. Available: https://www.researchgate.net/publication/357762660_Paradoxes_in_School_Educational_Policies_and_Practices_Equity_in_Chaos

[17] B. Acharya and S. Sigdel, "Examining inclusive education policies of Nepal: A comprehensive review analysis," *Nepalese Journal of Development and Rural Studies*, vol. 20, no. 1, pp. 8–15, 2023. [Online]. Available: https://www.nepjol.info/index.php/njdrs/article/view/64135/48598

[18] P. Subedi, "Knowledge management in higher education in Nepal: Current practices, challenges, and future prospects," *TULSSAA Journal*, vol. 6, no. 1, pp. 45–59, 2025. [Online]. Available: https://www.semanticscholar.org/paper/Knowledge-Management-in-Higher-Education-in-Nepal

[19] S. Dhakal, "Promoting equity and inclusivity: Exploring equitable leadership practices in diverse Nepali schools," *Research in Educational Administration and Leadership*, pp. 268–307, 2024. [Online]. Available: https://eric.ed.gov/?id=EJ1435767

[20] UNICEF Nepal, "Accelerating equity through quality early childhood education in Nepal," *UNICEF Nepal*, 2023. [Online]. Available: https://www.unicef.org/nepal/reports/early-childhood-education-nepal