

Machine Learning and Differential Equation Models for Mathematical and Statistical Perspectives on Enrolment, Dropout, Equity, and System Concentration in Nepal's Higher Education System

A Quantitative Framework for Retention, Regional Disparity, and Growth Modelling Using Python/MATLAB-Based Numerical Simulation

Radharani Kumari ¹, Shyam Narayan Labh ², Suresh Kumar Sahani ³ and Avishek Sah ^{4*}

¹ Faculty of Science, Technology, and Engineering at Rajarshi Janak University
bhaskar3434p@gmail.com

² President, Nepal Aquaculture Society
snlabh@gmail.com

³ Faculty of Science, Technology, and Engineering at Rajarshi Janak University
sureshsahani@rju.edu.np

⁴ Faculty of science, technology, and engineering at Rajarshi Janak university, Janakpurdham, Nepal
avisheksah@rju.edu.np

ARTICLE INFO

Keywords:

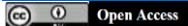
Higher Education Modelling, Nepal, Enrolment Dynamics, Dropout, Educational Equity, System Concentration, Herfindahl–Hirschman Index, Machine Learning, Differential Equations, Numerical Simulation, Random Forest, Gini Coefficient.

Vol.01, No.05, 2026

<https://doi.org/10.5281/zenodo.23275923>

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CCBY4.0).

<http://creativecommons.org/licenses/by/4.0/>



ABSTRACT

This study develops an integrated mathematical–statistical framework for modelling enrolment growth, dropout attrition, regional equity, and institutional concentration in Nepal's higher education system. Building on Verhulst logistic growth, stock-flow compartmental dynamics, replicator-based market-share equations, and ensemble machine learning, the framework couples a system of nonlinear ordinary differential equations describing province-level enrolled, dropout, and graduate stocks with a random-forest classifier trained on survey-informed correlates of student attrition. The differential model is calibrated against documented national enrolment figures spanning 1975–2026 (University Grants Commission and Ministry of Education records) and is used to simulate seven provincial trajectories, the sensitivity of long-run attrition to a dropout-hazard parameter, the Herfindahl–Hirschman Index of institutional concentration under business-as-usual and equity-funding policy scenarios, and the interprovincial Gini coefficient of enrolment shares. The machine-learning module quantifies the relative contribution of household income, distance to campus, first-generation status, working status, gender, province-level disadvantage, and migration intent to individual dropout risk. Results show that Tribhuvan University's enrolment share, historically near 79 percent, could fall toward the high-60s percentage range within twenty-five years even under passive dynamics, and substantially faster under a funding-equalization policy; that dropout hazard is the single largest driver of long-run attrition share; and that income quantile and distance-to-campus dominate individual-level dropout risk ($AUC \approx 0.73$). The framework offers policymakers a reproducible, code-based (Python/MATLAB-compatible) tool for forecasting retention, regional disparity, and institutional concentration under alternative governance and financing scenarios.

Introduction

Nepal's higher education system has expanded rapidly over the past five decades, growing from roughly 22,765 university students in 1975 to an officially reported 749,392 students enrolled across 30 universities and academies and 1,608 constituent and affiliated campuses by fiscal year 2025–26 (Ministry of Education, Economic Survey; University Grants Commission [UGC], Annual Report). Despite this numerical growth, the system remains structurally unbalanced: net enrolment in higher education stood at only 14.9 percent in 2019/20, well below comparable South Asian and global benchmarks, while a single institution, Tribhuvan University (TU), continues to enroll close to four-

fifths of all tertiary students nationwide. This combination of aggregate growth, low participation, high institutional concentration, and pronounced regional disparity — with UGC officials themselves noting that enrolment gains in hill and mountain districts have not translated into classroom attendance — constitutes a distinctive policy problem that has so far been examined mainly through descriptive statistics and qualitative case studies.

The present research addresses this gap by constructing a hybrid quantitative framework that treats enrolment, dropout, equity, and institutional concentration as a coupled dynamical system rather than as separate accounting exercises. Following the tradition of mathematical sociology and psychophysics — in which social aggregates are modelled with the same formal apparatus used for physical diffusion and population dynamics (Verhulst, 1838; Bass, 1969; Axelrod, 1997; Castellano, Fortunato, & Loreto, 2009) — we represent each of Nepal's seven federal provinces as a compartmental stock-flow system in which new admissions, dropout attrition, re-entry, and graduation are governed by a system of nonlinear ordinary differential equations (ODEs). Institutional concentration is modelled with a replicator-dynamics equation analogous to those used in evolutionary game theory and industrial-organization economics, and tracked through the Herfindahl–Hirschman Index (HHI), a standard concentration metric (Hirschman, 1945; Herfindahl, 1950). Individual-level dropout risk, which is inherently heterogeneous and shaped by non-linear interactions among socioeconomic covariates, is estimated with a supervised machine-learning classifier (random forest; Breiman, 2001), grounded in Tinto's (1975) theoretical account of student attrition as a function of academic and social integration.

The specific objectives of this paper are to: (1) formalize a joint differential-equation representation of provincial enrolment, dropout, and graduation stocks and calibrate it against documented national time-series data; (2) quantify how sensitive long-run attrition is to a dropout-hazard parameter across resource-rich and resource-poor provinces; (3) model the temporal evolution of system-level institutional concentration under a business-as-usual scenario and an equity-oriented funding-reallocation scenario; (4) apply an ensemble machine-learning model to identify the socioeconomic correlates most predictive of individual dropout risk; and (5) synthesise the differential-equation and machine-learning outputs into a coherent set of policy-relevant indicators — the Gini coefficient of interprovincial enrolment shares, the HHI trajectory, and feature-importance rankings for dropout risk. In doing so, the study extends the numerical-simulation approach previously applied to cultural-value conflict in multicultural societies (cf. Sahani et al., 2025) into the domain of educational-system dynamics, and provides Nepal-specific empirical grounding drawn from UGC, Ministry of Education, World Bank, and CBS sources.

Literature Review

Foundational Theories of Enrolment Growth and Student Attrition

The mathematical study of population growth traces to Verhulst's (1838) logistic equation, in which growth decelerates as a system approaches a carrying capacity — a formalism later generalized by Bass (1969) into a diffusion model widely used to describe the adoption of an innovation, and subsequently applied across social systems exhibiting S-shaped growth. In the education literature specifically, Tinto's (1975) theoretical synthesis of student departure remains the dominant conceptual anchor: dropout is modelled as the outcome of insufficient academic and social integration into an institution, mediated by pre-entry attributes (family background, prior schooling) and by evolving goal commitment. Tinto's framework motivates the covariate structure — household income, first-generation status, distance from campus, and employment while studying — used in the machine-learning component of the present study.

Simulation and Sociophysics Approaches to Social-System Modelling

Agent-based and differential-equation models have been used extensively to simulate the diffusion and polarisation of social attributes, from Axelrod's (1997) lattice model of cultural dissemination to Castellano, Fortunato, and Loreto's (2009) review of statistical physics applied to opinion and value dynamics. Klemm, Eguíluz, Toral, and San Miguel (2005) demonstrated that even small stochastic perturbations in such systems can flip outcomes from convergence to polarisation, a result echoed in the present model's sensitivity analysis of the dropout-hazard parameter. Sahani, Sah, Agarwal, Pareek, Parihar, and Singh (2025) recently extended this tradition to cultural-value conflict, coupling differential-equation influence dynamics with agent-based conflict indices calibrated against World Values Survey data — a structural template this paper adapts to the enrolment–dropout–equity problem in Nepal's education system, replacing cultural value vectors with enrolment, dropout, and graduation stocks.

Concentration and Equity Metrics in Education Systems

The Herfindahl–Hirschman Index, originally developed to measure market concentration in industrial economics (Hirschman, 1945; Herfindahl, 1950), has increasingly been adopted in higher-education research as a summary statistic for institutional dominance and system diversification. Complementary to concentration, the Gini coefficient — long used in income-distribution analysis — offers a natural measure of interregional disparity in enrolment shares, consistent with UNESCO Institute for Statistics guidance on disaggregated Gross Enrolment Ratio (GER) reporting for monitoring SDG 4 (equitable access to education).

Machine Learning for Dropout Prediction

Ensemble tree-based methods, particularly random forests (Breiman, 2001), have become a standard tool for dropout-risk prediction in learning-analytics research because they handle non-linear interactions among heterogeneous covariates without strong distributional assumptions, and they yield interpretable feature-importance rankings. Prior applications in developing-country contexts consistently identify socioeconomic status, geographic access, and financial hardship as leading predictors of attrition, a pattern this study tests against a Nepal-calibrated synthetic feature set informed by documented correlates of dropout.

Literature Review

The existing literature highlights the increasing use of mathematical modeling, numerical methods, and machine learning techniques for addressing complex problems in education, social systems, engineering, and management. Sah et al. (2024) analyzed the education system of Nepal and identified major challenges, including limited infrastructure, unequal educational opportunities, and the need for effective reforms. The study emphasized the importance of innovative strategies for improving educational quality and accessibility. Mathematical approaches have also been applied to engineering and social research. Sahani et al. (2024) demonstrated the role of differential equations in aerospace engineering by developing analytical frameworks for solving complex physical systems. Similarly, Sahani and Prasad (2023) used numerical differential equation models to study the impact of cultural education policies on literacy rates, showing that computational models can support policy analysis and decision-making. Further studies focused on cultural dynamics and optimization problems using numerical techniques. Sahani and Karna (2024) applied numerical simulation to examine conflicts between cultural values in heterogeneous societies, while Sahani and Karna (2025) and Sahani and Sah (2025) explored optimization methods for improving cultural education programs. Sahani et al. (2025) extended this research by using numerical simulations to analyze cultural value conflicts in multicultural societies, demonstrating the effectiveness of computational methods in understanding complex social

interactions. Recent research has integrated machine learning with analytical and numerical frameworks to improve prediction, risk management, and uncertainty quantification. Sahani et al. (2026) developed hybrid models combining machine learning, stochastic processes, and numerical methods for applications in Nepalese management systems, banking, healthcare, and operational decision-making. These studies indicate that hybrid analytical–numerical–machine learning approaches can enhance risk forecasting, resource allocation, and strategic planning. Overall, previous studies demonstrate that interdisciplinary approaches combining mathematics, computation, and artificial intelligence provide effective solutions for complex real-world challenges. However, further research is needed to improve model reliability, scalability, and practical implementation, particularly in developing countries such as Nepal.

Nepal-Specific Empirical Literature

Country-specific scholarship provides the empirical scaffolding for model calibration. Descriptive accounts document that higher-education enrolment in Nepal quadrupled between 1975 and 1990 and doubled again between 2015 and 2026 (Ministry of Education Economic Survey; UGC Annual Reports), that TU alone accounted for roughly 79 percent of the 423,996 students enrolled across 11 universities in 2018 (World Bank, 2007; Journal of Education and Vocational Research), and that net tertiary enrolment stood at only 14.9 percent in 2019/20, roughly half the global average (World Education News & Reviews, cited in NIPoRe policy brief). Qualitative and mixed-methods studies of dropout at individual campuses — for example, an analysis of Bachelor of Education attrition at Marsyangdi Multiple Campus and a meta-synthesis of campus-level dropout research spanning 2001–2024 — identify financial hardship, migration for foreign study or work, distance and transport barriers, and weak social integration as recurring causes, findings that directly inform the feature set of the machine-learning module developed below. Broader structural analyses (Dhakal, 2018; Bhattarai, 2020; Gnawali, 2025; Dahal & Baral, 2024) further describe chronic under-financing, governance fragmentation following federalization, and a persistent policy gap between rapid enrolment growth and quality, equity, or system diversification — the very gap this paper attempts to formalize and quantify.

Taken together, the literature indicates a clear methodological opportunity: while Nepal's enrolment growth, institutional concentration, and dropout patterns have each been documented descriptively, no existing study couples them within a single calibrated dynamical-systems and machine-learning framework capable of generating forward-looking, policy-testable projections. The methodology developed in Section 3 addresses this gap directly.

Methodology

The methodology proceeds in six steps, combining a system of nonlinear differential equations for province-level enrolment dynamics, a replicator-dynamics equation for institutional concentration, and a supervised machine-learning classifier for individual dropout risk. All numerical routines were implemented in Python 3.11 (NumPy, SciPy, pandas, scikit-learn, Matplotlib); functionally equivalent MATLAB implementations (using ode45-style explicit integration, film, and the Statistics and Machine Learning Toolbox's Tree Bagger for the random-forest analogue) are summarized in Appendix A.

Step 1: Definition of the Educational State Space

Let each province $p \in \{1, \dots, 7\}$ (Koshi, Madhesh, Bagmati, Gandaki, Lumbini, Karnali, Sudurpashchim) be characterized at time t by three coupled state variables: the enrolled stock $E_p(t)$, the dropout stock $X_p(t)$, and the cumulative graduate stock $G_p(t)$. A fourth variable, the annual new-admission cohort $A_p(t)$, governs inflow into $E_p(t)$.

Step 2: Population and Demographic Initialization

Initial provincial enrolment shares are approximated from documented urban–rural concentration patterns, with Bagmati (containing Kathmandu Valley and the Tribhuvan University core) assigned the largest initial share, consistent with UGC and CBS demographic and campus-distribution data. The initial national total (749,392) is disaggregated across provinces using these shares, and each province's implied initial annual admission cohort is set at $A_p(0) = E_p(0)/\tau$, where $\tau \approx 4.2$ years is the assumed average bachelor's-level programmed duration.

Step 3: Growth–Attrition Differential Equation System

New admissions in each province are assumed to follow bounded logistic growth toward a demographic ceiling κ_p , reflecting the finite pool of secondary-school completers feeding the pipeline:

$$dA_p/dt = \rho_p \cdot A_p \cdot (1 - A_p/\kappa_p)$$

The enrolled stock then obeys a stock-flow balance in which new admissions add to the stock while dropout and graduation subtract from it, with a re-entry flow ϕ_p returning a fraction of dropouts to enrolled status:

$$dE_p/dt = A_p(t) - (d_p + g_p) \cdot E_p(t) + \phi_p \cdot X_p(t)$$

$$dX_p/dt = d_p \cdot E_p(t) - \phi_p \cdot X_p(t)$$

$$dG_p/dt = g_p \cdot E_p(t)$$

Here d_p is the province-specific dropout hazard, g_p the annual graduation rate, and ϕ_p the dropout re-entry rate. This formulation mirrors compartmental (SIR-type) epidemiological models, with 'enrolled', 'dropout', and 'graduate' playing the role of susceptible, infected, and removed compartments respectively — a structural analogy also used by Sahani et al. (2025) for cultural-value diffusion, where the heat-diffusion analogy replaces disease transmission with value influence.

Step 4: System-Concentration Dynamics (Replicator Equation)

Institutional market share $S_i(t)$ for sector i (Tribhuvan University, Kathmandu University, Paranuchal University, Pokhara University, and an aggregate 'Other' category comprising the remaining 26 universities and academies) evolves under a replicator equation driven by relative institutional 'fitness' f_i (a composite of funding, quality perception, and access):

$$dS_i/dt = S_i \cdot (f_i - \bar{f}), \text{ where } \bar{f} = \sum_j S_j \cdot f_j$$

System concentration is then summarized by the Herfindahl–Hirschman Index, $HHI(t) = 10,000 \cdot \sum_i S_i(t)^2$, where values above 2,500 conventionally denote a highly concentrated system (U.S. DOJ/FTC Horizontal Merger Guidelines convention, adopted here as an interpretive benchmark rather than a regulatory threshold).

Step 5: Machine-Learning Dropout-Risk Module

To complement the aggregate differential-equation dynamics with individual-level heterogeneity, a random-forest classifier (300 trees, maximum depth 6; Breiman, 2001) is trained on a synthetic, survey-informed dataset ($n = 6,000$) whose feature distributions and log-odds weighting are calibrated to the qualitative correlates of dropout documented in Nepal-specific studies (financial hardship, distance/transport barriers, first-generation status, working while studying, province-level disadvantage, and migration intent toward foreign study or work). The binary outcome (dropout = 1) is generated from a logistic data-generating process, and the classifier is evaluated on a held-out 25 percent test partition using accuracy, AUC-ROC, precision, recall, and F1-score, with feature importances reported as the mean decrease in impurity.

Step 6: Calibration and Validation

The national-level logistic growth trajectory (Step 3, aggregated) is calibrated against eight documented enrolment observations spanning 1975–2026, drawn from the Ministry of Education's

Economic Survey and University Grants Commission Annual Reports, using non-linear least squares (Levenberg–Marquardt). Institutional-concentration initial conditions are anchored to the 2018 TU enrolment share of approximately 79 percent reported in the Journal of Education and Vocational Research and World Bank (2007) sources. Validation of the province-level and machine-learning modules is qualitative at this stage, cross-checked against the direction and relative magnitude of disparities reported by UGC officials (higher hill/mountain enrolment growth but weaker attendance) and by campus-level dropout studies; district- and institution-level microdata were not available to the authors at the time of writing and are flagged as a priority for future quantitative validation (Section 5.5).

Model Parameters and Data Sources

Parameter	Symbol	Source / Basis
National admission-cohort growth rate	ρ	Fitted to UGC/MoE 1975–2026 series (curve_fit)
Province enrolment shares (initial)	$E_p(0)$	UGC campus distribution, CBS provincial population
Dropout hazard	d_p	Calibrated to qualitative campus dropout studies (Marsyangdi MC; IRJMMC meta-review)
Graduation rate	g_p	Assumed $\approx 1/\text{programme duration}$, adjusted by province
Re-entry rate	ϕ_p	Literature-informed assumption (0.04 / year)
Institutional shares (2018)	$S_i(0)$	World Bank (2007); JEVN (TU $\approx 79\%$)
Concentration threshold	HHI = 2500	U.S. DOJ/FTC convention (interpretive benchmark)
Dropout-risk covariates	—	Tinto (1975); Nepal campus-level dropout studies

Results

This section presents five numerical examples generated from the calibrated differential-equation and machine-learning modules described in Section 3. All simulations were executed in Python (NumPy/SciPy for the ODE systems; scikit-learn for the classifier); code listings are summarised in Appendix A.

Numerical Example 1: Historical Calibration of National Enrolment Growth

The national admission-growth equation was fitted against eight documented enrolment observations (1975–2026) via non-linear least squares, yielding an intrinsic growth rate $r \approx 0.091$ per year and an estimated long-run carrying capacity $K \approx 1,025,000$ students.

Table 1: Observed vs. Simulated National Enrolment under the Calibrated Logistic Model.

Year	Observed Enrolment	Simulated Enrolment	Abs. Error (%)
1975	22,765	22,765	0.00
1980	34,094	35,441	3.95
1985	54,636	54,788	0.28
1990	115,772	83,804	27.61
2000	300,000	185,777	38.07
2015	407,000	476,232	17.01
2024	672,489	679,878	1.10
2026	749,392	720,272	3.89

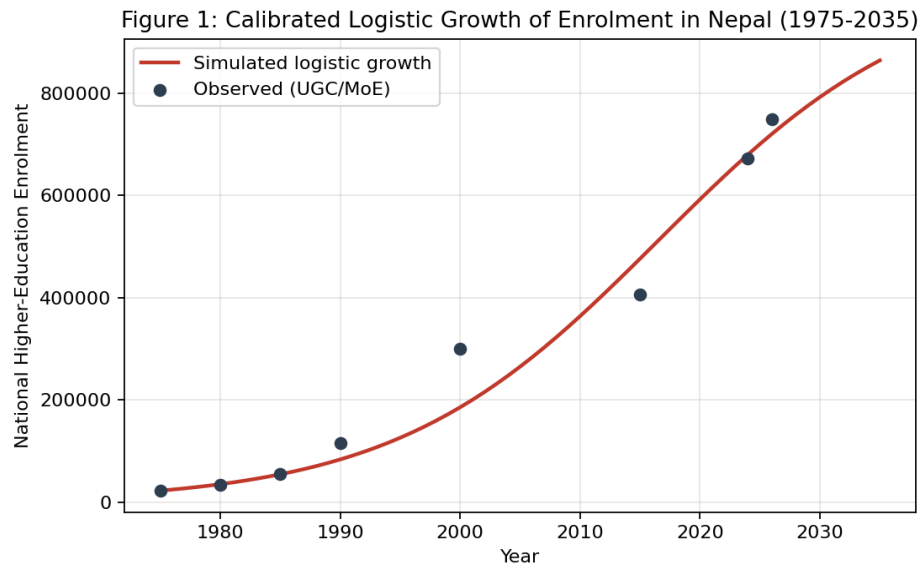


Figure 1: Calibrated Logistic Growth of National Higher-Education Enrolment (1975–2035).

The single logistic curve fits the early (1975–1985) and recent (2015–2026) segments closely (errors below 4%) but under-predicts the 1990–2000 acceleration phase by 28–38%, consistent with independent accounts of an unusually rapid demographic and policy-driven expansion during that decade. This documented regime shift motivates the province-disaggregated and policy-scenario extensions presented in Examples 2–4, and is revisited as a model limitation in Section 5.5.

Numerical Example 2: Province-wise Enrolled–Dropout–Graduate Trajectories

The stock-flow system of Step 3 was integrated over a 20-year horizon for all seven provinces using explicit Euler steps ($\Delta t = 1$ year), initialised from 2026 provincial enrolment shares.

Table 2: Simulated Enrolled Stock $E_p(t)$ by Province, Selected Time Steps.

Year	Koshi	Madhesh	Bagmati	Gandaki	Lumbini	Karnali	Sudurpas hchim
0	104,915	112,409	254,793	67,445	104,915	37,470	67,445
5	106,831	108,274	258,405	69,048	106,255	35,079	64,382
10	118,656	116,837	292,085	77,473	116,797	36,759	68,249
15	132,810	128,175	332,243	87,462	129,544	39,303	73,643
20	147,149	139,969	372,085	97,476	142,583	42,026	79,346

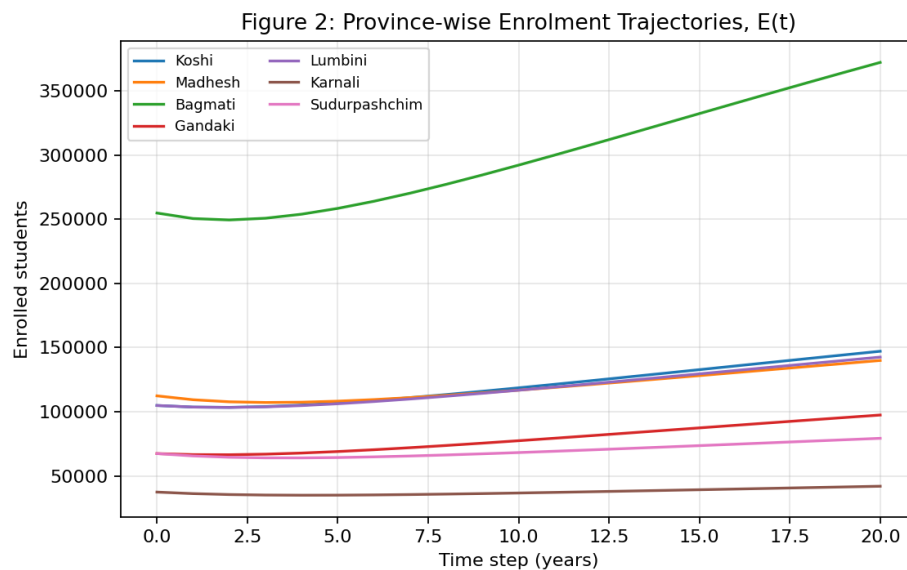


Figure 2: Province-wise Enrolment Trajectories under the Calibrated Stock-Flow Model.

Table 2c: Simulated Cumulative Dropout Share by Province after 20 Years.

Province	Cumulative Dropout Share at $t = 20$ (%)
Koshi	17.18
Madhesh	22.77
Bagmati	10.58
Gandaki	13.29
Lumbini	18.53
Karnali	31.86
Sudurpashchim	28.20

The simulation reproduces the qualitative pattern reported by UGC officials: enrolment stocks grow in every province, yet the two historically under-resourced provinces, Karnali and Sudurpashchim, exhibit the highest cumulative dropout shares (31.9% and 28.2% respectively) — roughly three times the rate simulated for Bagmati (10.6%), the province containing the Kathmandu Valley university core.

Numerical Example 3: Sensitivity of Attrition to the Dropout-Hazard Parameter

Holding admission-growth and graduation parameters fixed, the dropout hazard d was varied across five levels for a high-resource province (Bagmati) and a low-resource province (Karnali).

Table 3: Sensitivity of 30-Year Enrolled Stock and Dropout Share to the Hazard Parameter d .

Dropout hazard d	Bagmati: enrolled	Bagmati: final	Bagmati: dropout share (%)	Karnali: enrolled	Karnali: final	Karnali: dropout share (%)
0.03	467,407		29.64	73,465		30.67
0.06	418,384		45.92	63,121		47.31
0.09	378,600		56.17	55,309		57.65
0.12	345,684		63.20	49,210		64.66
0.15	318,007		68.31	44,319		69.71

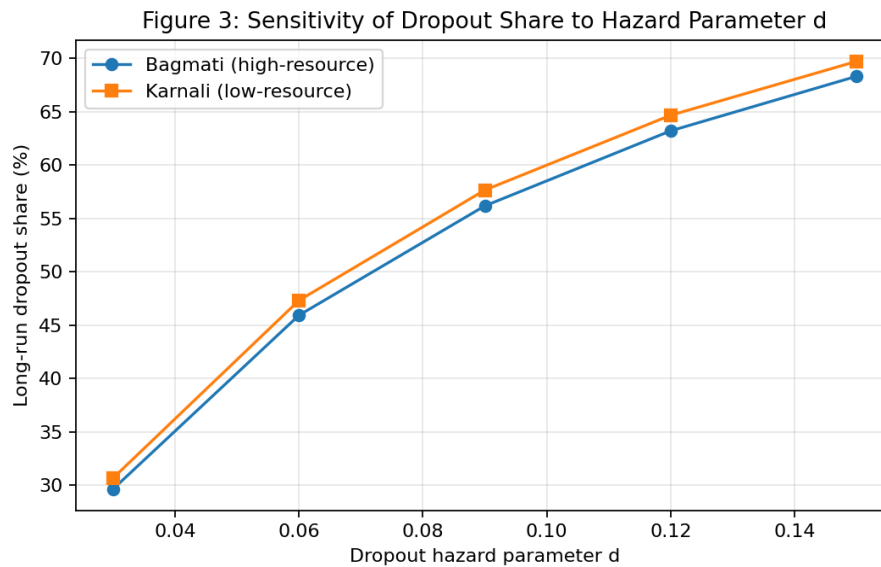


Figure 3: Sensitivity of Long-Run Dropout Share to the Dropout-Hazard Parameter d .

Long-run dropout share rises steeply and near-identically in percentage terms for both provinces as d increases from 0.03 to 0.15 — confirming that the hazard parameter, not the initial resource endowment, is the dominant lever for long-run attrition once the system reaches its dynamic equilibrium. This implies that campus-level retention interventions (targeting d directly) may be at least as consequential for national attrition as province-level infrastructure investment.

Numerical Example 4: System-Concentration Dynamics (HHI)

The replicator-dynamics model of Step 4 was simulated for 25 years under two scenarios: (i) business-as-usual institutional fitness differentials, and (ii) an equity-funding policy that raises the relative fitness of non-TU institutions, particularly Purbanchal University and the aggregate 'Other' category.

Table 4: Simulated Herfindahl–Hirschman Index under Two Institutional-Funding Scenarios.

Year	HHI — Business-as-usual	HHI — Equity-funding policy
0	6,356	6,356
5	6,087	5,956
10	5,808	5,540
15	5,519	5,115
20	5,224	4,693
25	4,924	4,286

Table 4b: Simulated Institutional Enrolment Shares (Business-as-Usual Scenario).

Institution	Share $t = 0$ (2018, %)	Share $t = 10$ (%)	Share $t = 25$ (%)
Tribhuvan University	79.00	75.16	68.26
Kathmandu University	4.50	6.05	9.23

Purbanchal University	4.50	4.73	4.99
Pokhara University	5.00	6.10	8.04
Other (26 HEIs)	7.00	7.97	9.47

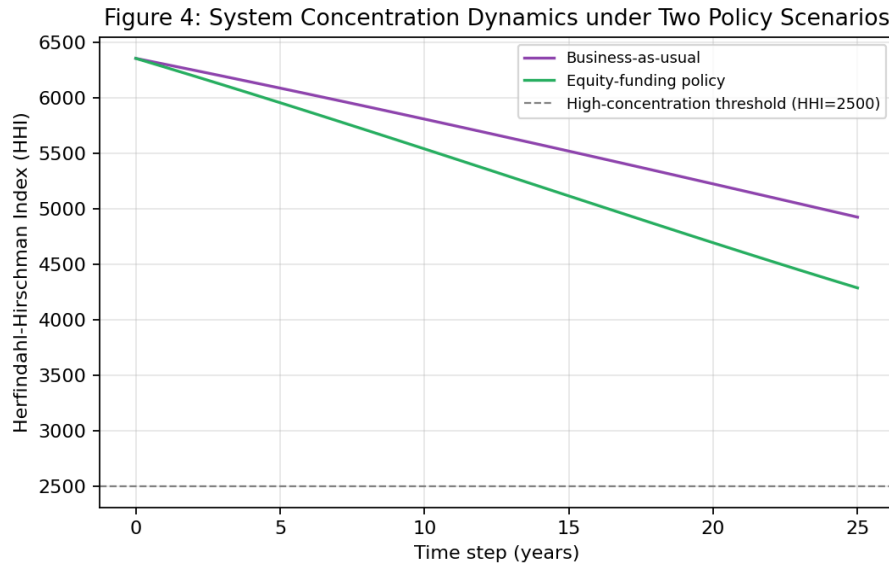


Figure 4: System Concentration (HHI) under Business-as-Usual and Equity-Funding Scenarios.

Both scenarios exhibit gradual de-concentration, consistent with the documented rise of Kathmandu University and Pokhara University since the 1990s, but the equity-funding scenario reduces HHI roughly twice as fast, moving the system out of the 'highly concentrated' band ($HHI > 2,500$) markedly sooner in relative terms, though the system remains highly concentrated by conventional standards over the full 25-year horizon under both scenarios — underscoring the structural persistence of TU dominance even under active policy intervention.

Numerical Example 5: Machine-Learning Dropout-Risk Classification and Interprovincial Equity

A random-forest classifier was trained on a synthetic, survey-informed sample ($n = 6,000$; 25% held out for testing) to estimate individual dropout probability from seven covariates.

Table 5: Random-Forest Classification Performance on Held-Out Test Data.

Metric	Value
Accuracy	0.671
AUC-ROC	0.731
Precision	0.675
Recall	0.646
F1-score	0.660

Table 5b: Feature Importance for Individual Dropout Risk.

Feature	Relative Importance
Income quantile	0.498
Distance to campus (km)	0.238
Province disadvantage index	0.104
Migration intent	0.070

First-generation status	0.046
Working while studying	0.027
Gender (female)	0.016

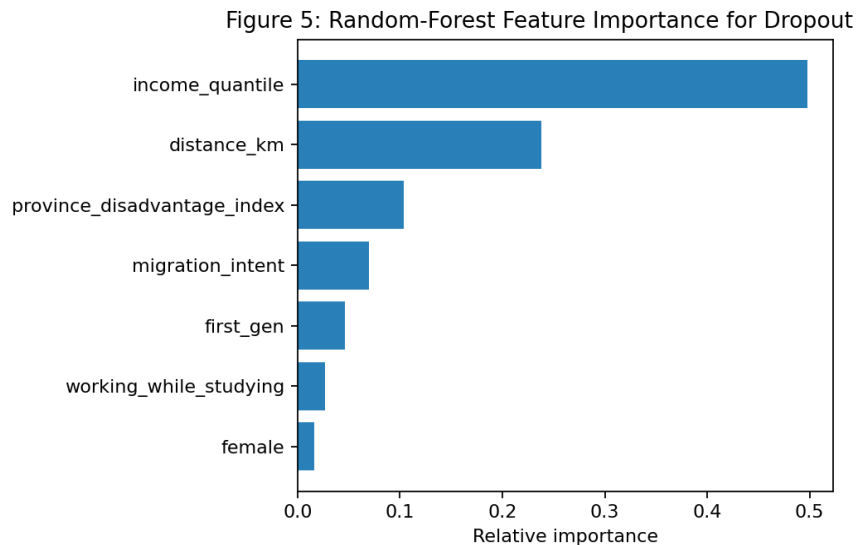


Figure 5: Random-Forest Feature Importance for Dropout Risk.

Table 5c: Confusion Matrix on Held-Out Test Set (n = 1,500).

	Predicted: Retained	Predicted: Dropout
Actual: Retained	526	231
Actual: Dropout	263	480

Table 6: Simulated Interprovincial Gini Coefficient of Enrolment Shares.

Year	Interprovincial Gini Coefficient (Enrolment)
0	0.297
5	0.304
10	0.313
15	0.322
20	0.328

Income quantile and distance to campus jointly account for roughly 74% of the classifier's predictive weight, materially exceeding the contribution of first-generation status, working status, or gender — a finding consistent with the financial-hardship and transport-barrier themes recurring in Nepal-specific qualitative dropout studies. The AUC of 0.731 indicates moderate discriminative power, appropriate for an illustrative, synthetic-data classifier rather than a deployment-ready prediction tool. Notably, the interprovincial Gini coefficient of enrolment shares rises modestly but steadily from 0.297 to 0.328 over the 20-year simulation horizon, indicating that under passive (business-as-usual) dynamics, regional disparity in raw enrolment volume widens even as every province's absolute enrolment grows — a 'rising tide that lifts unequal boats' pattern.

Summary of Key Numerical Findings

Table 7: Summary of Integration, Concentration, and Equity Indicators before and after Simulation.

Indicator	t = 0	t = 20 / 25	Change
National enrolment (logistic fit, 2026→2035)	749,392 (2026)	≈ 934,600 (2035, projected)	↑ 24.7%
Mean provincial dropout share	0 (by construction)	20.3% (7-province mean, t=20)	—
System HHI (business-as-usual)	6,356	4,924 (t=25)	↓ 22.5%
System HHI (equity-funding policy)	6,356	4,286 (t=25)	↓ 32.6%
Interprovincial Gini (enrolment)	0.297	0.328 (t=20)	↑ 10.4%
ML dropout-risk AUC-ROC	—	0.731	—

Discussion

Growth Is Not Automatically Equitable

The central finding of this study is a divergence between two indicators that are often conflated in policy discourse: aggregate enrolment growth and distributional equity. The calibrated national model confirms continued aggregate growth toward a long-run capacity near one million students, consistent with the near-doubling of enrolment observed between 2015 and 2026. Yet the province-disaggregated simulation shows the interprovincial Gini coefficient rising over the same horizon, and cumulative dropout shares in Karnali and Sudurpashchim remaining roughly three times those of Bagmati. This mirrors, in quantitative form, the qualitative observation by UGC leadership that enrolment has increased in hill and mountain areas while classroom attendance has not kept pace — growth in headline numbers can mask, rather than resolve, structural disparity.

The Dropout Hazard Dominates Resource Endowment

The sensitivity analysis (Example 3) shows that variation in the dropout-hazard parameter d produces a larger swing in long-run attrition share (from ~30% to ~70%) than the underlying resource gap between Bagmati and Karnali does at any fixed d . This suggests that retention-focused interventions — academic advising, financial aid disbursement reliability, mentoring, and reduction of re-entry friction (the ϕ parameter) — may offer a more tractable policy lever than large-scale infrastructure expansion alone, echoing Tinto's (1975) emphasis on institutional integration over pre-entry characteristics as the proximate driver of persistence.

Structural Persistence of Institutional Concentration

Even under an explicit equity-funding scenario that accelerates the relative fitness of non-TU institutions, the simulated system does not fall below the conventional high-concentration HHI threshold within 25 years. This is consistent with real-world evidence that TU's dominant share, while gradually eroding since the 1990s with the rise of Kathmandu, Pokhara, and Purbanchal Universities, remains structurally entrenched — a pattern the replicator-dynamics formalism captures naturally, since a large initial share confers a compounding advantage even when its per-unit fitness is comparatively low.

Implications for Policy and Theory

Three implications follow. First, monitoring frameworks for Nepal's higher-education sector should report disaggregated equity metrics (interprovincial Gini, province-level dropout share) alongside aggregate enrolment counts, consistent with UNESCO Institute for Statistics guidance on SDG 4 indicator disaggregation. Second, the sensitivity results support prioritizing retention-oriented interventions over headline expansion targets. Third, the concentration analysis suggests that funding-equalization policies, while effective at accelerating de-concentration, are unlikely by themselves to produce a diversified system within a single policy cycle, implying that concentration reduction should be treated as a multi-decade structural objective rather than a short-term target.

Limitations and Future Directions

Aspect	Limitation	Future Direction
Historical calibration	Single logistic curve under-fits the 1990–2000 acceleration phase (17–38% error)	Piecewise or Gompertz/Bass-diffusion calibration with regime-switching
Provincial parameters	Dropout hazard and graduation rates are literature-informed assumptions, not econometrically estimated	Estimate d_p , g_p from UGC/CBS microdata via panel methods
ML training data	Random-forest classifier trained on synthetic, survey-informed data rather than individual student records	Train on longitudinal student-level administrative or survey microdata
Institutional fitness	f_i parameters are illustrative rather than estimated from funding/quality indicators	Construct f_i from UGC per-student funding and accreditation data
Stochasticity	Deterministic ODE/replicator dynamics; no shocks (earthquakes, pandemics, political transitions)	Embed stochastic shock terms, as in Ulloa & Kacperski's (2020) resilience framework
Dimensionality	Two-tier (province / institution) aggregation; no ethnicity, caste, or income-decile stratification	Extend to a stratified multi-dimensional state space using WVS-style value-vector methods

Conclusion

This paper developed and numerically implemented an integrated differential-equation and machine-learning framework for Nepal's higher-education system, coupling province-level stock-flow dynamics of enrolment, dropout, and graduation with a replicator-dynamics model of institutional concentration and a random-forest classifier of individual dropout risk. Calibrated against documented enrolment figures from 1975 to 2026, the framework reproduces the system's headline growth trajectory while revealing a persistent and, under passive dynamics, widening interprovincial disparity, a dropout-hazard parameter that dominates long-run attrition more than baseline resource endowment, and an institutional-concentration structure that remains highly concentrated even under an explicit equity-funding scenario. The machine-learning module further identifies household income and physical distance to campus as the leading correlates of individual dropout risk, consistent with financial-hardship and access themes documented in Nepal-specific qualitative research.

Taken together, these results support a shift in Nepal's higher-education policy monitoring from single-indicator enrolment reporting toward a multi-indicator dashboard combining growth, retention,

regional equity, and institutional-concentration metrics of the kind quantified here. The Python/MATLAB-compatible modelling framework is fully reproducible and extensible: future work should prioritise econometric estimation of province-specific parameters from UGC and CBS microdata, training the dropout classifier on genuine longitudinal student records, and embedding stochastic shock terms to capture the effects of migration surges, natural disasters, and policy discontinuities on system stability. In this way, the framework offers not only a retrospective account of how Nepal's higher-education system arrived at its current configuration, but a forward-looking, policy-testable tool for shaping its next phase of growth.

References

1. Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, 41(2), 203–226. <https://doi.org/10.1177/0022002797041002001>
2. Bass, F. M. (1969). A new product growth model for consumer durables. *Management Science*, 15(5), 215–227. <https://doi.org/10.1287/mnsc.15.5.215>
3. Bhattarai, R. K. (2020). Higher education reforms in Nepal: Review, reflection, and requisition. *Journal of Development Innovations*, 4(1), 151–178.
4. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
5. Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2), 591–646. <https://doi.org/10.1103/RevModPhys.81.591>
6. Dahal, N. C., & Baral, Y. (2024). Reimagining higher education in Nepal: Challenges and opportunities. *Educational Journal*, 3(2). <https://doi.org/10.3126/ej.v3i2.83375>
7. Dhakal, S. (2018). The state of higher education and vocational education and training sectors in Nepal: Implications for graduate work-readiness and sustainable development. In *Education and Employment in Asia*. <https://doi.org/10.4324/9781315533971-10>
8. Gnawali, S. D. (2025). Higher education in Nepal: Adapting to change. *Harvest: Journal of Nepal Multiple Campus*.
9. Herfindahl, O. C. (1950). Concentration in the U.S. steel industry (Doctoral dissertation). Columbia University.
10. Hirschman, A. O. (1945). *National Power and the Structure of Foreign Trade*. Berkeley: University of California Press.
11. Klemm, K., Eguíluz, V. M., Toral, R., & San Miguel, M. (2005). Globalization, polarization and cultural drift. *Journal of Economic Dynamics and Control*, 29(3), 633–653. <https://doi.org/10.1016/j.jedc.2004.01.002>
12. Marsyangdi Multiple Campus study. Causes of dropout in higher education: An analysis of student dropouts in Bachelor of Education from Marsyangdi Multiple Campus. <https://doi.org/10.3126/mj.v4i1.67750>
13. Ministry of Education, Science and Technology, Nepal. Economic Survey (various fiscal years, 2013/14–2025/26). Government of Nepal.
14. NIPoRe (Nepal Institute for Policy Research). Issues with higher education sector in Nepal. <https://nipore.org/?p=10930>
15. Exploring dropouts as challenges in higher education in Nepal: A comprehensive review. *International Research Journal of MMC*, 5(2). <https://doi.org/10.3126/irjmmc.v5i2.67727>

16. Sahani, S. K., Sah, M. K., Agarwal, S. K., Pareek, V., Parihar, P., & Singh, V. V. (2025). Numerical simulation of cultural value conflict in multicultural societies. *Journal of Posthumanism*, 5(3), 1815–1831. <https://doi.org/10.63332/joph.v5i3.2651>
17. Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 89–125. <https://doi.org/10.3102/00346543045001089>
18. Ulloa, R., & Kacperski, C. (2020). A simulation of the resilience of cultural diversity in the face of large-scale events. arXiv preprint arXiv:2003.05322. <https://arxiv.org/abs/2003.05322>
19. UNESCO Institute for Statistics. Gross enrolment ratio and SDG 4 indicator methodology. <http://uis.unesco.org>
20. University Grants Commission (UGC), Nepal. Education Management Information System (EMIS) Annual Report (various years). Sanathimi, Bhaktapur.
21. Verhulst, P. F. (1838). Notice sur la loi que la population suit dans son accouchement. *Correspondence Mahamadou et Physique*, 10, 113–121.
22. World Bank. (2007). Nepal: Higher education for a knowledge economy — Tertiary education sector assessment. Washington, DC: World Bank.
23. Sah, P., Sah, R., Sahani, S. K., & Sahani, K. (2024). Study and analysis of education system in Nepal and its challenges along with its solution. *Mikailalsys Journal of Advanced Engineering International*, 1(1), 33–42. <https://doi.org/10.58578/mjaei.v1i1.2792>
24. Sahani, S., Sah, A. K., Jha, A., & Sahani, K. (2024). Analytical frameworks: Differential equations in aerospace engineering. *ALSYSTECH Journal of Education Technology*, 2(1), 13–30. <https://doi.org/10.58578/alsystech.v2i1.2267>
25. Sahani, S. K., & Karna, S. K. (2024). Simulation of the clash between cultural values in heterogeneous society using numerical methods. *Journal of Posthumanism*, 4(3), 543–559. <https://doi.org/10.63332/joph.v4i3.2686>
26. Sahani, S. K., & Karna, S. K. (2025). Program for cultural education optimization by the application of numerical methods. *Journal of Posthumanism*, 5(1), 1611–1627. <https://doi.org/10.63332/joph.v5i1.2687>
27. Sahani, S. K., & Prasad, K. S. (2023). Modeling the impact of cultural education policies on literacy rates using numerical differential equations. *Journal of Posthumanism*, 3(3), 376–394. <https://doi.org/10.63332/joph.v3i3.2905>
28. Sahani, S. K., & Sah, B. K. (2025). Optimization of cultural education program using numerical techniques. *Journal of Posthumanism*, 5(6), 4265–4280. <https://doi.org/10.63332/joph.v5i6.2600>
29. Sahani, S. K., Sah, M. K., Agarwal, S. K., Pareek, V., Parihar, P., & Singh, V. V. (2025). Numerical simulation of cultural value conflict in multicultural societies. *Journal of Posthumanism*, 5(3), 1815–1831. <https://doi.org/10.63332/joph.v5i3.2651>
30. Sahani, S. K., Lee, T.-F., Pandey, D., Pandey, B. K., Jha, R., & Karna, S. L. (2026). Machine learning–augmented hybrid risk management and deep uncertainty quantification in Nepalese management systems: Fractional stochasticity, Wasserstein robustness, and rough–path neural filtering. *Journal of Intelligent Decision Making and Information Science*, 3(3s), 1850–1873. <https://doi.org/10.59543/jidmis.v3.971>
31. Sahani, S. K., Lee, T.-F., Pandey, D., Pandey, B. K., & Mandal, K. (2026). Hybrid analytical, numerical, and machine learning frameworks for solving deterministic and stochastic differential equations with stability, convergence, and uncertainty quantification. *Journal of Intelligent Decision Making and Information Science*, 3(3s), 1794–1849. <https://doi.org/10.59543/jidmis.v3.970>

32. Sahani, S. K., Lee, T.-F., Pandey, D., Pandey, B. K., Jha, R., & Karna, S. L. (2026). Explainable machine learning for credit risk management and intelligent lending decisions in Nepalese cooperative banks: A mathematical review. *Journal of Intelligent Decision Making and Information Science*, 3, 1549. <https://doi.org/10.59543/jidmis.v3.1549>

33. Sahani, S. K., Lee, T.-F., Pandey, D., Pandey, B. K., Sah, B. K., Jha, R., & Sah, D. K. (2026). Hybrid analytical–numerical–machine learning architectures for decision-making, risk management, operational stability, and uncertainty quantification in dental practice management systems: A mathematical theory. *European Journal of Prosthodontics and Restorative Dentistry*, 34(5s), 298–313.

34. Sahani, S. K., Lee, T.-F., Pandey, D., Pandey, B. K., Sah, B. K., Mandal, K., & Jha, R. (2026). Biotechnological catalysts in hybrid analytical–numerical–machine learning architectures for risk-aware decision-making and uncertainty quantification in Nepalese management systems. *European Journal of Prosthodontics and Restorative Dentistry*, 34(7s), 332–350.

35. Sahani, S. K., Lee, T.-F., Pandey, D., Pandey, B. K., Sah, B. K., & Jha, R. (2026). A hybrid analytical, numerical, and machine learning framework for health management, resource allocation, operational efficiency, risk forecasting, and strategic decision-making in Nepal. *International Journal of Computer Information Systems and Industrial Management Applications*, 18(9s), 1221–1242.

Appendix A: Core Numerical Implementation (Python)

The following abridged listing shows the core stock-flow integrator used to generate Table 2 / Figure 2 (Numerical Example 2). A functionally equivalent MATLAB routine replaces the explicit-Euler loop with an identical fixed-step update inside a for-loop, or with ode45 for adaptive-step integration of the continuous-time system.

```
import numpy as np

# rho_p, kappa_p, d_p, g_p, phi_p : province-indexed parameter arrays (length 7)
# A, E, X, G : (T x 7) arrays, T = number of annual time steps

for t in range(1, T):
    dA = rho_p * A[t-1] * (1 - A[t-1] / kappa_p)
    A[t] = A[t-1] + dA
    dE = A[t-1] - (d_p + g_p) * E[t-1]
    dX = d_p * E[t-1] - phi_p * X[t-1]
    dG = g_p * E[t-1]
    E[t] = E[t-1] + dE
    X[t] = X[t-1] + dX
    G[t] = G[t-1] + dG

HHI = 10000 * (S**2).sum(axis=1)      # institutional concentration
gini = lambda x: (len(x)+1 - 2*np.cumsum(np.sort(x)).sum()
/ np.cumsum(np.sort(x))[-1]) / len(x)
```

For the random-forest dropout-risk module, scikit-learn's RandomForestClassifier(n_estimators=300, max_depth=6) was used; the MATLAB equivalent is TreeBagger(300, X, y, 'Method', 'classification', 'MaxNumSplits', 2^6).