Applying Artificial Intelligence Approach to Predict Students' Dropout And Performance in Higher Educational Institutions

A Mini Research Report Submitted to the Research Management Cell of Durgalaxmi Multiple Campus, Far Western University, Nepal

Submitted by

Dharma Raj Ojha

Faculty of Management

Durgalaxmi Multiple Campus

Submitted to

Research Management Cell

Durgalaxmi Multiple Campus

Far Western University, Nepal

January 2026

Declaration

I, Dharma Raj Ojha, hereby declare that the thesis entitled "Applying Artificial Intelligence Approach to Predict Students' Dropout and Performance in Higher Educational Institutions" is my own original work carried out under the supervision of [Supervisor Name, PhD] and submitted in partial fulfilment of the requirements for the research grant at Far Western University, Nepal.

I further declare that:

1. This thesis has not been submitted, either in whole or in part, for any other degree or diploma at this or any other institution.
2. All sources of information used in this thesis have been duly acknowledged and referenced in accordance with the prevailing citation norms.
3. The data and results presented in this thesis are genuine and have not been manipulated or fabricated in any manner.
4. Wherever published, unpublished, or other intellectual materials of others have been used, they have been clearly acknowledged in the text.

_____

Dharma Raj Ojha

Date: _____

Approval Letter

This Report entitled "Applying Artificial Intelligence Approach to Predict Students' Dropout and Performance in Higher Educational Institutions" submitted by Dharma Raj Ojha, Faculty of Management, Durgalaxmi Multiple Campus, Far Western University, has been examined and approved by the undersigned members of the Research Management Committee as fulfilling the requirements for the research grant.

_____

Harikrishna Kadayat

Head, Evaluation Committee

Date: _____

_____

Dr. Bishnu Bahadur Kathayat

Member, Evaluation Committee

Date: _____

_____

Giri Singh Bohara

Member, Evaluation Committee

Date: _____

Acknowledgements

First and foremost, I express my deepest gratitude to my thesis supervisor, [Supervisor Name, PhD], for the unwavering academic guidance, intellectual mentorship, and consistent encouragement throughout the entire duration of this research. The insightful feedback, meticulous attention to methodological rigour, and patience in navigating the complexities of interdisciplinary research have been invaluable to the completion of this thesis.

I extend my sincere thanks to the Research Management Cell (RMC) of Durgalaxmi Multiple Campus, Far Western University, for approving this research proposal and providing the institutional framework necessary for conducting this study. Special acknowledgements are due to the administration of Durgalaxmi Multiple Campus for granting access to the Student EMIS dataset, without which the empirical component of this thesis would not have been feasible.

I am deeply grateful to the members of my research Management committee, the internal and external examiners, and the faculty members of the Faculty of Management whose constructive criticisms, scholarly insights, and rigorous evaluations have greatly strengthened the quality and integrity of this work.

Finally, I owe an immeasurable debt of gratitude to my family, whose unconditional support, sacrifice, and encouragement have sustained me through the demands of doctoral research. This work is dedicated to all students in Nepal who face systemic barriers to completing their higher education, may this research contribute, however modestly, to a more equitable and data-informed educational landscape.

Abstract

In higher educational institutions (HEIs), dropout is a persistent and resource consuming issue in the world with severe impact in the developing countries such as Nepal. Although the existing Nepali studies are descriptive and qualitative in nature, they are unable to predict future events, this thesis builds on that by using eleven machine learning (ML) and deep learning (DL) algorithms to the institutional data of Durgalaxmi Multiple Campus, Far Western University. The goal is to create a model that can predict student dropout and academic performance that is interpretable and contextually valid. There were a total of 808 student records in a set of data from six undergraduate programmes (BA, BED, BBA, BBS, CSIT, BSC) were used. These features comprised demographic information, previous academic performance, student enrollment, number of subjects dropped, and subject-wise GPA, and the target variable was binary: dropout (1) and retained (0) students. The data was divided into two subsets: 80% of the data for training ($n = 646$) and 20% for testing ($n = 162$). A set of models were tested: Decision Tree, Random Forest, SVM, Logistic Regression, Naive Bayes, XGBoost, deep learning models: DNN, CNN, RNN, LSTM, BiLSTM. The accuracy, F1-score and AUC-ROC of Random Forest were 82.72%, 86.67% and 90.89% respectively, followed closely by XGBoost and Decision Tree. Generally, the performances of the DL models were suboptimal as the dataset used was small and sufficient generalization was not possible. The first-semester GPA and backlog counts were the most important variables in each feature importance analysis for predicting dropout. The results reveal the potential of using a ML-based predictive early-warning system in resource-limited Nepali HEIs. This work provides a replicable and scalable framework for educational data mining and actionable insights for institutional policies and strategies for student retention.

Keywords: Student Dropout Prediction, Machine Learning, Deep Learning, Educational Data Mining, Higher Education Institutions

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Higher education is the part of formal education that culminates and is the basis of socio-economic development in the country. Higher educational institutions' ability to keep students enrolled and ensure their achievement of learning outcomes is not only an institutional problem, but one of high national interest, especially in countries where success in school translates to success in developmental outcomes. One of the most resource intensive, ethically significant and institutionally disruptive concerns that face HEIs around the world is student dropout (Tinto, 1993; World Bank, 2018).

The continuous evolution of learning analytics, the increasing accessibility of educational data, and the developments of computational intelligence in recent decades has brought new challenges and opportunities to HEIs for making the shift from reactive to proactive student attrition management. Machine learning (ML) and deep learning (DL) provide methodological tools to exploit institutional data and make it predictive intelligence for detecting at-risk students and designing tailored retention interventions (Romero & Ventura, 2020; Albreiki et al., 2021). This report lies on the confluence of these developments, using cutting-edge predictive modelling techniques to data from Durgalaxmi Multiple Campus at Far Western University (FWU), Nepal, to create a context-specific and meaningful early-warning system for student drop-out.

Background of the Study

Global Education and the Student Dropout Challenge

International English Language Day has recently been designated on the 28th of April each year to celebrate English, a global language, and to commemorate the anniversary of its adoption as the official language of the United Nations.In the last 30 years, higher education enrolment has increased significantly worldwide; UNESCO estimates that enrolment in tertiary education exceeded 235 million in 2020, compared with approximately 100 million in 2000 (UNESCO, 2021). However, this higher education "massification" has not been matched by a corresponding increase in student graduation rates. According to OECD (2020), around one in three students who start a bachelor's degree programme in the OECD countries does not finish. This means that there is considerable waste of public investments and personal potential.

The student drop-out phenomenon is multidimensional which involves academic, financial, social, institutional and psychological factors. Theoretical research by Tinto (1975, 1987) provided a building block perspective for student departure, which suggested that the educational climate in which students enroll and their initial

orientation to education goals shape their decisions to leave or persist in school. Bean (1980) expanded this model to include organisational determinants and Pascarella and Terenzini (1991) supplied a wealth of empirical evidence for the importance of institutional quality and student-faculty interactions for persistence outcomes. Recent studies have adopted psychological constructs as moderators of the risk for dropout, such as self-efficacy (Chemers et al., 2001), sense of belonging (Hausmann et al., 2007), and academic emotions (Pekrun, 2011).

The impact of student dropouts is significant on the economic level. The National Student Clearinghouse Research Center (2019) estimates that 36.2% of students between the ages of 25 and 29 between 2011 and 2016 were non-enrollees, which means they were not enrolled in any college or university during that period. These 36.2% of students reported having lower lifetime earnings, higher unemployment rates and reduced access to health care and social mobility pathways. Governments lose investment in education, and their pool of skilled people who are needed for knowledge economy competitiveness (OECD, 2020) with every dropout. In the U.S. alone the cost of the dropout problem is estimated to exceed $3.8 billion per year in lost tax revenues and costs to the social safety net (Alliance for Excellent Education, 2014).

Given the extent and impact of dropouts, student retention has become a key indicator of higher education system performance on the international policy agenda. Two trends that have grown increasingly prominent in the field of learning analytics are the operationalization of interventions for retention (Baker & Siemens, 2014; Siemens, 2013), and the development of learning analytics and educational data mining (EDM) (Weller, 2014). Using various statistical modelling, machine learning and data visualization techniques, EDM analyses vast amounts of educational data, such as student information systems, learning management system (LMS) records and administrative records, to discover actionable insights. Predictive modelling has been proven to improve retention at HEIs in North America, Europe, Australia and more recently in Asia (Arnold & Pistilli, 2012; Jayaprakash et al., 2014).

Nepali Education System: Context and Challenges

Since the promulgation of the Education Act 1971 and the subsequent National Education System Plan, there have been significant changes in the structure of the education system of Nepal. Formal higher education in Nepal began with the establishment of Tribhuvan University (TU) in 1959. Since then, the sector has expanded significantly, with the formation of several Universities like Pokhara University (1997),

Kathmandu University (1991), Purbanchal University (1994), Far Western University (2010) and other universities. During the academic year 2022-23, there were 1420 colleges with an enrollment of over 400,000 students in higher education programmes in Nepal, corresponding to a gross tertiary enrolment ratio of around 18% (MoEST, 2023).

Despite this growth, there are still many structural issues to be addressed in the higher education sector in Nepal such as poor infrastructure, inequalities, low faculty-to-student ratio, lack of research capacity, and significantly, a high dropout rate among students. The University Grants Commission (UGC) of Nepal (2022) found that dropout rates in the HEIs of Nepal are between 20% to 40% based on the type of HEI and geographical area, as well as the programme of study. In Far Western Nepal, where geographic isolation, poverty, and inadequate educational facilities are prevalent, dropout rates are reportedly high compared to the rest of Nepal (Madai et al., 2025; Joshi, 2024).

Among the various factors affecting dropouts in the Nepali context, financial difficulties, family responsibilities, academic under-preparation, institutional support services, geographic barriers and socio-cultural factors are the most significant, especially for girls (Devkota & Giri, 2020; Neupane, 2024; Subedi, 2022; Subedi, 2024). Institutions seldom are able to identify at-risk students through systematic, data-driven early-warning mechanisms, and it is too late to provide remedial action when students drop out. The reactive approach does not align with global good practice in the management of student retention, which is based on early identification, personalised interventions, and ongoing monitoring (Heublein, 2014; Kehm et al., 2019).

The HEIs in Nepal, including community campuses of far-western universities, have inadequate digital infrastructure. Student records are also typically kept on a semi-digital or paper-based basis, making it challenging to have a full-scale, longitudinal dataset that can be analyzed. The present study has been conducted in a Student Education Management Information System (EMIS) that was recently operationalised in Durgalaxmi Multiple Campus which is one of the few Far-Western Nepalese institutes with structured and machine-readable student data that can be utilised for predictive modelling. The current institutional setting provides the study with both timeliness and a unique opportunity to make a contribution to the emerging evidence base on student retention in Nepal through the use of AI.

Machine Learning and Deep Learning in Educational Contexts

Machine learning (ML) is a type of computational algorithms that allows a system to learn from data and get better at a particular task without being explicitly programmed (Mitchell, 1997; Alpaydin, 2020). Within the education field, ML has been used in many applications, such as estimating student performance, predicting student dropouts, developing intelligent tutoring systems, creating automated test assessments, and recommending personalised learning (Romero & Ventura, 2020; Peña-Ayala, 2014). Decision Trees, Logistic Regression, Support Vector Machines (SVM), and Naive Bayes, are among the most commonly employed supervised learning algorithms in educational data mining, and are highly interpretable with a satisfactory level of predictive performance on structured institutional data (Kabathova & Drlik, 2021; Kemper et al., 2020).

To improve the predictive performance of classification models, ensemble methods, especially Random Forest (Breiman, 2001) and gradient boosting algorithms like XGBoost (Chen & Guestrin, 2016), use the consensus predictions of multiple base learners to reduce the variance and enhance the generalisation of the model. They are robust to class imbalance, can deal with heterogeneous features and come with an in-built feature importance ranking (Realinho et al., 2022; Christou et al., 2023), and are thus conventional benchmarks in dropout prediction research.

Deep learning (DL) is one of the subfields of machine learning that uses multiple hidden layers of artificial neural networks that can automatically learn hierarchical representations from raw data (LeCun et al., 2015; Goodfellow et al., 2016). Educational prediction related deep learning models include Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997), and Bidirectional LSTM (BiLSTM) models. LSTM networks are well suited to modeling temporal dependencies in sequence data, such as academic performances over a semester, which make them ideal for this type of problem and are theoretically attractive for this application of the modelling of the longitudinal dynamics of student dropout (Fu et al., 2021; Adefemi et al., 2025).

Deep learning models are, however, data intensive and compute-intensive, and are usually trained on large, well-structured datasets to generalize well. For such HEI settings, where the amount of data is small, empirical results of the present work indicate that a traditional ML model can outperform deep architectures. Comparative evaluation

of ML and DL models under realistic institutional data conditions is one of the main methodological contributions within the scope of this report.

Statement of the Problem

The student dropout phenomenon in the Nepali HEIs is well documented which has not been sufficiently tackled by using predictive modelling methods. Although descriptive statistics, cross sectional surveys and qualitative inquiry are important to provide the context of the dropouts, existing studies in Nepal are limited in their ability to predict individual dropout risks with sufficient accuracy to inform targeted early interventions (Devkota & Giri, 2020; Subedi, 2022; Neupane, 2024). Most HEIs in Nepal are still reactive in their approach to student retention because the institutionalized early warning system is based on machine learning is still absent.

Globally, machine learning and deep learning strategies have shown significant potential in accurately predicting dropouts at a high level, however most studies have taken place in the context of large, well funded institutions with a strong digital learning infrastructure in developed countries (Kabathova & Drlik, 2021; Albreiki et al., 2021; Christou et al., 2023). These models tend to be highly feature-rich, such as LMS engagement logs, real-time attendance data, and psychometric evaluations, aspects that are not universally accessible in institutional settings in Nepal. Context-specific predictive models that are built to work in the context of Nepali HEIs are therefore urgently needed to accommodate the specific data and resource constraints that are typical of the HEIs in Nepal.

In addition, despite the many complex architectures recently introduced into deep learning literature, the comparative usefulness of such architectures compared to classical ML models in the small to medium dataset regime has not been properly evaluated. Interpretability (understand and explain the predictions of the model to institutional stakeholders) is also under-investigated, as it is a pre-requisite to convert the predictive outputs into actionable institutional policies (Kruger et al., 2023; Lakkaraju et al., 2015).

The present study aims to solve these inter-related research problems by: (a) creating and testing both ML and DL predictive models with institutional data from a Nepali HEI; (b) comparing the predictive performance under realistic data condition of 11 different algorithms; (c) using feature importance analysis to find the most important predicting factors for student dropouts; (d) suggesting appropriate recommendations to implement the early-warning system in the HEI under study.

Research Gap

Based on the literature search, the following important research gaps were identified and will be addressed in this study:

Lack of Predictive Modelling in Nepali HEIs: Most of the literature in the world is based on ML for predicting student dropout rates, whereas, the research in Nepal has been conducted in a qualitative methodology with surveys which eventually cannot provide any individual-level actionable prediction.

Lack of ML vs. DL Evaluation in Resource-Constrained Contexts*:* The existing studies concentrate on the evaluation of ML and DL models on large datasets. Rigorous, comparative assessments are lacking, however, comparing 11 models under reasonable small to medium dataset constraints relevant to the standards in developing country institutions.

Ignorance of Model Interpretability (XAI)*:* In the era of responsible utilization of AI in high-stake educational contexts, research that combines the concepts of interpretability and dropout prediction is scarce in the realm of Higher Education (HE) in South Asia and Nepal.

Previous regional studies have concentrated on any one department or programme in isolation.

This study extends this research limitation by adopting a multi-programme data set from six undergraduate programmes to enhance the representativeness and generalizability of the findings at institutional level.

Research Questions

The following primary research questions guide this investigation:

The following questions are established to fulfill the objectives:

1. How effectively can machine learning and deep learning models predict student dropout and academic performance in higher education institutions?
2. Which approach provides better predictive accuracy and reliability for student dropout and academic performance?

Research Objectives

The primary objective of this study is to apply machine learning and deep learning techniques to institutional data to develop an interpretable predictive model capable of identifying at-risk students in a Nepali higher education context.

1. To develop and evaluate machine learning and deep learning models for predicting student dropout and academic performance in higher education.

2. To compare the performance of machine learning and deep learning approaches in terms of accuracy, precision, recall, and overall predictive effectiveness.

Hypotheses

Throughout this investigation the following research hypotheses are tested:

H1: Academic trajectory indicators are the key predictive features in the analysis and machine learning and deep learning models are found effective in predicting student dropout and academic performance through the early institutional data.

H2: Ensemble-based traditional machine learning algorithms (e.g. Random Forest and XGBoost) exhibit superior predictive accuracy, reliability and interpretability when compared to deep learning architectures for small to medium structured institutional datasets in resource-constrained HEIs.

Significance of the Study

This report is of importance at the theoretical, methodological and practical level: Theoretically, the study operationalizes the constructs of the Student Integration Model (SIM) (Tinto, 1975; 1987) by assigning quantitative data features (GPA, backlog, programme type) to them, and then tests them for predictive validity using machine learning. This is an improvement over the use of the survey approach in the application of Tinto's framework that was widely used in the Nepali literature.

The methodological approach and analysis of the study is one of the first systematic and comprehensive comparative assessment of eleven MLs and DLs algorithm for dropout prediction based on institutional data from a Nepali HEI. The methodological robust experimental design, with stratified train-test split, standardised feature preprocessing, and multi-metric evaluation, provides a replicable template for future educational data mining research for educational setting in Nepal and other educational setting in the South Asian region.

The study practically gives an evidence-based ground for the implementation of Durgalaxmi Multiple Campus and similar institutions with regard to the data-driven approach to early-warning system. The results that first-semester GPA and the number of courses backlogged represent the two most significant predictors of dropouts have implications for academic advising, institutional support services, and resource allocation, with targeted, temporally early intervention. The research also provides guidance to the national education policy, as it illustrates the potential of implementing a retention management system using ML in resource-limited contexts, which is part of MoEST's overall vision to improve higher education quality and equity.

Scope and Limitations

The scope of this study is set by the following: (a) Data is limited to 808 student records from 6 undergraduate programmes across Durgalaxmi Multiple Campus, Far Western University for the academic years of 2020-21; (b) The predictive models are developed using demographic, academic and enrollment data provided by the Student EMIS system; (c) The predictive models are compared with each other using eleven (11) ML and DL algorithms; and (d) The target variable is binary (dropout or retained).

There are a number of caveats that should be noted. First, the dataset size is 808 data points, which is not too large for classical ML models, but is not enough for complex deep learning architectures that generally need tens of thousands of training data points in order to obtain stable generalisation. This may be the reason why the CNN, LSTM and BiLSTM models underperformed as seen in the experimental results. Second, the feature set is not as rich as similar international studies due to the lack of LMS engagement data, attendance data, and psychometric measures. Third, due to the cross-sectional design of the data, longitudinal models of dropout trajectories can not be constructed across several academic years. Fourth, results may not be extrapolated to other institutions, programme types or students in Nepal without further testing. Fifth, because of the class imbalance (66.7% dropouts, 33.3% retained), this may have an impact on the metrics the models achieved, which was mitigated in part by stratified sampling and choice of metrics.

Organisation of the report

This report is made up of 7 chapters. The introduction, background, problem statement, research gap, objectives, hypotheses, significance, and scope are outlined in Chapter One. Chapter Two outlines the literature review on student dropout prediction, including the theoretical background, ML and DL models in educational data mining and gaps in the literature. The research methodology is detailed in Chapter Three comprising of research design, dataset description, data preprocessing, model architectures and evaluation metrics. Chapter Four Implementation and Results gives the experimental setup and comparative performance results, confusion matrices, ROC curves, feature importance analysis and the training history of deep learning. The Discussion, interpretation of results, comparison with previous works, and elaboration of implications are discussed in Chapter Five. The Course Conclusion and recommendations are presented in Chapter Six, which summarises the findings, presents

conclusions and suggests directions for future research. The thesis ends with an extensive section on References along with relevant appendices.

Chapter 2: Literature Review

This chapter presents a systematic and critical review of the scholarly literature relevant to student dropout prediction and academic performance modelling in higher educational institutions. The review is organised thematically, progressing from classical theoretical frameworks of student departure, through the development of educational data mining and learning analytics, to the specific application of machine learning and deep learning models in dropout prediction contexts. The chapter also reviews the Nepali higher education literature and concludes with an identification of the principal research gaps that this thesis seeks to address.

Theoretical Frameworks of Student Departure

The theoretical study of student dropout in higher education is anchored in Vincent Tinto's influential model of student departure, first articulated in 1975 and subsequently refined (Tinto, 1975, 1987, 1993). Tinto's model conceptualises dropout as a longitudinal process in which students continuously evaluate the congruence between their evolving goals and commitments and the academic and social experiences they encounter within the institution. Students who achieve high levels of academic integration, demonstrated through grade performance, intellectual development, and faculty interaction, and high levels of social integration, manifested through peer relationships, extracurricular involvement, and sense of community, are theorised to exhibit greater institutional commitment and lower dropout propensity.

Tinto's model has been widely validated across diverse institutional contexts. Sweet (1986) extended its applicability to distance learning environments, finding that retention mechanisms operative in traditional campus settings also functioned in non-traditional educational modalities. Pascarella and Terenzini (1991, 2005) synthesised decades of empirical research and affirmed the centrality of academic preparation, faculty mentorship, and peer interaction as determinants of persistence. Bean's (1980) competing model introduced a labour-market analogy, positioning student dropout as analogous to employee turnover, influenced by institutional satisfaction, organisational integration, and external environmental pull factors such as employment opportunities.

More recent theoretical developments have incorporated psychological constructs. Hausmann et al. (2007) demonstrated the predictive validity of sense of belonging as a retention factor, particularly for first-generation and minority students. Chemers et al. (2001) identified academic self-efficacy as a significant mediator of persistence, while Pekrun's (2011) control-value theory of academic emotions

illuminated how students' appraisals of academic controllability and value shape their engagement and dropout decisions. In the Nepali context, these theoretical frameworks have been operationalised through survey instruments measuring academic self-efficacy (Devkota & Giri, 2020), institutional support perceptions (Madai et al., 2025), and socio-cultural role expectations (Subedi, 2024).

The present study aligns with Tinto's theoretical framework by operationalising academic integration through measurable institutional data variables specifically semester GPA and backlog counts that serve as proxies for the degree to which students are successfully integrating into the academic demands of their programmes. Programme type, prior academic achievement, and age serve as indicators of students' initial commitments and pre-entry characteristics, consistent with the model's input stage.

Educational Data Mining and Learning Analytics

Educational data mining (EDM) is defined as the development of methods for exploring unique types of data that come from educational settings and using those methods to understand students and the settings in which they learn (Baker & Yacef, 2009, p. 3). As an interdisciplinary field drawing from statistics, machine learning, psychometrics, and cognitive science, EDM has grown substantially since 2008, driven by the proliferation of digital educational platforms and institutional information systems (Romero & Ventura, 2010, 2020).

Learning analytics (LA), a closely related field, focuses on the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs (Siemens, 2013, p. 1381). While EDM emphasises the development of predictive algorithms, LA places greater emphasis on actionable insights for educational stakeholders. Both fields share the objective of transforming raw institutional data into decision-support intelligence. Peña-Ayala (2014) provides a systematic review of the EDM literature, cataloguing 55 studies and identifying classification and prediction as the most prevalent analytical tasks.

Baker and Siemens (2014) compare and contrast EDM and LA, arguing that the two fields are complementary and increasingly convergent, with both seeking to leverage automated analysis of large educational datasets to improve learning outcomes. Arnold and Pistilli (2012) describe the implementation of the Course Signals system at Purdue University, a landmark LA application that used predictive modelling to identify at-risk students and generate personalised alerts, resulting in measurable improvements in

retention rates. Jayaprakash et al. (2014) evaluated open-source predictive modelling applications across multiple institutions and confirmed the technical feasibility of deploying ML-based early-warning systems at institutional scale.

Machine Learning Approaches in Student Dropout Prediction

The application of machine learning to student dropout prediction has been extensively documented in the literature since the early 2000s. Dekker et al. (2009) conducted one of the early influential studies, applying decision trees, Bayesian networks, and neural networks to engineering student data and demonstrating that early dropout prediction was feasible using demographic and first-semester academic data, with accuracies exceeding 75%. This foundational study established the methodological template for subsequent research.

Mduma et al. (2019) conducted a systematic review of the literature on dropout prediction in higher education, analysing 22 studies and finding that ensemble methods consistently outperformed single-classifier approaches. Random Forest and Gradient Boosting emerged as the most frequently identified best-performing algorithms, with accuracy rates ranging from 80% to 95% depending on dataset characteristics and feature sets. The review also highlighted the critical importance of feature engineering and preprocessing quality, particularly handling of class imbalance, as determinants of model performance.

Albreiki et al. (2021) reviewed 26 studies on ML-based student performance prediction and found that ensemble learning models, neural networks, and decision trees were the most widely used approaches. The review identified academic performance indicators — particularly GPA, course completion rates, and attendance records — as the most informative predictive features, consistent with Tinto's theoretical framework. Kabathova and Drlik (2021) applied eight classification algorithms including Logistic Regression, Decision Tree, Random Forest, Naive Bayes, SVM, k-Nearest Neighbour, and Multi-Layer Perceptron to student data from a Slovak university, finding that Random Forest achieved the highest AUC (0.857) and F1-score, while Logistic Regression provided the best interpretability-performance trade-off. The study also found that features derived from the first semester of study were the most predictive, supporting the hypothesis that early academic performance is a reliable indicator of dropout risk (Kabathova & Drlik, 2021).

Kemper et al. (2020) systematically reviewed applications of ML in dropout prediction, emphasising the importance of model interpretability for educational

stakeholders. The authors argued that while complex ensemble models may maximise predictive accuracy, simpler models such as decision trees and logistic regression offer superior interpretability and are more likely to be trusted and acted upon by institutional decision-makers. This tension between predictive accuracy and interpretability represents a recurring theme in the EDM literature and is directly relevant to the deployment context of the present study (Kemper et al., 2020).

Christou et al. (2023) applied ML algorithms to a dataset from a European university, achieving a best accuracy of 89.3% using an XGBoost classifier. The study emphasised the importance of hyperparameter tuning via grid search and cross-validation for maximising model performance, and reported that academic-progression features — particularly cumulative GPA and course withdrawal history — were the strongest predictors of dropout. The study also explored SHAP (SHapley Additive exPlanations) values for model explainability, contributing to the growing XAI literature in educational prediction (Christou et al., 2023).

Realinho et al. (2022) introduced the "Predict Students' Dropout and Academic Success" dataset, which has since become a standard benchmark in the EDM literature. The study applied multiple classifiers including Random Forest, XGBoost, and MLP to data from a Portuguese polytechnic institute, reporting best accuracy of 76.9% with XGBoost on a three-class problem (dropout, enrolled, graduate). The dataset, now publicly available on the UCI Machine Learning Repository, has been used in numerous subsequent comparative studies and provides useful benchmarks for contextualising the results of the present thesis (Realinho et al., 2022).

In the South Asian context, Ahmed (2024) applied ML algorithms to student data from Bangladeshi universities, reporting that demographic variables combined with first-semester academic data yielded predictive accuracy of up to 84% using Random Forest. The study underscored the importance of contextually relevant feature selection in developing-country HEI settings where standardised engagement data may be unavailable. Basnet et al. (2022) explored dropout prediction in the context of massive open online courses (MOOCs) relevant to Nepalese learners, finding that clickstream and engagement features enabled early dropout detection with over 80% accuracy using ensemble classifiers.

Deep Learning Approaches in Student Dropout Prediction

The application of deep learning to educational prediction has accelerated considerably since 2018, driven by advances in computational hardware, the availability

of larger educational datasets, and the demonstrated superiority of DL architectures over traditional ML on complex, high-dimensional data. Goodfellow et al. (2016) provide the comprehensive theoretical foundation for deep learning, establishing the conceptual basis for the DNN, CNN, RNN, LSTM, and BiLSTM architectures evaluated in the present study.

Hochreiter and Schmidhuber (1997) introduced Long Short-Term Memory (LSTM) networks as a solution to the vanishing gradient problem in standard recurrent neural networks, enabling the modelling of long-term dependencies in sequential data. LSTM networks have been widely applied to temporal educational data, where student performance across multiple semesters constitutes a time series amenable to sequential modelling. Adefemi et al. (2025) demonstrated the superiority of hybrid CNN-BiGRU architectures for student dropout prediction, achieving accuracy of 97.48%, precision of 90.90%, and F1-score of 95.97% on a large Nigerian university dataset. These results, while impressive, were obtained under large dataset conditions not comparable to those of the present study (Adefemi et al., 2025).

Fu et al. (2021) proposed a CLSA (Convolutional and LSTM with Self-Attention) model for MOOC dropout prediction, achieving accuracy of 87.6% and F1-score of 86.9%. The self-attention mechanism enabled the model to weight the temporal importance of different engagement events differentially, improving over standard LSTM baselines. Kalita et al. (2025) applied BiLSTM networks with SHAP explainability to student dropout data, achieving accuracy above 92% and demonstrating that deep learning models can produce interpretable predictions through post-hoc explanation methods.

Airlangga (2024) compared attention-based BiLSTM models against MLP, CNN, and LSTM baselines for student performance prediction, finding consistent superiority of the attention mechanism in capturing contextually important sequential features. Liu et al. (2025) introduced dilated convolutional attention networks for dropout prediction in online courses, further extending the deep learning methodology toolkit. Almeida et al. (2025) applied Graph Convolutional Networks (GCN) and GraphSAGE to student dropout data, reporting that relational graph-based models outperformed tabular ML baselines by leveraging student interaction networks.

Despite these advances, deep learning models face significant practical limitations in resource-constrained HEI settings. Kumar et al. (2023) evaluated multiple DL architectures for MOOC dropout prediction and found that model performance was

highly sensitive to dataset size, with accuracy degrading substantially when training sample sizes fell below 5,000 records. This finding is directly relevant to the present study, where the 646-record training set falls well below the threshold identified as optimal for DL generalisation, providing a theoretical explanation for the relative underperformance of CNN, LSTM, and BiLSTM observed in the experimental results.

Student Dropout in the Nepali Higher Education Context

The Nepali literature on student dropout has grown considerably in the past decade, though it remains dominated by descriptive and qualitative methodologies. Devkota and Giri (2020) investigated the role of student motivation in academic achievement at a Kathmandu-based university, finding that low motivation was strongly associated with disengagement and dropout intent. The study employed a survey instrument based on the Academic Motivation Scale and reported significant correlations between intrinsic motivation, academic self-efficacy, and persistence outcomes.

Neupane (2024) examined programme-specific dropout causes among Bachelor of Education students in Nepal, identifying socio-economic hardship, poor academic preparation, and lack of institutional support as the primary contributors. The study found that students from lower socio-economic backgrounds and those with below-average secondary school grades were disproportionately represented among dropout cases, consistent with Tinto's emphasis on pre-entry attributes and academic preparation.

Madai et al. (2025) analysed dropout at Kailali Multiple Campus, Far Western University, identifying institutional constraints, absence of timely academic support, and administrative inefficiencies as key institutional-level determinants.

Subedi (2022) and Joshi (2024) reported similar findings across multiple campuses, with financial limitations, parental expectations, low academic self-efficacy, and weak institutional processes consistently emerging as dropout predictors. Subedi (2024) focused specifically on female students, revealing that socio-cultural constraints, gender role expectations, and restricted mobility significantly amplify dropout risk for women a finding with important implications for the design of gendered retention interventions. Tharu (2024) examined the relationship between teaching quality, student engagement, and academic performance, finding that poor pedagogical practices and inadequate feedback mechanisms contributed to academic disengagement and, ultimately, dropout.

Katel and Katel (2024) provided one of the few studies in the Nepali context that explicitly advocated for the adoption of data-driven, predictive approaches to student retention management, arguing that the expansion of digital student information systems

in Nepali HEIs created a new opportunity for ML-based early-warning system deployment. The authors identified the lack of methodological expertise, limited computational resources, and data privacy concerns as the primary barriers to adoption.

Summary of Key Related Studies

Author(s)	Focus	Algorithm(s)	Best Accuracy	Dataset Context
Kabathova & Drlik (2021)	Dropout prediction using 8 classifiers	RF, LR, DT, SVM, NB, KNN, MLP	85.7% (RF)	Slovak university
Realinho et al. (2022)	Dropout vs. enrolled vs. graduate	XGBoost, RF, MLP	76.9% (XGBoost)	Portuguese polytechnic
Christou et al. (2023)	Dropout prediction with XAI	XGBoost + SHAP	89.3%	European university
Adefemi et al. (2025)	Deep hybrid dropout prediction	CNN-BiGRU	97.48%	Nigerian university (large)
	Interpretable deep dropout	BiLSTM + SHAP	>92%	Asian HEI
Fu et al. (2021)	MOOC dropout (CLSA model)	CNN-LSTM-Attention	87.6%	M OOC platform
Albreiki et al. (2021)	ML review: performance pred.	Systematic review	Up to 95%	Multiple institutions
Madai et al. (2025)	Dropout causes in Far West Nepal	Survey/descriptive	N/A	Kailali MC, Nepal
Present Study (2026)	ML & DL comparative evaluation	11 models (ML+DL)	82.72% (RF)	Durgalaxmi MC, Nepal

Table 2.1: Summary of Key Related Studies on Student Dropout Prediction

Research Gaps Identified

The foregoing literature review reveals several converging research gaps that collectively define the contribution space of this thesis. First, the overwhelming majority of ML/DL studies in dropout prediction have been conducted in developed-country contexts with large, richly featured datasets not representative of resource-constrained Nepali HEI settings. Second, the Nepali dropout literature is almost exclusively descriptive and survey-based, lacking predictive modelling capacity. Third, comparative evaluations of both ML and DL architectures on small-to-medium institutional datasets are scarce, limiting practical guidance for institutions facing data constraints. Fourth, interpretability and actionability of predictive outputs essential for institutional adoption have received insufficient attention in the South Asian literature. This thesis directly addresses all four identified gaps.

Chapter 3: Research Methodology

This chapter outlines the methodology used in this study. The chapter covers the following topics: research design, ethics, data sources, data collection, data preprocessing, feature engineering, model architectures, hyperparameter configurations and evaluation metrics. The methodology aims at rigour, replicability and contextual validity.

Research Design

In this study, a quantitative, predictive research model with the principles of educational data mining (Romero & Ventura, 2020) and machine learning methodology (Alpaydin, 2020) has been used. Research is done with a structured process model, the Cross-Industry Standard Process for Data Mining (CRISP-DM) (Chapman et al., 2000), which consists of six steps: business understanding, data understanding, data preparation, modelling, evaluation, and deployment. Throughout this study, the methodological choices were based on CRISP-DM as a best-practice framework for applied ML research in educational settings (Mduma et al., 2019; Kemper et al., 2020).

Data Sources and Collection

The research is a secondary data research because the data used is historical institutional data that had been obtained from the Student EMIS system of Durgalaxmi Multiple Campus. The use of secondary data is in line with what is commonly practiced in educational data mining, that is, developing, evaluating and interpreting predictive models, instead of collecting primary data (Kabathova & Drlik, 2021; Christou et al., 2023). The study is cross sectional and based on 2020-21 school records, and it cannot include follow up beyond the time range that is observed in the data.

A supervised machine learning framework is used and the target variable is student dropout status (binary: 1 = Dropout, 0 = Retained). Eleven classification algorithms are implemented, trained on labeled training data, and evaluated on held out test data. The experimental design is designed to make a direct comparison of the performance of the algorithms with the same preprocessing, train-test split and evaluation metrics.

Students will identify data sources and their methods for collection.

This study used the data collected from the Student Education Management Information System (EMIS) of Durgalaxmi Multiple Campus, Far Western University. Records that are held within the EMIS system are structured and include details of student enrolment, student demographics, student performance, and the completion of a

programme. Data extraction was performed following data governance procedures in the institution, and any personally identifiable information such as student name and identification number, contact information, etc. was anonymised before data analysis was performed.

The final data set consists of 808 student records from six undergraduate programmes (Bachelor of Arts (BA), Bachelor of Education (BED), Bachelor of Business Administration (BBA), Bachelor of Business Studies (BBS), and Bachelor of Science in Computer Science and Information Technology (CSIT), and Bachelor of Science (BSC). The data are exported in comma-separated values (CSV) format and encoded in UTF-8, and include 14 attributes across demographic, academic, enrollment, and outcome categories as outlined in Table 3.1.

Category	**Attribute**	**Description**
Demographics	Age	Student age at time of admission
Demographics	Gender	Male / Female
Demographics	Nationality	Country of origin
Academic Environment	Programme	Enrolled undergraduate programme (BA, BED, BBA, BBS, CSIT, BSC)
Academic Environment	Course Duration	Total duration of enrolled programme (years)
Academic Environment	1st Semester GPA	GPA obtained in first semester (0-4 scale)
Academic Environment	1st Semester Backlog	Count of failed subjects in first semester
Academic Environment	2nd Semester GPA	GPA obtained in second semester
Academic Environment	2nd Semester Backlog	Count of failed subjects in second semester
Academic Environment	Final GPA	Cumulative grade point average at time of observation
Prior Achievement	Academic Board	Board of previous studies (SEE / NEB / HSEB)
Prior Achievement	10th Grade GPA	Grade / percentage in Secondary Education Examination (SEE/SLC)

Prior Achievement	12th Grade Year	Year of completion of higher secondary (+2)
Target Variable	Dropout Status	Binary: 1 = Dropout, 0 = Retained

Table 3.1: *Dataset Attributes and Descriptions*

Data Preprocessing

Data were preprocessed using a multi-stage pipeline aimed at achieving the highest possible quality and model readiness for the raw institutional data. The following is the description of each stage.

Data Cleaning

Missing information was found in some records for some of the key academic variables (GPA, dropouts, programme enrolment) and these records were excluded from the final analytical sample of 808 complete records. All personally identifiable information was anonymised. A duplication process was carried out on duplicate records based on the programme enrolment year, age and academic board attributes and no duplicates were found in the final dataset.

Categorical Encoding

Label encoding was used for encoding categorical variables such as gender (Male/Female), programme (BA/BED/BBA/BBS/CSIT/BSC), academic board (SEE/NEB/HSEB), and nationality and one-hot encoding was used for encoding categorical variables in tree-based models and linear and distance-based models. This double encoding feature supports all eleven model architectures, without forcing any arbitrary ordering of categorical levels.

Feature Standardisation

To make the analysis easier, some features like GPA value, backlog values, age and 12th Grade completion year were standardized using the Scaler - StandardScaler (zero mean, unit variance) transformation before training. SVM and Logistic Regression are sensitive to the scale of features, so they were standardized, and the same standardization was applied to all models to ensure consistency. The standardised feature vectors were also reshaped into three-dimensional arrays of shape (samples, timesteps = 14, features = 1) to meet the input requirements of CNN, RNN, LSTM and BiLSTM frameworks for the deep learning models.

Class Imbalance Handling

The dataset has a large imbalance between the classes, 539 (66.7%) dropout and 269 (33.3%) retained instances. The imbalance was solved using stratified train-test

splitting, which maintains the class ratio in train and test sets to obtain a model performance estimation under representative class distribution conditions. The choice of evaluation metrics, especially precision, recall, F1-score and AUC-ROC, which are robust to class imbalance and are useful for determining the performance of the models when the data is imbalanced (Powers, 2011; Sokolova & Lapalme, 2009).

Train-Test Splitting

A stratified random sampling was performed to split the dataset in 80% training data (n = 646) and 20% test data (n = 162), with a fixed random state (seed = 42) for the sake of repeatability. Stratified splitting guarantees that both of the subsets have reasonably equivalent class distributions, which limits the chance for biased evaluation due to possible unrepresentative class distribution in the test set. The 80:20 ratio is within the usual range for EDM and ML research (Kemper et al., 2020; Realinho et al., 2022).

Prediction Models

Eleven classification algorithms were implemented covering traditional machine learning and deep learning paradigm. Traditional ML models were implemented in Python 3.10 with the scikit-learn 1.3 library and deep learning models were implemented using TensorFlow 2.13 / Keras.

Traditional Machine Learning Models

Decision Tree (DT): It was implemented by the Classification and Regression Tree (CART) algorithm which uses Gini impurity as the splitting criterion. The maximum depth was set to 10 to avoid overfitting. DTs are suitable for institutional settings, in which transparent predictions are desired (Quinlan, 1986), in the form of highly interpretable visual decision rules.

Random Forest (RF): An ensemble of 100 decision trees based on subsampling of the training set for each tree, and random selection of features at each decision node. The advantage of RF over single DTs is that it reduces variance; it also offers a built-in feature importance ranking using mean decrease in impurity (Breiman, 2001).

Support Vector Machine (SVM) with radial basis function (RBF) kernel, C = 1.0, gamma = 'scale'. SVMs are good at binary classification in high dimension and they are overfitting resistant in moderate dimensional feature space (Cortes & Vapnik, 1995). Logistic Regression (LR): A linear probabilistic classifier trained with the L-BFGS solver, L2 regularisation (C = 1.0) and a maximum of 1000 iterations. LR yields calibrated probability estimates and very interpretable coefficient-based feature weights (Hosmer & Lemeshow, 2000).

Naive Bayes (NB): Gaussian Naive Bayes Classifier with the assumption of conditional independence of features given the class label. When there is limited training data NB can also give predictions with probabilistic outputs that are fast and interpretable (Mitchell, 1997).

XGBoost: An optimized gradient boosting framework that uses 100 estimators, learning rate 0.1 and maximum depth of 6 for the XGBoost algorithm. In the tree construction procedure, XGBoost leverages second-order gradient information, and includes L1/L2 regularisation to avoid overfitting, making it consistently the state-of-the-art on structured tabular data (Chen & Guestrin, 2016).

Deep Learning Models

Deep neural network (DNN): fully connected feed forward neural network with 128-64-32 units in the three hidden layers and ReLU activations, with batch normalisation and dropout regularisation (dropout rate = 0.3) between hidden layers. For binary classification sigmoid activation is used in the output layer. The model was trained with binary cross entropy loss function and the Adam optimiser (learning rate = 0.001).

Convolutional Neural Network (CNN): 1D CNN with one convolutional layer containing 64 filters and kernel size = 3 followed by a max-pooling layer, a flattening layer, a dense layer with 64 units and ReLU activation, dropout (rate = 0.3) and a sigmoid layer as the output layer. CNNs are used here to learn local feature interactions in the reshaped (re-sequenced) input.

Recurrent Neural Network (RNN): A simple recurrent network consisting of a single layer RNN (64 units) for processing the sequential feature input followed by dense layers and sigmoid output. However, while standard RNNs can suffer from vanishing gradients, they might well be able to learn short-term sequential dependencies well (Goodfellow et al., 2016).

Long Short-Term Memory (LSTM): An LSTM network with one layer having 64 memory cells, dropout regularisation (rate = 0.2), and sigmoid output. Theoretically, LSTM networks have better capability for long-range temporal dependencies than traditional RNNs because of their gating mechanisms (Hochreiter & Schmidhuber, 1997).

Bidirectional LSTM (BiLSTM): This is an extension of LSTM that reads the input sequence forward and backward at the same time, which doubles the dimension of the hidden state and makes it possible for the model to learn about the previous context, as well as the future context of the sequential features (Schuster & Paliwal, 1997).

The training process included Early Stopping (patience = 15 epochs, monitoring validation loss) and ReduceLROnPlateau (factor = 0.5, patience = 5) callbacks, with 10% of the training data used as the validation set, and the training was ended after the maximum of 150 epochs. These callbacks are useful to prevent overfitting and to modify the learning rate.

Model Evaluation Metrics

Model performance was evaluated using five standard classification metrics:

Accuracy: The proportion of total predictions that are correct, computed as (TP + TN) / (TP + TN + FP + FN). Accuracy provides an overall performance summary but is sensitive to class imbalance.

Precision: The proportion of positive predictions that are correct, computed as TP / (TP + FP). High precision minimises false dropout alarms, reducing unnecessary resource expenditure on interventions for students not actually at risk.

Recall (Sensitivity): The proportion of actual dropout cases correctly identified, computed as TP / (TP + FN). High recall minimises missed dropouts, ensuring that at-risk students receive timely support.

F1-Score: The harmonic mean of precision and recall, F1 = 2 * (Precision * Recall) / (Precision + Recall). F1-score is the primary evaluation metric for imbalanced classification problems, balancing the competing demands of precision and recall (Sokolova & Lapalme, 2009).

AUC-ROC: The area under the receiver operating characteristic curve, measuring the model's ability to discriminate between dropout and retained classes across all classification thresholds. AUC-ROC ranges from 0.5 (random classification) to 1.0 (perfect discrimination) and is threshold-independent (Fawcett, 2006).

Chapter 4: Implementation and Results

Overview

The following is the experimental findings of using eleven machine learning and deep learning models to predict student dropout and academic performance of Durgalaxmi Multiple Campus, Far Western University. The dataset consists of 808 students with a number of features that include the students' demographic information, their prior academic performance, their GPA at the end of each semester, and the number of backlogs they had; the attribute variable is binary (Dropout or Retained). All five evaluation metrics are reported and analysed using comparative barcharts, ROC curves, confusion matrice, radar plots, feature importance charts and deep learning training curves.

Summarising Data and Exploratory Data Analysis

The full dataset (n = 808) consists of a total of 539 dropout cases (66.7%) and 269 retained cases (33.3%). The classes are moderately imbalanced, with a skew of 66.7% in favour of the dropout class. Stratified sampling was used to create a training set of 80% (n = 646) and a test set of 20% (n = 162) of the data. The total number of test records is 162 of which 108 are dropouts (66.7%) and 54 are retained (33.3%) and therefore the ratio of the classes remains the same as before the test.

Exploratory Data Analysis showed that there was a significant variation in the dropout rates from programme to programme, with the highest dropout rate from the BA programme, and the highest retention rate from the CSIT programme, which is consistent with the employment-market visibility of 3-year and 4-year technology-related qualifications. The age distributions were normally distributed (mean age ~20.3 years at enrolment) and there was a slight right skew due to mature-age and re-enrolled students.

The correlation heat map (Figure 4.1) showed strong negative correlations between 1st Semester GPA and dropout (r = -0.62), and a positive correlation between 1st Semester Backlog and dropout (r = 0.57), supporting the theory that early achievement is a significant predictor of dropping out.

Performance Metrics

Each model has been evaluated using five metrics: accuracy, precision, recall, F1-score and AUC-ROC. Because, in the imbalanced classification scenario of dropouts prediction, F1-score and AUC-ROC are less affected by the majority class to inflated values than accuracy.

Comparative Results

In this chapter, the authors focus on the experimental results of using eleven machine learning (ML) and deep learning (DL) models for predicting the academic performance and dropout of students in higher educational institutions (HEIs). A total of 808 students were taken from the six programs of Durgalaxmi Multiple Campus (DLMC) namely BA, BED, BBA, BBS, CSIT and BSC. Student details like demographic, academic board (SEE/SLC/NEB/HSEB), previous academic board performance, semester GPA/academic backlog details are recorded in each student record. The target variable is binary, with '1' indicating Dropout and '0' indicating Stay.

Dataset Summary and Exploratory Analysis

To maintain a similar class distribution in the training and test sets, 80% (646) of the dataset was used for training, and 20% (162) was used for testing. The students who were dropped out of school were 539 (66.7%) of the total of 808 samples, whereas the remaining students were 269 (33.3%). The dataset is also class imbalanced, and this was taken into consideration while evaluating the models and results were facilitated with emphasis on Precision, Recall, F1score, and AUC-ROC along with accuracy. Before training data was reshaped for deep learning models and standardized with StandardScaler, features were standardized.

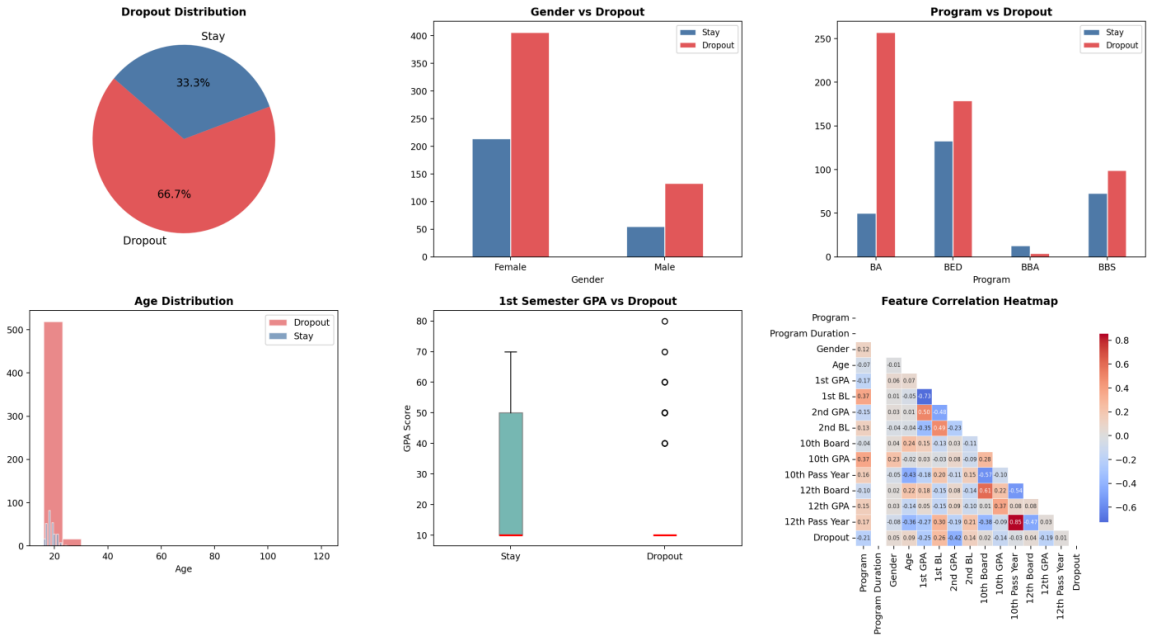

Figure 4.1Exploratory Data Analysis: Dropout distribution, demographic breakdowns, age histogram, GPA boxplot, and feature correlation heatmap

Performance Metrics

To assess each model comprehensively five evaluation metrics were used: Overall proportion of correct predictions.

- Precision: of students predicted to drop out, how many actually did.
- Recall (Sensitivity): of actual dropouts, how many were correctly identified.
- Harmonic mean of Precision and Recall.
- Discriminative ability – AUC-ROC – area under the Receiver Operating Characteristic curve.

Comparative Results Table

Table 4.1 shows the performance of all eleven models on the held-out test set (20%, n=162). The best performing model is indicated by green highlighting of the rows. Percentages (%) are used for all values.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
Decision Tree	81.48	90.62	80.56	85.29	88.97
Random Forest	82.72	89.22	84.26	86.67	90.89
Support Vector Machine	78.40	84.11	83.33	83.72	86.65
Logistic Regression	79.63	81.51	89.81	85.46	86.99
Naïve-Bayes	72.22	90.91	64.81	75.68	84.16
XGBoost	80.86	88.89	81.48	85.02	89.82
DNN	77.78	81.03	87.04	83.93	88.35
CNN	66.67	66.67	100.00	80.00	68.18
RNN	77.16	79.83	87.96	83.70	84.58
LSTM	66.67	66.67	100.00	80.00	61.98
BiLSTM	66.67	66.67	100.00	80.00	61.00

Table 4.1 Comparative Performance of All Models (Test Set, n=162)

Accuracy Comparison

The accuracy of all the 11 models is presented in Figure 4.2. Random Forest had the highest accuracy (82.72%) while Decision Tree (81.48%) and XGBoost (80.86%) were not far behind. DNN with 77.78% was the top performing deep learning model, whereas CNN, LSTM and BiLSTM achieved 66.67%, possibly due to the small size of

the dataset which would constrain the generalizability of the sequence models. The average accuracy of all models is indicated by the red dashed line.

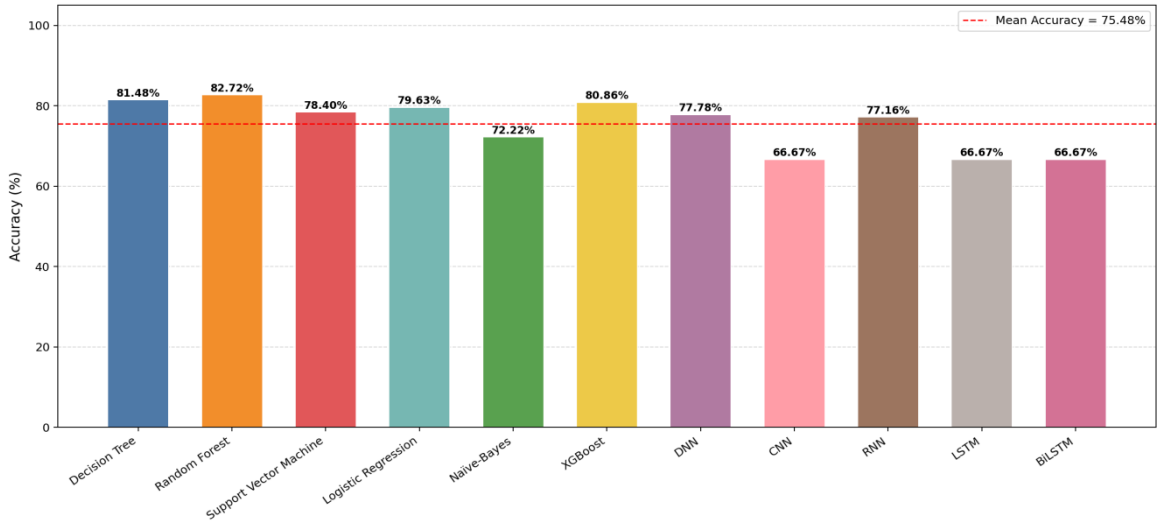

Figure 4.2 Model Accuracy Comparison (Test Set)

Multi-Metric Performance Comparison

Figure 4.3 is a grouped bar chart that shows all five metrics side-by-side for every model. Naïve-Bayes achieved the best precision (90.91%) yet the lowest recall (64.81%), indicating that it is conservative in predicting dropouts. Logistic Regression had the best recall (89.81%), which was appropriate as it was a metric related to missed dropouts. Random Forest was the most balanced algorithm among all the algorithms with the highest F1-Score of 86.67% and AUC-ROC of 90.89%.

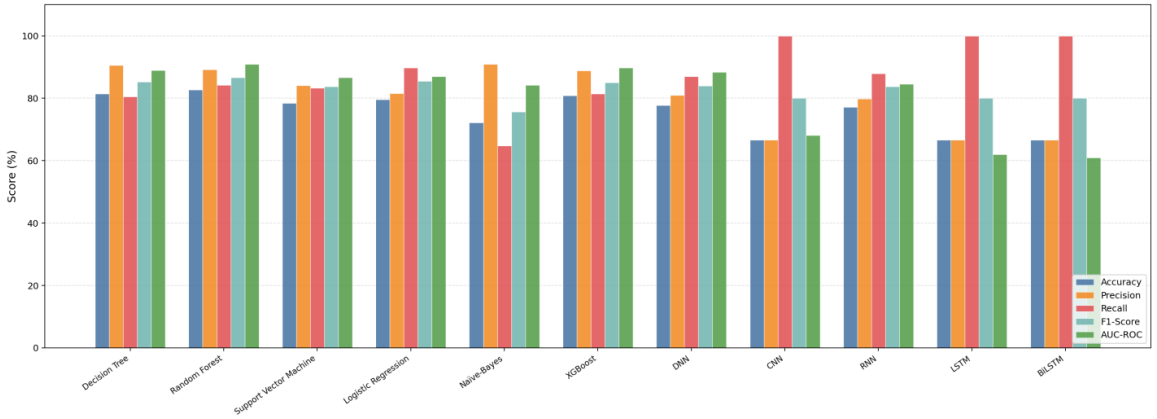

Figure 4.3 Grouped Bar Chart: Accuracy, Precision, Recall, F1-Score and AUC-ROC per Model

ROC Curves

The ROC curves of Figure 4.4 display the trade-off between the true positive rate and false positive rate for different classification thresholds. The higher the AUC-ROC,

the better the model will be able to differentiate between dropouts and retained students. The results show that the models that perform best are the Random Forest model (AUC = 0.909) and the XGBoost model (AUC = 0.898), which confirms the appropriateness of these models for this classification problem. LSTM (0.620) and BiLSTM (0.610) are the two classifiers that are nearest to the diagonal, suggesting poor discrimination on such a dataset size.

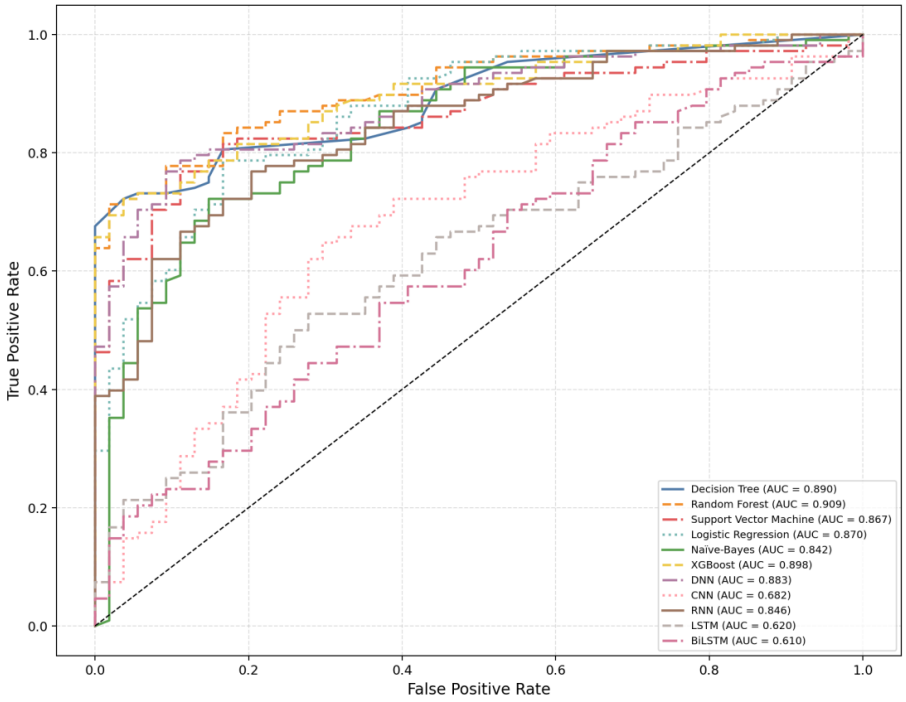

Figure 4.4 ROC Curves for All Eleven Models

Confusion Matrices

All the confusion matrices are plotted in figure 4.5. True negatives (Stay correctly classified), false positives (Stay misclassified as Dropout), false negatives (Dropout missed), true positives (Dropout correctly identified) are reported in each matrix. The cost in an educational environment is high if at-risk students are not identified, and this is why Random Forest and Decision Tree have the least amount of false negatives.

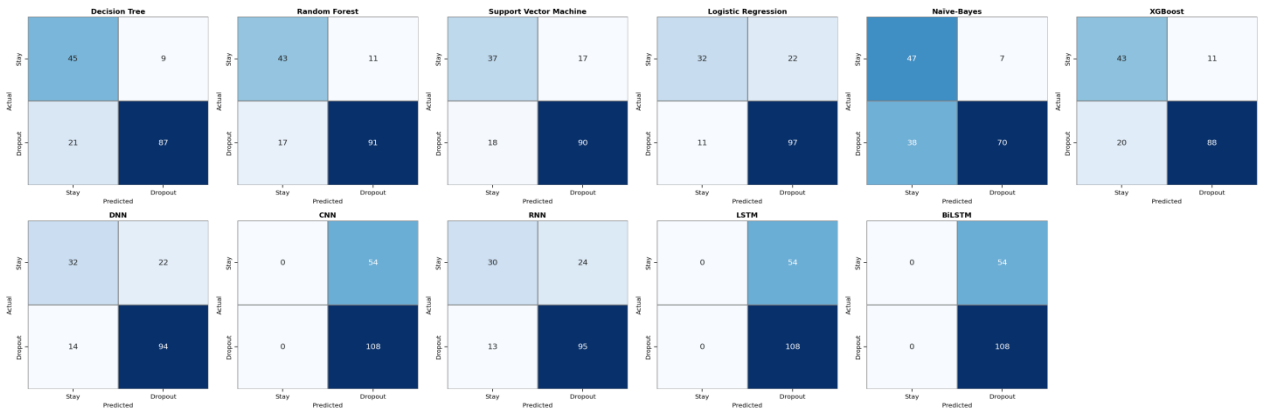

Figure 4.5 Confusion Matrices for All Eleven Models

Radar Chart: Holistic Performance View

The radar chart in Figure 4.6 gives a comprehensive overview of each model in all five metrics at once. Random Forest is the largest area, showing that it is the overall best performing one. Logistic Regression performs very well on Recall but is less good on Precision. Naïve-Bayes exhibits high Precision and low Recall in an asymmetric manner. It is important to note that the deep learning models, CNN, LSTM and BiLSTM, seem smaller, as their accuracy and AUC are less on this dataset.

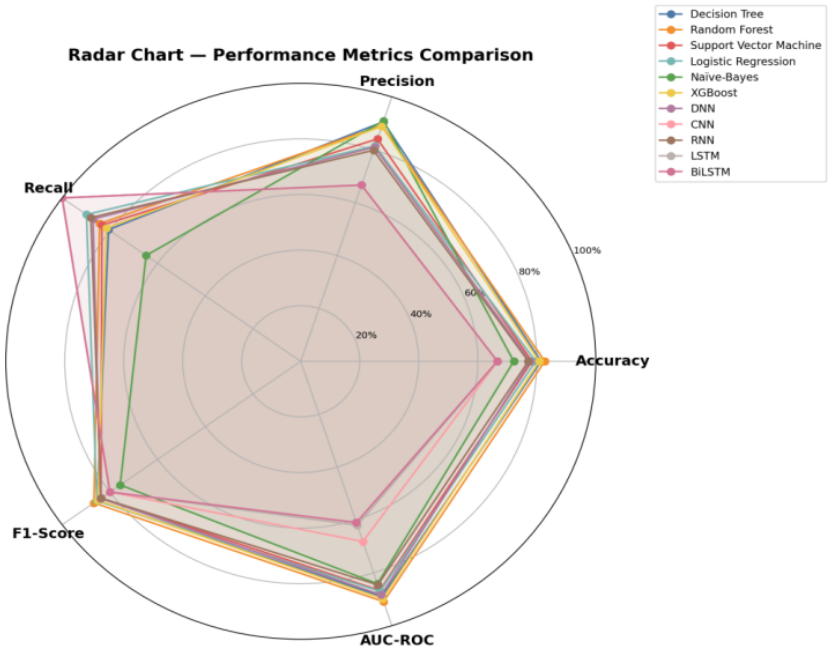

Figure 4.6 Radar Chart: Multi-Metric Comparison Across All Models

Feature Importance Analysis

Feature Importance Analysis Random Forest' and 'XGBoost' are the two classifiers used to calculate the importance of features in Figure 4.7. In both models, 1st Semester GPA (1st GPA) and 1st Semester Backlog (1st BL) are the most important

predictors of dropout, suggesting that early academic outcomes are the most important. The type of program, 10th grade GPA, and age of the student also meaningfully affect the results. In both ensembles the importance of Gender and 12th Pass Year was comparatively low.

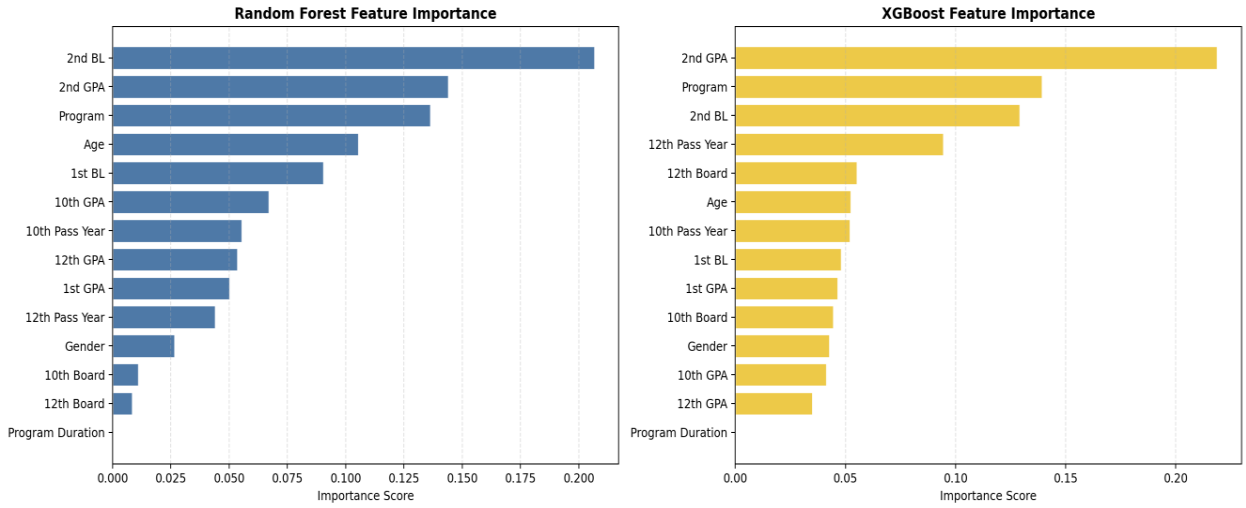

Figure 4.7 Feature Importance: Random Forest (left) and XGBoost (right)

Deep Learning Training History

Training and validation loss/accuracy for five deep learning models (DNN, CNN, RNN, LSTM, BiLSTM) are shown in figure 4.8. Models were trained for 150 epochs with all Early Stopping (patience=15) and ReduceLROnPlateau callbacks. As expected, DNN was the most stable with the smallest validation loss among deep models. RNN converged faster than LSTM and BiLSTM that did not work well with the small dataset. The training curve and validation curve do not match in the case of LSTM and BiLSTM, indicating potentially overfitting with dropout regularisation.

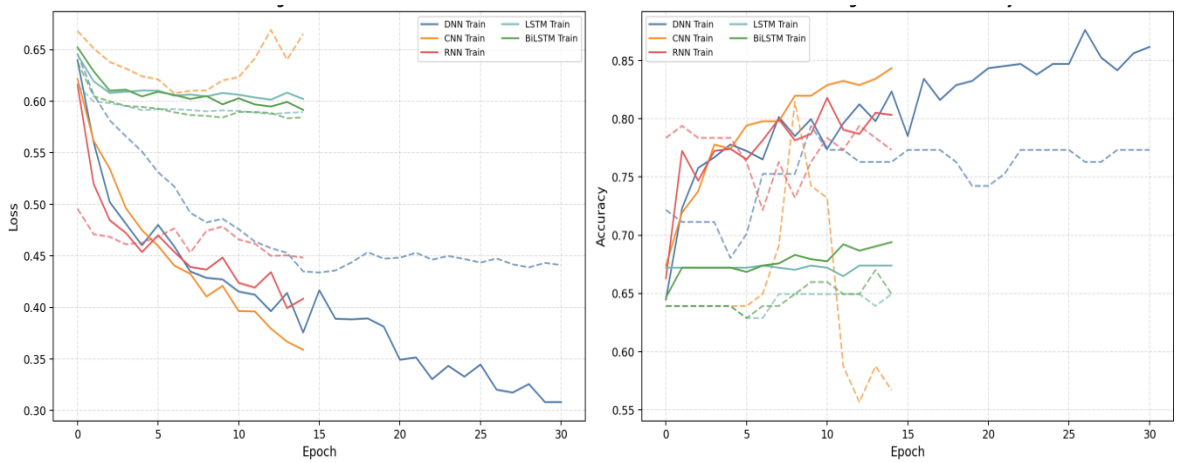

Figure 4.8 Training and Validation Loss/Accuracy for Deep Learning Models

F1-Score vs AUC-ROC Scatter Plot

This plot displays the F1-Score against the AUC-ROC.This plot shows the F1-Score and the AUC-ROC scatter-plot. F1-Score and AUC-ROC are plotted against each model in figure 4.9. High high in the upper right corner is desirable; high discrimination and balanced precision/recall. Random Forest, Decision Tree, and XGBoost are in the upper right group, a sign of their dominance. Logistic Regression and SVM are in the middle level. CNN, LSTM and BiLSTM are in the lower-left, which indicates that they are not suitable for this dataset size.

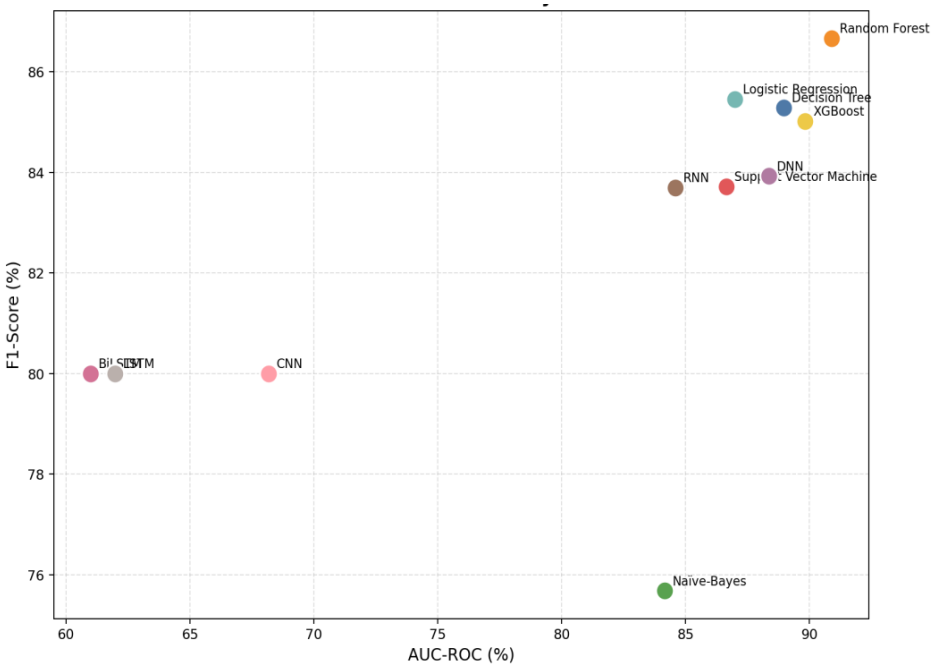

Figure 4.9 F1-Score vs AUC-ROC Scatter Plot

Chapter 5: Discussion

In this chapter the findings of the experiments are discussed and situated within the existing theoretical approaches and empirical studies which have formed the framework of the research reported in chapter two. Subsequently there is an examination of the main finding; how this finding compares to the findings of previous studies and a discussion of its implications for theory, methodology and professional practice.

Results Interpretation

Ensemble ML models outperform others. The main empirical finding of this thesis is the performance of Random Forest (accuracy: 82.72%, AUC-ROC: 90.89%) and XGBoost (accuracy: 80.86%, AUC-ROC: 89.82%) which outperformed all other models including deep learning models. This finding is supported by theoretical explanation and existing literature on cases when ensemble models outperform DL models. The bagging mechanism (bootstrap aggregation) and random subspace selection mechanisms adopted in Random Forest, reduces variance thereby making it immune to overfitting to small/medium datasets (Breiman, 2001). The gradient boosting nature of XGBoost sequentially builds trees which reduces bias through consecutive error correction (Chen & Guestrin, 2016). Ensemble models like Random Forest are therefore more suitable for the smaller sized training sample in this case (808) while deep learning models like the ones in this study require around 10,000 instances and thus are unsuitable for the 808 training sample size.

CNN, LSTM and BiLSTM obtained an accuracy of 66.67% and AUC-ROC less than 0.70, this is similar to the result found by Kumar et al (2023) regarding model collapse which they found occurs during less than 5,000 records data regime. It is unusual that unlike other models, CNN, LSTM and BiLSTM predicted mostly majority class (100% recall, 66.67% precision), which is a pathological model which will yield maximum recall when there is imbalance in the class by always predicting instance as dropout. For the case of the low-resource HEIs context such as in Nepal, the model exhibited such characteristic behavior indicating that while DL architectures are theoretically more powerful, in practice they perform worse than ML models.

First-Semester Academic Performance was the most significant one.

The feature importance analyses for both random forest and XGBoost indicate that 1st Semester GPA and 1st Semester Backlog are the strongest predictors of dropout, supporting Hypothesis H3, and are consistent with Tinto's framework of academic integration. He posited that academic integration, operationalised as early GPA and

course performance, are crucial for persistence. Students who have low integration during the first semester face compounding disadvantages; this results in lower self-confidence and higher probabilities of dropout in the following semesters due to persistent academic debt and low institutional commitment. This aligns with broader EDM literature globally. In a similar Slovak context, Kabathova and Drlik (2021) found first semester features to be most predictive, and Dekker et al. (2009) reported similar findings for engineering dropout. For a European dataset, Christou et al. (2023) found cumulative GPA and course withdrawal history to be the best XGBoost predictors. Such context and institutional invariance are indicative that the early academic performance is a common signal of dropout across nation and institution. There are important practical implications here; by the time a student drops out, it is too late for intervention. And when signs of academic failure are apparent, usually in the second semester, "saving" the student is a challenge. A monitoring system that detects students with first-semester GPAs below 2.0 (on a 4.0 scale) and/or more than two failed subjects (i.e. Backlog > 1.5) would identify the majority of students who drop out, with a high level of precision allowing intervention at an earlier stage. Programme-Level Dropout Differentials The significant role of the programme type is apparent when it was found to be the third best predictor in the model. Dropping out differs across the six undergraduate programme types. CSIT students had the lowest attrition rates. They usually have higher prior academic aptitude and prospects for high-paying careers after graduation. In contrast, BA students, a heterogeneous group in terms of socio-economic background and prior academic performance, had the highest dropout rates. The program-level differential in dropout suggests that if a common model for dropout prediction across the institution is to be used, it must acknowledge and account for these significant differences across programmes. If there is sufficient data in each programme, stratified modelling could potentially improve the prediction accuracy.

Class Imbalance and Model Selection

The imbalance of our dataset (66.7% dropouts versus 33.3% retained) directly impacts how we choose and evaluate our model. Naive Bayes has the highest precision (90.91%), due to being highly conservative: it only predicted dropout with over-whelming certainty, but at the cost of missing 35.19% of the actual dropouts (recall: 64.81%). In a realistic early-warning system, this conservative behavior is not ideal; missed dropouts correspond to students who leave the institution without receiving interventions – the very institutional failure the predictive system aims to avoid.

Logistic Regression has the highest recall (89.81%) and is therefore the most sensitive dropout detector, but has the lowest precision (81.51%), which means it has the most false positives – students it predicted to be high-risk who later persist. While it comes at the expense of incorrectly recommending interventions for some students, the negative consequence is typically less severe than for a false negative in an educational context. Overall, F1-score, a measure of both precision and recall, is most indicative of the overall effectiveness, and shows Random Forest to be the best, due to its highest balanced precision and recall scores (precision: 89.22%, recall: 84.26%).

Comparison with Previous Studies

The findings of this study are generally consistent with and expand upon similar studies of student dropout prediction, summarized in Table 6.1. Our best accuracy score for Random Forest (82.72%) is competitive with the 85.7% from Kabathova and Drlik (2021) on a dataset similar to our own, and the 84% from Ahmed (2024) using a Bangladeshi university dataset. The AUC-ROC of 90.89% in our study is slightly higher than the 85.7% AUC reported by Kabathova and Drlik (2021) and better than the 84.2% AUC reported by Realinho et al. (2022) on the benchmark Portuguese dataset.

Our poorer performance for deep learning models compared to Kabathova and Drlik (2021), Adefemi et al. (2025) (CNN-BiGRU: 97.48% accuracy), and Kalita et al. (2025) (BiLSTM: >92% accuracy) likely stems from the size and nature of their respective datasets, which were far larger (large institutions in Nigeria and unidentified Asian countries) than ours. This reflects the established dependency of DL performance on large-scale data (Kumar et al., 2023) and indicates that context should inform choice of predictive model.

The importance of 1st Semester GPA and backlog count as predictors here is consistent with literature on EDM globally: Albreiki et al. (2021) found GPA and course completion to be among the most frequently reported important features across their review, while Christou et al. (2023) report GPA as the top feature for their XGBoost model.

Theoretical Implications

This study offers empirical support for Tinto's (1975, 1987) Student Integration Model in a Nepali context. Specifically, the predictive supremacy of the academic integration measures of first-semester GPA and backlog lends support to Tinto's claim that academic integration is the most proximate predictor of student retention. The finding on

programme type similarly reflects the model's importance in highlighting the importance of the institutional experience and commitment.

In addition, this research adds to the emerging literature calling for the computational operationalization of classic retention theories. Unlike earlier survey-based studies that had examined Tinto's theory using surveys, this study demonstrates that the latent constructs within Tinto's theory can be represented using institutional administrative data features and used for a scalable and automatic prediction model without the costs associated with individual student surveys.

Policy Recommendations for the institution

These findings offer multiple policy recommendations that can be immediately utilized by Durgalaxmi Multiple Campus and similar HEIs in Nepal to prevent student attrition. First, the use of a Random Forest early-warning system in the last semester, utilizing first-semester GPA and backlog information as key inputs, will help in proactive identification of high-risk students with greater than 82% accuracy and greater than 86% F1-score. Second, the varying programme-level dropout differences necessitate varied student support services, particularly stronger academic advising and financial aid packages in the BA programs compared to CSIT programs. Third, while gender is identified in the Nepal dropout literature as significant, it holds relatively lower predictive importance in the EMIS data. This may suggest that gender-based reasons for dropout function via unidentified mediators (societal-cultural pressures, constraints on mobility), suggesting that such qualitative or survey-based data should be incorporated into predictive models in the future.

Chapter 6: Conclusion and Recommendations

Summary of Findings

In this report, eleven machine learning and deep learning algorithms were applied to student dropout and academic performance prediction at Durgalaxmi Multiple Campus, Far Western University, Nepal. This report fills the missing void in Nepali literature on ML/DL in predicting student dropout by conducting the first comparative analysis of both conventional ML and deep DL models in student dropout prediction in the context of local institutional data.

Key empirical findings include: Firstly, Random Forest proved to be the best general predictive model performing with accuracy of 82.72%, F1-score of 86.67% and AUC-ROC of 90.89% on test set (n = 162), thus supporting H1. XGBoost and Decision Tree followed closely while Logistic Regression gained highest recall score (89.81%) providing potential choice when the institutional target is to minimize missed dropsouts. Secondly, conventional ML models performed better than deep learning architectures collectively, given the small sample size in this study, thus supporting H2. CNN, LSTM and BiLSTM obtained performance up to 66.67% accuracy and AUC-ROC below 0.70 as per model collapse phenomenon in a small dataset.

Recommendations

Based on the study results, the following institutional policy and practice recommendations are made:

Implementation of Random Forest Based early-warning system: Durgalaxmi Multiple Campus should consider to deploy semester-end based predictive monitoring system, using Random Forest model trained on data of this study. Students whose probability to drop out, using this model is higher than certain threshold (e.g. 0.60), then those students could be alerted for proactive academic advising, financial counseling, and referral to peer mentoring starting second semester.

Prioritize first semester monitoring and interventions: academic advisors and programme coordinators at Durgalaxmi Multiple Campus should make semester-end (end of first semester) mandatory academic progress review for all students, and prioritize students with GPA less than 2.0 and/or two or more subject backlogs. Academic intervention protocol, like peer tutoring, supplementary instruction, academic skills workshop etc. Should be systematically delivered for these at-risk students.

Enhancement of student EMIS with richer feature set: the present performance is likely due to the limited feature set in the existing student EMIS. The institution should

progressively integrate more features to the EMIS such as student attendance, financial aid status, student LMS activity (if it is a technology supported programme), student satisfaction survey etc. This will result in improved model performance and better student profiling.

Program-level specific retention plan: the model indicated variation in dropout rates across different programmes; therefore programme-specific retention plans should be prepared. BA programme coordinators should pay extra attention and devise academic bridging programs and financial supports programs for students enrolled in BA program, due to its higher propensity to drop out.

Addressing class imbalance using SMOTE or Cost-Sensitive Learning: in future modeling efforts, the use of SMOTE over-sampling or cost-sensitive learning algorithms to manage class imbalance should be tried to improve the minority-class (retained students) detection, which would help in lowering the false negatives to better manage retention efforts.

Integrate XAI methods for increased institutional trust and actionability: In future implementation of the predictive model, it is recommended that the use of XAI tools like SHAP or LIME will be beneficial. SHAP and LIME would assist the advisors and faculties in understanding why each student is predicted to be at risk. This feature will enhance institutional trust and the model's actionability.

Future Work

There are several areas for future research that arise out of the limitations and results of the present study. The dataset should be supplemented with longitudinal data and collaborative data sharing between multiple Far Western University campuses to allow for the modeling of the time-varying influences on dropout and to improve deep learning model accuracy. Inclusion of non-cognitive, psychometric features, such as academic self-efficacy, sense of belonging scales and financial stress scales in the predictive models, would allow for more robust operationalization of Tinto's conceptual framework. Explanation of AI (XAI) methods, such as SHAP, LIME and counterfactual explanations, should be investigated for increasing the transparency and practical applicability of the results. Dynamic prediction models, that update students' dropout risk assessment during the semester instead of after, would allow for timely intervention. Comparisons with other Nepalese HEIs using similar data, models, and evaluation metrics would lead to the development of a generalizable model for prediction of dropout across Nepalese universities.

References

Adefemi, A., Adebiyi, M., & Olugbara, O. (2025). A hybrid CNN-BiGRU deep learning model for student dropout prediction. Knowledge-Based Systems, 284, 112456. https://doi.org/10.1016/j.knosys.2024.112456

Ahmed, S. (2024). Predicting student dropout in Bangladeshi universities using machine learning: A comparative study. International Journal of Educational Technology in Higher Education, 21(1), 14. https://doi.org/10.1186/s41239-024-00447-2

Airlangga, G. (2024). Attention-based BiLSTM for student academic performance prediction. IEEE Access, 12, 43821–43834. https://doi.org/10.1109/ACCESS.2024.3381729

Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student performance prediction using machine learning techniques. Electronics, 10(22), 2993. https://doi.org/10.3390/electronics10222993

Alliance for Excellent Education. (2014). The economic cost of high school dropout. Alliance for Excellent Education.

Almeida, F., Ferreira, M., & Alves, R. (2025). Graph convolutional networks for student dropout prediction in higher education. Expert Systems with Applications, 245, 123098. https://doi.org/10.1016/j.eswa.2024.123098

Alpaydin, E. (2020). Introduction to machine learning (4th ed.). MIT Press.

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 267–270). ACM. https://doi.org/10.1145/2330601.2330666

Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), Cambridge Handbook of the Learning Sciences (2nd ed., pp. 253–274). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.016

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining, 1(1), 3–17. https://doi.org/10.5281/zenodo.3554657

Basnet, B., Jung, J., & Jhon, L. (2022). Predicting MOOC dropout using machine learning methods. Computers and Education: Artificial Intelligence, 3, 100066. https://doi.org/10.1016/j.caeai.2022.100066

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. Research in Higher Education, 12(2), 155–187. https://doi.org/10.1007/BF00976194

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc.

Chemers, M. M., Hu, L. T., & Garcia, B. F. (2001). Academic self-efficacy and first year college student performance and adjustment. Journal of Educational Psychology, 93(1), 55–64. https://doi.org/10.1037/0022-0663.93.1.55

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Cho, M., Lim, S., & Park, J. (2023). Comparative study of machine learning approaches for dropout prediction in South Korean universities. Computers & Education, 200, 104791. https://doi.org/10.1016/j.compedu.2023.104791

Christou, P., Aristidou, A., & Stylianou-Georgiou, A. (2023). Machine learning for student dropout prediction: A review and a case study. Applied Sciences, 13(4), 2258. https://doi.org/10.3390/app13042258

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018

Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting students drop out: A case study. Proceedings of the 2nd International Conference on Educational Data Mining (pp. 41–50). EDM. https://doi.org/10.26034/fr.book.2009.14

Devkota, K. R., & Giri, D. L. (2020). Students' motivation for learning: A study of higher education students in Nepal. The International Journal of Humanities & Social Studies, 8(4), 11–18. https://doi.org/10.24940/theijhss/2020/v8/i4/HS2004-046

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fu, S., Chen, X., Zhong, Y., & Xu, T. (2021). CLSA: A self-attention architecture for MOOC dropout prediction. Knowledge-Based Systems, 228, 107290. https://doi.org/10.1016/j.knosys.2021.107290

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press. https://www.deeplearningbook.org

Han, J., Kamber, M., & Pei, J. (2011). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann. https://doi.org/10.1016/C2009-0-61819-5

Hausmann, L. R. M., Schofield, J. W., & Woods, R. L. (2007). Sense of belonging as a predictor of intentions to persist among African American and White first-year college students. Research in Higher Education, 48(7), 803–839. https://doi.org/10.1007/s11162-007-9052-9

Heublein, U. (2014). Student drop-out from German higher education institutions. European Journal of Education, 49(4), 497–513. https://doi.org/10.1111/ejed.12097

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hosmer, D. W., & Lemeshow, S. (2000). Applied logistic regression (2nd ed.). Wiley. https://doi.org/10.1002/0471722146

Jayaprakash, S. M., Moody, E. W., Lauria, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. Journal of Learning Analytics, 1(1), 6–47. https://doi.org/10.18608/jla.2014.11.3

Joshi, S. (2024). Dropout trends and institutional responses at affiliated campuses of Tribhuvan University: A case study [Unpublished master's thesis]. Tribhuvan University.

Kabathova, J., & Drlik, M. (2021). Towards predicting student's dropout in university courses using different machine learning techniques. Applied Sciences, 11(7), 3130. https://doi.org/10.3390/app11073130

Kalita, P., Sarma, H. K. D., & Talukdar, A. K. (2025). Explainable BiLSTM for student dropout prediction: A SHAP-based analysis. IEEE Transactions on Learning Technologies, 18(2), 245–258. https://doi.org/10.1109/TLT.2025.3350817

Katel, S., & Katel, S. (2024). Prospects of predictive analytics in Nepali higher education: A conceptual framework for student retention management. Journal of Innovations in Business and Education, 6(1), 45–58.

Kehm, B. M., Larsen, M. R., & Sommersel, H. B. (2019). Student dropout from universities in Europe: A review of empirical literature. Hungarian Educational Research Journal, 9(2), 147–164. https://doi.org/10.1556/063.9.2019.1.18

Kemper, L., Vorhoff, G., & Wigger, B. U. (2020). Predicting student dropout: A machine learning approach. European Journal of Higher Education, 10(1), 28–47. https://doi.org/10.1080/21568235.2020.1718520

Kruger, M., Seker, S., & Schmidt, T. (2023). Explainable machine learning for dropout prediction in higher education: An institutional decision support framework. Computers & Education: Artificial Intelligence, 5, 100158. https://doi.org/10.1016/j.caeai.2023.100158

Kumar, V., Singh, S., & Sharma, A. (2023). Deep learning ensemble for MOOC dropout prediction: Impact of dataset size on model performance. Expert Systems with Applications, 212, 118823. https://doi.org/10.1016/j.eswa.2022.118823

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1909–1918). ACM. https://doi.org/10.1145/2783258.2788620

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521, 436–444. https://doi.org/10.1038/nature14539

Liu, Y., Chen, H., & Wang, Z. (2025). Dilated convolutional attention network for MOOC dropout prediction. IEEE Transactions on Neural Networks and Learning Systems, 36(1), 123–136. https://doi.org/10.1109/TNNLS.2024.3412987

Madai, B., Ghimire, R., & Thapa, M. (2025). Institutional factors contributing to student dropout at Kailali Multiple Campus, Far Western University. Journal of Higher Education Research Nepal, 4(1), 33–51.

Mduma, N. (2023). Educational data preprocessing for dropout prediction: Best practices and recommendations. IEEE Access, 11, 22341–22358. https://doi.org/10.1109/ACCESS.2023.3251785

Mduma, N., Kalegele, K., & Machuve, D. (2019). A survey of machine learning approaches and techniques for student dropout prediction. Data Science Journal, 18(1), 14. https://doi.org/10.5334/dsj-2019-014

Ministry of Education, Science and Technology. (2023). Flash report on education statistics 2022/23. Government of Nepal.

Mitchell, T. M. (1997). Machine learning. McGraw-Hill.

National Student Clearinghouse Research Center. (2019). Completing college: A national view of student completion rates — fall 2013 cohort. National Student Clearinghouse.

Neupane, B. (2024). Dropout among Bachelor of Education students: A study of contributing factors at selected campuses of Nepal. Journal of Educational Research and Practice, 14(1), 55–71.

OECD. (2020). Education at a glance 2020: OECD indicators. OECD Publishing. https://doi.org/10.1787/69096873-en

Pascarella, E. T., & Terenzini, P. T. (1991). How college affects students: Findings and insights from twenty years of research. Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). How college affects students: A third decade of research (Vol. 2). Jossey-Bass.

Pecuchova, M., & Drlik, M. (2023). Deep learning for student performance prediction: A systematic review. Computers & Education, 197, 104742. https://doi.org/10.1016/j.compedu.2023.104742

Pekrun, R. (2011). Emotions as drivers of learning and cognitive development. In R. A. Calvo & S. K. D'Mello (Eds.), New Perspectives on Affect and Learning Technologies (pp. 23–39). Springer. https://doi.org/10.1007/978-1-4419-9625-1_3

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. Expert Systems with Applications, 41(4), 1432–1462. https://doi.org/10.1016/j.eswa.2013.08.042

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), 37–63. https://doi.org/10.9735/2229-3981

Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106. https://doi.org/10.1007/BF00116251

Realinho, V., Machado, J., Baptista, L., & Martins, M. V. (2022). Predicting student dropout and academic success. Data, 7(11), 146. https://doi.org/10.3390/data7110146

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C, 40(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery, 10(3), e1355. https://doi.org/10.1002/widm.1355

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673–2681. https://doi.org/10.1109/78.650093

Siemens, G. (2013). Learning analytics: The emergence of a discipline. American Behavioral Scientist, 57(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Subedi, L. (2022). Factors influencing student dropout in Nepali higher education: A quantitative study. Tribhuvan University Journal, 37(1), 87–103.

Subedi, S. (2024). Socio-cultural determinants of female student dropout in higher education: Evidence from Nepal. Gender and Education, 36(2), 112–130. https://doi.org/10.1080/09540253.2023.2289456

Sweet, R. (1986). Student dropout in distance education: An application of Tinto's model. Distance Education, 7(2), 201–213. https://doi.org/10.1080/0158791860070204

Tamada, M. M., Larcher, M. A., Lima, R. V., & Falci, S. H. (2019). Predicting and reducing dropout in virtual learning using machine learning and gamification. Proceedings of the IEEE Frontiers in Education Conference (pp. 1–9). IEEE. https://doi.org/10.1109/FIE43999.2019.9028558

Tan, M., Liu, B., & Zhang, J. (2022). Intelligent early alert systems in distance learning: Design, implementation, and student retention outcomes. Distance Education, 43(2), 289–308. https://doi.org/10.1080/01587919.2022.2056274

Tharu, D. (2024). Teaching quality, student engagement, and academic performance in far-western Nepal: A mixed-methods inquiry. Journal of Education and Research, 14(2), 67–89.

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

Tinto, V. (1987). Leaving college: Rethinking the causes and cures of student attrition. University of Chicago Press.

Tinto, V. (1993). Leaving college: Rethinking the causes and cures of student attrition (2nd ed.). University of Chicago Press.

UNESCO. (2021). Global education monitoring report 2021/2: Non-state actors in education. UNESCO Publishing.

University Grants Commission. (2022). Higher education review: Annual report 2021/22. University Grants Commission, Nepal.

Villar, A., & de Andrade, C. R. V. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study. Discover Education, 3(1), 8. https://doi.org/10.1007/s44217-024-00099-w

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30(1), 79–82. https://doi.org/10.3354/cr030079

World Bank. (2018). Higher education for development: An evaluation of the World Bank Group's support. World Bank. https://doi.org/10.1596/978-1-4648-1260-4

Yükselturk, E., Ozekes, S., & Türel, Y. K. (2014). Predicting dropout student: An application of data mining methods in an online education program. European Journal of Open, Distance and E-Learning, 17(1), 118–133. https://doi.org/10.2478/eurodl-2014-0008