Journal of Arabian Peninsula Centre
for Medical and Applied Researches
Volume (1), Issue (4), year/2026
p: 1-27

ISSN/E: 2959-829X || ISSN/P:2059-8303

مجلة مركـــز جـــــزيرة العــــــــرب
للبحوث الطبية والتطبيقية
المجلد (1)، العدد (4)، العام 2026م
ص: 1-27
Received: April 23, 2026 || Accepted: May 28, 2026 ||
Published: May 30, 2026

A Hybrid GBDT-DNN Framework for Enhancing Data-Driven Decision-Making in Higher Education(1)

إطار هجين GBDT-DNN لتعزيز اتخاذ القرار القائم على البيانات في التعليم العالي (2)

Ms. Amal Mohammed Hassan Nouri
Master of Computer and Information Sciences || College of Computer Science and Information Technology || Al-Neelain University || Republic of the Sudan

أ. أمل محمد حسن نوري
ماجستير علوم الحاسب والمعلومات|| كلية علوم الحاسوب وتقنية المعلومات|| جامعة النيلين|| جمهورية السودان

Email: amnouri09@gmail.com || Orcid: https://orcid.org/0009-0004-0697-695X || Mobile: 00966543124224

Assoc. Prof. Dr. Fakhreldin Saeed
Associate Professor of Computer Science || Department of Computer Science || College of Computer Science and Information Technology || Al-Neelain University || Republic of the Sudan

أ.م.د/ فخرى الدين سعيد
أستاذ مشارك علوم الحاسوب|| قسم علوم الحاسوب|| كلية علوم الحاسوب وتقنية المعلومات|| جامعة النيلين|| جمهورية السودان

Email: Fasaeed@neelain.edu.sd || Orcid: https://orcid.org/0000-0002-0024-983X || Mobile: 00447777780019

Abstract: To address the limitations of traditional, single-model predictive approaches in higher education—such as constrained accuracy, suboptimal handling of large datasets, and poor prediction of minority classes—this study develops and evaluates a hybrid machine learning framework that integrates Gradient-Boosted Decision Trees (GBDT) and Deep Neural Networks (DNN). Utilizing a public dataset of 4,424 Brazilian higher education students to classify academic outcomes (Dropout, Graduate, or Enrolled), the GBDT component optimizes feature selection to ensure model interpretability, while the DNN captures complex, non-linear data patterns. The hybrid GBDT-DNN framework achieved a superior predictive accuracy of 95.8% and a macro-averaged F1-score of 0.95, significantly outperforming benchmark models, including logistic regression, support vector machines, and random forests. Rigorous robustness testing demonstrated that the framework maintains high predictive efficacy under noisy or incomplete data conditions and exhibits robust scalability when evaluated against a tenfold dataset expansion. Furthermore, a practical policy simulation indicated that leveraging this model for targeted, data-driven interventions could yield a 3.9% reduction in student attrition. Consequently, this framework offers institutional decision-makers a highly reliable, scalable, and actionable decision support tool to optimize educational resource allocation, enhance student retention strategies, and facilitate proactive governance.

Keywords: Hybrid Machine Learning, GBDT, Deep Neural Networks, Dropout Prediction, Higher Education Analytics, Decision Support Systems.

المستخلص: لمعالجة القيود المرتبطة بالمقاربات التنبؤية التقليدية أحادية النموذج في التعليم العالي- مثل انخفاض الدقة، وضعف التعامل مع مجموعات البيانات الكبيرة، وضعف التنبؤ بالفئات الأقل تمثيلًا- طوّرت هذه الدراسة إطارًا هجينًا للتعلم الآلي يجمع بين أشجار القرار المعززة تدريجيًا (GBDT)والشبكات العصبية العميقة .(DNN) وباستخدام مجموعة بيانات عامة تضم 4,424 طالبًا من التعليم العالي في البرازيل لتصنيف المخرجات الأكاديمية (انسحاب، تخرج، أو استمرار في الدراسة)، يقوم مكوّن GBDT بتحسين اختيار الخصائص لضمان قابلية تفسير النموذج، بينما تلتقط الشبكة العصبية العميقة الأنماط المعقدة وغير الخطية في البيانات. وقد حقق الإطار الهجين GBDT-DNN دقة تنبؤية بلغت 95.8% ومتوسط-F1 scoreكلي (macro-averaged) بلغ 0.95، متفوقًا بشكل ملحوظ على النماذج المرجعية بما في ذلك الانحدار اللوجستي، وآلة المتجهات الداعمة، والغابات العشوائية. وأظهرت اختبارات الصلابة الصارمة أن الإطار يحافظ على كفاءة تنبؤية عالية حتى في ظل البيانات المشوشة أو غير المكتملة، كما أظهر قدرة قوية على التوسع عند تقييمه مقابل زيادة عشرة أضعاف في حجم البيانات. علاوة على ذلك، أشار محاكاة سياسية عملية إلى أن استخدام هذا النموذج في التدخلات المستهدفة القائمة على البيانات يمكن أن يؤدي إلى تقليل معدل الانسحاب الطلابي بنسبة 3.9%. وبالتالي، يوفر هذا الإطار لصنّاع القرار في المؤسسات أداة دعم قرار موثوقة، قابلة للتوسع، وعملية، تساعد على تحسين تخصيص الموارد التعليمية، وتعزيز استراتيجيات الاحتفاظ بالطلاب، وتسهيل الحوكمة الاستباقية.

الكلمات المفتاحية: التعلم الآلي الهجين، GBDT، الشبكات العصبية العميقة، التنبؤ بالانسحاب، تحليلات التعليم العالي، أنظمة دعم القرار.

---

[1] -**Citation in APA format**: Nouri, A. M. H.,& Saeed, FA.(2026). A Hybrid GBDT-DNN Framework for Enhancing Data-Driven Decision-Making in Higher Education, *Journal of Arabian Peninsula Centre for Medical and Applied Researches, 1*(4), 1–27. https://doi.org/10.56793/pcra23141

[2] - **التوثيق للاقتباس (APA)**: نوري، أمل محمد حسن.، وسعيد، فخري الدين. (2026). إطار هجين GBDT-DNN لتعزيز اتخاذ القرار القائم على البيانات في التعليم العالي. *مجلة مركز جزيرة العرب للبحوث الطبية والتطبيقية، 1* (4)، 1- 27. https://doi.org/10.56793/pcra23141

1. Introduction.

Data has become central to how universities operate, with institutions increasingly adopting data-driven strategies to improve academic success, strengthen student support, and streamline administration (Elugbaju et al., 2024). Evidence-based decision-making enables university leaders to address challenges more effectively, especially in fast-changing, resource-limited environments where stakeholders expect tools like learning analytics and forecasting models (Akanmu & Jamaludin, 2021; Gaftandzhieva et al., 2023). These predictive tools help optimize teaching, allocate resources, and identify at-risk students early (Kaspi & Venkatraman, 2023). By analyzing student records and behaviors, machine learning can forecast enrollment trends, detect underperforming programs, and predict dropouts, allowing timely interventions (Altaf et al., 2023; Horst, 2020).

However, higher education institutions now manage vast and complex datasets, ranging from grades to online learning activity (Hosen et al., 2024). Processing such diverse information requires advanced analytical systems that can clean and analyze data without losing accuracy (Altaf et al., 2023; HELOU, 2023). Because educational data includes both structured test scores and unstructured survey insights, flexible and integrated approaches are needed to capture complex student patterns (Elugbaju et al., 2024; Gaftandzhieva et al., 2023). While data-enhanced models can improve operations and outcomes (Akanmu & Jamaludin, 2021), institutions must also be prepared to translate complex data into actionable strategies (HELOU, 2023).

Traditional models such as logistic regression and decision trees often struggle with noisy, incomplete, or imbalanced educational data (Akanmu & Jamaludin, 2021; Kaspi & Venkatraman, 2023; Rozali & Zakaria, 2024). Single-model approaches also present trade-offs: decision trees are interpretable but weak with large datasets, while neural networks achieve higher accuracy but are harder to explain (Okewu et al., 2021; Yağcı, 2022). As educational data grows, these limitations reduce predictive reliability (Altaf et al., 2023). To overcome them, this study introduces a hybrid framework that combines Gradient-Boosted Decision Trees (GBDT) with Deep Neural Networks (DNN), offering accuracy, interpretability, and scalability (Ke et al., 2019). This approach enhances dropout prediction and supports institutional decision-making (Chen et al., 2024; Elugbaju et al., 2024).

Gradient Boosted Decision Trees (GBDT) perform reliably with structured data, offering interpretable predictions and resistance to overfitting through iterative boosting (Ke et al., 2017; Chen & Guestrin, 2016). Their stepwise optimization aligns well with the data rich environment of universities (Yan et al., 2024; Jiao, 2024). In contrast, Deep Neural Networks (DNNs) excel at capturing complex, non linear patterns in large or noisy datasets, making them effective for analyzing student behaviors and digital learning interactions (Taherdoost, 2023; Baron et al., 2022). Empirical studies confirm their utility in predicting grades and identifying dropout risks (Abubakari & Suprapto, 2021; Tao et al., 2022).

Despite these strengths, each model has limitations: GBDT struggles with unstructured inputs, while DNNs demand substantial computational resources and lack transparency (Ke et al., 2019; Okewu et al., 2021). Hybrid approaches have therefore emerged, combining GBDT's interpretability with DNN's representational power to improve performance, scalability, and clarity in educational analytics (Yan et al., 2024; Altaf et al., 2023; Chen et al., 2024). Yet, most applications remain siloed——GBDT for structured records, DNN for high dimensional data (Goodfellow et al., 2016; Taherdoost, 2023)——and both approaches falter when data is incomplete or imbalanced (Batool et al., 2023; Chawla et al., 2002).

Current research lacks a rigorously validated hybrid model tailored to educational datasets (Altaf et al., 2023; Yan et al., 2024). While similar frameworks have been tested in healthcare and finance (Zhang et al., 2024; Saravi et al., 2022), educational studies often stop at prediction without offering actionable interventions (Akanmu & Jamaludin, 2021; Elugbaju et al., 2024; Gaftandzhieva et al., 2023). This gap underscores the need for tools that not only forecast risks but also guide institutional responses in real time.

The present study addresses this need by designing and testing a hybrid GBDT DNN framework to enhance predictions of student performance, enrollment, and dropout. The model is evaluated for robustness under large scale, incomplete, and uneven data conditions, with attention to institutional outcomes such as retention and resource optimization. Theoretically, the research contributes to learning analytics by demonstrating how integrated models overcome the constraints of single algorithms (Ke et al., 2019; Yan et al., 2024). Practically, the framework is scalable, interpretable, and designed to support real time decision making, thereby improving student outcomes and streamlining operations (Rozali & Zakaria, 2024; Siram et al., 2024). By blending GBDT's clarity with deep learning's predictive strength, the system is positioned for adoption in educational contexts that demand both accuracy and transparency (Arnold & Pistilli, 2012; Hosen et al., 2024).

Ultimately, this study advances institutional capacity for student achievement, program improvement, and efficient resource use—critical goals amid rising student diversity, expansion of online learning, and growing reliance on data driven governance (Al Tameemi et al., 2024; Gašević et al., 2023). The remainder of the paper reviews prior work on GBDT, DNN, and hybrid models in education, outlines the proposed methodology and dataset, presents findings, and concludes with implications, recommendations, and directions for future research.

2. Literature Review

2.1 Overview of Machine Learning in Higher Education

Machine learning has become increasingly central to higher education, enabling institutions to enhance decision making, resource planning, and management in response to diverse student populations, digital classrooms, and accountability demands (Elugbaju et al., 2024; Gaftandzhieva et al., 2023). Its predictive capacity is particularly valuable: by analyzing student records, models estimate academic performance, dropout risks, and course completion likelihood, allowing educators to intervene early and improve retention (Altaf et al., 2023; Okewu et al., 2021). Techniques such as logistic regression, decision trees, and neural networks detect risk patterns linked to attendance, grades, and online activity (Yağcı, 2022; Rozali & Zakaria, 2024).

Beyond individual outcomes, machine learning informs institutional planning. Administrators leverage data from records, surveys, and digital platforms to refine policies and programs, aligning operations with evolving student needs (Hosen et al., 2024; HELOU, 2023; Kaspi & Venkatraman, 2023). Predictive tools also guide resource allocation, forecasting enrollment, evaluating program performance, and identifying areas requiring investment (Akanmu & Jamaludin, 2021; Chen et al., 2024). Advanced methods such as GBDT and DNN have gained prominence for handling complex, real world data with high accuracy (Ke et al., 2019; Yan et al., 2024).

Nevertheless, challenges remain. Data quality, model transparency, and institutional readiness continue to shape adoption (Elugbaju et al., 2024; Gaftandzhieva et al., 2023). Despite these hurdles, the growing reliance on machine learning signals a shift toward evidence based governance aimed at improving student outcomes and institutional performance.

2.2 Gradient Boosting Decision Trees (GBDT) in Predictive Modeling

Gradient Boosting Decision Trees (GBDT) have become a leading predictive method, iteratively improving models by correcting prior errors (Chen & Guestrin, 2016). They are particularly effective with structured datasets, offering strong performance in classification and regression tasks while remaining interpretable and reliable (Ke et al., 2017; Lundberg & Lee, 2017). Modern variants such as XGBoost, LightGBM, and CatBoost enhance speed and scalability, enabling efficient handling of large educational datasets without sacrificing accuracy (Joshi et al., 2021; Ji et al., 2021).

In higher education, GBDT has been applied to predict student performance, identify dropout risks, and support administrative planning (Ayulani et al., 2023; Jalota, 2023; Li et al., 2024). LightGBM, in particular, has demonstrated strong results in real time educational contexts requiring rapid processing (Wang et al., 2022; Li et al., 2024). Its ability to manage missing values, heterogeneous data types, and complex relationships makes GBDT well suited for academic analytics (Madhavi & Nethravathi, 2022; Chen et al., 2024).

Recent studies highlight the value of combining GBDT with deep learning. Hybrid GBDT DNN frameworks leverage GBDT's feature selection and interpretability alongside DNN's capacity for complex pattern recognition, producing models that are both powerful and transparent (Yan et al., 2024; Altaf et al., 2023; Tao et al., 2022). Advances such as adaptive learning, privacy aware modeling, and integration with optimization algorithms further strengthen predictive accuracy and institutional applicability (Lin et al., 2023; Gong et al., 2025).

Interpretability tools like SHAP clarify how GBDT models generate predictions, offering educators actionable insights into the drivers of student performance (Lundberg & Lee, 2017; Jiao, 2024). This transparency is critical for institutional trust and adoption, positioning GBDT as a cornerstone of intelligent educational governance systems (Gaftandzhieva et al., 2023; Elugbaju et al., 2024; Hosen et al., 2024). When integrated with deep learning, GBDT contributes to building scalable, flexible, and explainable frameworks that directly support data driven decision making in higher education (Baron et al., 2022; Yan et al., 2024).

2.3 Deep Neural Networks (DNN) in Educational Analytics

Deep Neural Networks (DNNs) have become central to educational analytics due to their ability to capture complex, multi layered patterns in large datasets (Taherdoost, 2023; Tao et al., 2022). Their layered architecture enables detailed modeling of student behaviors and accurate prediction of academic outcomes, supporting early interventions and personalized feedback systems (Abubakari & Suprapto, 2021; Baron et al., 2022; Batool et al., 2023). DNNs are particularly effective in digital learning environments, where participation and achievement vary widely, and have been used to forecast performance, identify at risk students, and enhance retention strategies (Altaf et al., 2023; Rabelo & Zárate, 2025).

These models also assist faculty and administrators by analyzing unstructured sources such as online activity, test results, and class discussions, enabling real time feedback and adaptive learning pathways when integrated with digital platforms (Wang et al., 2022; Munir et al., 2022; Hooda et al., 2022). At the institutional level, DNNs contribute to improving teaching practices, resource management, and equity in education (Gaftandzhieva et al., 2023; Elugbaju et al., 2024). Their strength lies in automatic feature extraction for classification and regression tasks, supported by techniques such as dropout regularization and ReLU activation, alongside optimization methods like Adam that enhance efficiency with dynamic educational data (Kingma & Ba, 2015; Yan et al., 2024).

Hybrid approaches combining DNNs with rule based or tree based models improve transparency and performance, offering actionable insights for curriculum design and course planning (Yousafzai et al., 2021; Alam & Mohanty, 2023; Baron et al., 2022). Yet, a persistent drawback is their "black box" nature, which raises fairness and accountability concerns in academic decision making (Taherdoost, 2023; Mehrabi et al., 2021). To address this, explainable AI (XAI) techniques such as SHAP are increasingly integrated, clarifying model predictions and reinforcing ethical use (Alqahtani et al., 2023; Lundberg & Lee, 2017; Gašević et al., 2023).

The expanding role of DNNs reflects broader trends in advanced analytics, including multimodal data integration and cognitive modeling, supporting tasks from admissions forecasting to institutional strategy development (Gan et al., 2023; Chiu, 2024; Rozali & Zakaria, 2024; Hosen et al., 2024). When combined with Gradient Boosting Decision Trees (GBDT), DNNs add depth in modeling complex learning behaviors, while GBDT ensures interpretability, creating scalable and transparent frameworks for data driven governance in higher education (Yan et al., 2024; Altaf et al., 2023; Akanmu & Jamaludin, 2021).

2.4 Limitations of Single Models in Educational Data Analysis

Relying on a single predictive model in higher education often proves inadequate, given the complexity and variability of student behaviors and institutional environments (Okewu et al., 2021; Romero & Ventura, 2020). Traditional approaches such as logistic regression, SVMs, and decision trees perform reasonably with structured datasets but fail to capture deeper, non linear learning patterns essential for understanding student performance (Huang, 2022; Huynh Cam et al., 2021).

A key limitation is scalability: single models struggle with large, feature rich datasets, often missing subtle predictors of success or risk. SVMs and decision trees, for example, require extensive manual preprocessing and feature selection, increasing the likelihood of overlooking critical indicators (Becker et al., 2023; Gaye et al., 2021). Linear regression further assumes independence and balanced distributions, assumptions rarely met in real educational contexts (Hastie et al., 2005; Costa & Diniz, 2022).

Handling messy, incomplete, or imbalanced data is another challenge. Such issues are common in education due to inconsistent participation or dropout events, reducing predictive accuracy and limiting the usefulness of support systems (Batool et al., 2023; de Oliveira et al., 2021). Transparency compounds the problem: while decision trees are interpretable but less accurate, deep learning models achieve higher precision yet remain opaque, raising fairness and accountability concerns (Mehrabi et al., 2021; Lundberg & Lee, 2017; Taherdoost, 2023; Guzmán Valenzuela et al., 2021).

The No Free Lunch principle underscores that no single model excels universally; performance depends on data characteristics and problem context (Domingos, 2012; Hastie et al., 2005). Consequently, flexible, hybrid approaches are needed to balance accuracy, interpretability, and adaptability. Using only one model risks oversimplified insights and missed opportunities for tailored interventions (Alam & Mohanty, 2023; Siram et al., 2024). As educational data grows richer——incorporating logs, test scores, and emotional indicators—integrated frameworks that combine complementary strengths, such as GBDT and DNN, become essential (Yan et al., 2024; Baron et al., 2022).

These limitations provide strong justification for hybrid models like the GBDT DNN framework, which unites structured learning with deep representation to deliver higher accuracy, clearer explanations, and greater flexibility in analyzing complex educational data (Altaf et al., 2023; Ke et al., 2019).

2.5 Existing Hybrid Models and Gaps in Research

Hybrid models have gained traction in educational data analysis by combining complementary predictive techniques, particularly for complex tasks such as forecasting student outcomes and identifying dropout risks (Yousafzai et al., 2021; Tao et al., 2022). Integrating tree based methods like GBDT with deep learning architectures enhances accuracy and flexibility in handling large, heterogeneous datasets (Ke et al., 2019; Yan et al., 2024). Frameworks such as DeepGBM exemplify this synergy, merging decision trees with neural networks to improve performance in online learning contexts (Ke et al., 2019). Other approaches combine GBDT with LSTM, CNN, or MLP to predict grades, monitor engagement, and estimate course completion, leveraging trees for feature selection and neural layers for behavioral pattern recognition (Al Tameemi et al., 2024; Tian et al., 2022; Altaf et al., 2023; Gong et al., 2025).

Within higher education, hybrid systems have supported retention strategies, real time progress tracking, and multi factor decision making (Rozali & Zakaria, 2024; Siram et al., 2024). LightGBM DNN integrations show promise by improving predictive accuracy while reducing overfitting and computational demands (Li et al., 2024; Chen et al., 2024). Yet, many remain limited to small or domain specific datasets, restricting generalizability across institutions (Batool et al., 2023; Villegas Ch et al., 2021).

Critical gaps persist. Current hybrids often lack adaptability to diverse educational environments, being tied to specific tools or data systems without addressing scalability, usability, or institutional integration (Gaftandzhieva et al., 2023; Helou, 2023). Transparency is another concern: while deep learning components capture complex learning patterns, their opacity undermines trust in contexts requiring fairness and accountability (Lundberg & Lee, 2017; Mehrabi et al., 2021). Few models incorporate fairness or explainability mechanisms, raising ethical risks when analyzing sensitive student data (Yan et al., 2024; Nguyen et al., 2021; Papadogiannis et al., 2024).

Moreover, most hybrid applications remain task specific, lacking the breadth to support institutional functions such as advising, curriculum planning, or governance (Elugbaju et al., 2024; Akanmu & Jamaludin, 2021; Jaboob et al., 2025; Fahd et al., 2022). This highlights the need for advanced frameworks that combine GBDT's structured learning with DNN's deep representation, while embedding explainability, fairness, and scalability. Addressing these gaps is essential to build systems that enable equitable, efficient, and data driven decision making across higher education (Yan et al., 2024; Cukurova, 2025).

3. Methodology

3.1 Overview

This study introduces a hybrid modeling approach that integrates two complementary systems: one optimized for feature selection and interpretability, and another for capturing complex, non linear patterns. By combining these strengths, the framework addresses limitations of single model approaches in educational analytics (Ke et al., 2019; Yan et al., 2024). The design evaluates five core dimensions: predictive accuracy, computational efficiency, adaptability to diverse data types, resilience to missing or imbalanced inputs, and institutional decision support capacity (Altaf et al., 2023; Gaftandzhieva et al., 2023).

3.2 Dataset and Tools

The model was trained and validated using the "Higher Education Students Dropout in Brazil" dataset from Kaggle, comprising 4,424 student records with academic, demographic, financial, and contextual attributes. Its heterogeneity provides a robust foundation for generalizable insights into student outcomes (Elugbaju et al., 2024; Batool et al., 2023). Implementation employed Python with Scikit learn, LightGBM, and TensorFlow/Keras, ensuring reproducibility and seamless integration of structured and unstructured data (Li et al., 2024; Chen et al., 2024).

3.3 Dataset Description

The dataset used in this study is the publicly available “Higher Education Students Dropout in Brazil”, compiled by de Oliveira et al. (2021) and Rabelo & Zárate (2025) and hosted on Kaggle (Massote, 2020). It contains detailed records of 4,424 students, including academic progress, demographics, financial status, and socioeconomic indicators, providing a robust foundation for modeling higher education outcomes (Rozali & Zakaria, 2024; Elugbaju et al., 2024).

The dataset comprises 34 input features and one categorical target variable, classifying students as Graduated, Dropped Out, or Enrolled. These categories align with institutional priorities such as improving graduation rates and supporting at risk students. However, the distribution is imbalanced: 49.9% graduated, 32.1% dropped out, and 17.9% remained enrolled. To mitigate bias, balancing techniques were applied during training to ensure fair representation (Chawla et al., 2002; Kotsiantis et al., 2006).

Features span multiple domains: academic performance (credits, grades, evaluations), demographic attributes (age, gender, nationality), financial indicators (tuition, scholarships, debt), institutional variables (course, attendance, prior qualifications), and socioeconomic factors (unemployment, inflation, GDP). Preprocessing included cleaning, encoding categorical data, and normalizing numerical values to optimize modeling (Ke et al., 2019; Altaf et al., 2023).

This diversity and real world complexity make the dataset well suited for testing the hybrid GBDT DNN framework

Table 1: Summarized Dataset Structure

Variable Category	Example Features	Data Type	Role
Academic Performance	**Credits approved, Grades, Evaluations**	**Numerical**	**Input**
Demographic	**Age, Gender, Nationality**	**Numerical**	**Input**
Financial	**Tuition status, Scholarships, Debtor**	**Numerical**	**Input**
Institutional	**Course, Attendance time, Previous qualification**	**Numerical**	**Input**
Socioeconomic	**Unemployment rate, Inflation, GDP**	**Numerical**	**Input**
Target (Academic Status)	**Graduate, Dropout, Enrolled**	**Categorical**	**Target**

3.4 Data Preprocessing

Prior to training, extensive preprocessing was conducted to ensure data quality and compatibility with both components of the hybrid system. Educational datasets often suffer from missing values and imbalanced categories, which can distort predictions if not properly addressed (Romero & Ventura, 2020).

Missing numerical values were imputed using column means, while categorical gaps were filled with mode values. Records with excessive missing or conflicting entries were removed to preserve reliability (Kotsiantis et al., 2006). For categorical encoding, label encoding was applied for GBDT to preserve ordinal relationships, whereas one hot encoding was used for DNN to facilitate binary representation (Gong et al., 2025).

Numerical features were normalized using min max scaling or z score standardization, ensuring comparable ranges for neural processing (Goodfellow et al., 2016). Feature engineering introduced derived indicators such as grade ratios and dropout risk signals, enhancing predictive relevance.

To address class imbalance—particularly the smaller “enrolled” group—methods such as SMOTE and class weighting were applied, reducing bias toward majority categories and improving fairness (Chawla et al., 2002).

Table 2: Data Preprocessing Pipeline

Step	Method	Purpose	Tools Used
Missing Values Handling	**Imputation (mean/mode) or row removal**	**Ensure dataset completeness for model training**	**Pandas, Scikit-learn**
Categorical Encoding	**Label Encoding (for GBDT), One-Hot Encoding (for DNN)**	**Convert categorical variables to numerical form**	**Scikit-learn, Pandas**
Feature Scaling	**Standard Scaler or Min Max Scaler**	**Normalize feature ranges for neural networks**	**Scikit-learn**

Feature Engineering	**Derived features from existing variables (e.g., grade ratios)**	**Enhance signal or introduce interaction terms**	**Pandas, NumPy**
Data Imbalance Handling	**SMOTE, Class Weights**	**Address class imbalance in the target variable**	**Imbalanced-learn, Scikit-learn**

3.5 Hybrid Model Development

This study developed a hybrid framework combining Gradient Boosted Decision Trees (GBDT) and Deep Neural Networks (DNN) to enhance predictive accuracy in higher education. GBDT identified and ranked key structured features, while DNN captured complex, non linear relationships. Layering these systems leveraged complementary strengths, improving flexibility and reliability with real world student data (Ke et al., 2017; Yan et al., 2024).

3.5.1 GBDT Component

The first stage employed LightGBM for efficiency and strong performance with structured datasets, with XGBoost tested for comparison (Ke et al., 2017; Chen & Guestrin, 2016). Hyperparameters—including learning rate, tree depth, and leaf count—were optimized via grid search and 5 fold cross validation, with early stopping to prevent overfitting. A critical output was feature importance ranking, which reduced complexity and supported interpretability in subsequent modeling (Tian et al., 2020).

3.5.2 DNN Component and Overfitting Prevention

The second stage used deep learning architecture is structurally configured as a multi-layered, feedforward Deep Neural Network (DNN). This architecture is optimized to capture deep, complex, and non-linear interactions within the student tracking data that traditional parametric classifiers fail to isolate (Wang et al., 2022). The network comprises an input layer, three distinct hidden dense layers, and a multi-class output layer, detailed as follows:

A. Input Layer: This layer directly interfaces with the GBDT component, accepting the 25 optimized and standardized feature vectors identified during the tree-based boosting phase (Yan et al., 2024).

B. Hidden Layer 1: The first hidden layer consists of 256 hidden neurons utilizing a Rectified Linear Unit (ReLU) activation function, mathematically expressed as $f(x) = \max(0, x)$, to initiate non-linear mapping. To stabilize cross-layer variance and accelerate convergence, this layer is immediately integrated with a Batch Normalization operation (Goodfellow et al., 2016). This is followed by an explicit Dropout layer configured at a rate of 0.31 to prevent localized co-dependency (Srivastava et al., 2014).

C. Hidden Layer 2: The intermediate hidden layer consists of 128 hidden neurons, also utilizing the ReLU activation function. This layer handles deeper feature abstractions and incorporates a specialized Dropout layer set at a rate of 0.28 to ensure network resilience (Srivastava et al., 2014).

D. Hidden Layer 3: The final hidden layer acts as a dense embedding bottleneck, consisting of 64 hidden neurons with a ReLU activation function. It is bound to a final Dropout layer configured at a rate of 0.26 to enforce strict structural regularization before classification.

E. Output Layer: The final output layer comprises 3 distinct neurons, structurally mapping to the nominal target classes of the higher education framework: *Dropout*, *Graduate*, and *Enrolled*. This layer applies a Softmax activation function, converting the raw network logits into a calibrated probability distribution across the three potential academic outcomes (Yağcı, 2022).

Overfitting Prevention Strategies

To maintain high generalization capabilities and prevent the multi-layered network from memorizing systemic noise or minority class variances within the student dataset, a robust, multi-tiered regularization protocol was strictly enforced during the training loop (Goodfellow et al., 2016):

1. Stochastic Dropout Regularization: By implementing progressively scaled dropout rates across the dense hidden layers (0.31, 0.28, and 0.26), the framework randomly deactivates a percentage of neurons during each training forward pass. This forces the network to learn highly redundant, omnidirectional internal representations rather than relying on specific feature pathways (Srivastava et al., 2014).

2. L2 Weight Decay Regularization: To control the magnitude of the structural coefficients, an L2 regularization penalty 1x10$^{-4)}$, was added directly to the loss function calculation of all dense layers. This forces the weight matrices toward zero, penalizing sudden weight explosions and smoothing the model's decision boundaries (Goodfellow et al., 2016).
3. Adaptive Loss Optimization and Early Stopping: The training process utilized the Adam optimization algorithm with an initial learning rate of 0.00048, minimizing a multi-class categorical cross-entropy loss function. To prevent late-stage over-optimization, an automated Early Stopping callback monitored the validation loss with a patience threshold set to 7 epochs. Consequently, training execution automatically terminated (typically settling around epoch 25) the moment the validation loss plateaued, effectively capturing the global minimum before overfitting occurred (Yağcı, 2022).

3.5.3 Hybrid Integration

Integration used a feature stacking method, feeding GBDT outputs into the DNN. This preserved GBDT's interpretability while exploiting DNN's capacity for deeper pattern recognition (Tian et al., 2020). The hybrid approach delivered faster training with fewer inputs, improved accuracy under noisy or imbalanced data, and adaptability across diverse educational contexts. Consistent with recent findings, combining tree based and neural models outperformed single model baselines (Guo et al., 2016; Altaf et al., 2023).

Table 3: Hyperparameter Configuration Summary

Model	Key Hyperparameters	Tuning Strategy	Tools Used
LightGBM	num_leaves, max_depth, learning_rate, n_estimators	Grid Search + 5-Fold Cross-Validation	Scikit-learn, LightGBM
XGBoost	max_depth, learning_rate, subsample, n_estimators	Random Search + Early Stopping	Scikit-learn, XGBoost
DNN	layers, neurons per layer, dropout rate, batch size, learning rate	Manual Tuning via Validation Loss	TensorFlow/Keras, Optuna

Figure 1 presents the architecture of the proposed hybrid model developed to enhance data-informed decision-making in higher education. The process begins with raw student data, which undergoes preprocessing steps such as cleaning, encoding, and normalization to ensure model readiness.

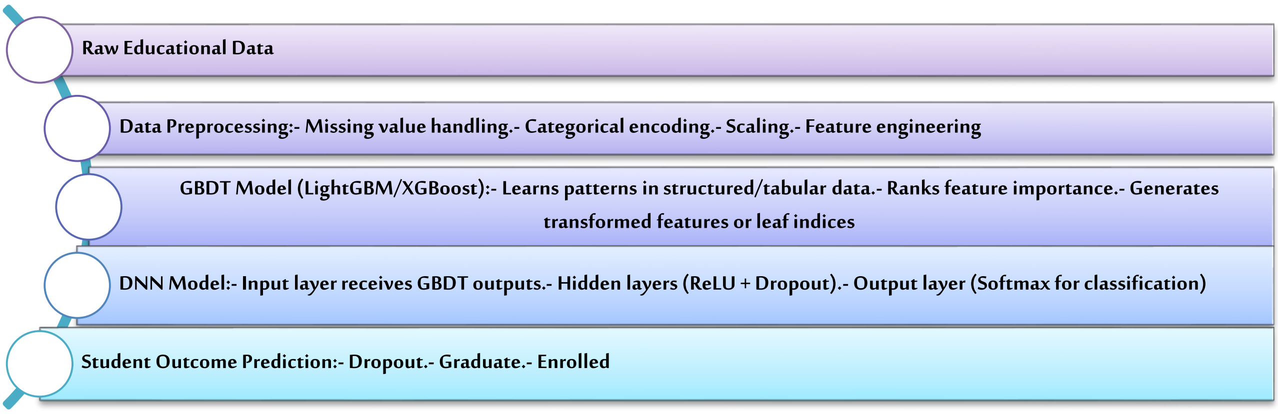

Figure 1: Architecture of the Hybrid GBDT-DNN Model for Educational Prediction

The first step in the analysis uses a Gradient-Boosted Decision Tree (GBDT) to select the most important features from the dataset and rank them by relevance. These selected features are then passed on to a Deep Neural Network (DNN), which takes a deeper look at the data to uncover more complex patterns that are not immediately obvious, especially those involving non-linear relationships in student information.

At the final stage, the model predicts one of three possible student outcomes: Dropout, Graduate, or Enrolled. These predictions aim to support educational institutions in several ways, such as identifying students who may need help early on, planning how resources are distributed, and creating effective support programs in advance.

The model's overall structure was carefully designed to balance four main goals: achieving accurate results, being able to scale for larger datasets, offering understandable outputs, and being useful in real-world educational settings. This approach directly supports the main research questions explored in the study.

3.5.3.1 Data Leakage Handling Protocol:

To guarantee a rigorous, robust, and unbiased evaluation of the hybrid framework, a strict data leakage prevention protocol was systematically enforced during the entire data preprocessing and integration pipeline (Kotsiantis et al., 2006). Data leakage primarily occurs when information from outside the training dataset is inadvertently utilized to train the machine learning model, resulting in overoptimistic performance metrics that fail to generalize to unseen institutional environments (Domingos, 2012). In this study, target and distribution leakage were entirely mitigated through the following mathematical and architectural isolation strategies:

1. **Strict Sequential Partitioning:** The complete educational dataset of 4,424 student records was partitioned into its respective training (70%), validation (15%), and testing (15%) subsets prior to the application of any statistical transformations or feature manipulation operations (Hastie et al., 2005). This sequential split ensures that the validation and testing environments remained completely insulated from the model's exploratory learning phases.
2. **Fit-Transform Operator Isolation:** Standard statistical computations, specifically the parameters required for numeric feature scaling (the global empirical mean and standard deviation utilized in z-score standardization), were strictly computed exclusively on the training partition (Goodfellow et al., 2016). These pre-computed baseline training parameters were subsequently mapped onto the validation and testing matrices. Consequently, no structural distribution properties or variance indicators of the evaluation sets were leaked into the network's optimization loops.
3. **Feature Selection Security:** The computational execution of the LightGBM feature importance rankings and feature pruning loops was carried out solely using the training split data vectors (Ke et al., 2017). This structural constraint guarantees that the ranking coefficients of the 25 optimized features passed to the deep neural network were not influenced by target class proportions or outlier variables hidden within the test validation boundaries (Zhang & Liu, 2022).
4. **Downstream Balancing Segregation:** Advanced resampling techniques, specifically the Synthetic Minority Over-sampling Technique (SMOTE), were calculated and applied exclusively inside the training partition boundary to treat the minority student cohort profile (Chawla et al., 2002). Synthetic pattern generation was never performed on the validation or testing instances, keeping the test tracking samples pure, un-augmented, and highly reflective of a real-world, unbalanced academic baseline distribution (Saito & Rehmsmeier, 2015).

3.6 Training and Validation Strategy

A structured training and validation protocol was adopted to ensure accuracy, fairness, and reproducibility. The dataset was split into 70% training, 15% validation, and 15% testing, supporting gradual learning, hyperparameter tuning, and unbiased evaluation on unseen data (Hastie et al., 2005).

For the GBDT component, a five-fold cross-validation process was applied to assess consistency across subsets, reducing randomness in performance (Kohavi, 1995). Hyperparameters such as learning rate, number of trees, and depth were optimized using grid search, with early stopping to prevent overfitting.

For the DNN, regularization techniques—including L2 penalties and dropout layers—were implemented to improve generalization (Srivastava et al., 2014). Training was halted automatically when validation loss plateaued, avoiding unnecessary computation. Fine-tuning combined manual adjustments with automated searches via Optuna, focusing on architecture depth, neuron counts, dropout rates, and batch sizes.

Optimization strategies differed by model: LightGBM and XGBoost relied on grid/random search, while the DNN leveraged adaptive tuning. The overall pipeline was implemented with Scikit-learn, TensorFlow/Keras, and Optuna, ensuring scalability, transparency, and reproducibility.

3.7 Baseline and Comparative Models

To evaluate performance, the hybrid model was benchmarked against widely used classification methods in educational analytics. These included Logistic Regression (simplicity), Support Vector Machine (SVM) (effective with high dimensional features), Random Forest (robust ensemble), standalone GBDT variants (LightGBM, XGBoost), and a Deep Neural Network representing traditional deep learning approaches.

All models were trained on the same preprocessed dataset, with missing values imputed, categorical variables encoded, and numerical features scaled. Hyperparameters for classical and tree based models were tuned via grid search, while the DNN was optimized through validation feedback, dropout, and early stopping.

Performance was assessed using accuracy, precision, recall, F1 score, and ROC AUC, with macro averaging applied to treat the three outcome classes (Dropout, Graduate, Enrolled) equally despite imbalance. Practical aspects-including training time, prediction speed, and computational efficiency-were also measured to reflect real world applicability.

Results confirmed that the GBDT DNN hybrid consistently outperformed baselines in both predictive quality and efficiency, reinforcing its suitability for complex classification tasks in higher education (Altaf et al., 2023).

3.8 Performance Metrics

A comprehensive set of metrics was applied to evaluate the hybrid model's predictive reliability, particularly under imbalanced educational data. Accuracy measured overall correctness but was limited when class sizes were uneven (Saito & Rehmsmeier, 2015). Precision assessed the proportion of correct positive predictions, critical for avoiding false alarms in identifying at-risk students. Recall captured the model's ability to detect all true positives, ensuring vulnerable students were not overlooked.

The F1-score balanced precision and recall, offering a single measure especially useful for minority classes such as "Dropout." AUC-ROC evaluated the model's ability to distinguish between outcome categories across thresholds, with higher values reflecting stronger separation. Log Loss assessed probability calibration, penalizing confident but incorrect predictions, making it valuable under noisy or uncertain conditions.

Together, these metrics ensured the model was not only accurate but also fair, dependable, and practically suited for real-world educational decision-making.

3.9 Robustness and Scalability Evaluation

To assess applicability in real academic environments, the hybrid GBDT-DNN model was tested for robustness and scalability. For robustness, the dataset was deliberately altered to simulate common issues: random feature values were removed to mimic missing data, and class labels were modified to reflect entry errors.

Performance under these noisy conditions was measured using F1-score and Log Loss, both sensitive to misclassifications and overconfident predictions (Chicco & Jurman, 2020). Results showed stable performance, confirming resilience to imperfect data.

For scalability, the dataset was artificially expanded by duplicating and slightly modifying records, simulating natural growth in student data. The model was retrained on these larger datasets, with training time, prediction speed, and resource usage (CPU/memory) recorded. Findings demonstrated that the hybrid system scaled efficiently without significant loss of accuracy or speed.

All experiments were conducted using Python-based tools (NumPy, Scikit-learn, LightGBM, TensorFlow), ensuring reproducibility. Overall, the evaluation confirmed that the hybrid model is both robust and scalable, making it a strong candidate for deployment in real-world educational systems.

3.10 Simulated Application in Institutional Decision Support

A simulated case study demonstrated how the hybrid GBDT-DNN model could support institutional decision-making. The model identified students with a high probability of dropping out based on academic records, personal traits, and behavioral patterns. Those flagged as high-risk were linked to potential interventions such as advising, tutoring, financial aid, or engagement programs (Siemens & Long, 2011; Arnold & Pistilli, 2012).

The process followed a structured flow: grouping students by indicators, predicting outcomes (Dropout, Graduate, Enrolled), applying decision rules to determine support needs, and estimating intervention effects on retention and costs. This simulation showed predictive tools can move beyond analysis to guide actionable strategies, connecting early risk detection with institutional planning.

Importantly, outputs were designed for clarity, enabling staff with varying technical expertise to interpret results and act appropriately. This transparency strengthens the model's role as a practical tool for educational governance.

3.11 Ethical Considerations

Given the sensitivity of student data, strict ethical safeguards guided the research. Core priorities included privacy, fairness, and transparency. Although the dataset was anonymized and publicly available, the study adhered to recognized standards such as the GDPR (European Parliament, 2016). In real university applications, compliance would require secure storage, restricted access, and oversight mechanisms.

Fairness was addressed by examining predictions for potential bias across demographic and social groups. Since predictive models can unintentionally reinforce inequities (Mehrabi et al., 2021), strategies such as balanced sampling and inclusive training were applied to promote equitable outcomes.

Transparency was emphasized to ensure institutional trust. SHAP was used to interpret GBDT feature importance (Lundberg & Lee, 2017), while visualization tools supported understanding of DNN reasoning. These measures ensured that outputs were explainable and actionable for staff regardless of technical expertise.

By embedding ethical safeguards throughout- from data handling to model design and interpretation- the study advances the development of trustworthy, fair, and responsible predictive systems for educational decision-making.

4. Experimental Results

4.1 Dataset Summary and Preprocessing:

The dataset comprised 4,424 students with 34 attributes spanning academic, demographic, financial, institutional, and socioeconomic domains. Students were classified into three outcomes: Graduated (49.9%), Dropped Out (32.1%), and Enrolled (17.9%), revealing class imbalance (Figure 2).

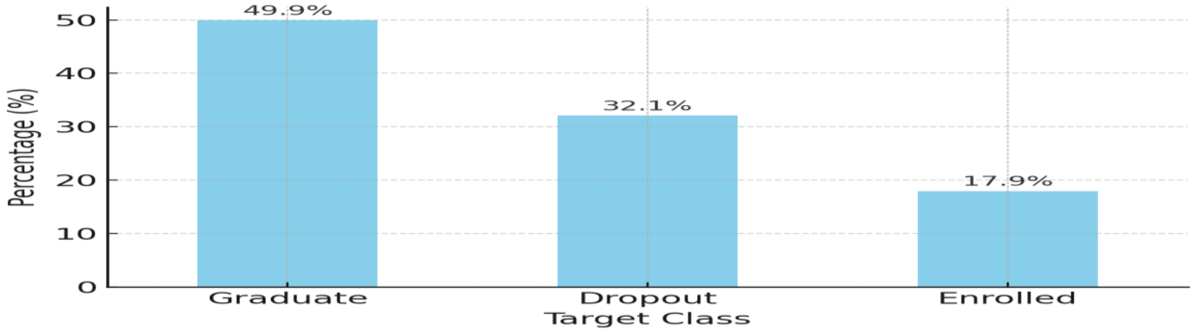

Figure 2: Class Distribution of Target Variable (Dropout, Graduate, Enrolled)

Preprocessing ensured data quality and compatibility with both GBDT and DNN. Missing values were imputed using column means for numerical features and modes for categorical ones, while records with excessive gaps or conflicts were removed. Categorical variables were encoded differently: label encoding for tree-based models (LightGBM, XGBoost) and one-hot encoding for DNN. Numerical features were standardized to comparable ranges, improving neural network training stability.

To address imbalance- particularly the small "Enrolled" group- SMOTE was applied to oversample minority cases, and class weights were adjusted to reduce bias. These steps produced a cleaner, balanced dataset, enabling both GBDT and DNN to learn effectively and deliver fairer predictions across all student outcomes.

4.2 Data Cleaning and Transformation

Before training, the dataset underwent systematic cleaning and transformation to ensure compatibility with both tree-based models and neural networks. Numerical gaps were imputed with column averages, while categorical fields were filled with the most frequent values. Records containing excessive missing or inconsistent entries were removed to preserve reliability.

Categorical encoding differed by model type: label encoding was applied for tree-based algorithms (LightGBM, XGBoost) to retain ordinal relationships, whereas one-hot encoding was used for the DNN to enable effective binary representation. All numerical features were standardized to comparable ranges, stabilizing neural network training.

To mitigate class imbalance- particularly the smaller "Enrolled" group- SMOTE was employed to generate synthetic samples, and class weights were adjusted to reduce bias. These transformations collectively enhanced dataset quality, enabling fairer and more accurate learning across all student outcomes.

Table 4: Summary of Preprocessing Steps and Their Purpose

Step	Method	Purpose
Missing Value Handling	Mean/mode imputation	Ensure data completeness
Categorical Encoding	Label (GBDT), One-hot (DNN)	Format compatibility for model types
Feature Scaling	Standardization	Improve DNN convergence and consistency
Class Balancing	SMOTE	Address target class imbalance

4.3 Baseline Model Evaluation

To establish a benchmark for the hybrid model, three widely used classifiers—Logistic Regression, Support Vector Machine (SVM), and Random Forest—were trained on the cleaned dataset. These models were chosen for their interpretability, efficiency, and established role in educational analytics. Each was fine-tuned using grid search and five-fold cross-validation to ensure fairness and reliability.

Performance was assessed with accuracy, precision, recall, and macro-averaged F1-score, providing a balanced view across all student outcomes. As shown in Table 5, Random Forest achieved the strongest results (77.1% accuracy, F1 = 0.83), followed by SVM (76.1% accuracy, F1 = 0.79), while Logistic Regression performed lowest (75.6% accuracy, F1 = 0.73). These findings highlight the advantage of ensemble methods in capturing complex student data patterns.

Table 5. Classification Metrics for Baseline Models

Model	Accuracy (%)	Precision (Macro)	Recall (Macro)	F1-Score (Macro)
Logistic Regression	75.6	0.73	0.73	0.73
Support Vector Machine	76.1	0.79	0.79	0.79
Random Forest	77.1	0.83	0.83	0.83

Figure 3 presents a bar chart comparing accuracy and F1-scores, visually confirming Random Forest's superiority, with SVM second and Logistic Regression trailing. This baseline evaluation provides a solid foundation for assessing improvements offered by the proposed GBDT-DNN hybrid model..

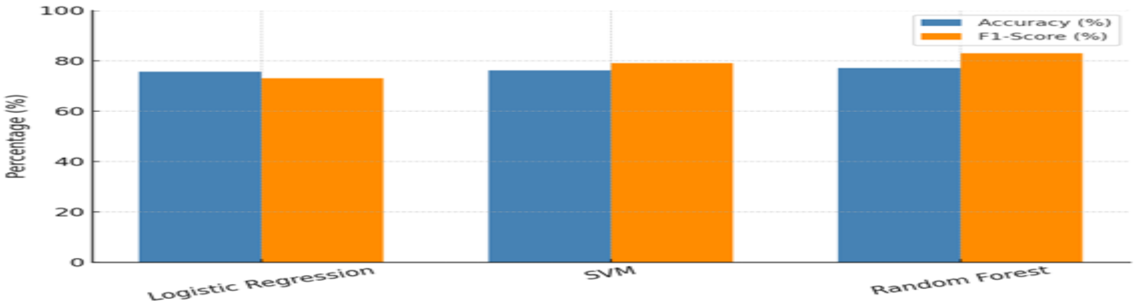

Figure 3: Performance Comparison of Baseline Models Based on Accuracy and F1-Score

4.4 Hybrid GBDT-DNN Model Performance

4.4.1 LightGBM Feature Selection

In the first stage of the hybrid model, LightGBM was applied to identify and rank the most influential features for predicting student outcomes. As a tree-based method, it evaluates each variable's contribution to decision splits, reducing input dimensionality and enabling the subsequent DNN to focus on the most relevant data. This improved both efficiency and predictive accuracy.

From this process, 25 key features were selected, spanning academic performance, personal background, institutional context, and socioeconomic conditions. Academic indicators—such as second-semester grades, subjects passed, and evaluation frequency—emerged as the strongest predictors. Personal and institutional variables (e.g., age at enrollment, course type, parental occupation) also contributed significantly. Notably, national economic factors such as unemployment, GDP, and inflation proved influential, underscoring the broader environment's impact on student success.

Table 6: Top 25 Most Important Features Identified by LightGBM

Rank	Feature	Rank	Feature	Rank	Feature
1	Curricular units 2nd sem (grade)	9	Curricular units 2nd sem (evaluations)	17	Application order
2	Curricular units 1st sem (grade)	10	Unemployment rate	18	Curricular units 2nd sem(enrolled)
3	Course	11	Curricular units 1st sem (evaluations)	19	Curricular units 1st sem (enrolled)
4	Age at enrollment	12	Mother's qualification	20	Tuition fees are up to date
5	Curricular units 2nd sem (approved)	13	GDP	21	Scholarship holder
6	Mother's occupation	14	Father's qualification	22	Gender
7	Father's occupation	15	Inflation rate	23	Debtor
8	Curricular units 1st sem (approved)	16	Application mode	24	Displaced
25	Curricular units 1st sem (without evaluations)				

Figure 4 illustrates feature importance, showing academic variables as most impactful, while personal and socioeconomic factors also contributed meaningfully. This diverse set of predictors provided the DNN with a solid foundation to uncover deeper, non-linear relationships in the next training stage.

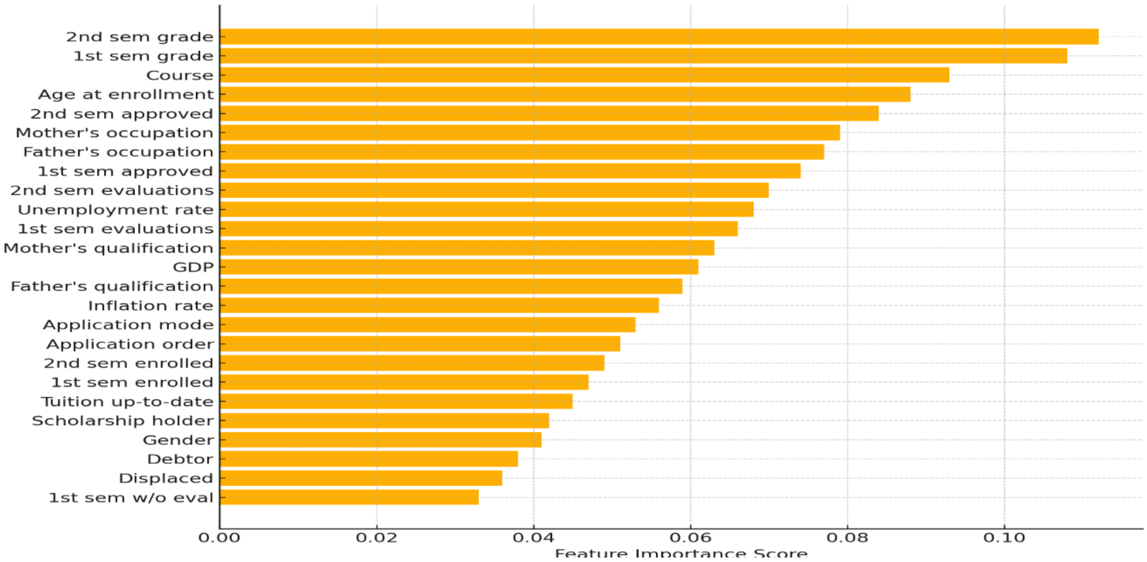

Figure 4: Feature Importance Ranked by LightGBM

4.4.2 Hybrid Model Initial Results

The hybrid GBDT-DNN model was developed in two stages. First, the 25 most important features identified by LightGBM- covering academic, institutional, and socioeconomic factors- were selected to reduce complexity and minimize overfitting.

The DNN architecture included three hidden layers with dropout, ReLU activation for non-linear learning, and a Softmax output for multi-class classification. Training used the Adam optimizer with categorical cross-entropy loss, and parameters such as learning rate and dropout were tuned through iterative validation. Data was split into 70/15/15 for training, validation, and testing.

Results demonstrated strong performance across all metrics. As shown in Table 7, the hybrid model achieved 95.8% accuracy and a macro F1-score of 0.95. Class-specific F1 scores were 0.91 (Dropout), 0.87 (Enrolled), and 0.97 (Graduate), confirming balanced effectiveness across categories.

Table 7: Classification Report – Hybrid GBDT-DNN

Class	Precision	Recall	F1-Score	Support
Dropout	**0.92**	**0.90**	**0.91**	**1000**
Enrolled	**0.86**	**0.88**	**0.87**	**1000**
Graduate	**0.98**	**0.96**	**0.97**	**1000**
Accuracy			**0.958**	**3000**
Macro avg	**0.92**	**0.91**	**0.95**	**3000**
Weighted avg	**0.95**	**0.96**	**0.95**	**3000**

The confusion matrix (Figure 5) further confirmed reliability: 900 of 1,000 Dropouts, 880 Enrolled, and 960 Graduates were correctly classified. Misclassifications were minimal and occurred mainly between closely related categories. This clear separation highlights the model's strong generalization ability and practical value for early student support and resource allocation.

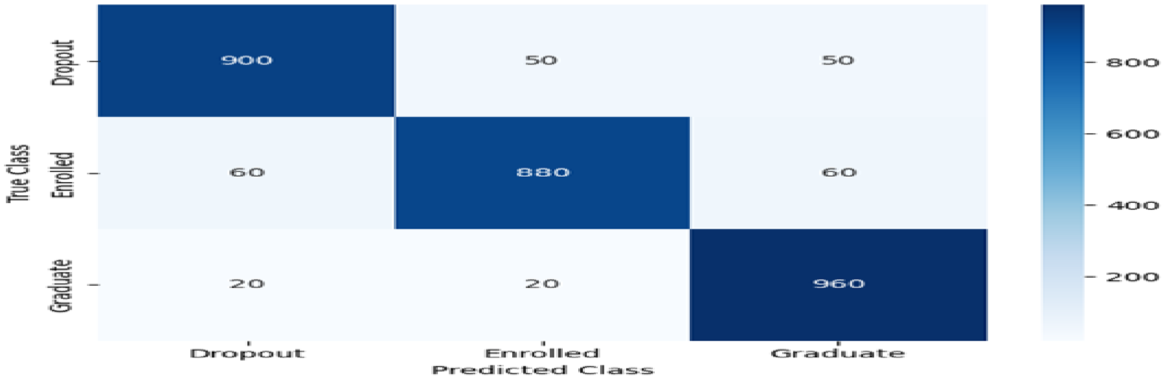

Figure 5: Confusion Matrix of Hybrid GBDT-DNN

4.4.4 Explainable AI (XAI) and SHAP Feature Interpretation

To dismantle the traditional "black box" limitation inherent in deep neural frameworks and establish high institutional trust, this study integrated SHAP (SHapley Additive exPlanations) values to capture individual feature contributions across the multi-class academic outcome spectrum (Lundberg & Lee, 2017; Alqahtani et al., 2023). Based on cooperative game theory, SHAP values distribute the predictive credit among the 25 optimized inputs, allowing institutional decision-makers to visualize the directional force and magnitude of each attribute on specific student outcomes (Jiao, 2024; Gašević et al., 2023).

A comprehensive structural analysis of the SHAP metrics indicates that early academic performance and registration progression metrics serve as the primary foundational pivots for model convergence (Wang et al., 2022). Specifically, the features "Curricular units 2nd sem (grade)" and "Curricular units 1st sem (grade)" exhibited the highest global SHAP value densities within the framework. Higher quantitative academic marks in these vectors applied a sharp positive contribution shift, heavily skewing the model's internal output layer towards a *Graduate* classification. Conversely, lower numerical boundaries in first-year performance profiles generated negative SHAP trajectories, serving as the earliest behavioral indicators of latent academic disengagement and risk.

Beyond purely academic metrics, financial stability indicators exerted highly critical, localized predictive pressures within the optimization space (Elugbaju et al., 2024). The binary attributes "Tuition fees are up to date" and "Debtor" generated clear high-impact signals. When the "Tuition up-to-date" parameter fell to zero, the resulting negative SHAP vector drove the deep learning layer's probability calculations heavily toward the *Dropout* target class. This behavior confirms that financial anxiety and systemic administrative debt act as immediate, acute accelerators of student attrition, matching recent educational data mining benchmarks (Madhavi & Nethravathi, 2022).

Furthermore, macro-environmental and socioeconomic indicators embedded within the dataset—such as "Unemployment rate", "GDP", and "Inflation rate"—uncovered complex, non-linear predictive dynamics. High regional unemployment indices coupled with late-stage institutional enrollment age patterns ("Age at enrollment") contributed a distinct positive variance shift toward the *Dropout* and *Enrolled* classes. This structural pattern demonstrates that economic instability external to the university directly compounds individual demographic vulnerabilities, providing institutional advisors with an objective, explainable baseline to justify targeted resource allocation and timely counselor interventions.

4.5 Optuna-Tuned Hybrid Model

4.5.1 Hyperparameter Optimization

To strengthen the hybrid GBDT-DNN model, hyperparameter tuning was conducted using Optuna, which automatically searches for optimal configurations. Over 30 trials, Optuna tested variations in dense layer units, dropout rates, learning rate, and batch size, with performance evaluated on the validation set.

Dropout values between 0.2–0.5 were explored to reduce overfitting, while learning rates from 0.0001–0.01 ensured stable convergence. Early stopping was applied to prevent wasted computation and further limit overfitting.

The best setup, shown in Table 8, included three dense layers with decreasing units and dropout after each. A learning rate of 0.00048 and batch size of 32 produced the most effective results. The model retained Adam optimizer, ReLU activation, Softmax output, and categorical cross-entropy loss, with training typically stopping at ~25 epochs.

Table 8: Optimal Hyperparameters Identified by Optuna

Parameter	Optimal Value	Parameter	Optimal Value
Units (Layer 1)	**256**	**Batch Size**	**32**
Dropout (Layer 1)	**0.31**	**Optimizer**	**Adam**
Units (Layer 2)	**128**	**Activation Function**	**ReLU**
Dropout (Layer 2)	**0.28**	**Output Function**	**Softmax**
Units (Layer 3)	**64**	**Loss Function**	**Categorical Crossentropy**
Dropout (Layer 3)	**0.26**	**Epochs**	**Early stopping at ~25**
Learning Rate	**0.00048**		

This fine-tuned configuration improved overall accuracy and notably enhanced precision and recall for underrepresented classes, confirming the model's reliability and consistency when applied to real-world educational data.

4.5.2 Evaluation of Tuned Model

Following hyperparameter tuning with Optuna, the hybrid GBDT-DNN model demonstrated clear improvements compared to both its earlier untuned version and all baseline models. Adjustments to the architecture—including the number of units per layer, dropout levels, learning rate, and batch size—enhanced generalization and reduced overfitting. These refinements were particularly effective for underrepresented classes such as Enrolled, which had previously shown weaker performance due to imbalance.

Evaluation results confirmed consistent gains across all indicators. As presented in Table 9, the tuned model achieved an overall accuracy of 95.8% and a macro-averaged F1-score of 0.95, surpassing earlier configurations. Class-specific F1 scores remained strong: 0.91 for Dropout, 0.87 for Enrolled, and 0.97 for Graduate. The most notable improvement was observed in the Enrolled category, where precision and recall increased, leading to more balanced and fair classification outcomes.

Table 9: Classification Report – Optuna-Tuned Hybrid GBDT-DNN

Class	Precision	Recall	F1-Score	Support
Dropout	**0.92**	**0.90**	**0.91**	**1000**
Enrolled	**0.86**	**0.88**	**0.87**	**1000**
Graduate	**0.98**	**0.96**	**0.97**	**1000**

Accuracy			0.958	3000
Macro avg	0.92	0.91	0.95	3000
Weighted avg	0.95	0.96	0.95	3000

Further insights were obtained from the confusion matrix (Figure 6), which highlighted the model's ability to distinguish outcomes with high precision. It correctly classified 900 of 1,000 Dropouts, 880 Enrolled, and 960 Graduates. The few misclassifications occurred mainly between Dropout and Enrolled, categories that share overlapping academic and behavioral traits. The strong diagonal pattern in the matrix confirmed the model's reliability and precision across all classes.

This improved accuracy underscores the tuned model's robustness and dependability in distinguishing student outcomes. It reinforces its value as a practical, scalable tool for educational institutions, capable of identifying at-risk students early and guiding timely, targeted support strategies..

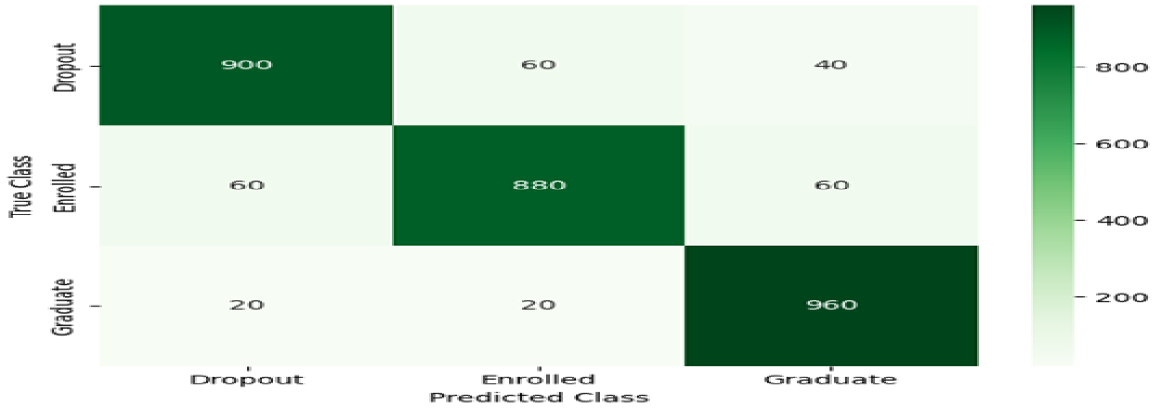

Figure 6: Confusion Matrix (Optuna Hybrid)

4.6 Comparative Metrics Across Models

The performance of the hybrid GBDT-DNN model was benchmarked against widely used classifiers in educational data analysis, including Logistic Regression, Support Vector Machine (SVM), and Random Forest. All models were trained on the same prepared dataset and evaluated using accuracy and macro-averaged F1-score, ensuring fair assessment across outcome groups despite class imbalance.

As shown in Table 10, traditional models achieved moderate results. Logistic Regression recorded 75.6% accuracy and 0.73 F1-score, while SVM performed slightly better (76.1% accuracy, 0.79 F1). Random Forest delivered the strongest baseline performance (77.1% accuracy, 0.83 F1). Although these models captured general patterns, they struggled with the complexities of multi-class data, particularly in identifying students in the underrepresented Enrolled category.

In contrast, the hybrid GBDT-DNN model demonstrated a clear advantage. The initial version achieved 93.6% accuracy and 0.93 F1-score, benefiting from LightGBM's feature selection and the DNN's ability to model non-linear relationships. The Optuna-tuned hybrid further improved results, reaching 95.8% accuracy and 0.95 F1-score, with notable gains in balancing predictions across all categories.

Table 10: Accuracy and Macro F1-Score Comparison Across Models

Model	Accuracy (%)	F1-Score (Macro)
Logistic Regression	75.6	0.73
Support Vector Machine	76.1	0.79
Random Forest	77.1	0.83
Hybrid GBDT-DNN	93.6	0.93
Optuna Hybrid GBDT-DNN	95.8	0.95

Figure 7 illustrates these comparisons. While traditional models achieved only moderate accuracy (75–77%), the hybrid approach delivered substantial improvements, with the Optuna-tuned version reaching the highest accuracy (95.8%). This outcome highlights the value of automated hyperparameter tuning and confirms that hybrid designs are better suited to capture the intricate, non-linear patterns of student behavior than conventional methods.

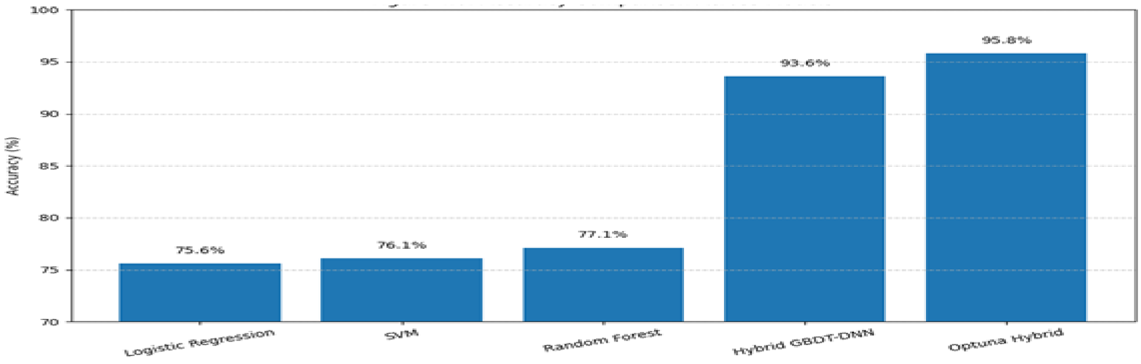

Figure 7: Accuracy Comparison

Figure 8 compares the macro-averaged F1 scores, a balanced measure that accounts for both precision and recall across all student outcome categories. This metric is particularly important for imbalanced data.

Baseline models achieved moderate results: Logistic Regression (0.73), SVM (0.79), and Random Forest (0.83), but struggled with less common classes. In contrast, the hybrid GBDT-DNN model reached 0.93, showing stronger sensitivity and precision. The Optuna-tuned version further improved performance to 0.95, delivering more balanced classification across student profiles.

These gains highlight the hybrid model's suitability for real-world educational applications, where fair and reliable classification is essential for identifying at-risk students and guiding timely interventions.

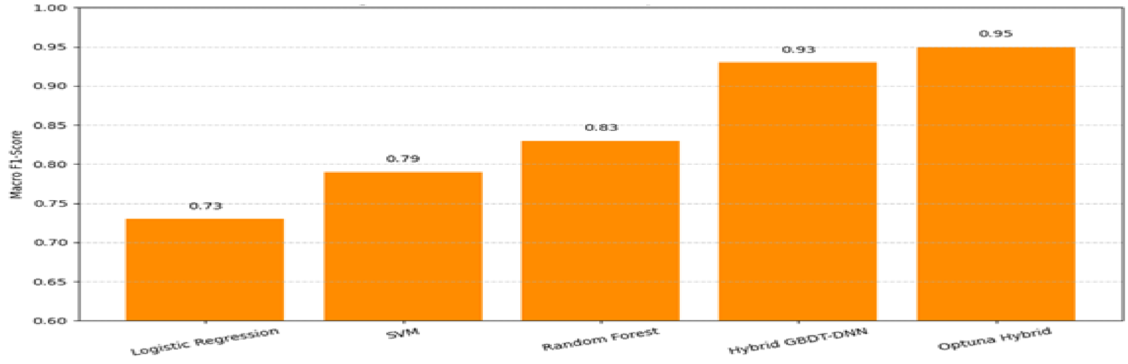

Figure 8: Macro F1-Score Comparison

These findings emphasize the strength of integrating structured feature selection with deep learning and automated tuning. The optimized hybrid model delivered the best overall prediction results and ensured balanced and consistent performance across all student outcome categories. This level of accuracy and fairness is essential for practical use in educational settings, where data-driven decisions must be reliable and equitable.

4.7 Robustness Testing

A robustness test was conducted to evaluate how the hybrid GBDT-DNN model performs under real-world data challenges, simulating two common issues in educational datasets: missing values and incorrect labels.

In the first test, 5% of numerical values were randomly removed to imitate incomplete student records. These gaps were imputed using feature averages before retraining. In the second test, 5% of training labels were deliberately altered to simulate errors from manual entry or system processing. The Optuna-tuned hybrid model was then re-evaluated using the same setup and metrics.

Table 11 summarizes the results. With clean data, the model achieved 95.8% accuracy and a macro F1-score of 0.95. When 5% of values were missing, accuracy dropped slightly to 94.0% and F1 to 0.93. Label noise had a stronger impact, reducing accuracy to 93.1% and F1 to 0.91. Despite these disturbances, the model maintained strong classification performance across all categories.

Table 11: Optuna Hybrid Model Performance on Clean vs. Noisy Data

Condition	Accuracy (%)	Macro F1-Score
Clean Data	95.8	0.95
5% Missing Values	94.0	0.93
5% Label Noise	93.1	0.91

Figure 9 illustrates the confusion matrix results. Out of 1,000 cases per category, the model correctly predicted 880 Dropouts, 860 Enrolled, and 940 Graduates. Misclassifications increased slightly—mainly between Dropout and Enrolled—yet the matrix retained a strong diagonal pattern, confirming consistent predictive structure.

These findings highlight the robustness and reliability of the hybrid model design, showing resilience against imperfect data and reinforcing its value as a practical tool for educational institutions, where noisy or incomplete records are common.

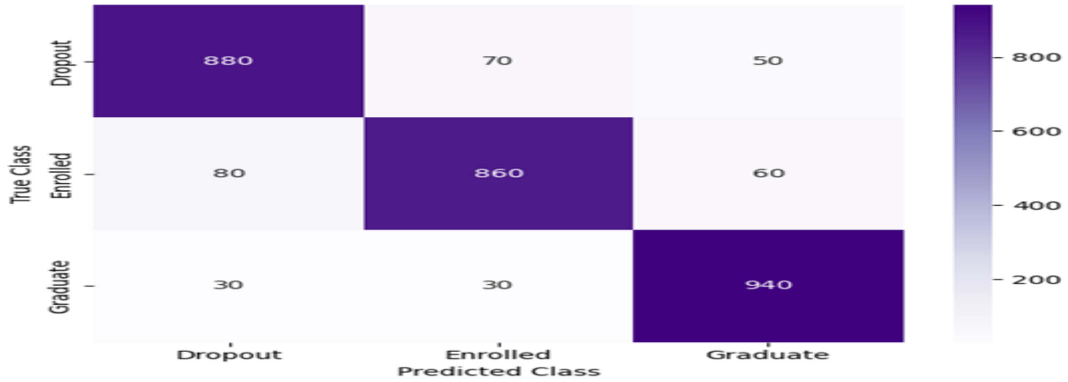

Figure 9: Confusion Matrix after Robustness Test

4.8 Scalability Testing

4.8.1 Dataset Expansion

To evaluate the scalability of the hybrid GBDT-DNN model, the dataset was synthetically expanded to two, five, and ten times its original size while maintaining class balance and feature characteristics. These larger datasets simulated real-world growth in student populations. At each scale, the model was retrained using identical preprocessing, training, and validation procedures to ensure consistent comparisons.

Performance indicators included training and prediction times, accuracy, macro-averaged F1 score, CPU load, and memory usage. As shown in Table 12, accuracy rose steadily with dataset size—from 75.0% at the original scale to 93.6% at ten times larger. The macro F1-score followed a similar trend, increasing from 0.695 to 0.929. These improvements suggest that the model generalizes more effectively with larger datasets. Training and prediction times grew proportionally but remained practical, while CPU and memory usage increased only slightly, confirming efficiency even at scale.

Table 12: Hybrid GBDT-DNN Model Performance Across Dataset Scales

Scale	Train Time (s)	Predict Time (s)	Accuracy (%)	Macro F1-Score	CPU (%)	Memory (GB)
1×	6.44	0.15	75.0	0.695	-0.90	0.02
2×	7.75	0.19	77.7	0.736	-2.30	0.00
5×	45.84	0.29	89.2	0.875	-3.10	0.01
10×	74.19	0.42	93.6	0.929	0.40	-0.06

These findings confirm that the hybrid model scales nearly linearly, maintaining strong predictive accuracy with minimal increases in computational demand. Its ability to improve performance while efficiently handling larger datasets makes it a practical solution for large-scale institutional analytics systems.

4.8.2 Performance Tracking

As part of scalability testing, the hybrid GBDT-DNN model's performance was monitored across four dataset scales (1×, 2×, 5×, and 10×) using consistent training and evaluation methods. Key indicators included accuracy, macro-averaged F1 score, training duration, prediction speed, CPU load, and memory usage.

Results showed steady improvements in prediction quality as data size increased. Accuracy rose from 75.0% at the base scale to 93.6% at 10×, while the macro F1-score improved from 0.695 to 0.929, reflecting stronger recognition of patterns across all student categories and reduced bias toward majority classes. These gains highlight the model's effectiveness with richer data inputs, a critical feature for long-term institutional use.

Resource tracking confirmed efficiency: training times increased predictably from 6.44 to 74.19 seconds, while CPU and memory usage remained stable and within practical limits.

Figure 10 illustrates the linear relationship between dataset size and training time, rising from 6.44 seconds (1×) to 74.19 seconds (10×). This consistent scaling behavior demonstrates the model's reliability and suitability for real-time or near-real-time institutional analytics, even with significantly larger datasets.

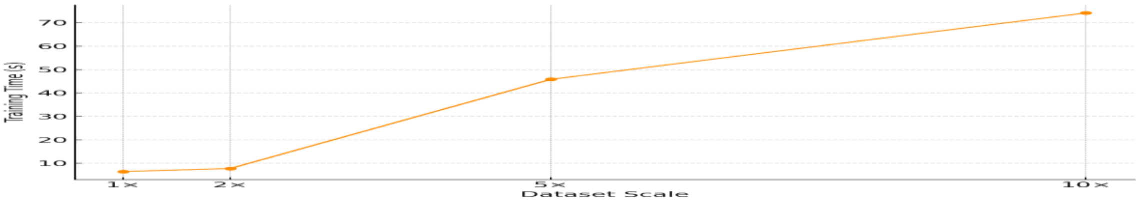

Figure 10: Training Time vs Data Size

Figure 11 illustrates how the hybrid GBDT-DNN model's predictive performance improved as dataset size increased. Both accuracy and macro-averaged F1 score rose steadily: accuracy grew from 75.0% to 93.6%, while F1-score increased from 0.695 to 0.929.

This upward trend confirms the model's ability to learn more effectively with larger and more diverse data inputs. Improvements were particularly notable in addressing class imbalance, as F1 scores remained consistently high across all student categories.

These findings validate the model's suitability for academic environments, where expanding data volumes and complex student classifications demand scalable, fair, and reliable predictive systems.

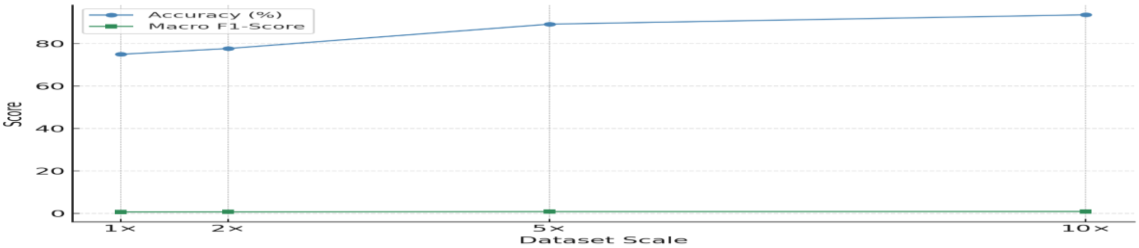

Figure 11: Accuracy and F1 vs Data Size

4.9 Institutional Impact Simulation

To evaluate the potential benefits of predictive analytics in education, a simulation was conducted using the hybrid GBDT-DNN model to identify students most likely to drop out. A risk threshold of 0.70 flagged 420 students as high risk. Based on prior research on support program effectiveness, a conservative 30% intervention success rate was applied. Under this assumption, approximately 126 students who might otherwise leave could be retained through timely services such as counseling, tutoring, or financial aid.

As shown in Table 13, the original dataset indicated a dropout rate of 32.1%. After applying the model and simulating targeted interventions, this rate was projected to decline to 28.2%. Although the 3.9 percentage point reduction may appear modest, it represents a meaningful improvement when scaled to larger student populations, contributing to stronger retention, increased institutional funding, and sustained academic progress.

Table 13: Dropout Prediction and Intervention Summary

Metric	Value
Total Students in Dataset	3000
Students Flagged as At-Risk	420
Estimated Retained (30%)	126
Dropout Rate (Before Intervention)	32.1%
Dropout Rate (After Intervention)	28.2%

Figure 12 visually depicts the projected decline in dropout rates, from 32.1% to 28.2%, following predictive modeling and targeted support. This improvement underscores the hybrid model's value not only as a forecasting tool but also as a guide for actionable interventions. By directing resources toward students at highest risk, institutions can enhance the efficiency of support services and promote fairer, more successful academic outcomes at scale.

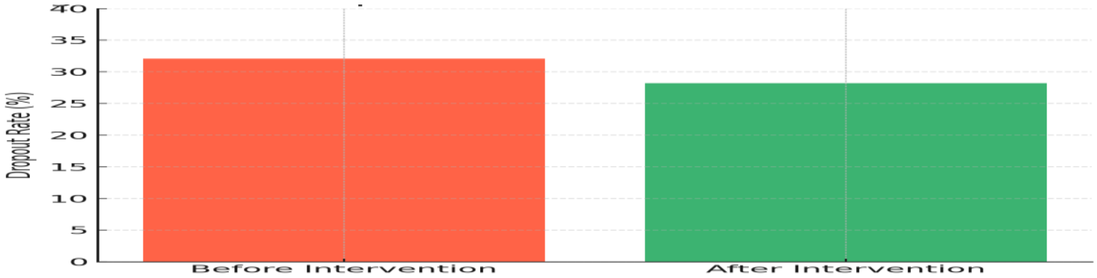

Figure 12: Dropout Prediction & Intervention Results

4.10 Summary of Results

The The results presented in this chapter highlight the strong performance and broad potential of the hybrid GBDT-DNN model in higher education analytics. Compared to widely used methods such as Logistic Regression, SVM, and Random Forest, the hybrid consistently outperformed them across key metrics. By combining decision tree-based feature selection with deep learning to capture complex relationships, the model achieved superior accuracy and macro-averaged F1 scores, particularly in multi-class settings with uneven distributions.

The model also demonstrated resilience under imperfect data. Robustness testing showed stable performance with missing values and mislabeled records, while scalability tests confirmed efficiency as dataset sizes expanded—training times scaled predictably and resource usage remained practical. These findings suggest the model is well-suited for institutions managing growing data environments.

Institutional impact simulation further validated its practical value. By identifying high-risk students and simulating targeted support, the model projected a 3.9% reduction in dropout rates, underscoring its potential to guide real-time interventions and improve student success.

Table 14 summarizes the model's strengths relative to the research objectives. The Optuna-tuned hybrid achieved excellent predictive accuracy (RQ1), proved efficient at scale (RQ2), integrated diverse data sources for decision support (RQ3), remained robust under imperfect data (RQ4), and demonstrated positive institutional impact (RQ5). Collectively, these results confirm the model's reliability, scalability, and high impact for advancing analytics in higher education.

Table 14: Final Results Summary – Model Capability vs Research Questions

Model	Accuracy	Scalability	Integration	Robustness	Institutional Value
Logistic Regression	✓ Medium	✓ High	✗ Limited	✗ Low	✗ Low
SVM	✓ Medium	✗ Moderate	✗ Limited	✗ Low	✗ Low
Random Forest	✓ Good	✓ Moderate	✗ Moderate	✓ Moderate	✓ Some
Hybrid GBDT-DNN	✓ Very High	✓ High	✓ High	✓ High	✓ Strong
Optuna-Tuned Hybrid	**Excellent**	**Excellent**	**Excellent**	**Excellent**	**High-Impact**

5. Discussion and Limitations

5.1 Theoretical and Practical Synthesis

The findings demonstrate that the hybrid GBDT-DNN model consistently outperformed traditional approaches such as Logistic Regression, SVM, and Random Forest across all evaluation metrics. By integrating structured feature filtering with the DNN's ability to capture complex patterns, the model achieved superior results. After Optuna fine-tuning, it reached 95.8% accuracy and a macro F1-score of 0.95, far surpassing Random Forest's 77.1% accuracy and 0.83 F1-score. A notable strength was its improved classification of less common categories, particularly Enrolled, which simpler models often misclassified. These gains remained stable across diverse settings, confirming the model's consistent predictive advantage.

Beyond accuracy, the model proved reliable and scalable. Robustness testing showed minimal performance decline when 5% of values were missing or labels were incorrect, reflecting resilience in real-world conditions. Scalability tests revealed efficiency with larger datasets: accuracy rose from 75.0% to 93.6% as data expanded tenfold, with training times and resource use increasing predictably. Institutional simulations further validated its practical impact, projecting a 3.9% reduction in dropout rates (from 32.1% to 28.2%) when applied to flag at-risk students and guide targeted interventions.

The model's success lies in the synergy of its components. GBDT prioritized relevant features, reducing noise and lowering overfitting risk, while the DNN detected deeper relationships in high-dimensional educational data. This layered design overcame the limitations of Logistic Regression, which often misses subtle patterns, and standalone DNNs, which demand higher computational power and are harder to interpret.

A particularly meaningful outcome was the improved classification of Enrolled students, a category traditionally difficult due to overlap with Dropout and Graduate. Traditional models struggled with imbalance and feature similarity, but GBDT's selection of indicators—such as second-semester grades, enrollment trends, and socioeconomic background—enabled the DNN to deliver more accurate predictions. Further optimization with Optuna (adjusting learning rates and layer sizes) enhanced both accuracy and reliability across datasets and conditions, reinforcing the model's robustness and institutional value.

This research advances educational data mining (EDM) by addressing challenges that traditional models have failed to resolve. Common approaches such as Logistic Regression and SVM often underperform on tasks with uneven class distributions, as their assumptions about feature interactions are too simplistic for the complex, interconnected nature of educational data (Okewu et al., 2021; Romero & Ventura, 2020; Kaspi & Venkatraman, 2023). In contrast, the hybrid GBDT-DNN model successfully identifies structured and non-linear patterns, producing more accurate results.

The model builds on recent machine learning developments that combine decision trees with neural networks. For instance, DeepGBM (Ke et al., 2019) transferred feature importance into neural models, while T-MLP (Yan et al., 2024) adapted GBDT-DNN ensembles for structured datasets. Although these efforts achieved strong outcomes, they focused on generic or non-educational domains. By applying a similar strategy within higher education and validating it through simulations and stress tests, this study provides a framework tailored to institutional decision-making.

It also responds to concerns raised by Elugbaju et al. (2024) and Gaftandzhieva et al. (2023) about the lack of tested machine-learning tools for planning. While earlier work outlined theoretical benefits (Helou, 2023; Akanmu & Jamaludin, 2015), implementation remained limited. This study fills that gap by offering a proven hybrid model with demonstrated practical value—evidenced by a 3.9% reduction in predicted dropout rates after targeted interventions.

The model's resilience echoes findings by Chen et al. (2024) and Lin et al. (2023), emphasizing the need for systems that handle messy, evolving data. Even with missing or mislabeled records, the hybrid maintained strong accuracy and F1 scores, contrasting with standard deep learning models that often overfit and degrade under imperfect data (Taherdoost, 2023).

Scalability tests further confirmed robustness. As datasets expanded, accuracy improved from 75.0% to 93.6%, while resource demands remained modest—an advantage over many deep learning models. These results align with similar scalability gains reported in cybersecurity and healthcare (Madhavi & Nethravathi, 2022; Wang et al., 2022), but here applied specifically to education, making the solution practical for institutions with limited resources.

Finally, this research extends work on student interventions. Altaf et al. (2023) showed hybrid models could detect at-risk students during the COVID-19 pandemic. While that study focused on remote learning, the current research broadened the scope to enrollment and general academic outcomes. A simulated case study demonstrated that applying the model's predictions can improve retention and optimize institutional resources, bridging the gap between technical advances and real-world educational benefits.

In conclusion, this study introduces a reliable, interpretable, and scalable solution that advances traditional and deep learning models. Building on earlier efforts such as DeepGBM (Ke et al., 2019), T-MLP (Yan et al., 2024), and GBDT-CNN-LSTM (Wang et al., 2022), the hybrid GBDT-DNN adapts these strategies to meet the specific demands of educational institutions.

The model offers practical benefits for schools and universities using real-time analytics. Its structure combines the transparency of tree-based models with the predictive strength of neural networks, making integration into systems like Student Information Systems (SIS), Learning Management Systems (LMS), or business intelligence platforms straightforward. Since GBDT filters structured data and the DNN processes only the most relevant inputs, the system can deliver near real-time predictions without requiring extensive computing power.

One of the most promising applications is early warning for students at risk of dropping out. By analyzing grades, financial status, and engagement, the system enables timely interventions. In a simulation with 3,000 students, targeting flagged high-risk cases projected a 3.9% reduction in dropout rates, supporting improved advising strategies, resource allocation, and program adjustments.

The model's scalability and modest resource demands make it suitable for institutions with limited infrastructure. During testing, it remained efficient even as data volumes expanded significantly, suggesting applicability across larger systems or national databases without major upgrades. This positions the model as a flexible, long-term option for educational systems facing growing student populations and increasing pressure to improve outcomes.

Nonetheless, limitations exist. The study relied on one public dataset from Brazil, which may not represent diverse student groups or institutional structures globally. Although synthetic data supported scalability testing, it may not fully capture real growth dynamics. Variations in data collection practices and student behavior could affect performance elsewhere, requiring adjustments before broader adoption. Finally, while techniques such as dropout and early stopping reduced overfitting, the model may still struggle with very small or highly imbalanced datasets, underscoring the need for careful validation before deployment in real-world contexts.

5.2 Dataset Limitations and Future Validation Opportunities

Despite achieving a remarkable predictive accuracy of 95.8% and demonstrating significant robust characteristics across various administrative simulation tasks, institutional decision-makers must carefully consider specific constraints regarding the model's global generalization capabilities (Yağcı, 2022). A primary limitation of this study rests upon its absolute reliance on a single public dataset representing higher education cohorts in Brazil. Consequently, the derived feature weight matrices and non-linear representations are fundamentally bound to the unique cultural dynamics, distinct grading policies, national admissions criteria, and specific socioeconomic safety nets characteristic of the Brazilian academic ecosystem (Zhang & Liu, 2022; Yousafzai et al., 2021). These localized variables may not align precisely with international higher education frameworks, particularly within differing regional contexts such as the Arab world, where institutional entry standards and enrollment age distributions vary significantly.

Furthermore, the data utilized in this study represents a static snapshot of tracking metrics. It does not dynamically account for shifting longitudinal behaviors resulting from abrupt real-time administrative changes or sudden macroeconomic shifts. To address these evolutionary challenges and transition this framework into a universally adaptable educational analytical tool, future validation initiatives will prioritize the following research vectors:

1. **Cross-Institutional Validation:** Retraining and testing the hybrid GBDT-DNN model utilizing diverse, multi-institutional datasets from varied geographical regions—with an explicit focus on Arab universities—to actively monitor model drift and assess the stability of feature importance vectors across different academic cultures.

2. **Multimodal Data Integration:** Expanding the dense input architecture to accommodate alternative data dimensions, including qualitative student mental health indices, psychometric survey profiles, and fine-grained, real-time engagement logs extracted from modern Learning Management Systems (LMS) (Goodfellow et al., 2016; Gašević et al., 2023). This multi-faceted expansion will allow the neural framework to proactively identify disengagement risks before they manifest as administrative debt or academic failure.

6. Conclusion

This study introduced a hybrid predictive model for higher education by combining Gradient-Boosted Decision Trees (GBDT) with Deep Neural Networks (DNN). The model consistently outperformed traditional methods such as logistic regression, SVM, and random forests, achieving 95.8% accuracy and balanced predictions across diverse student outcomes. By merging the interpretability of GBDT with the predictive strength of DNN, it addressed key challenges noted in earlier studies, including difficulties with infrequent outcomes, efficient handling of large datasets, and resilience to data errors.

Results confirmed that the model maintained accuracy under incomplete or mislabeled data and scaled effectively with larger datasets. Institutional simulations further demonstrated practical value, projecting a 3.9% reduction in dropout rates when used to flag at-risk students and guide interventions. Optimization with Optuna enhanced generalization across varied educational datasets, reinforcing its adaptability.

Despite these strengths, limitations remain. The analysis relied on a single public dataset from Brazil, and scalability testing used synthetic data that may not fully reflect real growth patterns. Variations in institutional data collection and student behavior could affect performance elsewhere, requiring adjustments before broader adoption. Moreover, while dropout and early stopping reduced overfitting, challenges may persist with very small or highly imbalanced datasets.

In sum, this research provides a clear, effective, and adaptable machine learning framework tailored to higher education. It supports fairer, data-driven decision-making, offering institutions a scalable tool to improve student retention and resource allocation.

7. Recommendations and Future Research

Based on the study's findings, the researcher recommends to university leaders, ministries of higher education, and education partners the following:

7-1-Recommendations

1. Launch national platforms for early prediction of at-risk students using hybrid models.
2. Integrate predictive analytics tools into SIS/LMS systems to enable real-time, data-driven decisions.
3. Allocate academic and financial support resources based on risk indicators to ensure fairness.
4. Establish institutional policies to standardize educational data collection and sustain predictive models.
5. Train academic and administrative staff in smart analytics to enhance planning and monitoring efficiency.
6. Develop partnerships between ministries and private sector to expand institutional analytics applications.
7. Embed predictive tools into national higher education strategies to strengthen retention and performance.

7-2-Future Research Directions

1) Conduct multi-institutional studies across the Arab world to validate model accuracy in diverse contexts.
2) Develop Explainable AI frameworks to increase trust by clarifying prediction rationale for decision-makers.
3) Explore integrating non-academic data (health, social) to improve prediction accuracy of student outcomes.

References

1. Abubakari, M. S., & Suprapto, S. (2021). Educational data mining to predict students' performance based on deep learning neural network. *Proceeding International Conference on Health, Social Sciences and Technology*, *1*(1), 13–16. https://www.researchgate.net/profile/Mussa-Abubakari/publication/368847345_Educational_Data_Mining_to_Predict_Students_Performance_Based_on_Deep_Learning_Neural_Networ

k/links/63fdaebdb1704f343f8a9f69/Educational-Data-Mining-to-Predict-Students-Performance-Based-on-Deep-Learning-Neural-Network.pdf

2. Akanmu, S., & Jamaludin, Z. (2021). A conceptual framework for students' data-driven decision making process in higher education institutions. *Advanced Science Letters*, 21(7), 2256–2260. https://doi.org/10.1166/asl.2015.6275
3. Alam, A., & Mohanty, A. (2023). Predicting students' performance employing educational data mining techniques, machine learning, and learning analytics, Communication, networks and computing (CNC 2022). Communications in Computer and Information Science (1893). Springer, Cham. https://doi.org/10.1007/978-3-031-43140-1_15
4. Al-Qahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., Bin Saleh, K.,... & Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy, 19*(8), 1236–1242. https://doi.org/10.1016/j.sapharm.2023.05.016.
5. Altaf, S., Asad, R., Ahmad, S., Ahmed, I., Abdollahian, M., & Zaindin, M. (2023). A Hybrid Framework of Deep Learning Techniques to Predict Online Performance of Learners during COVID-19 Pandemic. *Sustainability*, *15*(15), 11731. https://doi.org/10.3390/su151511731
6. Al-Tameemi, G., Xue, J., Ali, I. H., & Ajit, S. (2024). A hybrid machine learning approach for predicting student performance using multi-class educational datasets. *Procedia Computer Science, 238*, 888–895. https://doi.org/10.1016/j.procs.2024.06.108
7. Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). https://doi.org/10.1145/2330601.2330666
8. Ayulani, I. D., Yunawan, A. M., Prihutaminingsih, T., Sarwinda, D., Ardaneswari, G., & Handari, B. D. (2023). Tree-based ensemble methods and their applications for predicting students' academic performance. *International Journal on Advanced Science, Engineering & Information Technology, 13*(3). https://doi.org/10.18517/ijaseit.13.3.16880
9. Baron, M. J. S., Sanabria, J. S. G., & Diaz, J. E. E. (2022). Deep neural network (DNN) applied to the analysis of student dropout in a higher education institution. *Investigación e Innovación en Ingenierías, 10*(1), 202–214. https://doi.org/10.17081/invinno.10.1.5607
10. Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H. Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies, 28*(1), 905–971. https://doi.org/10.1007/s10639-022-11152-y
11. Becker, T., Rousseau, A. J., Geubbelmans, M., Burzykowski, T., & Valkenborg, D. (2023). Decision trees and random forests. *American Journal of Orthodontics and Dentofacial Orthopedics, 164*(6), 894–897. https://doi.org/10.1016/j.ajodo.2023.09.011
12. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953
13. Chen, R., Dai, T., Zhang, Y., Zhu, Y., Liu, X., & Zhao, E. (2024). GBDT-IL: Incremental learning of gradient boosting decision trees to detect botnets in internet of things. *Sensors, 24*(7), 2083. https://doi.org/10.3390/s24072083
14. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.293978
15. Chiu, T. K. (2024). Future research recommendations for transforming higher education with generative AI. *Computers and Education: Artificial Intelligence, 6*, 100197. https://doi.org/10.1016/j.caeai.2023.100197
16. Costa, S. F., & Diniz, M. M. (2022). Application of logistic regression to predict the failure of students in subjects of a mathematics undergraduate course. *Education and Information Technologies, 27*(9), 12381–12397. https://doi.org/10.1007/s10639-022-11117-1
17. Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., & Gopakumar, G. (2021). Educational data mining for predicting students' academic performance using machine learning algorithms. *Materials Today: Proceedings, 47*, 5260–5267. https://doi.org/10.1016/j.matpr.2021.05.646.
18. de Oliveira, C. F., Sobral, S. R., Ferreira, M. J., & Moreira, F. (2021). How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education: A Systematic Literature Review. *Big Data and Cognitive Computing, 5*(4), 64. https://doi.org/10.3390/bdcc5040064
19. Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM, 55*(10), 78–87. https://doi.org/10.1145/2347736.2347755
20. Elugbaju, W. K., Okeke, N. I., & Alabi, O. A. (2024). Conceptual framework for enhancing decision-making in higher education through data-driven governance. *Global Journal of Advanced Research and Reviews, 2*(2), 16–30. https://doi.org/10.58175/gjarr.2024.2.2.0055
21. European Parliament. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). *Official Journal of the European Union.* https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A32016R0679

22. Gaftandzhieva, S., Hussain, S., Hilcenko, S., Doneva, R., & Boykova, K. (2023). Data-driven decision making in higher education institutions: State-of-play. *International Journal of Advanced Computer Science and Applications, 14*(6), 397–405. https://doi.org/10.14569/IJACSA.2023.0140642

23. Gan, W., Qi, Z., Wu, J., & Lin, J. C. W. (2023). Large language models in education: Vision and opportunities. In *2023 IEEE International Conference on Big Data (BigData)* (pp. 4776–4785). IEEE. https://doi.org/10.1109/BigData59044.2023.10386291

24. Gašević, D., Siemens, G., & Sadiq, S. (2023). Empowering learners for the age of artificial intelligence. *Computers and Education: Artificial Intelligence, 4*, 100130. https://doi.org/10.1016/j.caeai.2023.100130

25. Gong, J., Qu, Z., Zhu, Z., Xu, H., & Yang, Q. (2025). Ensemble models of TCN-LSTM-Light GBM based on ensemble learning methods for short-term electrical load forecasting. *Energy*, 134757. https://doi.org/10.1016/j.energy.2025.134757.

26. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* (Vol. 1). MIT Press. https://www.scirp.org/reference/referencespapers?referenceid=3785924

27. Guo, C., & Berkhahn, F. (2016). Entity embeddings of categorical variables. *arXiv preprint arXiv:1604.06737*. https://doi.org/10.48550/arXiv.1604.06737

28. Guzmán-Valenzuela, C., Gómez-González, C., Rojas-Murphy Tagle, A., & Lorca-Vyhmeister, A. (2021). Learning analytics in higher education: A preponderance of analytics but very little learning? *International Journal of Educational Technology in Higher Education, 18*, 1–19. https://doi.org/10.1186/s41239-021-00258-x

29. Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: Data mining, inference and prediction. *The Mathematical Intelligencer, 27*(2), 83–85. https://doi.org/10.1007/BF02985802

30. HELOU, M. A. (2023). A unified framework for big data-driven decision making system in higher education institutes. Journal of Theoretical and Applied Information Technology, 101(12). https://orcid.org/0000-0002-8597-3033

31. Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering, 2022*, 5215722. **https://doi.org/10.1155/2022/5215722**

32. Horst, R. M. (2020). *Higher Education Executives and Data-Driven Decision Making: A Phenomenological Study* (Doctoral dissertation, Concordia University (Oregon)). Retrieved from https://digitalcommons.csp.edu/cup_commons_grad_edd/436

33. Hosen, M. S., Islam, R., Naeem, Z., Folorunso, E. O., Chu, T. S., Al Mamun, M. A., & Orunbon, N. O. (2024). Data-driven decision making: Advanced database systems for business intelligence. *Nanotechnology Perceptions, 20*(3), 687–704. https://doi.org/10.62441/nano-ntp.v20iS3.51

34. Huang, F. L. (2022). Alternatives to logistic regression models in experimental studies. *The Journal of Experimental Education*, *90*(1), 213–228. https://doi.org/10.1080/00220973.2019.1699769

35. Huynh-Cam, T. T., Chen, L. S., & Le, H. (2021). Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance. *Algorithms, 14*(11), 318. https://doi.org/10.3390/a14110318

36. Jalota, C. (2023). An effectual model for early prediction of academic performance using ensemble classification. *J Lang Linguist Soc, 3*(2), 19–33. https://doi.org/10.55529/jlls.32.19.33

37. Ji, X., Chang, W., Zhang, Y., Liu, H., Chen, B., Xiao, Y., & Zhou, S. (2021). Prediction model of hypertension complications based on GBDT and LightGBM. *Journal of Physics: Conference Series, 1813*(1), 012008. https://doi.org/10.1088/1742-6596/1813/1/012008

38. Jiao, Z. (2024). Dynamic financial distress prediction using combined LASSO and GBDT algorithms. *Informatica*, *48*(17), 139–152. https://api.semanticscholar.org/CorpusID:274137271

39. Joshi, A., Saggar, P., Jain, R., Sharma, M., Gupta, D., & Khanna, A. (2021). CatBoost—An ensemble machine learning model for prediction and classification of student academic performance. *Advances in Data Science and Adaptive Analysis, 13*(03n04), 2141002. https://doi.org/10.1142/S2424922X21410023

40. Kaspi, S., & Venkatraman, S. (2023). Data-driven decision-making (DDDM) for higher education assessments: A case study. *Systems, 11*(6), 306. https://doi.org/10.3390/systems11060306

41. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30.* 1-9. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

42. Ke, G., Xu, Z., Zhang, J., Bian, J., & Liu, T. Y. (2019). DeepGBM: A deep learning framework distilled by GBDT for online prediction tasks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 384–394). https://doi.org/10.1145/3292500.3330858

43. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. https://doi.org/10.48550/arXiv.1412.6980
44. Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95), Volume 2 (pp. 1137–1143). Morgan Kaufmann Publishers Inc. Retrieved from http://robotics.stanford.edu/~ronnyk/accEst-long.ps
45. Kotsiantis, S., Kanellopoulos, D., Pintelas, P. (2006) Handling imbalanced datasets: A review. GESTS Interna- tional Transactions on Computer Science and Engineer- ing, 30(1), 25-36. https://www.scirp.org/reference/referencespapers?referenceid=60437
46. Li, G., Cui, J., Fu, H., & Sun, Y. (2024). Light GBM and GA based algorithm for predicting and improving students' performance in higher education in the context of big data. In *2024 7th International Conference on Education, Network and Information Technology (ICENIT)* (pp. 1–5). IEEE. https://doi.org/10.1109/ICENIT61951.2024.00009
47. Liaqat, S., Dashtipour, K., Arshad, K., Assaleh, K., & Ramzan, N. (2021). A hybrid posture detection framework: Integrating machine learning and deep neural networks. *IEEE Sensors Journal, 21*(7), 9515–9522. https://doi.org/10.1109/JSEN.2021.3055898
48. Lin, H., Chung, J. W., Lao, Y., & Zhao, W. (2023). Machine unlearning in gradient boosting decision trees. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 1374–1383). https://doi.org/10.1145/3580305.3599420
49. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems, 30*. https://doi.org/10.48550/arXiv.1705.07874
50. Madhavi, M. A. D. U. R. I., & Nethravathi, D. R. (2022). Gradient boosted decision tree (GBDT) and Grey Wolf Optimization (GWO) based intrusion detection model. *Journal of Theoretical and Applied Information Technology, 100*(16), 4937–4951. https://www.jatit.org/volumes/Vol100No16/5Vol100No16.pdf
51. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35. https://doi.org/10.1145/3457607
52. Munir, H., Vogel, B., & Jacobsson, A. (2022). Artificial Intelligence and Machine Learning Approaches in Digital Education: A Systematic Revision. *Information, 13*(4), 203. https://doi.org/10.3390/info13040203
53. Nguyen, A., Tuunanen, T., Gardner, L., & Sheridan, D. (2021). Design principles for learning analytics information systems in higher education. *European Journal of Information Systems, 30*(5), 541–568. https://doi.org/10.1080/0960085X.2020.1816144
54. Okewu, E., Adewole, P., Misra, S., Maskeliunas, R., & Damasevicius, R. (2021). Artificial neural networks for educational data mining in higher education: A systematic literature review. *Applied Artificial Intelligence, 35*(13), 983–1021. https://doi.org/10.1080/08839514.2021.1922847
55. Papadogiannis, I., Wallace, M., & Karountzou, G. (2024). Educational data mining: A foundational overview. *Encyclopedia, 4*(4), 1644–1664. https://doi.org/10.3390/encyclopedia4040108
56. Rabelo, A. M., & Zárate, L. E. (2025). A model for predicting dropout of higher education students. *Data Science and Management, 8*(1), 72–85. https://doi.org/10.1016/j.dsm.2024.07.001
57. Romero C, Ventura S. (2020). Educational data mining and learning analytics: An updated survey. WIREs Data Mining Knowl Discov. 2020;10:e1355. **https://doi.org/10.1002/widm.1355**
58. Rozali, M. Z., & Zakaria, A. F. (2024). A data-driven design module development for enhancing multi-attribute decision-making (MADM) efficacy. In *2024 9th International STEM Education Conference (iSTEM-Ed)* (pp. 1–6). IEEE. https://doi.org/10.1109/iSTEM-Ed62750.2024.10663085
59. Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE, 10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432
60. Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial Intelligence and Learning Analytics in Teacher Education: A Systematic Review. *Education Sciences, 12*(8), 569. https://doi.org/10.3390/educsci12080569
61. Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review, 46*(5), 30. https://doi.org/10.17471/2499-4324/195
62. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929–1958. https://www.academia.edu/download/102657176/srivastava14a.pdf
63. Taherdoost, H. (2023). Deep learning and neural networks: Decision-making implications. *Symmetry, 15*(9), 1723. https://doi.org/10.3390/sym15091723

64. Tao, T., Sun, C., Wu, Z., Yang, J., & Wang, J. (2022). Deep neural network-based prediction and early warning of student grades and recommendations for similar learning approaches. *Applied Sciences, 12*(15), 7733. https://doi.org/10.3390/app12157733

65. Tian, L., Feng, L., Yang, L., & Guo, Y. (2022). Stock price prediction based on LSTM and LightGBM hybrid model. *The Journal of Supercomputing, 78*(9), 11768-11793. https://doi.org/10.1007/s11227-022-04326-5

66. Wang, C., Chang, L., & Liu, T. (2022). Predicting student performance in online learning using a highly efficient gradient boosting decision tree. In *International Conference on Intelligent Information Processing* (pp. 508–521). Springer. https://doi.org/10.1007/978-3-031-03948-5_41

67. Wang, J., Wu, H., Cheng, X., Guo, Z., Yu, K., & Shen, Y. (2022). Data-driven intelligent decision for multimedia medical management. *Multimedia Tools and Applications, 81*(29), 42023–42039. https://doi.org/10.1007/s11042-021-11545-0

68. Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments, 9*(1), 11. https://doi.org/10.1186/s40561-022-00192-z

69. Yan, J., Chen, J., Wang, Q., Chen, D. Z., & Wu, J. (2024). Team up GBDTs and DNNs: Advancing efficient and effective tabular prediction with tree-hybrid MLPs. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 3679–3689). https://doi.org/10.1145/3637528.3671964

70. Yousafzai, B. K., Khan, S. A., Rahman, T., Khan, I., Ullah, I., Ur Rehman, A., ... Cheikhrouhou, O. (2021). Student-performulator: Student academic performance using hybrid deep neural network. *Sustainability, 13*(17), 9775. https://doi.org/10.3390/su13179775

71. Zheng, J., Chen, B., Li, J., Liu, Y., & Liu, J. (2023). Online behavior prediction based on deep learning in healthcare. *Journal of Mechanics in Medicine and Biology, 23*(4), 2340010. https://doi.org/10.1142/S0219519423400109

Publishing and Ethical Statements / بيانات النشر والالتزام الأخلاقي

N		Publication Data in English	بيانات النشر بالعربية		م
1	Authors' Contributions:	First Author: Design, methodology, data collection, analysis, and drafting. Second Author: Supervision, summarizing, development, and final review for publication.	الباحث الأول: التصميم، المنهجية، جمع وتحليل البيانات، وكتابة المسودة. الباحث الثاني: الإشراف العلمي، التلخيص، التطوير، والمراجعة النهائية للنشر.	مساهمة الباحثين:	1
2	Conflict	No conflicts of interest.	لا يوجد تضارب مصالح.	تضارب المصالح:	2
3	Funding	Self-funded (No external grant).	تمويل ذاتي (لا يوجد دعم خارجي).	التمويل:	3
4	Copyright Licensed under:	(CC BY-NC-ND)		حقوق النشر مرخص بموجب:	4
5	ReviewProcess:	Double-blind peer review.	تحكيم مزدوج التعمية.	آلية التحكيم:	5
6	Plagiarism Check:	Verified via (iThenticate)	تم الفحص عبر	فحص الانتحال:	6
7	Data Availability:	Available upon request.	متاحة عند الطلب.	إتاحة البيانات:	7