

A novel AI-driven model for student dropout risk analysis with explainable AI insights

Sumaya Mustofa^{*}, Yousuf Rayhan Emon, Sajib Bin Mamun, Shabnur Anonna Akhy, Md Taimur Ahad

Department of Computer Science and Engineering, Daffodil International University, Savar, Dhaka, Bangladesh

ARTICLE INFO

Keywords:

Dropout
HLRNN
Hybrid logistic regression and neural network
Risk analysis
SHAP
LIME

ABSTRACT

The increasing number of students dropping out of school due to social, economic, personal (e.g., depression or persistent failure), and health issues is a growing concern for governments, educators, and guardians. Identifying and analyzing the factors contributing to student dropout is crucial. Various machine learning, analytical, and statistical models have been proposed to address this issue. However, the existing models have several limitations in providing a precise and automated system for predicting dropout risk and analyzing the factors behind this. Besides, generating a balanced dataset is also a limitation as 'Dropouts' are less than the 'Non-dropouts'. Moreover, selecting significant features contributing to student dropout and non-dropout is also very important in developing a model. However, this study introduces a comprehensive machine learning (ML) and explainable AI (XAI) based methodology to address these limitations. Firstly, the imbalanced dataset problem was handled using the Upsampling technique by adjusting the minority class 'Dropout'. Then, the feature selection method Recursive Feature Elimination (RFE) is used with Cross-Validation (CV) as the RFE-CV method to select the most significant features. After preprocessing, this study proposed a hybrid model named the Hybrid Logistic Regression and Neural Network (HLRNN) model, which predicts student dropout with 96% accuracy, outperforming other experimented models as well as the parent models Logistic Regression and Artificial Neural Network with 2% and 3% accuracy. Finally, the XAI model The SHapley Additive exPlanations (SHAP), and Local Interpretable Model-agnostic Explanations (LIME) are deployed to analyze the risk factors associated with student dropout. This approach aims to assist institutions and educational stakeholders in formulating policies for student retention, enabling early intervention to reduce dropout rates.

1. Introduction

Recent advancements in technology, particularly in machine learning (ML) and artificial intelligence (AI), have shown advancement in addressing the issue of student dropout. Machine learning algorithms can analyze complex student data from various sources, such as academic performance, attendance records, socio-economic status, and behavioral trends, to identify patterns that are often missed by traditional methods (Poudyal et al., 2020; Prasanth & Alqahtani, 2023). This allows educational institutions to detect at-risk students early and understand the multifaceted reasons behind their potential dropout (Martins et al., 2021; Mduma et al., 2019). ML has drawn a lot of interest in recent years as a potential solution to the dropout rate among students because it enables real-time monitoring and individualized intervention

strategies. For example, AI-driven models can recommend specific interventions, such as additional tutoring or financial assistance, based on the unique needs of each student. This not only increases retention rates but also optimizes the allocation of institutional resources (Mduma, 2023). Several machine learning algorithms have been proposed, including hybrid models, data balancing techniques, and ensemble methods, which have demonstrated high accuracy in predicting student dropout risk (Cho et al., 2023; Latif et al., 2021; Rastrollo-Guerrero et al., 2020; Ahmad Tarmizi et al., 2019). In Computer Science, using ML/AI for dropout prediction marks significant progress in creating models that can manage large and complex datasets while addressing the challenge of interpretability through tools like Explainable AI (e.g., SHAP). In education, these models offer practical insights that help schools and universities take timely action, ultimately leading to better

^{*} Corresponding author.

E-mail addresses: sumaya15-3445@diu.edu.bd (S. Mustofa), yousuf15-3220@diu.edu.bd (Y.R. Emon), sajib15-3435@diu.edu.bd (S.B. Mamun), shabnur15-3472@diu.edu.bd (S.A. Akhy), taimurahad.cse@diu.edu.bd (M.T. Ahad).

<https://doi.org/10.1016/j.caeai.2024.100352>

Received 11 August 2024; Received in revised form 28 November 2024; Accepted 21 December 2024

Available online 24 December 2024

2666-920X/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

student retention and overall educational quality improvement. Mduma, (2023) highlighted the potential of machine learning models to address this challenge, especially in developing countries where resources are limited. Additionally, Rastrollo-Guerrero et al. (2020) have reviewed various data mining and machine learning techniques, emphasizing their effectiveness in predicting student performance and dropout. In this study, Dropout is defined as students discontinuing their studies within the academic semester (KLITZKE & CARVALHAES, 2023; Ortiz-Lozano et al., 2020). Student dropout remains an issue in educational institutions worldwide, affecting student's academic success and the overall quality of education. The reasons for student dropout are multifaceted, including social, economic, personal, and health-related factors (Meijer et al., 2019; Prasanth & Alqahtani, 2023; Ahmad Tarmizi et al., 2019). Student dropout is a serious issue in universities since it harms students and educational establishments. During the COVID-19 pandemic, the lack of in-person interactions with instructors significantly contributed to increased dropout rates (Ifenthaler & Yau, 2020). However, this phenomenon underscores the importance of early prediction and intervention to prevent students from leaving higher education. Predicting dropout risk early allows for implementing support mechanisms and interventions, thereby increasing student retention rates and improving academic outcomes (Meijer et al., 2019). Moreover, understanding the factors influencing student failure and success, such as motivation, attendance, economic condition, mental, and social impact, is important for developing effective predictive models (Beaulac & Rosenthal, 2019).

Significant research gaps remain despite the progress in developing predictive models for student dropout risk prediction. Among other limitations, the following are mentionable.

- Most studies have focused on using either traditional statistical models or standalone machine learning models and few of the studies achieved remarkable accuracy (Cho et al., 2023; Deb et al., 2024; Martinho et al., 2013; Seo et al., 2024; Talebi et al., 2024; Thammasiri et al., 2014). However, exploring hybrid approaches that combine the strengths of multiple models and enhance prediction accuracy is still limited.
- Dropout datasets are often imbalanced, with far fewer dropouts than non-dropouts. Researchers tend to use 'SMOTE' as a data imbalance solution technique (Kim et al., 2023; Masood et al., 2024; Mulyani et al., 2019; Thammasiri et al., 2014; Villar and de Andrade, 2024). However, 'SMOTE' introduces noise when synthetic samples are generated which can decrease the performance of a machine learning model.
- The existing machine learning models are treated as black boxes. These models predict student dropout but make it difficult to interpret the results and understand the factors contributing to predictions. Hence, there's a gap in developing machine learning models that balance high accuracy with interpretability.
- Existing researchers provide limited ideas about what are the most influential factors that contribute to student dropout risk analysis so that educators and stakeholders can take necessary steps to prevent students from dropping out.

To address the gaps identified in existing research, this study aimed to explore a few research questions that can help to build more robust techniques to analyze student dropout risk and mitigate it.

RQ1: While most of the studies experimented with a standalone machine learning model, how can a hybrid machine learning model improve the prediction accuracy of student dropouts compared to existing methods?

RQ2: How does addressing class imbalance through up-sampling techniques without adding noise in the dataset and enhancing the performance of dropout prediction models?

RQ3: What critical factors influencing student dropout can be identified using Explainable AI tools such as SHAP and LIME?

RQ4: How can an interpretable and accurate model aid educators in implementing effective interventions to reduce dropout rates?

Specifically, this study proposes a Hybrid Logistic Regression and Neural Network (HLRNN) model to enhance the accuracy of dropout prediction. It tackles the issue of imbalanced datasets through an up-sampling technique, ensuring stable and reliable performance. Additionally, the study integrates Explainable AI tools, such as SHAP and LIME, to interpret the factors influencing predictions. These contributions aim to improve the accuracy and interpretability of student dropout prediction models, enabling institutions to implement timely, data-driven interventions to improve retention.

2. Literature review

Recent studies have utilized various machine learning (ML) algorithms to predict student dropout risk. The first group of researchers implemented different ML algorithms. Rastrollo-Guerrero et al. (2020) analyzed to show the modern techniques applied for predicting student performance, such as machine learning, collaborative filtering, recommender systems, and artificial neural networks. Latif et al. (2021) presented that Sequential Minimal Optimization (SMO) achieved a minimum mean absolute error of 2.350 for predicting final exam grades, while Linear Regression (LR) had a minimum mean absolute error of 1.978 for midterm exams. Ahmad Tarmizi et al. (2019) and Garg et al. (2021) highlighted the effectiveness of a Support Vector Machine (SVM) with a Polynomial Kernel and standard SVM in predicting student attrition and performance, respectively. Cho et al. (2023) utilized a Light Gradient Boosting Machine (LightGBM) to predict student dropout, achieving an F1 score of 0.840. Pecuchova and Drlik (2023) presented that ensemble methods like AdaBoost and XGBoost improved overall accuracy, recall, precision, and F1 scores by 2–4% compared to traditional classification algorithms. Mduma et al. (2019) highlighted the necessity of developing machine learning models that consider the unique challenges of developing countries. Beaulac and Rosenthal (2019) utilized random forests to predict whether students would obtain an undergraduate degree. Miguéis et al. (2018) presented random forests were superior to other classification techniques such as decision trees, support vector machines, naive Bayes, bagged trees, and boosted trees. Thammasiri et al. (2014) presented that the support vector machine combined with the SMOTE data-balancing technique achieved the best classification performance. However, this part of the study focused on either traditional statistical methods or machine-learning techniques. While studies like Pecuchova and Drlik (2023) demonstrated success using ensemble methods, few have incorporated models such as logistic regression alongside neural networks to enhance prediction performance. Our hybrid model addresses this gap by leveraging the advantages of both approaches.

The second group of researchers utilized Deep Learning and Neural Networks. Latif et al. (2021) included Multilayer Perceptron (MLP) in their study, which, along with other classifiers, was evaluated for predicting students' performance. Deep Neural Networks (DNN) were also considered by Cho et al. (2023) but did not outperform LightGBM. Martins et al. (2021) highlighted using Artificial Neural Networks (ANN) in predicting student performance, noting that state-of-the-art boosting algorithms performed better than standard methods but still fell short in identifying cases in minority classes. However, these studies focus on particularly deep learning, which are treated as black boxes, offering little insight into the factors behind their predictions. Though explainable AI tools like SHAP and LIME have been introduced in other domains, they have not been widely applied to student dropout prediction. By using SHAP and LIME in our model, we provide interpretable insights into dropout risks, ensuring that predictions are accurate and understandable to educators.

The third group of researchers focused on data features and sources. [Ahmad Tarmizi et al. \(2019\)](#) and [Bravo et al. \(2023\)](#), emphasized the importance of academic records, such as CGPA, course credits, and high school study results, in predicting dropout risk. [Bravo et al. \(2023\)](#) specifically noted that information from completed subjects provided the best feature model for dropout prediction. [Martinho et al. \(2013\)](#) focused on demographic attributes, behavior data, and socioeconomic data and applied a fuzzy-ARTMAP neural Network and achieved an accuracy of 85%. Family income, disability, and the number of dependents were significant predictors of student attrition in [Ahmad Tarmizi et al. \(2019\)](#). [Prasanth and Alqahtani \(2023\)](#) also included demographic information, behavior data, and extracurricular activities to enhance their predictive model. [Kostopoulos et al. \(2017\)](#) focused on the efficiency of active learning methodologies to predict dropout rates in distance education. [Meijer et al. \(2019\)](#) found that concrete pre-university behaviors were more predictive than psychological attributions such as self-efficacy. Moreover, the issue of class imbalance remains a challenge in dropout prediction datasets, with the minority class (dropouts) being underrepresented. While techniques like SMOTE have been used, they have not been applied consistently across all models. Our model incorporates up-sampling to address this imbalance, improving the accuracy of predictions for the dropout class and ensuring more reliable interventions for at-risk students.

The fourth group of researchers focused on accuracy and performance metrics. [Pecuchova and Drlik \(2023\)](#) achieved a 2–4% improvement in performance metrics using ensemble methods. [Cho et al. \(2023\)](#) reported an F1-score of 0.840 with LightGBM, demonstrating its superior performance in handling class imbalance. [Latif et al. \(2021\)](#) used MAE to evaluate the accuracy of their models, with SMO and LR showing the lowest errors for final and midterm exams, respectively. [Miguéis et al. \(2018\)](#) validated their model with a dataset of 2459 students and found that random forests achieved an accuracy above 95% in predicting students' performance levels early in their academic paths. [Thammasiri et al. \(2014\)](#) achieved a 90.24% overall accuracy using the support vector machine combined with the SMOTE data-balancing technique.

3. Methodology

3.1. Dataset description

A dataset (.csv) has been collected from a public repository to experiment with the proposed student dropout prediction model. The dataset contains 34 features (See [Table 1](#)) and 4424 samples per feature. The dependent feature, 'Target,' categorizes students into three groups: 'Enrolled,' 'Dropout,' and 'Graduate.' Since this research focuses on predicting student dropouts, the 'Graduate' and 'Enrolled' classes are combined into a single 'Non-dropout' class, resulting in 3003 samples labeled as 'Non-dropout' and 1421 samples labeled as 'Dropout'.

The independent features in the dataset capture demographic, socio-economic, and academic characteristics, providing insight into predictive factors for student retention and dropout. These include attributes such as marital status, tuition payment status, age at enrollment, and curricular performance metrics. The data originates from higher education, with a focus on student retention and completion.

In this study, the term 'Dropout' refers to students who discontinue their studies within the observation period of the semester. This definition is consistent with the dataset, which categorizes students as 'Dropout,' 'Enrolled,' or 'Graduate' based on their academic outcomes. Specifically, 'Dropout' refers to students who discontinued their studies within the observation period, 'Graduate' refers to those who completed their academic programs on time, and 'Enrolled' refers to those who continue their studies.

Table 1

Feature name, descriptions, and corresponding datatype of the experimented dataset.

Serial No.	Feature Name	Description
1	Marital status	Student's marital status. (Categorical)
2	Application mode	Application method used by the student. (Categorical)
3	Application order	Student applications are listed in order of submission. (Numerical)
4	Course	Courses taken by the student. (Categorical)
5	Daytime/evening attendance	Attendance at classes during the day or at night. (Categorical)
6	Previous qualification	Qualifications obtained before enrolling in higher education. (Categorical)
7	Nationality	A student's nationality. (Categorical)
8	Mother's qualification	Student's mother's qualification. (Categorical)
9	Father's qualification	Student's father's qualification. (Categorical)
10	Mother's occupation	Student's mother's occupation. (Categorical)
11	Father's occupation	Student's father's occupation. (Categorical)
12	Displaced	The student's status as a displaced person. (Categorical)
13	Educational special needs	Any special educational needs that the student may have. (Categorical)
14	Debtor	Debtor status of the student. (Categorical)
15	Tuition fees up to date	If the student has paid his or her tuition fees on time. (Categorical)
16	Gender	The student's gender. (Categorical)
17	Scholarship holder	Scholarship status of the student. (Categorical)
18	Age at enrollment	Age of the student when enrolled. (Numerical)
19	International	The student's status as an international student. (Categorical)
20	Curricular units 1st sem (credited)	The student credited the number of curricular units in the first semester. (Numerical)
21	Curricular units 1st sem (enrolled)	The student enrolled in the number of curricular units in the first semester. (Numerical)
22	Curricular units 1st sem (evaluations)	The student was evaluated on the number of curricular units in the first semester. (Numerical)
23	Curricular units 1st sem (approved)	The student was approved for the number of curricular units in the first semester. (Numerical)
24	Curricular units 1st sem (grade)	The student was graded on the number of curricular units in the first semester. (Numerical)
25	Curricular units 1st sem (without evaluations)	The student did not evaluate the number of curricular units in the first semester. (Numerical)
26	Curricular units 2nd sem (credited)	The student credited the number of curricular units in the second semester. (Numerical)
27	Curricular units 2nd sem (enrolled)	The student enrolled in the number of curricular units in the second semester. (Numerical)
28	Curricular units 2nd sem (evaluations)	The student was evaluated on the number of curricular units in the Second semester. (Numerical)
29	Curricular units 2nd sem (approved)	The student was approved for the number of curricular units in the Second semester. (Numerical)
30	Curricular units 2nd sem (grade)	The student was graded on the number of curricular units in the Second semester. (Numerical)
31	Curricular units 2nd sem (without evaluations)	The student did not evaluate the number of curricular units in the Second semester. (Numerical)
32	Unemployment rate	The unemployment rate in the students' area.
33	Inflation Rate	The inflation rate in the students' area.
34	GDP	GDP from the students' region.

3.2. Data preprocessing

After setting up the dataset for two categories, a few common pre-processing techniques were used to get the dataset ready, such as 'Upsampling', and 'Min-Max Scaler'. Then, the 'Recursive Feature Elimination (RFE)' feature selection method and cross-validation together as the 'RFECV' manual approach is applied to pick the best features.

3.2.1. Upsampling: handling imbalance dataset

The imbalance problem of the experimented dataset has been handled through the 'Upsampling' technique. Firstly, 'Non-dropout' and 'Dropout' classes have been differentiated as majority and minority classes respectively. Mathematically, 'Upsampling' involves replicating samples from the minority class using techniques like 'Resample' to balance the class distribution (Barr et al., 2022)

Upsampled Class (Dropout) = Original Class (Dropout) + Generated Samples
Upsampled Class (Dropout) = Original Class (Dropout) + Generated Samples

3.2.2. Recursive Feature Elimination (RFE) — cross-validation (CV): RFECV method

Recursive Feature Elimination is an iterative feature selection method that works recursively removing features from the dataset. It ranks the features based on their relation with the target variable and removes the least important features. This process continues repeatedly. To enhance the robustness of the RFE method, cross-validation (CV) has been implemented with RFE to select features. To cross-validate the data K-fold cross-validation has been used where $K = 5$. This method is a manual RFECV approach, unlike the Scikit-learn feature selection module (Zollanvari, 2023). The manual RFECV provides flexibility to better control the partitioning of data and evaluation process, allowing for a more thorough exploration of the importance of features at each iteration. It enables the definition of specific stopping criteria, evaluation metrics, and model selection better suited for the dataset. Applying the following approach, the dataset is finally prepared to carry out further experimental implementation.

3.3. Experimental methodology

Five machine learning classifiers: Naïve Bayes, Logistic Regression (LR), Random Forest, Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and two neural network classifiers: Artificial Neural Network (ANN) and Multi-Layer Perceptron (MLP) has been experimented into the pre-processed dataset. After the performance comparison, Logistic regression (LR) and Artificial Neural Network (ANN) were selected as the best performers. Then, a hybrid classification model Hybrid Logistic regression, and Artificial Neural Network (HLRNN) have been developed. After the training of the HLRNN model, weights will be saved as .h5 file for further prediction. This comprehensive approach ensures a robust and accurate student dropout prediction and classification model. In conclusion, an analysis of student dropout risk has been conducted utilizing the Explainable AI (XAI) method known as Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). By examining the dataset's features, it is essential to identify the most significant factors contributing to the HLRNN model. These factors represent the primary indicators or reasons for student dropout (See Fig. 1).

3.4. Proposed HLRNN model

In this study, a hybrid classification model has been implemented that combines Logistic Regression (LR) and an Artificial Neural Network (ANN) model named Hybrid Logistic Regression and Neural Network Model (HLRNN) (See Fig. 2). The HLRNN model incorporates the strengths of Logistic Regression (interpretable and efficient) and Artificial Neural Networks (able to capture complex patterns) to develop a more accurate model. Firstly, Logistic Regression (LR) is applied to the dropout dataset. Initially, the dataset contained 34 independent features ($x_1 \dots x_{34}$). By using the (LR), predict the output for each data point and generate a new column for each data's predicted output. Secondly, the new predicted output column (x_{35}) is concatenated with the original dropout dataset. Finally, ANN is used for the final classification by utilizing the new dropout dataset with ($x_1 \dots x_{35}$) independent features and predicting two probability values: dropout or not dropout. The proposed model has been designed to meet the following benefits.

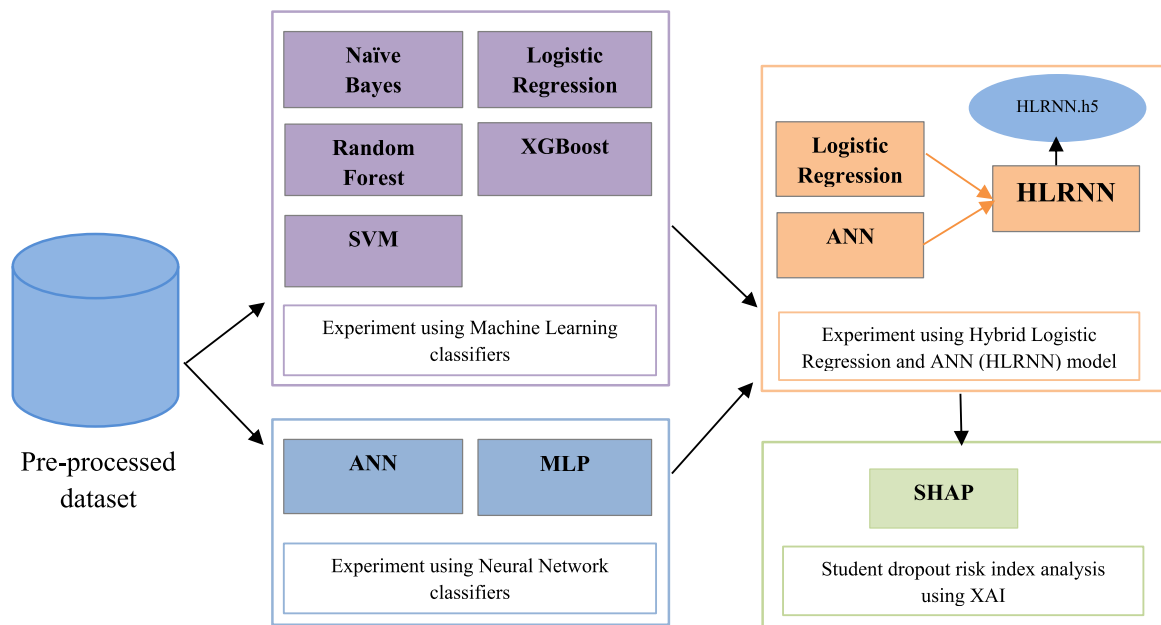


Fig. 1. Experimental process of student dropout prediction model.

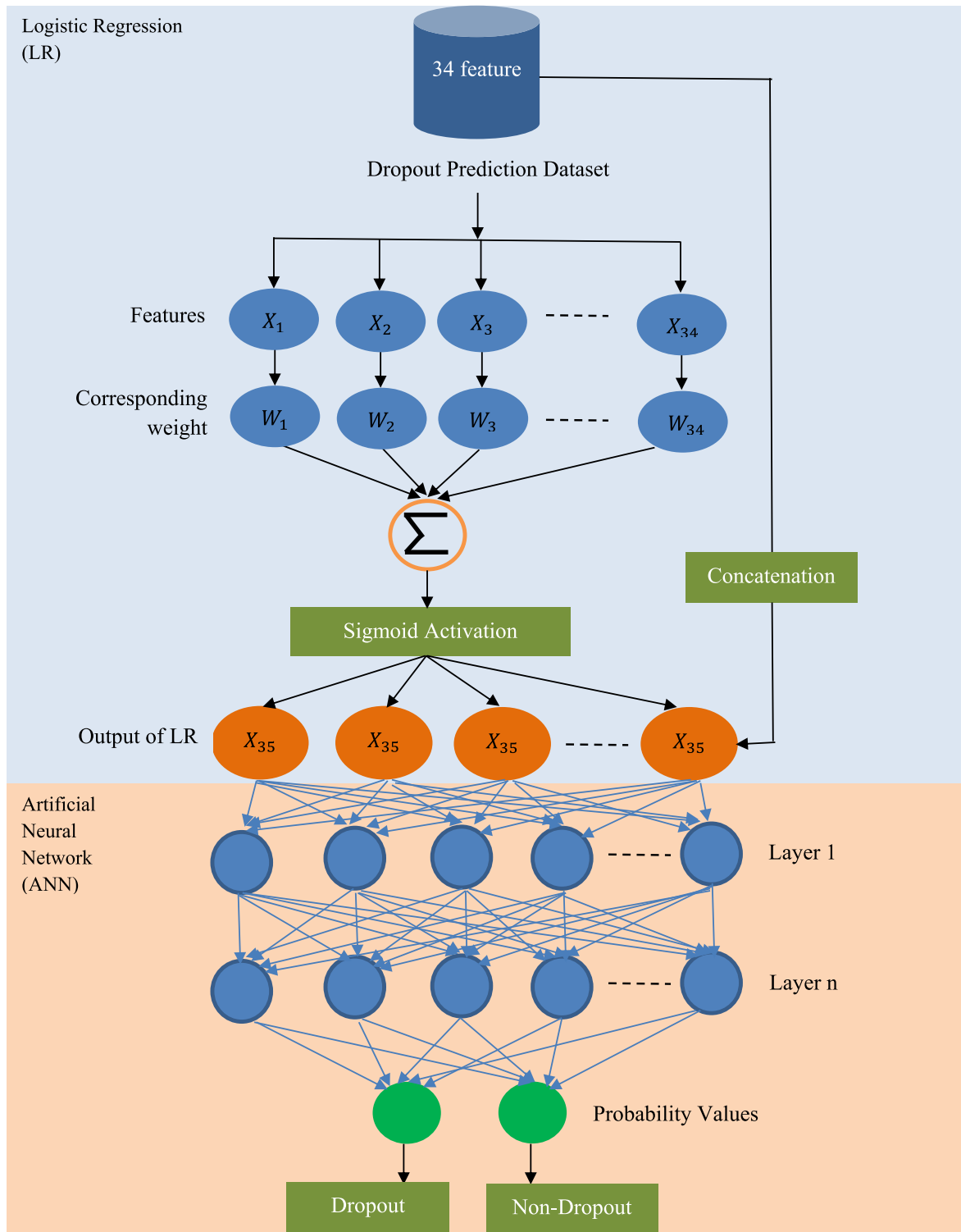


Fig. 2. Architecture of the proposed HLRNN model.

- The dropout problem involves multiple demographics, socio-economic, and academic features that have varying degrees of linearity and non-linearity. Combining Logistic Regression (LR) with an Artificial Neural Network (ANN), the HLRNN model enhances LR's interpretability and ANNs' non-linear pattern recognition capability, allowing it to capture complex patterns that a single model could miss.
- While LR or ANN can provide reasonable accuracy as a single model, a hybrid approach can improve accuracy. The initial LR layer

captures linear patterns and adds a probability-based feature (x_{35}), enhancing the ANN's input data.

- The HLRNN hybrid approach reduces the probability of overfitting to particular patterns within the data. By concatenating the new feature x_{35} that the LR model predicts, the HLRNN model enhances the stability of LR, potentially reducing overfitting compared to using a single ANN model.
- In educational contexts, explain ability is essential for administrators and educators to understand the factors leading to dropout

predictions. This combination supports explainable AI, which is particularly valuable in educational and policy-driven applications.

LR: To develop the HLRNN model, the preprocessed dataset has been trained with the Logistic Regression model. Then, using X_{train} and X_{val} , a probability value has been predicted for each instance of the training and validation data. The expected value that has been generated from probability is combined as a new feature with the original dataset.

Denoting the features of the student dropout dataset as x and their corresponding weight is W . There are 34 features in the dataset, hence, the feature range is $x_1 \dots x_{34}$, and the corresponding weight of those features is in the range $W_1 \dots W_{34}$. Since this study represents binary classification ('Dropout', 'Non-dropout'), the logistic regression model predicts the probability according to the following function and the y ($y_{probability_train}$, $y_{probability_val}$) is the probability output of given instance feature x (X_{train} , X_{val}). The y determining logistic regression function:

$$P(y|x) = \frac{1}{1 + e^{-(W_0 + W_1x_1 + W_2x_2 + \dots + W_{34}x_{34})}} \quad (1)$$

Here, W_0 is the intercepting term. According to the architectural orientation of the logistic regression, the logistic function 'sigmoid' has been used to map the predicted values of X_{train} and X_{val} to probability. The sigmoid function is denoted by $F(z)$.

$$F(z) = \frac{1}{1 + e^{-z}} = \frac{e^z}{1 + e^z} \quad (2)$$

Here, z is the linear combination of the features and their corresponding weight:

$$z = W_0 + W_1x_1 + W_2x_2 + \dots + W_{34}x_{34} \quad (3)$$

After training and collecting the probability value generated from equation (1) using Logistic Regression, the newly predicted probability column was added as the 35th feature of the original dataset. The new dataset feature range is $x_1 \dots x_{35}$ with corresponding weight range $W_1 \dots W_{35}$. Now, the new dataset with 35 features will be used to train the Artificial Neural Network (ANN) model.

ANN: Probability value generated from Logistic Regression has been added as x_{35} feature of the dataset. Finally, ANN applied as a classifier into the newly generated dataset for student dropout prediction. The ANN applied as following:

A neural network model follows two principles of forward and backward propagation. In forward propagation, the ANN model is driven by summation and activation functions. The summation function will be calculated by multiplying all the features x by their weight W and a bias b will be added too as equation (5).

$$Y = W_1x_1 + W_2x_2 + \dots + W_{34}x_{34} + b \quad (4)$$

This summation function Y is applied over the activation function similar to logistic regression (see equation (3)). Then the output of each two neuron will be multiplied with the weight of the next feature. This process will be continued according to the user-declared hidden layer. Dropout layer will be added after each hidden layer to prevent overfitting of the data. Finally, the output layer will use *softmax* as an activation function to obtain the probability distribution over the 'Dropout' and 'Non-dropout' classes. This procedure is represented by the following equations:

Input Layer: Let, the new dataset with features $x_1, x_2, x_3 \dots x_{35}$. These 35 features will be inputted as neurons into the ANN model.

Hidden Layer: Hidden layers will perform the summation and activation functionality on forward and backward propagation.

Hidden Layer 1
Input: $x_1, x_2, x_3 \dots x_{35}$
Output: $A_{1,1}, A_{1,2}, H_{1,3} \dots A_{1,n}$

(continued on next column)

(continued)

Layer Declaration: $H_{1,1}, H_{1,2}, H_{1,3} \dots H_{1,n}$

For neuron $H_{1,1}$:

$$z_{1,1} = W_{1,11}x_1 + W_{1,12}x_2 \dots W_{1,135}x_{35} + b_{1,1}$$

$$a_{1,1} = \text{Sigmoid}(z_{1,1})$$

For neuron $H_{1,2}$:

$$z_{1,2} = W_{1,21}x_1 + W_{1,22}x_2 \dots W_{1,235}x_{35} + b_{1,2}$$

$$a_{1,2} = \text{Sigmoid}(z_{1,2})$$

For neuron $H_{1,n}$:

$$z_{1,n} = W_{1,n1}x_1 + W_{1,n2}x_2 \dots W_{1,n35}x_{35} + b_{1,n}$$

$$a_{1,n} = \text{Sigmoid}(z_{1,n})$$

Hidden Layer 2

Input: $A_{1,1}, A_{1,2}, H_{1,3} \dots A_{1,n}$

Output: $A_{2,1}, A_{2,2}, H_{2,3} \dots A_{2,m}$

Layer Declaration: $H_{2,1}, H_{2,2}, H_{2,3} \dots H_{2,m}$

For neuron $H_{2,1}$:

$$z_{2,1} = W_{2,11}A_{1,1} + W_{2,12}A_{1,2} \dots W_{2,1n}A_{1,n} + b_{2,1}$$

$$a_{2,1} = \text{Sigmoid}(z_{2,1})$$

For neuron $H_{2,2}$:

$$z_{2,2} = W_{2,21}A_{1,1} + W_{2,22}A_{1,2} \dots W_{2,2n}A_{1,n} + b_{2,2}$$

$$a_{2,2} = \text{Sigmoid}(z_{2,2})$$

For neuron $H_{2,m}$:

$$z_{2,m} = W_{2,m1}A_{1,1} + W_{2,m2}A_{1,2} \dots W_{2,mn}A_{1,n} + b_{2,m}$$

$$a_{2,m} = \text{Sigmoid}(z_{2,m})$$

Layer forming continued till the last Hidden Layer L .

Output Layer: This layer formation will continue to any last layer L and finally, the output will be measured using the softmax function. The output of the last layer z_l will go through the softmax function and classification output y_l will be calculated as the student dropout prediction model HLRNNs' final output.

$$y_l = \text{Softmax}(z_l) \quad (5)$$

3.5. Algorithm of the proposed HLRNN model

Algorithm 1. Algorithm of the HLRNN model

1. **Input:** Pre-processed CSV data
2. **Output:** Student Dropout Prediction model
3. **Model and Parameter Initialization:** LR_model = LogisticRegression ()
ANN_model = Neural_Netwok (layers)
epochs = 300
batch_size = 48
val_split = 0.2
4. **# HLRNN model implementation**
5. *for* folds in k-fold cross-validation:
6. $data_split(X_{train}, y_{train}, X_{val}, y_{val}, X_{test}, y_{test})$
- 7.
8. **#Logistic Regression model**
9. LR_model = LR_model.train(X_{train}, y_{train})
- 10.
11. **#Predicting probability using the Logistic Regression model**
12. $y_{probability_train} = LR_model.predict(X_{train})$
13. $y_{probability_val} = LR_model.predict(X_{val})$
- 14.
15. **#Combining Predictive output with original feature**
16. $X_{train_combined} = ColumnStack(X_{train}, y_{probability_train})$
17. $X_{val_combined} = ColumnStack(X_{val}, y_{probability_val})$
- 18.
19. **#Logistic Regression model**
20. ANN_model.train($X_{train_combined}, y_{train}, epochs, batch_size, val_split$)
21. $y_pred = ANN_model.evaluate(X_{val_combined})$
- 22.
23. **# Collect the Final Evaluation of the HLRNN model**
24. Evaluate (y_{test}, y_{pred})
25. Save the model as.h5

3.6. Evaluation of model

The performance of this study has been measured based on a few

criteria such as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

Here, TP: True Positive, FP: False Positive, TN: True Negative, and FN: False Negative.

3.6.1. Error evaluation metrics

$$Mean Absolute Error = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}| \quad (10)$$

$$Root Mean Squared Error = \sqrt{\frac{\sum_{i=1}^N \|y(i) - \hat{y}(i)\|^2}{N}} \quad (11)$$

$$Relative Absolute Error = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i - \bar{y}_i|} \quad (12)$$

$$where \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (13)$$

3.6.2. Precision-recall (PR) curve

A Precision-Recall (PR) curve represents Recall ($\frac{TP}{TP+FN}$) values on the x-axis and Precision ($\frac{TP}{TP+FP}$) values on the y-axis. It represents the trade-off between Precision and Recall.

3.6.3. Receiver Operating Characteristic (ROC) curve

A Receiver Operating Characteristic (ROC) curve represents the False Positive Rate ($\frac{FP}{TN+FP}$) on the x-axis and True Positive Rate ($\frac{TP}{TP+FN}$) on the y-axis. It represents the trade-off between Sensitivity (True Positive Rate) and Specificity (1- False Positive Rate).

4. Result collection

4.1. Performance of the traditional machine learning algorithm

Among the five classifiers experimented with, Logistic Regression (LR) displayed the best accuracy with 94%, and similar precision, recall, and F1-score. Random Forest and XGBoost performed slightly less than the Logistic Regression and achieved 93% accuracy. Support Vector Machine (SVM) achieved an accuracy of 92% which is slightly less than the other better-performing model. On the other hand, Naïve Bayes performed poorly than other machine learning classifiers with 86% accuracy and 85% precision, recall, and F1-score respectively. Naïve Bayes

has an accuracy difference of 6% with the best-performing classifier (see Table 2).

The Mean Absolute Error (MAE) measures the average difference between actual and predicted values. The Relative Absolute Error (RAE) indicates the relative magnitude of the errors while the Root Mean Squared Error (RMSE) indicates the average magnitude of the faults. Naïve Bayes shows the most errors among all the classifiers with 14.93% MAE, 31.33% RAE, and 38.64% RMSE. Performance measurements show that Logistic Regression performs effectively with fewer mistakes with 27.51% RMSE, 6.31% MAE, and 16.54% RAE. Random Forest performs well with moderate error rates. With reduced MAE and RAE errors, XGBoost outperforms Random Forest. SVM has greater error rates compared to XGBoost and Random Forest (see Table 3).

The total effectiveness of the model is measured by the area under the PR curve (AUC-PR). The left side of Fig. 3 represents the precision-recall curve of the machine learning classifiers. The curve shows XGBoost holding the most AUC which is 90% and the second highest area is occupied by Logistic Regression which is 89%. Random Forest and Support vector machine (SVM) occupied 88% and 86% area respectively, Naïve Bayes has the least area at 77%. The performance of a classifier at various threshold levels is summarized by the area under the ROC curve (AUC-ROC). Similarly, to XGBoost, the Logistic Regression curve has the largest area under the curve of 93%. Naïve Bayes has the smallest AUC of 85%, while Random Forest and SVM occupied 92% area. As a result, XGBoost and Logistic Regression outperformed Naïve Bayes according to AUC (see Fig. 3).

Fig. 4 shows the confusion matrices for the experimented classifiers. Among all the traditional methods, Logistic Regression had the least number of inaccurate predictions of 23 values and the largest number of correct predictions of 340 values. The Naïve Bayes classifier accurately predicted 311 values and wrongly predicted 52 values. 336 values were properly predicted by Random Forest, whereas 27 values were wrongly predicted. XGBoost and SVM predicted values of 24 and 29 wrongly, whereas 339 and 334 numbers were accurately predicted. Thus, out of all the predictors, the logistic regression was the most accurate (see Fig. 4).

The classification matrix of Table 4, uses precision, recall, F1-score, and overall accuracy measures to evaluate the experimented classification models. The Non-Dropout and Dropout classes' precision values for Naïve Bayes were 0.85 and 0.87, while the corresponding recall values were 0.91 and 0.78. Logistic regression showed strong recall (0.96 and 0.91) and precision (0.93 and 0.94) for both classes, resulting in overall accuracy of 94%. Significant results were also shown by Random Forest, XGBoost, and SVM, with precision, recall, and F1-score values above 0.90 for both classes (see Table 4).

4.2. Performance of the neural network algorithm

Between Artificial Neural Network (ANN) and Multi-Layer Perceptron (MLP), ANN has obtained the highest accuracy with 93% as well as same amount of precision, recall and F1-score. On the other hand, MLP obtained 1% less accuracy, precision, recall, and F1-score than the ANN which is 92% (See Table 5).

The Artificial Neural Network (ANN) and Multilayer Perceptron (MLP) neural network classifiers' performance errors are shown in

Table 2
Performance of the experimented machine learning classifiers.

Algorithm Name	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Naïve Bayes	86	85	85	85
Logistic Regression	94	94	94	94
Random Forest	93	93	93	93
XGBoost	93	93	93	93
SVM	92	92	92	92

Table 3
The error rate of the experimented machine learning classifiers.

Algorithm Name	Mean Absolute Error (%)	Root Mean Squared Error (%)	Relative Absolute Error (%)
Naïve Bayes	14.93	38.64	31.33
Logistic Regression	6.31	27.51	16.54
Random Forest	7.58	28.96	18.12
XGBoost	7.20	28.33	19.31
SVM	8.37	31.73	21.88

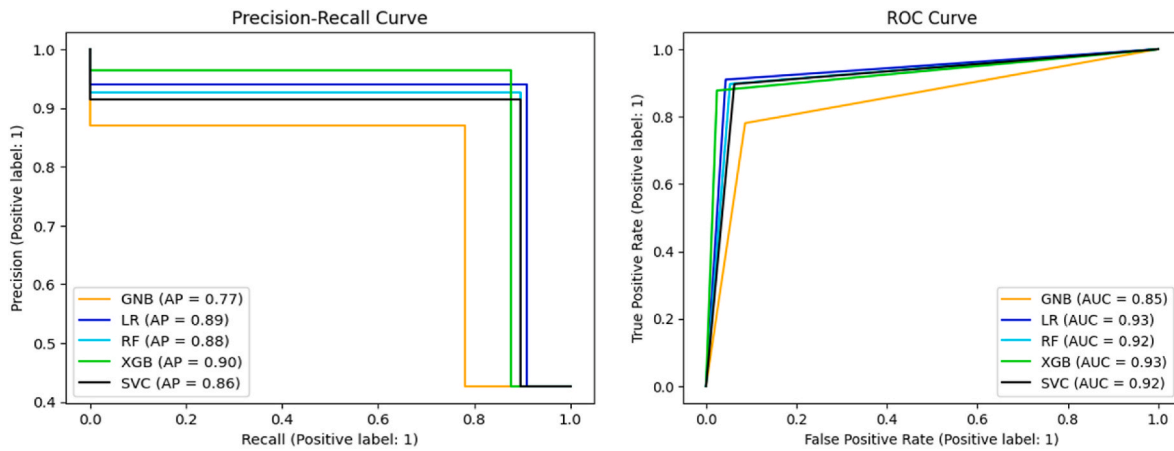


Fig. 3. Precision-Recall curve and ROC curve of the experimented machine learning classifier.

Table 6. ANN performs better than MLP in terms of Mean Absolute Error (MAE), with an error rate of 7.19% instead of 8.15%. On the other hand, both models perform similarly on the Root Mean Squared Error (RMSE), with values for ANN and MLP being 27.1% and 28.92%, respectively. Compared to ANN, MLP has a Relative Absolute Error (RAE) of 19.39%, while ANN has 16.0%. Rationally, MLP is a less error generator model (see Table 6).

The precision-recall (PR) curve illustrates the model's effectiveness. Hence, ANN is more effective compared to MLP since ANN has a larger area under the curve. ANN has 86% and MLP 87% area range in PR curve. Thus, the ANN model outperforms the MLP model by 1%. However, the ROC curve indicates that the MLP performs similar to ANN. Under the ROC curve, the ANN has 92% of the occupied area and the MLP has 92% (see Fig. 5).

The correctly and wrongly predicted values are represented by the confusion matrices of ANN and MLP. 336 and 334 values were successfully predicted by ANN and MLP, while 27 and 29 values were wrongly predicted. Despite a 1% accuracy difference, both the ANN and MLP have slightly differences in predictions (see Fig. 6).

Accuracy, precision, recall, and F1-score measures were used to evaluate the classification performance of the Artificial Neural Network (ANN) and Multilayer Perceptron (MLP) models. About the ANN model, the Non-Dropout and Dropout classes had precision values of 0.91 and 0.89, respectively, demonstrating a higher accuracy in detecting true positive values. Recall and F1-score have similar performance of ANN was 0.93 and 0.90, respectively. Both classes' F1-scores were 0.89 and 0.92, indicating a balanced performance. For both classes, the MLP model also showed strong performance, with precision, recall, and F1-score values in the 0.90–0.92 range (see Table 7).

4.3. Performance of the proposed HLRNN model

The accuracy, precision, recall, and F1-score metrics were used to evaluate the performance of the proposed model, HLRNN (Hybrid Logistic Regression and Neural Network). With an overall accuracy of 96%, the model demonstrated its efficiency in forecasting the prediction of student dropout. Consistently high at 0.96, the precision, recall, and F1-score values for the 'Non-Dropout' and 'Dropout' classes highlighted the models' well-balanced prediction skills (see Table 8).

HLRNN model showed, MAE and RMSE values of 9.0%, which represent the average magnitude of wrong predictions of the proposed model. A proportional measure of prediction accuracy about the actual values is indicated by the RAE of 10% for the HLRNN model (see Table 9).

The precision-recall curve illustrates the model's efficiency of the HLRNN model, the Voting ensemble has a larger area under the curve of 87%. Conversely, under the ROC curve, the HLRNN has 92% of the

occupied area. The AUC of the ROC curve decides whether the proposed model is more stable and balanced (see Fig. 7).

The confusion matrices represent the correctly and wrongly predicted values. The HLRNN model predicted 683 values successfully while 43 values were wrongly predicted (see Fig. 8).

The evaluation of HLRNN models considered precision, recall, and F1-score metrics for the 'Non-Dropout' and 'Dropout' classes. The accuracy of HLRNN in accurately identifying students who will not drop out was highlighted by its exceptionally high precision of 0.95 for 'Non-Dropout' and 0.96 for 'Dropout'. Recall scores of 0.90 for Dropout and 0.98 for Non-Dropout indicate that the model can effectively capture most cases of each class. A balanced performance is shown by the comparable F1-scores of 0.96 for Non-Dropout and 0.93 for Dropout (see Table 10).

5. Model validation: does the HLRNN model perform better?

This section presents and analyzes a performance comparison of all the experimented models. The x-axis of Fig. 9 represents different experimented models, while the y-axis indicates the accuracy percentage of these models. Naïve Bayes shows a poor accuracy of 86%. Due to its feature independence assumption, Naïve Bayes leads to lower performance. Logistic Regression, achieving the second-highest accuracy of 94%, is well-suited for binary classification problems. Random Forest and XGBoost both achieved an accuracy of 93%. SVM performed slightly less than the Random Forest and XGBoost model with 92% accuracy. Both ANN and MLP exhibit accuracies of 93% and 92%, respectively.

The proposed HLRNN model outperforms all other models with an accuracy of 96%, demonstrating its ability to capture temporal dependencies and complex patterns effectively, which are crucial for predicting dropout. The proposed model leverages the strengths of Logistic Regression in capturing linear relationships between features and targets, while also utilizing the ability of Artificial Neural Networks (ANNs) to identify non-linear relationships and complex patterns in the data (Goodfellow et al., 2016; Hosmer et al., 2013; Haykin, 1998). This approach effectively mitigates the limitations of each method. This model also outperformed its parent model Logistic Regression and ANN with 2% and 3% accuracy respectively (see Fig. 8).

6. Dropout risk factor analysis using SHAP and LIME

The SHapley Additive exPlanations (SHAP) is a method for interpreting Shapley values from cooperative game theory (Liu et al., 2024). In SHAP, global interpretations explain how general features impact overall predictions, while local interpretations clarify the specific reasons behind individual predictions for each instance. Fig. 10, SHAP value (x-axis) represents the impact of the feature on prediction. The

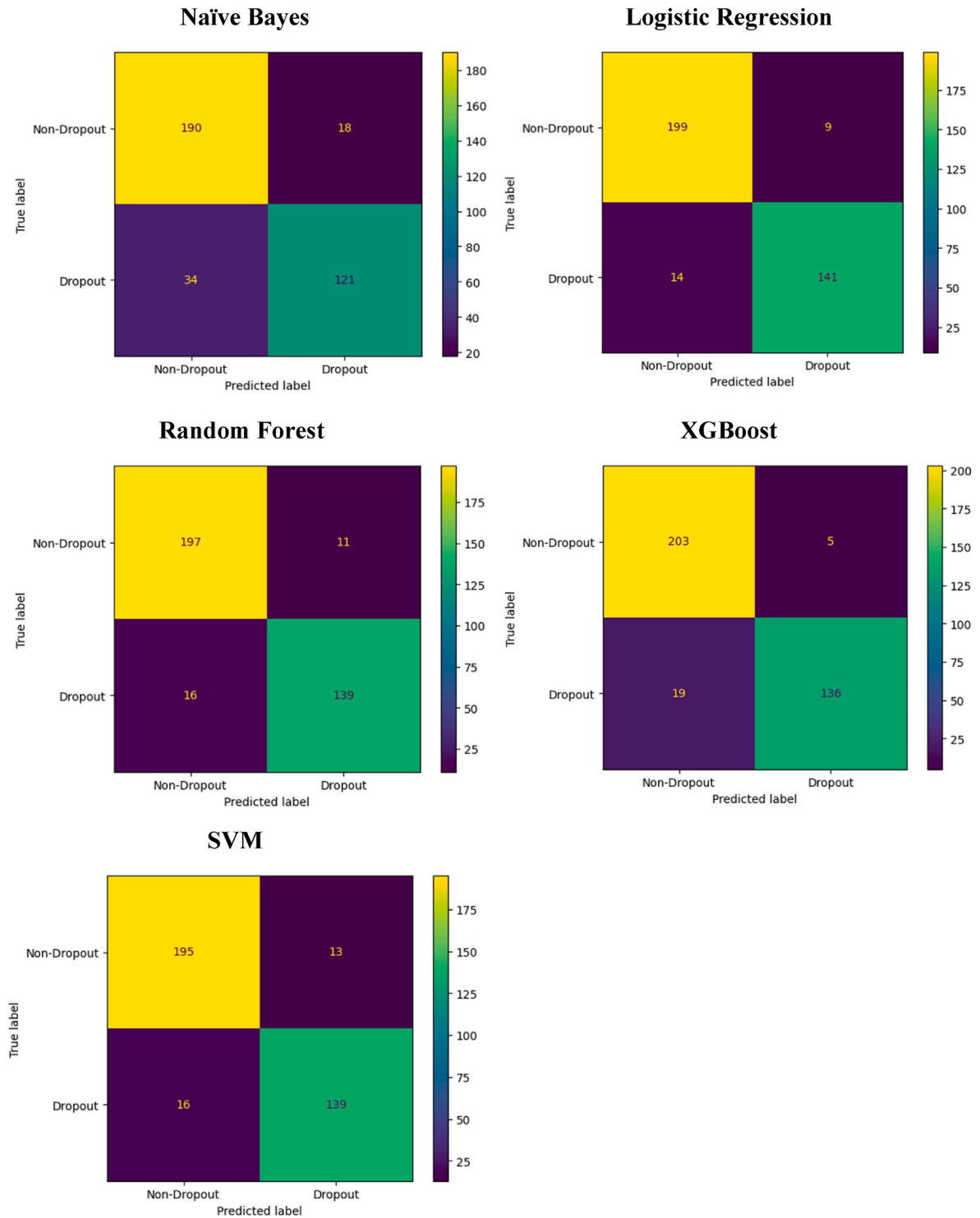


Fig. 4. Confusion matrix of the experimented machine learning classifiers.

positive values drive the prediction toward 'Dropout', while the negative values drive toward 'Non-Dropout'. The red dots represent high feature values, while the blue dots represent low feature values.

This analysis identified key features associated with student dropout or retention. It represents that the 'Gender' (16) feature is a significant risk factor for dropouts, with certain gender categories associated with higher risks. The 'Displaced' (12) students also showed a significantly higher risk of dropping out. Similarly, students with outstanding 'debts' (14) were at a greater risk of dropping out. Furthermore, Parental influence was also significant, higher levels in the mother's and father's

occupation (9–10) also impacted student retention. Dropout rates are influenced by 'nationality' (7), with certain nationalities having a higher risk of dropping out. In addition, students with 'educational special needs' (13) are more likely to drop out, highlighting the need for targeted support. There were complex impacts of features like 'Tuition Fees Up to Date' (15) and 'Scholarship Holder' (17), affecting both dropout and non-dropout predictions depending on individual values. However, broader economic indicators such as unemployment rates and GDP were relatively less significant (see Fig. 10).

Local Interpretable Model-agnostic Explanations (LIME) explain

Table 4

Classification report of the experimented machine learning classifiers.

Class Name	Precision	Recall	F1-score	support
Naïve Bayes				
Non-Dropout	0.85	0.91	0.88	208
Dropout	0.87	0.78	0.82	155
accuracy			0.86	363
macro avg	0.86	0.85	0.85	363
Weighted avg	0.86	0.86	0.86	363
Logistic Regression				
Non-Dropout	0.93	0.96	0.95	208
Dropout	0.94	0.91	0.92	155
accuracy			0.94	363
macro avg	0.94	0.93	0.93	363
Weighted avg	0.94	0.94	0.94	363
Random Forest				
Non-Dropout	0.92	0.95	0.94	208
Dropout	0.93	0.90	0.91	155
accuracy			0.93	363
macro avg	0.93	0.92	0.92	363
Weighted avg	0.93	0.93	0.93	363
XGBoost				
Non-Dropout	0.91	0.98	0.94	208
Dropout	0.96	0.88	0.92	155
accuracy			0.93	363
macro avg	0.94	0.93	0.93	363
Weighted avg	0.94	0.93	0.93	363
SVM				
Non-Dropout	0.92	0.94	0.93	208
Dropout	0.91	0.94	0.91	155
accuracy			0.92	363
macro avg	0.92	0.92	0.92	363
Weighted avg	0.92	0.92	0.92	363

Table 5

Performance of the experimented neural network classifiers.

Algorithm Name	Accuracy	Precision	Recall	F1 Score
ANN	0.93	0.93	0.93	0.93
MLP	0.92	0.92	0.92	0.92

Table 6

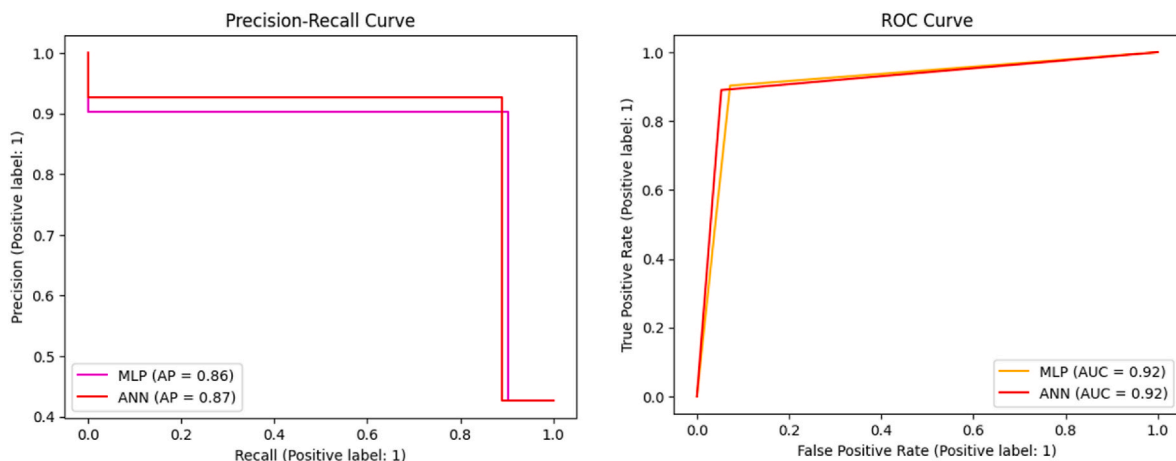
The Error rate of the experimented neural network classifiers.

Algorithm Name	Mean Absolute Error (%)	Root Mean Squared Error (%)	Relative Absolute Error (%)
ANN	7.19	27.1	16.0
MLP	8.15	28.92	19.39

individual predictions by approximating the model's behavior locally around the instance of interest. The positive features are the index of 'Non-dropout' risk where the negative features indicates 'Dropout' risk factor (see Table 11). By identifying which features contribute the most to a specific prediction, LIME identifies factors that influence the outcome.

This analysis represents several influential features related to student dropout risk. Lower age at enrollment (≤ 0.16) and 'Daytime or evening attendance' indicate that certain attendance types are linked to lower dropout risks. The 'student's nationality' (> -0.57) positively influences the prediction towards Non-Dropout, while higher levels of the 'mother's qualification' have a positive impact on retention. Certain 'gender' categories ($0.28 < \text{Gender} \leq 0.60$) are associated with a lower likelihood of dropout, and specific occupations of the mother link to an increased likelihood of Non-Dropout. Additionally, higher values for the father's occupation (> 0.16) suggest a positive impact on retention. Holding a scholarship (> 0.08) is associated with a lower dropout risk, and being current with tuition fees positively affects the prediction of Non-Dropout. Moreover, being classified as 'displaced' (≤ -0.20) negatively impacts the prediction, increasing the dropout likelihood. Students with educational special needs show a slight increase in dropout risk, while specific 'marital statuses' (≤ -1.16) contribute negatively, indicating a higher risk of dropout. A low 'application mode' value (-0.01) drives the prediction towards Dropout, and lower qualifications of the father ($-0.49 < \text{Father's Qualification}$) are increased dropout risk. Finally, being a 'debtor' negatively impacts the prediction, suggesting a higher likelihood of dropout (see Table 11).

The combined SHAP and LIME feature analysis reveals that key factors influencing dropout predictions include 'gender', 'displacement status', and 'debt status', with specific gender categories, displaced students, and those with 'debts' at a higher risk of dropping out. Research studies by Marczuk and Strauss, (2023), Banaag et al. (2024), Contini and Zotti, (2022) has been identified 'gender', 'displacement status', and 'debt status' factors as a major risk factor for student dropout too. Parental background, particularly higher qualifications and certain occupations of both 'mother and father occupation', also plays a protective role in retention, as does 'scholarship status' and 'tuition fees', both of which reduce dropout likelihood by alleviating financial stress. Additionally, 'nationality', 'educational special needs', and 'marital status' significantly impact dropout, with certain groups requiring more support. Factors like 'attendance type' (day-time/evening) and 'age at enrollment' indicate that younger students and specific attendance categories have a lower dropout risk. In contrast, lower values in 'application mode' and 'father's qualification' are associated with increased dropout likelihood. Broader economic indicators such as 'the unemployment rate' and 'GDP' are less

**Fig. 5.** Precision-Recall curve and ROC curve of the experimented neural network classifiers.

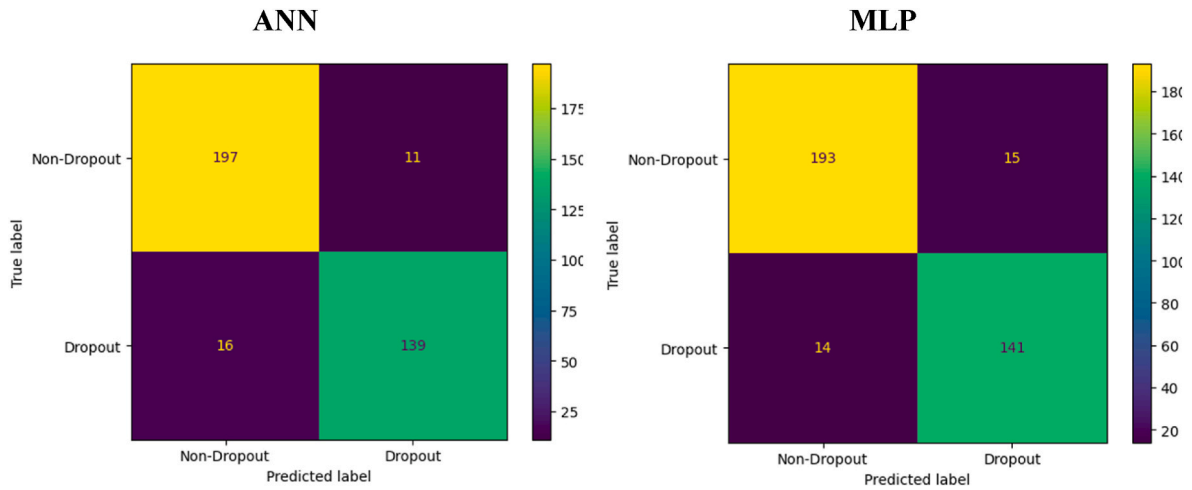


Fig. 6. Precision-Recall curve and ROC curve of the experimented neural network classifiers.

Table 7

Classification report of the experimented neural network classifier.

Class Name	Precision	Recall	F1-score	support
ANN				
Non-Dropout	0.93	0.93	0.93	208
Dropout	0.90	0.90	0.90	155
accuracy			0.92	363
macro avg	0.92	0.92	0.92	363
Weighted avg	0.92	0.92	0.92	363
MLP				
Non-Dropout	0.92	0.92	0.92	208
Dropout	0.90	0.91	0.91	155
accuracy			0.92	363
macro avg	0.90	0.90	0.90	363
Weighted avg	0.90	0.90	0.90	363

Table 8

Performance of the proposed HLRNN model.

Algorithm Name	Accuracy	Precision	Recall	F1 Score
HLRNN	0.96	0.96	0.96	0.96

Table 9

The error rate of the proposed HLRNN model.

Algorithm Name	Mean Absolute Error (%)	Root Mean Squared Error (%)	Relative Absolute Error (%)
HLRNN	4.0	4.0	10

influential, suggesting that personal and immediate socio-economic factors are more predictive of dropout. This analysis underscores the importance of targeted interventions for at-risk groups based on their unique profiles.

7. Discussion

The results of this study highlight the robustness and effectiveness of the proposed HLRNN model, achieving 96% accuracy, which is a significant improvement over its parent models and existing approaches. The HLRNN performed as an accurate and precise hybrid classifier achieving mentionable accuracy that removes the standalone model limitations. The 'SMOTE' was widely used to handle the data imbalance problem of the student dropout dataset it adds noise while balancing the classes which leads to less accurate detection of student dropout. This study presented the 'Upsampling' technique without adding any noise

while balancing the classes and it also increased the accuracy of the HLRNNs' model. Existing literature solely provides attention to dropout prediction whereas the interpretability of the model to identify the risk factors was crucial. In this study, the use of SHAP and LIME with the HLRNN model provided deep insights into critical dropout factors, such as 'gender', 'displacement status', 'debt status', 'parental background' 'scholarship status', and 'tuition fees' which align with real-world challenges faced by students. These findings underscore the importance of focusing on socio-economic and demographic factors in addressing dropout issues. Additionally, the 'RFE-CV' technique also enhances the models' performance through most relatable features that can heavily affect student dropout. Overall, the results validate the proposed model as a practical and interpretable tool, enabling early and effective interventions in diverse educational settings.

Numerous research studies have used different datasets and machine learning approaches to predict and classify student dropout. [Rabelo and Zárate \(2024\)](#) developed an ensemble model using Logistic Regression, Neural Networks, and Decision Tree model which achieved 89.7% accuracy to predict dropout in higher education. [Talebi et al. \(2024\)](#) used advanced deep learning technique CNN-LSTM which achieved 94% accuracy to predict dropout in MOOC dataset. [Seo et al. \(2024\)](#) predicted student dropout risk on Korean university using Light Gradient Boosting Machine (LGBM) that achieved 87.4 f1-score. [Deb et al. \(2024\)](#) experimented different techniques where Random Forests achieved highest f1-score of 0.78. [Bravo et al. \(2023\)](#) used gradient-boosting (XGB) approaches to achieve an accuracy range of 69%–85%. [Cho et al. \(2023\)](#) used LightGBM to analyze student dropout and achieved an 84% accuracy rate based on the f1-score. [Beaulac et al. \(2019\)](#) used Multinomial Logistic Regression and Random Forests to investigate dropout prediction, with accuracies of 42.63% and 47.41%, respectively. Achieving a high accuracy of 95%, [Miguéis et al. \(2018\)](#) applied Random Forest. [Martinho et al. \(2013\)](#) used a fuzzy-ARTMAP neural Network and achieved an accuracy of 85%. Although the existing studies showed impressive performance on student dropout risk prediction tasks, still most of the studies tend to apply single machine learning models which ended up achieving less accuracy. However, the proposed HLRNN model of this research has the highest accuracy, outperforming all the existing research, obtaining 96% accuracy (see [Table 12](#)).

However, beyond accuracy, its practical applicability lies in its ability to provide explainable insights through SHAP and LIME, enabling educational institutions to identify the key factors driving student dropout. Unlike traditional machine learning models that act as "black boxes," the HLRNN model helps educators understand which features are most important for predicting dropout risk.

The SHAP and LIME models' common feature analysis indicates

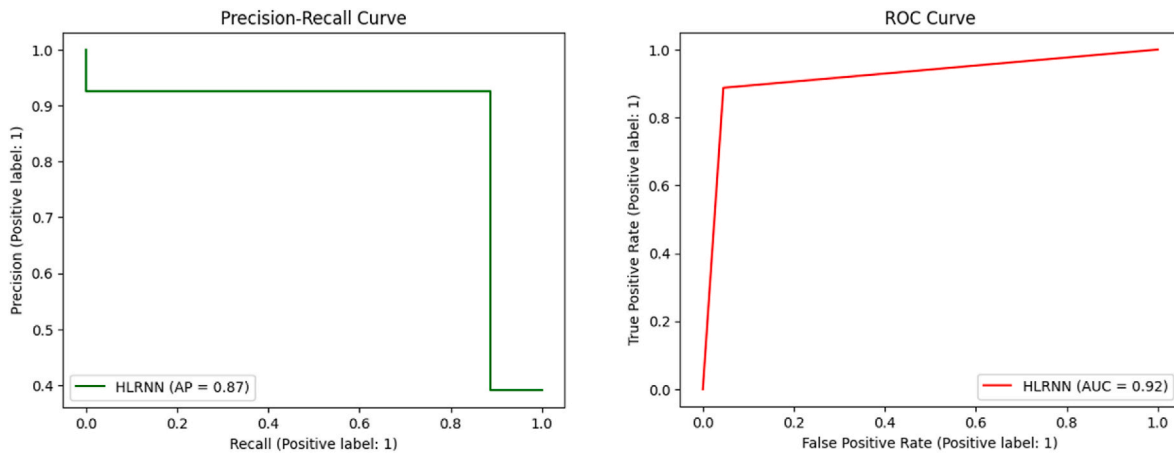


Fig. 7. Precision-Recall curve and ROC curve of the proposed HLRNN model.

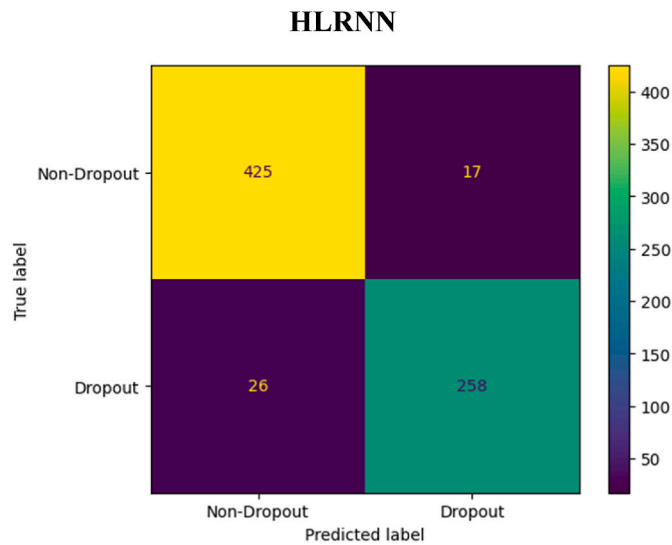


Fig. 8. Confusion matrix of the proposed HLRNN model.

Table 10

Classification report of the proposed HLRNN model.

Class Name	Precision	Recall	F1-score	support
Non-Dropout	0.95	0.98	0.96	442
Dropout	0.96	0.90	0.93	284
accuracy			0.96	726
macro avg	0.95	0.94	0.94	726
Weighted avg	0.95	0.95	0.95	726

specific gender categories, displaced students, and those with debts at a higher risk of dropping out. While other studies only propose an AI-based solution that contributes to an algorithmic approach (Bravo et al., 2023; Cho et al., 2023), this study addresses the risk factors directly. The educational institution and stakeholder can use their dataset to predict the features that influence student dropout. After training the HLRNN model the best weights will be saved as a *HLRNN.h5* file. Using that *.h5* file, a new prediction from a new dataset can easily be done. Students' regular performance and physical, mental, and financial data can be collected and analyzed using the *HLRNN.h5* file to detect whether the student is likely to dropout. The HLRNN model can make a probable decision about that particular student. Besides providing decisions about a student, the XAI approach can disclose the most influential features or reasons that uplift student dropout. The institution's

authority and stakeholders can work on the crucial risk factors to increase student retention.

Finally, we can announce that this study effectively addressed the research questions and research objectives. The proposed HLRNN model achieved 96% accuracy, outperforming Logistic Regression and ANN by 2% and 3% accuracy, respectively, demonstrating the strength of hybrid models in improving dropout predictions. The proposed hybrid model outperformed the existing models too. Regarding the second question, the Up-sampling techniques addressed class imbalance, enhancing model reliability across minority and majority classes. For the third question, SHAP and LIME identified critical dropout factors like 'gender', 'displacement status', 'debt status', 'parental background' 'scholarship status', and 'tuition fees' providing actionable insights for targeted interventions. Finally, in line with the fourth question, the study enables educators to implement data-driven strategies to reduce dropout rates and improve retention by using the *HLRNN.h5* file and XAI insights.

7.1. Implications of our findings

This study proposes a rigorous methodological approach. The methodology incorporates RFE-CV feature selection, HLRNN modeling, and Explainable AI (SHAP and LIME) to support data-informed decision-making in educational contexts as technological tools to empower educators and policymakers. This aligns with the constructivist educational theory of Technological Pedagogical Content Knowledge (TPACK). TPACK integrates AI-based techniques with pedagogical decision-making to enhance dropout prediction and classification in teaching.

In the context of Behaviorism, the proposed HLRNN model is behaviorist since it observes student data related (e.g. performance, demographic features, gender, etc.) to predict dropout risks with 96% accuracy. The HLRNN model rewards or penalizes the importance of features during training, similar to behaviorist principles of reinforcement.

The UDL and SDT ensure equality in access to education and handling external challenges. By analyzing the most important factors (financial factors, gender) using XAI, the stakeholders and educational institutions can support the students who intend to dropout due to unavoidable causes. This approach helps to provide basic human rights to be educated to all students equally.

7.2. Future research

In future work, we aim to expand the scope of this research by applying the proposed HLRNN model to new datasets across diverse educational settings to evaluate its adaptability and generalizability.

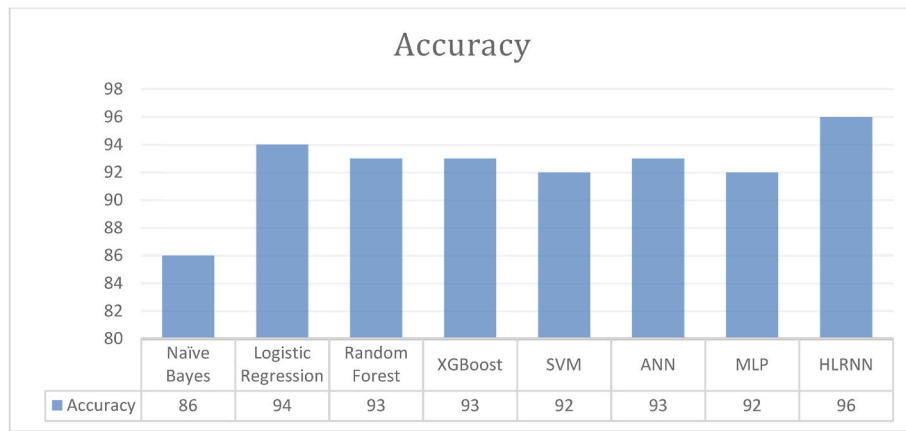


Fig. 9. Performance comparison of the experimented models.

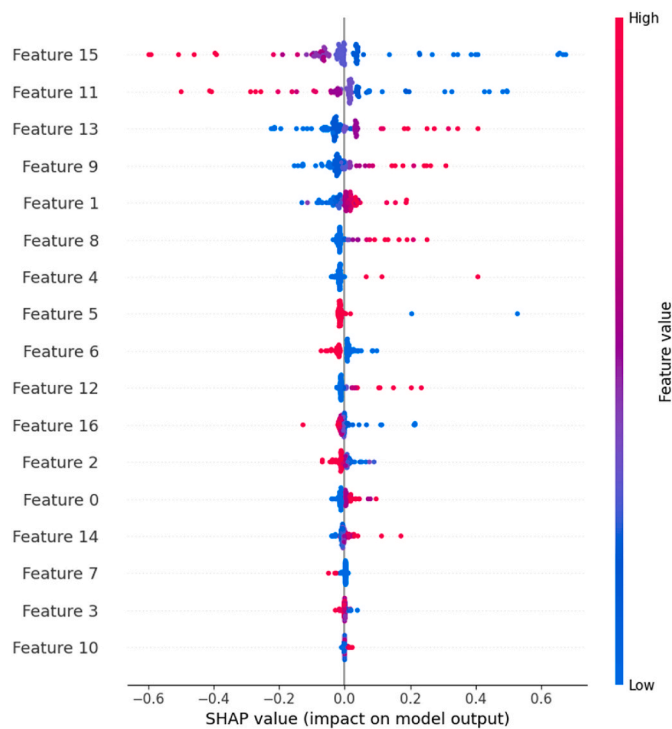


Fig. 10. SHAP summary plot for dropout risk factor analysis.

This will allow us to explore the model's effectiveness in varying cultural, institutional, and academic contexts, further strengthening its applicability to real-world scenarios.

Additionally, as the current study developed a hybrid approach integrating machine learning and neural network-based techniques, future research will investigate advanced architectures such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Vision Transformers (ViTs). These models have demonstrated strong capabilities in capturing complex patterns, modeling sequential dependencies, and handling large-scale data, making them promising candidates for enhancing the robustness and accuracy of dropout prediction.

We also plan to develop a comprehensive Decision Support System (DSS) that incorporates the HLRNN model and other advanced architectures to provide educators, policymakers, and institutions with a user-friendly tool for identifying at-risk students. This system will integrate real-time data inputs, predictive analytics, and actionable insights using Explainable AI techniques like SHAP and LIME. The DSS will

Table 11

LIME feature importance table for dropout analysis.

Feature	Value
Age at enrollment	0.02
Daytime/evening attendance	−0.40
Nationality	1.77
Mother's qualification	−0.30
Gender	0.60
Mother's occupation	−0.47
Father's occupation	0.47
Scholarship holder	1.15
Tuition fees up to date	−0.40
Displaced	−0.20
Educational special needs	−0.28
Marital status	−1.16
Application mode	−0.01
Father's qualification	−0.10
Course	0.59
Application order	0.80
Debtor	−0.09
Previous qualification	0.45

Table 12

Performance comparison with the existing research based on the context of dropout risk prediction.

Authors' Name	Techniques	Accuracy
Rabelo and Zárate (2024)	Ensemble technique	89.7%
Talebi et al. (2024)	CNN-LSTM	94%
Seo et al. (2024)	Light Gradient Boosting Machine (LGBM)	87.4 (f1-score)
Deb et al. (2024)	KNN, Random Forests, Logistic Regression	0.75 (f1-score) 0.78 (f1-score) 0.75 (f1-score)
Bravo et al. (2023)	Gradient-Boosting (XGB)	69%–85%
Cho et al. (2023)	LightGBM (Light Gradient Boosting Machine)	84% (f1-score)
Beaulac et al. (2019)	Random forests	47.41%
	Multinomial logistic regression	42.63%
Miguéis et al. (2018)	Random Forest	95%
Martinho et al. (2013)	Fuzzy-ARTMAP Neural Network	85%
This study (2024)	Proposed HLRNN model	96%

enable timely interventions to support students and improve retention rates.

Lastly, we will investigate the ethical and fairness implications of predictive models in education, ensuring that the developed framework

minimizes biases and promotes equity in decision-making. These advancements aim to create an inclusive and adaptive system that supports diverse educational environments while addressing the global challenge of reducing student dropout rates.

9. Conclusion

This study presents a novel AI-based approach for predicting student dropout, employing a Hybrid Logistic Regression and Neural Network model (HLRNN) with 96% accuracy and dropout risk analysis using the SHAP and LIME model. This approach presents the importance of machine learning techniques to identify at-risk students early, enabling timely interventions to enhance student retention and academic success. The findings of this study, the importance of predictive models in diverse educational environments, contributing to the broader goal of reducing student dropout rates and improving educational outcomes globally. This approach is implemented in such a way that aligns with important educational theories and pedagogies such as 'TPACK', 'Behaviorism', 'UDL', and 'SDT' which offers a reliable solution for student dropout risk.

We will apply our HLRNN model to new datasets, further investigating its adaptability and generalizability in various education settings. As the proposed study developed a hybrid approach involving machine learning and neural network-based techniques, advanced models such as CNN, LSTM networks, and Vision Transformers should also be researched further. These models have shown promise in the modeling of complex patterns and sequential dependencies and could further improve the robustness and accuracy of the student dropout prediction. We intend to experiment with some of these newer architectures in the hope of coming up with a more inclusive and adaptive framework for the analysis of student dropout risks.

CRedit authorship contribution statement

Sumaya Mustafa: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Yousuf Rayhan Emon:** Writing – review & editing, Writing – original draft. **Sajib Bin Mamun:** Writing – review & editing, Validation. **Shabnur Anonna Akhy:** Writing – original draft. **Md Taimur Ahad:** Validation, Supervision.

Contribution

- Firstly, the imbalanced dataset problem has been assessed using the 'Upsampling' technique. Here, the minority class 'Dropout' has been adjusted with the majority class 'Non-dropout'.
- A feature selection method named 'RFE-CV' has been applied to select the most significant features. The RFE feature selection method has been applied through a cross-validation technique to select the most robust features.
- In this study, a thorough experiment involving five machine learning models and two neural network models was conducted. Following the experiment, the highest-performing Logistic Regression (LR) and Artificial Neural Network (ANN) models were combined to create the proposed HLRNN model.
- The proposed HLRNN model can predict student dropout with 96% accuracy. This model outperformed its parent models, Logistic Regression and ANN, with 2% and 3% accuracy, respectively. This model also outperformed other existing AI-based student dropout prediction models.
- The HLRNN model developed in this study demonstrates a 96% accuracy in predicting student dropout risk. By analyzing key indicators within the collected data, the *HLRNN.h5* file can provide a reliable probability score, indicating whether a student is at high risk of dropping out or likely to continue. This prediction capability enables educators and institutions to make proactive, data-driven

decisions, allowing them to implement timely interventions for at-risk students.

- The Explainable AI model named SHAP and LIME has been used on the HLRNN model to analyze features that are influencing student dropout risk factors. While executing two XAI approaches, the 'gender', 'displacement status', 'debt status', 'parental background', 'scholarship status', and 'tuition fees' features indicate higher dropout risk. Incorporating two XAI model analyses and selecting the most common features as dropout risk factors makes the insight stronger.
- Our approach is implemented in such a way that aligns with important educational theories and pedagogies such as 'TPACK', 'Behaviorism', 'UDL', and 'SDT'. Based on these theories and pedagogies, policymakers and educators can offer reliable support to the students to attain their basic human rights.
- This study provides a practical framework for educational institutions to implement machine learning techniques in dropout prediction. The proposed machine learning approach is designed to be scalable and adaptable to different educational contexts, offering a valuable tool for policymakers and educators.

Data availability

The dataset is available in the Kaggle Repository. Dataset Link: <https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention>.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sumaya Muatofa

List of Acronyms

HLRNN:	Hybrid Logistic Regression and Neural Network
SHAP	SHapley Additive exPlanations
LIME	Local Interpretable Model-agnostic Explanations
SVM	Support Vector Machine
XAI	Explainable AI
RFE-CV	Recursive Feature Elimination – Cross Validation
ANN	Artificial Neural Network
LR	Logistic Regression
MLP	Multilayer Perceptron
TPACK	Technological Pedagogical Content Knowledge
UDL:	Universal Design for Learning
SDT	Self-Determination Theory

References

- Ahmad Tarmizi, S. S., Mutalib, S., Abdul Hamid, N. H., Abdul-Rahman, S., & Md Ab Malik, A. (2019). A case study on student attrition prediction in higher education using data mining techniques. *Soft Computing in Data Science: 5th International Conference, SCDS 2019, Iizuka, Japan, August 28–29, 2019, Proceedings*, 5, 181–192. Springer Singapore.
- Banaag, R., Sumodevilla, J. L., & Potane, J. (2024). Factors affecting student drop out behavior: A systematic review. *International Journal of Educational Management and Innovation*, 5(1), 53–70.
- Barr, J. R., Sobel, M., & Thatcher, T. (2022). Upsampling, a comparative study with new ideas. In *2022 IEEE 16th international conference on semantic computing (ICSC)* (pp. 318–321). IEEE.
- Beaulac, C., & Rosenthal, J. S. (2019). Predicting university students' academic success and major using random forests. *Research in Higher Education*, 60, 1048–1064.
- Bravo, D. P., Alves, M. A. Z., Ensina, L. A., & de Oliveira, L. E. S. (2023). Evaluating strategies to predict student dropout of a bachelor's degree in computer science. In *Anais do XI symposium on Knowledge discovery, mining and learning* (pp. 1–8). SBC.
- Cho, C. H., Yu, Y. W., & Kim, H. G. (2023). A study on dropout prediction for university students using machine learning. *Applied Sciences*, 13(21), Article 12004.
- Contini, D., & Zotti, R. (2022). Do financial conditions play a role in university dropout? New evidence from administrative data. In *Teaching, research and academic careers:*

- An analysis of the interrelations and impacts (pp. 39–70). Cham: Springer International Publishing.
- Deb, S., Sammy, M. S. R., Tusher, A. N., Sakib, M. R. S., Hasan, M. F., & Aunik, A. I. (2024). Predicting student dropout: A machine learning approach. In *2024 15th international conference on computing communication and networking technologies (ICCCNT)* (pp. 1–7). IEEE.
- Garg, A., Lilhore, U. K., Ghosh, P., Prasad, D., & Simaiya, S. (2021). Machine learning-based model for prediction of student's performance in higher education. In *2021 8th international conference on signal processing and integrated networks (SPIN)* (pp. 162–168). IEEE.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Ifenthaler, D., & Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research & Development*, 68, 1961–1990.
- Kim, S., Choi, E., Jun, Y. K., & Lee, S. (2023). Student dropout prediction for university with high precision and recall. *Applied Sciences*, 13(10), 6275.
- Klitzke, M., & Carvalhaes, F. (2023). Student dropout in a Brazilian public university: A survival analysis. *Educação em Revista*, 39, Article e37576.
- Kostopoulos, G., Kotsiantis, S., Ragos, O., & Grapsa, T. N. (2017). Early dropout prediction in distance higher education using active learning. In *2017 8th international conference on information, intelligence, systems & applications (IISA)* (pp. 1–6). IEEE.
- Latif, G., Alghazo, R., Pilotti, M. A., & Brahim, G. B. (2021). Identifying "at-risk" students: An AI-based prediction approach. *International Journal of Computing and Digital System*.
- Marczuk, A., & Strauss, S. (2023). Does context matter? The gendered impact of study conditions on dropout intentions from higher education. *Zeitschrift für Erziehungswissenschaft*, 26(5), 1349–1371.
- Martinho, V. R. D. C., Nunes, C., & Minussi, C. R. (2013). An intelligent system for prediction of school dropout risk group in higher education classroom based on artificial neural networks. In *2013 IEEE 25th international conference on tools with artificial intelligence* (pp. 159–166). Ieee.
- Martins, M. V., Tolledo, D., Machado, J., Baptista, L. M., & Realinho, V. (2021). Early prediction of student's performance in higher education: A case study. *Trends and Applications in Information Systems and Technologies*, 1 9, 166–175. Springer International Publishing.
- Masood, S. W., Gogoi, M., & Begum, S. A. (2024). Optimised SMOTE-based imbalanced learning for student dropout prediction. *Arabian Journal for Science and Engineering*, 1–15.
- Mduma, N. (2023). Data balancing techniques for predicting student dropout using machine learning. *Data*, 8(3), 49.
- Mduma, N., Kalegele, K., & Machuve, D. (2019). A survey of machine learning approaches and techniques for student dropout prediction. *Data Science Journal*, 18, 14–14.
- Meijer, E., Cleiren, M. P., Dusseldorp, E., Buurman, V. J., Hogervorst, R. M., & Heiser, W. J. (2019). Cross-validated prediction of academic performance of first-year university students: Identifying risk factors in a nonselective environment. *Educational Measurement: Issues and Practice*, 38(1), 36–47.
- Miguéis, V. L., Freitas, A., Garcia, P. J., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, 115, 36–51.
- Mulyani, E., Hidayah, I., & Fauziati, S. (2019). Dropout prediction optimization through smote and ensemble learning. In *2019 international seminar on research of information technology and intelligent systems (ISRITI)* (pp. 516–521). IEEE.
- Ortiz-Lozano, J. M., Rua-Vieites, A., Bilbao-Calabuig, P., & Casadesús-Fa, M. (2020). University student retention: Best time and data to identify undergraduate students at risk of dropout. *Innovations in Education & Teaching International*.
- Pecuchova, J., & Drlik, M. (2023). Predicting students at risk of early dropping out from course using ensemble classification methods. *Procedia Computer Science*, 225, 3223–3232.
- Poudyal, S., Mohammadi-Aragh, M. J., & Ball, J. E. (2020). Data mining approach for determining student attention pattern. In *2020 IEEE frontiers in education conference (FIE)* (pp. 1–8). IEEE.
- Prasanth, A., & Alqahtani, H. (2023). Predictive modeling of student behavior for early dropout detection in universities using machine learning techniques. In *2023 IEEE 8th international conference on engineering technologies and applied sciences (ICETAS)* (pp. 1–5). IEEE.
- Rabelo, A. M., & Zárate, L. E. (2024). *A model for predicting dropout of higher education students*. Data Science and Management.
- Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review. *Applied sciences*, 10(3), 1042.
- Seo, E. Y., Yang, J., Lee, J. E., & So, G. (2024). Predictive modelling of student dropout risk: Practical insights from a South Korean distance university. *Heliyon*, 10(11), Article e30960.
- Talebi, K., Torabi, Z., & Daneshpour, N. (2024). Ensemble models based on CNN and LSTM for dropout prediction in MOOC. *Expert Systems with Applications*, 235, Article 121187.
- Thammasiri, D., Delen, D., Meesad, P., & Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, 41(2), 321–330.
- Villar, A., & de Andrade, C. R. V. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study. *Discover Artificial Intelligence*, 4(1), 2.
- Zollanvari, A. (2023). Feature selection. In *Machine learning with Python: Theory and implementation* (pp. 283–302). Cham: Springer International Publishing.