

# Toward a New Hybrid System for the Prediction of Student Academic Performance

## Younes Zidani

M2S2I Laboratory, ENSET Mohammedia Hassan II University (UH2), Casablanca, Morocco  
zidani\_younes@yahoo.fr (corresponding author)

## Younes Zahrou

M2S2I Laboratory, ENSET Mohammedia Hassan II University (UH2), Casablanca, Morocco  
younes.zahrou@gmail.com

## Salah Nissabouri

M2S2I Laboratory, ENSET Mohammedia Hassan II University (UH2), Casablanca, Morocco  
SALAH.NISSABOURI@univh2c.ma

## Moulay El Houssine Ech-Chhibat

M2S2I Laboratory, ENSET Mohammedia Hassan II University (UH2), Casablanca, Morocco  
echchhibate@gmail.com

## Khalifa Mansouri

M2S2I Laboratory, ENSET Mohammedia Hassan II University (UH2), Casablanca, Morocco  
khalifa.mansouri@enset-media.ac.ma

Received: 4 June 2025 | Revised: 5 July 2025, 27 July 2025, and 17 August 2025 | Accepted: 20 August 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.12563>

## ABSTRACT

Accurately predicting Student Academic Performance (SAP) is a complex task due to the interplay of cognitive, behavioral, and institutional factors. This paper presents a method for predicting SAP using a hybrid classification algorithm combining four complementary algorithms: Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANNs). We employ a dataset of 32,000 students from the Open University of the United Kingdom for training and testing our hybrid system. The dataset contains information on student registration, courses, assessments, and interactions with resources. The algorithm is implemented using MATLAB software. We evaluate our system on several standard datasets and compare its performance with that of individual algorithms. The proposed method achieves an accuracy of 95.35%, a precision of 97.62%, a recall of 96.33%, and an F1-score of 96.97%. We used McNemar's test to assess the statistical significance of the model's performance, which reveals a very low p-value, indicating the robustness of the results. This system will help teachers predict students' exam results and take proactive measures to prevent academic failure.

**Keywords-**Student Academic Performance (SAP); prediction; learning analytics; Machine Learning (ML); Random Forest (RF); Logistic Regression (LR); Artificial Neural Network (ANN); Support Vector Machine (SVM)

## I. INTRODUCTION

Students often encounter various personal and academic challenges, which can sometimes lead to issues such as dropping out or failing exams. Predicting academic performance can be a complex task because it depends on various factors such as study habits, motivation, intelligence, and external circumstances. Although it is challenging to accurately predict an individual student's academic

performance [1, 2], there are some common indicators and methods that can provide insights into student performance. Analyzing a student's past academic performance can offer valuable insights. For example, examining their previous grades, test scores, and class participation helps identify patterns or trends. If a student has consistently performed well or struggled in certain subjects, it can provide an indication of their future performance [3, 4]. Teachers who have worked closely with students can provide insights into their

performance. They can assess a student's work ethic, participation, and understanding of the subject matter. Teacher evaluations, combined with other data, can offer a more holistic prediction of academic performance. These tests measure a student's aptitude and knowledge in specific areas and can provide insight into their potential success in related subjects, especially at the secondary level [5].

Socioeconomic factors also provide additional context. Students from disadvantaged backgrounds may face challenges that affect their academic performance. It is important to account for such factors when predicting academic success [1, 6]. Although there are many factors that contribute to student success in a course, four main factors are particularly important:

- Resilience and engagement factors [7].
- Emotional intelligence [8].
- Attitudes and behavioral skills [9].
- Family background factors [10].

Many studies have been conducted to identify the critical factors influencing academic performance [11]. Authors in [12] revealed that the student's weekly homework score, weekly practice score, and weekly quiz score are important elements that impact the learners' final academic success. Additionally, individual factors such as prior knowledge, learning styles, and personal circumstances [13] can also influence student success. Recognizing and addressing these factors on an individual level can further support students in achieving success in their courses.

The proposed hybrid model addresses two critical gaps in existing ensemble methods for Student Academic Performance (SAP) prediction:

- Complementary algorithm integration: Unlike traditional ensembles (e.g., bagging/boosting with homogeneous base learners), our approach combines heterogeneous algorithms (Logistic Regression (LR) for linear patterns, Random Forest (RF) for non-linear relationships, Support Vector Machine (SVM) for high-dimensional separability, and Artificial Neural Network (ANN) for complex feature interactions) to exploit their complementary strengths. This diversity mitigates individual model biases (e.g., SVM's sensitivity to imbalanced data or LR's linearity constraint).
- Dynamic probability averaging: Prior hybrid models often use majority voting, which discards confidence scores. Our "soft voting" mechanism averages class probabilities from all four models, preserving uncertainty estimates and improving robustness to noisy data (e.g., mislabeled student records).

## II. REVIEW OF ACADEMIC PERFORMANCE PREDICTORS

Based on [14] and [15], Educational Data Mining (EDM) is a relatively new field that seeks hidden patterns in different educational situations. There are five basic types of EDM: prediction, clustering, relationship mining, distillation of data

for human judgment, and discovery using models [16]. In prediction, the aim is to create a system capable of determining a predicted (dependent) variable using a combination of predictor (independent) variables. There are mainly three types of prediction: classification, regression, and density estimation [17]. SAP prediction is an expanding area of study, using various approaches from Artificial Intelligence (AI) and Machine Learning (ML). These methodologies facilitate the early identification of at-risk students, enhance classroom supervision, and personalize learning courses. Authors in [3] applied learning management systems to forecast achievement levels and provided students who were expected to perform poorly with a digital learning intervention. Authors in [4] examined the Bidirectional Long Short-Term Memory (BiLSTM) network, which, when combined with an attention mechanism, can effectively forecast student performance (grades) using past data. By addressing current research challenges focused on sophisticated feature categorization and prediction, the authors employed the most advanced BiLSTM in conjunction with an attention mechanism system. Authors in [18] employed both traditional statistical analysis and neural network-based modeling for predicting student performance. Traditional statistical analyses were used to identify variables influencing students' performance.

Authors in [19] conducted a study to create accurate systems using ML to estimate the combined influence of school and teacher variables, as well as background and extracurricular factors, on student learning. In particular, the goal was to create precise ANN-based prediction systems. The study in [20] had two objectives: the first was to evaluate a methodical approach for using ANNs in higher education to forecast academic achievement; the second was to highlight the importance of several factors in academic performance. Authors in [21] employed RF, nearest neighbor, LR, SVM, Naïve Bayes (NB), and K-Nearest Neighbor (KNN) algorithms to estimate students' final exam grades. In [22], a Multilayer Perceptron (MLP) and Cumulative Grade Point Average (CGPA) were used to forecast the likelihood that students would be 'not at risk' or 'at risk' in their degree. Authors in [23] used an improved multi-label classifier that forecasts qualitative values for the impact of different criteria on student performance, based on a regression system, which improves SAP prediction accuracy of, as measured by future grades in various courses. Three dynamically weighted techniques, including collaborative filtering, Lasso linear regression, and fuzzy set rules, were used to forecast student performance. In [24], the authors established connections between expected student conduct in the LMS course and academic motivation, using students from all years at the Osijek Faculty of Education. SVM, Decision Tree (DT) and ANN were the three ML classifiers employed. The study in [25] aimed to evaluate the performance of deep neural network transfer learning in predicting student achievement in higher education.

A summary of widely used academic performance prediction systems, together with their reported gaps, limitations, and key performance metrics, is provided in Table I, which highlights the remaining challenges associated with these techniques [26-28].

TABLE I. REVIEW OF ACADEMIC PERFORMANCE PREDICTION SYSTEMS

System	Ref.	Gaps / Limitations	Model used	Dataset used	Performance metrics reported	Key findings
RF	[27]	Performance drops with large datasets; clustering becomes challenging	Parameters such as node centralities and similarities were used in supervised ML classifiers, namely Gradient Boosting and RF, for link prediction	Six real-world network datasets (Hamsterster, Facebook NIPS, Caltech, Ca-GrQc, DBLP, Epinions)	Accuracy, precision, recall, F1-score, Area Under the Curve (AUC), information gain	Combining node centralities with structural similarity features improves link prediction accuracy, with Gradient Boosting and RF yielding the highest performance across network types
LR	[29]	Captures only a small portion of relevant knowledge; many components not included in prerequisite description.	Cumulative knowledge-based regression models (CKRM) with different course-knowledge spaces for next-term grade prediction, using historical student course grade data and course relationships	Dataset from the College of Science and Engineering, University of Minnesota (Fall 2002–Spring 2015)	Root Mean Squared Error (RMSE), percentage of predicted grades	CKRM methods achieved statistically significant improvements; textual-based models (CKRMtext) identify hidden or informal knowledge components crucial for student performance
SVM	[30]	Large or imbalanced datasets can lead to biased predictions	Two-step classification: motivation classification dividing learners into three groups using an activity index, followed by grade classification using SVM with different kernels	Person-Course dataset AY2013 from HarvardX-MITx (10 courses with detailed learning activity records including video play and forum activities)	Accuracy, RMSE, R-square values	earners can be effectively classified into motivation-based categories; method improves grade prediction accuracy and detects surrogate exam-takers
Graph Regularized Robust Matrix Factorization (GRMF)	[31]	Moderate performance (~65.4% accuracy); insufficient for very complex or noisy environments	GRMF integrates two side-information graphs from student and course background data into robust low-rank matrix factorization using L1-norm, optimized via Majorization Minimization	MovieLens 100K, public image data, NPU educational dataset (1,325 students, 832 courses, 2013–2017)	RMSE, L1-norm Error, Peak Signal to Noise Ratio (PSNR), accuracy	GRMF combining robust matrix factorization with side-information graphs outperforms baseline methods, showing robustness across application domains
Collaborative filtering	[32]	Highly dependent on historical data and similarity between students; less effective for new or atypical students	Enhanced collaborative filtering with outlier elimination and GPA-based similarity filtering	48,741 course grades from 2,524 students at Istanbul Sehir University (2010–2015)	Mean Absolute Error (MAE), F-error, weighted average of MAE and F-error	Proposed enhancements improved prediction accuracy by 16%; user-based collaborative filtering outperforms item-based in accuracy

As shown in Table I, the previous methods require a large amount of data. A key question that arises is how to optimize the dataset analysis, especially when using deep learning and data mining. In addition, many researchers overlook how unbalanced data can reduce prediction accuracy.

Class imbalance is a common problem in ML where one class in the dataset has significantly more examples than the others. This can lead to biased models that perform poorly on the minority class, as they tend to favor the majority class. Imbalance handling methods are techniques used to mitigate this problem. Several methods exist, such as Adaptive Synthetic Sampling (ADASYN), Synthetic Minority Oversampling Technique (SMOTE), SMOTE combined with Edited Nearest Neighbor (SMOTE-ENN), etc. [33].

In our study, in which we develop a hybrid method for predicting SAP using LR, RF, SVM, and ANN, we used the SMOTE method, as it is the most used technique and offers the best results [34]. We further improved it by combining

SMOTE with a soft-voting ensemble, which helps maintain strong performance even in rare failure situations.

### III. METHOD

#### A. Dataset Analysis

##### 1) Dataset Collection

The dataset was obtained from the Open University Learning Analytics Dataset (OULAD) [35], which contains information on approximately 32,000 students. Figure 1 provides an overview of the OULAD database structure, showing three main types of data: module presentation (green), which includes course and assessment information; student activities (purple); and student's demographic details (yellow). The dataset encompasses information on courses, students, and their interactions within a Virtual Learning Environment (VLE) across seven specifically chosen modules. It includes a detailed collection from a relational database made up of linked data tables with unique identifiers and provided as several comma-separated values (.csv) files.

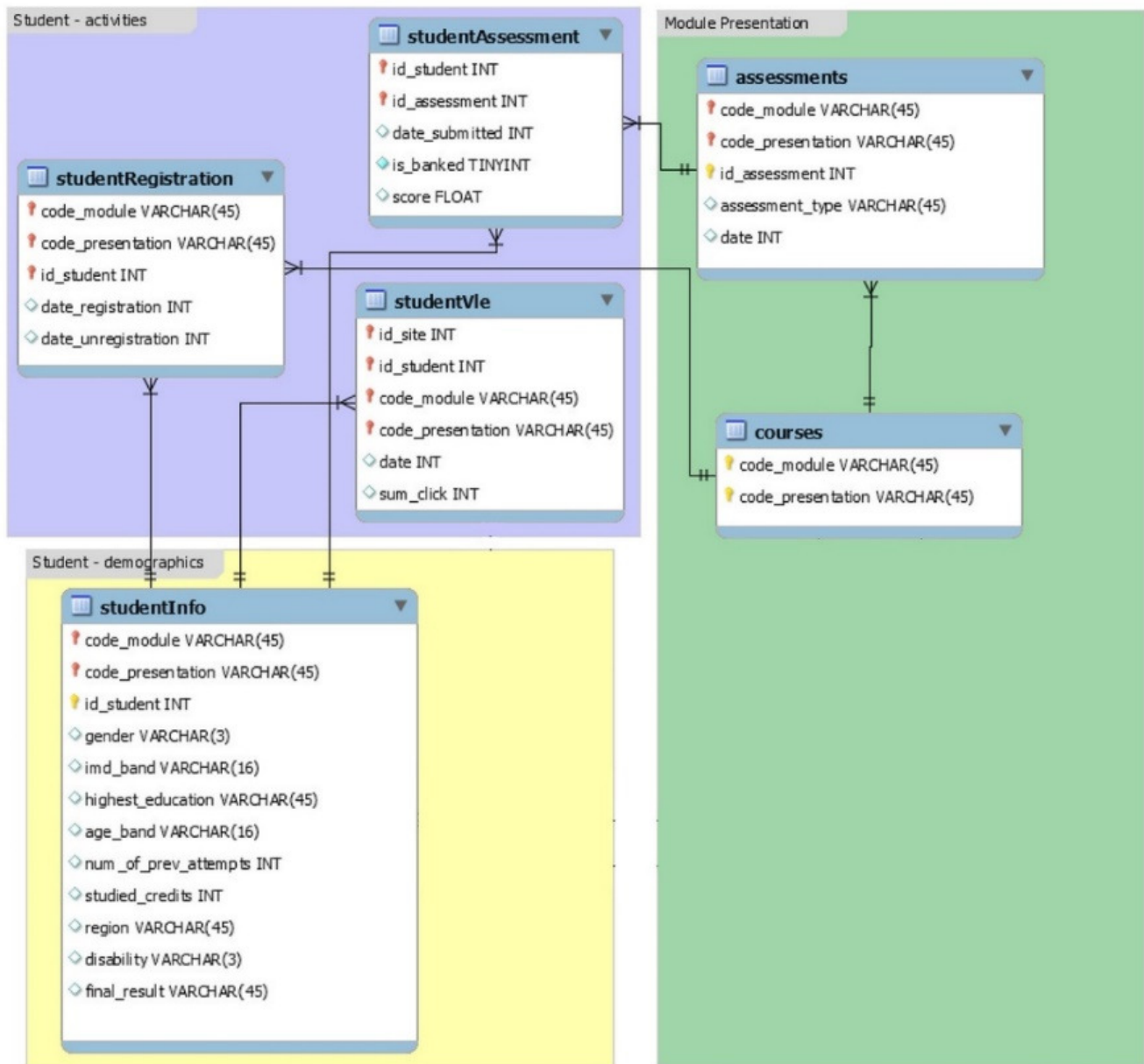


Fig. 1. Structure of the OULAD dataset.

To better understand the characteristics of the dataset and guide the modelling process, we conducted an Exploratory Data Analysis (EDA) [14, 36]. This analysis included a detailed examination of both numerical and categorical variables. We observed a wide range of values and variability in student engagement for numerical features such as the number of clicks, assessment scores, and studied credits. For example, the number of clicks ranged from 0 to over 70,000, whereas assessment scores were generally concentrated between 40 and 80. We also analyzed categorical variables, such as gender, region, age band, highest education, and disability status. Most students were between 26 and 45 years old, and the majority held either a higher education qualification or an A-level equivalent. Approximately 9% of

students reported having a disability. The distribution of the target variable (final result) revealed a class imbalance, with about 76% of students passing and 24% failing, which justified the use of SMOTE in the training phase. A correlation analysis showed that features like the number of clicks and assessment scores were moderately correlated with academic success, whereas previous attempts and studied credits had weaker associations. Finally, we examined the pattern of missing values: categorical variables with missing entries were filled with 'Unknown,' and numerical variables with missing values were set to zero, indicating no recorded activity. This EDA phase provided valuable insights and supported the design of our preprocessing and modeling strategies.

## 2) Inputs

The hybrid algorithm for predicting SAP takes as its input a set of detailed data on students and their academic backgrounds. As input, the system receives twelve main characteristics that describe each student. These twelve characteristics are 'code\_module,' 'code\_presentation,' 'gender,' 'region,' 'highest\_education,' 'imd\_band,' 'age\_band,' 'disability,' 'num\_of\_prev\_attempts,' 'studied\_credits,' 'sum\_click,' and 'score.'

## 3) Outputs

After processing, the algorithm produces several useful outputs. The main one is a simple binary prediction that indicates whether the student is likely to pass (value 1) or fail (value 0) their course, as shown in Figure 2. For each

prediction, the system also provides a precise probability score that quantifies the level of confidence in that prediction, expressed as a percentage.

Prediction: Success  
Probability: 71.45%

Fig. 2. Output example of the proposed hybrid student performance prediction system.

## B. Description and Architecture of the Hybrid System

Figure 3 shows the workflow of the proposed student performance prediction hybrid system. The hybrid model integrates four classifiers (LR, RF, SVM, and neural network) via soft voting.

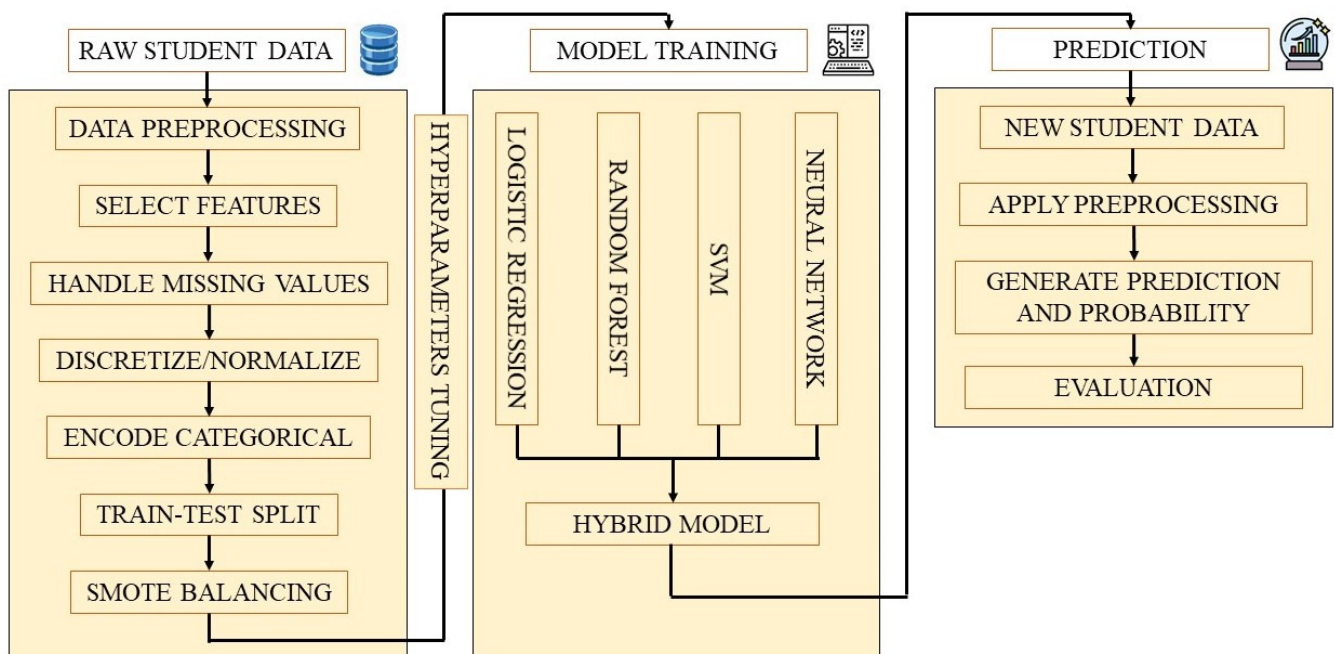


Fig. 3. Workflow of the proposed hybrid student performance prediction system.

The dataset underwent a thorough preprocessing pipeline to ensure data quality and model compatibility. First, missing values were addressed through strategic imputation: categorical variables such as education level and age band were filled with 'Unknown,' whereas numerical features, including assessment scores and interaction clicks, were set to zero, indicating no recorded activity. The target variable, final\_result, was converted into a binary classification task, mapping 'Pass' to 1 (success) and 'Fail' to 0 (failure). Next, feature engineering was applied to normalize and encode the data. Numerical variables, including previous attempts, studied credits, and engagement metrics, were standardized using StandardScaler to ensure consistent scaling. Categorical features, such as gender, region, and disability status were transformed via OneHotEncoding, expanding them into binary columns while avoiding multicollinearity by dropping the first category. To mitigate class imbalance, SMOTE was applied exclusively to the training set, generating synthetic samples for the minority class

(failures) while preserving the original distribution in the test set. This ensured balanced learning without artificially inflating model performance. Finally, the data were split into training (80%) and testing (20%) sets using stratified sampling, maintaining proportional representation of outcomes in both subsets. The preprocessing pipeline was rigorously validated to confirm no data leakage, proper scaling, and realistic synthetic sample generation. This systematic approach ensured robust model training while maintaining the integrity of the evaluation process.

The training phase is based on four separate algorithms trained in parallel. LR provides a simple but effective linear approach. RF captures non-linear relationships and is robust to outliers. The SVM employs a Radial Basis Function (RBF) kernel and is configured to return probability estimates. Finally, an MLP with a hidden layer of 50 neurons is used to model complex relationships in the data.

The voting classifier then combines these four systems using a "soft" strategy to form the hybrid system. We used the voting classifier because it outperforms the other models, such as bagging, boosting, and stacking, in terms of overall results [37, 38]. It combines the predictions of different models. The voting classifier can minimize the biases or difficulties of a particular model and provide more precise and trusted predictions using the collective decision-making of several models. It proposes different voting procedures such as hard voting, soft voting, and weighted voting, allowing customization according to specific tasks and model characteristics [39]. This sophisticated approach is not content with a simple majority vote but calculates the average of the probabilities predicted by each individual system. Each system contributes identically to the final prediction, with a given weight of 1 for each.

Various libraries are used, including scikit-learn (for ML systems and preprocessing), pandas (for data manipulation), matplotlib/seaborn (for visualization), imbalanced-learn (for handling class imbalance via SMOTE), and patool (for file extraction).

### C. Hyperparameter Tuning Process

Dependence on hyperparameters, which are configuration parameters governing the learning process, is a key factor influencing the accuracy of ML models. To improve model performance, the optimal hyperparameters must be identified. There are several algorithms for Hyperparameter Optimization (HPO) [40]. To optimize the hyperparameters of individual classifiers (LR, RF, SVM, and neural networks), we used grid search (GridSearchCV) in Scikit-learn with fivefold cross-validation. This method allows us to systematically evaluate combinations of hyperparameters and select those that maximize the F1-score.

During the prediction phase, the trained model uses new data to make predictions. In the case of a new user, it is in this substep that new information is provided for students for whom no academic outcome (pass/fail) is yet known. These are raw data, similar in structure to the data used for training but distinct and new to the model. For example, these could be data from a new cohort of students at the start of the semester, for whom we wish to anticipate results to enable early interventions if necessary. The hybrid model must rigorously apply the same preprocessing steps used during the "Data Preprocessing" phase of training to correctly interpret and process the "New Student Data." This includes:

- Selection of the same characteristics (module\_code, presentation\_code, gender, region, etc.).
- Handling of missing values in the same way (for example, filling 'imd\_band' with 'Unknown' and 'sum\_click' with 0).
- Normalization of numerical characteristics (num\_of\_prev\_attempts, studied\_credits, sum\_click, etc.) using the same parameters (mean and standard deviation) as those learned from the training data.

- Encoding of categorical variables (code\_module, gender, region, etc.) using the same mapping (OneHotEncoder) learned from the training set.

This consistency ensures that the new data are transformed into the exact format expected by the model, avoiding errors and ensuring reliable predictions.

Once the 'New Student Data' have been correctly preprocessed, they are submitted to the hybrid model, which has been trained beforehand. The model then uses its acquired knowledge to produce two types of results for each new student:

- Raw prediction: A binary result, such as 'Pass' (1) or 'Fail' (0).
- Associated probability: The model provides a probability that the student belongs to the predicted class. For example, if the model predicts 'Pass,' it might indicate a probability of 0.85 (85%) that the student will pass.

The final step is evaluation, in which metrics such as accuracy, precision, recall, and F1-score are calculated on the actual data, and the confusion matrix is generated to visualize classification errors.

### D. Mathematical Formulation

The final prediction combines the average probabilities:

$$P_{\text{hybrid}}(y|x) = \frac{1}{4} \sum_{i=1}^4 w_i p_i(y|x) \quad (1)$$

with:

- $M$ :  $\{m_{\text{LR}}, m_{\text{RF}}, m_{\text{SVM}}, m_{\text{NN}}\}$ , the set of models used in the hybrid system.
- $p_i$ : the predicted probability from model  $m_i$  that instance  $x$  belongs to class  $y$  (pass/fail).
- $w_i = 1$  (uniform weights for equal contribution).

This formula takes a weighted average of the predictions of four models. Each predicts a probability  $p_i(y|x)$  and its contribution is weighted by  $w_i$ . The result is a combined (hybrid) prediction that is more robust than that of a single model.

## IV. EVALUATION OF THE PROPOSED SYSTEM PERFORMANCE

The system's performance was assessed using a confusion matrix, classification accuracy, precision, recall, and F1-score.

### A. Confusion Matrix

Figure 4 shows the confusion matrix, a fundamental tool for evaluating classification performance. In the upper left corner, the value of 1202 indicates the true negatives. These are cases where the system correctly predicted "Failure" and the actual value was indeed "Failure." Next to it, the value of 104 represents the false positives, where the system predicted "Success" whereas the actual value was "Failure." In the bottom left, the number 163 corresponds to the false negatives, meaning the system predicted "Failure" when the actual value

was "Success." Finally, in the bottom right corner, the value of 4274 represents the true positives where the system correctly predicted "Success" and the actual value was also "Success."

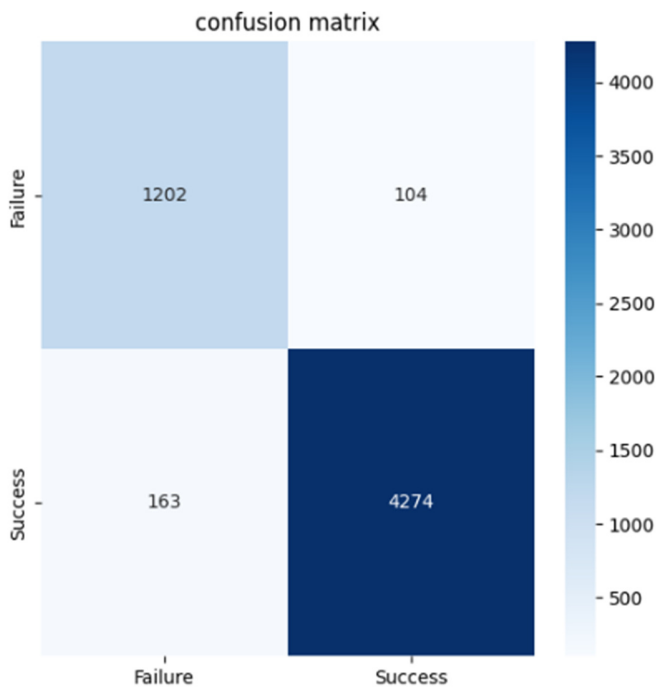


Fig. 4. Confusion matrix of the proposed hybrid student performance prediction system.

### B. Performance of the Proposed Model

Table II shows the evaluation metrics of the proposed model evaluated according to four classic classification metrics: accuracy, precision, recall, and F1-score [41].

TABLE II. EVALUATION METRICS OF THE PROPOSED HYBRID STUDENT PERFORMANCE PREDICTION MODEL

Metrics	Accuracy	Precision	Recall	F1-score
Proposed model	0.9536	0.9762	0.9633	0.9697

Accuracy is the proportion of correct predictions (TP + TN) to the total number of occurrences (TP + TN + FP + FN) [21]. The model achieved a value of 95.36%, indicating that it correctly predicted approximately 95.36% of cases and makes few errors overall.

The precision is defined as the percentage of correctly predicted positive observations relative to all predicted positive observations [42]. The model achieved a score of 97.62%, indicating that the vast majority of positive predictions were correct.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall is the proportion of correctly predicted positive observations to all actual positive observations. Recall is inversely related to the number of false negatives [43].

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

With a recall of 96.33%, the model detects true positives very effectively. In other words, it correctly identifies the majority of students who actually pass.

F1-score is the balance between precision and recall, which is determined by taking their weighted averages [42].

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

This metric reaches 96.97%, demonstrating a good balance between the two and highlighting the robustness of the model in handling potential imbalances between classes.

The results indicate that the proposed hybrid model is highly effective and well balanced across the various metrics. This performance suggests that the model is reliable for predicting SAPs while minimizing critical classification errors. This robustness is due to the combination of the complementary strengths of the algorithms integrated into the hybrid model.

### C. Statistical Significance Test

The McNemar's test is a common non-parametric test that requires only two numbers [44]. It is well suited for evaluating binary classifiers using cross-validation [45]. The McNemar's test showed a really low p-value ( $p = 0.0000$ ). This p-value should be reported as  $p < 0.001$  in accordance with scientific conventions. This indicates that there is a significant difference between the performance of our hybrid model and that of the best single model, RF. This robust statistical result supports our hypothesis that combining multiple algorithms in a hybrid architecture does indeed produce more reliable predictions than individual models considered in isolation.

### D. Performance Comparison of the Hybrid System with Previous Studies Using the OULAD Dataset

#### 1) Accuracy

Figure 5 shows a diagram illustrating the accuracy rates obtained by different ML and deep learning systems.

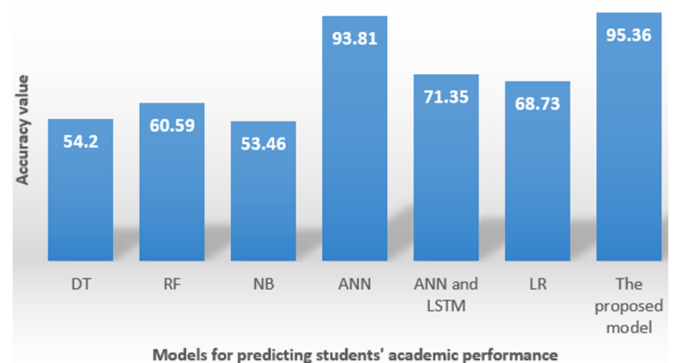


Fig. 5. Comparison of the proposed hybrid system accuracy with previous studies using the OULAD dataset.

Among the evaluated ML models for predicting SAP, DT achieved 54.2% accuracy, RF reached 60.59%, NB obtained 53.46%, and LR delivered 68.73% [46]. The ANN also performed well, achieving 93.81% accuracy [47]. The combination of ANN and Long Short-Term Memory (LSTM) did not significantly improve accuracy, reaching 71.35% [48]. The proposed system demonstrated exceptional performance with 95.36%, surpassing all conventional approaches.

## 2) Precision

Figure 6 presents the precision values obtained by different learning systems, enabling us to assess their ability to avoid false positives. The proposed system achieves the highest performance, with a precision of 97.62%, outperforming all the other systems. This indicates that the proposed system is particularly effective at providing correct positive predictions, which is essential in contexts where false alarms must be minimized.

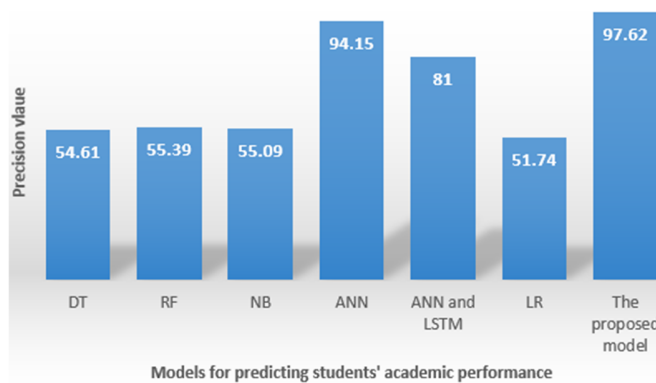


Fig. 6. Comparison of the proposed hybrid system precision with previous studies using the OULAD dataset.

DT attained 54.61% precision, followed closely by RF at 55.39% and NB at 55.09%, whereas LR showed the lowest precision among conventional methods at 51.74% [46]. The ANN also performed very well, with 94.15% [47], confirming the effectiveness of this approach for accurate classification. However, the proposed model achieved remarkable precision of 97.62%, demonstrating superior predictive capability.

## 3) Recall

In terms of recall, as shown in Figure 7, LR achieved a recall value of 68.73%, outperforming RF at 60.59%, DT at 54.2%, and NB at 53.46% [46]. The ANN achieved a recall value of 95.13% [47]. The combination of ANN and LSTM had a particularly low recall score of just 53% [48], indicating that a large proportion of positive cases are not detected by this system.

The proposed system performed best, with a recall of 96.33%, confirming its ability to detect almost all positive instances.

## 4) F1-Score

Figure 8 shows the F1-score values obtained by the same learning systems. The proposed system achieved the highest score of 96.97%, confirming not only its ability to predict

accurately (high precision) but also its capacity to identify relevant cases (high recall). It is the most reliable and balanced system among those tested. The ANN followed closely, with a score of 95.83% [47], demonstrating its effectiveness in contexts where a compromise between precision and recall is required.

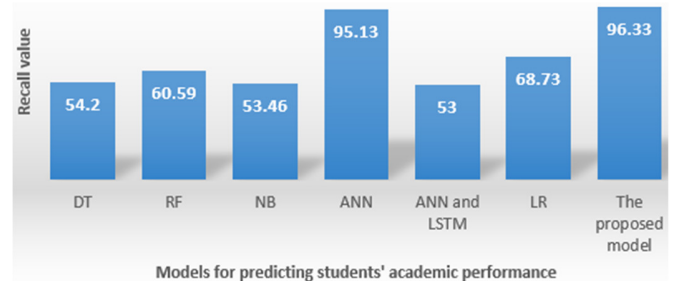


Fig. 7. Comparison of the proposed hybrid system recall with previous studies using the OULAD dataset.

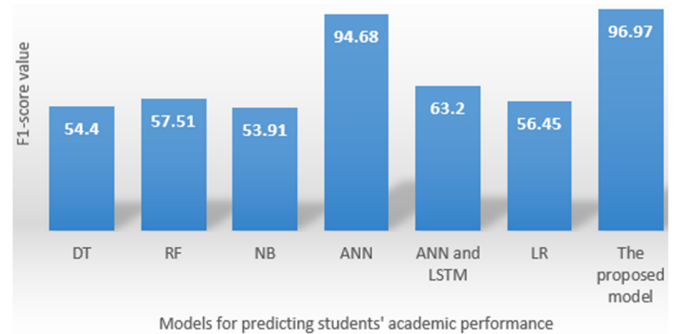


Fig. 8. Comparison of the proposed hybrid system F1-score with previous studies using the OULAD dataset.

These results indicate that other systems struggle to achieve a satisfactory balance between precision and recall, limiting their applicability for complex tasks.

## V. CONCLUSION

Predicting Student Academic Performance (SAP) is a complex task due to the interplay of cognitive, behavioral, and institutional factors. In this paper, we proposed a hybrid system to predict SAP, combining Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANNs). The system predicts whether a student is likely to succeed or fail based on specific inputs, including academic session, gender, region of origin, educational level, and socioeconomic status, providing both a binary outcome and a probability score for the prediction.

This hybrid approach leverages the complementary strengths of each model while mitigating individual weaknesses. The system achieved an overall accuracy of 95.35% and an F1-score of 96.97%. Using the Synthetic Minority Oversampling Technique (SMOTE) to balance the classes improved the detection of students at risk of failure, which is critical for targeted educational interventions. McNemar's test yielded a low p-value ( $p < 0.001$ ), providing

strong statistical support for the hybrid approach over individual models.

Beyond performance metrics, the hybrid system provides immediate practical benefits for educational leaders. For example, teachers can identify at-risk students early and implement targeted interventions, such as adaptive teaching or additional lessons. Institutions can integrate the system into learning platforms to optimize resource allocation, enhance student support, and reduce dropout rates.

## VI. LIMITATIONS AND FUTURE WORK

The system has several limitations. First, its performance heavily depends on the quality of the training data. If the data do not accurately represent the diversity of students or learning situations, the predictions could be biased. In addition, the system does not account for potentially important factors such as personal motivation or social interactions between students, limiting its ability to capture the full complexity of the learning process. Another limitation is the system's interpretability. Since it combines four separate algorithms, it may be difficult for non-experts to understand, making it challenging to explain predictions to teachers or students. Furthermore, its computational complexity requires powerful hardware, which could hinder large-scale deployment.

There are several ways to improve the proposed system. Incorporating behavioral indications, such as forum interactions or time spent on exercises, could enhance predictive accuracy. Including information about students' prior academic history would also strengthen the model. From a technological perspective, further optimization could involve experimenting with more complex architectures, integrating various techniques, or analyzing data over time to monitor student performance more dynamically.

## CONFLICT OF INTEREST

The authors declare that they have no known conflicts of interest that could have appeared to influence the work reported in this paper.

## AVAILABILITY OF DATA AND MATERIALS

The dataset generated during and/or analyzed during the current study is available from the corresponding author upon reasonable request.

## REFERENCES

- [1] M. T. Batayneh, "Economic, Social, And Academic Factors Affecting Academic Achievement of Students at The University of Taibah," *Res Militaris*, vol. 13, no. 1, pp. 1660–1679, 2023.
- [2] A. Kukkar, R. Mohana, A. Sharma, and A. Nayyar, "Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms," *Education and Information Technologies*, vol. 28, no. 8, pp. 9655–9684, Aug. 2023, <https://doi.org/10.1007/s10639-022-11573-9>.
- [3] M. L. Bernacki, M. M. Chavez, and P. M. Uesbeck, "Predicting achievement and providing support before STEM majors begin to fail," *Computers & Education*, vol. 158, Dec. 2020, Art. no. 103999, <https://doi.org/10.1016/j.compedu.2020.103999>.
- [4] B. K. Yousafzai *et al.*, "Student-Performer: Student Academic Performance Using Hybrid Deep Neural Network," *Sustainability*, vol. 13, no. 17, Sep. 2021, Art. no. 9775, <https://doi.org/10.3390/su13179775>.
- [5] S. Hussain and M. Q. Khan, "Student-Performer: Predicting Students' Academic Performance at Secondary and Intermediate Level Using Machine Learning," *Annals of Data Science*, vol. 10, no. 3, pp. 637–655, Jun. 2023, <https://doi.org/10.1007/s40745-021-00341-0>.
- [6] C. F. Rodríguez-Hernández, E. Cascallar, and E. Kyndt, "Socio-economic status and academic performance in higher education: A systematic review," *Educational Research Review*, vol. 29, Feb. 2020, Art. no. 100305, <https://doi.org/10.1016/j.edurev.2019.100305>.
- [7] J. C. Ayala and G. Manzano, "Academic performance of first-year university students: the influence of resilience and engagement," *Higher Education Research & Development*, vol. 37, no. 7, pp. 1321–1335, Nov. 2018, <https://doi.org/10.1080/07294360.2018.1502258>.
- [8] C. MacCann, Y. Jiang, L. E. R. Brown, K. S. Double, M. Bucich, and A. Minbashian, "Emotional intelligence predicts academic performance: A meta-analysis," *Psychological Bulletin*, vol. 146, no. 2, pp. 150–186, Feb. 2020, <https://doi.org/10.1037/bul0000219>.
- [9] B. A. Al-Sheeb, A. M. Hamouda, and G. M. Abdella, "Modeling of student academic achievement in engineering education using cognitive and non-cognitive factors," *Journal of Applied Research in Higher Education*, vol. 11, no. 2, pp. 178–198, Jan. 2019, <https://doi.org/10.1108/JARHE-10-2017-0120>.
- [10] M. Goga, S. Kuyoro, and N. Goga, "A Recommender for Improving the Student Academic Performance," *Procedia - Social and Behavioral Sciences*, vol. 180, pp. 1481–1488, May 2015, <https://doi.org/10.1016/j.sbspro.2015.02.296>.
- [11] S. Huang and N. Fang, "Work in progress — Prediction of students' academic performance in an introductory engineering course," in *2011 Frontiers in Education Conference*, Rapid City, SD, USA, 2011, pp. S4D-1-S4D-3, <https://doi.org/10.1109/FIE.2011.6142729>.
- [12] O. H. T. Lu, A. Y. Q. Huang, J. C. H. Huang, A. J. Q. Lin, H. Ogata, and S. J. H. Yang, "Applying Learning Analytics for the Early Prediction of Students' Academic Performance in Blended Learning," *Journal of Educational Technology & Society*, vol. 21, no. 2, pp. 220–232, Apr. 2018.
- [13] S. Mammadov, "Big Five personality traits and academic performance: A meta-analysis," *Journal of Personality*, vol. 90, no. 2, pp. 222–255, Apr. 2022, <https://doi.org/10.1111/jopy.12663>.
- [14] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications*, vol. 33, no. 1, pp. 135–146, Jul. 2007, <https://doi.org/10.1016/j.eswa.2006.04.005>.
- [15] L. Juhaňák, J. Zounek, and L. Rohlíková, "Using process mining to analyze students' quiz-taking behavior patterns in a learning management system," *Computers in Human Behavior*, vol. 92, pp. 496–506, Mar. 2019, <https://doi.org/10.1016/j.chb.2017.12.015>.
- [16] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601–618, Nov. 2010, <https://doi.org/10.1109/TSMCC.2010.2053532>.
- [17] R. Baker and G. Siemens, "Educational Data Mining and Learning Analytics," in *The Cambridge Handbook of the Learning Sciences*, 2nd ed., R. K. Sawyer, Ed. Cambridge, UK: Cambridge University Press, 2014, pp. 253–272, <https://doi.org/10.1017/CBO9781139519526.016>.
- [18] E. T. Lau, L. Sun, and Q. Yang, "Modelling, prediction and classification of student academic performance using artificial neural networks," *SN Applied Sciences*, vol. 1, no. 9, Aug. 2019, Art. no. 982, <https://doi.org/10.1007/s42452-019-0884-7>.
- [19] M. F. Musso, E. C. Cascallar, N. Bostani, and M. Crawford, "Identifying Reliable Predictors of Educational Outcomes Through Machine-Learning Predictive Modeling," *Frontiers in Education*, vol. 5, Jul. 2020, Art. no. 104, <https://doi.org/10.3389/educ.2020.00104>.
- [20] C. F. Rodríguez-Hernández, M. Musso, E. Kyndt, and E. Cascallar, "Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation," *Computers and Education: Artificial Intelligence*, vol. 2, Jan. 2021, Art. no. 100018, <https://doi.org/10.1016/j.caeai.2021.100018>.

- [21] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, Mar. 2022, Art. no. 11, <https://doi.org/10.1186/s40561-022-00192-z>.
- [22] Z. Ahmad and E. Shahzadi, "Prediction of Students' Academic Performance Using Artificial Neural Network," *Bulletin of Education and Research*, vol. 40, no. 3, pp. 157–164, Dec. 2018.
- [23] A. Alshantit and A. Namoun, "Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification," *IEEE Access*, vol. 8, pp. 203827–203844, 2020, <https://doi.org/10.1109/ACCESS.2020.3036572>.
- [24] I. Đurđević Babić, "Machine learning methods in predicting the student academic motivation," *Croatian Operational Research Review*, vol. 8, no. 2, pp. 443–461, Dec. 2017, <https://doi.org/10.17535/crorr.2017.0028>.
- [25] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Transfer Learning from Deep Neural Networks for Predicting Student Performance," *Applied Sciences*, vol. 10, no. 6, Mar. 2020, Art. no. 2145, <https://doi.org/10.3390/app10062145>.
- [26] S. TK and Midhunchakkavarthy, "Academic Performance Prediction of At-Risk Students using Machine Learning Techniques," in *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering*, Greater Noida, India, 2023, pp. 1222–1227, <https://doi.org/10.1109/ICACITE57410.2023.10183199>.
- [27] E. A. Abbas, N. A. Hussein, and R. Alubady, "Machine Learning for Link Prediction between Nodes in Complex Networks," *Fusion: Practice and Applications*, vol. 18, no. 1, pp. 41–55, 2025, <https://doi.org/10.54216/FPA.180104>.
- [28] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, pp. 86716–86727, 2024, <https://doi.org/10.1109/ACCESS.2024.3416838>.
- [29] Q. Hu and H. Rangwala, "Course-Specific Markovian Models for Grade Prediction," in *22nd Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, Melbourne, Australia, 2018, pp. 29–41, [https://doi.org/10.1007/978-3-319-93037-4\\_3](https://doi.org/10.1007/978-3-319-93037-4_3).
- [30] B. Xu and D. Yang, "Motivation Classification and Grade Prediction for MOOCs Learners," *Computational Intelligence and Neuroscience*, vol. 2016, no. 1, Jan. 2016, Art. no. 2174613, <https://doi.org/10.1155/2016/2174613>.
- [31] Y. Zhang, Y. Yun, H. Dai, J. Cui, and X. Shang, "Graphs Regularized Robust Matrix Factorization and Its Application on Student Grade Prediction," *Applied Sciences*, vol. 10, no. 5, Mar. 2020, Art. no. 1755, <https://doi.org/10.3390/app10051755>.
- [32] A. Cakmak, "Predicting Student Success in Courses via Collaborative Filtering," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 5, no. 1, pp. 10–17, Mar. 2017, <https://doi.org/10.18201/ijisae.2017526690>.
- [33] C. Kaope and Y. Prityanto, "The Effect of Class Imbalance Handling on Datasets Toward Classification Algorithm Performance," *MATRIK*, vol. 22, no. 2, pp. 227–238, Mar. 2023, <https://doi.org/10.30812/matrik.v22i2.2515>.
- [34] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, <https://doi.org/10.1613/jair.953>.
- [35] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open University Learning Analytics dataset," *Scientific Data*, vol. 4, no. 1, Nov. 2017, Art. no. 170171, <https://doi.org/10.1038/sdata.2017.171>.
- [36] C. Romero and S. Ventura, "Data mining in education," *WIREs Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 12–27, Jan. 2013, <https://doi.org/10.1002/widm.1075>.
- [37] S. T. Jishan, R. I. Rashu, N. Haque, and R. M. Rahman, "Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique," *Decision Analytics*, vol. 2, no. 1, Mar. 2015, Art. no. 1, <https://doi.org/10.1186/s40165-014-0010-2>.
- [38] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 40–46, Jun. 2021, <https://doi.org/10.1016/j.ijcce.2021.01.001>.
- [39] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble Learning for Disease Prediction: A Review," *Healthcare*, vol. 11, no. 12, Jun. 2023, Art. no. 1808, <https://doi.org/10.3390/healthcare11121808>.
- [40] T. G. T. Tran and T. N. Tran, "Enhancing Regression Accuracy and Reliability with Ensemble Learning and Iterative Hyperparameter Optimization," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 24541–24548, Aug. 2025, <https://doi.org/10.48084/etasr.11266>.
- [41] A. B. Altamimi, "Big Data in Education: Students at Risk as a Case Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11705–11714, Oct. 2023, <https://doi.org/10.48084/etasr.6190>.
- [42] E. C. Mutlu, T. Oghaz, A. Rajabi, and I. Garibay, "Review on Learning and Extracting Graph Features for Link Prediction," *Machine Learning and Knowledge Extraction*, vol. 2, no. 4, pp. 672–704, Dec. 2020, <https://doi.org/10.3390/make2040036>.
- [43] S. Haghani and M. R. Keyvanpour, "A systemic analysis of link prediction in social network," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 1961–1995, Oct. 2019, <https://doi.org/10.1007/s10462-017-9590-2>.
- [44] S. Kim and W. Lee, "Does McNemar's test compare the sensitivities and specificities of two diagnostic tests?," *Statistical Methods in Medical Research*, vol. 26, no. 1, pp. 142–154, Feb. 2017, <https://doi.org/10.1177/0962280214541852>.
- [45] R. Saha, "Comparing Classification Models on Kepler Data." arXiv, 2019, <https://doi.org/10.13140/RG.2.2.11232.43523>.
- [46] Y. Jedidi *et al.*, "Predicting students' academic performance and modeling using data mining techniques," *Mathematical Modeling and Computing*, vol. 11, no. 3, pp. 814–825, Sep. 2024.
- [47] H. Chavez, B. Chavez-Arias, S. Contreras-Rosas, J. M. Alvarez-Rodriguez, and C. Raymundo, "Artificial neural network model to predict student performance using nonpersonal information," *Frontiers in Education*, vol. 8, Feb. 2023, Art. no. 1106679, <https://doi.org/10.3389/educ.2023.1106679>.
- [48] F. A. Al-azazi and M. Ghurab, "ANN-LSTM: A deep learning model for early student performance prediction in MOOC," *Heliyon*, vol. 9, no. 4, Apr. 2023, Art. no. e15382, <https://doi.org/10.1016/j.heliyon.2023.e15382>.