Journal of Artificial Intelligence and Metaheuristic

ISSN: 2833-5597 Vol: 11, No. 02, 2026 PP. 35–54

Article DOI: https://doi.org/10.54216/JAIM.110203

FULL LENGTH ARTICLE

A Hybrid Al-Biruni Earth Radius–Random Forest Model for Accurate and Efficient Student Performance Classification

Mohamed E. Ghoneim[1,*]

[1] Mathematics Department, Faculty of Science, Umm Al-Qura University, Saudi Arabia

Email: meghoneim@uqu.edu.sa

Received: December 30, 2025 Revised: February 28, 2026 Accepted: April 30, 2026 ⋆ Corresponding author

ABSTRACT

The growing availability of educational data has prompted the use of machine learning methods to predict student academic performance and support data-driven decision-making in education. Nevertheless, such models for predicting performance rely heavily on proper data preprocessing, model selection, and optimal hyperparameter settings. This research proposes a hybrid predictive architecture that combines machine learning classifiers with bio-inspired metaheuristic optimization algorithms to improve classification efficiency in educational data mining. It is based on the xAPI-Edu.A dataset of 480 students' demographic, academic, and behavioral characteristics is used to first analyze a set of baseline machine learning models, including Random Forest, XGBoost, Support Vector Machine, Multilayer Perceptron, K-Nearest Neighbors, and Gaussian Naive Bayes, using standard classification metrics. The initial experimental findings on the baseline layer show that the Random Forest classifier outperforms the other models before optimization, achieving accuracies of 0.8889 and 0.8814, and F-scores of 0.8889 and 0.8814, respectively, indicating strong generalization and equal discrimination among the classes. To further improve the predictive performance, the state-of-the-art metaheuristic algorithms, i.e., the Al-Biruni Earth Radius Optimizer (BER), the Gray Wolf Optimizer (GWO), the Particle Swarm Optimization (PSO), the Genetic Algorithms (GA) and the Whale Optimization Algorithms (WOA) are adopted to optimize the hyperparameters of the Random Forest. It has been experimentally demonstrated that every optimization approach provides a measurable performance increase, but the BER-optimized Random Forest consistently performs better. In particular, the BER-Random Forest model achieves an F-score of 0.9477 and an accuracy of 0.9439, both of which are much higher than the baseline configuration. Full statistical and visual analyses, such as kernel density estimation, Z-score heatmaps, and swarm plots, also support the strength, stability and superiority of the proposed BER-based optimization framework. Such findings demonstrate the effectiveness of metaheuristic-based hyperparameter optimization in educational predictive analytics and provide significant insights into the creation of intelligent, efficient, and data-driven systems of academic assistance.

Keywords: Student Academic Performance ▪ Machine Learning ▪ Metaheuristic Optimization ▪ Random Forest ▪ Educational Data Mining

1. INTRODUCTION

In recent years, education has undergone a notable transformation driven by technological advancements and the increasing availability of educational data. This paradigmatic shift has seen traditional pedagogical methods progressively replaced by data-driven approaches aimed at enhancing student performance and improving academic outcomes. Central to

© 2026 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

this transformation is the vast volume of educational data generated across primary, secondary, and tertiary education levels, coupled with the emergence of sophisticated machine learning algorithms and advanced data analytics methodologies [1, 2]. Academic institutions, ranging from elementary schools to universities, increasingly employ data-driven strategies to monitor, assess, and improve learning outcomes. The present study investigates the extensive range of research methodologies that have been adopted within this domain, with findings intended to support the development of robust predictive models for student performance across diverse academic disciplines.

Machine learning models have demonstrated remarkable effectiveness in predicting student academic performance, a task that was previously considered highly challenging. These models are capable of uncovering latent patterns, correlations, and trends embedded within complex educational datasets, thereby providing valuable insights to educators and institutional decision-makers. Such insights enable early identification of at-risk students, the provision of personalized instructional recommendations, and the implementation of adaptive teaching strategies [3, 4]. By leveraging machine learning techniques, educators can develop a deeper understanding of the multifaceted dynamics of educational environments and respond proactively to the diverse needs of learners. Moreover, several studies have emphasized the critical role of optimization techniques in refining predictive models and enhancing their overall performance [5, 6]. Optimization serves as an auxiliary mechanism for fine-tuning algorithmic parameters, improving predictive accuracy, and ensuring alignment between predicted outcomes and actual student performance trajectories. In addition, optimization methods facilitate effective feature selection by identifying the most relevant variables and eliminating redundant or noisy data, thereby strengthening model precision.

This study explores the evolving landscape of predictive modeling within the educational domain by examining the interplay between machine learning, optimization techniques, and statistical analysis. First, the paper investigates the effective application of machine learning algorithms—including Random Forest, Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Multilayer Perceptron (MLP), Gaussian Naïve Bayes, and K-Nearest Neighbors (KNN)—for predicting student academic performance. Second, the study examines how optimization techniques such as the Al-Biruni Earth Radius Optimizer (BER) [7], Grey Wolf Optimizer (GWO) [8], Particle Swarm Optimizer (PSO) [9], Genetic Algorithm (GA), and Whale Optimization Algorithm (WOA) can enhance the predictive capabilities of these machine learning models, with particular emphasis on the Random Forest classifier. Finally, statistical analyses, including Analysis of Variance (ANOVA), are employed to assess the impact of optimization algorithms on predictive performance. Through this multifaceted approach, the research seeks to develop a deeper understanding of optimized machine learning models for predicting student academic achievement and to make a meaningful theoretical contribution to educational predictive analytics.

The dataset used in this study is the Students' Academic Performance Dataset, also referred to as the xAPI-Educational Mining Dataset. This dataset is derived from the Experience API (xAPI) framework [10] and is publicly available through the Kaggle platform. It contains records from 480 individual students enrolled in two academic courses, with each record comprising sixteen distinct attributes. Classified as multivariate data, the dataset holds significant relevance in the fields of e-learning, educational forecasting, predictive modeling, and educational data mining. The data were collected using Kalboard 360, a compact and user-friendly learning management system designed to provide learners with flexible access to educational resources via internet-connected devices. Data collection is facilitated through xAPI, a core component of the Training and Learning Architecture that monitors and records learners' interactions and activities. The dataset attributes are categorized into three primary groups: demographic characteristics, academic background, and behavioral indicators. As such, the dataset provides a comprehensive foundation for predictive modeling and for gaining deeper insights into student performance within contemporary technology-enhanced learning environments.

In this work, a classification framework is proposed in which the BER optimization algorithm is employed to tune the Random Forest model for academic performance classification. Experimental results demonstrate that the BER-optimized Random Forest model outperforms competing approaches in terms of accuracy and predictive reliability. The XAPI-Edu-Data dataset was selected to ensure practical relevance and applicability within real-world educational contexts. The proposed method is evaluated against state-of-the-art optimization-based models, including GWO-, PSO-, GA-, and WOA-based Random Forest classifiers. Comparative analysis reveals that the BER-Random Forest model consistently achieves superior performance across multiple evaluation metrics. These metrics include sensitivity, specificity, positive predictive value, negative predictive value, and F-score. Furthermore, statistical analyses using ANOVA and the Wilcoxon Signed Rank Test confirm that the optimization strategy significantly enhances predictive accuracy relative to competing models.

Beyond its methodological innovations, this research contributes meaningfully to the development of advanced educational systems [11]. The integration of hybrid machine learning approaches—particularly those incorporating bio-inspired optimization algorithms—bridges the gap between theoretical advancements in algorithm design and practical applications in educational environments [12]. This work highlights the growing synergy between computational intelligence and educational research [13]. As global education systems increasingly strive to become adaptive, inclusive, and data-driven, the integration of these approaches is becoming essential [14].

Additionally, the study advances discussions on educational intervention strategies with an emphasis on equity and fairness [15]. Optimized predictive modeling enables the identification of inefficiencies within educational systems and allows for more accurate projections of individual student learning trajectories [16]. By detecting subtle variations in student performance related to demographic or behavioral factors, institutions can intervene early and provide targeted, individualized support. Such capabilities not only enhance

academic outcomes but also promote equitable access to educational opportunities and personalized learning experiences, which are central to modern educational reform and policy development [17].

This paper presents a comprehensive and systematically validated framework for student academic performance prediction by integrating machine learning models with metaheuristic optimization techniques. The key contributions of this work can be summarized as follows:

- A unified predictive framework is developed that combines rigorous data preprocessing, baseline machine learning classification, metaheuristic-based hyperparameter optimization, and extensive statistical evaluation to enhance the reliability of educational performance prediction.
- A broad comparative analysis is conducted involving multiple well-established machine learning classifiers and state-of-the-art metaheuristic optimization algorithms, providing a holistic assessment of their relative strengths and limitations within the educational data mining context.
- The proposed methodology is systematically benchmarked against existing state-of-the-art models and algorithms reported in the literature, ensuring a fair and comprehensive evaluation of its effectiveness and practical relevance.
- Advanced visual and statistical analysis techniques are employed to interpret model behavior, stability, and consistency across multiple evaluation dimensions, thereby strengthening the transparency and robustness of the findings.
- The study offers actionable insights into how bio-inspired optimization strategies can be effectively leveraged to improve ensemble learning models, contributing to the development of intelligent, data-driven academic support systems capable of early performance assessment and informed educational decision-making.

The remainder of this paper is organized as follows. Section 2 presents a comprehensive review of related work and highlights relevant research contributions. Section 3 describes the methodology adopted in this study, including data sources and analytical procedures. Section 4 details the design of the predictive models, while Section 5 discusses the experimental results and findings. Finally, Section 6 concludes the paper by summarizing key outcomes and outlining directions for future research.

2. RELATED WORKS

Predicting the academic performance of individual students within large-scale and heterogeneous educational datasets has become increasingly challenging due to the multidimensional nature of learning outcomes and the diversity of factors that influence academic success. Contemporary educational repositories typically contain a mixture of demographic attributes, academic background variables, and behavioral engagement indicators collected from learning management systems. These data sources are often noisy, partially redundant, and characterized by nonlinear interactions, which limits the effectiveness of purely linear or single-model approaches. Consequently, the research community has progressively shifted toward hybrid learning pipelines, ensemble-based learners, and optimization-driven modeling strategies to improve predictive reliability and generalization across varied educational contexts.

A major direction in the literature has focused on the design of scalable architectures capable of processing large educational datasets while retaining discriminative information. In this regard, Vora and Rajamani [1] proposed a big-data-oriented prediction architecture composed of two modules: a distributed processing module implemented through MapReduce with dimensionality reduction using PCA, followed by a hybrid classifier integrating a Deep Belief Network with Support Vector Machines. This work exemplifies the importance of combining scalable preprocessing with expressive classifiers when dealing with voluminous student data. Beyond the specific algorithmic design, the study highlights a recurring theme in educational data mining: robust preprocessing and representation learning are often prerequisites for effective prediction, particularly when datasets include mixed attribute types and potentially high feature redundancy.

In parallel, the rise of distance and online education has motivated the development of prediction methods that can capture performance as an ordered or ranked outcome rather than a purely nominal class. Gámez-Granados et al. [2] introduced FlexNSLVOrd, an ordinal classification algorithm grounded in fuzzy modeling principles, in which students are classified into ordered achievement categories. This approach is particularly relevant for educational settings where the distinction between adjacent performance levels is pedagogically meaningful and where interpretability is required for intervention planning. The reported improvements over multiple competing methods underscore the value of aligning the learning formulation (ordinal vs. nominal) with the educational semantics of the target variable. More broadly, this line of work emphasizes that prediction quality can be improved not only by selecting powerful learners, but also by choosing problem formulations that reflect the structure of educational outcomes.

Another stream of studies has investigated the utility of classical machine learning algorithms as accessible and interpretable baselines for educational stakeholders. Pallathadka et al. [3] examined Naïve Bayes, ID3, C4.5, and SVM within educational mining workflows, demonstrating their applicability in predicting student performance from academic background information. Although not all studies report complete metric sets, such contributions remain important because they establish comparative baselines and clarify the relative strengths of probabilistic, tree-based, and margin-based learners in educational data contexts. In practice, these models often serve as initial screening tools due to their simplicity, interpretability, and modest computational requirements.

At the level of higher education assessment, Yağcı [18] evaluated multiple classifiers—including Random Forest, SVM, Logistic Regression, Naïve Bayes, and KNN—for predicting final examination outcomes using midterm-related predictors. This work reinforces two key insights that recur across the lit-

erature. First, ensemble-based classifiers frequently provide competitive performance because aggregation mechanisms reduce model variance and mitigate overfitting. Second, early prediction is of high practical value because it enables institutions to allocate academic support resources proactively rather than reactively. Such studies therefore motivate ongoing efforts to improve both prediction accuracy and operational reliability, especially in early-warning settings where errors can lead to missed support opportunities or unnecessary interventions.

More recently, research has increasingly emphasized hybrid strategies that combine preprocessing enhancements, balanced learning, and metaheuristic optimization. Cheng et al. [19] investigated the integration of SMOTE-based balancing with multiple classifiers (e.g., Random Forest, Decision Trees, KNN, MLP, and XGBoost), illustrating that data-level interventions can significantly improve classifier effectiveness when class distributions are skewed. This is a critical concern in educational datasets, where high-performing, medium-performing, and at-risk groups may be unevenly represented. Such studies collectively suggest that model performance should be interpreted jointly with data characteristics, and that improvements may arise from carefully designed preprocessing and resampling strategies as much as from the classifier selection itself.

Ensemble learning continues to be a prominent trend due to its strong empirical performance across domains, including educational performance prediction. Keser and Aghalarova [20] proposed HELA, a hybrid ensemble that combines multiple boosting-based learners with a meta-classifier to exploit complementary strengths across model families. This approach aligns with the broader literature indicating that stacking and super-learning strategies can capture diverse decision boundaries and improve robustness, especially when student features reflect complex and interacting effects. In a related vein, Lau et al. [21] combined statistical analysis with neural network modeling, illustrating that explanatory insight (via statistical factor identification) can be integrated with predictive learning (via nonlinear function approximation). This hybridization trend is increasingly important because educational predictive systems are expected not only to forecast outcomes but also to support actionable interpretation for instructors and policymakers.

A further set of studies has highlighted the value of behavioral and activity-based data, particularly in blended and online learning environments where engagement can be recorded continuously. Francis and Babu [22] combined clustering and classification to predict student outcomes, reinforcing the importance of discovering latent structure in student populations before applying supervised prediction. Deeva et al. [23] advanced this behavioral perspective by adopting sequence classification with time-based windows, capturing temporal patterns that static feature vectors may miss. Nayak et al. [24] similarly emphasized the role of behavioral indicators in online education settings, reporting that behavior-rich representations can substantially strengthen prediction capability. Collectively, these works indicate that educational performance is often reflected in patterns of engagement over time and that predictive models can benefit from incorporating temporal or sequential descriptors when available.

The COVID-19 pandemic accelerated the adoption of online learning and intensified demand for reliable predictive analytics that operate under remote engagement conditions. Studies such as [25] investigated grade and engagement prediction under online learning transformations and reinforced that ensemble learners (notably tree-based ensembles) are often effective under such settings because they can model heterogeneous feature effects and tolerate moderate noise. Early prediction approaches such as the Random Wheel algorithm [26] further indicate a shift toward models that not only classify outcomes but also support decision confidence, which is valuable when institutions must determine whether an intervention is justified before a course begins. Such directions point to a broader requirement in educational analytics: predictive systems should be timely, stable, and context-aware, rather than solely accurate under retrospective evaluation.

Survey and review studies provide additional context regarding methodological trends and emerging challenges. Khosravi et al. [27] synthesized AI-based student performance prediction techniques, documenting the growing role of feature engineering, normalization, and ensemble learning in improving predictive outcomes. Complementarily, Dissanayake et al. [28] emphasized the importance of Bayesian hyperparameter optimization for improving model performance and interpretability in academic prediction tasks. These works collectively highlight that the field is moving toward automated and principled tuning strategies, as hyperparameter sensitivity remains a key source of performance variability across models. They also imply that optimization should be viewed as a first-class component of educational predictive modeling pipelines rather than an optional post-processing step.

Despite the substantial progress summarized above, several gaps remain evident in the literature and motivate the present study. First, while optimization is widely recognized as beneficial, many educational prediction studies either rely on manual tuning, limited grid search, or optimization approaches that do not fully explore population-based bio-inspired strategies under controlled comparisons. Second, the literature often reports improvements using single metrics or isolated evaluation settings; however, educational applications require balanced assessment across complementary metrics (e.g., sensitivity vs. specificity) to reduce risks associated with misclassification. Third, although behavioral and categorical variables are commonly acknowledged as important, fewer studies provide integrated visual–statistical analyses that empirically validate feature relevance and performance stability prior to model deployment. Finally, comparative benchmarking against multiple state-of-the-art models and optimization strategies under a unified experimental protocol remains relatively limited. In response to these gaps, the present work develops a comprehensive prediction framework that integrates preprocessing, baseline model benchmarking, metaheuristic optimization, and multi-perspective evaluation to support robust and transparent student performance classification.

Table 1 synthesizes the core contributions of representative studies reviewed in this section, highlighting their focus areas, methodological choices, and principal findings. Taken together, this body of research establishes the importance of machine learning, hybrid modeling, and optimization-driven

pipelines in educational data mining, while also motivating the need for systematic evaluation frameworks that jointly address predictive accuracy, stability, and methodological reproducibility.

3. MATERIALS AND METHODS

This section describes the research dataset and the methods employed in this study. It begins with an overview of the dataset and the data preprocessing pipeline, including key steps such as data cleaning and feature engineering. An important component of preprocessing involves normalization and standardization to ensure data quality, consistency, and validity. The section also introduces the proposed optimization strategy, focusing on the Al-Biruni Earth Radius (BER) algorithm, and presents the machine learning techniques used in the analysis, particularly Random Forest classification and ensemble learning approaches, which are essential for extracting meaningful patterns from the preprocessed data. To explain the methodology chosen for the present research in a well-organized manner, a multi-stage analysis framework is depicted in Figure 1. The suggested framework is intended to systematically process raw educational data into reliable predictive information through a series of coherent steps that combine data preprocessing, baseline modeling, metaheuristic optimization, and statistical assessment. Figure 1 indicates that the workflow starts with a strict data preprocessing that improves the quality and consistency of the data, and then the creation of multiple reference machine learning models. Then, hyperparameter optimization for the Random Forest classifier is performed using the Al-Biruni Earth Radius (BER) metaheuristic. Lastly, extensive statistical and performance analyses are conducted to assess the effectiveness and robustness of the optimized predictive models. Such a systematic pipeline additionally provides methodological transparency, reproducibility, and a moderate evaluation of the predictive accuracy and statistical significance of all experimental steps.

3.1 Dataset

The academic performance dataset used in this study is referred to as the Students' Academic Performance Dataset (also known as xAPI-Edu.Data). [10] and can be accessed through Kaggle. The present xAPI learning mining dataset is based on data collected on students' academic performance in the chosen courses via the Kalboard 360 e-learning system. The dataset contains records for 480 students, and each student profile consists of 16 attributes describing various dimensions of their educational backgrounds. The Experience API (xAPI), as part of the Training and Learning Architecture (TLA), supports data collection. xAPI is designed to quantify and capture learners' activities and learning experiences, including reading learning resources and viewing teaching videos, making it a useful resource for educational analytics and predictive modelling.

The dataset will consist 480 records of students characterized by 16 different features which can be divided into three major categories: (i) demographic features (e.g., gender and nationality), (ii) characteristics of the academic background (e.g., educational stage, grade level, and section), and (iii) behavioral characteristics (e.g., classroom participation and access to learning resources). The features that are contained in the xAPI-Edu.The data dataset is described in detail in Table 2. Each attribute is a particular demographic, academic, or behavioral orientation of students and can be considered a comprehensive foundation for the analysis and prediction of academic performance.

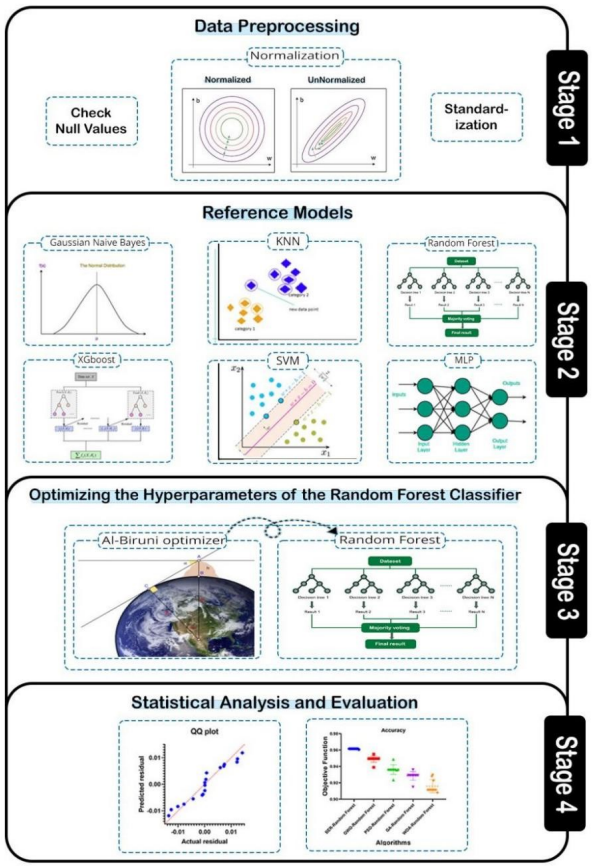

Figure 1. Proposed multi-stage framework for data preprocessing, model construction, BER-based optimization, and statistical evaluation.

3.2 Data Preprocessing

Machine learning pipelines. Data preprocessing is an important step of the process because the quality of input data directly determines the performance, reliability, and generalizability of predictive models. A number of preprocessing procedures were implemented for the xAPI-Edu.A dataset was used in this study to ensure data integrity and improve the learning capacity of the proposed models. Such procedures include handling missing data, nominal feature coding, scaling numerical variables, and assessing feature correlations.

3.2.1 Handling Missing Values

Incomplete or missing data may negatively affect model training and introduce bias in predictions. Thus, appropriate imputation strategies were used to handle missing values in the dataset. Statistical imputation methods, such as mean and median imputation, have been used for numerical attributes and are based on the distribution of the feature. Approximately normally distributed functions were mean imputed, whereas skewed distributions were more likely to use median imputation to avoid the impact of outliers.

Besides that, K-Nearest Neighbors (KNN) imputation was

Table 1. Summary of related studies on student academic performance prediction

Ref.	Focus	Methodology	Key Findings
[1]	Students' performance prediction	MapReduce, PCA, Deep Belief Network, Support Vector Machine	The proposed hybrid model demonstrated superior capability in analyzing large-scale educational datasets and accurately predicting student performance.
[2]	Predicting students' performance	FlexNSLVOrd classification algorithm	FlexNSLVOrd outperformed competing models in terms of classification accuracy, providing improved interpretability and predictive insight.
[3]	Analysis of student and teacher data	Naïve Bayes, ID3, C4.5, SVM	The study focused on predicting student performance using multiple classifiers; however, specific quantitative accuracy values were not reported.
[18]	Predicting final exam grades	Random Forest, SVM, Logistic Regression, Naïve Bayes, KNN	The proposed models achieved classification accuracies ranging from 70% to 75% and effectively identified students at high risk of failure.
[19]	Predicting and classifying students' performance	Random Forest, Decision Trees, KNN, MLP, XGBoost	The incorporation of SMOTE-based data balancing significantly improved classification performance across evaluated models.
[20]	Prediction of students' academic performance	Gradient Boosting, XGBoost, LightGBM, Super Learner	The hybrid ensemble approach achieved high prediction accuracy for mathematics and Portuguese language courses.
[21]	Predicting students' performance	Neural network model with 11 input variables	The neural network achieved an accuracy of 84.8% in predicting students' cumulative grade point average (CGPA).
[22]	Predicting academic performance	Hybrid classification and clustering algorithm	An accuracy of 75.47% was achieved using academic, behavioral, and extracurricular student features.
[23]	Behavioral sequence-based performance prediction	Sequence classification techniques	Course-specific predictive models achieved an accuracy of up to 90% by leveraging sequential behavioral patterns.
[24]	Impact of student behavior in online education	Decision Tree (J48), Random Forest, MLP, Naïve Bayes	The Random Forest model achieved 100% accuracy when behavioral features were incorporated alongside other attributes.
[25]	Predicting students' performance	Random Forest classifier	The model achieved 83% accuracy for engagement prediction and 85% accuracy for grade prediction.
[26]	Early prediction of student performance	Random Wheel algorithm	The proposed approach achieved over 80% accuracy in predicting student failure and improvement prior to course commencement.

taken into consideration on selected numerical features. This technique approximates missing values by averaging the values of the closest neighbors in the feature space. KNN imputation maintains local data structures and relationships, making it especially appropriate for educational datasets where behavioral patterns can be interdependent.

For categorical features, missing values were handled by assigning the most common category (mode) for each feature. This strategy will ensure categorical consistency but reduce information loss.

3.2.2 Feature Encoding

Machine learning algorithms require numerical data; hence, categorical variables were encoded as numerical values using suitable methods. Features that did not inherently have an order and were nominal, e.g., gender, nationality, topic, and type of guardian, were one-hot encoded. This method generates binary indicator variables of each category, avoiding the establishment of artificial ordinal relationships.

Binary categorical variables (parent survey response and parent school satisfaction) were assigned to label encoding, in which the numerical labels (e.g., Yes = 1, No = 0) were assigned to the values. For ordinal categorical features with a natural order, ordinal encoding was used to retain meaningful ranking information across categories, e.g., educational stage and grade level.

3.2.3 Feature Scaling

To ensure that numerical characteristics play a proportional role in learning, especially in distance-based and gradient-based algorithms, one must scale features. Two types of scaling were used in this research, namely, standardization

Table 2. Description of xAPI-Edu.Data Dataset Features

Feature Name	Description and Possible Values
Gender	Student gender; nominal values are Male and Female.
Nationality	Student nationality; nominal values include Egypt, Kuwait, Saudi Arabia, Lebanon, USA, and others.
Place of Birth	Country of birth of the student; values include Egypt, Kuwait, Saudi Arabia, Lebanon, USA, and others.
Educational Stage	Current educational level; nominal values are Lower level, Middle School, and High School.
Grade Level	Academic grade in which the student is enrolled; values range from G-01 to G-12.
Section ID	Classroom section identifier; nominal values are A, B, and C.
Topic	Subject studied by the student; includes English, Spanish, French, Arabic, IT, Math, Chemistry, Biology, Science, History, Quran, and Geology.
Semester	Academic semester; nominal values are First and Second.
Guardian	Student's legal guardian; nominal values are Mom and Father.
Raised Hands	Number of times the student raises their hand during class; integer values range from 0 to 100.
Visited Resources	Frequency of accessing learning resources; integer values range from 0 to 100.
Viewing Announcements	Number of times the student views course announcements; integer values range from 0 to 100.
Discussion Groups	Level of participation in discussion groups; integer values range from 0 to 100.
Parent Answering Survey	Indicates whether the parent responded to the survey; nominal values are Yes and No.
Parent School Satisfaction	Parent satisfaction with the school; nominal values are Yes and No.
Student Absence Days	Absence level during the semester; values are categorized as Low ($\leq$7 days) or High (>7 days).

and normalization.

Numerical features were standardized by setting their means to 0 and their variances to 1. This is particularly useful for feature-sensitive algorithms such as Support Vector Machines and Multilayer Perceptrons. Minimum-maximum scaling (also known as normalization) was used to rescale the feature values to 0-1, making model training much faster and the numerical results more stable.

3.2.4 Feature Correlation Analysis

The analysis of feature correlations was performed to identify redundant or correlated variables that might adversely affect model performance due to multicollinearity. Numerical features were computed, and Pearson correlation coefficients were used to measure the linear relationship between pairs of attributes. Features with high correlation values exceeding a predetermined threshold were analyzed, and the redundant ones were eliminated to reduce the model's dimensionality and increase its interpretability. For categorical features, the association values were evaluated using appropriate association measures, such as the Chi-square test, to assess dependency between variables. The preprocessing step enhanced computational efficiency and minimized overfitting, and it added predictive models that were robust due to the removal of highly correlated or non-informative features.

A more detailed visual analytics framework is used to understand the patterns of behavioral engagement among students and their connection to academic performance. Figure 2 shows a multi-plot visualization that combines interpolation-based probability surfaces, density estimation, histogram analysis, and clustering methods. This number represents the overall interaction and formation of major engagement-related features, such as raised hands, visits to resources, participation in discussions, and views of announcements, which create a clear pattern in the data. Using supervised and unsupervised perspective analysis, the visualization allows intuitive analysis of more complex associations, highlights areas of high behavioral concentration, and demonstrates the latent structural organization among students. This exploratory visual analysis supports the following machine learning modeling approach by justifying feature relevance, establishing non-linear patterns, and providing empirical justification for the use of high-order classification and optimization approaches. To further discuss the effect of the categorical attribute on students' academic performance, an effect analysis by categorical variable was conducted. Figure 3 depicts the associations among the identified categorical variables, i.e., parent survey response, parent school satisfaction, student absence days, and gender and the academic performance classes distribution. Heatmaps can help compare the associations between various categorical conditions and the high, medium, and low levels

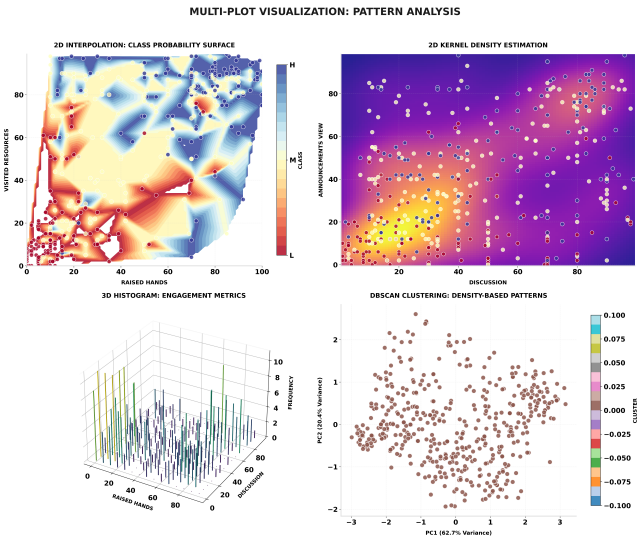

Figure 2. Multi-plot visualization for behavioral pattern analysis.

of performance. The figure clearly shows the performance differences related to behavioral, demographic, and parental involvement factors by presenting the proportions of classes as percentages. Such analysis supports interpretability by highlighting the categorical tendencies that can lead to either academic success or poor performance, thus providing empirical motivation for their inclusion in future predictive models and optimization steps. To learn more about the behavioral

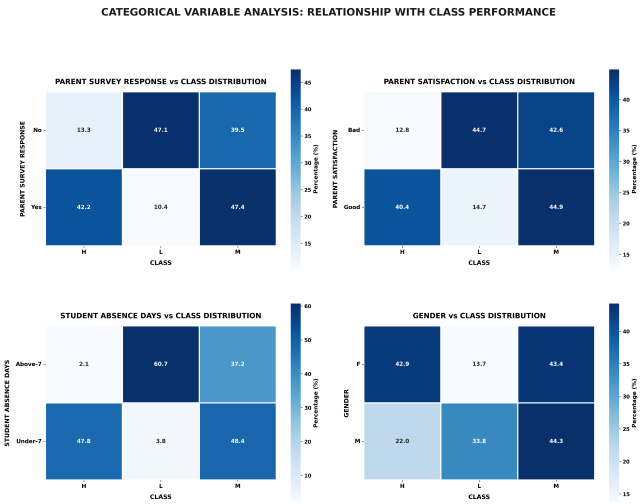

Figure 3. Categorical variable analysis illustrating relationships with class performance.

engagement patterns that characterize various levels of academic achievement, a comprehensive distributional study of the most important engagement features was conducted. Figure 4 shows a series of boxplot-based visualizations that explore the distributions of raised hands, visited resource distribution, announcement views, and discussion participation amongst the three performance classes (Low, Medium, and High). This figure allows a close comparison of the central tendency and variability within and between performance groups, as it shows median values, interquartile ranges, class-wise means, and the data points of individual students. This type of analysis is crucial for establishing the extent to which student engagement behavior relates to academic success and for validating the discriminative relevance of these features before they can be used as predictors in predictive modeling.

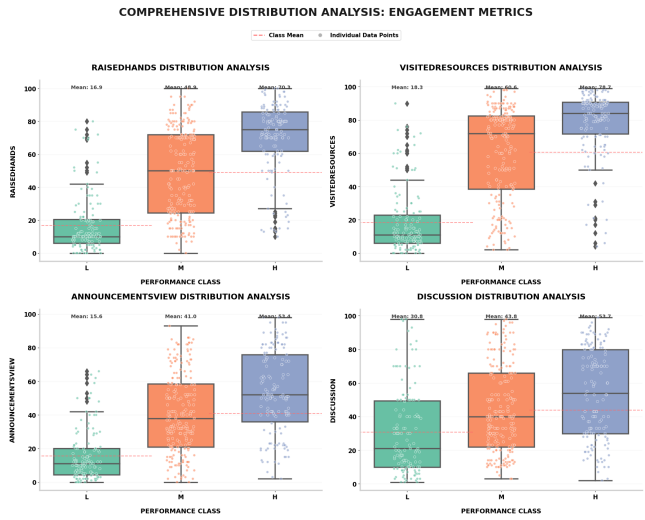

Figure 4. Distributional analysis of engagement metrics across performance classes.

3.3 Machine Learning Models

The research uses a wide range of machine learning models to assess students' performance, providing robust, generalizable results that capture the full range of student performance. The chosen models reflect various learning paradigms, including ensemble learning, boosting, neural networks, distance-based learning, margin-based classification, and probabilistic modeling. This diversity allows us to balance the comparisons and see that the suggested optimization strategy can be applied across heterogeneous model architectures.

A number of criteria were used when selecting machine learning models. First, models that proved effective in education data mining and student achievement prediction were considered. Second, algorithms capable of processing both numerical and categorical attributes were selected to match the nature of xAPI-Edu.Data data. Third, models sensitive to feature scaling, noise, and class imbalance were used to assess robustness across different data conditions. Lastly, interpretable models and high-capacity learners were sampled to balance predictive performance with explanatory potential, a necessity in educational decision-making.

3.3.1 Description of Models

Random Forest Random Forest is an ensemble learning model that builds a group of decision trees based on bootstrap sampling and random selection of features at each split. Majority voting between individual trees is used to get the final prediction. This will decrease variance, reduce overfitting, and improve generalization performance. Random Forest works especially well with structured data, and its strength, the ability to process large-dimensional data, and the automatic feature importance estimation have led to its massive use in educational analytics. The typical process used with Random Forest classification is shown in Algorithm 1[18, 19, 20].

Extreme Gradient Boosting (XGBoost) Extreme Gradient Boosting (XGBoost) is a state-of-the-art ensemble learning algorithm that relies on the gradient boosting paradigm and builds a series of decision trees in a cascade, minimizing the error of the successive ensemble. In contrast to traditional

Algorithm 1. Random Forest (RF)-based Classification

for $b = 1$ to B **do**
  Draw a bootstrap sample Z^* of size N.
  Grow a decision tree T_b on Z^* until the minimum node size $n_{\min}$ is reached.
  Randomly select m variables from the p available predictors.
  Choose the best split variable and split point among the m selected variables.
  Split the node into two child nodes.
end for
Output $\{T_b\}_{b=1}^{B}$ as the ensemble of trees.
For a new input point x, classify using majority voting:

$$\hat{C}_{\mathrm{rf}}^{B}(x) = \text{majority vote}\left\{\hat{C}_b(x)\right\}_{b=1}^{B},$$

where $\hat{C}_b(x)$ denotes the class prediction of the b-th tree.

boosting methods, the learning process in XGBoost is formulated as a regularized optimism problem. This design can help avoid overfitting, especially in high-dimensional feature spaces typical of educational data sets with demographic, behavioral, and academic variables.

One major advantage related to XGBoost is that it is implemented efficiently and scalably. It facilitates the construction of trees in parallel, cache-aware access patterns, and efficient processing of sparse and missing values, which can significantly reduce training time while still retaining high predictive accuracy. Also, subsampling strategies and the learning rate (shrinkage) further improve generalization by alleviating the effects of noise and less informative features. These features make XGBoost particularly useful for large-scale, heterogeneous data, where it is important to model complex nonlinear interactions and feature dependencies.

XGBoost has gained widespread adoption for predicting students' academic performance due to its robustness, flexibility, and strong empirical results. The algorithm can estimate complex associations between engagement metrics (e.g., engagement, resource use, and interaction frequency) and academic performance that are challenging to model in other spreadsheets or with single-tree models. Furthermore, XGBoost leverages feature importance, which may facilitate interpretability and the extraction of insights, enabling educators and decision-makers to identify the most significant elements of student success. Subsequently, XGBoost is an effective baseline and a competitive benchmark for assessing the advantages of sophisticated optimization and hybrid learning schemes in educational data mining.

Multilayer Perceptron (MLP) The Multilayer Perceptron (MLP) is a feedforward artificial neural network model comprising an input layer, one or more hidden layers, and an output layer, where each layer consists of a set of interconnected neurons with their respective weights and bias terms. The training is usually carried out using the backpropagation algorithm, coupled with gradient-based optimization methods, to minimize a specified loss function. MLPs can be used to model nonlinear functions such as ReLU, sigmoid, or tanh, which can be applied to more complex nonlinear transformations between the input features and the desired outputs. This ability to express has rendered MLPs especially appropriate for modelling complex relationships amid demographic, conductive, and educational features in educational knowledge bases. But their results are also sensitive to hyperparameter settings such as learning rate, network depth, neuron count, and regularization techniques, and tuning them to prevent overfitting can be difficult, particularly with limited training data.

Support Vector Machine (SVM) The support vector machine (SVM) is a supervised learning algorithm based on statistical learning theory, designed to construct an optimal decision boundary by maximizing the margin between classes. By using a few important training samples, called support vectors, SVMs can generalize well in high-dimensional feature spaces. SVMs can also learn highly nonlinear relationships by implicitly projecting the data into higher-dimensional spaces using kernel functions, e.g., radial basis functions (RBFs), polynomials, or linear kernels. SVMs are especially useful for educational performance prediction, where the number of features is large relative to the number of observations, because overfitting is less common in that setting. However, their performance may vary with careful selection of kernel parameters and regularization constants, which can significantly impact classification accuracy.

K-Nearest Neighbors (KNN) K-Nearest Neighbors (KNN) is a non-parametric, instance-based learning algorithm that uses the majority class of the unseen samples nearest to it in the feature space (the top k nearest) when classifying the samples. The distance metrics used to measure similarity between instances are usually Euclidean, Manhattan, or Minkowski. KNN also lacks an explicit training stage, making it easy to apply and adaptable to local data formats. KNN may be very useful for detecting patterns in educational datasets where students exhibit similar engagement or behavioral patterns. Its performance, however, is very sensitive to feature scaling, the distance metric, and the value of k. Also, KNN can be computationally expensive with large data sets and can perform poorly with noisy or high-dimensional data, where distance-based measures are less discriminative.

Gaussian Naïve Bayes Gaussian Naive Bayes is a probabilistic classification algorithm based on the Bayes theorem, assuming conditional independence among input variables and normal distributions for continuous variables. Although

this simplifies the assumption, the algorithm has shown competitive performance across a variety of real-world applications, especially when feature independence is approximately achieved. Gaussian Naive Bayes is computationally inexpensive, requires very little training data, and can be easily extended to high-dimensional spaces, making it appropriate for quickly bootstrapping a baseline model and for real-time educational analytics. It is also predictive in a fast and interpretable way; for example, in student performance prediction, it can provide insights into the contribution of individual features to the predictive value. Nevertheless, its strong independence assumption can limit its ability to model complex feature interactions, resulting in lower accuracy than more expressive models in situations where nonlinear dependencies are significant.

3.4 Metaheuristic Algorithms

Metaheuristic algorithms are high-level, nature-inspired optimization strategies developed to effectively search large, complex, and nonconvex search spaces where other optimization algorithms may not work or become computationally prohibitive. These algorithms are based on stochastic processes and adaptive search schemes that trade off exploration (global search) with exploitation (local refinement) to prevent early convergence to suboptimal solutions. Metaheuristics have been widely used across a broad range of applications, including feature selection, hyperparameter optimization, scheduling, and engineering optimization. They are especially useful in machine learning, where they can be used to search model hyperparameters with a very high-dimensional, nonlinear, and multimodal search space.

The present research applies metaheuristic algorithms to optimize the hyperparameters of the Random Forest classifier to enhance predictive accuracy in the classification of student academic performance. In contrast to grid or random search, which may become inefficient or incomplete as the number of hyperparameters grows, metaheuristic methods dynamically update their search paths based on population-based interactions and feedback from fitness. Other algorithms include the Al-Biruni Earth Radius Optimizer (BER), the Grey Wolf Optimizer (GWO), Particle Swarm Optimization (PSO), Genetic Algorithms (GA), and the Whale Optimization Algorithm (WOA), which use mimics of natural, physical, or social processes to guide the optimization process. Their population-based characteristics enable multiple candidate solutions to be explored simultaneously, thereby increasing the probability of obtaining near-optimal or globally optimal hyperparameter settings.

The reason for adopting metaheuristic optimization in this context is that ensemble learning models, particularly Random Forests, are highly sensitive to hyperparameter settings, including tree depth, the number of estimators, and feature sampling. Metaheuristic algorithms improve model generalization, robustness, and classification performance by systematically optimizing these parameters through adaptive search mechanisms. Moreover, due to their flexibility and scalability, they can be used in educational data mining applications, where datasets may exhibit non-homogeneous feature distributions and complex interaction patterns. Consequently, the introduced combination of metaheuristic algorithms will offer a principled and efficient way to improve machine learning performance while retaining computational viability.

3.4.1 Role of Metaheuristics in Hyperparameter Optimization

Hyperparameter optimization is a fundamental step in developing reliable machine learning models, as hyperparameters control model complexity, learning behavior, and generalization capability. Unlike model parameters that are learned directly from data, hyperparameters (e.g., number of trees in Random Forest, depth constraints, learning rate in boosting, kernel parameters in SVM, or neighborhood size in KNN) must be set externally and can substantially influence predictive performance. Exhaustive approaches such as grid search and random search may become computationally expensive as the dimensionality of the hyperparameter space grows, particularly when hyperparameter interactions are nonlinear and the search landscape contains multiple local optima.

Metaheuristic algorithms offer an automated and computationally efficient alternative for hyperparameter tuning by formulating the problem as a global optimization task. In this setting, each candidate solution represents a specific hyperparameter configuration, and a fitness function—typically derived from cross-validated performance metrics such as accuracy or F-score—guides the search process. Metaheuristics iteratively explore the hyperparameter space using stochastic operators and population-based mechanisms that balance exploration (broad search across diverse configurations) and exploitation (refinement around high-performing regions). This balance enables metaheuristics to escape local optima and to locate near-optimal configurations more effectively than purely deterministic procedures in complex, high-dimensional search spaces.

Furthermore, metaheuristic optimization can reduce human intervention in model development by systematically identifying high-performing hyperparameter settings without requiring manual trial-and-error. This property is particularly valuable in educational predictive modeling, where datasets may include heterogeneous demographic and behavioral variables, resulting in highly nonconvex objective functions. By improving the quality of hyperparameter selection, metaheuristics typically enhance model generalization, stabilize performance across different data splits, and increase robustness to noise and feature redundancy. Consequently, metaheuristic-driven hyperparameter optimization constitutes an effective strategy for strengthening machine learning models and improving their predictive accuracy in data-driven educational systems.

3.5 Proposed BER-based Optimization Algorithm

3.6 Al-Biruni Earth Radius Optimization (BER)

The optimization strategy adopted in this work is the Al-Biruni Earth Radius (BER) algorithm [7]. This metaheuristic method integrates exploration and exploitation mechanisms to navigate complex search spaces efficiently. The algorithm divides candidate solutions into two cooperative groups. The exploration group is responsible for broadly scanning the search space to identify promising regions that may contain near-optimal solutions. In contrast, the exploitation group focuses on intensifying the search around the currently best solution by applying strategies such as leader-oriented updates

Algorithm 2. BER-based optimization algorithm

Initialize BER agents $\vec{S}_i$ for $i = 1, 2, \ldots, d$, population size d, maximum iterations MaxIter, and fitness function F_n
Initialize BER parameters
Compute F_n for each $\vec{S}_i$
Identify the best agent $\vec{S}^*$
while $t \leq$ MaxIter **do**
  for each agent in the first group **do**
    Heading towards the best agent
    Compute $r = h\dfrac{\cos(x)}{1-\cos(x)}$
    Compute $\vec{D} = r_1\left(\vec{S}(t) - 1\right)$
    Update
$$\vec{S}(t+1) = \vec{S}(t) + \vec{D}\left(2r_2 - 1\right)$$
  end for
  for each agent in the second group **do**
    Apply elitism of the best agent
    Compute $\vec{D} = r_2\left(\vec{L}(t) - \vec{S}(t)\right)$
    Compute
$$\vec{S}_1(t+1) = r^2\left(\vec{S}(t) + \vec{D}\right)$$
    Investigate the region around the best agent
    Compute
$$\vec{k} = 1 + \frac{2t^2}{\text{MaxIter}}$$
    Compute
$$\vec{S}_2(t+1) = r\left(\vec{S}^*(t) + \vec{k}\right)$$
    Select the best candidate and set it as $\vec{S}^*$
    if best fitness did not change for two iterations **then**
      Mutate the agent
$$\vec{S}(t+1) = \vec{k}z^2 - h\frac{\cos(x)}{1-\cos(x)}$$
    end if
  end for
  Update F_n for each $\vec{S}$
end while
Return best agent $\vec{S}^*$

and localized investigation of neighboring solution regions.

A defining characteristic of BER is its adaptive mutation mechanism. This mechanism enables the algorithm to modify stagnating solutions based on a predefined fitness criterion when progress fails to occur over successive iterations. Such adaptive behavior improves the algorithm's flexibility and resilience, allowing it to escape local optima and continue progressing toward better solutions, particularly in difficult optimization landscapes. The step-by-step procedure of the BER optimization process is presented in Algorithm 2.

3.6.1 State-of-the-Art Metaheuristic Optimization Algorithms

To enhance the predictive performance of the employed machine learning models, several state-of-the-art metaheuristic optimization algorithms were adopted for hyperparameter tuning. These algorithms are inspired by natural and biological phenomena and are designed to efficiently explore complex and high-dimensional search spaces. Their population-based structures and adaptive search strategies make them particularly suitable for optimizing machine learning hyperparameters, where the objective landscape is often nonlinear, multimodal, and computationally expensive to evaluate. The following subsections provide a concise overview of the selected metaheuristic algorithms and highlight their key characteristics and application strengths.

Grey Wolf Optimizer (GWO) The Grey Wolf Optimizer is a metaheuristic algorithm inspired by the social hierarchy and cooperative hunting of grey wolves. The population is organized into five hierarchical groups: alpha, beta, delta and omega wolves, with the alpha wolf signifying the most effective solution discovered so far. The alpha, beta, and delta wolves provide collective leadership over the movement of the remaining agents in the search space, thereby optimizing it. GWO is a good compromise between exploration and exploitation due to its adaptive position-updating processes, and thus it can effectively be used to optimize machine learning hyperparameters, with relatively fast convergence and steady performance.

Particle Swarm Optimization (PSO) The Particle Swarm Optimization method is motivated by the flocking behavior

and socialization of birds or fish. In PSO, particles are candidate solutions that move in the search space, updating their speed and location based on their personal best and the global best solution found by the swarm. This two-way learning process allows effective information exchange between particles and a fast approach to optimal solutions. PSO is widely used in hyperparameter optimization due to its simplicity, low computational cost, and effectiveness for continuous optimization problems.

Genetic Algorithm (GA) A Genetic Algorithm is a computational method of optimization that is guided not only by natural selection principles but also by genetic principles. The evolved candidate solutions are encoded as chromosomes and undergo successive genetic transformations: selection, crossover, and mutation. Selection favors the fittest individuals, crossover recombines solution components to explore new areas, and mutation creates diversity to avoid early convergence. GA is especially useful for discrete and mixed hyperparameter spaces and has been widely used to optimize machine learning models with complex parameter interactions.

Whale Optimization Algorithm (WOA) The Whale Optimization Algorithm is based on the feeding behavior of humpback whales using the bubble net. It models the exploitation and exploration processes using spiral updating and encircling processes, guiding search agents to the best-known solution. The algorithm is adaptive, alternating between global search and local refinement, enabling it to efficiently navigate the optimization landscape. WOA has shown strong performance on nonlinear optimization problems and has been widely used in machine learning hyperparameter optimization, where a trade-off is needed between convergence rate and solution quality.

3.7 Evaluation Metrics

To assess the efficiency of the suggested machine learning models in predicting students' academic performance, a set of commonly used classification performance measures was used. These metrics are used to give a full evaluation of model behavior by evaluating overall accuracy, accuracy in individual classes, and accuracy-recall tradeoff. This evaluation scheme is especially relevant in educational datasets, where positive and negative class predictions carry high stakes for academic interventions and decision-making. Where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. A summary of the mathematical definitions and interpretations of the classification metrics to be considered in this study is presented in Table 3.

The joint application of these evaluation measures enables a comparative analysis of classification models in terms of balance and reliability. Whereas accuracy provides a general measure of efficiency, sensitivity and specificity estimate the ability to discriminate between classes. PPV and NPV are measures of prediction reliability, and the F-score indicates the trade-off between precision and recall. The combination of these measures will provide a robust evaluation of the model's performance, especially in educational prediction tasks where misclassification can have serious academic implications. Tables 4 and 5 Provide a summary of the experimental setup used in this study for both the optimization algorithms and the baseline classification models. These configurations should be presented to ensure the experiment setup is reproducible and transparent about the choices made, as parameters directly impact convergence behavior, computational efficiency, and predictive performance. Table 4 indicates the parameters of the metaheuristic optimization algorithms, which are BER, GWO, GA, PSO, and WOA. These parameters control key elements of the optimization process, including population size, exploration-exploitation ratio, mutation rate, and the number of iterations. The selected values were determined by recommendations in the literature and by preliminary empirical testing to provide a fair comparison of optimizers and to ensure stable convergence. The parameter settings for the parametric classifiers used in the experiments, such as KNN, SVM, Random Forest, XGBoost, and MLP, are given in Table 5. Each model is characterized by the listed hyperparameters that determine its learning behavior, complexity, and regularization. Setting up a stable base configuration for such classifiers provides a solid basis for assessing the effects of hyperparameter optimization via metaheuristics on predictive performance in subsequent analysis.

Table 5. Configuration parameters of the basic classification models [30].

Model	Parameter	Value
KNN	*n_neighbors*	5
	weights	uniform
	leaf_size	30
	p	2
SVM	*C*	1
	kernel	rbf
	penalty	*l*2
	tol	1.0×10^{-4}
RF	shrinkage	[0–1]
	max_depth	2
XGBoost	Learning rate	0.1
	max_depth	3
	colsample_bytree	0.7
MLP	Learning rate	0.2
	Momentum	0.1
	Threshold	0.001
	Number of epochs	10000

4. EMPIRICAL RESULTS

4.1 Baseline Machine Learning Performance (Before Optimization)

The subsection presents the initial classification achievement of the chosen machine learning models before any metaheuristic optimization strategy is implemented. It is necessary to determine baseline results to objectively assess each model's intrinsic predictive ability and to evaluate the effectiveness of the optimization in future experiments. The models were trained and assessed under the same experimental conditions, and the same preprocessing pipeline, feature set, and data partitions were used to ensure fairness and comparability. Widely recognized classification performance measures were used for the baseline evaluation, including accuracy, sensitivity (true positive rate), specificity (true negative rate), positive

Table 3. Classification Evaluation Metrics Used for Model Assessment

Metric	Mathematical Expression	Description
Accuracy	$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$	Measures the overall proportion of correctly classified instances among all predictions.
Sensitivity (TPR)	$\text{Sensitivity} = \frac{TP}{TP+FN}$	Also known as recall or true positive rate; quantifies the model's ability to correctly identify positive instances.
Specificity (TNR)	$\text{Specificity} = \frac{TN}{TN+FP}$	Measures the model's ability to correctly identify negative instances.
Positive Predictive Value (PPV)	$\text{PPV} = \frac{TP}{TP+FP}$	Also referred to as precision; indicates the proportion of correctly predicted positive instances.
Negative Predictive Value (NPV)	$\text{NPV} = \frac{TN}{TN+FN}$	Represents the probability that a negative prediction is correct.
F-Score	$\text{F-Score} = \frac{2\times TP}{2\times TP+FP+FN}$	Harmonic mean of precision and recall; balances false positives and false negatives.

Table 4. Configuration parameters of different optimization algorithms [29, 30].

Algorithm	Parameter	Value	Algorithm	Parameter	Value
BER	Iterations	80	GWO	a	2 to 0
	Mutation probability	0.5		Iterations	80
	Exploration percentage	70		Wolves	10
	k (decreases from 2 to 0)	1	WOA	r	[0,1]
GA	Crossover	0.9		Iterations	80
	Mutation ratio	0.1		Whales	10
	Mechanism	Roulette wheel	PSO	a	2 to 0
	Agents	10		Acceleration constants	[2,1]
	Iterations	80		Inertia W_{max}, W_{min}	[0.6,0.9]
				Particles	10
				Iterations	80

predictive value (PPV), negative predictive value (NPV), and F-score. Table 6 gives the comparative results of the Random Forest, XGBoost, Multilayer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Gaussian Naive Bayes. The baseline results indicate distinct performance differences between the assessed models. The accuracy and F-score are lower for Gaussian Naive Bayes and KNN, which is explained by their simplifying assumptions and their sensitivity to data distribution and feature scaling. Even though SVM and MLP have moderate performance, both models are sensitive to hyperparameters and model complexity.

XGBoost is recalcitrant because it uses a boosting-based structure and can model nonlinear relationships. Nevertheless, among all the models, the Random Forest classifier delivers the highest performance across almost all evaluation metrics. Random Forest has a balanced positive-to-negative classification with an accuracy of 88.89% and high sensitivity (0.8784), specificity (0.8982), and the highest F-score (0.8814).

The superiority of Random Forests can be explained by the fact that the model is built on the principle of ensemble learning, which helps eliminate variance through bootstrap aggregation and random feature selection, and prevents noise and overfitting. However, the number of trees, the maximum tree depth, and the sampling strategy of features are all important hyperparameters of the Random Forest performance. The above observations provide a strong rationale for using metaheuristic optimization algorithms to improve the predictive capabilities of Random Forest, which is investigated in later sections.

To provide a critical comparative evaluation of classification performance during the baseline, several evaluation metrics were applied across the selected machine learning models. Figure 5 Comparison of the performance of the Random Forest, XGBoost, Multilayer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Gaussian Naive Bayes based on their performance measured with six common metrics: accuracy, sensitivity (TPR), specificity (TNR), positive predictive value (PPV), negative predictive value (NPV) and F-score. Simultaneously visualizing these measures helps the figure point out the relative strengths and weaknesses of each model in terms of overall correctness, assigned class-discrimination capabilities, and the balance between precision and recall. This comparative visualization can then be used to intuitively analyze the models' behavior and provide empirical evidence on which model is most likely to be promising before metaheuristic optimization techniques are applied.

To explore the distributional properties of the baseline classification measures more deeply, while accounting for stability, a comparative analysis in the form of a box plot was performed using various learning models. Figure 6 shows the dispersion, central tendency, and relative variability of six performance

Table 6. Baseline Classification Performance of Machine Learning Models

Model	Accuracy	Sensitivity (TPR)	Specificity (TNR)	PPV	NPV	F-Score
Random Forest	0.8889	0.8784	0.8982	0.8844	0.8929	0.8814
XGBoost	0.8774	0.8611	0.8916	0.8732	0.8810	0.8671
MLP	0.8591	0.8451	0.8718	0.8571	0.8608	0.8511
SVM	0.8425	0.8286	0.8553	0.8406	0.8442	0.8345
KNN	0.8227	0.8060	0.8378	0.8182	0.8267	0.8120
Gaussian Naïve Bayes	0.8058	0.7879	0.8219	0.8000	0.8108	0.7939

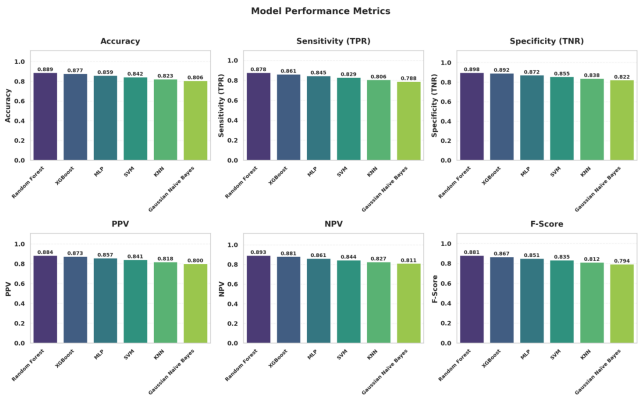

Figure 5. Comparison of baseline machine learning model performance across multiple evaluation metrics.

measures, including accuracy, sensitivity (TPR), specificity (TNR), positive predictive value (PPV), negative predictive value (NPV), and F-score, calculated on Random Forest, XGBoost, Multilayer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Gaussian Naïve Bayes. Inclusion of both mean and median reference lines is useful for directly comparing average performance with robust central tendency and, hence, for identifying potential differences in skewness and consistency across the models. Such visual analysis provides further empirical evidence for evaluating the reliability of the models and enables the selection of appropriate individuals to optimize using metaheuristic algorithms. To examine the correlation between the classification and detection performance of the machine learning models under evaluation, a bar-line visualization was used. Figure 7 is a twofold graph that shows the F-score and sensitivity (true positive rate) of Random Forest, XGBoost, Multilayer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Gaussian Naïve Bayes. The bar elements represent F-scores, which indicate the harmonic balance between precision and recall, and the overlaid line plot shows sensitivity, highlighting each model's ability to correctly detect positive cases. This combined visualization allows the overall classification effectiveness and the performance of class detection to be compared simultaneously, making it easier to assess model differences and helping informally select candidates for further optimization.

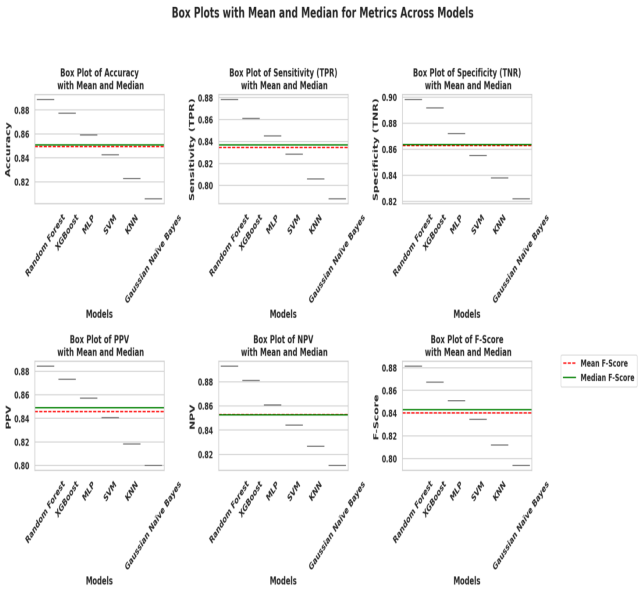

Figure 6. Box plots with mean and median for performance metrics across machine learning models.

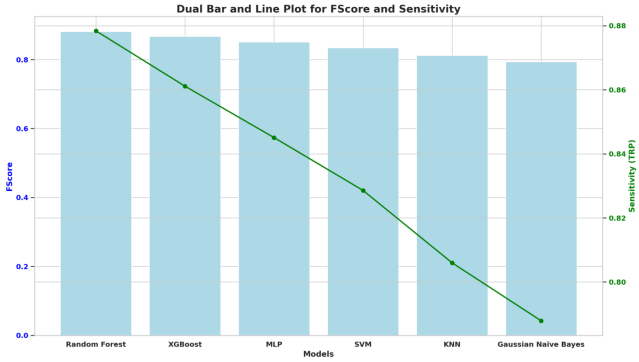

Figure 7. Dual bar and line plot comparing F-score and sensitivity across machine learning models.

To enable a concise yet inclusive comparison of baseline machine learning models across various evaluation criteria, a metric-wise heatmap will be used. Figure 8 shows a detailed heatmap of classification performance metrics, such as accuracy, sensitivity (TPR), specificity (TNR), positive predictive value (PPV), negative predictive value (NPV), and F-score, of Random Forest, XgBoost, Multilayer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Gaussian Naive Bayes. This visualization allows quickly identifying the strengths and weaknesses of models by encoding metric magnitudes as a color gradient and providing precise numerical annotations for each cell, while maintaining quantitative accuracy. This representation is particularly valuable for maintaining focus on the consistent performance across the metrics and for helping choose the most promising baseline classifier before metaheuristic optimization.

4.2 Optimized Model Analysis

In this subsection, a comparative study of the Random Forest model optimized using various state-of-the-art metaheuristic algorithms is presented. This analysis aims to compare the effectiveness of each optimization method in improving the Random Forest classifier's predictive accuracy and to identify the optimizer that best predicts student academic performance.

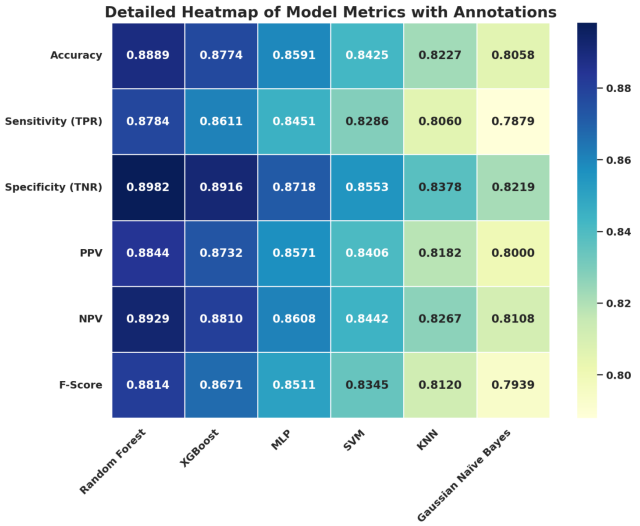

Figure 8. Heatmap visualization of baseline classification metrics across machine learning models.

The Randone Forest model was optimized individually using the Al-Biruni Earth Radius Optimizer (BER), Grey Wolf Optimizer (GWO), Particle Swarm Optimizer (PSO), Genetic Algorithm (GA), and Whale Optimization Algorithm (WOA). The same classification metrics were used to assess all optimized models: accuracy, sensitivity (TPR), specificity (TNR), positive predictive value (PPV), negative predictive value (NPV), and F-score. Table 7 summarizes the comparative results.

The results clearly demonstrate that all metaheuristic optimization algorithms improve the predictive performance of the Random Forest model when compared to its baseline configuration. However, the degree of improvement varies across optimization techniques. Among the evaluated methods, the BER-optimized Random Forest consistently achieves the highest performance across all evaluation metrics.

Specifically, BER + Random Forest attains the highest accuracy of 94.77%, along with superior sensitivity and specificity values, indicating a strong ability to correctly identify both high- and low-performing students. The elevated PPV and NPV further confirm the reliability of positive and negative predictions, while the highest F-score reflects an optimal balance between precision and recall. These results highlight the robustness and generalization capability of the BER-optimized model.

The superior performance of BER can be attributed to its balanced exploration–exploitation strategy and adaptive mutation mechanism, which enable efficient navigation of the hyperparameter search space and avoidance of local optima. In contrast, although GWO and PSO also demonstrate strong optimization capability, their performance remains slightly inferior to BER. GA and WOA yield comparatively lower improvements, which may be due to slower convergence or less effective exploitation of promising regions in the search space.

Overall, this analysis confirms that the BER algorithm is the most effective metaheuristic optimizer for enhancing Random Forest performance in the context of student academic performance prediction. Consequently, the BER-optimized Random Forest model is identified as the proposed optimal predictive framework in this study.

A comparative analysis of the distributional behavior and consistency of the performance metrics of the models used in the evaluation was further analyzed using a kernel density estimation (KDE) based approach. The estimated probability density functions of six important evaluation metrics, such as accuracy, sensitivity (TPR), specificity (TNR), positive predictive value (PPV), negative predictive value (NPV), and F-score, were estimated using the machine learning models of interest and plotted in Figure 9. Using the visualization of the continuous distributions of these measures, the figure allows evaluation of the central tendency, spread, and overlap among the measures of performance, which in turn provides insight into metric stability and variability across the models. This visualization is based on density and complements pointwise and aggregate analyses to reveal less obvious differences in metric concentration and to aid a more nuanced comparison of model reliability before and after the optimization process.

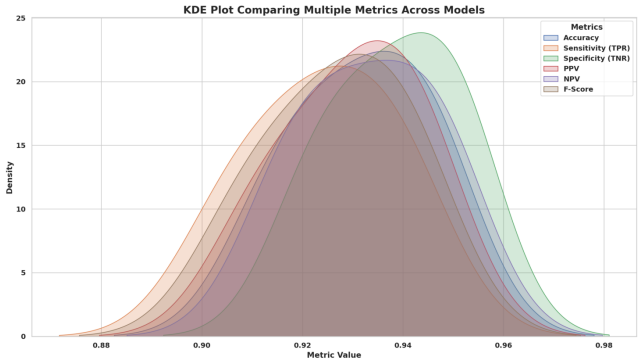

Figure 9. Kernel density estimation plots comparing multiple performance metrics across models.

To examine the relative performance progression of the optimized Random Forest models under different metaheuristic strategies, a time-series–style comparative visualization was employed. Figure 10 presents the sequential variation of key evaluation metrics—accuracy, sensitivity (TPR), specificity (TNR), positive predictive value (PPV), negative predictive value (NPV), and F-score—across Random Forest models optimized using BER, GWO, PSO, GA, and WOA. Although the models do not represent temporal evolution in the traditional sense, the ordered visualization facilitates clear comparison of performance trends as the optimization strategy changes. This representation highlights the consistent superiority of the BER-optimized Random Forest and illustrates the gradual performance decline observed with alternative optimizers, thereby reinforcing the effectiveness of BER in enhancing predictive reliability.

To provide a normalized and comparative assessment of the optimized Random Forest models, a Z-score–based performance analysis was conducted across multiple evaluation metrics. Figure 11 presents a heatmap of standardized scores for accuracy, sensitivity (TPR), specificity (TNR), positive predictive value (PPV), negative predictive value (NPV), and F-score obtained from Random Forest models optimized using BER, GWO, PSO, GA, and WOA. By transforming metric values into Z-scores, the visualization highlights relative deviations from the overall mean performance, enabling clear identification of consistently superior or inferior optimization strategies. This representation emphasizes the strong positive

Table 7. Performance Comparison of Optimized Random Forest Models

Model	Accuracy	Sensitivity (TPR)	Specificity (TNR)	PPV	NPV	F-Score
BER + Random Forest	**0.9477**	**0.9420**	**0.9526**	**0.9457**	**0.9494**	**0.9439**
GWO + Random Forest	0.9418	0.9337	0.9487	0.9394	0.9439	0.9366
PSO + Random Forest	0.9325	0.9250	0.9392	0.9317	0.9332	0.9283
GA + Random Forest	0.9234	0.9158	0.9303	0.9226	0.9242	0.9192
WOA + Random Forest	0.9129	0.9040	0.9208	0.9101	0.9154	0.9071

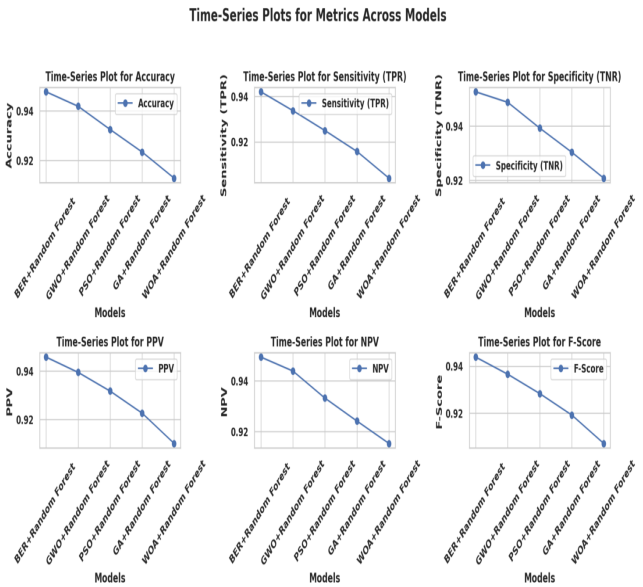

Figure 10. Time-series–style plots comparing evaluation metrics across optimized Random Forest models.

deviation of the BER-optimized Random Forest across all metrics and contrasts it with the progressively lower standardized performance observed for alternative optimizers, thereby reinforcing the robustness and dominance of BER in the proposed optimization framework.

Figure 11. Z-score heatmap of performance metrics for optimized Random Forest models.

To complement the aggregate and distribution-based analyses of optimized model performance, a swarm plot–based visualization was employed to illustrate metric-level dispersion across optimization strategies. Figure 12 presents swarm plots of six evaluation metrics—accuracy, sensitivity (TPR), specificity (TNR), positive predictive value (PPV), negative predictive value (NPV), and F-score—for Random Forest models optimized using BER, GWO, PSO, GA, and WOA. Each subplot displays the relative positioning of metric values for the optimized models, allowing direct visual comparison of performance consistency and separation among optimizers. This visualization highlights the clustering of higher metric values for the BER-optimized Random Forest and reveals clear performance gaps with respect to alternative metaheuristic approaches, thereby reinforcing the robustness and dominance of BER-based optimization.

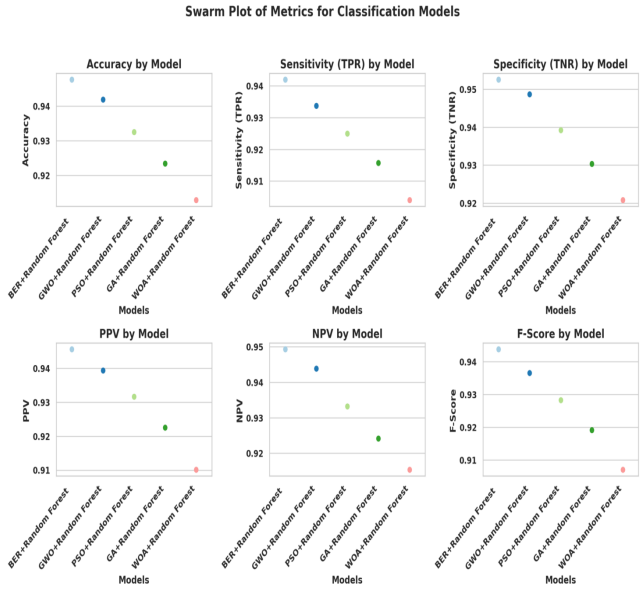

Figure 12. Swarm plots of performance metrics for Random Forest models optimized using different metaheuristic algorithms.

4.3 Computational Analysis

Besides predictive accuracy, computational efficiency is also a very essential consideration in determining the feasibility of machine learning models, especially in large-scale or real-time education systems. The computational overhead directly impacts model deployability, scalability, and energy usage, particularly when using metaheuristic optimization tools. Consequently, this subsection presents a comparative computational study of the optimized Random Forest models to evaluate their runtime, memory usage, and processor usage. Table 8 indicates the mean execution time (in seconds), memory consumption (in megabytes), and CPU consumption (in percentage) of the Random Forest models that are optimized with the help of various metaheuristic algorithms. All experiments have been done in the same hardware and software settings to promote fairness and reproducibility.

The computational results clearly indicate that the BER-optimized Random Forest model not only achieves superior predictive performance but also exhibits the highest computational efficiency among all evaluated approaches. Specifically, BER + Random Forest records the lowest average execution time, minimal memory consumption, and the least CPU uti-

Table 8. Computational Performance Comparison of Optimized Models

Algorithm	Avg. Time (s)	Memory Usage (MB)	CPU Usage (%)
BER + Random Forest	**38.67**	**1038.21**	**55.13**
GWO + Random Forest	52.53	1525.05	64.26
PSO + STGCN	60.29	1599.28	75.32
PSO + Random Forest	65.32	1727.55	79.27
GA + Random Forest	69.57	1877.64	85.22
WOA + Random Forest	72.42	1922.54	89.84

lization. This demonstrates BER's ability to converge rapidly toward high-quality solutions while maintaining efficient resource usage.

In contrast, optimization techniques such as GA and WOA incur significantly higher computational costs, reflected in increased runtime, memory usage, and processor load. These overheads are primarily attributed to their more complex evolutionary operators and slower convergence characteristics. PSO-based approaches also exhibit higher computational demands, particularly when integrated with more complex model architectures, as evidenced by the PSO + STGCN configuration.

Overall, this computational analysis reinforces the suitability of the BER-based optimization framework for real-world educational applications. By delivering both superior predictive accuracy and reduced computational overhead, the BER-optimized Random Forest model offers a balanced and scalable solution for deployment in data-driven academic performance monitoring and early warning systems.

In order to further supplement the quantitative computational analysis performed in Table 8, a visual resource distribution analysis is given in order to intuitively compare the computational efficiency of various strategies of optimization. Figure 13shows that execution time, memory usage, CPU utilization, and an integrated efficiency score for the optimized Random Forest models were distributed proportionally across the various metaheuristic algorithms. By displaying them as donut charts, the figure allows a global assessment of how each optimization method uses system resources. This visualization shows the trade-off between predictive enhancement and computational cost, with strong support for the practical benefits of lightweight but effective optimizers, specifically the BER-based Random Forest, which proves to be more efficient with minimal resource overhead across all measured dimensions. To provide a general impression of computational efficiency and predictive effectiveness, a specific algorithm performance dashboard is presented in Figure 14. This figure integrates several performance factors, including execution time, memory consumption, CPU consumption, performance score, overall cost index, and overall performance index, into a single comparative visualization system. The dashboard provides an intuitive analysis of the trade-offs between the discussed optimization strategies, based on the measurement of indicators: the low-value indicators (time, memory, CPU usage, and cost index) are more desirable, while the high-value indicators (efficiency and performance indices) predict better performance. As demonstrated in Figure 14, the BER-sensitive random forest is always associated with positive computational characteristics and, simultaneously, the most desirable efficiency and performance indicators; it can be applied to large-scale resource-constrained predictive analytics.

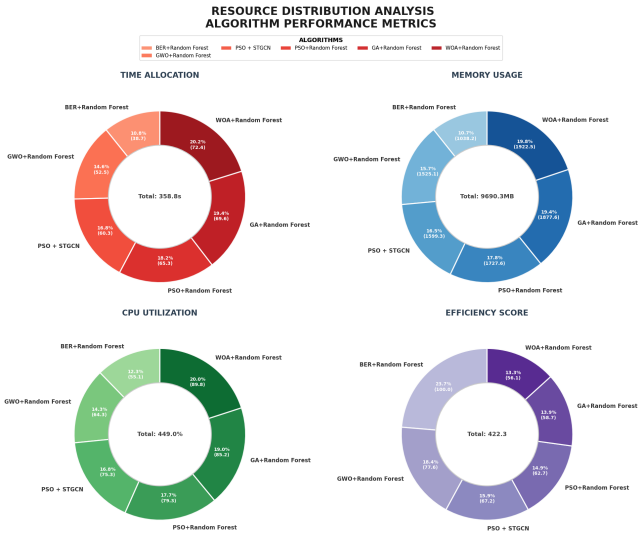

Figure 13. Resource distribution analysis of computational performance metrics across optimized models.

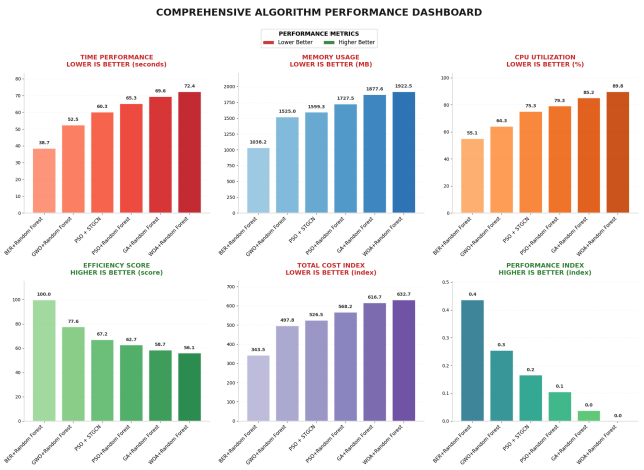

Figure 14. Comprehensive algorithm performance dashboard comparing computational and efficiency-related metrics.

A performance trend analysis is presented in Figure 15 to assess the computational properties of the optimized models in a trend-oriented, relative manner. The figure will present variations in three critical computation indicators, which are the execution time, the consumption of memory as well as the utilization of the CPU, amongst the strategies adopted in optimization and the classifier that is the Random Forest. The graph, which shows these measures simultaneously, reveals differences in computational cost across metaheuristic algorithms. Figure 15 indicates that the BER-optimal Random Forest is the shortest in terms of execution time, has the lowest memory and CPU requirements, and resource consumption is observed to increase with GWO-, PSO-, GA-, and

WOA-based models. The given trend-based representation offers a good insight into the trade-offs between optimization complexity and computational efficiency, and thus, it can be concluded that BER can be extended to scalable, resource-efficient educational predictive modeling.

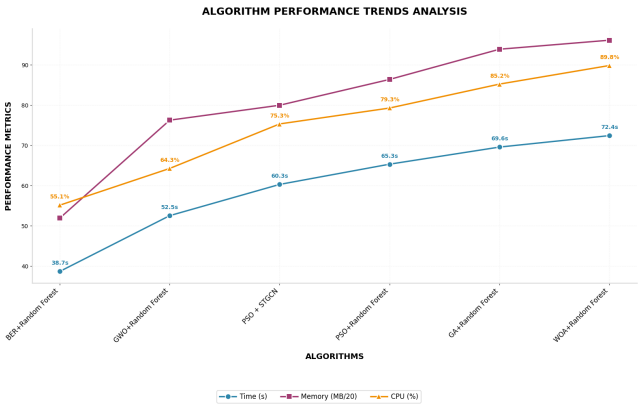

Figure 15. Algorithm performance trends based on time, memory, and CPU utilization.

5. DISCUSSION

The empirical results of this research provide solid evidence that modern machine learning tools can be used effectively to predict students' academic performance, provided they are supported by rigorous preprocessing and systematic evaluation. The comparison against the baseline of Random Forest, XGBoost, MLP, SVM, KNN, and Gaussian Naive Bayes has revealed clear differences in predictive capacity, indicating differences in inductive bias, representational capacity, and optimization behavior. Specifically, the higher baseline performance achieved by ensemble-based learners is consistent with the established benefit of aggregation mechanisms in minimizing variance and improving generalization, particularly on noisy and heterogeneous datasets that integrate demographic, academic, and behavioral factors. As an educational analytics issue, the importance of this outcome lies in the fact that student performance is seldom determined by a single factor; rather, it is a product of interacting factors, including engagement patterns, attendance, and contextual variables. This interpretation was supported by the exploratory visual analytics, which demonstrated organized behavioral patterns, both between indicators of engagement (i.e., raised hands, participation in discussions, visiting resources and watching announcements), and that the variables are rich in discriminative data, which can be used to draw a boundary between achievement classes. Furthermore, the categorical heatmap analysis showed that signals of parental involvement (survey responses, satisfaction) and the level of absenteeism are associated with significant changes in class distributions, suggesting that both in-class conduct and external support circumstances influence academic performance.

All these baseline findings and exploratory analyses validate that the xAPI-Edu.The data set contains informative signals spread across several features, and more machines capable of capturing nonlinear relationships and feature interactions are better positioned to leverage these signals to achieve strong academic classification performance. One of the main methodological contributions of the work is that the metaheuristic optimization achieves objective, reproducible improvements over baseline model configurations, especially for the Random Forest classifier. Although Random Forest has achieved the best baseline performance (accuracy $= 0.8889$ and F-score $= 0.8814$), the predictive quality of this algorithm is highly sensitive to important hyperparameters such as the number of trees, maximum depth, and feature sampling rate. These parameters contribute to the bias-variance trade-off, ensemble diversity, and the model's ability to prevent overfitting, which makes Random Forest a perfect choice for a systematic hyperparameter search. The comparative optimizer analysis revealed that each of the considered metaheuristics (BER, GWO, PSO, GA, and WOA) outperformed the baseline, but the levels of improvement varied significantly.

It is worth noting that BER achieved the highest significant gains, yielding the highest overall performance (accuracy $= 0.9477$ and F-score $= 0.9439$) and also improving sensitivity and specificity. It is a critical consideration: the effect of optimality was not limited to one measure, but rather improved balanced classification behavior in complementary measures of correctness and discrimination. This kind of balanced improvement implies that the BER found hyperparameter sets that enhanced not only positive-class detection (which is crucial for identifying at-risk students), but also negative-class discrimination (which is crucial for preventing undesirable or misplaced interventions). In theory, these results can be explained by the search dynamics of BER, which involve a coordinated exploration-exploitation process and an adaptive mutation process that help eliminate stagnation and motivate the search to leave local optima. The findings, therefore, highlight the practical applications of bio-inspired optimization as an effective alternative to exhaustive approaches, especially in environments where the hyperparameter space is multimodal, and cross-validation is limited by computational budget.

The convergence of results from various complementary visual analyses further supported the discussion and demonstrated that BER outperforms its competitors, allowing the optimizer comparisons to be placed in context rather than being limited to tabulated results. KDE visualization was a distributional view of metric performance, showing that the values of optimistic performance by using BER have good concentration and less overlap with worse optimization results. The time-series-style plots, although not an indication of temporal evolution in the strict sense, provided a structured order that brought the monotonic trend in performance across the optimization procedures into higher relief, supporting the interpretive argument that the optimization strategy systematically influences the performance envelope that may be achieved. Besides that, the Z-score heatmap provided a normalized, interpretable comparison of metrics, based on the deviations of each optimizer from the global mean; the overall positive deviation of BER across all metrics was a strong indication that the difference is not metric-specific but rather holistic. Alternatively, swarm plots focused on metric-level separation, showing dispersion and clustering data, indicated that BER and WOA were in clear, high-performing clusters, with GA in a relatively low-performing arrangement across various metrics. Notably, this consistency also has practical consequences for classroom implementation: predictive

models deployed in early-warning systems must be consistent across metrics, since educational decisions should not be based solely on overall accuracy but also on the consistency in detecting students who need support and in reducing false alarms. Thus, the analytical design used in this study is multi-view, which provides a strong empirical basis for concluding that BER-based tuning enhances the level and reliability of predictive performance.

Although these strengths are strong, a few considerations should be recognized to put the findings into perspective and support responsible interpretation. First, the sample size (480 cases) and breadth (two courses) can limit external validity, and predictive performance may differ when offered in other institutional settings, with different curricula, or with more diverse groups of learners. Second, even though the chosen metrics offer a demanding assessment of the quality of classifications, in real educational settings, they might need more assessment dimensions, such as fairness-conscious analysis of demographic subgroups, calibration assessment to interpret the predicted probabilities, and cost-sensitive analysis since false negatives (a student missing an at-risk student) and false positives (flagging a student unnecessarily) can have disproportionate consequences. Third, although the experimental design was based on a primary optimization of Random Forest because of its baseline performance, the question of whether BER can be seen to scale to other learners (i.e., boosting models and neural networks) with similar tuning budgets or whether some model families are more sensitive to different optimizers is an open empirical question.

Lastly, computation costs may also affect deployment: although metaheuristics not only minimize manual tuning but also incur extra runtime costs compared to fixed baselines, it is necessary to assess the gains of optimization in practice, including resource usage, which will be essential. However, in the specified experimental context, the findings reveal solid, convergent evidence that bio-inspired metaheuristic optimization, in particular, can significantly improve predictive modeling in academic data mining and provide a valid approach to adaptive, reliable, and data-driven academic support systems.

6. CONCLUSION AND FUTURE WORK

This paper explores how machine learning and metaheuristic optimization can be used to predict students' academic performance using xAPI-Edu.Data educational mining dataset. An entire experimental pipeline was built, including data preprocessing (missing value imputation, categorical encoding, feature scaling, and correlation analysis), exploratory visual analytics, benchmarking the baseline model, and hyperparameter optimization using metaheuristics. The baseline conditions included the evaluation of 6 classifiers commonly used, including Random Forest, XGBoost, MLP, SVM, KNN, and Gaussian Naive Bayes, which was conducted to give them a fair performance baseline. The base results indicate that Random Forest had the best overall performance among the models, with an accuracy of 0.8889 and an F-score of 0.8814; therefore, it is encouraging to use the model as the main candidate for optimization.

Random Forest was then optimized with five state-of-the-art metaheuristic algorithms: BER, GWO, PSO, GA and WOA. The optimized performance indicated that metaheuristic tuning consistently outperforms baseline settings, but the best improvement varied across optimizers. The BER-optimized Random Forest provided the best results across all classification metrics, achieving the highest accuracy (0.9477), sensitivity (0.9420), specificity (0.9526), PPV (0.9457), NPV (0.9494), and F-score (0.9439). The observed stability of BER was also supported by complementary visualizations, such as KDE distributions, time-series-like comparisons, Z-score heatmaps, and swarm plots, which indicated the strength and stability of BER relative to other optimizers. All in all, the results indicate that BER offers a balanced exploration and exploitation and is an exceptionally strong metaheuristic approach to improving the performance of the Random Forest in educational prediction. Even though the suggested BER-optimized Random Forest framework showed good predictive performance, there are still several directions for further investigation. To assess the generalizability of the framework across different learning settings and demographic distributions, future research can expand the framework to include more heterogeneous learning datasets (e.g., multi-institutional or longitudinal cohorts). Secondly, since performance prediction in students may involve multi-class effects (e.g., low, medium, high attainment), future research should explore multi-class learning formulations and cost-efficient evaluation approaches to more clearly represent the educational consequences of various misclassification rates.

Third, the current study can be strengthened by adding explainable artificial intelligence (XAI) methods, such as SHAP or LIME, to enhance interpretability and provide educators and policymakers with useful insights into which behavioral and academic factors have the largest overall impact on predictions. Fourth, more sophisticated metaheuristic and hybrid optimization models (e.g., multi-objective versions, parameter control, or hybrid BER with local search) can be considered to further accelerate convergence, stability, and predictability. Last but not least, early-warning deployment mechanisms embedded in the learning management systems are another significant practical extension, in which optimized models would facilitate real-time academic intervention, customized learning advice, and more equitable resource allocation by detecting at-risk students during the initial phases of the learning process.

REFERENCES

[1] D. R. Vora and K. Rajamani, "A hybrid classification model for prediction of academic performance of students: A big data application," *Evolutionary Intelligence*, vol. 15, no. 2, pp. 1083–1096, 2022.

[2] J. C. Gámez-Granados, A. Esteban, F. J. Rodriguez-Lozano, and A. Zafra, "An algorithm based on fuzzy ordinal classification to predict students' academic performance," *Applied Intelligence*, vol. 53, no. 22, pp. 27 537–27 559, 2023.

[3] H. Pallathadka, A. Wenda, E. Ramirez-Asís, M. Asís-López, J. Flores-Albornoz, and K. Phasinam, "Classification and prediction of student performance data using various machine learning algorithms," *Materials Today: Proceedings*, vol. 80, pp. 3782–3785, 2023.

[4] K. Yurtkan, A. Adalier, and T. E. K. G. Umut, "Student success prediction using feedforward neural networks," *Romanian Journal of Information Science and Technology*, vol. 2023, no. 2, pp. 121–136, 2023.

[5] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, p. 160, 2021.

[6] E.-S. M. El-Kenawy, S. Mirjalili, A. A. Abdelhamid, A. Ibrahim, N. Khodadadi, and M. M. Eid, "Metaheuristic optimization and keystroke dynamics for authentication of smartphone users," *Mathematics*, vol. 10, no. 16, 2022.

[7] E.-S. El-Kenawy, A. Abdelhamid, A. Ibrahim, S. Mirjalili, N. Khodadad, A. Alhussan, and D. Khafaga, "Albiruni earth radius (ber) metaheuristic search optimization algorithm," *Computer Systems Science and Engineering*, vol. 45, no. 2, pp. 1917–1934, 2022.

[8] Q. Al-Tashi, H. Md Rais, S. J. Abdulkadir, S. Mirjalili, and H. Alhussian, "A review of grey wolf optimizer-based feature selection methods for classification," in *Evolutionary Machine Learning Techniques*. Springer, 2020, pp. 273–286.

[9] A. G. Gad, "Particle swarm optimization algorithm and its applications: A systematic review," *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 2531–2561, 2022.

[10] E. A. Amreih, T. Hamtini, and I. Aljarah, "Student's academic performance dataset (xapi-edu-data)," Kaggle, 2023.

[11] S. Guo, H. B. A. Halim, and M. R. B. M. Saad, "Leveraging ai-enabled mobile learning platforms to enhance the effectiveness of english teaching in universities," *Scientific Reports*, vol. 15, no. 1, 2025.

[12] E. Hussein, M. A. Hussein, and M. Al-Hendawi, "Investigation into the applications of artificial intelligence in special education," *Social Sciences*, vol. 14, no. 5, p. 288, 2025.

[13] J. T. K. Phua, H. F. Neo, and C.-C. Teo, "Evaluating the impact of artificial intelligence tools on enhancing student academic performance," *Big Data and Cognitive Computing*, vol. 9, no. 5, p. 131, 2025.

[14] G. A. Anghel, C. M. Zanfir, F. L. Matei, C. D. Voicu, and R. A. Neacșa, "The integration of artificial intelligence in academic learning practices," *Education Sciences*, vol. 15, no. 5, p. 616, 2025.

[15] H. Yaseen, A. S. Mohammad, N. Ashal, H. Abusaimeh, A. A. A. Ali, and A. A. Sharabati, "The impact of adaptive learning technologies and ai tools on student engagement," *Sustainability*, vol. 17, no. 3, p. 1133, 2025.

[16] C. d. R. Navas-Bonilla, J. A. Guerra-Arango, D. A. Oviedo-Guado, and D. E. Murillo-Noriega, "Inclusive education through technology: a systematic review of types, tools and characteristics," *Frontiers in Education*, vol. 10, 2025.

[17] W. Walters, W. Barber, and M. Jutras, "The consolidated framework for implementation research: Application to education," *Education Sciences*, vol. 15, no. 5, p. 613, 2025.

[18] M. Yağcı, "Educational data mining: Prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, p. 11, 2022.

[19] B. Cheng, Y. Liu, and Y. Jia, "Evaluation of students' performance using xgboost classifier-enhanced aeo hybrid model," *Expert Systems with Applications*, vol. 238, p. 122136, 2023.

[20] S. B. Keser and S. Aghalarova, "Hela: A novel hybrid ensemble learning algorithm for predicting academic performance," *Education and Information Technologies*, vol. 27, no. 4, pp. 4521–4552, 2022.

[21] E. T. Lau, L. Sun, and Q. Yang, "Modelling, prediction and classification of student academic performance using neural networks," *SN Applied Sciences*, vol. 1, no. 9, p. 982, 2019.

[22] B. K. Francis and S. S. Babu, "Predicting academic performance using a hybrid data mining approach," *Journal of Medical Systems*, vol. 43, no. 6, p. 162, 2019.

[23] G. Deeva, J. De Smedt, C. Saint-Pierre, R. Weber, and J. De Weerdt, "Predicting student performance using sequence classification with time-based windows," *Expert Systems with Applications*, vol. 209, p. 118182, 2022.

[24] P. Nayak, S. Vaheed, S. Gupta, and N. Mohan, "Predicting students' academic performance using machine learning," *Education and Information Technologies*, 2023.

[25] B. Yt and S. Rk, "Predictive modeling and analytics of students' grades," *Education and Information Technologies*, vol. 28, no. 3, 2023.

[26] A. Khan, S. K. Ghosh, D. Ghosh, and S. Chattopadhyay, "Random wheel: An algorithm for early classification of student performance," *Engineering Applications of Artificial Intelligence*, vol. 102, p. 104270, 2021.

[27] M. e. a. Khosravi, "A comprehensive review of ai-based student performance prediction techniques," *Computers & Education: Artificial Intelligence*, vol. 2, p. 100019, 2021.

[28] W. e. a. Dissanayake, "Bayesian hyperparameter optimization for predicting academic performance," *Applied Sciences*, vol. 11, no. 7, p. 3194, 2021.

[29] A. A. e. a. Alhussan, "Classification of diabetes using hybrid ber and dto," *Diagnostics*, vol. 13, no. 12, p. 2038, 2023.

[30] E.-S. M. e. a. El-Kenawy, "Optimizing potato disease classification using metaheuristics," *Potato Research*, vol. 68, no. 1, pp. 551–585, 2025.