Elevating Student Performance Prediction using Extra-Trees Classifier and Meta-Heuristic Optimization Algorithms

Yangbo Li[1]*, Mengfan He[2]
Department of Computer Science and Technology, Henan Institute of Technology, Xinxiang Henan, 453003, China[1]
Office of the President, Henan Institute of Technology, Xinxiang Henan, 453003, China[2]

***Abstract*—In the highly competitive landscape of academia, the study addresses the multifaceted challenge of analyzing voluminous and diverse educational datasets through the application of machine learning, specifically emphasizing dimensionality reduction techniques. This sophisticated approach facilitates educators in making data-informed decisions, providing timely guidance for targeted academic improvement, and enhancing the overall educational experience by stratifying individuals based on their innate aptitudes and mitigating failure rates. To fortify predictive capabilities, the study employs the robust Extra-Trees Classifier (ETC) model for classification tasks. This model is enhanced by integrating the Gorilla Troops Optimizer (GTO) and Reptile Search Algorithm (RSA), cutting-edge optimization algorithms designed to refine decision-making processes and improve predictive precision. This strategic amalgamation underscores the research's commitment to leveraging advanced machine learning and bio-inspired algorithms to achieve more accurate and resilient student performance predictions in the mathematics course, ultimately aiming to elevate educational outcomes. Analyses of G1 and G3 showcase the efficacy of the ETRS model, demonstrating 97.5% Accuracy, F1-Score, and Recall in predicting the G1 values. Similarly, the ETRS model emerges as the premier predictor for G3, attaining 95.3% Accuracy, Recall, and F1-Score, respectively. These outcomes underscore the significant contributions of the proposed models in advancing precision and discernment in student performance prediction, aligning with the overarching goal of refining educational outcomes.**

Keywords—Student performance; mathematics; machine learning; Extra-Trees Classifier; Gorilla Troops Optimizer; Reptile Search Algorithm

I. Introduction

A. Background

The achievement of academic success by students is a core aim in education and a crucial component of any country's educational agenda. Emphasizing the significance of quality education as a driver for societal transformation, educational institutions are compelled to give priority to the development of students who excel not only in academic and non-academic evaluations but also acquire vital practical skills to remain competitive in the job market. Education, central to societal progress, reflects the shared aspirations for well-being and advancement [1]. The emphasis on the caliber of students graduating from schools has emerged as a significant worry. As underscored by Spinath [2], [3], academic success occupies a central position, serving as a gauge for intellectual education and an essential requirement for personal and societal well-being. In this context, Martín asserts that academic achievement goes beyond intellectual quotient (IQ), encompassing diverse dimensions to encompass the cognitive, psychomotor, and affective aspects of students' development [4], [5], [6].

The main advantage of data mining is its capability to meticulously analyze large datasets and formulate rules that can attract the interest of pertinent stakeholders. Additionally, it has the potential to unveil previously unknown and valuable insights that significantly enhance decision-making. Machine learning (ML) algorithms, particularly noted for their efficacy in classification tasks, stand as a focal point in various research pursuits [7], [8], [9]. As per the findings of Sharma, Himani, and Kumar [10], decision tree algorithms are widely acknowledged as effective tools for classification purposes. Decision trees, which are structured models comprising root nodes, branches, and leaf nodes, serves the functions of predicting outcomes. These trees exhibit versatility in handling both numerical and categorical data, are easily comprehensible, and can be visually represented. Their pivotal role extends to the identification of group characteristics, exploration of relationships between variables, and application in predicting various educational outcomes, including student performance. Jorda and Raqueno [11] underscore the significance of diverse decision tree algorithms such as C&R Tree, CHAID, C 5.0, and QUEST, emphasizing their role in the development of classification systems [12], [13], [14].

B. Related Works

Many scholars have conducted thorough investigations into the diverse factors that impact student success across different academic levels [15], [16], [17], [18]. Numerous studies in this realm have employed data mining techniques, specifically classification algorithms, to improve the overall quality of higher education systems and to forecast student performance. This section highlights a selection of pertinent studies; particularly those centered on the utilization of decision trees and classification methods in assessing students' academic performance [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29].

As an illustration, Al-Radaideh et al. [30] conducted a study proposing a decision tree classification model aimed at assisting school management in selecting appropriate academic tracks for students, thereby streamlining decision-making processes. Thammasiri et al. [31] introduced a model designed to predict inadequate academic performance among freshmen. By combining support vector machines with SMOTE (Synthetic Minority Over-sampling Technique), they achieved an impressive accuracy rate of 90.24%, effectively addressing class imbalance issues. Mustafa et al. [32] utilized the CRISP framework to assess students' data in C++ courses, conducting a comparative analysis of classifiers such as ID3, C4.5 decision trees, and Naive Bayes. The C4.5 decision tree exhibited superior performance, offering insights into the attributes influencing student performance. Nguyen and Peter [33] investigated the efficacy of decision trees and Bayesian networks in forecasting the academic performance of undergraduate and postgraduate students, with their findings indicating the superior performance of decision trees. Sunita and LOBO L.M.R.J [34] showcased the practicality of data mining in the realm of education by employing classification and clustering algorithms to predict student performance and group students accordingly. Bichkar and R. R. Kabra [35] developed classification models geared towards identifying at-risk students among first-year engineering students. Bharadwaj and Pal [36] applied the ID3 decision tree algorithm to predict student divisions based on various academic indicators. Edin Osmanbegovic et al. [37] formulated a model aimed at predicting student academic success, specifically addressing challenges related to data dimensionality. Despite Naïve Bayes achieving the highest accuracy at 76.65%, the model fell short in effectively handling the class imbalance issue. Surjeet and Pal [38] utilized various decision tree algorithms to forecast the performance of first-year engineering students, focusing on the identification of those at risk of failure. Mahfuza and Shovon [13] proposed a hybrid approach that combines clustering and classification to categorize students into high, medium, and low standards, enabling informed decisions about their academic performance and ultimately enhancing their final examination results. Kabakchieva [39] compared data mining algorithms for predicting student performance and classifying students as strong or weak, with the neural network achieving high accuracy for the strong class. Carlos et al. [40] employed machine learning to create a model for predicting student failure, achieving a notable accuracy of 92.7% with the ICRM classifier. The summary of several related studies was reported in Table I.

TABLE I. LITERATURE REVIEW

No.	Author (s)	Models	Accuracy	Reference
1	Al-Radaideh et al.	DTC	87.9%	[30]
2	Nguyen and Peter	DTC	82%	[33]
3	Bichkar and R. R. Kabra	DTC	69.94%	[35]
4	Edin Osmanbegovic et al.	NBC	76.65%	[37]
5	Carlos et al. [40]	ADTree	97.3%	[40]
6	Kabakchieva	DTC	72.74%	[39]

C. Objective

Employing the Extra-Trees Classifier (ETC) technique, this research had the primary objective of developing a robust Machine learning model for predicting student performance, leveraging data from reliable sources. In creating these models, the study introduced an innovative approach by seamlessly integrating two optimization algorithms: the Gorilla Troops Optimizer (GTO) and the Reptile Search Algorithm (RSA). The decision to integrate these optimization algorithms stems from their complementary strengths. GTO, inspired by gorilla troop foraging behavior, balances global and local search strategies, while RSA adapts to changing environments efficiently. This novel combination of techniques aimed to enhance the accuracy and precision of the predictive model, ultimately contributing to more effective student performance forecasts in an educational context. The ETC model is employed in predicting and classifying student performance due to its robustness and effectiveness. ETC minimizes overfitting, enhances accuracy, and handles diverse data patterns. This model is particularly valuable in educational contexts where the prediction of student outcomes requires a versatile and resilient algorithm capable of capturing nuanced relationships within complex datasets. Subsequently, in Section II, the material and methodology of the research are prepared; Section III contains information about the evaluation methods, results, and discussion of prediction models; and finally, the results of classification models. In the end, Section IV concludes the important findings of the study.

II. MATERIALS AND METHODOLOGY

A. Extra-Trees Classifier (ETC)

Geurts et al. [41] introduced the Extra Trees Classifier as a modification of the Random Forest algorithm. This model, acknowledged as a highly randomized tree classifier or redundant tree classifier, operates through the utilization of an ensemble learning approach. The Extra-Trees algorithm constructs an ensemble of decision or regression trees via the conventional top-down procedure. Its primary distinctions from other tree-based ensemble methods lie in two key aspects: firstly, it randomly selects cut-points for node splits, and secondly, it employs the entire learning sample for the growth of the trees.

The Extra-Trees algorithm employs a randomized splitting procedure for numerical attributes, controlled by parameters K (number of randomly selected attributes at each node) and n_{min} (minimum sample size for node splitting). The method utilizes the full original learning sample multiple times to create an ensemble model with M trees. Predictions are aggregated through majority vote or arithmetic average for classification and regression, respectively. The approach aims to reduce variance by explicit randomization of cut points and attributes, outperforming other methods. Using the full learning sample minimizes bias. Despite a complexity of $N \log N$, the simplicity of the node-splitting procedure contributes to computational efficiency. Parameters K, n_{min}, and M influence attribute selection, noise averaging, and variance reduction, respectively. While adaptable, default settings are preferred for computational advantages and method autonomy.

The process of dividing attributes in Extra-Trees is outlined as follows:

Split a node (S)
Input: for the node designated for splitting, present the local learning subset S. The resulting output is either a split, denoted as [$a < a_c$], or no result.
– If Stop split(S) is TRUE then return nothing.
– Otherwise select K attributes $\{a_1, \ldots, a_k\}$ among all non-constant (in S) candidate attributes;
– Draw K splits $\{s_1, \ldots, s_k\}$, where s_i= Pick a random split (S,a_i), $\forall_i$= 1, . . ., K;
– Return a split s_* such that Score (s_*, S) = maxi=1..., K Score (s_i, S)
Pick a random split (***S***,a)
Inputs: a subset S and an attribute a
Output: a split
– Let a^s_{max} and a^s_{min} denote the maximal and minimal value of a in S;
– Draw a random cut-point a_c uniformly in [a^s_{max}, a^s_{min}];
– Return the split [a < a_c].
Stop split (***S***)
Input: a subset S
Output: a Boolean
– If |S| < n_{min}, then return TRUE;
– If all attributes are constant in S, then return TRUE;
– If the output is constant in S, then return TRUE;
– Otherwise, return FALSE.

B. Gorilla Troops Optimizer (GTO)

The genesis of the GTO technique can be traced back to the observation and analysis of social intelligence within gorilla groups in their natural habitats [42]. In this methodology, every gorilla is considered a potential solution, and the optimal solution at each optimization stage is identified as the silverback gorilla. The optimization process is delineated into two key phases: exploration and exploitation. To stimulate exploration, three strategies are implemented, with one of them entailing the migration of gorillas to unexplored areas. The objective of this migration strategy is to augment the exploration process, as elucidated in Eq. (1).

$$GX(t+1) = (UL - LL) \times r_1 + LL, Rnd < O \quad (1)$$

The second approach entails transitioning to a different gorilla group, contributing to the equilibrium between exploration and exploitation, as articulated in Eq. (2).

$$GX(t+1) = (r_2 - C) \times X_r(t) + L \times H, Rnd \geq 0.5 \quad (2)$$

The third tactic involves relocating to the designated site, primarily focused on augmenting the GTO's capacity to explore varied optimization spaces, as elucidated in Eq. (3).

$$GX(t+1) = X(i) - L \times \Big(L \times \big(X(t) - GX_r(t)\big) + r_3 \times \big(X(t) - GX_r(t)\big)\Big), Rnd < 0.5 \quad (3)$$

where, $GX(t+1)$ signifies the potential solution position of a gorilla in the subsequent iteration, while $X(t)$ is the current position vector of the gorilla. Rnd, r_1, r_2, and r_3 are random values in the range of $[0-1]$. The parameter o denotes the probability of opting for the migration strategy to an unfamiliar position and must be predetermined between 0 and 1 prior to initiating the optimization process. X_r represents a randomly chosen member from the gorilla group, while GX_r denotes the potential solution vector position of the gorilla, chosen at random. LL and UL represent the lower and upper limits of the variables, respectively. Additionally, C, L, and H can be defined through mathematical expressions as per Eq. (4) to Eq. (6).

$$C = (\cos(2 \times r_4) + 1) \times (1 - t/\max(t)) \quad (4)$$

$$L = C \times l \quad (5)$$

$$H = Z \times X(t) \quad (6)$$

Here, r_4 signifies a random value within the range of 0 to 1, while l represents a random value within the range of -1 to 1, and Z denotes a random value that ranges from $-C$ to C. If the value of $Rand$ is below p, the first strategy is executed. Conversely, if $Rand$ is greater than or equal to 0.5, the second strategy is applied, and if $Rand$ is less than 0.5, the third strategy is chosen. The optimal solution acquired during the exploration phase is subsequently identified as the silverback.

To enhance exploitation, the GTO methodology incorporates two strategies. The initial strategy entails tracking the silverback, which is the designated gorilla, symbolizing the optimal solution. This strategy is activated when the value of the parameter C exceeds the random parameter W. The silverback assumes the role of a leader guiding other gorillas in foraging for food. This behavior can be expressed mathematically through Eq. (7) and Eq. (8), where, $g = 2^L$.

$$GX(t+1) = L \times M \times (X(t) - X_{sb}) + X(t) \quad (7)$$

$$M = \left(\left|\left(\frac{1}{n}\right)\sum_{i=1}^{n} GX_i(t)\right|^g\right)^{1/g} \quad (8)$$

Here, X_{sb} signifies silverback gorilla.

The alternate exploitation strategy centers around vying for the adult female gorilla. This course is selected when the value of the parameter C falls below the random parameter W. In their native environment, young male gorillas engage in intense competition to win the favor of a female gorilla. This conduct can be mathematically articulated through Eq. (9).

$$GX(t+1) = X_{sb} - (X_{sb} \times Q - X(t) \times Q) \times V \quad (9)$$

$$Q = 2 \times r_5 - 1 \quad (10)$$

$$V = \gamma \times I \quad (11)$$

Here, Q signifies the force of impact, with r_5 representing a random value within the range of $[0-1]$. V is a vector signifying the intensity of aggression during a conflict, and γ is a predetermined value established before initiating the optimization process. I denotes the influence of aggression on the solution's dimensions. The optimal solution derived from the exploitation phase is then assigned the role of the new silverback. This designation could either be retained from the chosen gorilla during the exploration phase or newly selected.

The pseudo-code of GTO is provided below [43]:

Algorithm 1. The pseudo-code of GTO.
GTO setting
Inputs: The population size n and maximum number of iterations T and parameters γ and O
Outputs: The location of the Gorilla and its fitness value
Initialization
Initialize the random population $X_i (i = 1,2, \dots, n)$
Calculate the fitness values of the Gorilla
Main Loop
while (stopping condition is not met) do
Update the C
Update the L
Exploration phase
for (each Gorilla (X_i)) do
Update the location of Gorilla
end for
% Create group
Calculate the fitness values of the Gorilla
if GX is better than X, replace them
Set X_{sb} as the location of $silverback$ (best location)
% Exploitation phase
for (each Gorilla (X_i)) do
if ($\|C\| \geq 1$) then
Update the location of Gorilla
else
Update the location of Gorilla
end if
end for
% Create group
Calculate the fitness values of the Gorilla
if New Solutions are better than previous solutions, replace them
Set X_{sb} as the location of $silverback$ (best location)
end while
Return $X_{BestGorilla}$, $bestFitness$

C. Reptile Search Algorithm

The Reptile Search Algorithm (RSA) draws inspiration from the foraging behaviors observed in crocodiles within their natural environment [44]. It operates by alternating between encircling and hunting search phases, with the transition between these phases achieved by dividing the total number of iterations into four segments [45], [46].

1) Initialization phase: *The Reptile Search Algorithm commences by stochastically generating an initial set of solution candidates using the following equation:*

$$X_{ij} = rnd \times (UB - LB) + LB \quad j = 1,2, \dots, n \tag{12}$$

In the initialization matrix (X_{ij}) mentioned earlier, the variable j corresponds to the population size, indicating the number of rows in the matrix. LB and UB denote the lower and upper bound constraints, respectively, and rnd signifies randomly generated values employed in the initialization process.

2) Exploration (Encircling phase): The encircling phase primarily involves navigating an area with a high density of potential solutions. In this phase, movements inspired by crocodile behaviors, such as high walking and belly walking, plays a crucial role. It is essential to emphasize that these movements are not directly focused on capturing prey; instead, their purpose is to explore a wide search space within the optimization process.

$$X_{ij}(\vartheta + 1) = OPT_j(\vartheta) \times \left(-\rho_{(ij)}(\vartheta)\right) \times \varepsilon - \left(R_{ij}(\vartheta) \times rnd\right), \quad \vartheta \leq \frac{N}{4} \tag{13}$$

$$X_{ij}(\vartheta + 1) = OPT_j(\vartheta) \times X_{(r_1,j)} \times ES(\vartheta) \times rnd, \vartheta \leq \frac{2N}{4} \, and \; \vartheta > \frac{N}{4} \tag{14}$$

Here, $OPT_j(\vartheta)$ represents the optimal solution obtained at the jth position, where ϑ denotes the current iteration number, and N is the maximum number of iterations. $\rho_{(ij)}$ represents the value generated by the hunting operator for the ith solution at the jth position. The parameter ε explains the sensitivity, influencing the exploration accuracy. R_{ij} is utilized to reduce the search space area. The calculations for $\rho_{(ij)}$ and R_{ij} are as follows:

$$\rho_{(ij)} = OPT_j(\vartheta) \times P_{(i,j)} \tag{15}$$

$$\rho_{(ij)} = \frac{OPT_j(\vartheta) - P_{(r_2,j)}}{OPT_j(\vartheta) + \alpha} \tag{16}$$

Here, the variable r_1 is a randomly generated number within the range of $[1 - T]$, where T represents the total count of candidate solutions. $X_{(r_1,j)}$ signifies a randomly chosen position for the jth solution. Similarly, r_2 is another randomly generated number ranging from $[1 - T]$, and α denotes a small-magnitude value. $ES(\vartheta)$ denoted as Evolutionary Sense, is a probability-based ratio. The mathematical expression of Evolutionary Sense can be articulated as follows:

$$ES(\vartheta) = 2 \times r_3 \times \left(1 - \frac{1}{N}\right) \tag{17}$$

In this scenario, the variable r_3 represents a randomly generated numerical value. The calculation of $P_{(i,j)}$ is determined using the following formula:

$$P_{(i,j)} = \varepsilon + \frac{X_{(i,j)} - A(X_i)}{OPT_j(\vartheta) \times (UB - LB) + \alpha} \tag{18}$$

$$AVG(X_i) = \frac{1}{T} \sum_{j=1}^{T} X_{(i,j)} \tag{19}$$

where, $AVG(X_i)$ represents the average position of the ith solution.

3) Exploitation (Hunting phase): The hunting phase involves two key strategies: hunting coordination and cooperation. These strategies play a crucial role in local-scale exploration, resembling the pursuit of optimal solutions, similar to hunting prey. The hunting phase is segmented based on the current iteration number. The hunting coordination strategy operates when the iteration number ϑ is within $\vartheta \leq \frac{3N}{4}$ and $\vartheta > \frac{2N}{4}$, while the hunting cooperation strategy is applied when $\vartheta \leq N$ and $\vartheta > \frac{3N}{4}$. These strategies incorporate stochastic coefficients to explore the local search space and generate optimal solutions systematically. The exploitation phase is guided by Eq. (9) and Eq. (10) to facilitate this process.

$$X_{ij}(\vartheta+1) = OPT_j(\vartheta) \times P_{ij}(\vartheta) \times rnd,$$

$$\vartheta \leq \frac{3N}{4} \ and \ \vartheta > \frac{2N}{4} \quad (20)$$

$$X_{ij}(\vartheta+1) = OPT_j(\vartheta) - \rho_{(ij)}(\vartheta) \times \alpha - R_{ij}(\vartheta) \times rnd, \vartheta \leq Nand \ \vartheta > \frac{3N}{4} \quad (21)$$

The RSA process is illustrated in Fig. 1.

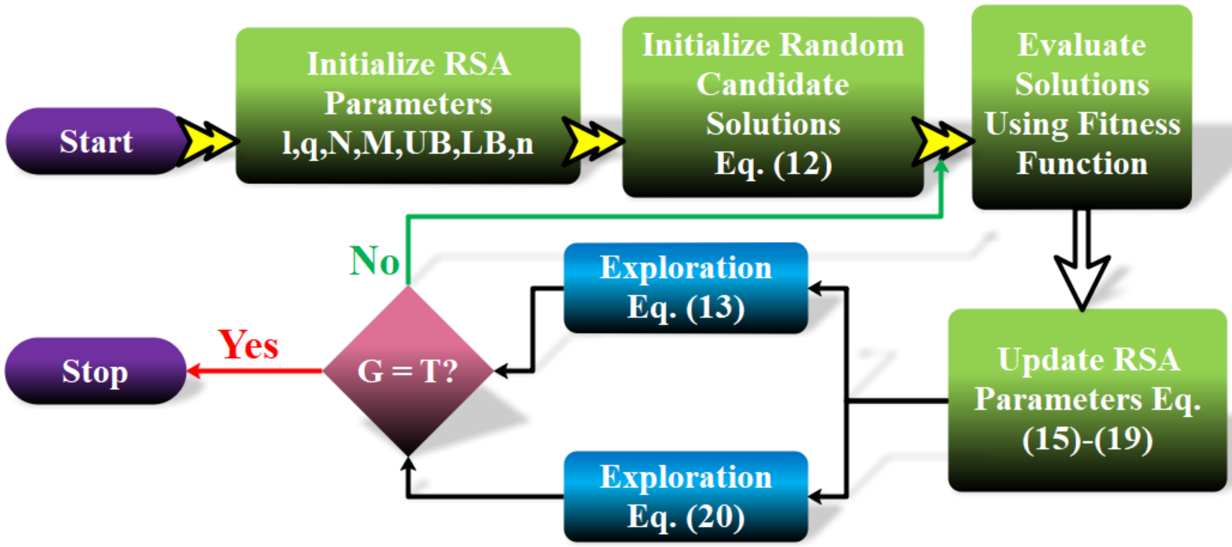


Fig. 1. Flowchart of RSA.

D. Data Processing

The principal objective of this study is to formulate a robust methodology for the accurate evaluation of students' academic performance, considering various contextual factors that exert influence. To achieve this goal, meticulous preprocessing of the initial dataset is imperative. In this research, a dataset related to education in Portugal was employed, consisting of 33 distinct characteristics [47], [48] [49]. These features were selected to effectively depict the academic performance of a total of 395 students, considering the information and circumstances of each individual throughout the academic period. The initial step involves the conversion of textual data into numerical values, a foundational prerequisite for the execution of machine learning tasks, facilitating effective data analysis and the application of advanced statistical techniques. The dataset encompasses a diverse range of variables with potential impacts on academic outcomes, including sex, school, urban or rural residency (address), age, family size (famsize), guardian, parental cohabitation status (Pstatus), parental education and occupations (Medu, Fedu, Mjob, and Fjob), home-to-school travel time (traveltime), weekly study time (studytime), school choice motivation (reason), current health status, past class failures (failures), weekday (Dalc), and weekend (Walc) alcohol consumption, engagement in extra paid classes, participation in supplementary education (schoolsup), family educational support (famsup), attendance at nursery school, involvement in extracurricular activities, aspirations for higher education, access to the internet, student absences, involvement in romantic relationships, quality of family relationships, free time, and frequency of socializing. To optimize the dataset's suitability, the preprocessing phase incorporated the application of random permutation (randperm) to mitigate biases, along with normalization procedures aimed at standardizing parameter scales. This research aims to predict and categorize students' academic performance, utilizing the G1 and G3 variables, with G3 representing final grades segmented into four distinct levels: Excellent (16–20), Good (14–16), Acceptable (12–14), and Poor (0–12). The methodology seeks to establish a comprehensive framework for comprehending and assessing academic performance within various contextual factors, contributing to improvements in educational practices and policy development. Fig. 2, presented in the article, illustrates a correlation matrix detailing relationships among input and output variables, highlighting the positive influence of parental education, especially maternal education, on academic performance. Additionally, factors such as daily and weekly alcohol consumption, prior academic failures, and student age demonstrate discernible impacts on school grades, underscoring the critical importance of both study time and parental education as pivotal factors contributing to academic success.

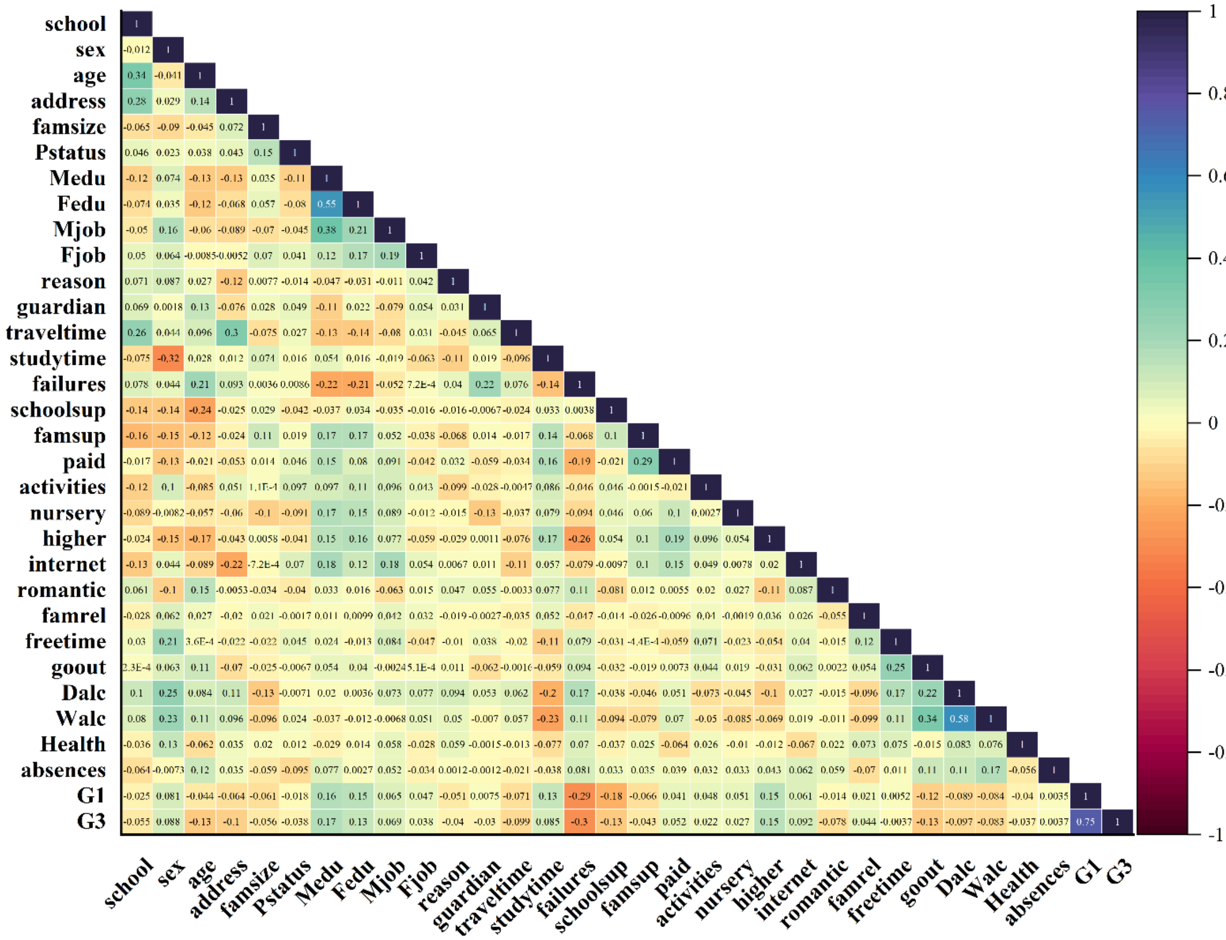


Fig. 2. Correlation matrix for the input and output variables.

III. Results and Discussion

A. Evaluation of Models' Applicability

In the evaluation of classification problems, the metric commonly employed to assess a model's overall performance is Accuracy. This metric relies on four key components: True Positives (TP) for correct positive predictions, True Negatives (TN) denoting accurate negative predictions, False Positives (FP) representing inaccurate positive predictions, and False Negatives (FN) indicating incorrect negative predictions. However, the applicability of Accuracy diminishes in scenarios involving imbalanced data, where it tends to favor the majority class, limiting its interpretability. To overcome this limitation, three additional evaluation metrics, including Recall, F1-Score, Precision, Matthew's correlation coefficient (MCC), and Area under the curve (AUC), are frequently utilized. These metrics offer a more nuanced understanding of a model's performance, particularly in the presence of imbalanced class distributions. Expressed through mathematical equations, typically numbered from 22 to 26, these metrics collectively contribute to a refined and comprehensive assessment of the effectiveness of a classification model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{22}$$

$$Precision = \frac{TP}{TP+FP} \tag{23}$$

$$Recall = TPR = \frac{TP}{P} = \frac{TP}{TP+FN} \tag{24}$$

$$F1\_score = \frac{2\times Recall \times Precision}{Recall+Precision} \tag{25}$$

$$MCC = \frac{(TP*TN)-(FP+FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{26}$$

B. Hyperparameters and Convergence Results

In machine learning, hyperparameters, external configurations that encompass elements such as learning rates and regularization strengths, play a pivotal role in shaping a model's behavior. Unlike parameters, hyperparameters are predetermined and are not directly acquired from the data. The optimization of model performance hinges on the essential process of tuning hyperparameters, demanding experimentation, and the application of optimization techniques. Table II meticulously delineates the hyperparameter values associated with ETRS and ETGT models, specifically max_depth, min_samples_split, min_samples_leaf, and max_leaf_nodes. This comprehensive presentation significantly bolsters the transparency and reproducibility of models in machine learning research, offering crucial insights for a more profound comprehension and accurate replication of model configurations.

TABLE II. RESULTS OF HYPERPARAMETERS

Target	Hyperparameter	ETRS	ERGT
G1	max_depth	24	722
	min_samples_split	0.028	0.511
	min_samples_leaf	0.0125	0.092
	max_leaf_nodes	250	10
G3	max_depth	275	975
	min_samples_split	0.001	0.234
	min_samples_leaf	0.0015	0.0631
	max_leaf_nodes	4790	10

This research endeavors to optimize the Extra-Trees Classifier's (ETC) hyperparameters, tailoring its performance to specific datasets and problem domains. The optimization process involves utilizing the Gorilla Troops Optimizer (GTO) and Reptile Search Algorithm (RSA), representing a substantial advancement in enhancing the predictive capabilities of this foundational machine learning algorithm. Evaluating the optimization performance involves assessing how selected algorithms impact the Accuracy of ETC through iterations. Fig. 3 depicts two convergence curves, namely ETRS and ETGT, using a stair form with four steps of 50 iterations each. In G1 prediction, ETRS initiates with lower accuracy, but within the first 90 iterations, it consistently outperforms ETGT. The dynamics reverse in the second stage, and after the 90th iteration, both models perform similarly. Notably, around the 125th iteration, ETRS exhibits a marked increase in Accuracy. Conversely, in G3 estimation, both models start with similar Accuracy values. ETRS outperforms ETGT from 0 to 25 iterations, and then ETGT surpasses ETRS from 25 to approximately 75 iterations. Between 75 and 120 iterations, the performance of both models aligns. After the 125th iteration, ETRS experiences a distinctive surge, ultimately concluding the convergence process with a higher accuracy rate than the ETGT model.

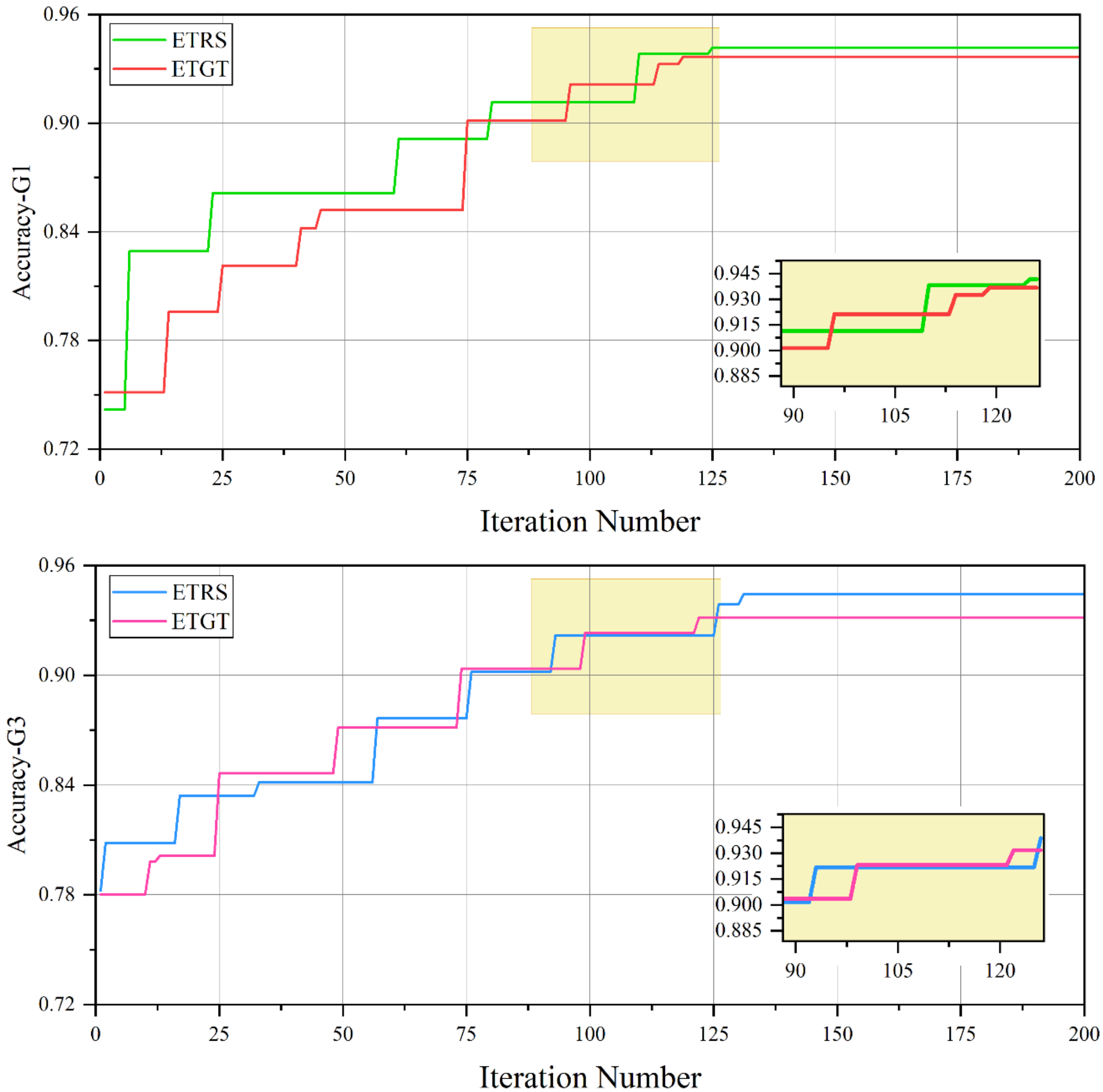


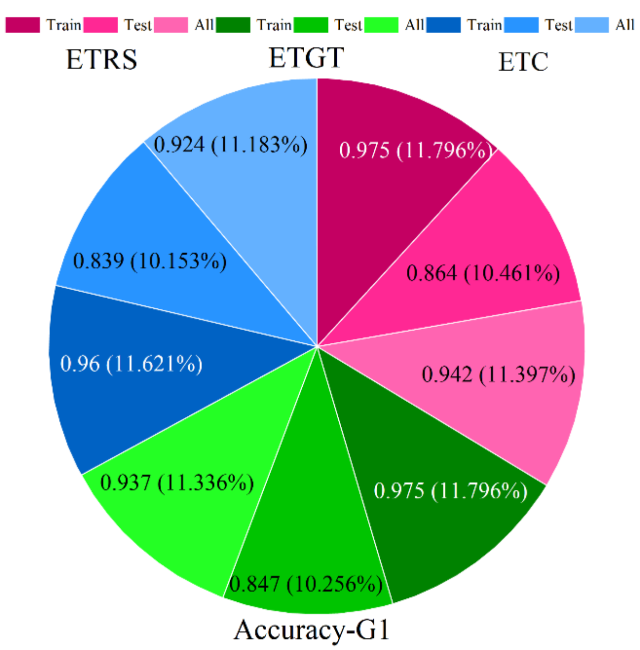
Fig. 3. Convergence curve of models. Prediction and classification results.

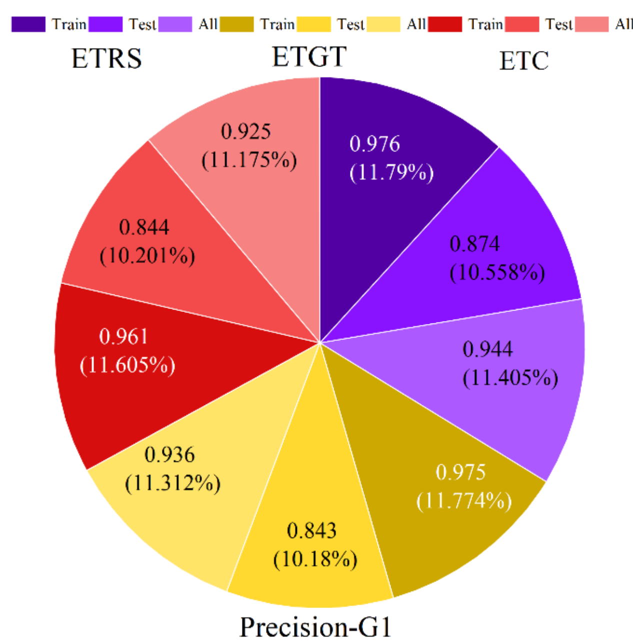
In the pursuit of predicting future academic achievements in mathematics through machine learning algorithms, this investigation integrates a diverse array of student information, with a specific emphasis on their short-term and final grades (G1 and G3). The dataset assumes a pivotal role in the training and evaluation of three models based on Extra-Trees Classifier (ETC), namely ETC, ETRS, and ETGT. Within this section, the study systematically computes performance metrics such as Accuracy, Precision, Recall, F1-score, MCC, and AUC at each prediction stage. This meticulous analysis aims to discern the most effective prediction model, offering valuable insights for enhancing students' academic success. All relevant metric values, encompassing all, train, test, and model, are detailed in Table III and illustrated in Fig. 4.

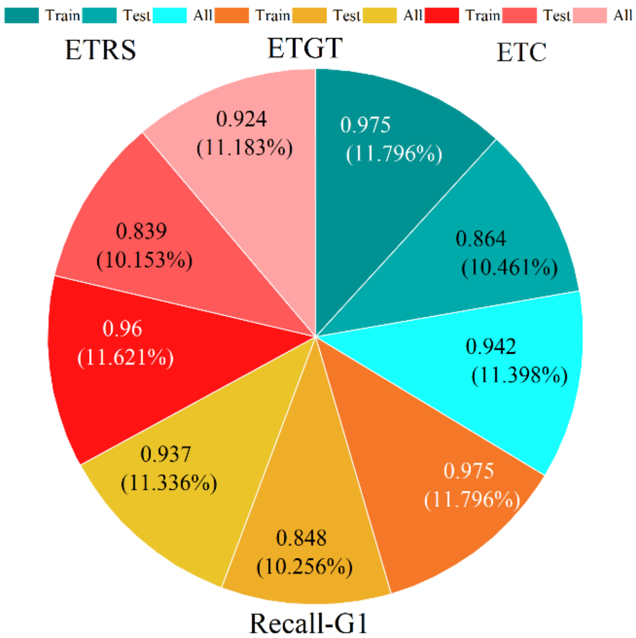
Regarding G1 prediction, ETRS and ETC exhibit the strongest and weakest prediction performance, achieving maximum and minimum Accuracy values of 0.975 and 0.839, respectively. ETRS attains maximum Precision, Recall, F1-score, MCC, and AUC values of 0.976, 0.975, 0.975, 0.957, and 0.931, affirming its high accuracy in positive predictions. The performance of the other hybrid model (ETGT) aligns with ETRS in the training phase Accuracy but experiences a lower value in the testing phase. For G3 prediction, the comparison among the three models reveals ETRS as the strongest predictor with maximum Accuracy, Recall, and F1-score values of 0.953 and 0.921 for MCC. ETGT, with a 0.7% lower Accuracy, secures the second position in the ranking when compared to ETC.

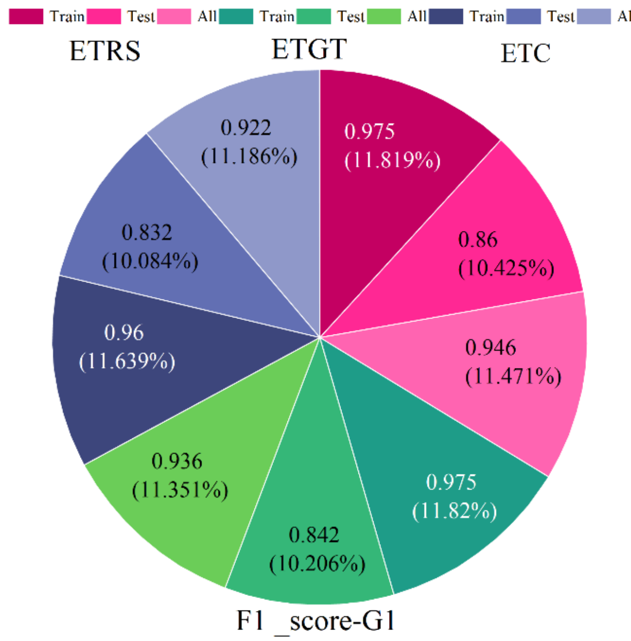
TABLE III. RESULT OF PRESENTED MODELS

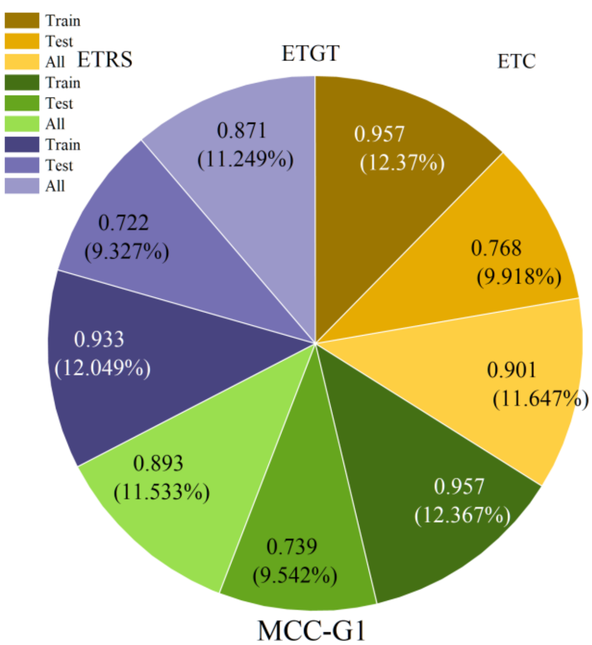
Target	**Model**	**Section**	**Index values**					
			Accuracy	***Precision***	***Recall***	***F1 _Score***	***MCC***	***AUC***
G1	ETC	Train	0.960	0.961	0.960	0.960	0.932	0.913
		Test	0.839	0.844	0.839	0.832	0.722	
		All	0.924	0.925	0.924	0.922	0.871	
	ETRS	Train	0.975	0.976	0.975	0.975	0.957	0.931
		Test	0.864	0.874	0.864	0.860	0.768	
		All	0.942	0.944	0.942	0.946	0.902	
	ETGT	Train	0.975	0.975	0.975	0.975	0.957	0.936
		Test	0.847	0.843	0.848	0.842	0.739	
		All	0.937	0.936	0.937	0.936	0.893	
G3	ETC	Train	0.957	0.957	0.957	0.956	0.926	0.922
		Test	0.847	0.851	0.848	0.844	0.737	
		All	0.924	0.924	0.924	0.923	0.871	
	ETRS	Train	0.953	0.955	0.953	0.953	0.921	0.945
		Test	0.924	0.924	0.924	0.923	0.872	
		All	0.944	0.946	0.944	0.944	0.906	
	ETGT	Train	0.946	0.946	0.946	0.946	0.908	0.932
		Test	0.898	0.900	0.898	0.896	0.827	
		All	0.932	0.933	0.932	0.931	0.884	

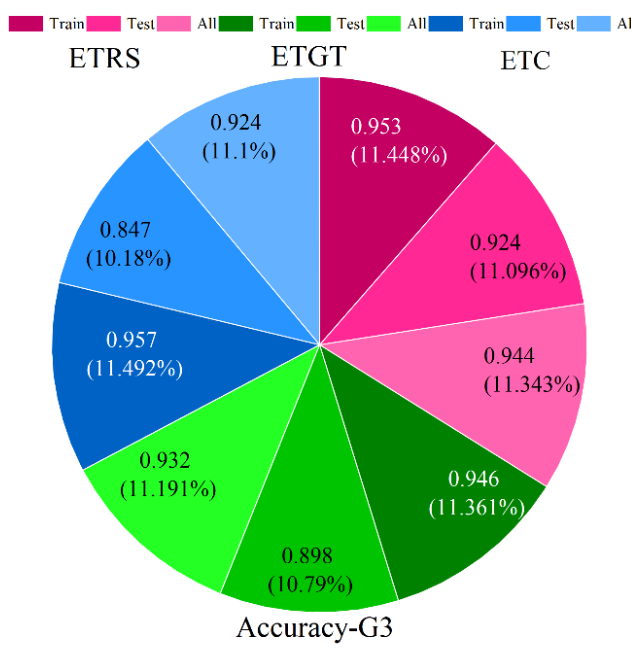





Train Test All Train Test All Train Test All
ETRS
ETGT
ETC
0.975 (11.796%)
0.864 (10.461%)
0.942 (11.398%)
0.975 (11.796%)
0.848 (10.256%)
0.937 (11.336%)
0.96 (11.621%)
0.839 (10.153%)
0.924 (11.183%)
Recall-G1


Train Test All Train Test All Train Test All
ETRS
ETGT
ETC
0.975 (11.819%)
0.86 (10.425%)
0.946 (11.471%)
0.975 (11.82%)
0.842 (10.206%)
0.936 (11.351%)
0.96 (11.639%)
0.832 (10.084%)
0.922 (11.186%)
F1 _score-G1


Train
Test
All ETRS
Train
Test
All
Train
Test
All
ETGT
ETC
0.957 (12.37%)
0.768 (9.918%)
0.901 (11.647%)
0.957 (12.367%)
0.739 (9.542%)
0.893 (11.533%)
0.933 (12.049%)
0.722 (9.327%)
0.871 (11.249%)
MCC-G1


Train Test All Train Test All Train Test All
ETRS
ETGT
ETC
0.953 (11.448%)
0.924 (11.096%)
0.944 (11.343%)
0.946 (11.361%)
0.898 (10.79%)
0.932 (11.191%)
0.957 (11.492%)
0.847 (10.18%)
0.924 (11.1%)
Accuracy-G3

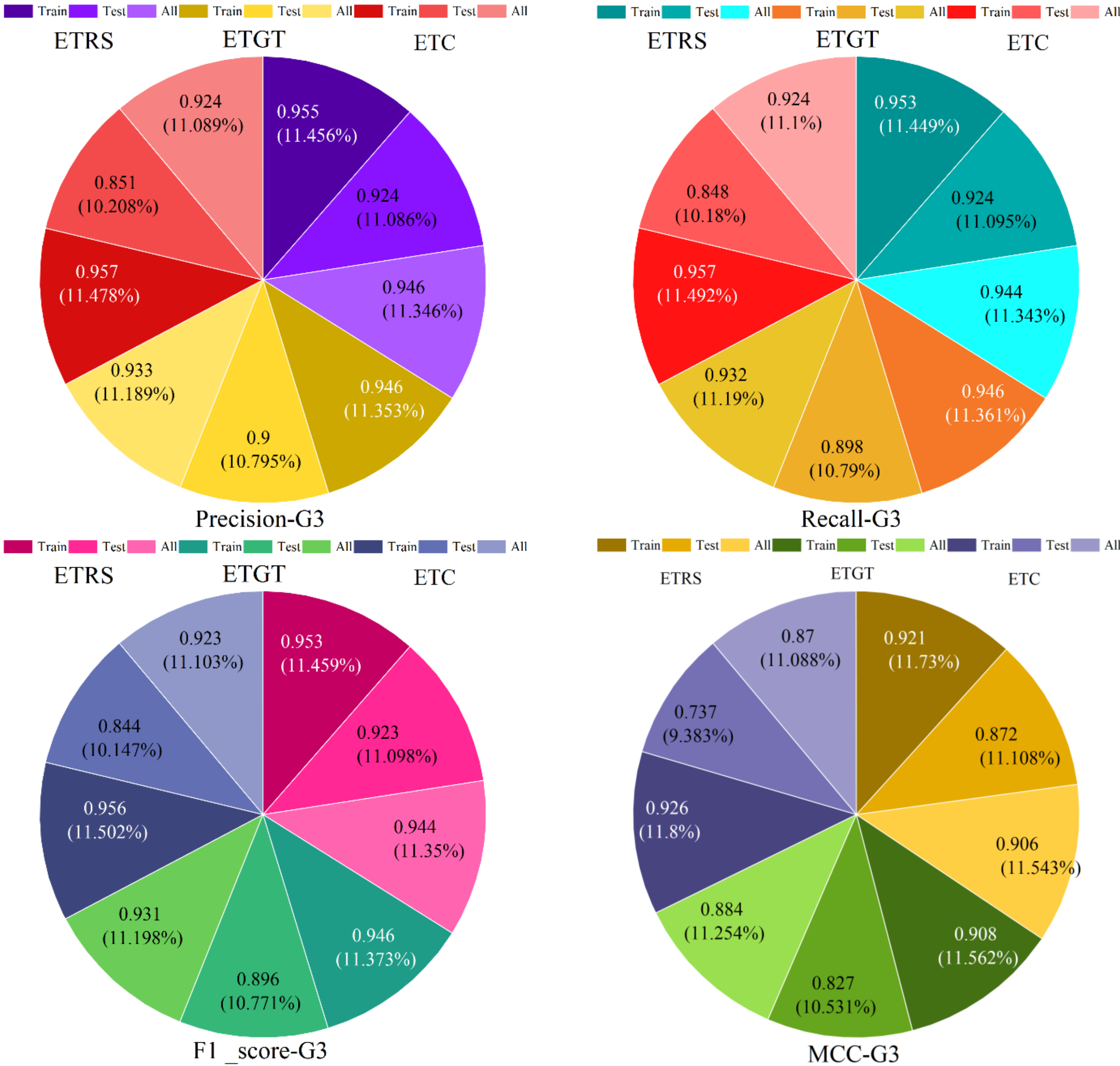


Fig. 4. Pie chart plot for the evaluation of developed models.

Tables IV and V presents the Precision, Recall, F1_score, MCC metrics for students categorized by G1 and G3 grades. These tables offer insights into the model's performance, revealing its accuracy in positive predictions, ability to capture true positives, and overall effectiveness in classifying students based on academic performance levels.

1) G1

a) Excellent: This cohort constitutes around 10% of the dataset, featuring 41 high-achieving students. Despite the ETC and ETRS models exhibiting impeccable Precision (1), the optimized ETGT model shows a slight difference of approximately -5.5%. However, with a Recall value of 0.8293, the ETGT model excels in accurately identifying instances within this top-performing group, surpassing the other models.

b) Good: Among the 54 students in this group, the ETGT model emerged as the superior classifier, achieving Precision, Recall, and F1_score values of 0.9444. Notably, ETRS exhibited the weakest performance based on Precision values, whereas, considering MCC, Recall, and F1-Score, the ETC model demonstrated the least optimal performance levels.

c) Acceptable: The ETRS model showcased superior applicability compared to other models, registering maximum values for all metrics (Precision = 0.984, Recall = 0.912, F1-Score = 0.947, and MCC=0.928). In contrast, the ETGT hybrid model occupied the last position, exhibiting the lowest values across all metrics.

d) Poor: This group pertains to students who have faced academic failure, and the 232 students classified within this category demand heightened institutional focus for improvement. Notably, the performance of all models in classifying this group proves to be optimal, surpassing the other three classes with Precision exceeding 90%. Once again, ETRS stands out as the best model, exhibiting the highest metric values in this context.

TABLE IV. EVALUATION INDEXES OF THE DEVELOPED MODELS' PERFORMANCE IN G1

Model	Grade	Index values			
		Precision	***Recall***	***F1-score***	***MCC***
ETC	Excellent	1.000	0.805	0.892	0.873
	Good	0.878	0.796	0.835	0.812
	Acceptable	0.922	0.868	0.894	0.892
	Poor	0.924	0.991	0.956	0.887
ETRS	Excellent	1.000	0.781	0.877	0.937
	Good	0.839	0.870	0.855	0.831
	Acceptable	0.984	0.912	0.947	0.928
	Poor	0.947	0.996	0.971	0.872
ETGT	Excellent	0.9444	0.8293	0.8831	0.847
	Good	0.9444	0.9444	0.9444	0.935
	Acceptable	0.8923	0.8529	0.8722	0.906
	Poor	0.9458	0.9784	0.9619	0.873

2) G3

a) Excellent: This subset comprises 40 high-achieving students, representing nearly 10% of the entire dataset under scrutiny. While the Precision values imply that the standalone ETC model showcases flawless predictive capability with a score of 0.971, the optimized models exhibit slightly higher scores (less than 1%). However, a thorough evaluation based on MCC, Recall and F1- score underscores the superiority of the ETRS model, attaining values of 0.883 0.9 and 0.9351.

b) Good: Among this cohort of 60 students, representing 15% of the total 395 studied students, the ETGT model showcased superior performance, particularly evident in Precision values. Furthermore, a comprehensive evaluation considering MCC, Recall and F1-Score affirmed the model's excellence, boasting MCC, Recall and F1-Score values of 0.9, 0.8667 and 0.9123, respectively.

c) Acceptable: Upon scrutinizing the outcomes, it is apparent that the Reptile Search Algorithm (RSA) outperformed the Gorilla Troops Optimizer (GTO) in optimizing the Extra-Trees Classifier (ETC) for G3 classification. The RSA demonstrated higher success, yielding a Recall of 0.9516 and an F1-score of 0.908.

d) Poor: In the classification of students within the Poor category, the ETRS model demonstrated superior performance among the three models. It achieved maximum values across all metrics, notably excelling in the Recall evaluator with a value of 0.9828.

TABLE V. EVALUATION INDEXES OF THE DEVELOPED MODELS' PERFORMANCE IN G3

Model	Grade	Index values			
		Precision	***Recall***	***F1-score***	***MCC***
ETC	Excellent	0.971	0.825	0.892	0.856
	Good	0.895	0.850	0.872	0.850
	Acceptable	0.885	0.871	0.878	0.885
	Poor	0.934	0.974	0.954	0.884
ETRS	Excellent	0.973	0.9	0.9351	0.883
	Good	0.9245	0.8167	0.8673	0.847
	Acceptable	0.8551	0.9516	0.908	0.942
	Poor	0.9703	0.9828	0.9765	0.929
ETGT	Excellent	0.9722	0.875	0.9211	0.855
	Good	0.963	0.8667	0.9123	0.900
	Acceptable	0.8852	0.871	0.878	0.880
	Poor	0.9303	0.9742	0.9518	0.914

C. Discussion

1) G1: In Fig. 5, the visual representation illustrates the distribution of students across categories, enabling a comprehensive comparison between measured data and the outcomes of classification effectiveness. Specifically focusing on forecasting student performance in G1 scores, individual graphs for each category (Poor, Acceptable, Good, and Excellent) are presented. It is noteworthy that, as per the studied dataset, the total number of students amounts to 395. The subsequent sections meticulously evaluate the models based on recorded figures, revealing 232 individuals in the Poor category, 68 in the Acceptable category, 54 in the Good category, and 41 in the Excellent category. The ETRS model emerges as the most effective classifier for the Poor and Acceptable categories, showcasing precise predictions. However, in the Good and Excellent groups, notable differences are observed between the two hybrid models, with ETRS displaying weaker performance in classifying datasets for these higher-performing groups.

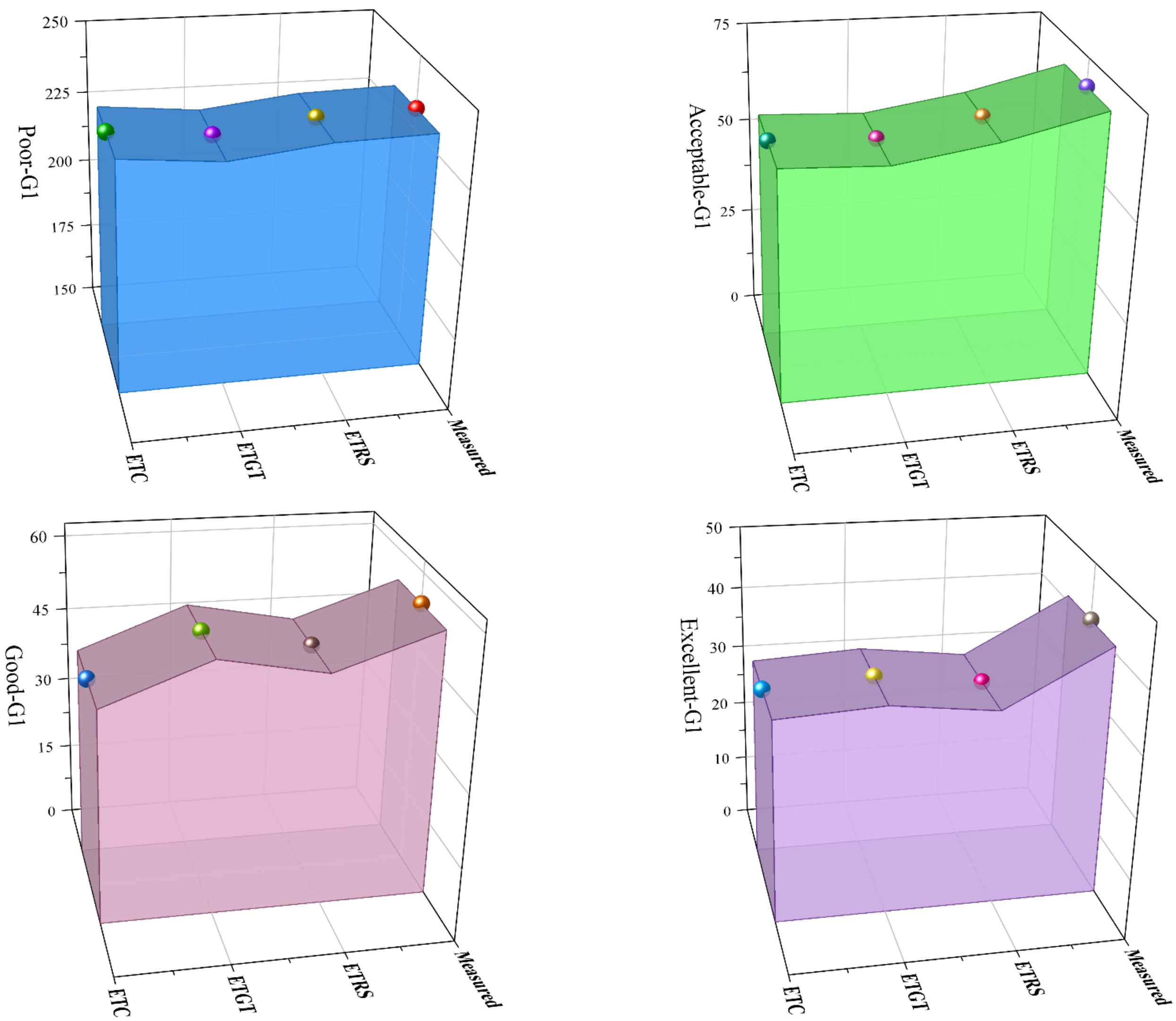


Fig. 5. 3D wall plot for the developed models' accuracy for G1.

Fig. 6 presents a confusion matrix, offering insights into the accurate categorization of students and instances of misclassifications. Within the ETRS model, 372 students were accurately classified across grades, including 32 in Excellent, 47 in Good, 62 in Acceptable, and 231 in Poor, with 23 misclassifications. In contrast, the ETGT model had 25 misclassifications, while the straightforward ETC model accurately classified 365 students and misclassified 30 students.

2) G3: According to Fig. 7, the recorded student figures for the Poor, Acceptable, Good, and Excellent categories were 233, 62, 60, and 40, respectively. Interestingly, the standalone ETC model demonstrated superior performance in the Poor and Good categories compared to the two hybrid models. Subsequently, the ETRS model emerged as the more effective classifier, excelling in the categorization of students into Excellent and Acceptable groups.

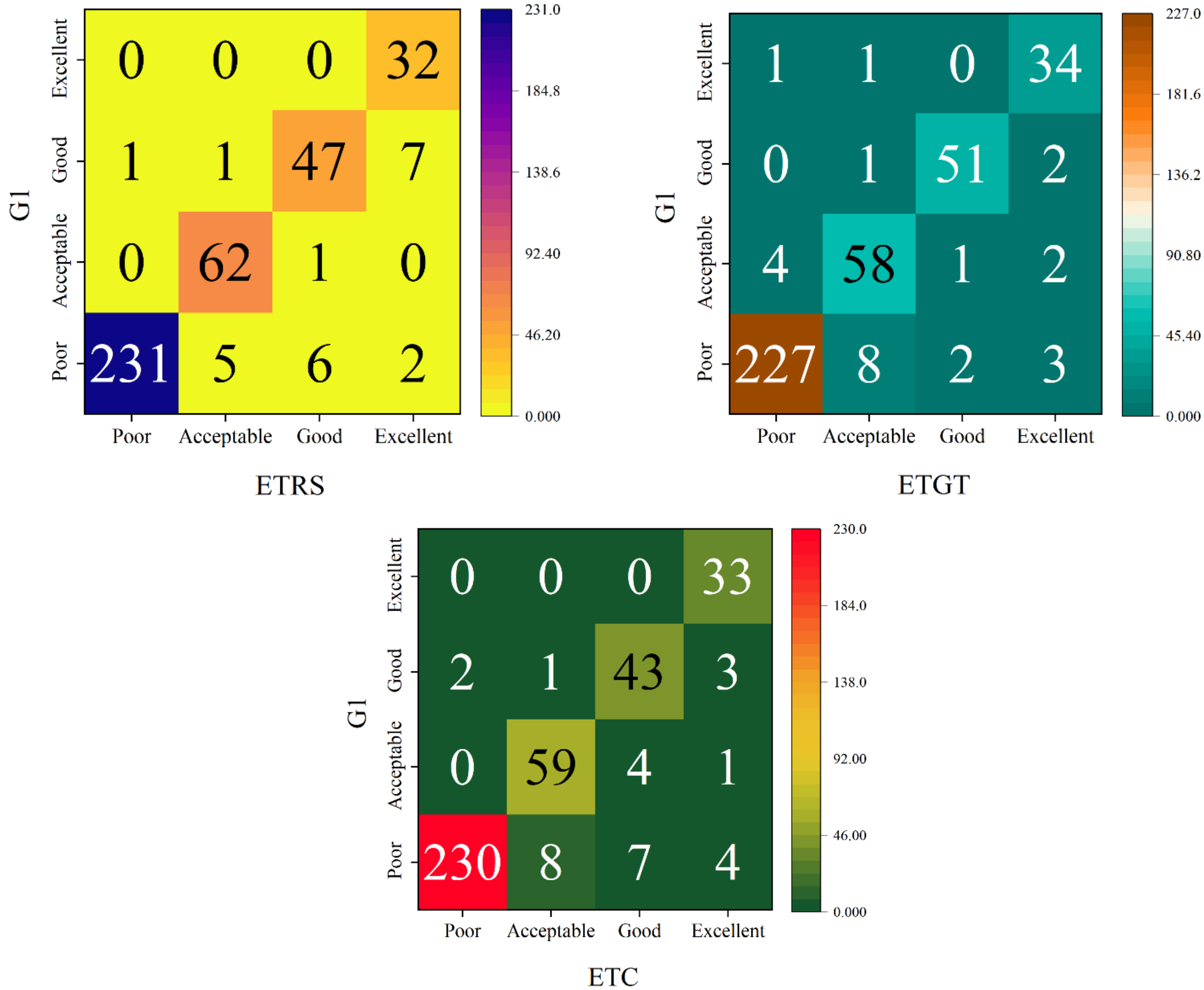
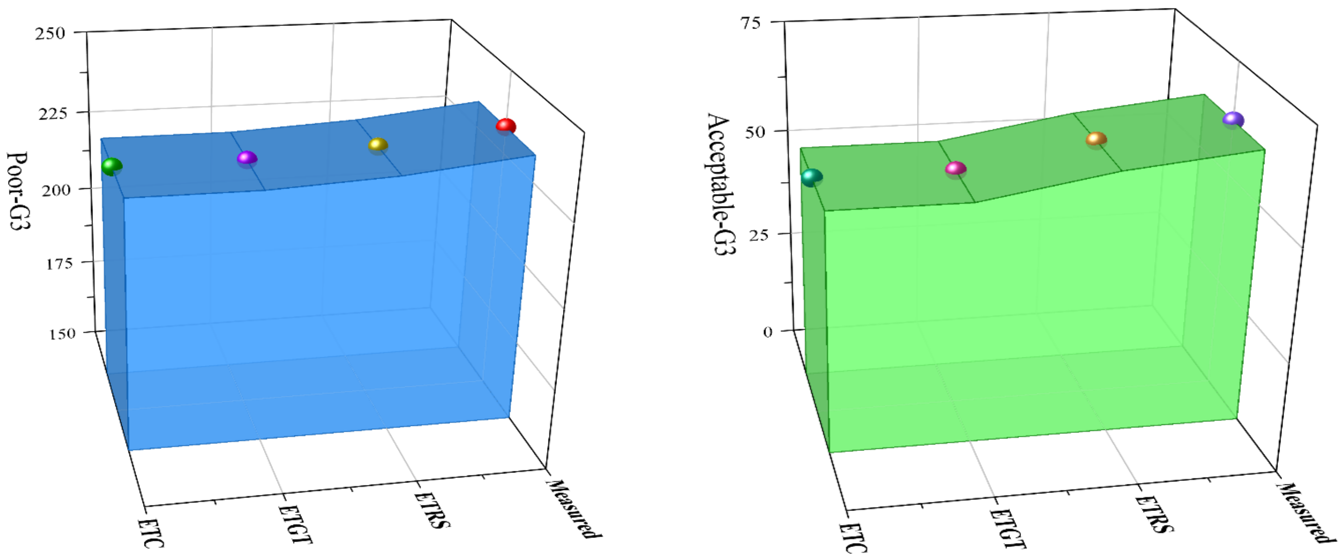


Fig. 6. Confusion matrix for each model's accuracy for G1.



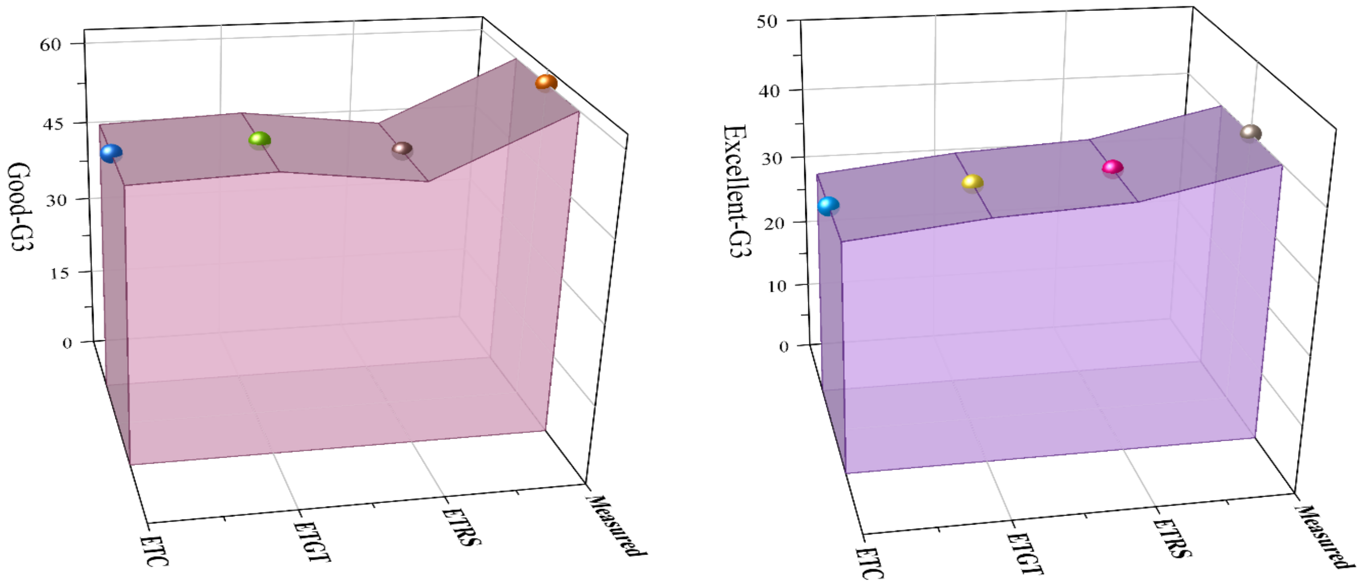


Fig. 7. 3D wall plot for the developed models' accuracy for G3.

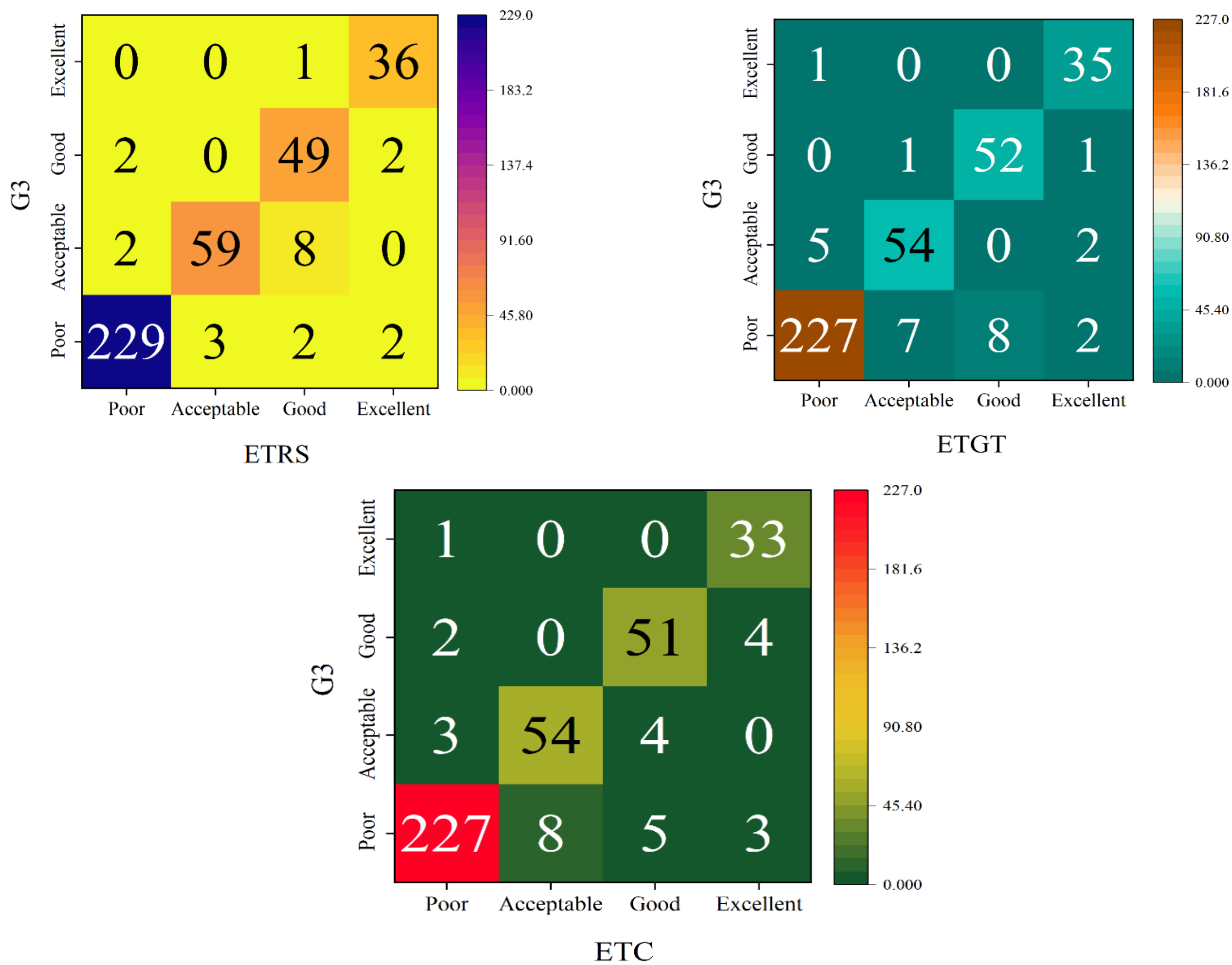


Fig. 8. Confusion matrix for each model's accuracy for G3.

Examining Fig. 8, the ETRS model accurately categorized 373 students into their respective grades, with only 22 misclassifications. In contrast, the ETGT model achieved 368 correct predictions but experienced 27 misclassifications. A thorough comparison reveals that the ETRS model outperformed both the ETGT and ETC models in terms of overall performance.

D. Comparing Previous Studies vs. Present Research Study

Table VI summarizes the results of four existing studies in the field of student performance. According to these works, the highest accuracy was related to the employment of the DTC model in the Nguyan and Peter's study [33] with 82%, while in the present study, the combination of ETC model and RSA algorithm, achieved high value of 0.975 for G1 and 0.953 for G3. As can be seen, this work reached the highest accuracy among others.

TABLE VI. COMPARING RESULTS OF EXISTING STUDIES AND PRESENT WORK

Author (s)	**Models**	**Accuracy**
Kabakchieva [39]	DTC	72.74%
Bichkar and R. R. Kabra [35]	DTC	69.94%
Nguyen and Peter [33]	DTC	82%
Edin Osmanbegovic et al. [37]	NBC	76.65%
Present study for G1	ETRS	0.975
Present study for G3	ETRS	0.953

E. Generelizability and Limitation of Proposed Model

It is acknowledged that our study was conducted on a specific dataset and that the results may not be directly applicable to other educational contexts. However, it is believed that the potential to be generalized to other settings is possessed by the proposed approach of combining the ETC model with optimization algorithms, as long as the following conditions are met: The data is sufficiently large and representative of the target population The data contains relevant features that can capture the factors influencing students' academic performance The data is preprocessed and cleaned to ensure its quality and validity The optimization algorithms are tuned and adapted to the characteristics of the data.

The optimization algorithms are powerful tools that can improve the performance of predictive models, but they also have some limitations that need to be taken into account. Some of the common limitations and potential drawbacks of optimization algorithms are then discussed, such as:

- The dependence on the quality and quantity of the data. The data is relied on by optimization algorithms to learn and optimize the objective function, but the data may not be sufficient, representative, or accurate enough to capture the true complexity and variability of the problem domain. This may lead to overfitting, underfitting, or bias in the optimization results. Iterative processes are often involved by optimization algorithms that require a large amount of computation and memory resources, especially for high-dimensional and nonlinear problems.
- This may limit the applicability and efficiency of the optimization algorithms in real-world scenarios, where time and space constraints are important factors.
- The sensitivity to the choice of parameters and initial conditions.
- Optimization algorithms often have several parameters and initial conditions that need to be specified by the user or tuned by some methods.
- The convergence, stability, and quality of the optimization results may have a significant impact on these parameters and initial conditions, but they may not be easy to determine or adapt to different problems or datasets.
- The lack of guarantees and robustness.

IV. CONCLUSION

In the context of education, the deployment of data-driven predictive models takes center stage in this investigation. It accentuates the critical need to integrate both qualitative and quantitative factors for the prediction and evaluation of students' academic performance. Illustrating the effectiveness of data mining methodologies like clustering, classification, and regression, the study addresses the multifaceted challenges proactively encountered by undergraduate students. The insights gleaned offer valuable guidance for policymakers, educational institutions, and students alike, with the shared objective of enhancing future academic outcomes. Moreover, the study introduces a cutting-edge strategy by merging the Extra-Trees Classifier (ETC) model with optimization algorithms, namely Gorilla Troops Optimizer (GTO) and Reptile Search Algorithm (RSA). This innovative approach demonstrates the potential of combining machine learning techniques and optimization algorithms to enhance the precision and effectiveness of predictive models. The outcome is a resilient toolkit designed to tackle the dynamic challenges inherent in students' academic journeys. The comprehensive evaluation undertaken in the study, involving the division of models into training and testing sets, unveils the considerable potential of these hybrid models to augment the classification capabilities of the ETC model. This improvement manifests in noteworthy enhancements in Accuracy and Precision. Upon scrutinizing the results, it becomes apparent that the recognition of the potential to significantly enhance the classification capabilities of the ETC model by these hybrid models is growing. In the context of G1 values, the enhancement of Accuracy, Recall, and F1-Score, achieved through the implementation of RSA and GTO optimization algorithms on the ETC model, was notable. The utilization of RSA and GTO resulted in a 1.56% improvement. The ETRS model, showcasing a remarkable Accuracy rate of 0.975, effectively and precisely classified the majority of students.

On the other hand, the ETGT and ETC models experienced misclassification rates of 6.33% and 7.6%, respectively. Turning attention to G3 values, the application of RSA and GTO optimization algorithms to the ETC model yielded significant improvements in all metrics' values in the testing phase. Nonetheless, this application led to a small reduction in the results of the training phase. The rate of Accuracy values' enhancement for RSA was 9.09%, and for GTO was 6.02% in the testing phase. When the 395 students were categorized based on their final grades, the prowess of the RSA algorithm in enhancing classification Accuracy became evident. The ETRS model demonstrated an impressive Accuracy rate of 0.953, adeptly classifying the majority of students. These findings underscore the efficacy of both RSA and GTO optimization algorithms in refining the predictive capabilities of the ETC model. Notably, the RSA algorithm exhibited a particularly commendable performance, showcasing its exceptional ability to enhance Accuracy, especially in the classification of students based on final grades (G3). The results suggest that the integration of these optimization algorithms holds promise for refining and optimizing the performance of models in educational contexts, contributing to more accurate and reliable student classifications.

References

[1] L. M. Mphale and M. B. Mhlauli, "An Investigation on students' academic performance for junior secondary schools in Botswana," European Journal of Educational Research, vol. 3, no. 3, pp. 111–127, 2014.

[2] K. Kriegbaum, N. Becker, and B. Spinath, "The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis," Educ Res Rev, vol. 25, pp. 120–148, 2018.

[3] L. D. Yulianto, A. Triayudi, and I. D. Sholihati, "Implementation Educational Data Mining For Analysis of Student Performance Prediction with Comparison of K-Nearest Neighbor Data Mining Method and Decision Tree C4. 5: Implementation Educational Data Mining for Analysis of Student Performance Prediction with Comparison of K-Nearest Neighbor Data Mining Method and Decision Tree C4. 5," Jurnal Mantik, vol. 4, no. 1, pp. 441–451, 2020.

[4] N. Martin Sanz, I. Rodrigo, C. Izquierdo GarcÃa, and P. Ajenjo Pastrana, "Exploring Academic Performance: Looking beyond Numerical Grades.," Universal Journal of Educational Research, vol. 5, no. 7, pp. 1105–1112, 2017.

[5] M. Pandey and V. K. Sharma, "A decision tree algorithm pertaining to the student performance analysis and prediction," Int J Comput Appl, vol. 61, no. 13, pp. 1–5, 2013.

[6] D. K. Kolo and S. A. Adepoju, "A decision tree approach for predicting students' academic performance," 2015.

[7] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: a review," Procedia Manuf, vol. 35, pp. 698–703, 2019.

[8] M. Apolinar-Gotardo, "Using decision tree algorithm to predict student performance," Indian J Sci Technol, vol. 12, p. 5, 2019.

[9] Y. S. Alsalman, N. K. A. Halemah, E. S. AlNagi, and W. Salameh, "Using decision tree and artificial neural network to predict students' academic performance," in 2019 10th international conference on information and communication systems (ICICS), IEEE, 2019, pp. 104–109.

[10] H. Sharma and S. Kumar, "A survey on decision tree algorithms of classification in data mining," International Journal of Science and Research (IJSR), vol. 5, no. 4, pp. 2094–2097, 2016.

[11] E. R. Jorda and A. R. Raqueno, "Predictive model for the academic performance of the engineering students using CHAID and C 5.0 algorithm," Int. J. Eng. Res. Technol, pp. 917–928, 2019.

[12] A. K. Verma and T. N. Singh, "Intelligent systems for ground vibration measurement: a comparative study," Eng Comput, vol. 27, pp. 225–233, 2011.

[13] M. H. I. Shovon and M. Haque, "An Approach of Improving Students Academic Performance by using k means clustering algorithm and Decision tree," arXiv preprint arXiv:1211.6340, 2012.

[14] A. B. Raut and M. A. A. Nichat, "Students performance prediction using decision tree," International Journal of Computational Intelligence Research, vol. 13, no. 7, pp. 1735–1741, 2017.

[15] T. M. Ogwoka, W. Cheruiyot, and G. Okeyo, "A model for predicting students' academic performance using a hybrid of K-means and decision tree algorithms," International Journal of Computer Applications Technology and Research, vol. 4, no. 9, pp. 693–697, 2015.

[16] S. Wiyono, D. S. Wibowo, M. F. Hidayatullah, and D. Dairoh, "Comparative study of KNN, SVM and decision tree algorithm for student's performance prediction," (IJCSAM) International Journal of Computing Science and Applied Mathematics, vol. 6, no. 2, pp. 50–53, 2020.

[17] S. Wiyono, T. Abidin, D. S. Wibowo, M. F. Hidayatullah, and D. Dairoh, "Comparative study of machine learning knn, svm, and decision tree algorithm to predict students' performance," International Journal of Research-Granthaalayah, vol. 7, no. 1, pp. 190–196, 2019.

[18] V. Matzavela and E. Alepis, "Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments," Computers and Education: Artificial Intelligence, vol. 2, p. 100035, 2021.

[19] R. Hasan, S. Palaniappan, A. R. A. Raziff, S. Mahmood, and K. U. Sarker, "Student academic performance prediction by using decision tree algorithm," in 2018 4th international conference on computer and information sciences (ICCOINS), IEEE, 2018, pp. 1–5.

[20] A. Hamoud, "Selection of best decision tree algorithm for prediction and classification of students' action," American International Journal of Research in Science, Technology, Engineering & Mathematics, vol. 16, no. 1, pp. 26–32, 2016.

[21] P. Strecht, J. Mendes-Moreira, and C. Soares, "Merging Decision Trees: a case study in predicting student performance," in Advanced Data Mining and Applications: 10th International Conference, ADMA 2014, Guilin, China, December 19-21, 2014. Proceedings 10, Springer, 2014, pp. 535–548.

[22] S. Sivakumar and R. Selvaraj, "Predictive modeling of students performance through the enhanced decision tree," in Advances in Electronics, Communication and Computing: ETAEERE-2016, Springer, 2018, pp. 21–36.

[23] A. K. Srivastava, A. Chaudhary, A. Gautam, D. P. Singh, and R. Khan, "Prediction of students performance using KNN and decision tree-a machine learning approach," Strad, vol. 7, no. 9, pp. 119–125, 2020.

[24] F. Chiheb, F. Boumahdi, H. Bouarfa, and D. Boukraa, "Predicting students performance using decision trees: Case of an Algerian University," in 2017 International Conference on Mathematics and Information Technology (ICMIT), IEEE, 2017, pp. 113–121.

[25] P. Guleria, N. Thakur, and M. Sood, "Predicting student performance using decision tree classifiers and information gain," in 2014 International conference on parallel, distributed and grid computing, IEEE, 2014, pp. 126–129.

[26] A. Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting student performance in higher education institutions using decision tree analysis," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, pp. 26–31, 2018.

[27] G.-H. Wang, J. Zhang, and G.-S. Fu, "Predicting student behaviors and performance in online learning using decision tree," in 2018 seventh international conference of educational innovation through technology (EITT), IEEE, 2018, pp. 214–219.

[28] P. Cheewaprakobkit, "Predicting student academic achievement by using the decision tree and neural network techniques," Human Behavior, Development And Society, vol. 12, no. 2, pp. 34–43, 2015.

[29] A. B. Adeyemo and G. Kuye, "Mining students' academic performance using decision tree algorithms," Journal of Information Technology Impact, vol. 6, no. 3, pp. 161–170, 2006.

[30] Q. A. Al-Radaideh, A. Al Ananbeh, and E. Al-Shawakfa, "A classification model for predicting the suitable study track for school students," Int. J. Res. Rev. Appl. Sci, vol. 8, no. 2, pp. 247–252, 2011.

[31] D. Thammasiri, D. Delen, P. Meesad, and N. Kasap, "A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition," Expert Syst Appl, vol. 41, no. 2, pp. 321–330, 2014.

[32] Q. A. Al-Radaideh, E. M. Al-Shawakfa, and M. I. Al-Najjar, "Mining student data using decision trees," in International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.

[33] N. T. Nghe, P. Janecek, and P. Haddawy, "A comparative analysis of techniques for predicting academic performance," in 2007 37th annual frontiers in education conference-global engineering: knowledge without borders, opportunities without passports, IEEE, 2007, pp. T2G-7.

[34] S. B. Aher and L. Lobo, "Data mining in educational system using weka," in International conference on emerging technology trends (ICETT), 2011, pp. 20–25.

[35] R. R. Kabra and R. S. Bichkar, "Performance prediction of engineering students using decision trees," Int J Comput Appl, vol. 36, no. 11, pp. 8–12, 2011.

[36] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," arXiv preprint arXiv:1201.3417, 2012.

[37] E. Osmanbegovic and M. Suljic, "Data mining approach for predicting student performance," Economic Review: Journal of Economics and Business, vol. 10, no. 1, pp. 3–12, 2012.

[38] S. K. Yadav and S. Pal, "Data mining: A prediction for performance improvement of engineering students using classification," arXiv preprint arXiv:1203.3832, 2012.

[39] D. Kabakchieva, "Student performance prediction by using data mining classification algorithms," International journal of computer science and management research, vol. 1, no. 4, pp. 686–690, 2012.

[40] C. Márquez-Vera, A. Cano, C. Romero, and S. Ventura, "Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data," Applied intelligence, vol. 38, pp. 315–330, 2013.

[41] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Mach Learn, vol. 63, pp. 3–42, 2006.

[42] M. F. Isham et al., "Bearing Fault Diagnosis Using Extreme Learning Machine Based on Artificial Gorilla Troops Optimizer," in Advances in Intelligent Manufacturing and Mechatronics: Selected Articles from the Innovative Manufacturing, Mechatronics & Materials Forum (iM3F 2022), Pahang, Malaysia, Springer, 2023, pp. 87–103.

[43] B. Abdollahzadeh, F. Soleimanian Gharehchopogh, and S. Mirjalili, " Artificial gorilla troops optimizer: a new nature - inspired metaheuristic algorithm for global optimization problems," International Journal of Intelligent Systems, vol. 36, no. 10, pp. 5887–5958, 2021.

[44] L. Abualigah, M. Abd Elaziz, P. Sumari, Z. W. Geem, and A. H. Gandomi, "Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer," Expert Syst Appl, vol. 191, p. 116158, 2022.

[45] M. K. Khan, M. H. Zafar, S. Rashid, M. Mansoor, S. K. R. Moosavi, and F. Sanfilippo, "Improved Reptile Search Optimization Algorithm: Application on Regression and Classification Problems," Applied Sciences, vol. 13, no. 2, p. 945, 2023.

[46] I. Al-Shourbaji, N. Helian, Y. Sun, S. Alshathri, and M. Abd Elaziz, "Boosting ant colony optimization with reptile search algorithm for churn prediction," Mathematics, vol. 10, no. 7, p. 1031, 2022.

[47] K. Hasib, F. Rahman, R. Hasnat, and M. G. R. Alam, A Machine Learning and Explainable AI Approach for Predicting Secondary School Student Performance. 2022. doi: 10.1109/CCWC54503.2022.9720806.

[48] P. Cortez and A. M. G. Silva, "Using data mining to predict secondary school student performance," 2008.

[49] S. S. Shreem, H. Turabieh, S. Al Azwari, and F. Baothman, "Enhanced binary genetic algorithm as a feature selection to predict student performance," Soft comput, vol. 26, no. 4, pp. 1811–1823, 2022.