Optimizing Student Performance Prediction: A Data Mining Approach with MLPC Model and Metaheuristic Algorithm

Qing Hai[1], Changshou Wang[2]*
Department of Water Resources and Civil Engineering, Hetao College, Bayan Nur 01500, Inner Mongolia, China[1]
Department of Agriculture, Hetao College, Bayan Nur 01500, Inner Mongolia, China[2]

***Abstract*—Given the information stored in educational databases, automated achievement of the learner's prediction is essential. The field of educational data mining (EDM) is handling this task. EDM creates techniques for locating data gathered from educational settings. These techniques are applied to comprehend students and the environment in which they learn. Institutions of higher learning are frequently interested in finding how many students will pass or fail required courses. Prior research has shown that many researchers focus only on selecting the right algorithm for classification, ignoring issues that arise throughout the data mining stage, such as classification error, class imbalance, and high dimensionality data, among other issues. These kinds of issues decreased the model's accuracy. This study emphasizes the application of the Multilayer Perceptron Classification (MLPC) for supervised learning to predict student performance, with various popular classification methods being employed in this field. Furthermore, an ensemble technique is utilized to enhance the accuracy of the classifier. The goal of the collaborative approach is to address forecasting and categorization issues. This study demonstrates how crucial it is to do algorithm fine-tuning activities and data pretreatment to address the quality of data concerns. The exploratory dataset utilized in this study comes from the Pelican Optimization Algorithm (POA) and Crystal Structure Algorithm (CSA). In this research, a hybrid approach is embraced, integrating the mentioned optimizers to facilitate the development of MLPO and MLCS. Based on the findings, MLPO2 demonstrated superior efficiency compared to the other methods, achieving an impressive 95.78% success rate.**

***Keywords*—Educational data mining; multilayer perceptron classification; pelican optimization algorithm; crystal structure algorithm; student performance**

I. Introduction

A. Background

Providing high-quality education to students is the primary goal of higher education establishments [1]. One strategy for achieving a better quality standard in a higher education program is to forecast pupils' academic success and intervene soon to raise pupil achievement and teacher quality [2]. Data mining techniques may be used to retrieve the useful knowledge concealed inside the educational data collection [3]. Against the backdrop of higher education, the current research aims to evaluate the potential of data-mining approaches by providing a data-mining model [4]. This activity aims to assess pupils' performance through categorization [5]. It is necessary to continuously assess how well pupils do in every topic. to pinpoint where the learner lost their grade [6]. This makes it easier for the educator to take the required steps, such as giving the student greater focus on that specific topic, teaching in a way that the student can understand quickly, giving tests, etc., all of which eventually raise the student's academic standing and quality [7]. Educational Data Mining (EDM) is the term for data mining within the education framework. Analytics has been used more in the previous few decades in educational settings [8], [9].

B. Related Works

On the provided dataset, six classifiers were used. At 79.23%, the ID3 had the highest accuracy [10]. The class mismatch challenge was beyond the model's ability to solve. To identify weak pupils, a model of ensembles such as classifiers (NB, SVM, and KNN) was suggested [11]. In addition to the common score-based evaluation, the data collection includes a characteristic referred to as standard-based grading evaluation. Comparing the outcomes of the suggested approach via six independent classifiers led to the conclusion that the ensemble model's accuracy was greater than the others at 85%. A multilayer classification model was put forth to overcome the multiple classifications issue regarding student performance prediction [12]. A methodology to give an early categorization of first-year students with poor educational outcomes was suggested by Dech Thammasiri et al. [13]. The class imbalance challenge was solved by applying four classifications and three balancing techniques. According to the results, the combination of SMOTE and support vector machines produced a maximum general precision of 90.24%. Students' performance in an online class may be predicted using information from their learning portfolios, according to one proposed early warning system [14]. The results showed that approaches based on time were more precise than those independent of time. Test the framework did not in offline mode. Using time-dependent properties, functioning might be reduced in offline mode.

Earlier research suggested that data mining algorithms only worked effectively with huge data sets; however, this study provided evidence that data mining may also be used for smaller datasets [15]. A model for predicting learner achievement was presented in this study. Several decision tree techniques were used for a small dataset containing students' academic data (Reptree, J48, M5P). According to the results, the Reptree had the best accuracy, exceeding 90 percent. The suggested model does not support class balance issues and data with large complexity. By grouping students into binary classes (successful/unsuccessful), Dorina et al. [16] presented a prediction model for students' performance. The suggested model was built using the research methodology of the Cross-Industry Standard Procedure for Data Mining, or CRISP-DM [17]. The provided dataset was subjected to the categorization methods OneR [18], MLP, J48, and IBK. The results showed that the MPL model was the most accurate at 73.59 percent in determining which students passed, while the other three models did a better job of determining which students failed. Issues with class balance and large complexity data were unsolvable for the model.

To overcome issues with disparities in classes and data complexity, Carlos et al. [19] focused on a machine learning-based failure of students' prognosis model. The dataset was utilized to execute ten classifiers. The accuracy of the ICRM classifier was found to be 92.7%, surpassing the performance of the other classifiers. The evaluation of the proposed model's performance was not conducted across various educational levels due to the distinct student characteristics associated with each level of education. Another EDM challenge is predicting which students will drop out of their classes [20]. Four data mining techniques with six characteristic pairings were employed in this study. The outcome reveals that, in data classification, superior performance was achieved when utilizing the support vector machine model that combined the variables. Adding a characteristic, achieved scores of prerequisite courses, in a data set was the study's restriction since it was feasible that the student had become more knowledgeable about the prerequisites for any course while studying for any other course. Research on pupil achievement prediction was carried out by Ajay et al. [21]. The main importance of the study was the introduction of a new social element, known as the CAT. The text elucidates the first categorization of Indians into four distinct groups based on their social standing and other variables that influenced student admission. The dataset underwent classification using four methods, namely R, MLP, J48, and IB1. Based on the available data, it can be shown that the IBI model has the highest level of accuracy, reaching 82%. Create an enhanced iteration of the ID3 method, which forecasts academic achievement in students [22]. The ID3 model's intention to choose those qualities as a node with additional values was one of its weaknesses. Consequently, the produced tree lacked efficiency. The suggested model resolves such an issue. This model generated the Pass and Failure output types. J48, wID3, and Naïve Bayes classifiers were used, and the outcomes were contrasted. An accuracy rate of 93% was attained with the wID3A model to forecast student achievement in courses presented in [23]. This study used three decision tree classifiers: Reptree, Hoeding tree, and J48. Reptree obtained the greatest accuracy of 91.47%. Problems with class balance and large dimensionality data were unsolvable for the model.

Through solving the data complexity issue, Edin Osmanbegovic et al. [24] was created a model to estimate the academic progress of students in a given course. This study evaluated many machine learning classifiers, such as NB, MLP, and j48. Based on the results, it can be observed that the Naïve Bayes model achieved the highest level of accuracy, reaching 76.65%. The issue of class imbalance is not addressed by the suggested model. In this paper, a model for predicting students' academic success was presented [25]. This study examined the classification methods with three different feature arrangements: J48, Decision Stump, Reptree, NB, and ANN. A high accuracy of 90.51% was attained with the J48 classifier. To forecast student abandonment, the suggested method took into account three numbers of courses that were assessed: dropout, persisting, and completing. Ten models of categorization were evaluated. According to the research's findings, for all three student classes, the Naïve Bayes algorithm achieved the greatest prediction values.

C. *Objective*

The fundamental aim of this research was to develop a robust machine-learning framework tailored to forecast student performance in Portuguese language courses, leveraging dependable data reservoirs. Through the strategic utilization of the Multilayer Perceptron Classification (MLPC) methodology, this study embarked on a path of innovation, ingeniously amalgamating two optimization algorithms: the Pelican Optimization Algorithm (POA) and the Crystal Structure Algorithm (CSA). This unique integration sought to improve both the accuracy and precision of the estimative model, thereby enriching the efficacy of prognostications regarding student performance. MLPC is chosen for predicting and classifying student performance in Portuguese language learning due to its ability to capture complex patterns inherent in language acquisition processes. By accommodating non-linear relationships between various factors influencing language proficiency and automatically learning feature representations from diverse datasets, MLPC offers scalability and robust generalization to unseen data. Moreover, its capacity for fine-tuning and potential for interpretability allows for continuous model improvement and insights into the determinants of student performance. Consequently, MLPC is a valuable tool for educators and stakeholders in effectively assessing and addressing student needs in Portuguese language education.

II. MATERIALS AND METHODS

A. *Data Gathering*

As previously elucidated, the prognostication of students' academic performance is shaped not only by their quiz outcomes, fulfilment of homework assignments, and engagement in class activities but also by the external circumstances they encounter outside the confines of the educational institution. For example, their family situation, the size of their family ($famsize$), family support ($famsup$), their health status, the amount of time spent on social media, their parents' occupation ($Fjob/Mjob$), and other relevant factors. Each of these terms influences the students' conditions in the classroom. However, the educational system's responsibility is to diagnose these factors, treat students according to their situations, act according to their talents, address their weaknesses, and capitalize on their strengths. The following diagram delineates the interplay between input and output variables. Notably, the school manifests a direct correlation with sex, suggesting the insignificance of students' gender. Likewise, while travel time lacks a direct association with students' failure, it does exert a marginal effect.

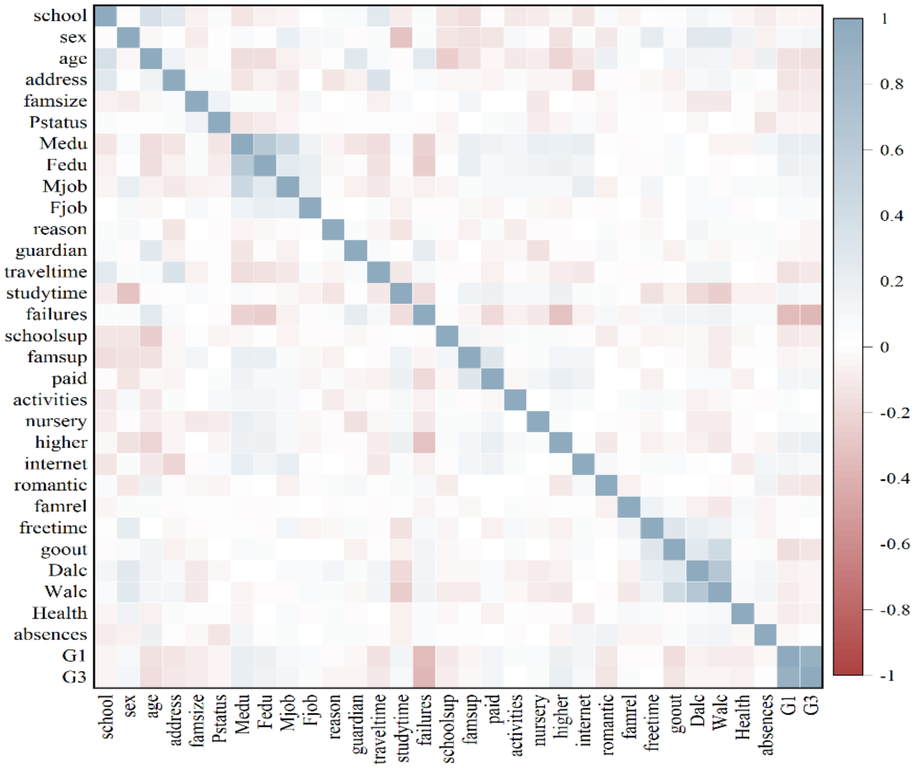


Fig. 1. Correlation matrix for the input and output variables.

Furthermore, study time exhibits no correlation with gender or school. In conclusion, although the diagnosis of these elements initially influences the prediction of students' performance, it is crucial to emphasize that none of these factors operates in isolation; instead, their effectiveness relies on the collaborative engagement of each student. Fig. 1 exhibits the correlation matrix for the in/output variables.

B. Multilayer Perceptron Classifier (MLPC)

Based on the concepts of neural network design, the Multilayer Perceptron Classifier (MLPC) is a particular kind of feed-forward artificial neural network (ANN) classifier. The MLPC in this configuration is made up of several layers of nodes, each of which is intimately linked to the network's next layer. Because of its architecture, the network can analyze and alter incoming data over a series of layers, which makes it possible for the MLPC to identify intricate patterns and correlations in the data. The nodes in the input layer of the MLPC represent the input data. Every node after it in the network uses its weight (shorthand for w) and an offset b to conduct a linear selection of the input as the data moves through the network. After this combination, an activation function transfers the input to the output. For improved clarity and illustration, this procedure may be concisely described in matrix form in the case of an MLPC with $K+1$ layers [26].

$$y(x) = f_k(\ldots f_2(w_2^T f_1(w_1^T + b_1) + b_2) \ldots + b_k) \quad (1)$$

Nodes inside the middle layer use the logistic or stochastic algorithm:

$$f(z_i) = \frac{1}{1 + e^{-z_i}} \quad (2)$$

The results of the layer's nodes use the $softmax$ feature:

$$f(z_i) = \frac{e_{z}^{\,i}}{\sum_{K-1}^{N} e^{z_k}} \quad (3)$$

The number of classes and nodes in the output layer has the matches.

C. Crystal Structure Algorithm (CSA)

Crystals are minerals with a structured composition that exhibit three regularly repeating or ordered crystalline dimensions. Crystalline solids can take on various sizes and shapes, and their properties may be either isotropic or anisotropic [27]. Crystals consist of small particles with well-defined shapes. Numerous physical and chemical compositions have been explored and suggested through experimentation. Moreover, crystals' complex symmetries and characteristics have profoundly influenced diverse human creations, including mechanisms, structures, and artworks. This article employs the Bravais model to explain the crystal structure. In this model, infinite lattice geometry is examined, and the periodic arrangement described by the lattice geometry, along with the vector of the lattice positions, is defined as follows:

$$l = \sum m_i e_i \quad (4)$$

where e_i is the minimum vector of the principal crystal directions, m_i is the and i is the angular number of the

$crystal$. Here, the basic idea of $Crystal$ is presented with appropriate modifications for the $CryStAl$ mathematical model. In this model, every candidate solution of the optimization method is likened to a distinct crystal space. To initiate the $cycle$, an arbitrary number of precious stones is selected.

$$\begin{bmatrix} cl_1 \\ cl_2 \\ \vdots \\ cl_i \\ \vdots \\ cl_m \end{bmatrix} = \begin{bmatrix} x_1^1 \dots & x_1^j \dots & x_1^p \\ x_2^1 \dots & x_2^2 \dots & x_2^p \\ \vdots & \vdots & \vdots \\ x_i^1 \dots & x_i^2 \dots & x_i^p \\ \vdots & \vdots & \vdots \\ x_m^1 \dots & x_m^2 \dots & x_m^p \end{bmatrix}, \begin{cases} i = 1,2,3,\dots,m \\ j = 1,2,3,\dots,p \end{cases} \quad (5)$$

where m is the candidate solution, and p is the dimension of the problem. Within the search space, the initial positions of these crystals are determined randomly by:

$$x_i^j(0) = x_{i,min}^j + \delta\left(x_{i,max}^j - x_{i,min}^j\right), \begin{cases} i = 1,2,3,\dots.,m \\ j = 1,2,3,\dots,p \end{cases} \quad (6)$$

where, $x_i^j(0)$ characterizes the starting gem position, the least and greatest permitted values are characterized as $x_{i,max}^j$ and $x_{i,min}^j$ separately, the jth choice variable of the ith candidate arrangement is within the indicated ς. Based on the crystallographic concept of the $base$, the primary crystals are all corner crystals. cl_{main} randomly determined considering the first generated crystal. In addition, the c_l the current value is ignored, and a random extraction method is set for each tread. $Crystals$ with optimal configuration determined by cl_z. S_u represents the mean of randomly selected crystals. To monitor the position of a candidate solution in the search space, four types of update procedures are established based on fundamental network principles:

Simple cubic;

$$cl_{new} = cl_{main} + cl_{old} \quad (7)$$

Best crystal cubicle;

$$cl_{new} = l_1 cl_{lmain} + l_2 cl_z + cl_{old} \quad (8)$$

Mean crystal cubicle;

$$cl_{new} = l_1 cl_{lmain} + l_2 S_u + cl_{old} \quad (9)$$

M&B crystal cubicle;

$$cl_{new} = cl_{old} + l_1 cl_{lmain} + l_2 cl_z + l_3 S_u \quad (10)$$

In the above formula, the old position is given by cl_{old} and the new position is denoted by cl_{new} and the random numbers are denoted by $l, l_1, l_2, and\ l_3$. Mining and exploration are the two main elements of metaheuristics, and it is worth mentioning that they have beesn tested in Eq. (7) to (10), where global and local searches are performed simultaneously. To deal with variable solutions x_i^j that violate the variable limit requirements, a mathematical flag is created that requires adjustment of the variable limits, causing problems with x_i^j they are exceeding the variable range. The termination criteria depend on the maximum number of iterations, which determines when the optimization process concludes after a fixed number of iterations [28], [29].

D. Pelican Optimization Algorithm (POA)

The researchers identified a population-based optimization method, known as the POA, which draws inspiration from pelicans [30]. The method employs a simulation of evolutionary processes within an ecological system, wherein pelicans are seen as single entities within a larger population. Every person represents a possible solution and provides optimization recommendations, which arise from adjusting the issue variable according to the position of each person in the search area. In order to ensure the variety of the population and improve the global search capacity, each member is randomly initialized within the stated upper and lower limits of the issue during the population initialization procedure, as illustrated in Eq. (11).

$$x_{i,j} = l_j + rand.\left(u_j - l_j\right), i = 1,2,\dots,N, \\ j = 1,2,\dots,m \quad (11)$$

Where N is the number of population members, m is the number of issue variables, $rand$ is a random integer in the interval $[0, 1]$, l_j is the jth lower bound, and u_j is the jth upper limit of problem variables. The values of the variables indicated by the ith candidate solution are represented by the variables $x_{i,j}$. Eq. (12) uses a matrix known as the population matrix to identify the pelican population members in the proposed POA. The columns of this matrix show the suggested values for the issue variables, and each row indicates a potential solution.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}_{N\times m} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots & & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m} \end{bmatrix}_{N\times m} \quad (12)$$

If X_i is the ith pelican, and X is the pelican population matrix. A potential fix for the stated issue is the planned POA, in which every member of the population is a pelican. Thus, assessing the given issue's objective function is possible by considering each potential solution. The objective function vector in Eq. (13) is used to derive the values obtained for the objective function.

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N\times 1} \quad (13)$$

where, F_i is the objective function value of the ith candidate solution, and F is the objective function vector. To update potential answers, the suggested POA mimics the tactics and behaviour of pelicans during hunting and assault. There are two phases to simulating this hunting strategy: i Approaching the prey (the period of exploration). ii Winging during the exploitation phase on the water's surface.

1) Phase 1 (exploration phase): approaching the prey: The initial stage of the process involves the pelicans locating the prey and then approaching it. Search space scanning and the exploration capability of the suggested POA in locating various search space regions are made possible by modelling this pelican approach. The fact that the prey's position is produced is crucial to POA.

$$x_{i,j}^{P_1} = \begin{cases} x_{i,j} + rand.\left(p_j - I.x_{i,j}\right), & F_P < F_i; \\ x_{i,j} + rand.\left(x_{i,j} - p_j\right), & else, \end{cases} \tag{14}$$

Where x^{P1}. In the context of Eq. (14), the importance of the variable can be observed $x_{i,j}^{P_1}$, an updated state of the pelican in the jth dimension is represented by the result of stage 1, and this can be the ith pelican. To introduce additional diversity and exploration, the value of I is introduced as a random number ranging between one and two. Also, the parameter pj, the position of the prey, is employed to be denoted jth dimension, while F_P the objective function value of the prey is represented. By incorporating Eq. (15), the process can be effectively simulated and modelled.

$$X_i = \begin{cases} X_i^{P_1}, F_i^{P_1} < F_i; \\ X_i \quad else, \end{cases} \tag{15}$$

Where X^{P_1}. This is the updated status for the F^{P_1} and ith pelican. The goal function is based on values pertaining to the phase.

2) Phase 2: winging on the water surface (exploitation phase): In the subsequent stage, the pelicans gather their meal in their throat pouches after reaching the water's surface and spreading their wings to push the fish upward. This tactic helps pelicans catch more fish in the assaulted region. As a result of simulating this pelican behaviour, the suggested POA converges to more advantageous locations inside the hunting region. The exploitation potential and local search power of POA are enhanced by this method. From a mathematical perspective, the algorithm must look at the points surrounding the pelican position to converge to an optimal solution. Eq. (16) simulates the hunting behaviour of pelicans mathematically.

$$x_{i,j}^{P_2} = x_{i,j} + R\left(1 - \frac{t}{T}\right).(2.rand - 1).x_{,j} \tag{16}$$

Where X^{P_2}, based on phase 2, i,j represents the ith pelican's new state in the jth dimension. $x_{i,j}, s$ neighbourhood radius is given by $R\left(1 - \frac{t}{T}\right)$, which is equal to 0.2. T represents the maximum number of iterations and iteration counter. The exponent $R\left(1 - \frac{t}{T}\right)$ reflects the local search radius for the population members' neighbourhoods. Close to every participant to arrive at an improved answer. This coefficient works well on the POA exploitation power to reach the ideal global solution. Since this coefficient is highly valued in the first iterations, a bigger region is considered around each member. The $R\left(1 - \frac{t}{T}\right)$ The coefficient falls as the method replicates more, resulting in smaller radii for each neighbourhood member. For the POA to converge to solutions that are closer to the global (and even precisely global) ideal based on the utilization notion, this enables us to scan the region surrounding each member of the population in smaller and more precise stages. Eq. (17) models the new pelican posture, which has also been accepted or rejected at this stage by successful updating.

$$X_i = \begin{cases} X_i^{P_2}, F_i^{P_2} < F_i; \\ X_i, \quad else, \end{cases} \tag{17}$$

where X^{P_2}. This is the updated status for the F^{P_2} and ith pelican. Its goal function is value-based, and i on stage 2.

3) Steps repetition, pseudo-code, and flowchart of the proposed POA: The best candidate solution up to this point will be updated after all population members have been updated based on the first and second phases, the population's new status, and the values of the goal function. When the algorithm reaches the next iteration, the stages of the suggested POA are based on Eq. (15) to (17) are repeated until the execution is finished. Lastly, a quasi-optimal solution to the given issue is offered using the best candidate solution found throughout the algorithm rounds.

E. Performance Evaluators

When evaluating a classifier's performance, it is essential to consider multiple criteria to obtain a thorough insight into its effectiveness. These criteria function as metrics, providing insights into various aspects of the classifier's performance and enabling a nuanced assessment. Here are some crucial factors to consider:

- Accuracy: A frequently employed metric measures the classifier's efficiency by determining the percentage of accurately predicted samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

- Recall: Recall quantifies the proportion of correctly predicted positive instances in relation to all actual positive instances.

$$Recall = TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{19}$$

- Precision: Precision centres on the precision of positive predictions, evaluating the probability that instances identified as positive are indeed accurate. This metric is particularly valuable when the cost of false positives is significant.

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

- F1-score: The combination of Precision and Recall yields a composite measure recognized as the f1-score.

$$F1\ score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{21}$$

In Eq. (18) to (21), TP represents a positive prediction that correctly corresponds to the actual positive outcome. FP denotes a positive prediction when the actual outcome is negative. FN is used to indicate a negative prediction when the actual outcome is positive, while TN represents a negative prediction that accurately aligns with the actual negative outcome.

III. Result

A. Hyperparameters and Convergence Curve Results

Hyperparameters are external settings that encompass vital factors like learning rates and regularization strengths, exerting significant influence over a behavior of model. They are predetermined values and are not directly inferred from the dataset itself. Maximizing model performance relies heavily on the indispensable task of fine-tuning hyperparameters, which necessitates rigorous experimentation and the adept application of optimization methodologies. The results of the hyperparameters for MLPC-based hybrid models (MLPO and MLCS) are represented in Table I for G2 and Table II for G3 values. The hyperparameter of the MLPC-based models is Layer_size. This comprehensive exposition substantially enhances the transparency and reproducibility of models within the field of machine learning research, furnishing invaluable insights that deepen understanding and facilitate precise replication of model configurations.

TABLE I. Result of Hyperparameters for G2

Layer of MLPC	*Models*	*Hyperparameter*	
		MLPO	*MLCS*
Layer 1	*Layer_size*	*74*	*71*
Layer 2	*Layer_size*	*22*	*69*
	Layer_size	*27*	*16*
Layer 3	*Layer_size*	*46*	*29*
	Layer_size	*37*	*56*
	Layer_size	*13*	*31*

TABLE II. Result of Hyperparameters for G3

Layer of MLPC	*Models*	*Hyperparameter*	
		MLPO	*MLCS*
Layer 1	*Layer_size*	*32*	*53*
Layer 2	*Layer_size*	*23*	*86*
	Layer_size	*19*	*99*
Layer 3	*Layer_size*	*12*	*34*
	Layer_size	*13*	*19*
	Layer_size	*12*	*20*

This study aims to forecast learners' academic achievement throughout the educational program to improve their skills and increase their chances of success. The MLPC model, which combines the two optimizers known as POA optimization and CSA, is presented to achieve this aim. The model has a favourable impact on the prediction of pupil achievement. In this study, two novel models, MLCS and MLPO, are generated by integrating the foundational model, MLPC, with optimizers to enhance prognostic capabilities further. This section encompasses a comparative analysis to determine the relative effectiveness of each model over the others. A hybrid model's convergence is typically understood to signify that it has reached its peak throughout the training process. When a machine learning algorithm gets to a point where more iterations of training do not significantly improve the model's performance on the training set, it stabilizes its parameters and becomes a convergent state. This is especially true for complex models like hybrid models. In the context of hybrid models, characterized by incorporating multiple model or technique types, achieving convergence necessitates verification that each constituent functions as intended and that the model achieves overall prediction consistency. The monitoring of convergence during the training phase commonly involves observing effectiveness indicators or examining loss functions on a validated dataset. Rapid convergence is imperative for a hybrid model to comprehend knowledge structures and effectively generalize its findings to novel, unseen data. Fig. 2 and 3 comprehensively compare models across two distinct targets, G2 and G3, encompassing three layers. In the initial layer of the G2 target, the MLCS model achieves stability at a core value of 0.889 within 90 iterations, in contrast to the MLPO model, which attains stability at a point of 0.899 in 120 iterations.

Although the MLPO model maintains a higher accuracy than the MLCS model in the second layer, achieving stability at 0.939 within 130 iterations, compared to the MLCS model's accuracy of 0.927 measured in 128 iterations. Examination of the third layer underscores the MLPO model's superior accuracy in the G2 target, reaching an estimated value of 0.919 within 130 iterations, in contrast to the MLCS model, with a measured value of 0.904 in 150 iterations. Incidentally, in the first layer of the G3 target, the MLCS model, with an accuracy of 0.861 measured in 148 iterations, is surpassed by the MLPO model, which achieves a higher accuracy of 0.878 within 150 iterations. Subsequently, in the second layer, the MLCS model reaches a level of 0.901 in the 90th iteration; however, the MLPO model outperforms it with an accuracy of 0.914 measured in the 148th iteration. Ultimately, in the third layer, the MLPO model attains an accuracy of 0.891 after 150 iterations. Conversely, the MLCS model exhibits weaker performance with a lower accuracy measurement than the MLPO model. Ultimately, the current line plot reveals that higher accuracy is achieved by the POA optimizer when combined with the base model across three layers of two targets.

B. Results of Predictive Models

Table III delineates measured values of accuracy, precision, recall, and F1 score across three phases—namely, train, test, and all—within the G2 and G3 targets, each comprising three layers. For instance, during the training phase, the MLPO model exhibits values of 0.910, 0.913, 0.910, and 0.909 for accuracy, precision, recall, and F1-score, respectively. In contrast, the MLCS model in the training phase records values of 0.906, 0.910, 0.906, and 0.905 for the corresponding metrics. This comparative analysis underscores that the MLPO model consistently attains higher accuracy in each of the four metrics than the MLCS model in the same phase. However, during the test phase for both models, precision emerged as the metric with the highest value, specifically registering at 0.888 for the MLPO model and 0.863 for the MLCS model.

Incidentally, in the second layer, the aggregate precision value in the MLPO model is 0.940, surpassing the corresponding value of 0.927 in the opposing model. Notably, the MLCS model maintains uniform values across all phases for three metrics—accuracy, precision, and recall—except for the F1-score, where it records a lower value of 0.926, indicating inferior accuracy compared to other metrics. In the final layer, the MLPO model achieves accuracy values of 0.931 and 0.890 in the train and test phases, respectively.

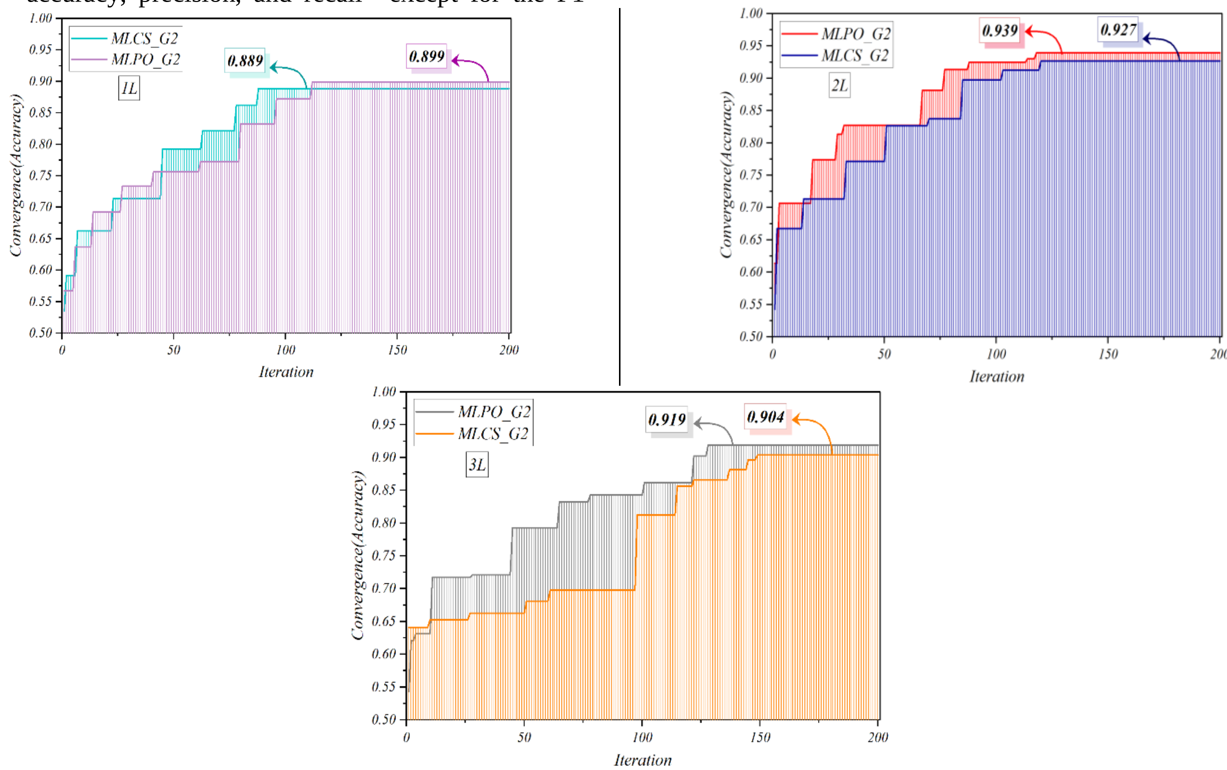


Fig. 2. Line plot for convergence of hybrid models for G2 values.

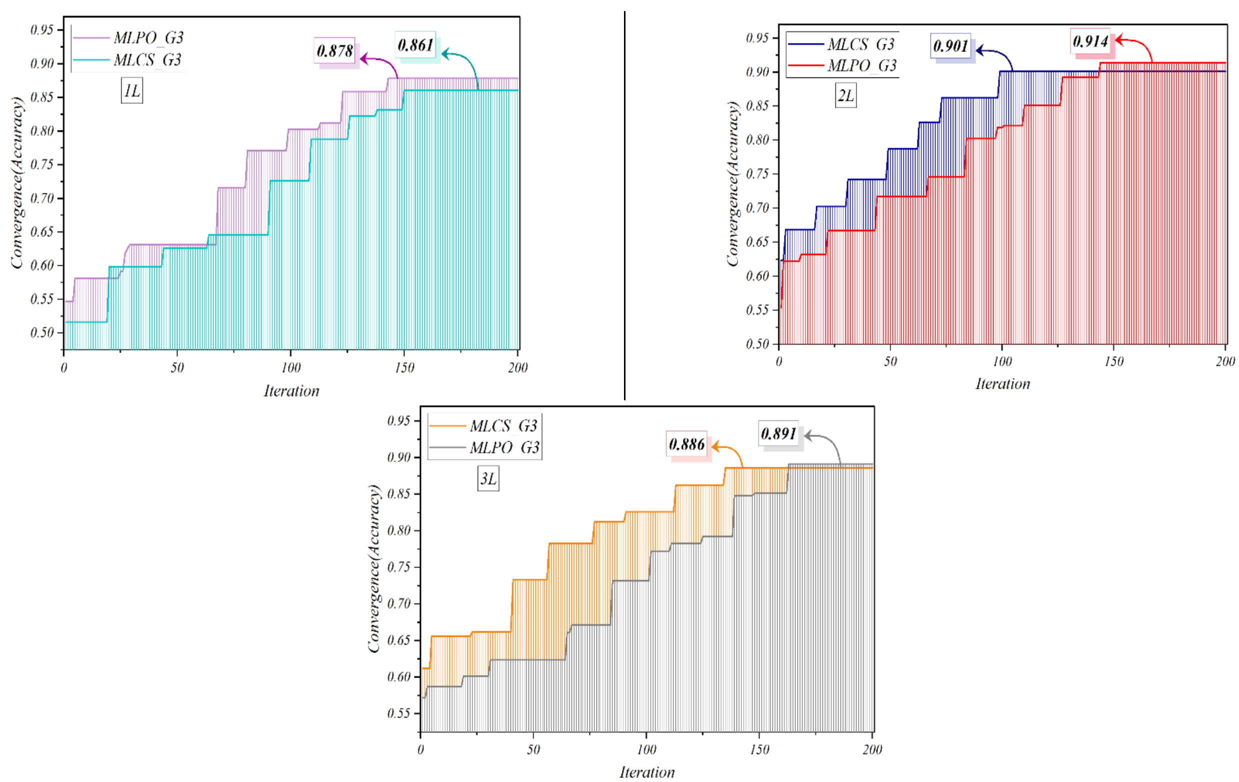


Fig. 3. Line plot for convergence of hybrid models for G3 values.

TABLE III. RESULT OF PRESENTED MODEL FOR G2

Model	Section	Index values			
		Accuracy	***Precision***	***Recall***	***F1_score***
MLPO (1)	Train	0.910	0.913	0.910	0.909
	Test	0.873	0.888	0.873	0.870
	All	0.899	0.905	0.899	0.898
MLCS (1)	Train	0.906	0.910	0.906	0.905
	Test	0.847	0.863	0.848	0.846
	All	0.889	0.892	0.889	0.888
MLPO (2)	Train	0.953	0.954	0.953	0.953
	Test	0.907	0.9079	0.907	0.903
	All	0.939	0.940	0.939	0.939
MLCS (2)	Train	0.946	0.949	0.946	0.945
	Test	0.881	0.891	0.881	0.878
	All	0.927	0.927	0.927	0.926
MLPO(3)	Train	0.931	0.934	0.931	0.932
	Test	0.890	0.893	0.890	0.886
	All	0.919	0.921	0.919	0.919
MLCS (3)	Train	0.921	0.921	0.921	0.920
	Test	0.864	0.873	0.864	0.865
	All	0.904	0.905	0.904	0.904

In contrast, the MLCS model exhibits values of 0.921 and 0.864 for the corresponding terms in the train and test phases, illustrating that, in both terms and phases, the MLPO model consistently outperforms the MLCS model in accuracy. In summary, it is noteworthy that the MLPO model demonstrates superior performance compared to the MLCS model. Examining two models within the G3 target across three layers reveals that the MLPO model consistently maintains higher accuracy than its counterpart. For instance, in the recall term during the training phase, the MLPO model achieves a value of 0.892, whereas the MLCS model records a slightly lower value of 0.888. Similarly, in the F1-score metric, the MLPO model attains a value of 0.890, surpassing the MLCS model's F1-score of 0.886. This subtle comparison unequivocally underscores the MLPO model's superior accuracy compared to the MLCS model. Concerning the second layer, the accuracy values in the test phase for the MLPO and MLCS models are 0.864 and 0.856, respectively. Additionally, the precision values of the MLPO and MLCS models are 0.873 and 0.866, respectively. This observation signifies that the accuracy of the MLPO model surpasses that of the opposing model. Moreover, the MLPO model exhibits superior performance in the third layer. For a more thorough comprehension, it is noteworthy that the accuracy and precision values of the MLPO model across all phases are higher than the corresponding values of the MLCS model, with the accuracy comparison being 0.891 > 0.886. Ultimately, the accuracy of the MLPO model surpasses that of the MLCS model in each layer of both targets. This comparison is presented in Table IV for further examination.

C. *Results of Classification Processes*

The comparison between the MLPO and MLCS models in two targets is illustrated in Tables V and VI, elucidating the layer-wise accuracy of each model. Analogous to the preceding tables, these tables contrast grades instead of phases. Notably, the precision values of the MLPO model in the excellent grade across the first, second, and third layers are recorded as 0.77, 0.88, and 0.84, respectively. This observation suggests that the model's optimal performance is evident in the second layer, outperforming the other layers. The recall values for the good and acceptable grades in the MLPO model are 0.64 and 0.88 in the first layer, 0.76 and 0.88 in the second layer, and 0.76 and 0.88 in the third layer. This implies that the performance of the MLPO model is consistent in the second and third layers, while it is comparatively lower in the first layer. In the first layer, MLCS exhibits precision values of 0.74 and 0.96 for the acceptable and poor grades, respectively. In the second layer, the corresponding precision values are 0.85 and 0.96; in the third layer, they are 0.86 and 0.94, respectively. This analysis indicates that the model achieves higher accuracy in the third layer for the acceptable grade, surpassing the accuracy in the first and second layers. However, in the case of poor grades, the functionality is optimal in the first and second layers, contrasting with the third. It is pertinent to note that this comparison pertains to the G2 target. Contrastingly, within the G3 target, the recall values for the MLPO model in the first layer are 0.76 for excellent grade and 0.65 for good grade; in the second layer, they are 0.87 for excellent grade and 0.78 for good grade, and in the last layer, they are 0.81 for excellent grade and 0.83 for good grade. These statistics reveal that the model demonstrates heightened accuracy in the excellent grade of the second layer compared to the other layers. However, in the context of the good grade, the second layer exhibits superior functionality compared to the first and third layers.

TABLE IV. RESULT OF PRESENTED MODEL FOR G3

Model	Section	Index values			
		Accuracy	***Precision***	***Recall***	***F1_score***
MLPO (1)	Train	0.892	0.893	0.892	0.890
	Test	0.847	0.846	0.848	0.841
	All	0.878	0.879	0.879	0.876
MLCS (1)	Train	0.888	0.885	0.888	0.886
	Test	0.822	0.8278	0.822	0.822
	All	0.861	0.858	0.861	0.859
MLPO (2)	Train	0.935	0.937	0.935	0.935
	Test	0.864	0.873	0.864	0.866
	All	0.914	0.918	0.914	0.914
MLCS (2)	Train	0.921	0.924	0.921	0.921
	Test	0.856	0.866	0.856	0.858
	All	0.901	0.905	0.901	0.902
MLPO (3)	Train	0.917	0.920	0.917	0.918
	Test	0.831	0.845	0.831	0.832
	All	0.891	0.896	0.891	0.893
MLCS (3)	Train	0.906	0.909	0.906	0.905
	Test	0.839	0.862	0.839	0.841
	All	0.886	0.893	0.886	0.886

TABLE V. PERFORMANCE OF PRESENTED MODELS BASED ON THE GRADES IN G2

Model	Grade	Index values		
		Precision	***Recall***	***F1-score***
MLPO (1)	Excellent	0.77	0.86	0.81
	Good	1.00	0.64	0.78
	Acceptable	0.85	0.88	0.86
	Poor	0.95	0.96	0.95
MLCS (1)	Excellent	0.81	0.77	0.79
	Good	0.92	0.67	0.77
	Acceptable	0.74	0.88	0.80
	Poor	0.96	0.96	0.96
MLPO (2)	Excellent	0.88	0.94	0.91
	Good	0.93	0.76	0.83
	Acceptable	0.89	0.88	0.89
	Poor	0.97	0.98	0.98
MLCS (2)	Excellent	0.87	0.88	0.88
	Good	0.96	0.73	0.83
	Acceptable	0.85	0.88	0.86
	Poor	0.96	0.98	0.97
MLPO (3)	Excellent	0.84	0.87	0.86
	Good	0.96	0.76	0.85
	Acceptable	0.82	0.88	0.85
	Poor	0.97	0.97	0.97
MLCS (3)	Excellent	0.81	0.82	0.82
	Good	0.96	0.82	0.89
	Acceptable	0.86	0.89	0.88
	Poor	0.94	0.95	0.94

TABLE VI. PERFORMANCE OF PRESENTED MODELS BASED ON THE GRADES IN G3

Model	Grade	Index values		
		Precision	***Recall***	***F1-score***
MLPO (1)	Excellent	0.80	0.76	0.78
	Good	0.90	0.65	0.75
	Acceptable	0.76	0.83	0.79
	Poor	0.93	0.96	0.94
MLCS (1)	Excellent	0.72	0.74	0.73
	Good	0.83	0.75	0.79
	Acceptable	0.76	0.68	0.72
	Poor	0.93	0.96	0.94
MLPO (2)	Excellent	0.82	0.87	0.84
	Good	0.94	0.78	0.85
	Acceptable	0.80	0.88	0.84
	Poor	0.97	0.96	0.96
MLCS (2)	Excellent	0.78	0.81	0.79
	Good	0.91	0.73	0.81
	Acceptable	0.78	0.90	0.84
	Poor	0.97	0.96	0.96
MLPO (3)	Excellent	0.71	0.81	0.76
	Good	0.80	0.83	0.81
	Acceptable	0.83	0.80	0.81
	Poor	0.98	0.95	0.96
MLCS (3)	Excellent	0.72	0.79	0.75
	Good	0.96	0.68	7941.00
	Acceptable	0.79	0.87	0.83
	Poor	0.95	0.95	0.95

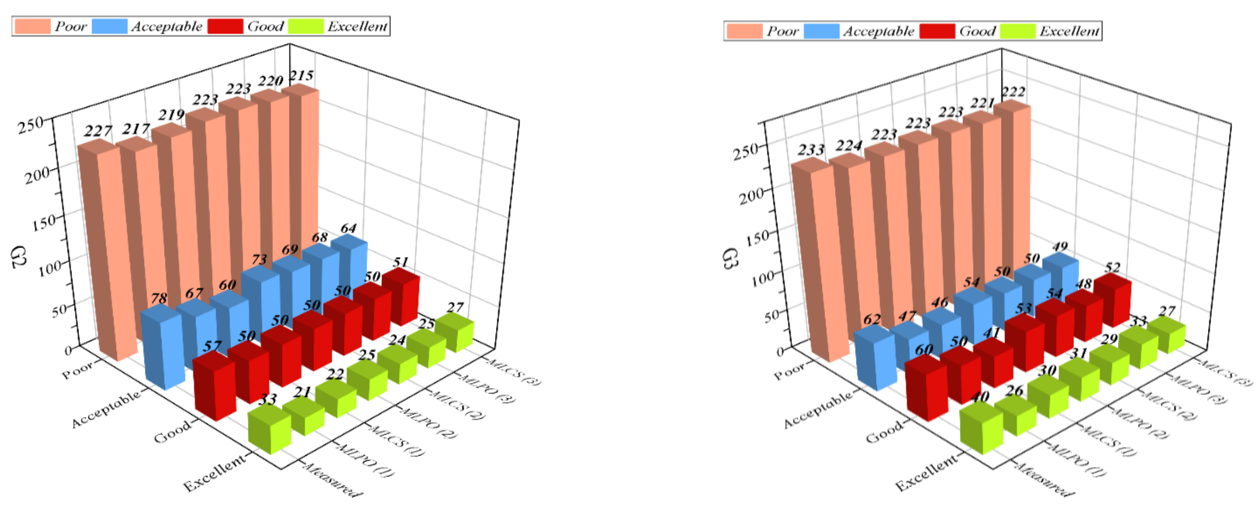


Fig. 4. 3D Bars plot for comparing the measured and predicted values.

Conversely, the recall values of the MLCS model in the first layer are 0.74 for excellent grade and 0.75 for good grade; in the second layer, they are 0.81 for excellent grade and 0.73 for good grade, and in the last layer, they are 0.79 for excellent grade and 0.68 for good grade. These figures suggest that the MLCS model exhibits enhanced functionality in the excellent grade of the second layer. However, concerning the good grade, it is noteworthy to emphasize that this model in the first layer attains higher accuracy than the second and third layers.

Fig. 4 delineates the comparison between each layer's measured and predicted values for the MLPO and MLCS models in the G2 and G3 targets. The illustration of this plot reveals that the accuracy of the MLPO model in the first layer attains 21 out of 33 measured values. In the second layer, it achieves 25 out of 33 measured values; in the last layer, it similarly attains 25 out of 33 measured values. This observation underscores that the MLPO model demonstrates the highest accuracy in the excellent grade for the second and third layers, surpassing the accuracy observed in the first layer. Nevertheless, in the good grade, the accuracy of the MLPO and MLCS models across all three layers is recorded at 50 out of 57 measured values, except for MLCS in the third layer, which achieves a measured value of 51 out of 57. It is discernible that this model exhibits superior functionality in the third layer compared to other models across all three layers.

Regarding the acceptable grade, the MLPO model attains the highest accuracy in the second layer, with 73 out of 78 measured values. The second-highest accuracy in the same layer is observed for the MLCS model, recording 69 out of 78 measured values. In contrast, the third-highest accuracy is attributed to the MLPO model in the first layer, achieving a measured value of 67 out of 78. The comparative analysis elucidates that superior performance is evident in both models within the second layer when contrasted with other models in different layers. Nevertheless, within the same target, the measured value of the MLCS model in the third layer amounts to 215 out of 227, indicative of the lowest measured value across all layers among the models. The second-highest performance is attributed to the MLPO model in the first layer, achieving a measured value of 217 out of 227, while the third-highest performance is observed for the MLCS model in the same layer.

On the contrary, ascendancy is asserted by both the MLPO and MLCS models in the second layer, attaining 223 out of 227 measured values. In the subsequent target, parity is observed between the MLCS and MLPO models in the second layer and the MLCS model in the first layer, registering 223 out of 233 measured values. However, the MLPO model in the first layer stands out with the highest accuracy, recording 224 out of 233 measured values, particularly notable in the context of the poor grade. For the acceptable grade, equivalence is noted as both models in the third and second layers exhibit identical measured values of 50 out of 62, representing the maximum accuracy among models across all three layers. The MLCS model in the third layer achieves the second-highest performance, recording 49 out of 62 measured values.

Conversely, the least favourable measured value is attributed to the MLCS model in the first layer. Notably, in the good grade, optimal performance is observed in the MLCS model within the second layer, achieving the highest measured value of 54 out of 60. In contrast, the least favourable performance in this grade is associated with the same model but in the first layer, registering 41 out of 60 measured values. In the highest grade, excellent, the highest accuracy is attained by the MLPO model in the third layer, achieving 33 out of 40 measured values. The second-highest performance in this grade is noted for the MLPO model in the second layer, with a measured value of 31 out of 40. In contrast, the third-highest performance is attributed to the MLCS model in the first layer, recording 30 out of 40 measured values. The lowest performance in the excellent grade is associated with the MLPO model in the first layer, registering 26 out of 40 measured values.

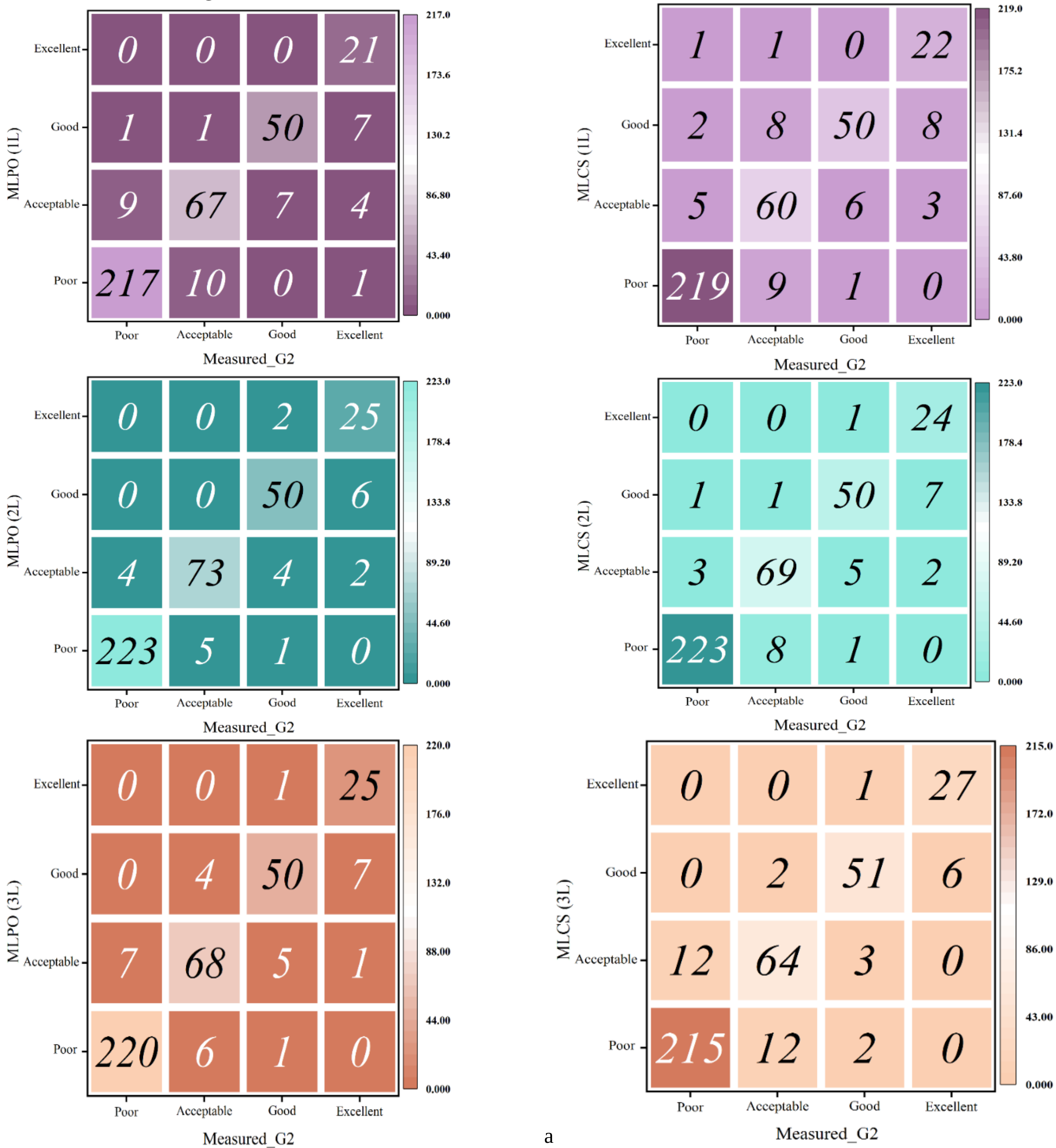
The accuracy of the models in the confusion matrix across all three layers for two targets is depicted in Fig. 5 and 6. The performance of the MLPO models in the G2 target, specifically in layer one, is observed. In instances characterized by a suboptimal grade, the recorded value is 217 out of 227, reflecting a difference of 4.5%. Additionally, nine students are misclassified in an acceptable grade and one in a good grade. Similarly, within the same layer, the value for acceptable grades is 67 out of 78, indicating a difference of 15.17%. In this context, ten students are misclassified as having a poor grade and one as having a good grade. The MLCS model's measured value in the second layer of a suboptimal grade is 219 out of 223, reflecting a marginal difference of 1.81%. This outcome entails misclassifying five students in an acceptable grade, two in a good grade, and one in an excellent grade.

Conversely, the measured value for acceptable grades in the first layer is 60 out of 62, with a difference of 3.28%, accompanied by the misclassification of nine students with poor grades, eight with good grades, and one with excellent grades. Furthermore, in the second layer, the measured value of the MLPO model in a good grade is 50 out of 60, demonstrating an 18.18% difference and involving the misclassification of two students in an excellent grade, four in an acceptable grade, and one in a good grade. Incidentally, within the G2 target of an excellent grade, the MLPO model exhibits a difference of 27.59% in the second layer, entailing the misclassification of six students in a good grade and two students in an acceptable grade. Conversely, the MLCS model in the G3 target of the current grade manifests a 50% difference, accompanied by the misclassification of seven students in a good grade and two in an acceptable grade. Additionally, in the G2 target, the MLPO model in a poor grade of the second layer demonstrates a 5.74% difference, resulting in the misclassification of seven students in an acceptable grade. Simultaneously, within an acceptable grade, it showcases a 6.62% difference, leading to the misclassification of five students with poor grades.

Regarding the good grade, it is imperative to note that it exhibits a 13.8% difference, resulting in the misclassification of four students in an acceptable grade, one in a poor grade, and two in an excellent grade. In the same target and layer, the MLCS model demonstrates a 1.78% difference in the poor grade category, leading to the misclassification of three students in an acceptable grade and one in a good grade. Meanwhile, a 12.24% difference is observed for the acceptable grade, entailing the misclassification of one student in a good grade and eight students in a poor grade. In the third layer of the G2 target, the MLPO model is observed to misclassify seven students with a good grade and one student with an acceptable grade, reflecting a 27.59% difference. This denotes the performance of the current model in an excellent grade.

Conversely, the MLCS model in the same target, layer, and grade exhibits a 20% difference, accompanied by the misclassification of six students with good grades. Upon reaching the G3 target and assessing the models' performance in the first layer, a discernible discrepancy is observed between the MLCS model and the MLPO model in the context of an excellent grade. Specifically, the MLPO model manifests a substantial 42.4% difference in an excellent grade, accompanied by the misclassification of eleven students in a good grade, one in an acceptable grade, and two in a poor grade. In contrast, the MLCS model exhibits a 28.57% difference, misclassifying eight students with good grades and two with acceptable grades. In the second layer, the MLPO model demonstrates a 12.39% difference, with the misclassification of two students in an excellent grade, two in an acceptable grade, and one in a poor grade.

Conversely, the MLCS model in the same layer features a 10.53% difference, entailing the misclassification of three students with excellent grades and three with acceptable grades. Nevertheless, a singular examination of one stage for each model might suggest that the MLCS model exhibits superior accuracy compared to the alternative model. However, when considering the comprehensive assessment across all layers, it becomes apparent that the functionality of the MLPO model surpasses that of the MLCS model in each respective layer.



a

Fig. 5. Confusion matrix for accuracy of each model for G2 values.

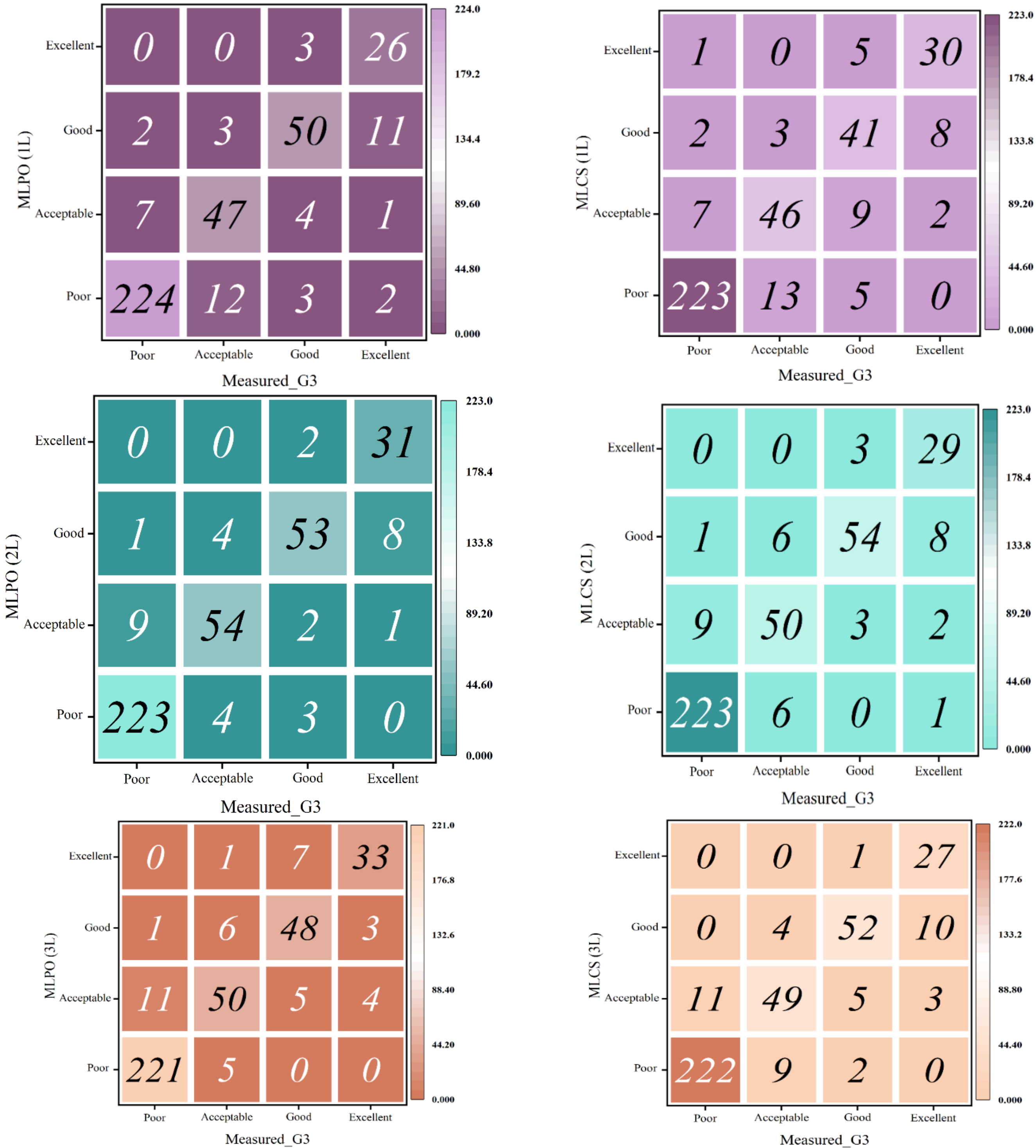


Fig. 6. Confusion matrix for accuracy of each model accuracy for G3 values.

The subsequent column plot illustrates the percentage achievements of the developed models. Specifically, within the G2 target, the MLPO model in the second layer attains the highest accuracy at 0.93924, followed by the MLCS model in the same layer with a percentage of 0.92658, securing the second rank. The MLPO model in the third layer holds the third rank with a percentage of 0.91899. This concise comparison indicates that the MLPO model in the second layer exhibits superior functionality compared to the other layers. Nevertheless, the MLPO model in the second layer is characterized by superior precision relative to the other models, achieving a percentage of 0.9395. The MLCS model in the second layer and the MLPO model in the third layer secure the second and third ranks, respectively, with percentages of 0.9274 and 0.9214, respectively.

Additionally, the recall and F1-score values of the MLPO model, standing at 0.9392 and 0.9385, surpass those of the alternative models. In summary, the performance of MLPO L2 not only outperforms the MLCS model but also exceeds its performance in other layers. Upon a cursory examination, it becomes evident that MLPO L2 in the G3 target attains elevated accuracy, precision, recall, and F1-score values. The column plots in Fig. 7 and 8 illustrate the achievement percentage for developed models as assessed by evaluators.

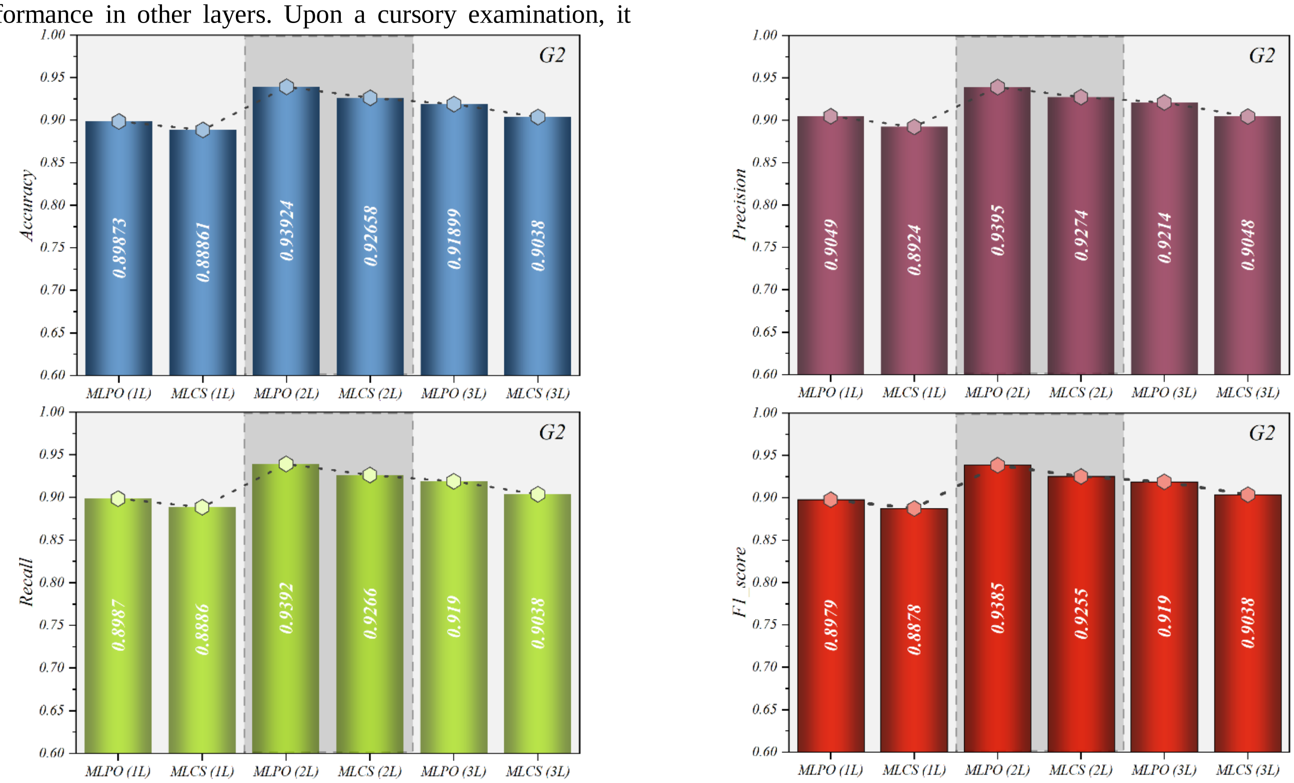


Fig. 7. Column plots the achievement percentage for developed models of G2 prediction values based on evaluators.

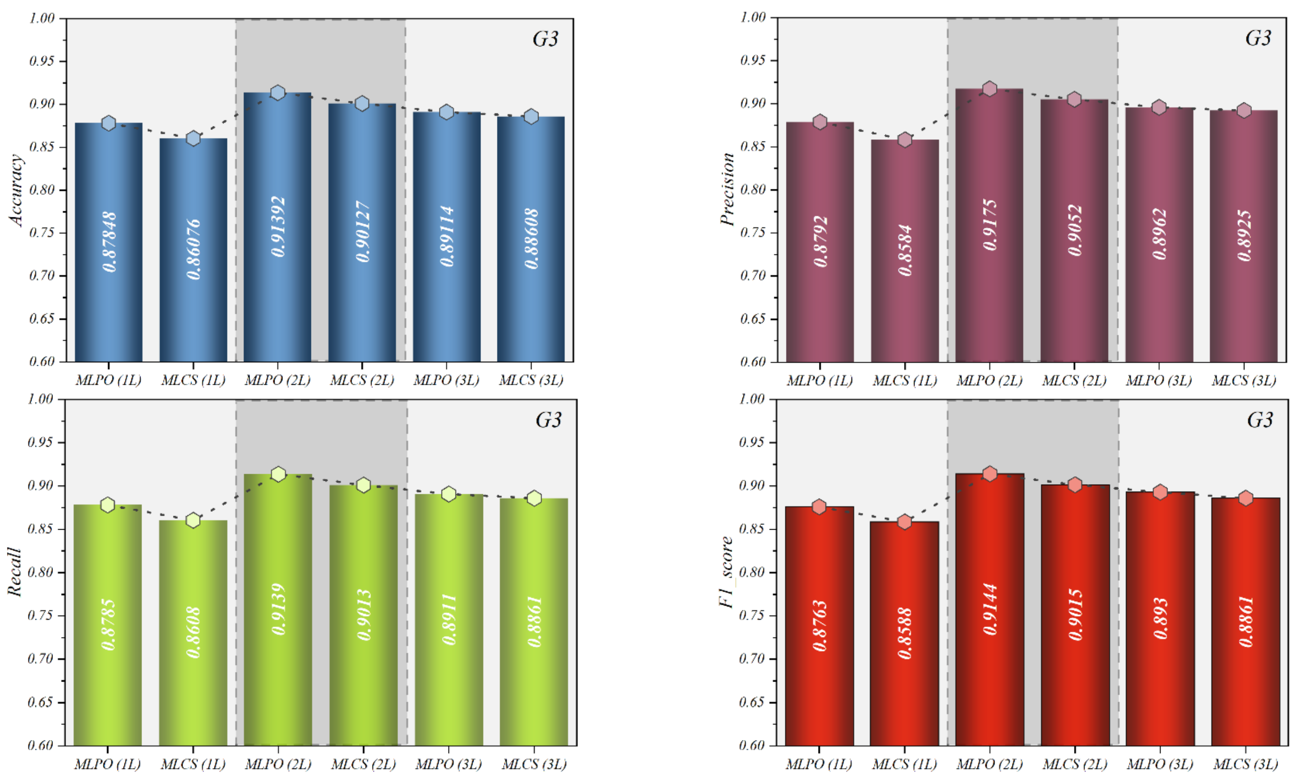


Fig. 8. Column plots the achievement percentage for developed models of G3 prediction values based on evaluators.

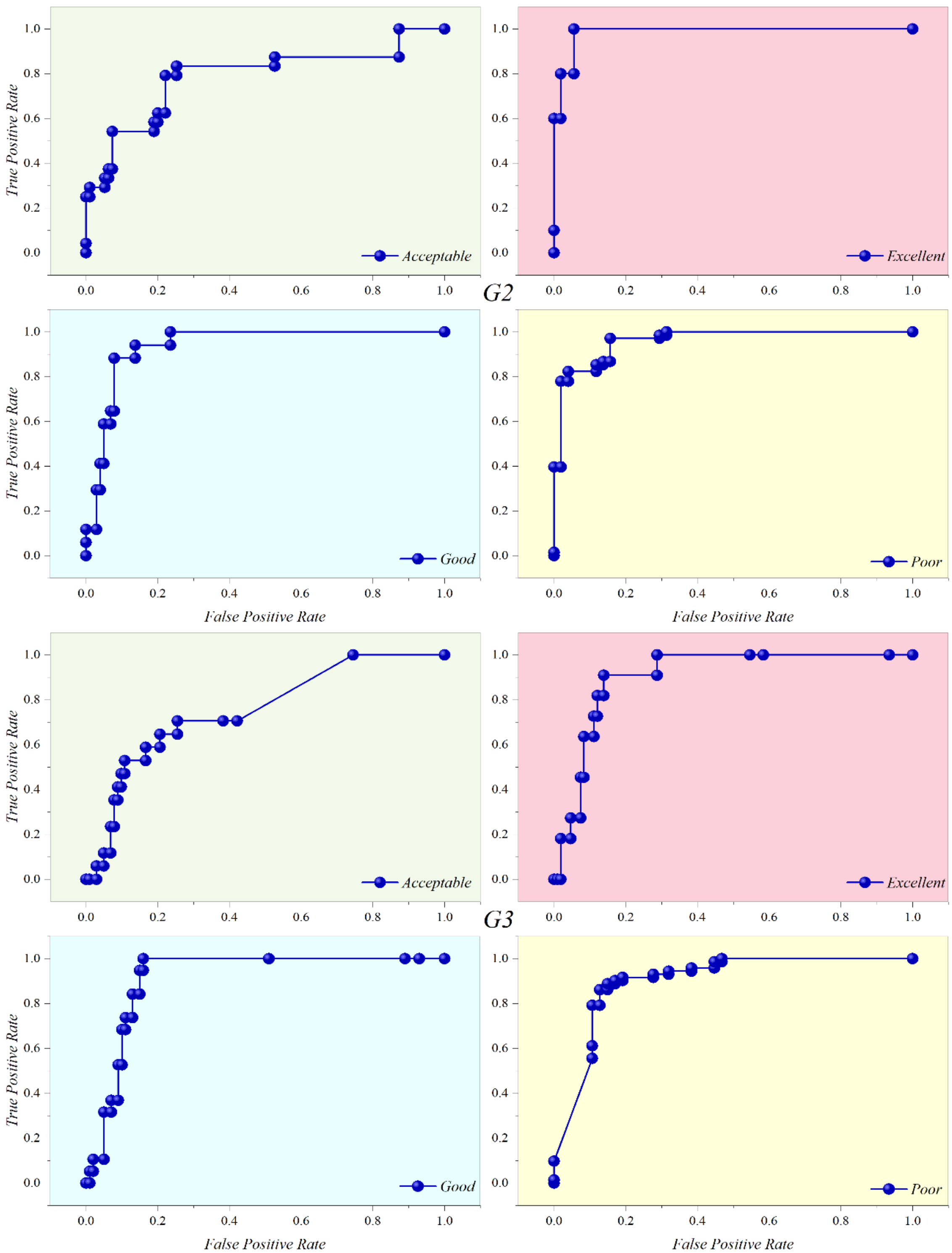


Fig. 9. The result of the ROC curve.

A binary classification model's performance at different classification thresholds is represented graphically by the Receiver Operation Characteristic (ROC) curve presented in Fig. 9. It shows how different threshold values affect the trade-off between the genuine positive rate (sensitivity) and the false positive rate (1-specificity). The following are important ideas about ROC curves: True Positive Rate (Sensitivity): The percentage of real positive cases the model accurately predicts is the true positive rate. To compute it, divide the number of true positives (TP) by the quantity of false negatives (FN), or TP / (TP + FN). False Positive Rate (1-Specificity): This refers to the percentage of real negative cases the model mispredicts as positive. The formula for calculating it is FP / (FP + TN), where FP stands for false positives and TN for true negatives.

Threshold: Predictions in binary classification models are frequently predicated on a probability threshold. Positive observations are those whose estimated probability falls above the threshold; negative observations fall below it.

By changing this threshold and seeing how the true positive rate and false positive rate change in tandem, ROC curves are produced. Area Under the ROC AUC-ROC curve: This concisely indicates the classifier's overall effectiveness and potential classification levels. It offers a solitary scalar value that symbolizes the model's overall performance. On the other hand, AUC-ROC values of 1.0 and 0.5 suggest models with performance comparable to random chance and are regarded as ideal, respectively. On the other hand, the following convergence curve shows the optimal model (MLPO 2), whose grade has the highest accuracy and a rank of false positive rate that approaches 1.0. The current plot demonstrates that the performance of the best model within an acceptable range is deemed unsatisfactory, as evidenced by its attainment of a true positive rate of 1.0 after a false positive rate of 0.8. The best model exhibits improvement in both poor and good performance categories, yet it remains insufficient. An examination of the excellent performance category reveals that the vector achieves a true positive rate of 1.0 before a false positive rate of 0.2.

Consequently, the optimal performance of MLPO 2 is realized in the excellent grade of the G2 target. On the contrary target, superior performance is observed in the good grade by the best model, with a true positive rate of 1.0 occurring before a false positive rate of 0.2. Following the good grade, the subsequent rank is assigned to the excellent grade, while the poor and acceptable grades occupy the third and fourth ranks, respectively.

IV. Discussion

A. Comparing Previous Studies vs. Present Study

The findings from three articles investigating student performance in the literature—specifically, one by Bichkar and R. R. Kabra [31], another by Edin Osmanbegovic et al. [32], and a third by Nguyen and Peter [33] —are succinctly summarized in Table VII. Notably, the research conducted by Nguyen and Peter, employing the DTC model, demonstrated the highest accuracy rate of 82%. In contrast, within this particular present study, which endeavors to forecast and classify students' performance in Portuguese language based on their G2 and G3 scores, the combination of the MLPC model and POA optimization algorithm yielded remarkable results. The accuracy metrics recorded were 95.3% for G2 and 93.5% for G3. Consequently, the proposed methodology achieved notably more reliable outcomes compared to prior studies, underscoring its efficacy in enhancing predictive accuracy and classification precision.

TABLE VII. Result of Presented and Published Studies

Author (s)	Model	Accuracy
Bichkar and R. R. Kabra [31]	DTC	69.94%
Edin Osmanbegovic et al. [32]	NBC	76.65%
Nguyen and Peter [33]	DTC	82%
Present study for G2	MLPC+POA (2)	95.3%
Present study for G3	MLPC+POA (2)	93.5%

V. Conclusion

To sum up, utilizing the MLPC model in conjunction with the Pelican Optimization and Crystal Structure Algorithm optimizers presents a viable method for predicting achievement among learners. Using these sophisticated methodologies enhances the precision and efficacy of evaluating learning objectives. This study highlights the possibilities for subsequent breakthroughs in educational analytics while also improving the accuracy of achievement prediction using complex machine learning algorithms and optimization techniques. The combined use of Crystal Structure Algorithm Optimizers, Pelican Optimization, and MLPC demonstrates a strong foundation for forecasting and comprehending students' academic performance. This opens the door for more knowledgeable and focused interventions in educational settings.

Nevertheless, within the scope of this study, an evaluation of the performance of MLPO and MLCS models is conducted across three distinct layers of the G2 and G3 targets. The findings reveal that the MLPO model in the second layer of both targets demonstrates superior accuracy, precision, recall, and F1-score, registering percentages of 0.9324 and 0.91392 for accuracy, 0.9395 and 0.9139 for precision, 0.9393 and 0.9139 for recall, and 0.9385 and 0.9144 for F1-score, respectively. Conversely, MLPO L2 achieves the highest accuracy in the G3 target, specifically in the acceptable grade, with a measured value of 54 out of 62, in contrast to MLCS L1, which records the lowest accuracy in the same grade and target, with a measured value of 46 out of 62. This comparison suggests that the MLPO model L2, given its elevated accuracy, can predict student performance with a high degree of precision. There is substantial potential for the educational system to utilize this model for advancements in this domain. Notably, the outcomes of this predictive process can be applied in real-world scenarios, yielding consistent results.

References

[1] K.-L. Tsui, V. Chen, W. Jiang, F. Yang, and C. Kan, “Data mining methods and applications,” in Springer handbook of engineering statistics, Springer, 2023, pp. 797–816.

[2] Q. H. Ngô and N. M. Trịnh, “A University Student Dropout Detector based on Academic Data-A case study at FPT University.” FPTU Hà Nội, 2023.

[3] S. Chatterjee, T. P. Singh, S. Lim, and A. Mukhopadhyay, “Social Media and Crowdsourcing”.

[4] P. T. T. Thao et al., “55 Khoa Học Giáo Dục Việt Nam,” 2023.

[5] S. Dutta and S. Mandi, "Can One Deep Model Be Effective in Multiple Domain? a Case Study with Public Datasets," EasyChair, 2023.

[6] A. Rawal and B. Lal, "Predictive model for admission uncertainty in high education using Naïve Bayes classifier," Journal of Indian Business Research, vol. 15, no. 2, pp. 262–277, 2023.

[7] S. Sengupta, "Search for Articles," 2023.

[8] S. Jordão, D. Durães, and P. Novais, "Performance Analysis of Models Used to Predict Failure in Secondary School," in International Conference on Data Science and Artificial Intelligence, Springer, 2023, pp. 339–348.

[9] S. Khan and M. Shaheen, "From data mining to wisdom mining," J Inf Sci, vol. 49, no. 4, pp. 952–975, 2023.

[10] S. Chanmee and K. Kesorn, "Semantic decision Trees: A new learning system for the ID3-Based algorithm using a knowledge base," Advanced Engineering Informatics, vol. 58, p. 102156, 2023.

[11] S. Bum, I. B. Iorliam, E. O. Okube, and A. Iorliam, "Prediction of Student's Academic Performance Using Linear Regression," NIGERIAN ANNALS OF PURE AND APPLIED SCIENCES, vol. 2, pp. 259–264, 2019.

[12] A. M. Shahiri and W. Husain, "A review on predicting student's performance using data mining techniques," Procedia Comput Sci, vol. 72, pp. 414–422, 2015.

[13] 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.

[14] 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 w," Jurnal Mantik, vol. 4, no. 1, pp. 441–451, 2020.

[15] H. Zhou, Z. Wu, N. Xu, and H. Xiao, "PDR-SMOTE: an imbalanced data processing method based on data region partition and K nearest neighbors," International Journal of Machine Learning and Cybernetics, pp. 1–16, 2023.

[16] D. Kabakchieva, K. Stefanova, and V. Kisimov, "Analyzing university data for determining student profiles and predicting performance," in Educational Data Mining 2011, 2010.

[17] Y. Fan, Y. Liu, H. Chen, and J. Ma, "Data Mining-based Design and Implementation of College Physical Education Performance Management and Analysis System.," Int. J. Emerg. Technol. Learn., vol. 14, no. 6, pp. 87–97, 2019.

[18] J.-M. Trujillo-Torres, H. Hossein-Mohand, M. Gómez-García, H. Hossein-Mohand, and F.-J. Hinojo-Lucena, "Estimating the academic performance of secondary education mathematics students: A gain lift predictive model," Mathematics, vol. 8, no. 12, p. 2101, 2020.

[19] 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.

[20] S. Kotsiantis, "Educational data mining: a case study for predicting dropout-prone students," Int J Knowl Eng Soft Data Paradig, vol. 1, no. 2, pp. 101–111, 2009.

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

[22] A. O. Ogunde and D. A. Ajibade, "A data mining system for predicting university students' graduation grades using ID3 decision tree algorithm," Journal of Computer Science and Information Technology, vol. 2, no. 1, pp. 21–46, 2014.

[23] S. Alturki and N. Alturki, "Using educational data mining to predict students' academic performance for applying early interventions," Journal of Information Technology Education: JITE. Innovations in Practice: IIP, vol. 20, pp. 121–137, 2021.

[24] 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.

[25] B. Olukoya, "Using ensemble random forest, boosting and base classifiers to ameliorate prediction of students' academic performance," vol. 6, p. 654, Mar. 2023.

[26] X. Zhang, R. Xue, B. Liu, W. Lu, and Y. Zhang, "Grade prediction of student academic performance with multiple classification models," in 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), IEEE, 2018, pp. 1086–1090.

[27] S. Talatahari, M. Azizi, M. Tolouei, B. Talatahari, and P. Sareh, "Crystal structure algorithm (CryStAl): a metaheuristic optimization method," IEEE Access, vol. 9, pp. 71244–71261, 2021.

[28] S. A. Farooqui et al., "Crystal Structure Algorithm (CryStAl) Based Selective Harmonic Elimination Modulation in a Cascaded H-Bridge Multilevel Inverter," Electronics (Basel), vol. 10, no. 24, p. 3070, 2021.

[29] J. C. Thomas, A. R. Natarajan, and A. Van der Ven, "Comparing crystal structures with symmetry and geometry," NPJ Comput Mater, vol. 7, no. 1, p. 164, 2021.

[30] T. Sağ, "PVS: a new population-based vortex search algorithm with boosted exploration capability using polynomial mutation," Neural Comput Appl, vol. 34, no. 20, pp. 18211–18287, 2022.

[31] 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.

[32] 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.

[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.