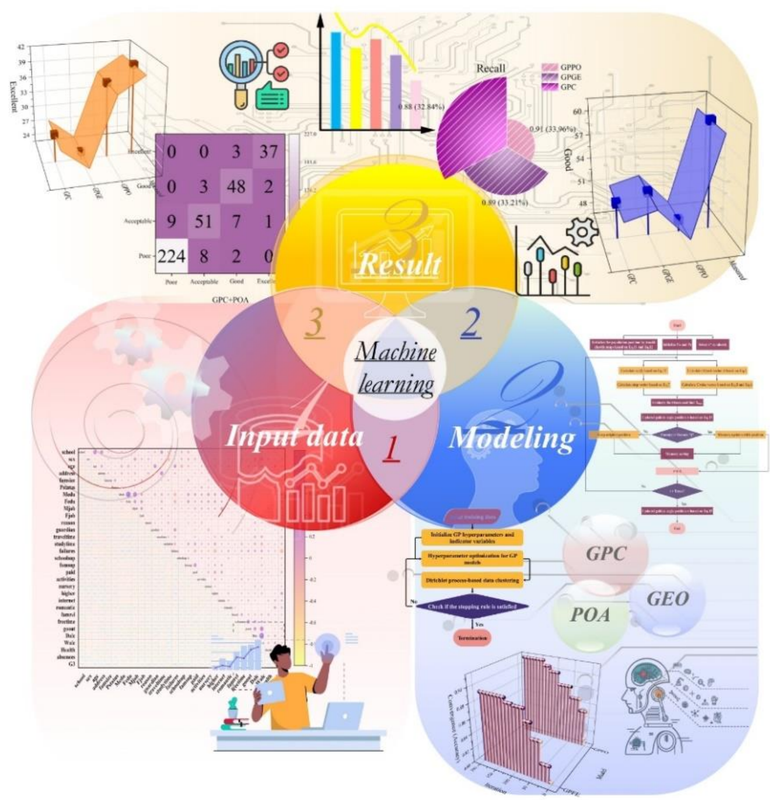
Student Performance Estimation Through Innovative Classification Techniques in Education

Hui Fan[1], Guoping Zhu[2]*, Jianhua Zhan[3]

School of Marxism, Guangzhou Xinhua University, Guangzhou 510520, Guangdong, China[1]
School of Marxism, Wannan Medical College, Wuhu 241002, Anhui, China[2]
School of Marxism, South China University of Technology, Guangzhou 510641, Guangdong, China[1, 2]
School of Japanese Studies, Shanghai International Studies University, Shanghai 200083, China[3]

***Abstract*—In the current era of intense educational competition, institutions must effectively classify individuals based on their abilities, proactively forecast student performance, and work towards enhancing their forthcoming examination outcomes. Providing early guidance to students is crucial in helping them focus their efforts on specific areas to boost their academic achievements. This analytical approach supports educational institutions in mitigating failure rates by utilizing students' previous performance in relevant courses to predict their outcomes in a specific program. Data mining encompasses a variety of techniques used to reveal hidden patterns within vast datasets. In the context of educational data mining, these methods are applied within the educational sphere, with a specific emphasis on analyzing data from both students and educators. These patterns can offer significant value for predictive and analytical objectives. In this study, Gaussian Process Classification (GPC) was employed for the prediction of student performance. To improve the model's accuracy, two cutting-edge optimizers, namely the Golden Eagle Optimizer (GEO) and the Pelican Optimization Algorithm (POA), were incorporated. When assessing the model's performance, four widely used metrics were utilized: Accuracy, Precision, Recall, and F1-score. The results of this study underscore the effectiveness of both the POA and GEO optimizers in enhancing GPC performance. Specifically, GPC+GEO demonstrated remarkable effectiveness in the Poor grade, while GPC+POA excelled in the Acceptable and Excellent category. This highlights the positive impact of these optimization techniques on the model's predictive capabilities.**

Keywords*—Student performance; Gaussian Process Classification; Golden Eagle Optimizer; Pelican Optimization Algorithm



Graphical Abstract

I. INTRODUCTION

One of the fundamental difficulties with every nation's instructive organization lies in the exact evaluation of students' academic achievements [1]. This precise assessment is instrumental for educational administrators in pinpointing and addressing issues within the educational system. Academic performance encompasses the array of actions undertaken by students throughout their academic journey [2]. The critical post-implementation phase of educational programs is the assessment of students. Assessment is the procedure via which the accomplishment of instructive goals for both the teacher & the student is ascertained. Student assessment takes place through various methods [3]. The evaluation approaches can include techniques such as examination, conduct scrutiny, evaluation of schemes, assessment of documents and summaries, and the use of theoretical development tests for measurement [4].

Predicting students' performance early on is beneficial for enhancing learning results [5]. The capacity to anticipate a pupil's theoretical achievements clasps significance by way of driving modifications in college theoretical policies notifies teaching techniques, assesses the proficiency and efficiency of education, offers respected input to educators and learners, and modifies knowledge environments [6]. At the commencement of the educational journey, accurately identifying underperforming students is valuable. Educational institutions utilize data mining methods to analyze available data, a practice commonly referred to as Educational Data Mining (EDM) [7]. While data mining supports knowledge discovery, it is important to note that MLA delivers the indispensable tackles for this procedure. Correct prediction of pupil presentation is valuable as it enables the early detection of underperforming students [8], [9]. Educational Data Mining (EDM) aids educational institutions in enhancing and innovating learning approaches through the analysis of pertinent educational data [10]. In practice, forecasting a pupil's theoretical success is crucial for every one of their educational progress, yet it can be challenging due to the influence of numerous factors on student performance [11]. The constant evolution of technology has opened up novel avenues for the expansion and enhancement of educational systems. Recent research indicates that the machine learning (ML)-based methods employed in this study have proven to be highly efficient [12].

In the realm of educational institutions, a multitude of researchers have utilized statistical approaches and ML algorithms to forecast student performance [13]. Ogunde et al. [14] initiated the development of a system that utilizes the Iterative Dichotomiser (ID3) decision tree methodology and input data to anticipate grades. The authors suggest that their approach shows substantial potential for precise forecasting of students' ultimate graduation results. Bharadwaj et al. [15] utilized data sourced from a prior student database, integrating elements such as student attendance, class participation, participation in seminars, and assignment scores to make projections regarding semester-end outcomes. Their results revealed that decision tree analysis produced the highest accuracy, followed by K-nearest neighbor (KNN) classification [16]; on the other hand, Bayesian classification systems demonstrated the least accuracy. Duzhin and Gustafsson [17] introduced an ML technique to account for students' prior knowledge. Their approach relies on symbolic regression and incorporates historical university scores as non-experimental input data. This classification method has the potential to aid the Ministry of Education in enhancing student performance through early predictions. Naïve Bayes [18] displays traits of conditional independence, which makes it proficient at estimating class conditional probabilities. Watkins et al. [1] introduced a technique called SENSE (Student Performance Quantifier using Sentiment Analysis) to enrich the content of secondary school reports by leveraging natural language processing. Sentiment analysis [19] can play a significant role in impacting student performance.

Several studies have demonstrated the practical application of data mining in education. In order to forecast student performance and arrange the students appropriately, Sunita and LOBO L.M.R.J. [20] used classification and clustering methods. Thammasiri et al. [21] created a model to predict poor academic performance in first-year students, and by integrating support vector machines with SMOTE, they were able to achieve an astounding accuracy of 90.24%. Using classification algorithms, Bichkar and R. R. Kabra [22] concentrated on identifying first-year engineering students who were at risk. Surjeet and Pal [23] employed decision tree algorithms to forecast the performance of first-year engineering students, with a specific emphasis on identifying those at risk of failure. The C4.5 decision tree outperformed other classifiers and offered insights into factors influencing student performance, according to Mustafa et al. [24], who evaluated student data in C++ courses using the CRISP framework. Using academic markers as a basis, Bharadwaj and Pal [25] predicted student divisions using the ID3 decision tree method. Nguyen and Peter [26] compared decision trees and Bayesian networks in predicting undergraduate and postgraduate academic performance, with decision trees demonstrating superior performance.

Table I offers an overview of many relevant investigations.

TABLE I. LITERATURE REVIEW

No.	Author (s)	Models	Accuracy	Reference
1	Carlos et al.	$ADTree$	97.3%	[27]
4	Edin Osmanbegovic et al.	NBC	76.65%	[28]
3	Al-Radaideh et al.	DTC	87.9%	[29]
2	Nguyen and Peter	DTC	82%	[26]
5	Bichkar and R. R. Kabra	DTC	69.94%	[22]

The literature review revealed the high effectiveness of ML methods for assessing and appraising students' performance [30]. Although various studies [31], [32], [33], [34], [35] have employed diverse models, each with distinct conditions and characteristics tailored to the specific problem context, also, there are gaps in the literature in the field of utilizing GPC model in integration with several optimization algorithms. Therefore, the main purpose of this study is to succeed in a framework to forecast students' academic performance by amalgamating ML models with meta-heuristic algorithms, considering the unique educational circumstances throughout their academic journey. In this research, substantial variations of Gaussian Process Classification (GPC) algorithms have been included to assist educators and parents in predicting the performance of new students and improving next year's outcomes. Additionally, to ensure the utmost reliability in the results, both POA and GEO techniques were integrated, leading to the attainment of promising outcomes. The proposed framework not only enhances prediction accuracy but also provides practical applications for educators and parents, empowering them with valuable insights into students' performance and potential areas for improvement. Anticipating promising outcomes, this research offers a systematic approach to leveraging advanced techniques for educational prediction, thereby facilitating more effective decision-making and ultimately improving student outcomes.

The structure of the remaining sections in this article is as follows:

- Section II outlines the research methodology, encompassing an explanation of evaluation metrics, ML-based classifiers, meta-heuristic algorithms, and an overview of the dataset used in the study.
- In Section III, the outcomes of the case study were investigated and analyzed using actual data. This section is subdivided into three parts: results pertaining to the initial dataset, results related to the balanced (edited) data, and findings associated with the application of hybrid models on the balanced dataset.
- Finally, Section IV presents the concluding remarks.

II. Datasets and Methodology

A. Data Gathering

In this research, a dataset pertaining to the Portuguese educational system was employed. This dataset comprises 33 distinct attributes thoughtfully selected to provide a precise representation of students' academic advancement, considering their unique characteristics and situations [36]. The dataset was generated by merging data obtained through two survey techniques with the academic records of the students. These attributes encompass a broad spectrum of student-related factors, including demographics such as gender, age, school attended, and residential type (address). Additionally, these attributes encompass parental characteristics such as parents' cohabitation status (Pstatus), educational background, and occupation (Medu, Mjob, Fedu, Fjob). Other factors considered include the student's parent, family attributes like the size of the family (famsize), the caliber of familial connections (famrel), & various attributes for example, the cause of selecting an educational institution (rationale), commuting duration to an educational institution (journey duration), weekly schoolwork hours (studytime), past academic setbacks (disappointments), contribution in supplementary academic programs (schoolsup), (famsup), engagement in extracurricular actions (activities), attendance in paid classes (paidclass), internet accessibility (internet), attendance at nursery school (school), aspirations for higher education (higher), romantic relationship status (romantic), availability of leisure time after school (freetime), socializing preferences (goout), weekday alcohol consumption (Dalc), weekend alcohol consumption (Walc), and the present physical condition of the individual (well-being).

Furthermore, besides these characteristics, there are 3 extra attributes, namely G1, G2, & final, which signify the grades of students across 3 assessment stages throughout their learning journey, ranging from 0 (the minimum grade) to twenty (the maximum grade). G3 signifies the students' ultimate score. These 3 attributes, al chose as pattern results along with the count of college nonappearances (nonappearances), were selected as model outputs, serving as reliant on parameters. For grading purposes, they were categorized into 4 groups: zero–twelve: Deprived, twelve –fourteen: Acceptable, fourteen – sixteen: Respectable, and sixteen –twenty: Outstanding. In Fig. 1, as anticipated, cells along the central axis are displayed in red, indicating a correlation value of 1. From the visual representation above, it is evident that the three attributes, G2, G1 & last, all of which are considered reliant on parameters and represent students' grades, exhibit the highest correlation values among themselves.

B. Gaussian Process Classification (GPC)

Gaussian process priors offer expressive nonparametric function models. To conduct classification using this prior, the process is compressed through a sigmoidal inverse-link function, and a Bernoulli likelihood is applied to the data based on the transformed function values [37]. The binary class observations are denoted as y = $\{y_1, y_2, \dots, y_N\}$, and the input data is organized into a design matrix X = $\{x_1, x_2, \dots, x_N\}$. The covariance function is computed for all pairs of input vectors to create the covariance matrix K_{nn} following the standard process. This results in a prior distribution for the values of the Gaussian Process function at the input points: $p(f) = N(f \mid 0, K_{nn})$.

The $probit$ inverse link function is represented as $\emptyset(x) = \int_{-\infty}^{x} N(f|0, K_{nn})$, and the Bernoulli distribution $, B(y_n \mid \emptyset(f_n)) = \emptyset(f_n)^{y_n}.(1 - \emptyset(f_n))^{\{1-y_n\}}$. This leads to the joint distribution of data and latent variables [38].

$$p(y, f) = \prod_{n=1}^{N} B(y_n \mid \emptyset(f_n)) N(f|0, K_{nn}) \tag{1}$$

The primary emphasis lies in the approximation of the posterior distribution of function values, labeled as $p(f \mid y)$. Furthermore, there is a requirement for an approximation of the marginal likelihood, p(y), to enable the optimization or marginalization of covariance function parameters. Various methods for approximation have been suggested, but all of them necessitate computation on the order of $O(N^3)$. Fig. 2 illustrates the structure of the GPC model.

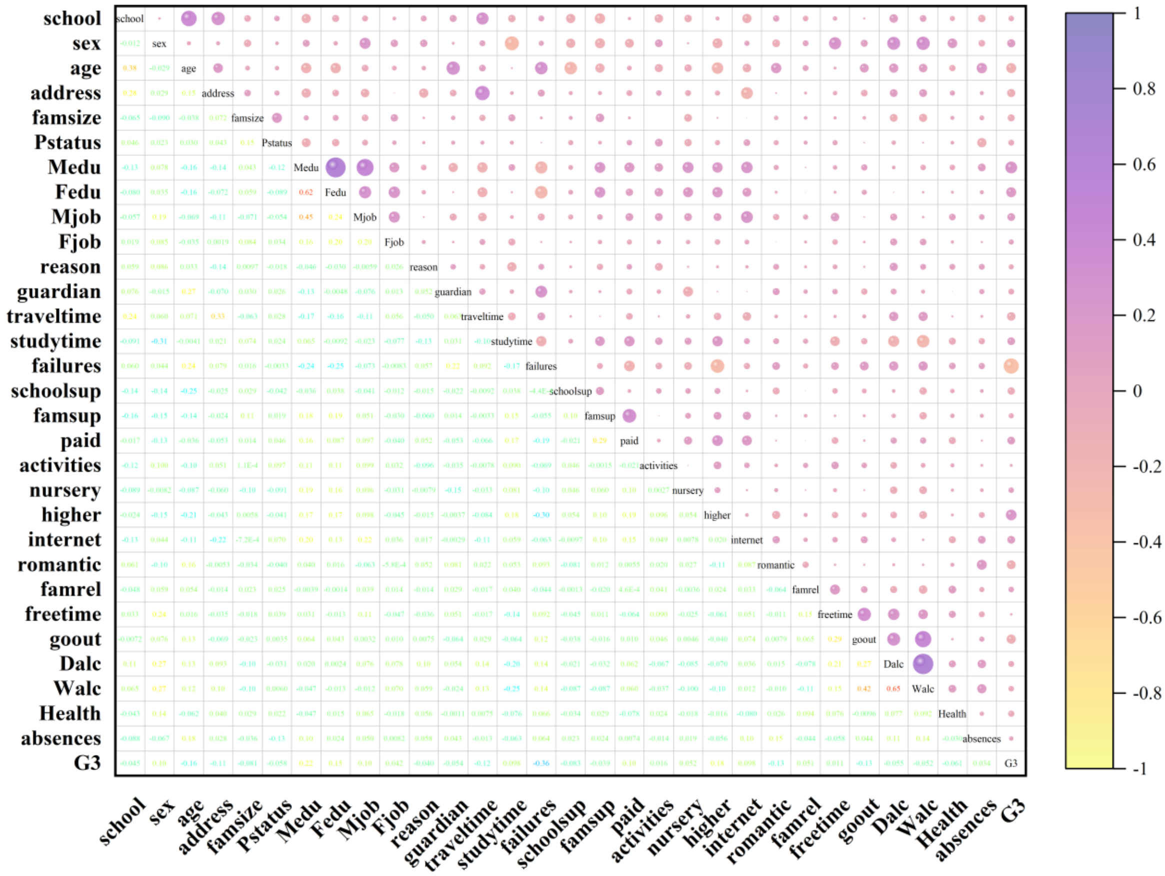


Fig. 1. The output and input variables' correlation matrix.

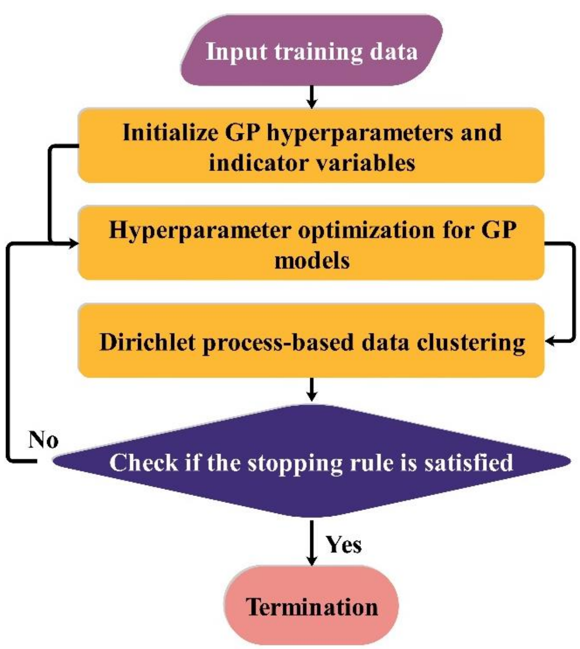


Fig. 2. The flowchart of GPC.

C. Pelican Optimization Algorithm (POA)

In the year 2022, the POA, a novel nature-inspired approach, was introduced by Dehghani and Trojovský. This algorithm is influenced by the social behavior and hunting strategies of pelicans [39]. Pelicans, characterized by their large size and elongated beaks, have a sizeable throat pouch that they use for capturing and consuming prey. They typically live in significant colonies, making up the population of concern. The individuals within this population are randomly initialized using the following equation:

$$x_{i,j} = l_j + rand.\left(u_j - l_j\right), \quad i = 1,2,3, \dots, N, j = 1,2,3, \dots, m \tag{2}$$

Within this equation, $x_{i,j}$ denotes the value of the $j-th$ variable as indicated by the ith candidate solution. The parameters N and m correspond to the count of individuals in the population and the total number of problem variables, respectively. Furthermore, l_j and u_j signify the lower and upper boundaries of the problem variables. The term "rand" represents a random number within the [0,1] range. The population matrix, representing the individuals within the candidate solutions, is formed using Eq. (3):

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}_{N\times m} = \begin{bmatrix} X_{1,1} & \dots & X_{1,j} & \dots & X_{1,m} \\ \vdots & & \vdots & & \vdots \\ X_{i,1} & \dots & X_{i,j} & \dots & X_{i,m} \\ \vdots & & \vdots & & \vdots \\ X_{N,1} & \dots & X_{N,j} & \dots & X_{N,m} \end{bmatrix}_{N\times m} \tag{3}$$

The objective function is computed based on the expression given in Eq. (4).

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N\times m} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N\times 1} \quad (4)$$

The objective function vector, labeled as F, comprises individual objective function values denoted as F_i for each candidate's solution. The hunting process of pelicans is bifurcated into two phases: exploration and exploitation. In the exploration phase, pelicans approach their prey, whereas, during the exploitation phase, they gracefully glide along the water's surface. During the initial stage of the exploration phase, pelicans close in on the prey by identifying its position, which is randomly generated. The stochastic nature of the prey's location enhances the exploration capacity of the POA [40]. The mathematical expression of the initial phase is depicted in Eq. (5):

$$x_{i,j}^{p_1} = \begin{cases} x_{i,j} + rand.\left(p_j - I.x_{i,j}\right), & F_p < F_i; \\ x_{i,j} + rand.\left(x_{i,j} - p_j\right), & else, \end{cases} \quad (5)$$

Let $x_{i,j}^{p_1}$ signify the revised state of the ith pelican in the jth dimension following the first phase. This update is contingent on a random variable I, which can assume values of 1 or 2, p_j indicating the prey's location in the $j-th$ dimension and F_p denoting the prey's objective function value. In the POA algorithm, a pelican's new position is deemed acceptable if it enhances the objective function value at that particular position. This procedure, termed effective updating, safeguards the algorithm against converging to suboptimal regions. Mathematically, this concept can be expressed as follows:

$$x_i = \begin{cases} X_i^{p_1}, F_i^{p_1} < F_i \\ X_i \quad\quad else \end{cases} \quad (6)$$

The mathematical depiction of the hunting process is as follows: $X_i^{p_1}$ signifies the updated state of the $i-th$ pelican after the second phase, and $F_i^{p_1}$ represents the objective function value of the pelican derived from this phase. During the second phase, pelicans enhance their prospects of capturing more fish by lifting them upward through wing expansion while on the water's surface [41]. Following this, they ensnare the prey within their throat pouches. Consequently, this phase substantially enhances the effectiveness of the POA algorithm, facilitating the convergence of enhanced solutions within the hunting region.

$$x_{i,j}^{p_1} = x_{i,j} + R.\left(1 - \frac{t}{T}\right).(2.rand - 1).x_{i,j} \quad (7)$$

During the second phase, the updated state of the ith pelican in the $j-th$ dimension indicated as $X_{i,j}^{p_2}$, is determined considering several factors. One of these factors is the constant R, which is set to 0.2. The neighborhood radius of $x_{i,j}$ is influenced by the term $\left(1 - \frac{t}{T}\right)$, where t represents the iteration count, and T is the maximum number of iterations. Moreover, an effective updating procedure is implemented in this phase, where the new pelican position, as per Eq. (8), may be either accepted or rejected.

$$x_i = \begin{cases} X_i^{p_2}, F_i^{p_2} < F_i \\ X_i \quad\quad else \end{cases} \quad (8)$$

$X_i^{p_2}$ signifies the revised condition of the $i-th$ pelican and $F_i^{p_2}$ indicates the respective objective function value for that pelican. Once all individuals in the population have been updated, the subsequent iteration commences, and the series of steps outlined by Eq. (5) to (8) are reiterated until the entire execution process is completed [42]. The POA flowchart, which is displayed in Fig. 3, illustrates the iterative process.

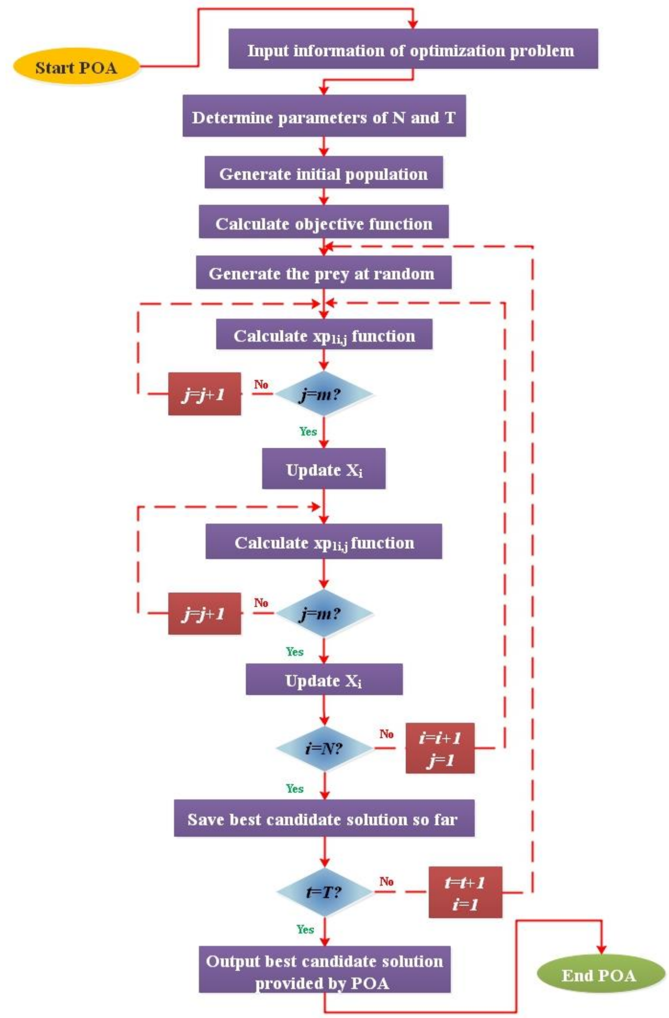


Fig. 3. The flowchart of POA.

D. Golden Eagle Optimizer (GEO)

GEO is inspired by the spiral flight pattern of golden eagles. Each golden eagle remembers its most rewarding locations visited thus far. It combines both gliding in search of food and hunting prey simultaneously. The ROD image is subjected to segmentation, which involves breaking it down into distinct regions using a geometrically active multilevel contour. This segmentation process enables precise scrutiny and disease diagnosis. Initially, the chosen experimental image is processed with multiple threshold levels (Th) using the geometrically active multi-contour method. Data preprocessing is carried out using the GEO and Shannon entropy method (GEO + SE). GEO + SE enhances ROD by combining similar pixel values determined by Th allocation. Entropic techniques are commonly utilized for the evaluation of medical images. A hypothetical RGB image is considered with dimensions M*N. In this case, the pixel at (x, y) is defined as:

$$F(x,y) while\ x \in \{1,2,3,\dots,M\}\ and\ y \in \{1,2,3,\dots,N\}$$

Given that T represents the gray level of the experimental image, with the entire range of gray values spanning from 0 to T-1, denoted as R, as follows:

$$F(x,y) \in R^{\forall}(x,y) \in picture \tag{9}$$

Here is the description of the standardized histogram (bar chart) for the image:

$$J = \{j0.j1,\dots jR1\} \tag{10}$$

The previously mentioned equation can be formulated as follows using the geometrically active multi-contours method:

$$J(Th) = j0(th_1) + j1(th_2),\dots,jR-1(th_{k-1}) \tag{11}$$

$$Th *= \max\{J(Th)\} \tag{12}$$

$Th *$ represents the selected threshold. Eq. (11) employs Shannon entropy. The GEO typically demands fewer initial parameters for allocation compared to other established methods. The required data is usually extracted from the preprocessed image using the segmentation approach. In this paper, this task is achieved using the widely recognized DRLS technique. A dynamic bounding box is integrated into DRLS, adapting its dimensions based on the region to be extracted. The adjustments in the dimensions of this box align with the boundaries of the ROD, depending on the extent of the repetitive process. Once the predefined repetition level is complete, the adjustments cease, and the extracted ROD is presented. There is no doubt that this approach outperforms the methods employed in the articles, namely the turning point and Chan-Vese methods. Initially, normalization is conducted in this phase, and subsequent results are extracted. A subset of the data is utilized as training data for the vector machine model, which is then constructed using this dataset. Weight tests for this algorithm are calculated to assess its performance in this context further. Each image contains a multitude of reference data points gathered from diverse sensors. In a comparative analysis, segmented discs were scrutinized alongside expert observational data images. The initial phase entails the computation of image similarity metrics such as GEOccard, Dice, FPR, and FNR, following the methods detailed in the articles. The mathematical formula is displayed below. Furthermore, Fig. 4 shows the flowchart of the GEO.

$$Jaccard\ (I_g, I_m) = I_g \cap I_m / I_g \cup I_m \tag{13}$$

$$Dice\ (I_g, I_m) = 2(I_g \cap I_m)/|I_g| \cup |I_m| \tag{14}$$

$$FPR\ (I_g, I_m) = \ (I_g/I_m)/\ (I_g \cup I_m) \tag{15}$$

$$FNR\ (I_g, I_m) = \ (I_m/I_g)/\ (I_g \cup I_m) \tag{16}$$

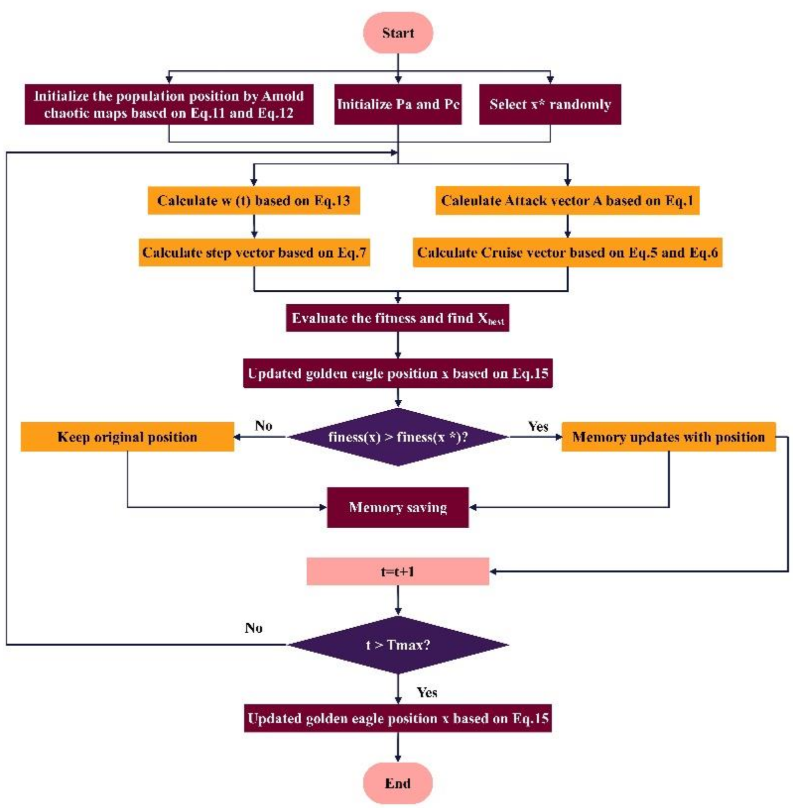


Fig. 4. The flowchart of GEO.

E. Performance Evaluators

In the evaluation of classifier performance, various assessment criteria are at one's disposal. Accuracy, a commonly used metric, evaluates the classifier's effectiveness by measuring the percentage of samples correctly predicted. In addition to Precision, Accuracy, and Recall are widely employed metrics. Recall measures the proportion of correctly predicted positive instances among all actual positive instances, while Precision assesses the likelihood that positive predictions are correct. The combination of Precision and Recall produces a composite measure known as the f1-score.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

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

$$F1\ score\ = \frac{2 \times Recall\ \times\ Precision}{Recall + Precision} \tag{20}$$

In these formulas, TP denotes a positive prediction that accurately matches the true positive outcome. FP represents a positive prediction when the actual outcome is negative. TN indicates a negative prediction that correctly corresponds to the true negative outcome. FN is used to indicate a negative prediction when the actual outcome is positive.

III. Result and Discussion

A. Prediction and Classification Results

Fig. 5 offers a comprehensive illustration of the convergence curve for the proposed models, providing a visual representation of the algorithm's progression towards its predefined objective. This curve meticulously traces the accuracy performance metric across a sequence of iterations. The shape and trends within this curve offer valuable insights into the optimization process. A steep descent in the curve indicates rapid convergence, signifying swift progress towards the objective. Conversely, a flattened or erratic curve suggests potential challenges in attaining the optimal solution. These hurdles may encompass tasks like parameter refinement, managing computational intricacies, and enhancing the algorithm's efficiency. Convergence curves play a pivotal role in the field of algorithm assessment. They act as a guiding tool for researchers and professionals, helping them gauge the algorithm's performance and aiding in the intricate process of parameter fine-tuning. These curves also reveal the subtle balance between the requirement for speed and the quest for Precision in diverse computational tasks.

Focusing on the convergence curves of the two models, GPC+POA and GPC+GEO, as depicted in Fig. 5, a in the convergence curves becomes evident. Notably, the curve representing GPC+POA starts with a more favorable initial accuracy point compared to GPC+GEO. Moreover, it reaches its optimal outcome swiftly within a smaller number of iterations in contrast to GPC+GEO. This observation suggests that, as iterations progress, GPC+POA demonstrates greater efficiency for the specified task.

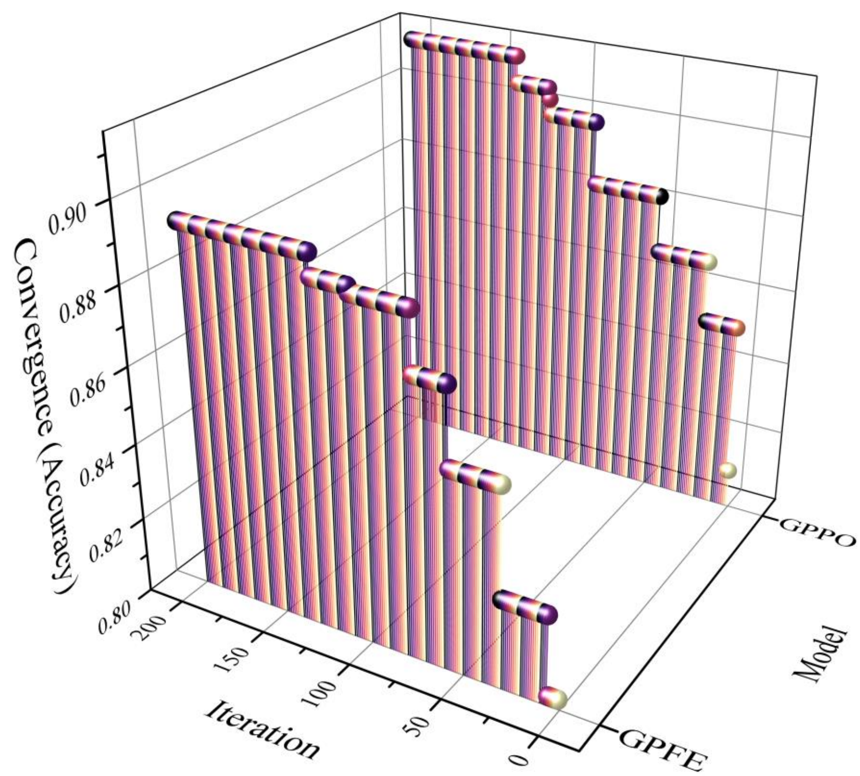


Fig. 5. Convergence curve of hybrid models.

TABLE II. RESULT OF PRESENTED MODELS

Model	Index values		
	GPC	GPC+GEO	GPC+POA
Accuracy	0.884	0.894	0.911
Precision	0.879	0.90	0.912
Recall	0.882	0.892	0.905
F1 _core	0.884	0.889	0.914

Table II provides a comprehensive overview of the models evaluated in this study, namely GPC+POA, GPC+GEO, and GPC. Their visual representation can be found in Fig. 6. The key performance metrics, including accuracy, Precision, recall, and F1-score, are examined for each model. Starting with GPC+POA, this model impresses with an exceptional accuracy of 0.91, indicating its ability to classify a substantial portion of the dataset accurately. Moreover, its Precision and recall both stand at 0.91, emphasizing its proficiency in correctly identifying positive instances. The F1-score of 0.91 highlights a remarkable balance between Precision and recall, further confirming GPC+POA's effectiveness. Moving to GPC+GEO, this model showcases strong overall performance with an accuracy of 0.8937. Its precision value of 0.9 suggests a low rate of false positives, and a recall of 0.89 indicates its capability to detect actual positive instances correctly. The F1-score of 0.89 signifies a well-balanced trade-off between Precision and recall in GPC+GEO.

Lastly, the base GPC model demonstrates respectable results with an accuracy of 0.8835. Its Precision and recall, both at 0.88, indicate a good balance between correctly identifying positive instances and minimizing false positives. The F1 score of 0.88 underscores its well-rounded performance in terms of Precision and recall. In the discussion, it becomes evident that GPC+POA leads the pack, excelling in scenarios where Precision and recall are of utmost importance, such as medical diagnoses or critical decision-making contexts. GPC+GEO closely follows, offering a balanced approach that suits applications requiring a trade-off between Precision and recall. The base GPC model, while still delivering a strong performance, is a reliable choice for more general applications where a well-rounded performance is required. Ultimately, the choice of the model should align with the specific needs and priorities of the task at hand, with GPC+POA, GPC+GEO, and GPC offering valuable options catering to different scenarios.

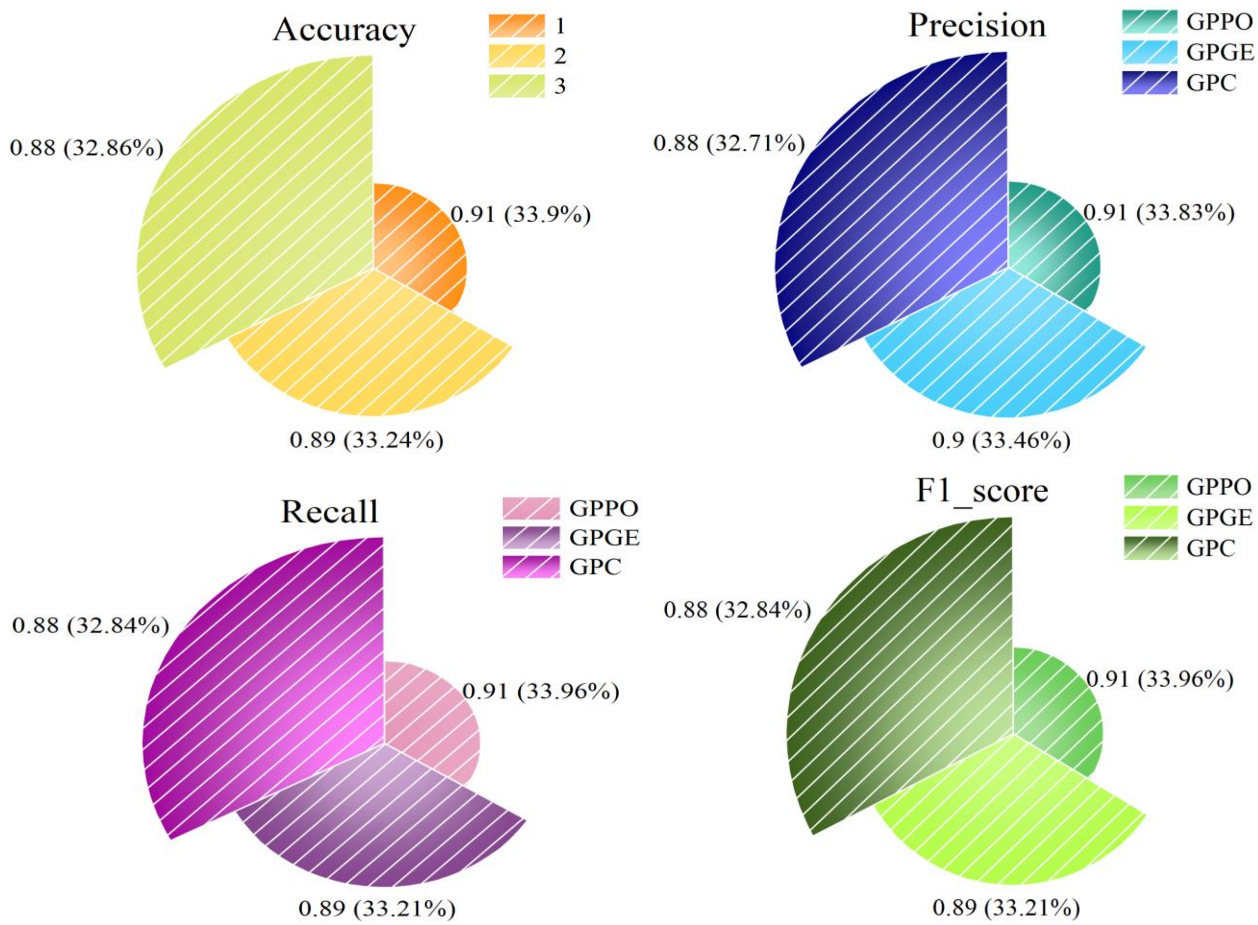


Fig. 6. Radial comparison of developed models based on metrics.

The created models' performance evaluation indicators on the basis of grades are displayed in Table III. These models, GPC+POA, GPC+GEO, and GPC, are assessed across various grade categories, including Excellent, Good, Acceptable, and Poor. The evaluation metrics considered are Precision, recall, and F1-score in each grade category. When examining the performance of GPC+POA, it is evident that this model excels in the Excellent grade category with a precision of 0.93, a recall of 0.93, and an F1-score of 0.93. In the Good category, GPC+POA maintains a high precision of 0.91 but experiences a slight decrease in recall to 0.8, resulting in an F1-score of 0.85. The Acceptable and Poor categories also display strong performance, with particularly impressive results in the Poor category, where the model achieves a precision, recall, and F1-score of 0.96.

Shifting focus to GPC+GEO, this model demonstrates outstanding Precision in the Excellent grade category, reaching a perfect score of 1. However, its recall in the Excellent category is 0.6, leading to an F1-score of 0.75. In the Good and Acceptable categories, GPC+GEO performs well, with balanced Precision and recall, resulting in F1-scores of 0.81. The Poor category maintains a high precision and recall, with an F1-score of 0.96. Finally, the base GPC model's performance is assessed. In the Excellent category, GPC achieves a precision of 0.88, but the recall is relatively lower at 0.7, resulting in an F1-score of 0.78. The Good category exhibits a balanced precision and recall, with an F1-score of 0.82. The Acceptable category presents a similar pattern, with an F1-score of 0.79. In the Poor category, GPC maintains strong Precision and recall, leading to an F1 score of 0.94. It is important to relate these results to the previous table (Table I), which evaluated the models based on general performance metrics. The results in Table III provide a more nuanced view of the models' performance across different grade categories. GPC+POA consistently achieves high Precision, recall, and F1 scores across all grade categories, highlighting its effectiveness in various scenarios. GPC+GEO shows strengths in Precision but faces challenges in the recall, particularly in the Excellent category. The base GPC model also exhibits solid performance, with well-balanced Precision and recall in most grade categories.

Overall, these findings emphasize that the choice of the model should align with the specific needs of the task, considering both general and grade-based performance metrics. GPC+POA excels in Precision and recall across all grade categories, while GPC+GEO and GPC offer balanced performance suitable for various applications.

For a comprehensive evaluation of the model's predictive capabilities and for facilitating model comparisons, Fig. 7 illustrates a bar chart displaying the four distinct grades. This visual representation effectively conveys the models' proficiency in predicting the observed values for each grade, offering insights into their relative performance. When examining the Poor grades, it is noteworthy that the GPC+GEO hybrid model accurately predicted 227 out of 233 measured values, surpassing both the GPC+POA and GPC models in terms of correct predictions. Shifting the attention to the Acceptable grade, the performance of the hybrid models closely aligns, with only a 1 percent difference between them. GPC+POA correctly predicted 51 out of 62 measured values, which is quite similar to GPC+GEO, with 50 correctly predicted values.

In contrast, the GPC model falls behind the hybrid models with only 47 correctly predicted values. When evaluating the good grade, it is evident that the GPC+GEO model outperforms the others by correctly predicting 52 out of 60 measured values, with the GPC model coming close to 51 predicted values. However, in the highest grade, Excellent, the GPC+POA model excels by correctly predicting 37 out of 40 measured values. Notably, in the Excellent grade, the GPC+GEO model lags behind the GPC model's performance. Overall, it is challenging to determine the clear superiority of the models due to their varying performances in different grade categories.

TABLE III. PERFORMANCE EVALUATION INDICES FOR THE DEVELOPED MODELS BASED ON GRADES

Model	**Grade**	**Index values**		
		Precision	Recall	F1-score
GPC+POA	Excellent	0.93	0.93	0.93
	Good	0.91	0.8	0.85
	Acceptable	0.75	0.82	0.78
	Poor	0.96	0.96	0.96
GPC+GEO	Excellent	1	0.6	0.75
	Good	0.76	0.87	0.81
	Acceptable	0.82	0.81	0.81
	Poor	0.96	0.97	0.96
GPC	Excellent	0.88	0.7	0.78
	Good	0.78	0.85	0.82
	Acceptable	0.82	0.76	0.79
	Poor	0.93	0.96	0.94

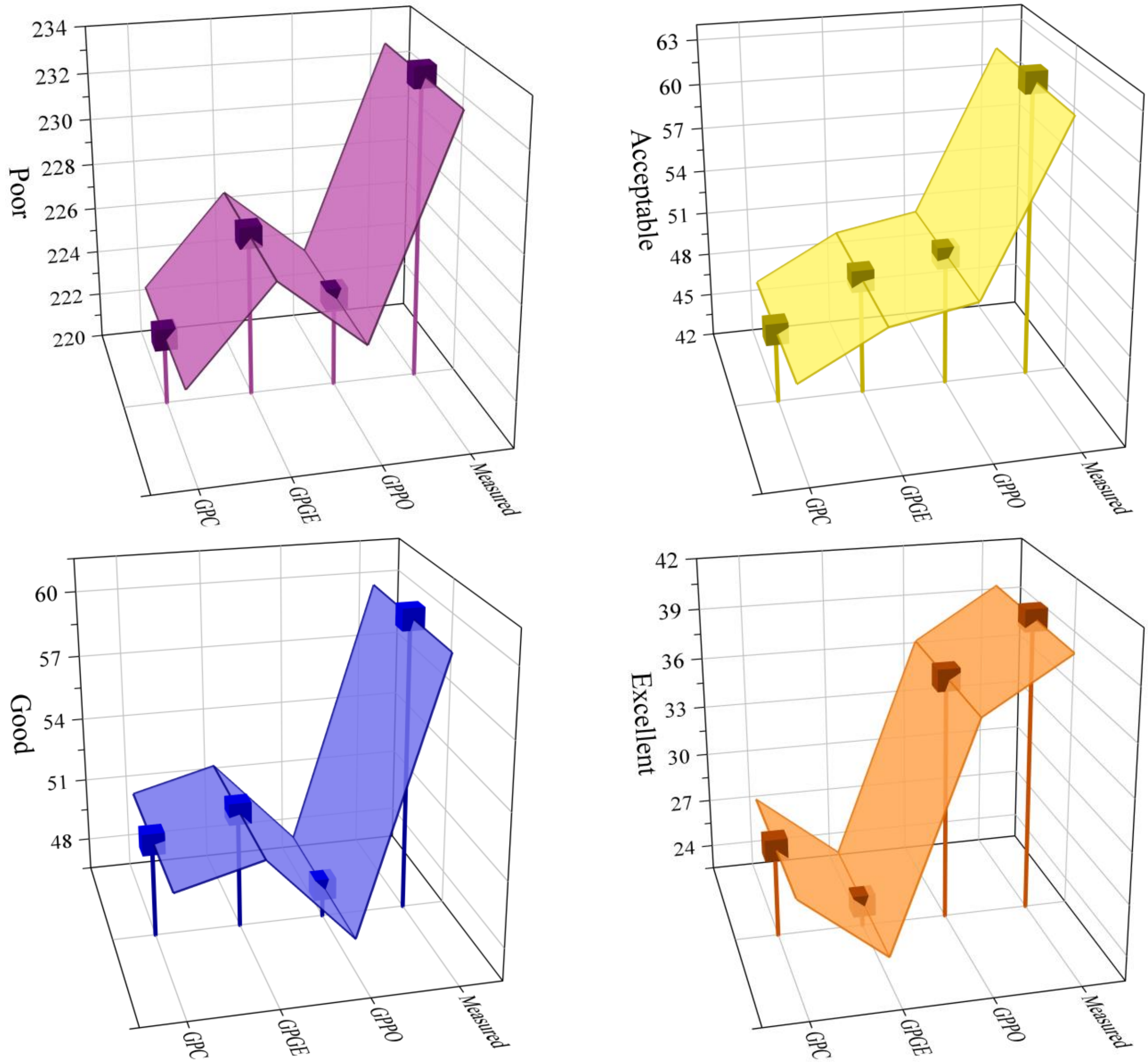


Fig. 7. Column plot for the measured and predicted values.

In Fig. 8, the confusion matrix visually depicts the correspondence between observed and predicted classes by the models. The vertical axis shows the expected classes, and the horizontal axis shows the observed classes. It is evident from this visual representation that the cells along the matrix's main diagonal contain a more significant number of values compared to the remaining cells.

For example, GPC+GEO considers a model that demonstrates a strong ability to make accurate predictions, particularly in the Poor grade. To illustrate this, when dealing with 233 students categorized as Poor, GPC+GEO successfully predicted 227 of them within this category, with only six students being misclassified. This results in the model accurately predicting 97.40% of the observed data within the Poor category. In the case of the Acceptable, Good, and Excellent classes, GPC+GEO achieves prediction accuracies of 80.64%, 86.46%, and 60%, respectively. On the other hand, GPC+POA also displays a high accuracy rate in correctly predicting the Poor, Acceptable, Good, and Excellent classes, with accuracy percentages standing at 96.13%, 82.25%, 80%, and 92.5%, respectively. Similarly, GPC delivers accuracies of 95.70%, 75.80%, 85%, and 70% for the corresponding classes. These examples highlight the models' prediction capabilities in various grade categories, emphasizing their accuracy in predicting student performance across a wide spectrum of Poor, Acceptable, Good, and Excellent classifications.

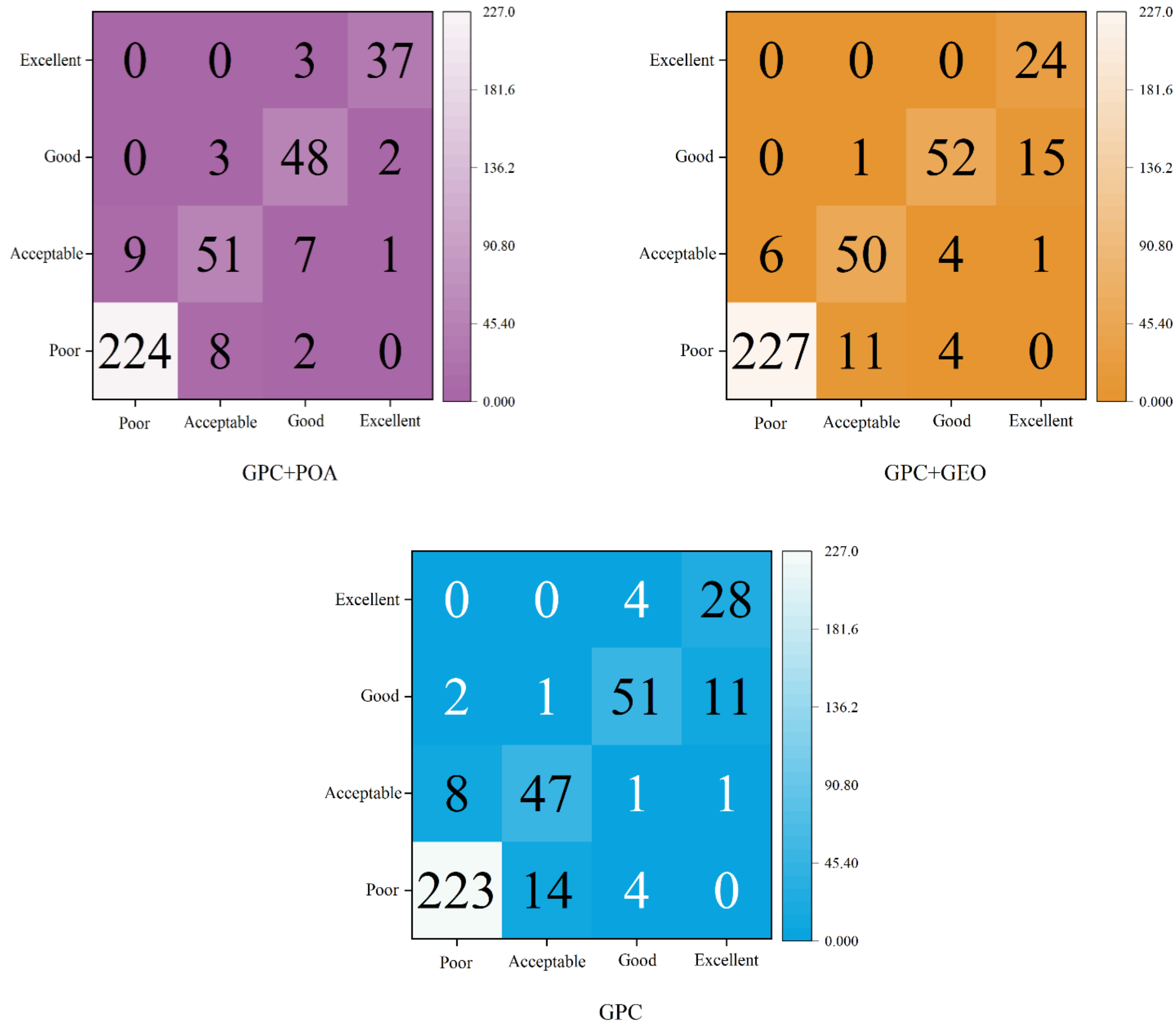


Fig. 8. Confusion matrix for the models' accuracy.

B. Discussion

1) Sensitivity analysis: The impact of input parameters on output values is assessed through the SHAP (Shapley Additive Explanations) sensitivity analysis. Based on the results of this analysis, the significance of the variables has been identified. Fig. 9 illustrates the outcomes of the SHAP-based sensitivity analysis for student performance prediction. According to this figure, it is observed that Freetime, Failures, Medu, Schoolsup, and Fedu experience the highest impact on the G3 values across all categories. Within the Excellent category, these features are found to exert the most substantial influence on the target values. However, for the Good and Acceptable categories, these inputs are not identified as the most influential. In summary, all the inputs are observed to have effects on the G3 values, and through parameter optimization, it is feasible to achieve the highest values.

2) Comparing previous studies vs present research study: A thorough synopsis of the conclusions from four groundbreaking research in the field of student performance is given in Table IV. Among these investigations, Nguyan and Peter's inquiry [26] achieved the best accuracy rate of 82% by using the DTC model. This is noteworthy. However, in the current study, a novel approach integrating the GPC model and POA algorithm yielded exceptional results, achieving a noteworthy accuracy score of 0.911 for G3 prediction. This stands out as the highest accuracy achieved among all referenced works, underscoring the effectiveness and superiority of the proposed methodology.

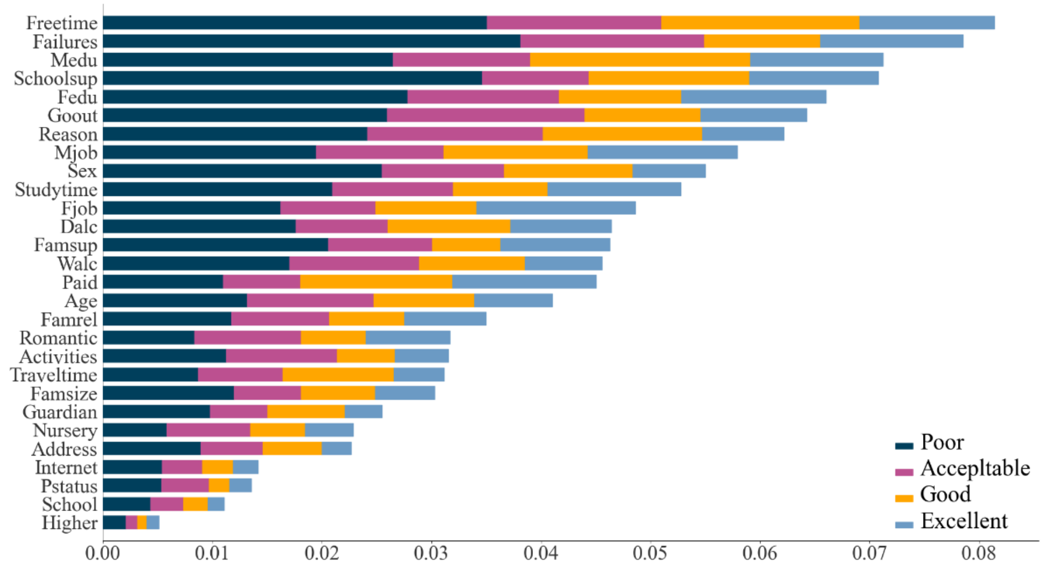


Fig. 9. The results of the SHAP-based sensitivity analysis for assessing the features' impact on output parameters

TABLE IV. EXTENSIVE STUDY RESULTS COMPARED TO THE CURRENT WORK

Author (s)	Models	Accuracy
Bichkar and R. R. Kabra [22]	DTC	69.94%
Kabakchieva [43]	DTC	72.74%
Edin Osmanbegovic et al. [28]	NBC	76.65%
Nguyen and Peter [26]	DTC	82%
Present study for G3	GPC+POA	91.1%

IV. CONCLUSION

In the realm of education, predicting student performance is a critical task, as it holds the potential to revolutionize the way educational institutions operate and provide valuable insights for educators, administrators, and policymakers. This study delved into the world of student performance estimation by harnessing innovative classification techniques, offering an array of promising models such as GPC, GPC+POA, and GPC+GEO. The results of this research have shed light on the capabilities and performance of these models across different educational contexts. GPC, a fundamental model, displayed commendable performance in accurately predicting student grades across various categories, showcasing its reliability in providing a well-rounded assessment of student performance. However, it was the hybrid models, GPC+POA and GPC+GEO, that truly stood out. These models demonstrated their prowess in achieving high Precision, recall, and F1 scores, which are crucial for applications demanding a fine balance between correctly identifying positive instances and minimizing false positives. GPC+POA excelled in predicting Excellent grades, while GPC+GEO showcased its strength in Poor and Good grades, emphasizing the flexibility of these models across different educational scenarios. One of the key takeaways from this study is the importance of model selection based on the specific requirements of the educational task at hand. GPC+POA and GPC+GEO offer tailored solutions for scenarios where Precision, recall, and F1 scores play a pivotal role. In contrast, GPC remains a reliable choice for more general applications. The performance evaluation, as reflected in the results, further demonstrated the versatility of these models in addressing the unique challenges posed by different student performance grades. GPC+POA, for example, showcased superior accuracy in the Excellent grade, while GPC+GEO excelled in the Poor grade. This versatility in handling various performance categories is a testament to the potential of these models to cater to diverse educational settings. As a parting thought, it is essential to recognize the evolving landscape of education and the role that innovative classification techniques can play in shaping its future. These techniques not only provide accurate predictions but also contribute to informed decision-making processes, thus enabling institutions to allocate resources efficiently and support struggling students proactively. In conclusion, this study has provided a glimpse into the exciting possibilities of using innovative classification techniques to estimate student performance. The hybrid models, in particular, have exhibited their potential to enhance the educational landscape by delivering accurate and context-specific predictions. As the field of education continues to evolve, the integration of these innovative techniques may very well hold the key to unlocking a brighter and more data-driven future for both students and educators alike.

ACKNOWLEDGMENT

This work was supported by the Comprehensive Reform and Ideological and Political Ability Enhancement Program for "Three-Whole Parenting" in Universities in Anhui Province "A Study of Digital Enabled Precision Ideological and Political Education in Medical Schools" (sztsjh-2023-6-9).

REFERENCES

[1] J. Watkins, M. Fabielli, and M. Mahmud, "Sense: a student performance quantifier using sentiment analysis," in 2020 International Joint Conference on Neural Networks (IJCNN), IEEE, 2020, pp. 1–6.

[2] Z. Xu, H. Yuan, and Q. Liu, "Student performance prediction based on blended learning," IEEE Transactions on Education, vol. 64, no. 1, pp. 66–73, 2020.

[3] P. Shruthi and B. P. Chaitra, "Student performance prediction in the education sector using data mining," 2016.

[4] F. Ünal, "Data mining for student performance prediction in education," Data Mining-Methods, Applications and Systems, vol. 28, pp. 423–432, 2020.

[5] H. Hamsa, S. Indiradevi, and J. J. Kizhakkethottam, "Student academic performance prediction model using decision tree and fuzzy genetic algorithm," Procedia Technology, vol. 25, pp. 326–332, 2016.

[6] P. Chaudhury and H. K. Tripathy, "An empirical study on attribute selection of student performance prediction model," International Journal of Learning Technology, vol. 12, no. 3, pp. 241–252, 2017.

[7] H. Chanlekha and J. Niramitranon, "Student performance prediction model for early identification of at-risk students in traditional classroom settings," in Proceedings of the 10th International Conference on Management of Digital EcoSystems, 2018, pp. 239–245.

[8] B.-H. Kim, E. Vizitei, and V. Ganapathi, "GritNet: Student performance prediction with deep learning," arXiv preprint arXiv:1804.07405, 2018.

[9] H. Al-Shehri et al., "Student performance prediction using support vector machine and k-nearest neighbor," in 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), IEEE, 2017, pp. 1–4.

[10] I. Khan, A. R. Ahmad, N. Jabeur, and M. N. Mahdi, "A Conceptual Framework to Aid Attribute Selection in Machine Learning Student Performance Prediction Models.," International Journal of Interactive Mobile Technologies, vol. 15, no. 15, 2021.

[11] P. M. Arsad and N. Buniyamin, "A neural network students' performance prediction model (NNSPPM)," in 2013 IEEE International Conference on Smart Instrumentation, Measurement, and Applications (ICSIMA), IEEE, 2013, pp. 1–5.

[12] H. Hassan, S. Anuar, and N. B. Ahmad, "Students' performance prediction model using meta-classifier approach," in Engineering Applications of Neural Networks: 20th International Conference, EANN 2019, Xersonisos, Crete, Greece, May 24-26, 2019, Proceedings 20, Springer, 2019, pp. 221–231.

[13] H. Lu and J. Yuan, "Student performance prediction model based on discriminative feature selection," International Journal of Emerging Technologies in Learning (Online), vol. 13, no. 10, p. 55, 2018.

[14] 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.

[15] B. K. Bhardwaj and S. Pal, "Data Mining: A prediction for performance improvement using classification," arXiv preprint arXiv:1201.3418, 2012.

[16] M. M. R. Khan, M. A. B. Siddique, and S. Sakib, "Non-intrusive electrical appliances monitoring and classification using K-nearest neighbors," in 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET), IEEE, 2019, pp. 1–5.

[17] F. Duzhin and A. Gustafsson, "Machine learning-based app for self-evaluation of teacher-specific instructional style and tools," Educ Sci (Basel), vol. 8, no. 1, p. 7, 2018.

[18] K. M. Hasib et al., "A survey of methods for managing the classification and solution of data imbalance problem," arXiv preprint arXiv:2012.11870, 2020.

[19] K. M. Hasib, N. A. Towhid, and M. G. R. Alam, "Online review based sentiment classification on Bangladesh airline service using supervised learning," in 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), IEEE, 2021, pp. 1–6.

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

[21] 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.

[22] 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.

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

[24] 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.

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

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] M. Chitti, P. Chitti, and M. Jayabalan, "Need for interpretable student performance prediction," in 2020 13th International Conference on Developments in eSystems Engineering (DeSE), IEEE, 2020, pp. 269–272.

[31] 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.

[32] S. O. Oppong, "Predicting Students' Performance Using Machine Learning Algorithms: A Review," Asian Journal of Research in Computer Science, vol. 16, no. 3, pp. 128–148, 2023.

[33] 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.

[34] 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.

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

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

[37] H. Nickisch and C. E. Rasmussen, "Approximations for binary Gaussian process classification," Journal of Machine Learning Research, vol. 9, no. Oct, pp. 2035–2078, 2008.

[38] J. Hensman, A. Matthews, and Z. Ghahramani, "Scalable variational Gaussian process classification," in Artificial Intelligence and Statistics, PMLR, 2015, pp. 351–360.

[39] P. Trojovský and M. Dehghani, "Pelican optimization algorithm: A novel nature-inspired algorithm for engineering applications," Sensors, vol. 22, no. 3, p. 855, 2022.

[40] N. Alamir, S. Kamel, T. F. Megahed, M. Hori, and S. M. Abdelkader, “Developing hybrid demand response technique for energy management in microgrid based on a pelican optimization algorithm,” Electric Power Systems Research, vol. 214, p. 108905, 2023.

[41] W. Tuerxun, C. Xu, M. Haderbieke, L. Guo, and Z. Cheng, “A wind turbine fault classification model using broad learning system optimized by improved pelican optimization algorithm,” Machines, vol. 10, no. 5, p. 407, 2022.

[42] F. N. Al-Wesabi et al., “Pelican Optimization Algorithm with federated learning driven attack detection model in the Internet of Things environment,” Future Generation Computer Systems, 2023.

[43] 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.