Predictive Modeling of Student Performance Through Classification with Gaussian Process Models

Xiaowei ZHANG[1]*, Junlin YUE[2]
Department of Educational Science and Music, Luoyang Institute of Science and Technology, luoyang 471023, China[1]
Department of Foreign Language Teaching and Research, Luoyang Normal University, Luoyang 471934, China[2]

***Abstract*—In the contemporary educational landscape, proactively engaging in predictive assessment has become indispensable for academic institutions. This strategic imperative involves evaluating students based on their innate aptitude, preparing them adequately for impending examinations, and fostering both academic and personal development. Alarming statistics underscore a notable failure rate among students, particularly in courses. This article aims to employ predictive methodologies to assess and anticipate the academic performance of students in language courses during the G2 and G3 academic exams. The study utilizes the Gaussian Process Classification (GPC) model in conjunction with two optimization algorithms, the Population-based Vortex Search Algorithm (PVS) and the COOT Optimization Algorithm (COA), resulting in the creation of GPPV and GPCO models. The classification of students into distinct performance categories based on their language scores reveals that the GPPV model exhibits the highest concordance between measured and predicted outcomes. In G2, the GPPV model demonstrated the notable 51.1% correct categorization of students as Poor, followed by 25.57% in the Acceptable category, 14.17% in the good category, and 7.7% in the Excellent category. This performance surpasses both the optimized GPCO model and the singular GPC model, signifying its efficacy in predictive analysis and educational advancement.**

Keywords—Academic performance; language; hybrid algorithms; Gaussian Process Classification; population-based Vortex Search Algorithm; COOT Optimization Algorithm

I. INTRODUCTION

Artificial Intelligence (AI) has significantly contributed to the growth and productivity of various industries, including transportation, communication, commerce, and finance. However, its impact on the education sector requires further refinement, particularly in the enhancement of AI-based learning systems designed to support students in both classroom and remote settings, as well as individuals with special needs [1], [2]. In software development, there is a notable emergence of instructional software tailored to individual learning needs. These innovative tools not only foster connectivity among learners but also provide access to a vast array of digital materials. Furthermore, they support decentralized learning tools, creating a dynamic and engaging learning environment that caters to diverse educational requirements [1].

In educational contexts, traditional machine learning approaches have been extensively employed, including support vector machine (SVM) [3], [4], decision tree [5], and matrix factorization (MF) [6], [7], along with their respective extensions [8], [9]. The SVM algorithm, in particular, is a frequently applied method known for its superior performance [3], [10]. However, its original design for binary classification poses limitations, as it does not inherently account for the relative importance of feature vector elements. While extensions for multiple classification problems exist, they may not consistently yield optimal results. The advent of deep learning [11], [12], [13], [14], [15] has brought about notable advancements in the educational domain. Despite its promising performance, the issue of overtraining looms large, especially when the dataset size is not sufficiently large. Recent studies propose that a moderately deep Artificial Neural Network (ANN) [16] can offer comparable accuracy without succumbing to overtraining, making it a viable alternative in educational applications. This nuanced understanding of traditional and contemporary machine learning methods provides valuable insights for selecting appropriate models tailored to specific educational contexts and datasets.

Analyzing and mitigating factors influencing student performance is a paramount concern for educational institutions aiming to reduce student failure rates. Educational data mining (EDM) emerges as a pivotal technique in this endeavour [17]. EDM encompasses the development of methodologies tailored to handle diverse data types within educational systems, ultimately enhancing students' learning outcomes [18]. Through the amalgamation of statistical, machine learning, and data mining approaches, EDM endeavours to extract and modify information from educational data, facilitating informed decision-making in the educational domain. The primary objective of EDM is to glean valuable insights from educational data, enabling effective decision-making to enhance educational outcomes [18]. By harnessing the power of predictive modelling, EDM can forecast students' academic achievement at an early stage [19]. This multifaceted approach encompasses various strategies to analyze and interpret educational data, offering institutions valuable tools for proactive intervention and support to improve overall student success.

The data mining project, initiated in 2013, constitutes a comprehensive effort aimed at extracting valuable insights from existing datasets to inform university management strategies, particularly in understanding student dynamics and refining university marketing policies. A thorough review of the literature indicates a sustained interest in these issues, with 63 researchers exploring various facets in recent years. Luan [20] delves into the potential applications of data mining in higher education, emphasizing resource efficiency and academic effectiveness. Several papers [20], [21], [22], [23] scrutinize student typology and targeted marketing through data mining models. Similarly, DeLong et al. [24] and Nandeshwar and Chaudhari [25] focus on student types, marketing strategies, and enrollment prediction models based on admissions data, utilizing diverse data mining methods. Dekker et al. [26] concentrate on predicting student dropouts, contributing to the broader discourse on enhancing university management and decision-making through data-driven approaches in higher education.

In this comprehensive investigation, we utilize the Gaussian Process model along with two advanced optimization algorithms—the Population-based Vortex Search Algorithm (PVS) and the COOT Optimization Algorithm (COA)—to enhance the model and develop distinct versions (GPPV and GPCO). The primary objective is to categorize students' performance in language courses based on G2 and G3 exam results into four performance grades: poor, acceptable, good, and excellent. We rigorously evaluate and compare the predictive performance of these models using key classification metrics such as Accuracy, Precision, Recall, and F1-score.

This study aims to provide a thorough understanding of the research findings, with subsequent sections delving into the impact of carefully selected input data on model outcomes. Beyond categorizing students, the analysis extends to examining the models' predictive capabilities, highlighting their strengths and limitations. Furthermore, the article offers detailed explanations of the Gaussian Process model, the Population-based Vortex Search Algorithm, and the COOT Optimization Algorithm, providing readers with a comprehensive perspective on the methodologies employed in this study.

II. Data Selection and Preparation

Data mining, known as database knowledge discovery, involves the systematic extraction of valuable insights and patterns from large datasets. This process employs various techniques and algorithms to uncover hidden knowledge, contributing to informed decision-making and meaningful analysis. This study relies on an extensive dataset from previous research covering various variables. These include school, sex (female or male), age, residence, family size (famsize), parental cohabitation status (Pstatus), and details about the mother's and father's education and occupations (Medu, Fedu, Mjob, Fjob). The dataset also explores reasons influencing school choice, such as proximity, reputation, and course preferences. Additionally, it delves into aspects like the student's guardian (whether it is the father, mother, or another guardian), travel time to school (traveltime), weekly study hours (studytime), past failures, participation in educational support programs (schoolsup), family educational support (famsup), engagement in paid classes and extracurricular activities, attendance at nursery school, aspirations for higher education, internet access at home, romantic relationships, family relationships (famrel), free time, going out with friends, alcohol consumption in weekdays (Dalc) and weekends(Walc), health status, and school absences.

In Fig. 1, a visual representation illustrates the influence of each specified parameter on the outcomes of both G2 and G3 tests. The visualization employs a color-coded scheme, with red denoting the most positive impact (+1) and blue representing the most negative influence (-1). As the colours gradually fade, approaching zero, the corresponding influence diminishes. The shapes depicted tend towards circular, with a tendency towards elongation (oval) when nearing the maximum or minimum values. Examining the graph's details, it is evident that the most substantial positive effect is associated with the impact of each parameter on itself, visually depicted by the diameter of the shape. Following this, the influence of G2 on G3 emerges as particularly significant. In the final two lines of the figure, portraying the impact of each parameter on G2 and G3, it becomes apparent that failures exert the most pronounced negative impact, aligning with logical expectations. Furthermore, aspirations for higher education exhibit the most positive influence on these tests. Notably, the majority of effects are proximal to zero, signifying a circular and faint representation.

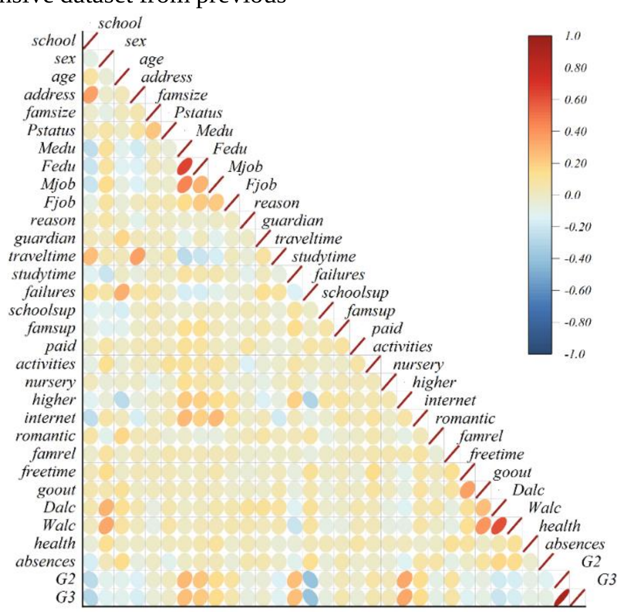

Fig. 1. Correlation matrix for the input and output variables.

III. Gaussian Process Classifier (GPC)

A nonparametric probabilistic classification model based on Gaussian procedure regression is the GPC [27]. Here, the value of an underlying latent function connected to the input data exhibits a monotonic correlation with the probability that the incoming data falls into a certain class. Initially establishing a prior for this latent function, the available information aids in deducing the values of hyperparameters that control various aspects of the function as well as the latent function's posterior distribution.

Consider a dataset comprising observations denoted as $x = (x_1, x_2, \ldots\ldots, x_N)$ and $y = (y_1, y_2, \ldots\ldots, y_n)$. Here, $x_t = \left(x_{t,1}, x_{t,2}, \ldots, x_{t,P}\right)^T$ represents the vector of p inputs at time t, and y_t corresponds to the associated binary response, i.e., $y_t \in \{1, -1\}$ for $t = 1, \cdots, n$. To streamline the representation of the response distribution y_t, it is expedient to introduce the concept of an unobserved "latent function," denoted as f_t.

$$p(y_t = 1) = Lf(f_t) \tag{1}$$

In the given context, Lf serves as the link function, and $f_t = X_t'\beta$, where β denotes the coefficient vector. This implies that the likelihood of $y_t = 1$ is contingent on a nonlinear function, specifically the result of the linear combination of the input data X_t. To illustrate, a logistic model for a binary target can be conceptualized as follows:

$$p(y_t = 1) = [1 = \exp(-X_t'\beta)]^{-1} \tag{2}$$

The link function is defined as $Lf(z) = (1 + exp(-z))^{-1}$.

Given inputs $X_1, X_2, \cdot\cdot\cdot, X_n$, let the latent functions $f = [f(X_1), \cdot\cdot\cdot, f(X_n)]$ follow a multivariate Gaussian distribution. Its mean and covariance functions entirely specify a Gaussian process. In simpler terms:

$$f \mid X \sim GP\ (\mu(X), V(X, X')) \tag{3}$$

Here, the mean vector $\mu(X)$ is represented as $[\mu(X_1), \cdot\cdot\cdot, \mu(X_n)]$, while $V(X, X')$ signifies the $n \times n$ covariance matrix of f, where the (i, j)-th element $V_{i,j}$ is expressed as $V(\mathrm{X}_i, \mathrm{X}_j)$. It is pertinent to note that for this article, $\mu(X)$ is considered to be zero, specifically $\mu(X) = 0$.

$$f \mid X \sim GP\ (0, V(X, X')) \tag{4}$$

The covariance function, $V(X, X')$, is pivotal in defining the relationship between latent variables, determining the response at a single input, X_i, is influenced by responses at another input, X_j. Different kernel functions, such as the Automatic Relevance Determination (ARD) exponential kernel function outlined in Eq. (5), can be utilized to define the covariance function $V(\cdot,\cdot)$. These kernel functions introduce varying levels of smoothness and structural characteristics, offering flexibility to tailor the covariance function to capture the intricate relationships between latent variables more effectively.

$$V\left(X_i, X_j\right) = \sigma_0^2 exp\left(-\frac{1}{2}\sum_{l=1}^{n}\frac{\left(X_{i,l} - X_{j,l}\right)^2}{\lambda_l^2}\right) \tag{5}$$

The hyper-parameters λ_l's represent the characteristic length scales and the scale parameter σ_0^2 signifies the relevance of different regions. In essence, the characteristic length scale provides a concise measure of the distance between input values X_i, indicating the range within which response values become uncorrelated.

Let X^* represent the input data in the testing or prediction dataset. The latent value corresponding to the testing data is denoted as f^* and expressed as $X^{*\prime}\beta$, can be derived. The posterior probabilistic prediction of the probability $\pi^* = p(y^* = 1)$ can be calculated using Eq. (6).

$$p(y^* = 1 \mid X, Y, X^*) = \int Lf(f^*)p(f^* \mid X, Y, X^*)df^*, \tag{6}$$

Where

$$p(f^* \mid X, Y, X^*) = \int p(f^* \mid X, X^*, f)p(f \mid X, Y)df \tag{7}$$

The probability $p(f \mid X, Y)$ can be computed using the formula presented in Eq. (8).

$$p(f \mid X, Y) = \frac{1}{Z}\, p(f \mid X) \prod_{i=1}^{n} p(y_i \mid f_i) \tag{8}$$

Here, $p(f \mid X)$ denotes the Gaussian prior distribution of f, $GP\ (0, V(X, X')$. The normalization term Z corresponds to the marginal likelihood, expressed as $Z = \int p(f \mid X) \prod_{i=1}^{n} p(y_i \mid f_i)$. For binary classification, a probit likelihood is employed, where $p(y_i \mid f_i)$ represents the density function of a standard normal distribution.

The computational complexity of Gaussian process methods is typically $O(n^3)$ due to the inversion of the covariance matrix. However, various sparse approximation techniques, including Markov Chain Monte Carlo (MCMC), Laplace approximation (LA), expectation propagation (EP), and variational inference (VI), have been introduced to mitigate this complexity. For a detailed exploration of these approximation methods in the context of Gaussian Process Classification (GPC), refer to [28]. Additionally, [29] provides a comprehensive review of various sparse approximation methods.

In the investigation conducted, the Expectation Propagation (EP) algorithm from [30] is employed to approximate the Gaussian posterior distribution specified in Eq. (6). Eq. (8), representing the posterior probability function $p(f \mid X, Y)$, can be approximated using Eq. (9).

$$\begin{aligned} p(f \mid X, Y) &= \frac{1}{Z}\, p(f \mid X) \prod_{i=1}^{n} t_i\left(f_i \middle| \tilde{Z}_i, \tilde{\mu}_i, \tilde{\sigma}_2^2\right) \\ &= \mathcal{N}(\mu, \textstyle\sum) \end{aligned} \tag{9}$$

The site parameters, expressed as $\tilde{\mu}_i$ and $\tilde{\sigma}_i^2$, are integral components within the normalized Gaussian distribution denoted by the notation $\mathcal{N}$.

$$\mu = \sum \tilde{\textstyle\sum}^{-1} \tilde{\mu} \tag{10}$$

And

$$\textstyle\sum = (V^{-1} + \tilde{\sum}^{-1})^{-1} \tag{11}$$

Matrix $\widetilde{\Sigma}$ is a diagonal matrix characterized by the elements $\widetilde{\Sigma}_{ii} = \tilde{\sigma}_i^2$, and $\tilde{\mu} = (\tilde{\mu}_i, \ldots, \tilde{\mu}_N)$.

In accordance with Eq. (7) and (9), Eq. (12) presents the Gaussian approximation to the posterior distribution outlined in Eq. (6).

$$P(y^* = 1 \mid X, Y, X^*) = \Phi\left(\frac{V^{*\prime}\left(K + \widetilde{\Sigma}\right)^{-1}\tilde{\mu}}{\sqrt{1 + V(X^*, X^*) - V^{*\prime}\left(V + \widetilde{\Sigma}\right)^{-1}V^{*\prime}}}\right) \quad (12)$$

Here $V^* = V(X, X^*)$.

IV. Optimization Algorithms

A. Coot Optimization Algorithm (COA)

The Coot Optimization Algorithm (COA) draws inspiration from the collective behaviours of Coots. Water birds are known for various movements on the water, including random, chain, leader-driven, and leader-adjusted behaviours. In its metaheuristic optimization approach, COA initializes a population randomly based on Eq. (13) [31]:

$$CP(i) = rand\ (1, N) \times (Ub - Lb) + Lb \quad (13)$$

$CP(i)$ represents the positions of an individual coot, where N is the problem's dimensionality or the number of variables. Ub and Lb indicate the upper and lower limits of the exploration space for the search.

$$Ub = [Ub_1, Ub_2, \ldots, Ub_N]\ , Lb = [Lb_1, Lb_2, \ldots, Lb_N] \quad (14)$$

Following the initial population setup, the positions of the coots are then modified using four distinct movement patterns.

1) Random movement: The position S for this particular movement is initially randomized according to the equation outlined in Eq. (15):

$$S = rand(1, N) \times (Ub - Lb) + Lb \quad (15)$$

To avoid entrapment in local optima, the position is updated following the procedure defined in Eq. (16):

$$CP(i) = CP(i) + A \times R_2 \times (S - CP(i)) \quad (16)$$

The variable R_2 is a randomly generated number falling within the interval [0,1], while A is computed using the equation specified in Eq. (17):

$$A = 1 - W \times \left(\frac{1}{Iter}\right) \quad (17)$$

$Iter$ denotes the maximum allowable number of iterations, and W represents the current iteration count.

2) Chain movement: To perform the chain movement, the average position of two coot birds can be computed using the formula provided in Eq. (18):

$$CP(i) = \frac{CP(i-1) + CP(i)}{2} \quad (18)$$

where, $CP(i-1)$ represents the position of the second coot in the sequence.

3) Adjusting position according to the leader: In each subgroup, the position of a coot bird is modified in alignment with the leader's position, resulting in the follower moving closer to the leader. The leader is chosen based on the equation specified in Eq. (19).

$$q = 1 + (i\ MOD\ NW) \quad (19)$$

where, q denotes the index of the leader, i represents the number of the follower coot bird, and NW signifies the overall count of leaders in the group.

In this specific movement, the position of a coot is adjusted following the formula provided in Eq. (20):

$$CP(i) = LP(q) + 2 \times R_1 \times Cos(2R\pi) \times (LP(q) - CP(i)) \quad (20)$$

The notation $CP(i)$ represents the current position of the coot bird, and $LP(q)$ stands for the position of the selected leader. The parameters R_1, a random number within [0, 1], and R, a random number within [-1, 1], play a role in the position update computation outlined in Eq. (20).

4) Leader movement: The positions of leaders undergo updates according to Eq. (21), which are aimed at transitioning from local optimal positions to global optimal positions.

$$LP(i) = \begin{cases} B \times B_3 \times Cos(2\pi R) \times \left(gBst - LP(i)\right) + gBst & B_4 < 0.5 \\ B \times B_3 \times Cos(2\pi R) \times \left(gBst - LP(i)\right) - gBst & B_4 \geq 0.5 \end{cases} \quad (21)$$

In this scenario, $gBst$ represents the optimal position, and B_3 and B_4 are random numbers within the interval [0, 1]. The value B is determined through the calculation outlined in Eq. (22):

$$B = 2 - L \times \left(\frac{1}{Iter}\right) \quad (22)$$

B. Population-based Vortex Search Algorithm (PVSA)

The Vortex Search algorithm, a metaheuristic known for its efficient exploitation capabilities, revolves around a single solution [32]. It swiftly generates new candidate solutions using a Gaussian distribution, clustering them around a central point. However, this approach may face premature convergence issues in specific situations despite efforts to maintain diversity in the search space. Population-based approaches, particularly effective in exploration phases, excel in investigating unexplored positions by generating fresh coordinates based on accumulated knowledge from previous iterations [33].

1) Initializing: In the algorithm's setup phase, crucial control parameters such as population size (psize), vortex size (vsize), termination criteria, and mutation probability (η_m) are defined. The psize parameter determines the total candidate solutions generated in one iteration, with $vsize$ being half of $psize$. Initially, the candidate solutions (CS) count equals $vsize$, extending subsequently to $psize$. The algorithm halts upon reaching the predetermined maximum number of function evaluations ($maxFEs$). In the second phase, the polynomial mutation operation is contingent on the probability parameter η_m. Furthermore, μ_0 and q_0 are computed sequentially using Eq. (23) and Eq. (24).

$$\mu_0^i = \frac{up_i + low_i}{2} \tag{23}$$

$$q_0^i = \sigma_0^i = \frac{\max(up_i) - \min(low_i)}{2} \tag{24}$$

up means upper, and low means lower.

2) First phase: In the initial iteration of this stage, a population of $psize$ individuals is randomly generated. Subsequent iterations limit random generation to only half of the population, denoted as $vsize$. The stage concludes with the update of the central point (μ), achieved by replacing it with the best-discovered solution. This update employs a Gaussian distribution to generate half of the population, following the principles outlined in Eq. (25) of the original VS algorithm. While one-half of the population undergoes exploitation focused on the best centre, the other half is updated using a population-based approach with selection pressure. Solutions exceeding a specified limit are adjusted to fall within the designated range.

$$s_i^t(x_i^t \Big| \mu_t, v) = ((2\pi)^d |v|)^{-\left(\frac{1}{2}\right)} e^{\left(-1/2\left(x_i^t - \mu_t\right)^T v^{-1}(x_i^t - \mu_t)\right)} \tag{25}$$

$$\begin{aligned} s_i(low_i \vee s_i)up_i &\rightarrow s_i \\ &= rand \times (up_i - low_i) + low_i \end{aligned} \tag{26}$$

The original VS algorithm utilizes the initial centre point (μ_0) to generate the initial population, though it is not directly part of that population. A modification to the VS algorithm has given rise to PVSA algorithm variants. In $PVSA_a$, μ_0 is included in the initial population, while $PVSA_b$ excludes it. In the first iteration of $PVSA_a$, μ_0 serves as the initial candidate solution ($POP(1)$) in the population, with the remaining $psize - 1$ candidate solutions ($POP(2{:}psize)$) generated randomly. In contrast, the initial population of $PVSA_b$ is formed by randomly generating $psize$ candidate solutions ($POP(1{:}psize)$).

3) Second phase: Population-based algorithms, distinct from single-solution counterparts, leverage interactions among candidate solutions across iterations to adapt their positions in the search process. These algorithms encapsulate the experiences of candidates collectively or individually in vector form, fostering effective information exchange. Take the PVSA algorithm as an example, employing a proportional selection approach amalgamated from the observer bee phase of the ABC algorithm with tailored adjustments for problem minimization. Eq. (27) computes the selection probability vector (pb) for each candidate solution.

$$pb_i = csum_i / csum_{psize} \tag{27}$$

$$csum_i = \sum_{j=1}^{i} normp_i \ \ and$$

$$normp_i = p_i / \sum_{i=1}^{psize} p_i \ \ and$$

$$p_i = 0.9 \times \left(max\{\vec{f}\} - f_i\right) + 0.1$$

The variable f symbolizes the fitness metric linked to the $i-th$ solution, while $max\{f^{\rightarrow}\}$ signifies the maximum fitness value within the existing population. p_i represents the rescaled fitness measure of the $i-th$ solution concerning minimization. The probabilities derived from the normalization of p values are denoted as $normp$, ensuring their confinement within the range of $[0.5-1]$.

In the latter segment of the population, encompassing solutions identified as CS_i where i lies within the range of $vsize + 1\ to\ psize$, a neighbouring solution is randomly chosen from the entire population. The selection is influenced by the $prob$ vector. Employing Eq. (28), the value of a randomly selected dimension is altered to generate a novel solution designated as CS_{new}. Following this modification, the adjusted dimension's value is scrutinized to determine if it surpasses predefined limits, as specified in Eq. (29).

$$\begin{aligned} &CS_{new} = CS_{cur}\ then\ CS_{new}^i = CS_{cur}^i + \\ &(CS_{cur}^i - CS_{neigbour}^i) \times (r - 0.5) \times 2 \end{aligned} \tag{28}$$

$$CS_{new} = \begin{cases} low_i, & CS_{new}^i < low_i \\ CS_{new}^i, & lower_i \le CS_{new}^i \le up_i \\ up_i, & CS_{new}^i > up_i \end{cases} \tag{29}$$

The evaluation of the newly generated solution, denoted as CS_{new}, involves incorporating a randomly chosen number r from the range of 0.5 to 1. Subsequently, the newly computed fitness is juxtaposed with the fitness of the current solution, CS_{cur}. If CS_{new} demonstrates a superior fitness compared to CS_{cur}, it supplants the latter.

However, in cases where CS_{new} falls short of surpassing CS_{cur}, a mutant solution designated as CS_{mut}, is crafted using polynomial mutation. This mutation process adheres to the procedural steps expounded in Eq. (30).

$$CS_{mut} = CS_{cur} + \delta_q \times (up - low)$$

$$\delta_q = \begin{cases} \left[\frac{2r + (1-2r)}{(1-\delta_1)^{\eta_m + 1}}\right]^{\frac{1}{\eta_m + 1}}, & if\ \ r \le 0.5 \\ 1 - \left[\frac{2(1-r) + 2(r-0.5)}{(1-\delta_2)^{\eta_m + 1}}\right]^{\frac{1}{\eta_m + 1}}, & otherwise \end{cases} \tag{30}$$

$$\delta_1 = \frac{CS_{cur} - low}{up - low}$$

$$\delta_2 = \frac{up - CS_{cur}}{up - low}$$

In these situations, a random number rnd is generated for each dimension between 0.5 and 1. The polynomial mutation operator, known for overcoming local optima in metaheuristics, introduces perturbations into the solution. A selection process favours the superior solution between CS_{cur} and CS_{mut}. After this step, the central point μ is updated with the best solution.

After assessing Eq. (31), the radius size for the subsequent generation diminishes at the conclusion of the ongoing generation. The PVS algorithm persists until it reaches the maximum number of function evaluations. Initially, $vsize$ solutions within the reduced radius are duplicated, and in the subsequent phase, random data is incorporated into the solutions constituting the remaining population.

$$r_t = \sigma_0 \times \frac{1}{x} \times \Gamma(x, a_t)$$
$$where \;\; a_t = \frac{(MaxFEs - Fes)}{MaxFEs} \tag{31}$$
$$then \;\; if \;\; (a_t \leq 0) a_t = 0.1$$

V. Performance Evaluation Metrics

The evaluation of the classification performance of the developed models is delineated through the presentation of statistical metrics, as detailed in Eq. (32) – Eq. (35):

$$A = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \tag{32}$$

$$P = \frac{Tp}{Tp + Fp} \tag{33}$$

$$R = TpR = \frac{Tp}{P} = \frac{Tp}{Tp + Fn} \tag{34}$$

$$F1\_score = \frac{2 \times R \times P}{R + P} \tag{35}$$

where A, P , and R represent Accuracy, Precision, and Recall, Tp (True positives) denotes the occurrences where the model accurately predicted the outcome. Conversely, Fp (False positives) represents instances where the model's forecasts were incorrect. Tn (True negatives) refers to situations where the model made accurate predictions, while Fn (False negatives) pertains to instances where the model inaccurately predicted the outcome.

VI. Clustering

Clustering is a fundamental technique in data analysis and machine learning, utilized to identify and group similar data points within a dataset. This process involves partitioning data into clusters, where items in the same cluster share common characteristics, while those in different clusters are distinct from each other. The primary goal of clustering is to uncover inherent structures in data without prior knowledge of category labels, making it an unsupervised learning method. Various clustering algorithms have been developed, each with its unique approach to grouping data. K-means clustering, one of the most widely used methods, partitions data into K distinct clusters by minimizing the variance within each cluster. It is particularly effective for datasets where clusters are spherical and of similar size. Hierarchical clustering, another popular technique, builds a tree-like structure of nested clusters, offering a visual representation of data organization. This method is advantageous when the number of clusters is not predetermined. In the context of educational assessment, clustering can be a powerful tool. By applying clustering algorithms to student performance data, educators can identify distinct groups of students with similar learning behaviors and academic outcomes. For instance, clustering students based on their scores during the G2 and G3 exams can reveal patterns that are not immediately apparent through traditional grading methods. These insights can inform targeted interventions, tailored support, and personalized learning plans, ultimately enhancing the educational experience. Moreover, clustering combined with predictive models, like the Gaussian Process Classification (GPC) discussed earlier, can further refine the categorization of student performance. By integrating clustering techniques with optimization algorithms such as the Population-based Vortex Search Algorithm (PVS) and the COOT Optimization Algorithm (COA), educators can achieve more accurate and actionable predictions. This hybrid approach not only improves the precision of student assessments but also supports strategic decision-making in educational institutions.

VII. Convergence Assessment

In this study, the GPC underwent optimization through the application of metaheuristic optimization algorithms, specifically PVS and CO. The integration of these algorithms with GPC led to the development of hybrid models, termed GPPV and GPCO. To assess the convergence of these optimized models, a robust evaluation method was employed, which involved generating a convergence curve, exemplified in Fig. 2. This curve illustrates the trajectory of Accuracy measurements across 200 iterations. Upon comparative analysis of line plots for G2 and G3 in Fig. 2, a distinct observation emerges. Notably, both the GPPV and GPCO models stabilize before the 150th iteration. Further examination reveals that throughout this convergence process, the GPPV consistently outperforms its GPCO counterpart in both G2 and G3 scenarios.

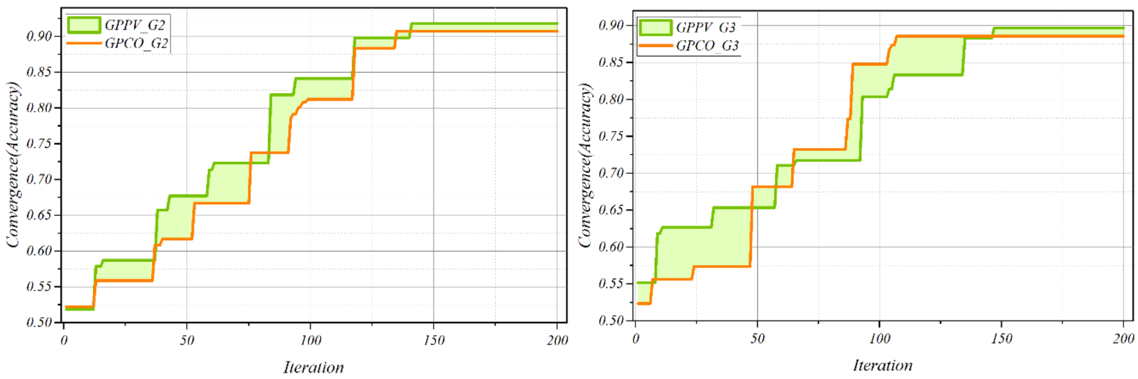

Fig. 2. Line plot for convergence of hybrid models.

VIII. Significance of Model

The significance of the models developed in this study lies in their potential to revolutionize educational assessment and intervention strategies. The choice of the Gaussian Process Classification (GPC) model, enhanced with the Population-based Vortex Search Algorithm (PVS) and the COOT Optimization Algorithm (COA), was driven by several key factors that make these methods particularly suitable for addressing the complexities of educational evaluation. Firstly, the GPC model is renowned for its robustness and flexibility in handling non-linear relationships, which are often present in educational data. This makes it highly effective in capturing the nuanced patterns of student performance. Secondly, the integration of PVS and COA enhances the GPC model's predictive accuracy and reliability. These optimization algorithms are designed to efficiently search large solution spaces and find optimal parameter settings, which is crucial for developing precise predictive models in complex domains like education. Furthermore, these advanced methodologies offer more than mere categorization—they provide a comprehensive framework for educational institutions to identify students who may need additional support or resources early on. By accurately predicting which students are likely to struggle, educators can implement targeted interventions to address specific needs, thereby improving overall educational outcomes. This proactive approach can help reduce the alarming failure rates, particularly in language courses, and promote both academic and personal development among students. Moreover, the use of PVS and COA exemplifies the integration of cutting-edge computational techniques in educational research. This not only enhances the predictive power of the models but also paves the way for future studies to explore and implement similar methodologies in different educational contexts. In summary, the models developed in this study are significant for their ability to provide accurate, early predictions of student performance, enabling targeted interventions and fostering improved educational outcomes. The innovative use of optimization algorithms further underscores the potential of advanced computational techniques in enhancing educational assessment and intervention strategies.

IX. Results and Discussion

The primary objective of this study was to introduce and evaluate three predictive models employing a classification methodology, to forecast students' academic performance in language courses and strategically enhance subsequent grades. Among these models, one utilized a Gaussian Process Classifier (GPC), while the remaining two were derived through the optimization of GPC via the Population-based Vortex Search Algorithm (PVS) and COOT Optimization Algorithm (CO). The dataset underwent partitioning into 70% for training and 30% for model testing. Evaluation metrics, including Accuracy, Precision, Recall, and F1-score, were computed and presented for both the training and testing phases of all models in Tables I and II. It is noteworthy that metric values during the training phase exceeded those observed in the testing phase across all models. The GPPV model exhibited superior performance, recording the highest values across all metrics for G2 and G3 (Accuracy G2 = 0.918, Accuracy G3 = 0.897, Precision G2 = 0.925, Precision G3 = 0.900, Recall G2 = 0.918, Recall G3 = 0.897, F1-score G2 = 0.916, and F1-score G3 = 0.896).

These input variables encompass a diverse array of data types, including nominal, numeric, and binary, thereby providing a comprehensive and informative dataset for the study. Moreover, the academic year comprises three final exams, with a focus on predicting the last two exams, specifically G3 and G2. These grades, ranging from zero (indicating the lowest score) to 20 (representing the highest attainable score), are reported by the school. To further classify the reported scores, students are segmented into four distinct categories based on their G3 and G2 performance: Poor (0-12 range), Acceptable (12-14 range), Good (14-16 range), and Excellent (16-20 range). According to Fig. 3, the examination of the G2 distribution for the GPPV model unveils a dominant majority of students (51.1%) falling within the Poor category. Subsequently, 25.57% are classified as Acceptable, 14.17% as Good, and 7.7% as Excellent. Transitioning to the G3 analysis for the GPPV model, 46.37% of students are categorized as Poor, followed by 23.72% in the Acceptable range, 17.25% in the good category, and 12.63% in the Excellent category. These distribution patterns in Fig. 3 delineate the varying proportions of students across performance categories, providing insights into the effectiveness of the GPPV model in predicting academic outcomes in both G2 and G3 assessments.

TABLE I. RESULT OF PRESENTED MODEL FOR G2

Model	phase	Index values			
		Accuracy	***Precision***	***Recall***	***F1_score***
GPC	Train	0.894	0.909	0.894	0.885
	Test	0.887	0.904	0.887	0.881
	All	0.892	0.907	0.892	0.884
GPCO	Train	0.914	0.918	0.914	0.910
	Test	0.892	0.903	0.892	0.892
	All	0.908	0.914	0.908	0.905
GPPV	Train	0.914	0.921	0.914	0.911
	Test	0.928	0.934	0.928	0.928
	All	0.918	0.925	0.918	0.916

TABLE II. RESULT OF PRESENTED MODEL FOR G3

Model	phase	Index values			
		Accuracy	***Precision***	***Recall***	***F1_score***
GPC	Train	0.883	0.889	0.883	0.881
	Test	0.862	0.865	0.862	0.858
	All	0.877	0.882	0.877	0.874
GPCO	Train	0.901	0.905	0.901	0.899
	Test	0.851	0.852	0.851	0.848
	All	0.886	0.887	0.886	0.884
GPPV	Train	0.905	0.909	0.905	0.904
	Test	0.877	0.882	0.877	0.874
	All	0.897	0.900	0.897	0.896

Table III presents a comparative analysis of the accuracy results from existing studies alongside the findings of the present work. The focus of this comparison is to highlight the advancements in predictive modeling accuracy achieved in the current study compared to previous research. In previous studies, various predictive models were employed, including Decision Tree Classification (DTC) and Naive Bayes Classification (NBC). Kabakchieva [34] utilized DTC, achieving an accuracy of 72.74%. Similarly, Bichkar and R. R. Kabra [35] employed DTC and reported an accuracy of 69.94%. Nguyen and Peter [36] also used DTC, but with a significantly higher accuracy of 82%. On the other hand, Edin Osmanbegovic et al. [37] implemented NBC, achieving an accuracy of 76.65%. The present study introduces a novel approach using the Gaussian Process Classification model optimized with the Population-based Vortex Search Algorithm (GPPV) to predict student performance in language courses during the G2 and G3 exams. The results demonstrate a substantial improvement in predictive accuracy. For G2, the GPPV model achieved an accuracy of 91.8%, while for G3, the accuracy was 89.7%. These findings indicate a significant enhancement in prediction accuracy compared to the models used in prior studies. The GPPV model's superior performance can be attributed to the advanced optimization techniques incorporated, which likely contribute to its higher precision in categorizing students' performance levels. In summary, the present study's use of the GPPV model represents a notable advancement in the field of educational predictive analytics. The increased accuracy rates for both G2 and G3 exams underscore the model's potential to more effectively assess and anticipate student performance, thus providing valuable insights for educational institutions aiming to improve academic outcomes.

TABLE III. COMPARING RESULTS OF EXISTING STUDIES AND PRESENT WORK

Author (s)	Models	Accuracy
Kabakchieva [34]	DTC	72.74%
Bichkar and R. R. Kabra [35]	DTC	69.94%
Nguyen and Peter [36]	DTC	82%
Edin Osmanbegovic et al. [37]	NBC	76.65%
Present study for G2	GPPV	0.918
Present study for G3	GPPV	0.897

Tables IV and V comprehensively present the values corresponding to Precision, Recall, and F1-score indices, serving as evaluative metrics for the classification performance of the developed models concerning distinct student categories in both G2 and G3 assessments. In the ensuing analysis, a meticulous examination of Precision values elucidates nuanced distinctions among the models across performance categories. In the Excellent group of G2, GPCO and GPC models exhibit comparable performances, whereas the GPPV model surpasses both, achieving a Precision value of 0.91. Conversely, in the Good and Poor groups, the GPC model demonstrates superior performance. Notably, in the Acceptable group, GPCO and GPPV, with a Precision of 0.78, outperform the singular GPC model. In the context of the G3 test, all three models demonstrate optimal Precision in the Good group with a maximum value of 1, while their performances converge at 0.84 in the Excellent group.

Furthermore, discerning comparisons between Acceptable and Poor categories reveal the GPPV model's superior performance. Upon evaluating Recall and F1-score, the GPPV model consistently outperforms its counterparts in both G2 and G3 predictions. Moreover, a comprehensive comparison involving Recall, F1-score, and Precision collectively substantiates the superior performance of the GPPV model across G2 and G3 assessments in contrast to other models.

The crux of the confusion matrix lies in its fundamental principle: accurately predicted instances align along the main diagonal, while misclassifications diverge from this central axis. In Fig. 4, particularly within the context of G2, a meticulous examination reveals misclassifications in the GPPV model, totaling 53 instances, compared to 60 in the GPCO model and a relatively higher count of 67 in the single GPC model. Acknowledged as the superior performer, the GPPV model exhibits the fewest misclassifications, thus boasting superior predictive accuracy. Likewise, in G3, the GPPV model excels with 67 misclassifications, surpassing the GPCO model with 74 and the single GPC model with 80 misclassifications. This consistent pattern underscores the efficacy of the GPPV model, validating its superior performance in minimizing misclassifications and enhancing predictive accuracy across both G2 and G3 scenarios. Fig. 5 provides a visual representation of the confusion matrix, offering a clear portrayal of the accuracy of each model. Through meticulous analysis and comparison, the GPPV model emerges as the frontrunner, demonstrating its reliability and efficacy in predictive modeling tasks.

TABLE IV. EVALUATION INDEXES OF THE DEVELOPED MODELS' PERFORMANCE BASED ON GRADES IN G2

Model	**Grade**	**Index values**		
		Precision	***Recall***	***F1_score***
GPC	Excellent	0.86	0.94	0.90
	Good	1.00	0.37	0.54
	Acceptable	0.69	0.88	0.78
	Poor	0.97	0.97	0.97
GPCO	Excellent	0.85	0.95	0.89
	Good	0.97	0.57	0.72
	Acceptable	0.78	0.85	0.81
	Poor	0.97	0.97	0.97
GPPV	Excellent	0.91	0.91	0.91
	Good	1.00	0.62	0.76
	Acceptable	0.78	0.92	0.85
	Poor	0.96	0.98	0.97

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

Model	**Grade**	**Index values**		
		Precision	***Recall***	***F1_score***
GPC	Excellent	0.84	0.84	0.84
	Good	1.00	0.66	0.79
	Acceptable	0.74	0.77	0.75
	Poor	0.93	0.99	0.96
GPCO	Excellent	0.84	0.83	0.83
	Good	1.00	0.73	0.85
	Acceptable	0.79	0.79	0.79
	Poor	0.92	0.99	0.96
GPPV	Excellent	0.84	88.96	0.86
	Good	1.00	0.73	0.85
	Acceptable	0.79	0.79	0.79
	Poor	0.95	0.99	0.97

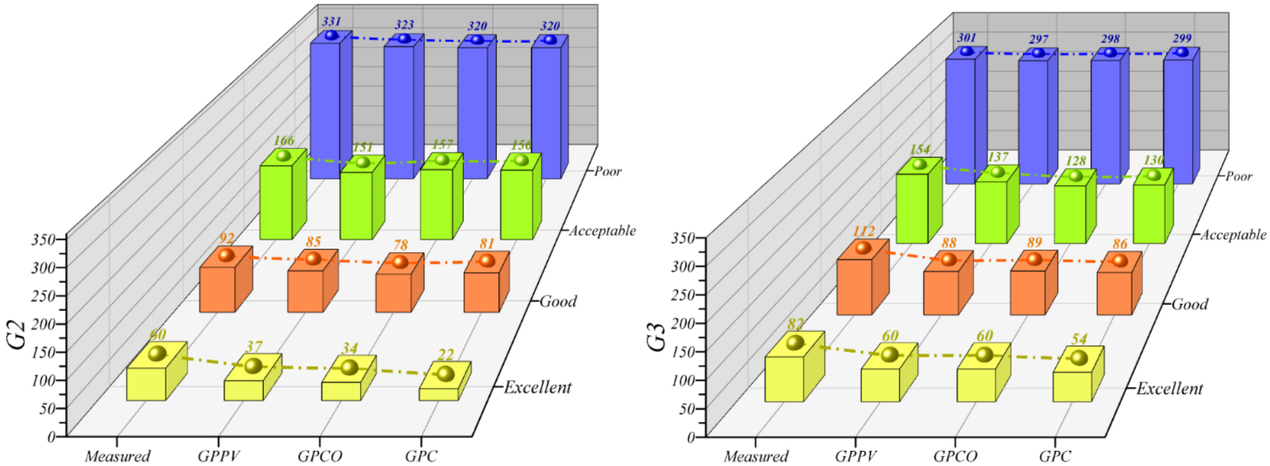

Fig. 3. 3D Bar plot for the comparison between the measured and predicted values.

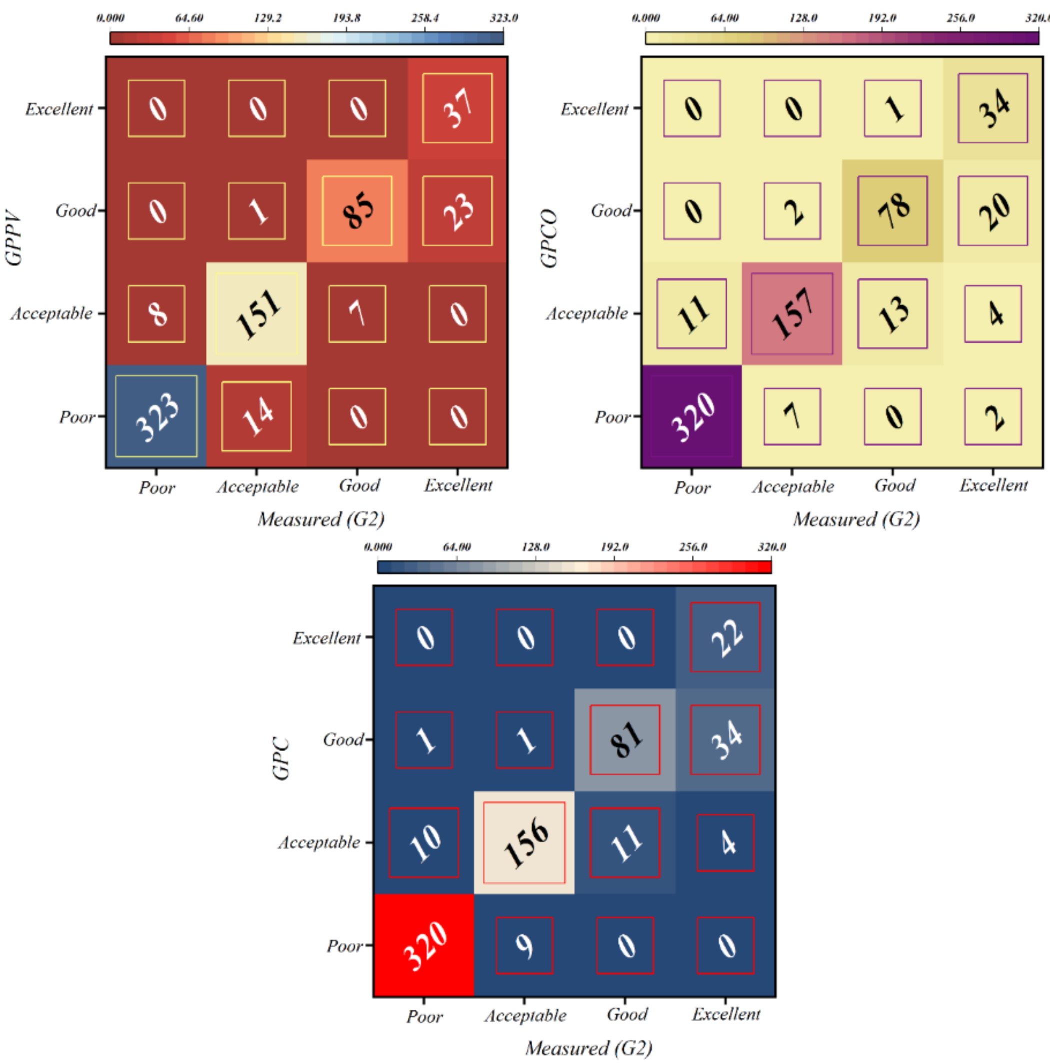

Fig. 4. Confusion matrix for each model's accuracy.

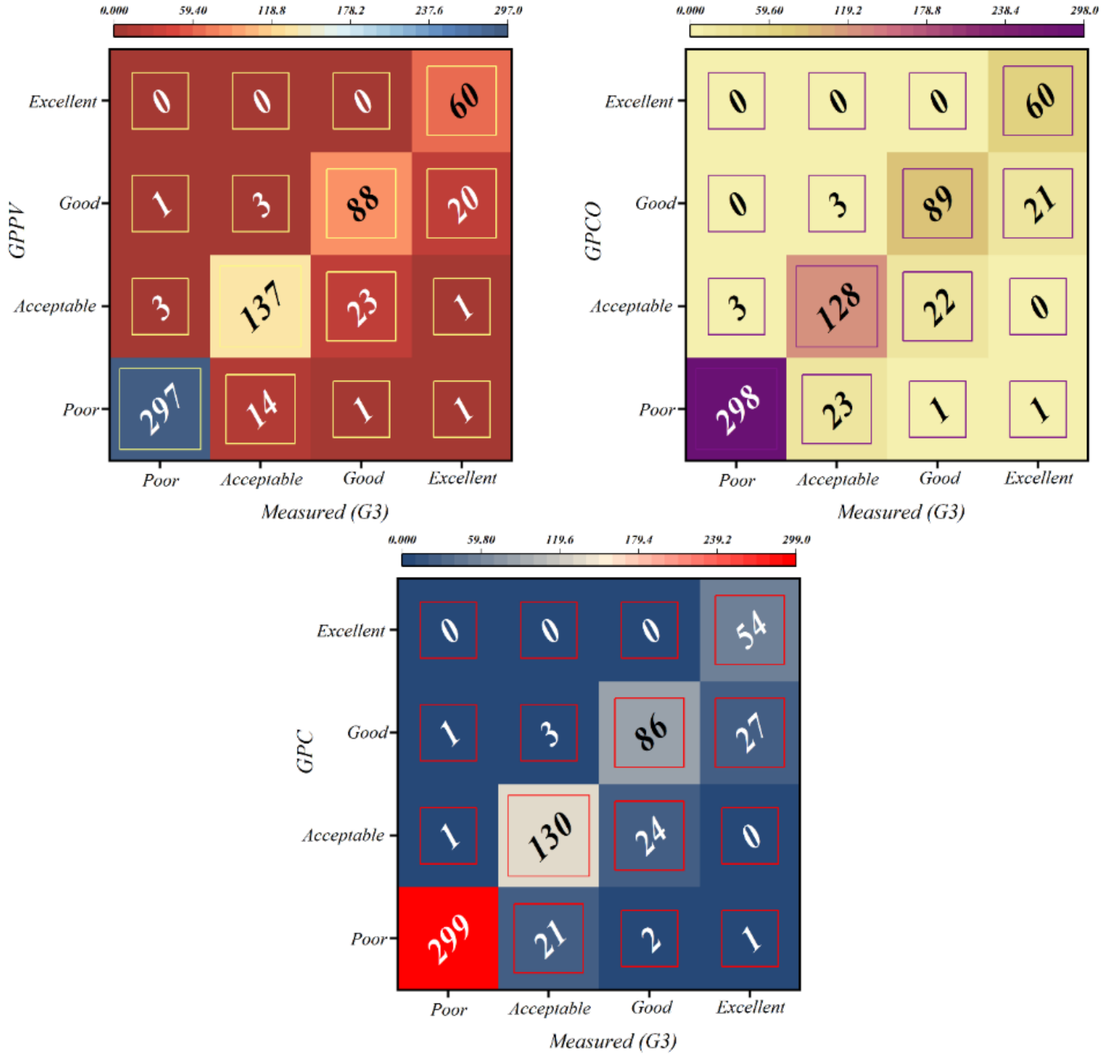

Fig. 5. Confusion matrix for each model's accuracy.

The depiction of model performance in Fig. 6, presented as a 3D Scatter plot mapping achievement percentages based on evaluative metrics, offers a nuanced understanding of the developed models. Through its volumetric representation, where higher cube numbers correspond to superior model performance, it underscores the thoroughness of the assessment. When comparing G2 and G3, the discernible prominence of the GPPV model above others signifies its superior performance. This superiority is further evidenced by the GPPV model's achievement of the highest precision predictions in both G2 (0.9245) and G3 (0.9002). The visual insights gleaned from the 3D Scatter plot not only accentuate the consistent prominence of the GPPV model but also affirm its superior Precision, thereby cementing its position as the optimal model for attaining accurate and reliable predictions across a spectrum of evaluation metrics and academic contexts. In essence, the 3D Scatter plot serves as a powerful tool for visually dissecting and comprehending the intricate nuances of model performance. Its depiction of achievement percentages based on evaluative metrics provides researchers with a holistic view, enabling them to discern patterns and trends that may not be immediately apparent through other means. This aids in making informed decisions regarding the selection and refinement of models, ultimately contributing to the advancement of predictive modeling in academic and research domains.

Fig. 7 showcases the Receiver Operating Characteristic (ROC) curve, a pivotal tool meticulously delineated to evaluate the superior GPPV model with clarity. This visually immersive representation enables a nuanced comprehension of the model's performance across various thresholds. The Area Under the ROC Curve (AUC), a widely recognized metric, serves as a comprehensive gauge of predictive accuracy and classification efficacy for the GPPV model, with a perfect test achieving an AUC of 1. Upon comparing G2 and G3, the Excellent group dominates both predictions, achieving an AUC of 1 and boasting the largest area under the curve. Following closely, the Poor group exhibits a slightly smaller difference in G3 and a marginally larger gap in G2. In G2, subsequent rankings feature the Good model followed by the Acceptable model, whereas in G3, the positions of the Acceptable and Good models interchange, with the former claiming the third rank and the latter securing the fourth position.The ROC curve's depiction elucidates the model's discriminative ability across varying thresholds, offering valuable insights into its performance characteristics. By analyzing the AUC metric, researchers can assess the model's overall predictive power and its ability to distinguish between classes. This comprehensive evaluation aids in informed decision-making regarding model selection and refinement, contributing to enhanced predictive modeling outcomes in diverse applications and research endeavors.

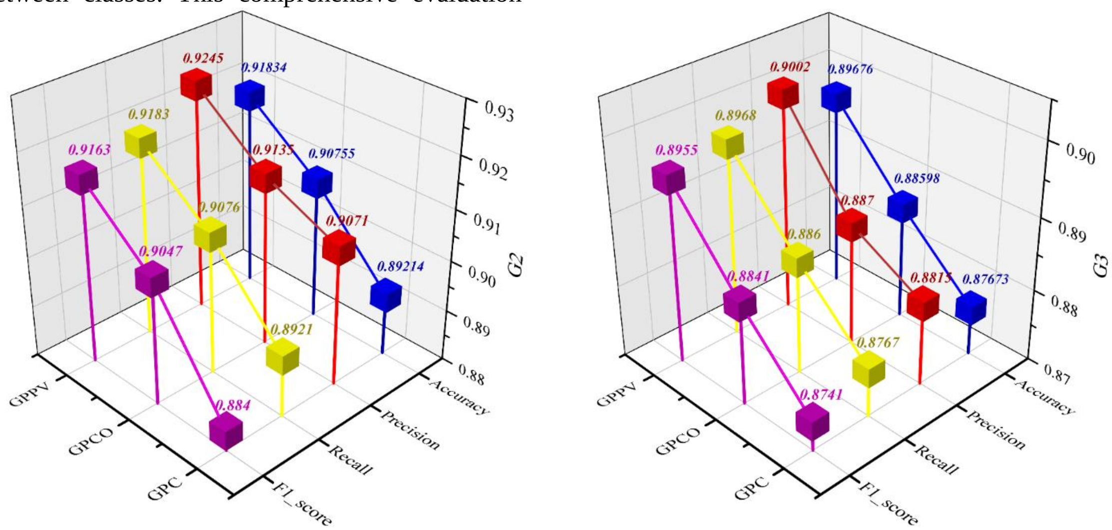

Fig. 6. 3D Scatter plots the percentage of achievement for developed models based on evaluators.

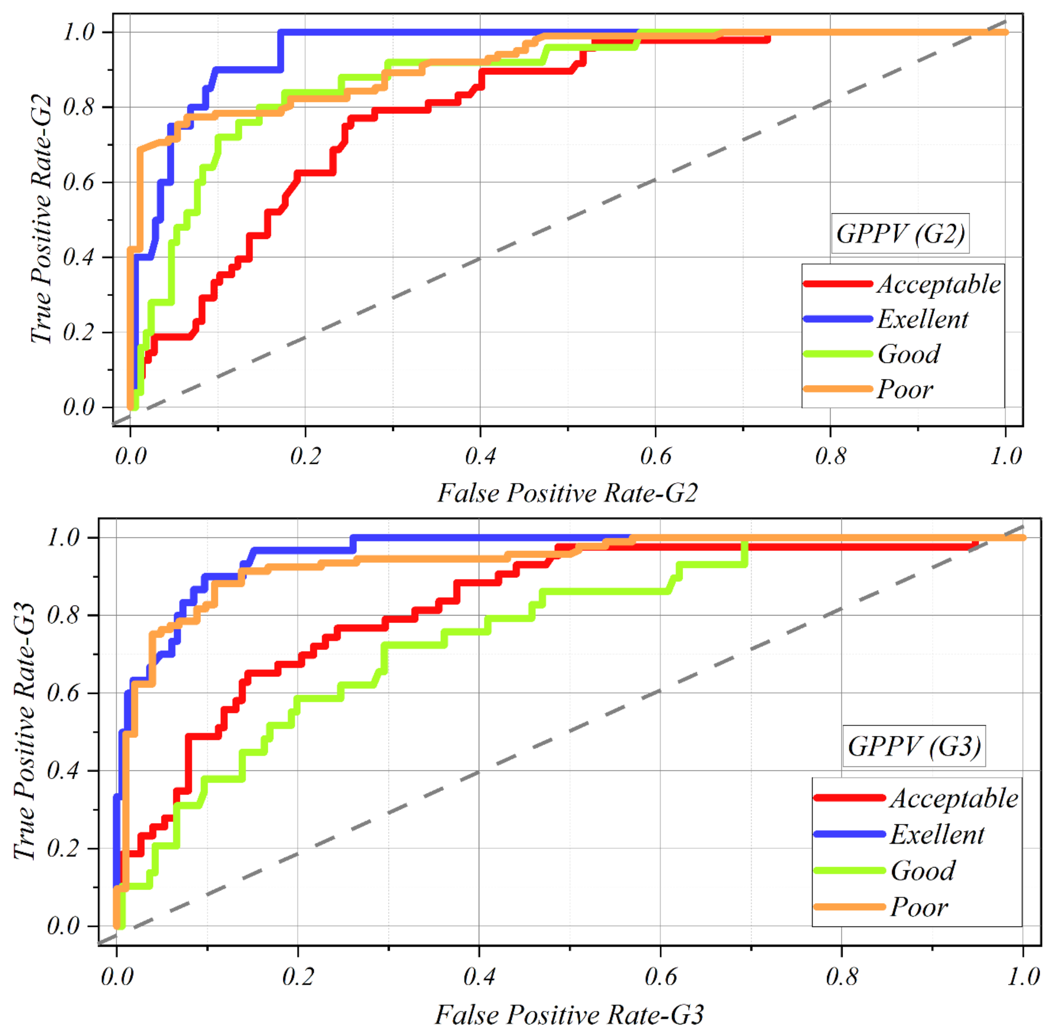

Fig. 7. The result of the ROC curve validation.

The validation of this study represents a pivotal step forward in affirming the efficacy and applicability of advanced predictive models within the realm of educational assessment and intervention. At the core of this validation process lies a meticulous examination of the methodologies employed. The deliberate integration of the Gaussian Process Classification (GPC) model with the Population-based Vortex Search Algorithm (PVS) and the COOT Optimization Algorithm (COA) was predicated on their demonstrated capabilities in handling non-linear relationships and navigating complex parameter spaces. This methodological selection was not arbitrary but grounded in empirical evidence and theoretical underpinnings, ensuring that the models were robustly equipped to address the multifaceted nature of educational data. The validation extends beyond the theoretical realm into practical implementation. Through rigorous testing and evaluation, the study demonstrated the tangible impact of these models in real-world educational settings. By accurately identifying students in need of additional support and facilitating targeted interventions, the integrated approach showcased its ability to significantly enhance educational outcomes. Furthermore, the proactive nature of these interventions, particularly in addressing elevated failure rates, underscores the practical relevance and urgency of this research. By actively mitigating academic challenges and fostering holistic student development, the study exemplifies a paradigm shift towards more personalized and effective educational practices. Moreover, the incorporation of advanced optimization algorithms such as PVS and COA not only enhances the predictive power of the GPC model but also highlights the potential of cutting-edge computational methodologies in driving innovation within the educational landscape.

In conclusion, the validation of this study serves as a testament to the transformative potential of advanced predictive models in shaping the future of educational assessment and intervention, paving the way for continued exploration and advancement in this vital field.

X. Conclusion

In the pursuit of advancing academic excellence and refining educational practices, this research emphasizes the instrumental role played by data mining and classification algorithms, specifically focusing on Gaussian Process models, in deciphering and foreseeing student performance in language courses. Diverging from conventional methodologies, this study introduces an inventive approach that integrates meta-heuristic optimization algorithms, notably the Population-based Vortex Search and COOT Optimization Algorithms (PVS and COA). These optimizers led to elevating the Precision and accuracy of student performance models, contributing a novel dimension to the existing body of literature.

The extensive evaluation, encompassing vital metrics such as Accuracy, Precision, Recall, and F1-score, illuminates the considerable potential of these meta-heuristic algorithms in refining classification outcomes. Moreover, the stratification of 649 students based on their final grades exposes the superior performance of the GPPV model, showcasing a remarkable capacity to accurately categorize the majority of students (596 correct in G2 and 582 in G3), in contrast to the comparatively lower correct classifications by GPCO and GPC. Beyond contributing to the existing knowledge base, this study offers valuable insights for educators and institutions striving to optimize educational processes, foster academic success, and thereby advance societal development and progress.

Despite its innovative approach, this study has several limitations. It focuses solely on language courses, which may limit the generalizability of its findings to other subjects. The dataset, while extensive, may not fully represent the diversity of student populations, potentially affecting the model's applicability in different educational contexts. Additionally, the reliance on historical data means the models may not adapt well to future changes in curricula or teaching methodologies. The meta-heuristic algorithms, although improving accuracy and precision, still leave room for errors, as not all student performance variability can be captured. Furthermore, external factors such as socio-economic conditions, psychological well-being, and classroom dynamics, which can significantly impact student performance, are not accounted for in the models. Finally, the computational complexity of the optimization algorithms may pose practical challenges for their implementation in real-world educational settings.

References

[1] B. P. Woolf, H. C. Lane, V. K. Chaudhri, and J. L. Kolodner, “AI grand challenges for education,” AI Mag, vol. 34, no. 4, pp. 66–84, 2013.

[2] A. S. Drigas and R.-E. Ioannidou, “A review on artificial intelligence in special education,” Information Systems, E-learning, and Knowledge Management Research: 4th World Summit on the Knowledge Society, WSKS 2011, Mykonos, Greece, September 21-23, 2011. Revised Selected Papers 4, pp. 385–391, 2013.

[3] M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart, “Predicting MOOC dropout over weeks using machine learning methods,” in Proceedings of the EMNLP 2014 workshop on analysis of large scale social interaction in MOOCs, 2014, pp. 60–65.

[4] C. Cortes and V. Vapnik, “Support-vector networks,” Mach Learn, vol. 20, pp. 273–297, 1995.

[5] J. Chen, J. Feng, X. Sun, N. Wu, Z. Yang, and S. Chen, “MOOC dropout prediction using a hybrid algorithm based on decision tree and extreme learning machine,” Math Probl Eng, vol. 2019, 2019.

[6] M. Sweeney, H. Rangwala, J. Lester, and A. Johri, “Next-term student performance prediction: A recommender systems approach,” arXiv preprint arXiv:1604.01840, 2016.

[7] M. Sweeney, J. Lester, and H. Rangwala, “Next-term student grade prediction,” in 2015 IEEE International Conference on Big Data (Big Data), IEEE, 2015, pp. 970–975.

[8] Z. Ren, X. Ning, A. S. Lan, and H. Rangwala, “Grade Prediction Based on Cumulative Knowledge and Co-Taken Courses.,” International Educational Data Mining Society, 2019.

[9] S. Morsy and G. Karypis, “Cumulative knowledge-based regression models for next-term grade prediction,” in Proceedings of the 2017 SIAM International Conference on Data Mining, SIAM, 2017, pp. 552–560.

[10] S. M. Saqlain et al., “Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines,” Knowl Inf Syst, vol. 58, pp. 139–167, 2019.

[11] P. M. Arsad and N. Buniyamin, “Prediction of engineering students’ academic performance using Artificial Neural Network and Linear Regression: A comparison,” in 2013 IEEE 5th Conference on Engineering Education (ICEED), IEEE, 2013, pp. 43–48.

[12] Q. Hu and H. Rangwala, “Academic performance estimation with attention-based graph convolutional networks,” arXiv preprint arXiv:2001.00632, 2019.

[13] J. Whitehill, K. Mohan, D. Seaton, Y. Rosen, and D. Tingley, “Delving deeper into MOOC student dropout prediction,” arXiv preprint arXiv:1702.06404, 2017.

[14] W. Feng, J. Tang, and T. X. Liu, “Understanding dropouts in MOOCs,” in Proceedings of the AAAI Conference on Artificial Intelligence, 2019, pp. 517–524.

[15] M. Fei and D.-Y. Yeung, “Temporal models for predicting student dropout in massive open online courses,” in 2015 IEEE international conference on data mining workshop (ICDMW), IEEE, 2015, pp. 256–263.

[16] I. A. Basheer and M. Hajmeer, “Artificial neural networks: fundamentals, computing, design, and application,” J Microbiol Methods, vol. 43, no. 1, pp. 3–31, 2000.

[17] T. L. de ANDRADE, S. J. Rigo, and J. L. V. Barbosa, “Active Methodology, Educational Data Mining and Learning Analytics: A Systematic Mapping Study.,” Informatics in Education, vol. 20, no. 2, 2021.

[18] C. Romero and S. Ventura, “Educational data mining and learning analytics: An updated survey,” Wiley Interdiscip Rev Data Min Knowl Discov, vol. 10, no. 3, p. e1355, 2020.

[19] L. C. Liñán and Á. A. J. Pérez, “Educational Data Mining and Learning Analytics: differences, similarities, and time evolution,” RUSC. Universities and Knowledge Society Journal, vol. 12, no. 3, pp. 98–112, 2015.

[20] A. M. Serban and J. Luan, Knowledge management: Building a competitive advantage in higher education. Jossey-Bass, 2002.

[21] C. M. Antons and E. N. Maltz, “Expanding the role of institutional research at small private universities: A case study in enrollment management using data mining,” New directions for institutional research, vol. 2006, no. 131, pp. 69–81, 2006.

[22] J. Luan, “Data mining applications in higher education,” SPSS Executive, vol. 7, 2004.

[23] Y. Ma, B. Liu, C. K. Wong, P. S. Yu, and S. M. Lee, “Targeting the right students using data mining,” in Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, 2000, pp. 457–464.

[24] C. DeLong, P. Radclie, and L. Gorny, “Recruiting for retention: Using data mining and machine learning to leverage the admissions process for improved freshman retention,” in Proc. of the Nat. Symposium on Student Retention, 2007.

[25] A. Nandeshwar and S. Chaudhari, “Enrollment prediction models using data mining,” Retrieved January, vol. 10, p. 2010, 2009.

[26] G. W. Dekker, M. Pechenizkiy, and J. M. Vleeshouwers, “Predicting students drop out: A case study,” in Proceedings of the 2nd International Conference on Educational Data Mining, EDM 2009, July 1-3, 2009. Cordoba, Spain, 2009, pp. 41–50.

[27] C. E. Rasmussen and C. K. I. Williams, Gaussian processes for machine learning, vol. 1. Springer, 2006.

[28] 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.

[29] J. Quinonero-Candela and C. E. Rasmussen, “A unifying view of sparse approximate Gaussian process regression,” The Journal of Machine Learning Research, vol. 6, pp. 1939–1959, 2005.

[30] T. P. Minka, “A family of algorithms for approximate Bayesian inference.” Massachusetts Institute of Technology, 2001.

[31] I. Naruei and F. Keynia, “A new optimization method based on COOT bird natural life model,” Expert Syst Appl, vol. 183, p. 115352, 2021.

[32] B. Doğan and T. Ölmez, “A new metaheuristic for numerical function optimization: Vortex Search algorithm,” Inf Sci (N Y), vol. 293, pp. 125–145, 2015.

[33] 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.

[34] 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.

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

[36] 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.

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