Hybrid Machine Learning Framework for Type 2 Diabetes Prediction Using Metaheuristic Optimization Algorithms

Naiyue Zhang[1], Ying Liu[2,*] and Zheng Zhang[3]
[1]Department of Computer Application Engineering, Hebei Software Institute, Baoding 071000, China
[2]Department of Internet Commerce, Hebei Software Institute, Baoding 071000, China
[3]BeiJing HuaDe Eye Hospital, Beijing 100020, China
E-mail: chanyu13579@163.com
[*]Corresponding Author

Keywords: diabetes, machine learning**,** gaussian process classification, henry gas solubility enhancement schemes (HGSO), and metaheuristic algorithms

The general basis of diabetes prediction using machine learning involves the application of algorithms that take an overall look at multiple features like BMI and glucose levels, age, genetic predispositions, and other conditions that may predict the likelihood of developing diabetes. The data-driven schemes, such as neural networks or DTs, find patterns in past data and use these to provide reliable predictions about future diabetes cases. These schemes keep learning and improving; they grow with new inputs. ML now helps in early detection by the use of large datasets, thus enabling early actions such as lifestyle changes or medical therapies. Finally, it enhances healthcare by providing individualized risk assessment and thus enables timely actions to diminish the burden of diabetes. In addition, the application of ML schemes, including Gaussian Process Classification-GPC, Linear Discriminant Analysis-LDA with Henry Gas Solubility Optimization-HGSO, Chaos Game Optimization-CGO, and Chef-Based Enhancement scheme-CBOA, has greatly benefited the process of prediction. These schemes were combined with optimizers, guided by the objective of this work, which deals with predicting the type of diabetes and the diagnosis of persons vulnerable to it. This was a strategic fusion aimed at creating new hybrid schemes with increased precision in prediction. Further analysis showed that the GPCB model was the best, with an impressive 0.981 during training. By contrast, the GPCG and GPHG schemes are relatively less accurate, with an accuracy of 0.963 and 0.946, respectively. These results justify the utility of the integrated approach, where advanced ML algorithms were able to generate predictive schemes superior in terms of accuracy and efficiency compared to the classical methods.

Povzetek: V članku je opisan sistem za napovedovanje sladkorne bolezni tipa 2 s pomočjo strojnega učenja. Algoritem GPCB združuje klasifikacijo Gaussovega procesa z metahevrističnimi optimizacijskimi algoritmi za kvalitetno diagnozo.

1 Introduction

Type 2 diabetes, another name for diabetes, is a long-term metabolic illness marked by elevated blood glucose levels because of either the pancreas's insufficient production of insulin or its inability to utilize insulin effectively [1]. Insulin is a hormone produced by the pancreas that regulates blood sugar levels by allowing the absorption of glucose into the cells to use it as energy [2]. Whenever the mechanism is disturbed, glucose builds up in the circulation and causes hyperglycemia [3]. Diabetes is sorted into 3 types: type I, type II, and gestational diabetes [4]. Type 1 diabetes, which is frequently diagnosed in childhood or adolescence, is caused by the immune system erroneously targeting and killing the insulin-generating beta cells in the pancreas [5]. This involves lifetime insulin treatment to control blood sugar levels. Type 2 diabetes, the most prevalent kind, usually develops in adulthood and is often associated with overweight, lack of exercise, and genetic risk [6], [7]. Type 2 diabetes develops when the body becomes resistant to or cannot produce sufficient insulin to meet its needs, thereby resulting in high blood sugar levels [8]. Gestational diabetes develops during pregnancy when fluctuations in hormones compromise insulin activity, increasing the risk of complications for both mother and child [9], [10]. Diabetes' persistent High blood sugar levels can cause a stream of issues affecting many organ systems [11]. These include cardiovascular disorders including strokes and heart attacks; nerve damage; diabetic neuropathy; kidney damage causing numbness, tingling (diabetic nephropathy), and discomfort; as well as eye disturbances that can cause blindness due to diabetic retinopathy, if not addressed [12], [13]. Diabetes also raises the risk of ulcers in feet and amputations owing to impaired circulation and damage to nerves [14].

Management includes frequent testing of blood glucose, proper nutrition, regular physical activity, and insulin therapy or medication when necessary. Other treatments for type 2 diabetes include weight loss and smoking cessation. People with diabetes need training and support, as enabling them with skills for optimum self-

management reduces complications, reflecting a collaborative approach by all involved [15]: providers of healthcare, the patient, and family members [16], [17]. Type 2 is a complex metabolic condition that casts ripples in personal life since it has myriad implications for many facets [18], [19]. It presents physically as a constellation of symptoms that include chronic thirst and frequent urination, fatigue, and unexpected weight gain or loss [20], [21]. This chronic fight against blood sugar becoming normal turns out to be an everyday obsession with food intake, medication routines, and even social interactions [22]. Besides the physical discomforts, type 2 diabetes also has a great psychological and emotional impact. The constant monitoring required to manage the disease can lead to feelings of anxiety, stress, and depression. The fear of complications is huge, with every increase or decrease in glucose triggering a snowball effect of questions about what this could mean for long-term health and well-being.

Type 2 diabetes can negatively affect social relationships and interactions. Even going out for meals may become a maze of counting carbohydrates and administering insulin, while social events may become distressing in their demand to explain dietary restrictions or personally withdraw to check blood glucose levels [23]. The stigma associated with diabetes can also make people feel isolated or humiliated, disrupting interpersonal interactions [24]. Besides that, type 2 diabetes may lead to serious financial burdens. Pharmaceutical treatment, apparatus for blood glucose monitoring, and frequent medical consultations are not cheap, especially when insurance coverage is inadequate. Further, loss of working days due to poor health or visiting doctors may affect earnings and professional development [25]. Notwithstanding such constraints, persons with type 2 diabetes often show remarkable resilience and resourcefulness [26]. Most learn to manage the complexity of their disease through education, proactive self-management, and support networks and feel empowered by taking responsibility for their health. However, the pervasive nature of type 2 diabetes ensures its impacts are felt at all levels of life, making comprehensive approaches to prevention, treatment, and care of utmost importance.

Machine learning algorithms can predict the risk a person has for diabetes and even define which type of diabetes the person is most probable to get, considering his or her medical history, life style habits, biomarkers, and genetic trends. These algorithms are trained on large datasets consisting of data from diabetic and non-diabetic patients through a method called supervised learning. The computers learn to find, through patterns and links in data, small signs and risk factors associated with different types of diabetes [27]. For example, ML schemes for the diagnosis of type 2 diabetes consider age, BMI, family medical history of diabetes, cholesterol levels, blood pressure, and glucose tolerance. These combined indicators may, therefore, enable the model to project the likelihood of a person developing type 2 diabetes over a specific period [28]. Other ML methods, including DT, LR, and SVM, might also classify individuals into types of diabetes based on sets of different variables. This will enable individual risk assessments and prevention methods based on an individual profile, and in time, will allow healthcare professionals to offer more personalized and effective preventative treatment [29].

1.1 Objectives

This article proposes developing a scheme for diagnosing types of diabetes and predicting the likelihood of a person being affected with it. In order to solve this issue, the use of ML schemes including LDA and GPC is chosen, along with 3 optimizers: CGO, HGSO, and CBOA. The integration of these optimizers with the schemes leads to some new hybrid model generation, which is supposed to give better performance in the prediction process. Further, these newly designed hybrid schemes are evaluated for their performances using different plots and tables. It is expected that through their dense analysis, information about the most effective performance of the different schemes can be extracted, along with the potential deficit in functionality among them. Such an inclusive strategy will provide thorough knowledge about various schemes' strengths and flaws that help in formulating approaches related to the diagnosis and prediction of diabetes.

Gaussian Process Classification (GPC) and Linear Discriminant Analysis (LDA) were picked owing to their complimentary capabilities in modeling classification challenges. GPC is a non-parametric, probabilistic model that captures complicated, nonlinear interactions and offers uncertainty estimates, making it suited for the nuanced and high-risk nature of diabetes prediction. Conversely, LDA is a basic yet powerful linear classifier that performs well when class distributions are nearly Gaussian. Its interpretability and minimal computing cost make it suitable for baseline comparison. LDA is good for efficiency and understanding, while GPC is good for making strong, adaptable models of complicated health data patterns. Together, they make a balanced framework.

2 Material and methods

2.1 Data collection

Prior to model training, the dataset underwent several preprocessing procedures to enhance data quality and model performance. Missing values were addressed using mean imputation for numerical features. Outliers were detected and mitigated using z-score normalization. All continuous features were standardized to zero mean and unit variance. Categorical variables, if any, were encoded using one-hot encoding. Feature selection was conducted using mutual information to retain only the most relevant predictors. The final dataset was randomly shuffled and split into training and testing sets using an 70:30ratio to ensure unbiased model evaluation. Fig. 1 displays the far-reaching consequences of diabetes on a person's life, spanning blood pressure to pregnancy, as it affects an

individual's well-being and lifestyle in general. This study tries to make meaning out of the interaction of diabetes with these major determinants, therefore, basically determining the trend of the illness.

- High blood pressure worsens diabetes complications by essentially destroying blood vessels and organs. High blood pressure and atherosclerosis accelerate the narrowing of arteries, which limits blood flow, thereby worsening the common diabetes consequences of heart disease, stroke, and kidney failure. Hypertension further increases the risk for diabetic retinopathy, which can cause visual impairment or even total blindness. It also leads to peripheral artery disease, which raises the chances of foot ulcers and amputations in diabetic patients. Good management of blood pressure through lifestyle modifications, medication, and regular checks is of utmost importance in effective management and reduction of adverse effects of diabetes on general health. Pregnancy complicates the care of diabetes because of fluctuating hormonal changes and increased insulin resistance.
- Gestational diabetes may be developed during pregnancy, increasing the risk for complications in both mother and child, including macrosomia, preeclampsia, and anomalies at birth. Women with previous diabetes have difficulties managing blood sugar levels, again increasing risks for adverse outcomes such as preterm birth and cesarean section delivery. Close monitoring, dietary modification, and medication may be necessary to achieve appropriate risk reduction and optimal health for both mother and fetus. Such cooperation between obstetricians, endocrinologists, and diabetes educators forms the very foundation for the best pregnancy outcomes among women with diabetes.

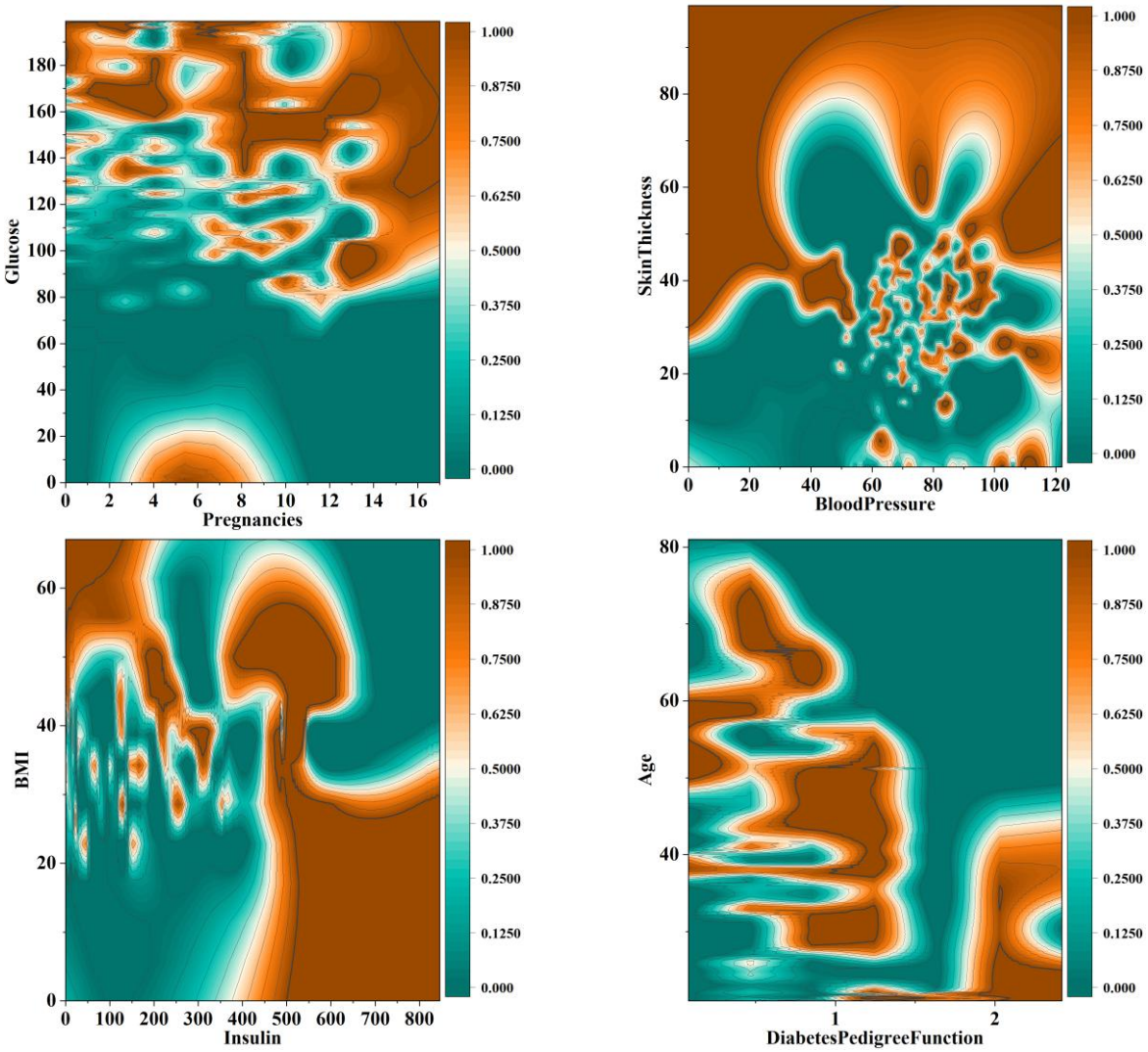


Figure 1: The plot illustrating the Contour - color fill between the input and output

2.2 Linear discriminant analysis (LDA)

Linear Discriminant Analysis (LDA) is a statistical approach used to separate two or more classes by identifying a linear combination of characteristics that best differentiates them. It assumes that the different classes create data based on Gaussian distributions with the same covariance matrix. LDA is computationally efficient, interpretable, and particularly successful when the relationship between features and labels is nearly linear, making it suited for baseline comparison in medical classification problems like diabetes prediction.

LDA assumes that the 2 categories' matrices of covariance are similar [30], and one of the 2 categories has a greater average than the other, as seized $\mu_1 < \mu_2$. One of these examples is the one provided for $x \in R$ classes:

$$\Sigma_1 = \Sigma_2 = \Sigma. \tag{1}$$

$$\begin{aligned} &\frac{1}{\sqrt{(2\pi)^d|\Sigma|}} exp\left(-\frac{(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)}{2}\right)\pi_1 \\ &= \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} exp\left(-\frac{(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)}{2}\right)\pi_2, \\ &\Rightarrow exp\left(-\frac{(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)}{2}\right)\pi_1 \\ &= exp\left(-\frac{(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)}{2}\right)\pi_2, \\ &\overset{(a)}{\Rightarrow} -\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \ln(\pi_1) \\ &= -\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2) + \ln(\pi_2) \end{aligned} \tag{2}$$

The simple logarithm of the equation's sides is found by (a). The equation may be written as:

$$\begin{aligned} &(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) = (x^T - \\ &\mu_1^T)\Sigma^{-1}(x-\mu_1) = x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_1 - \\ &\mu_1^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1 \overset{(a)}{=} x^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1 - \\ &2\mu_1^T\Sigma^{-1}x \end{aligned} \tag{3}$$

Where (a) is because $x^T\Sigma^{-1}\mu_1 = x^T\Sigma^{-1}x$ since Σ^{-1} is balanced and $\Sigma^{-T} = \Sigma^{-1}$. As a result, it is observed:

$$\begin{aligned} &-\frac{1}{2}x^T\Sigma^{-1}x - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \mu_1^T\Sigma^{-1}x + \ln(\pi_1) \\ &= \frac{1}{2}x^T\Sigma^{-1}x - \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 + \mu_2^T\Sigma^{-1}x + \ln(\pi_2) \end{aligned} \tag{4}$$

As an outcome of multiplying both sides of the equation by 2, the expression that follows is obtained:

$$\begin{aligned} &2(\Sigma^{-1}(\mu_2-\mu_1))^T x + (\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2) \\ &+ 2\ln\left(\frac{\pi_2}{\pi_1}\right) = 0 \end{aligned} \tag{5}$$

The equation of a line may be represented as $a^T x + b = 0$. T As a result, if the Gaussian distributions of the 2 classes are considered, and the covariance matrices are considered to be equal, a line displays the categorization choice border. This approach is called LDA because the choice border between the 2 classes is linear. The expressions were relocated to the correct side, which related to the second class, to create Eq. (5). Therefore, if used $\delta(x)$: $\mathbb{R}^d \to \mathbb{R}$ as the left-hand side calculation (function) in Eq. (6).

$$(x) := 2\left(\Sigma^{-1}(\mu_2-\mu_1)\right)^T x \tag{6}$$

$$+(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)) + 2\ln\left(\frac{\pi_2}{\pi_1}\right)$$

An instance x's intended class is:

$$\hat{C}(x) = \begin{cases} 1, if\ (x) < 0, \\ 2, if\ \delta(x) > 0. \end{cases} \tag{7}$$

When both categories have identical priors, $\pi_1 = \pi_2$, Eq. (5) takes a particular form:

$$\begin{aligned} &2\left(\Sigma^{-1}(\mu_2-\mu_1)\right)^T x + (\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2) \\ &= 0, \end{aligned} \tag{8}$$

Whose statement on the left can be interpreted as $\delta(x)$ in Eq. (7).

2.3 Gaussian process classification (GPC)

Gaussian Process Classification (GPC) puts a Gaussian process prior over a latent function to predict the chance of being in a certain class. This lets GPC capture nonlinear patterns in big datasets in a flexible way and measure how uncertain predictions are, which is very important for medical diagnostics. GPC is better for risk-sensitive predictions like figuring out how likely someone is to have diabetes since it changes its complexity dependent on the input. This is different from fixed parametric models.

Given a set of N training input points, in typical classification using Gaussian methods, procedure $X = [x_1, \ldots, x_N]^T$ and their associated class designations $Y = [Y_1, \ldots, Y_N]^T$, one would like to forecast the class participation percentage of a fresh test point $x_\times$. This may be accomplished by utilizing a latent function f, which is then mapped onto the [0; 1] interval utilizing the probit operator. For binary classification, use the notion that y belongs to {0,1}, where 1 displays the positive class and 0 displays the negative. Therefore, the likelihood of class membership $p(y = 1|x)$ might be expressed as Φ(f(x)), where Φ(.) is the probit purpose. Gaussian procedure classification is then performed by applying a GP prior to the latent function of $f(x)$. A GP [31] is a random procedure completely described by a mean function $m(x) = \mathbb{E}[f(x)]$ and a positive definite covariance method $\mathbb{k}(x; \acute{x}) = \mathbb{v}[f(x); f(\acute{x})]$. To project an additional test point $x_\times$, first calculate the range of the related latent variable $f_\times$.

$$p(f_\times|x_\times, X, y) = \int p(f_\times|x_\times, X, f)\, p(f|X, y)df \tag{9}$$

Where $f = [f_1, \ldots, f_N]^T$, and then using this distribution, calculate the class participation distribution:

$$\begin{aligned} &p(y_\times = 1|x_\times, X, y) \\ &= \int \Phi(f_\times)\, p(f_\times|x_\times, X, y)df_\times \end{aligned} \tag{10}$$

2.4 HGSO

The following subsection describes the motivation for HGSO, which depends on the act of Henry's law.

2.4.1 Henry's Law

In 1803, William Henry created Henry's Law, a gas law [32]. Henry's law reads as follows: "At a temperature that remains constant, the amount of a given gas that dissolves

in a given type and volume of liquid is inversely related to the partial pressure that exists for that gas in equilibrium with that liquid." Consequently, Henry's law is greatly dependent on temperature [33] and displays that a gas's solubility (Sg) is directly proportional to its relative pressure (Pg), as represented in the subsequent equation:

$$S_g = H \times P_g \tag{11}$$

Where H is Henry's stable, which is particular to the given gas-solvent mixture at a certain temperature, and P_g is the gas's relative pressure.

$$\frac{dlnH}{l(1/T)} = \frac{-\nabla sol^E}{R} \tag{12}$$

Furthermore, the impact of temperature dependency on Henry's law variables has to be addressed. The Van't Hoff equation describes how Henry's law constants vary when a system's temperature varies:

$$H(T) = \exp(B/T) \times A \tag{13}$$

Where H is an expression of 2 parameters, A as well as B, which are the 2 factors that determine H's T dependency. In addition, one can generate a function based on H at the standard temperature $T = 298.15K$.

$$H(T) = H^{\theta} \times exp\left(\frac{-\nabla sol^E}{R}\left(1/T - 1/T^{\theta}\right)\right) \tag{14}$$

The Van't Hoff formula applies if $-\nabla sol^E$ is a stable, hence Eq. (14) may be rewritten as follows:

$$H(T) = exp\left(-c \times \left(1/T - 1/T^{\theta}\right) \times H^{\theta}\right) \tag{15}$$

2.4.2 HGSO mathematical scheme

This part describes the mathematical formulas for the suggested HGSO method. The mathematical procedures are outlined below:

Step 1: Initialization process.

The count of gases (population size N) and the placements of gases have been set up using the subsequent equation:

$$X_i(t+1) = X_{min} + r \times (X_{max} - X_{min}) \tag{16}$$

where t is the repetition time, X_{min} and X_{max} are the issue bounds, r is a random number between 0 and 1, and X_i is the location of the ith gas in population N. The below equation is used to establish the count of gasses i, Henry's constant of type j ($H_j(t)$) partial pressure $P_{i,j}$ of gas i in cluster j, and $-\nabla sol^E/R$steady value of type j (C_j).

$$H_j(t) = l_1 \times rand(0,1), P_{i,j} = l_2 \times rand\ (0,1), C_j = l_3 \times rand(0,1) \tag{17}$$

where l_1, l_2, and l_3 are designated as constants with corresponding amounts of $5E-02$, 100, and $1E-02$.

Step 2: Clustering.

In proportion to the count of gas types, the entire number of agents is split into equal clusters. Every cluster has the same Henry's constant measurement (H_j) since they all contain the same gases.

Step 3: Evaluation.

The gas having the largest equilibrium state among the others of its sort is identified by analyzing each cluster j. The optimal gas for the entire colony is then determined by rating the gasses.

Step 4: Update Henry's coefficient.

Eq. (18), which updates Henry's factor, is as follows:

$$H_j(t+1) = H_j(t) \times exp\left(-C_j \times \left(\frac{1}{T(t)} - \frac{1}{T^{\theta}}\right)\right), T(t) = \exp(-t/iter) \tag{18}$$

T displays the temperature, T^{θ} displays a constant equal to 298.15, iter is the overall count of cycles, and H_j is Henry's factor for cluster j in this equation.

Step 5: Update solubility.

The following formula is used to modify the solubility:

$$S_{i,j}(t) = K \times H_j(t+1) \times P_{i,j}(t) \tag{19}$$

$S_{i,j}$ is the soluble content of gas i in cluster j, $P_{i,j}$ is the amount of partial pressure on gas i in cluster j, and K is a value that is constant.

Step 6: Update position.

The position was revised below:

$$X_{i,j}(t+1) = X_{i,j}(t) + F \times r \times \gamma \times \left(X_{i,best}(t) - X_{i,j}(t)\right) + F \times r \times \alpha \times \left(S_{i,j}(t) \times X_{best}(t) - X_{i,j}(t)\right) \tag{20}$$

$$\gamma = \beta \times exp\left(-\frac{F_{best}(t)+\varepsilon}{F_{i,j}(t)+\varepsilon}\right), \varepsilon = 0.05$$

Where $X_{i,j}$ displays the location of gas i in cluster j, and r and t are the random constant and cycle time, respectively. The best gas in cluster j is indicated by X_{best}, while the best gas in the entire swarm is shown by $X_{i,best}$. In addition, γ displays gas $j's$ capacity to interact with other gases in cluster i, α displays the effect of other gases on gas i in cluster j and is equal to 1, and β is a constant. The fitness of gas i in cluster j is denoted by $F_{i,j}$, whereas F_{best} displays the fitness of the best gas in the overall system. F is the flag that modifies the direction of the search agent and gives variety ($\pm$). $X_{i,best}$and X_{best} are the 2 parameters that control the exploration and exploitation capabilities. Particularly, $X_{i,best}$ displays the best gas i in cluster j, whereasX_{best} displays the best gas in the whole swarm.

Step 7: Escape from local optimum.

The purpose of this phase is to leave the local optimum. The count of worst agents N_w can be chosen and ranked using the following equation:

$$N_w = N \times (rand(c_2 - c_1) + c_1), c_1 = 0.1\ and\ c_2 = 0.2 \tag{21}$$

The count of search agents is denoted by N.

Step 8: Update the position of the worst agents.

$$G_{(i,j)} = G_{\min(i,j)} + r \times \left(G_{\max(i,j)} - G_{\min(i,j)}\right) \tag{22}$$

In Eq. (22), $G_{(i,j)}$ displays gas i's position in cluster j, r is a random integer, and $G_{\min(i,j)}$ and $G_{\max(i,j)}$ represent the problem boundaries. The steps of the process are depicted in Fig. 2.

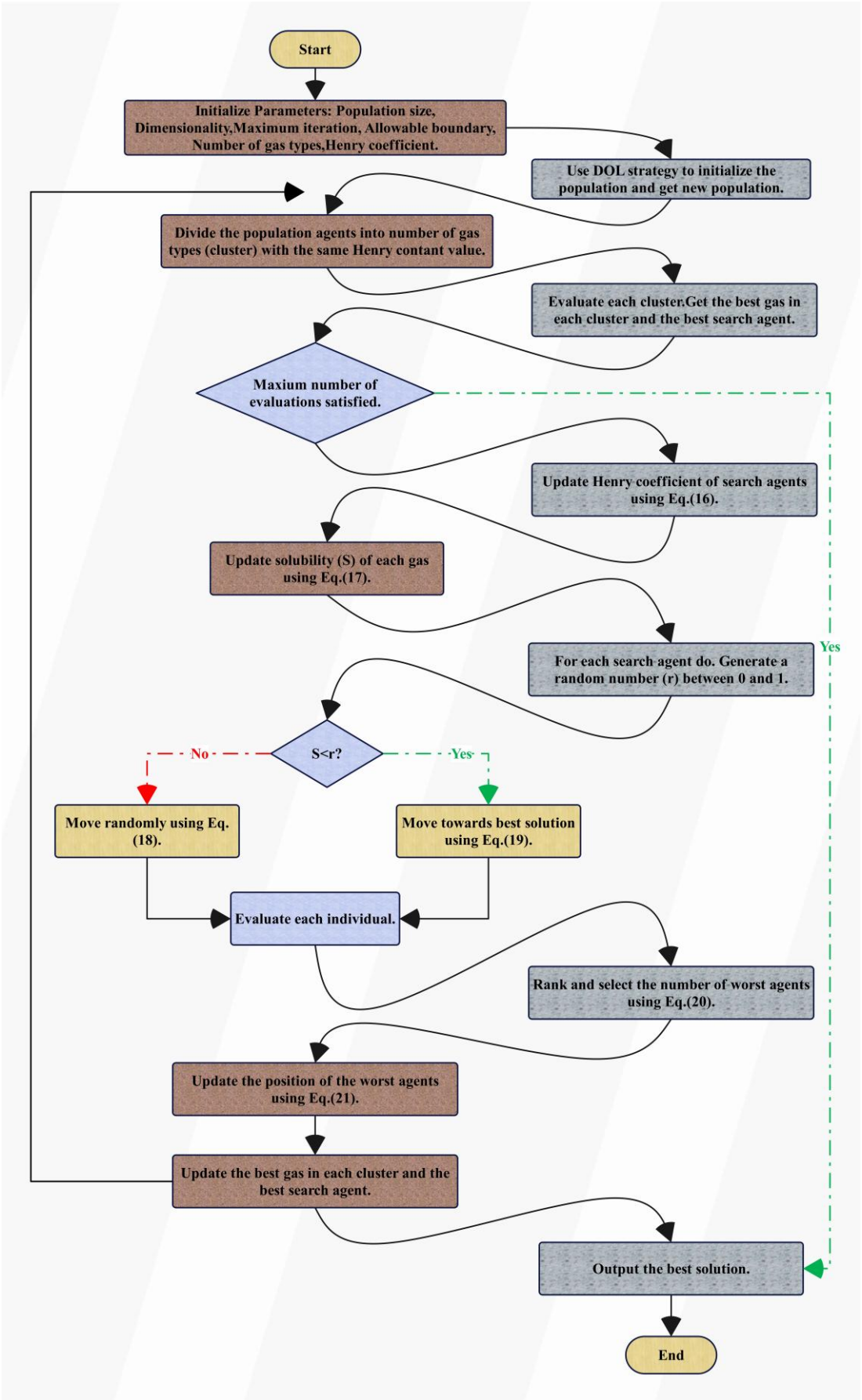


Figure 2: The flowchart of the HGS.

2.5 Chaos game optimization (CGO)

The reasons behind the groundbreaking metaheuristic algorithm known as CGO and its computational architecture are covered in this section.

2.5.1 Mathematical model

This section presents an optimization technique based on the ideas of chaos theory. The mathematical foundation of the CGO algorithm is developed based on the basic concepts of fractals and chaotic games. The CGO algorithm considers several solution candidates (X) that suggest certain able seeds within a Sierpinski triangle because many natural evolution algorithms keep an array of solutions that evolve through random modifications and selections. Each solution candidate (X_i) in this method contains a set of choice factors (x_i^j) that represent where the eligible seeds are located inside a Sierpinski triangle. The enhancement scheme uses the Sierpinski triangle to explore potential solutions. In the enhancement scheme, the Sierpinski triangle is used to look for possible solutions. The quantitative treatment of these aspects is given below:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^j & \dots & x_1^d \\ x_2^1 & x_2^2 & \dots & x_2^j & \dots & x_2^d \\ \vdots & \vdots & \dots & \vdots & \ddots & \vdots \\ x_i^1 & x_i^2 & \dots & x_i^j & \dots & x_i^d \\ \vdots & \vdots & \dots & \vdots & \ddots & \vdots \\ x_n^1 & x_n^2 & \dots & x_n^i & \dots & x_n^d \end{bmatrix}, \begin{cases} i = 1,2,..,n. \\ i = 1,2,\dots,d. \end{cases} \quad (23)$$

For each seed in the Sierpinski triangle (search area), the count of permissible seeds, or potential solutions, is n; and d is the seed's size. Random selection is used to determine where these appropriate seeds are initially placed in the search space.

$$x_i^j(0) = x_{min}^i + rand \cdot \left(x_{i,max}^j - x_{i,min}^j\right), \begin{cases} i = 1,2,\dots,n. \\ j = 1,2,\dots,d. \end{cases} \quad (24)$$

The beginning position of the eligible seeds is defined by x_i^j; $x_{i,max}^j$ as well as $x_{i,min}^j$ indicate the maximum and lowest permitted values for the ith solution candidate's jth choice variable; rand is a random integer within the range [0,1]. The way dynamical systems, often known as self-similar and self-organizing systems, behave, as was previously described, and display specific fundamental patterns serves as the foundation for the core ideas of chaos theory. The fundamental dynamical system patterns according to chaos theory are exhibited by eligible seeds, which are acquired beginning positions. It is possible to ascertain whether these seeds are suitable to function as fundamental patterns (self-similarity) for an optimization issue by employing potential solutions (X). The candidates for the solutions with the greatest and worst fitness values as well as the lowest and highest levels of eligibility are connected.

The basic idea of this mathematical model is to create the general shape of a Sierpinski triangle by producing several appropriate seeds inside the search area. In this way, fresh seeds are also produced via the Sierpinski triangle technique. An intermediate triangle with three seeds is created as follows for each appropriate seed in the search field X_i:

- Positioning of the previously identified Global Best (GB),
- The average group's location (MG_i),
- The ith resolution competitor (X_i) is the chosen seed.

Although the mean values of randomly chosen eligible seeds with an equal chance of integrating the currently regarded starting eligible seed (X_i) are reflected in the MG_i, the GB is the best solution candidate with the highest eligibility levels. Together with the identified eligible seed (X_i), the GB and MG_i create a Sierpinski triangle. In order to generate some more seeds that can be regarded as fresh eligible seeds for finishing the Sierpinski triangle, a temporary triangle is made inside the search area for each of the first eligible seeds, as was previously indicated. Four strategies are suggested to accomplish this aim. The ith permanent triangle (ith repetition) includes a Sierpinski triangle's three vertices [GB (green seed), MG_i (red seed), and X_i (blue seed)] in addition to the n appropriate seeds that were accessible in the previous cycle. This homemade triangle uses the chaotic game principle to produce fresh seeds using one die and three seeds. X_i is used to hold the first seed, GB for the second, and MG_i for the third. For the first seed, a die with three green and three red faces was utilized. Upon rolling the dice, the seed in the $X_i$is shifted to the MG_i (red face) or the GB (green face) based on the resulting color. This element is replicated using a random number generation method that generates just 2 values, 0 as well as 1, enabling the choice of red or green faces. When the green face is visible, the X_i seed advances in the direction of the GB; it moves toward the MG_i. Even if each green or red face has an equal chance of appearing in the game, the potential of getting two equivalent random integers for the GB and the MG_i is also taken into account. The direction of the X_i's seed advancement is a line segment that connects the GB with the MG_i. The flow of seeds within the search area must be restricted because of the chaotic game method; hence, this component is controlled by certain at-random factorials that were created:

$$Seed_i^1 = X_i + \alpha_i \times (\beta_i \times GB - \gamma_i \times MG_i), i = 1,2,\dots,n. \quad (25)$$

X_i displays the ith resolution candidate, GB denotes the global best discovered thus far, and MG_i displays the mean of a few selected, qualified seeds. While β_i and γ_i indicate a random integer between 0 and 1 to enable die rolling, α_i is a randomly generated factorial to reflect seed movement limitations. Three blue and three red-faced dice are used for the next seed (GB). Either the MG_i (red face) or the X_i (blue face) receives the seed in the GB, depending on the color that emerges from rolling the dice.

The model used in this section is the same as the original seed. If a blue face emerges, the seed travels to the X_i; if a red face appears, the seed goes to the MG_i. Another seed, like the first, can travel towards a location on the connecting lines between $X_i$and MG_i. This motion is restricted by randomly produced factorials.

$$Seed_1^2 = GB + \alpha_i \times (\beta_i \times X_i - \gamma_i \times MG_i), i = 1,2,\dots,n. \quad (26)$$

where each of the variables β_i and γ_i is a random value of 0 or 1 to simulate the option of rolling a die, and α_i is the randomly generated factorial for characterizing the mobility limitations of the seeds. The remaining requirements are the same as those listed for the initial seed. The third seed is employed to roll a die with green and blue faces, MG_i. The seed is directed toward either the X_i (blue face) or the GB (green face) depending on the color. An approach for generating random numbers is used to duplicate this element. It yields just 2 values, 0 and 1, so that users may select between the blue or green faces. Additionally, the lines connecting the X_i and GB can be followed by the seed. Some random factorials are also used to achieve this goal, such as:

$$Seed_1^3 = MG_i + \alpha_i \times (\beta_i \times X_i - \gamma_i \times GB), i = 1,2,\dots,n. \quad (27)$$

In order to generate the fourth seed, an additional method is employed to carry out the modification stage in the qualifying seeds' position updates within the search area. Changes in this seed's position are made depending on arbitrary adjustments made to the randomly chosen decision criteria. Eq. (28) depicts a schematic depiction of the specified procedure for the 4th seed; it has the following mathematical representation:

$$Seed_i^4 = X_i\left(x_i^k = x_i^k + R\right), \quad k = [1,2,\dots,d]. \quad (28)$$

Where k is an integer at random in the interval $[1,\ d]$ and R is a random number with uniform distribution in the region $[0, 1]$. Four formulations for αi, which controls the mobility limitations of the seeds, are provided in order to alter the exploration and exploitation rate of the CGO algorithm.

$$\alpha_i = \begin{cases} Rand \\ 2 \times Rand \\ (\delta \times Rand) + 1 \\ (\varepsilon \times Rand) + (\sim\varepsilon) \end{cases} \quad (29)$$

In this case, δ as well as ε are indeterminate numbers in the interval [0,1], and Rand is a randomly dispersed, equally distributed number in that interval. Given the self-similarity problems in the fractals, the eligibility of the new and existing seeds should be jointly assessed to decide if the additional seeds ought to be included in the search space's overall count of eligible seeds. The best new solution candidates are retained after being vetted; seeds with the lowest fitness values, or the lowest degrees of self-similarity, are removed. It is important to note that the mathematical method reduces the mathematical model's complexity by using substitution. Actually, the entire form of the Sierpinski triangle has been completed using all of the qualifying seeds found in the search region. To cope with the solution variables x_i^j breaching the boundaries of the factors, a mathematical flag is constructed. For the variables that violate the technique, a boundary change is ordered if the x_i^j is beyond the parameter's range. The most repetitions that can be done in which the optimization process takes place serves as the basis for the termination criterion.

2.6 Chef-Based Enhancement scheme (CBOA)

A metaheuristic method called CBOA was just introduced by [34]. The CBOA's mathematical representation and natural architecture are covered in this section.

2.6.1 Mathematical model of CBOA

Below is a presentation of the CBOA mathematical model using the situation from Section 2.1. First, the initialization stage of the algorithm is initiated, much like in other metaheuristics. There are 2 populations as a result of the CBOA: elite agents and candidate solutions. Therefore, as shown by Eq. (30), a matrix may be used to represent the CBOA members.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} x_{1,1} & \dots & x_{1,dim} \\ & \ddots & \\ x_{N,dim} & \dots & x_{N,dim} \end{bmatrix}_{N\times dim} \quad (30)$$

where N is the population size, dim is the issue length ($a \in [1, N], b \in [1, dim]$), X is the CBOA population matrix, and $x_{a,b}$ indicates the value of the bth problem parameter for the ath CBOA member. CBOA members' locations are established using Eq. (31):

$$x_{a,b} = LOW_b + rand \cdot (UP_b - LOW_b) \quad (31)$$

Where rand is an arbitrary number in the range of [0, 1], LOW_b and UP_b are the lower and upper limits of the bth problem factor, correspondingly. Each member's goal function may be determined and expressed as a vector according to Eq. (32):

$$Fit = \begin{bmatrix} FitxX_1 \\ \vdots \\ FitX_N \end{bmatrix}_{N\times 1} \quad (32)$$

Fit symbolizes the values of objective functions, whereas $FitX_a$ displays the value of a member. The objective function's value is used as the selection criteria for selecting the best candidate solution. The optimal member of the population and potential solution is the one that has the highest value for the objective function. It's time to complete the CBOA's processing steps after the algorithm has been launched. The CBOA is composed of two demographic groups: elite agents and candidate solutions. These two groups' update procedures are different. Its elements are changed at each cycle, and the values of the aim function are computed and evaluated. As a result, the best member is changed after each repetition. Upon comparing the values of the objective function, elite agents are selected from among the CBOA members with the highest values. The values of the goal function are used to sort the population matrix in decreasing order.

$$SX = \begin{bmatrix} SX_1 \\ \vdots \\ SX_{NC} \\ \vdots \\ SX_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} sx_{1,1} & & sx_{1,dim} \\ \vdots & & \vdots \\ sx_{NC,1} & & sx_{NC,dim} \\ sx_{NC+1,1} & \cdots & sx_{NC+1,dim} \\ \vdots & & \vdots \\ sx_{N,1} & & sx_{N,dim} \end{bmatrix}_{N\times dim} \quad (33)$$

$$SFit = \begin{bmatrix} SFitX_1 \\ \vdots \\ SFitX_{NC} \\ SFitX_{NC+1} \\ \vdots \\ SFitX_N \end{bmatrix}_{N\times 1} \quad (34)$$

Where NC is the count of chef instructors, SX denotes the sorted demographic matrix, and SFit displays the ascending objective function value vector. Following that, changes will be made in 2 steps for each group, from 1 to NC and $NC + 1\ to\ N$. NC has started to represent one-fifth of the entire population in the first group division. For instance, NC = 6 if there are 30 populations in the beginning. All cycles or the end of the epochs result in the availability of a single chef.

Step 1- Updating for chef instructors:

Chef instructors use the two best chef instructors' strategies to hone their culinary skills. At first, they try to acquire chef educator methods by imitating the best elite agent. This plan describes the global exploration and capabilities of the CBOA. The primary benefit of this upgrade is that before instructing candidate solutions, chef educators may test their skills against the best chefs. This method allows for the upgrading of candidate solutions, not only the most gifted individuals. By doing this, it prevents the algorithm from being stuck in the local optimum and promotes more precise and effective scanning over the many search space regions. In this example, freshly established cooking teacher posts are filled using Eq. (35).

$$sx_{a,b}{}^{(CFS)} = sx_{a,b} + rand \cdot \left(BestC_b - Ind \cdot sx_{a,b}\right) \quad (35)$$

$sx_{a,b}{}^{C}F\ S$ specifies the first strategy for switching chef instructors, and CFS indicates the new role for the ath-ordered member in the bth manage. The best chef instructor in the bth coordinate, or SX_1 in the SX matrix, is represented by $BestC_b$. I nd is a randomly chosen number from the set {1,2}, and rand is an arbitrary number in the interval [0,1]. Eq. (36) is used to determine this condition:

$$SX_a = \begin{cases} SX_a^{(CFS)}, SFit_a^{(CFS)} < Fit_a \\ SX_a,\ else \end{cases} \quad (36)$$

In this equation, $SFit_a^{(CFS)}$ displays the objective function of $SX_a^{(CFS)}$, and Fita is the fitness function ath member. Based on the second method, each culinary teacher strives to develop their abilities via individual practice. This method intends to increase CBOA's exploitation capabilities and local search. Every elite agents culinary expertise identifies the factors needed to get the aim function's ideal value. This updating technique is beneficial since every person searches for better opportunities in the vicinity, independent of the location of other community members. This idea is to use Eqs. (37) to (38) to produce a random position around each culinary instructor in the search space for each issue variable $b \in [1, dim]$. If this random site increases the goal function's value, it can be updated. Eqs. (39) to (40) are used to model this scenario.

$$LOW_b^{(loca)} = LOW_b^{(local)}/iter \quad (37)$$

$$UP_b^{(local)}/iter \quad (38)$$

Here, $LOW_b^{(local)}$and $UP_b^{(local)}$ show the local boundaries of the bth issue variable, where $iter$ is a parameter for repetition.

$$sx_{a,b}^{(CSS)} = sx_{a,b} + LOW_b^{(local)} + rand.\left(UP_b^{(local)} - LOW_b^{(local)}\right), j - 1, NC, J = 1, \dots, dimm \quad (39)$$

$$SX_a = \begin{cases} SX_a^{(CSS)}, SFit_a^{(CSS)} < Fit_a \\ SX_a, else \end{cases} \quad (40)$$

$SX_a^{(CSS)}$ is the new location for the ath-ranked membership according to the chef's next strategy called $C\ SS$, $sx_{a,b}^{(CSS)}$ displays its bth manage, and $SFit_a^{(CSS)}$is the goal variable value.

Step 2- candidate solutions ' updates As per the CBOA, candidate solutions pursuing culinary arts use these three methods to enhance their cooking abilities:

A chef trains each student, randomly assigning them to a class. This method has the benefit of having a chef mentor the pupils, which helps them acquire new skills. It alludes to users who have moved to the other search zone in the technique. If the best chef instructor teaches pupils, on the other hand, there won't be a worldwide search since there will be a computational bias in favor of the best. The guidance and training of the elite agent determine each culinary student's new role. This situation is expressed in Eq. (41).

$$sx_{a,b}^{(SFS)} = sx_{a,b} + rand \cdot \left(CI_{R_{a,b}} - Ind \cdot sx_{a,b}\right) \quad (41)$$

Based on the learner's initial strategy, known as SFS, the updated position for the ath-sorted member is expressed as $sx_{a,b}^{(SFS)}$, where $CI_{R_{a,b}}$ is the elite agent and R is an arbitrary index in the interval $[0, NC]$. New locations are found using Eq. (42).

$$SX_a = \begin{cases} SX_a^{SFS}, SFit_a^{(SFS)} < Fit_a \\ SX_a,\quad else \end{cases} \quad (42)$$

$SFit_a^{(SFS)}$ is the ultimate value for SFS.

The CBOA's technique involves treating every factor as a skill. Each student learns and mimics one of the chef instructor's skills. An instructor chosen at random from the collection CI_R is used (R is selected from $[1, NC]$). This is comparable to changing just one variable instead of every possible answer in terms of algorithms. This enhances global exploration and search. In order to recreate this situation, the first lead instructor, represented by the CI_{R_a} vector, is randomly selected for each culinary learner sx_a (a CBOA member selected at random from Ra's index

from $[1, NC]$). To represent a talent of the selected head instructor, the cth coordinate of the vector of sx_a, the culinary pupil, is picked at random from $[1, dim]$. CI_{R_c} is this value. In this case, Eq. (43) may be used to calculate the new location:

$$sx_{a,b}^{(SSS)} = \begin{cases} CI_{R_{a,b}}, & b = c \\ sx_{a,b}, & else \end{cases} \quad (43)$$

where b is the problem size ($[1, dim]$), a matches the population and takes a value in the range of $[NC + 1, NC + N]$, c is a random integer selected from $[1, dim]$, and SSS is the student's next strategy. Consequently, the location update is established using Eq. (44).

$$SX_{a,b} = \begin{cases} SX_a^{(SSS)}, FitS_a^{(SSS)} < Fit_a \\ SX_a, else \end{cases} \quad (44)$$

$SX_i^{(SSS)}$ relates to the new position of ath ranked member based on SSS.

Using one of the two last methods, personal activities or research, each culinary student aims to grow personally. This is the algorithm's exploitation stage. The benefit of this approach is that it makes local search stronger while also allowing the algorithm to find more practical answers that are closer to previously discovered solutions. When every obstacle is viewed as a skill, kids will work to improve these skills in order to become more fit. Thus, Eq. (45) is used to find new locations.

The selection of HGSO, CGO, and CBOA stems from their distinct abilities to enhance exploration and exploitation during model optimization critical in high-dimensional, nonlinear domains like diabetes prediction. HGSO draws on thermodynamic principles to escape local optima, improving convergence reliability. CGO leverages fractal-inspired chaotic dynamics, offering effective global search in complex spaces. CBOA mimics human learning strategies to balance global and local refinement. While these optimizers are general-purpose, their adaptability makes them suitable for fine-tuning model parameters in sensitive health-related tasks. These schemes were integrated to boost classification performance beyond what standalone models achieve. Although formal ablation studies were not conducted here, the comparative evaluation highlights clear improvements in predictive metrics, justifying their inclusion.

$$sx_{a,b}^{(STS)} = \begin{cases} sx_{a,b} + LOW_b^{(local)} + rand \cdot \left(UP_b^{(local)} - LOW_b^{(local)}\right) \\ sx_{a,b}, \ else \end{cases} \quad (45)$$

where r dim is a random number chosen from $[1, dim]$ and $sx_{a,b}^{(STS)}$ displays the updated calculated state of the ath member based on the student's third strategy (STS). Eq. (46) displays the changes:

$$SX_{a,b} = \begin{cases} SX_a^{(STS)}, & FitS_a^{(STS)} < Fit_a \\ sx_{a,b}, else \end{cases} \quad (46)$$

Fit $SX_a^{(STS)}$ displays the desired function value of $SX_a^{(STS)}$ as STS. Culinary learners and elite agents discuss $CBOA$ tactics.

2.7 Performance evaluator

A variety of indicators are utilized to assess classifier performance. The term "accuracy" refers to the proportion of accurately predicted observations. Three commonly used metrics are recall, accuracy, and precision. Total accuracy, which encompasses both real negatives and positives, is referred to as accuracy. Unbalanced datasets can lower accuracy. Recall finds only positives and assumes minimal mistakes. The F1 score is helpful in schools with different distributions since it balances recollection and accuracy. It can handle both false negatives and real positives. These measures assist in estimating the efficacy of ML schemes.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (47)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (48)$$

$$\text{Recall} = \text{TPR} = \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (49)$$

$$\text{F1 score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (50)$$

where in the further analysis the sign TP designates the case of a positive forecast of the good luck, FP - the abbreviation of fall positive - is used in the case when the outcome of a case is bad. In the case when the forecast is negative and the real result is really negative TN gives the same result. The FN means a bad forecast when the real result is good.

3 Result and discussion

The results obtained from these hybrid schemes are represented comprehensively with various graphs and tables. These tools systematically compare and contrast each model's performance for an in-depth assessment of the functions of each model. From a careful study of the results represented in the graphs and tables, insightful analysis is performed to identify the best model that performs well in terms of predictive accuracy and suitability for the prediction process. Moreover, this review also points out schemes with flaws or limits, adding a critical perspective to the work, especially in respect of their applicability to real-life scenarios. This strong assessment methodology allows researchers to make informed decisions on model selection and optimization for prediction tasks, helping to advance not only the science but also practical applications behind predictive modeling.

3.1 Convergence curve

The convergence curve has a significant influence on prediction processes since it displays the rate at which a scheme learns. A steep slope in the convergence curve displays that convergence happens fast, and hence, the

model quickly learns the pattern and forecasts stabilize. In contrast, a shallow curve indicates slower convergence, which means the model takes longer to comprehend patterns, and hence, the predictions are highly unpredictable throughout training. This helps to understand this curve for optimizing the training tactics and finding a balance between underestimating and overfitting. The suggestions made include those of learning rate changes, batch size changes, and model topology for best prediction performance with no convergence or wasted time in unnecessary training. The convergence curve in Fig. 3 illustrates and compares the results of the hybrid schemes presented. Fig. 3 displays the convergence behavior of each hybrid model across iterations, revealing learning stability and showing which schemes reach optimal accuracy most efficiently during training. It can be seen from this figure that, among the LDCB, LDCG, and LDHG schemes, the LDCG model, which has reached an accuracy of 0.930, has been outperformed by the LDCB model with 0.968 accuracy, whereas its accuracy is higher than that of the LDHG model, which stands at 0.921. Similarly, among the GPHG, GPCG, and GPCB schemes, the GPHG schemes showed an accuracy of 0.942, proving that their accuracy is the lowest compared to the GPCG model, which had an accuracy of 0.960, and the GPCB model, which had an accuracy of 0.980. Their optimal condition was achieved after 60 cycles.

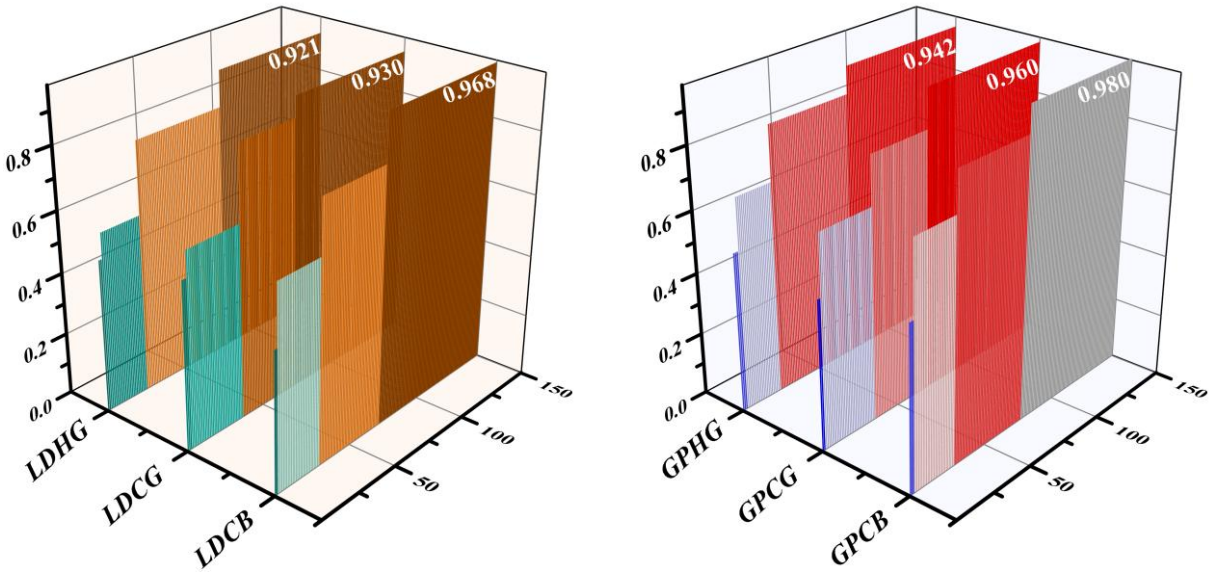


Figure 3: 3D The convergence curve for the 3 schemes

3.2 Schemes comparison

Table 1 displays the outcomes of both the LDR and GPC schemes, as well as their respective hybrid forms in different phases. Table 1 summarizes the accuracy, precision, recall, and F1-scores of all models during training, testing, and overall phases, enabling side-by-side evaluation of classifier performance. In the training phase, it becomes apparent that the functionality of the LDR model, boasting an accuracy of 0.916, falls short than another base model, GPC, achieving 0.937 accuracy in the same phase. Similarly, its hybrid counterpart, the LDHG model, with an accuracy of 0.926, also lags behind the GPHG model with 0.946 accuracy. Furthermore, the precision value of the GPCG model, reaching 0.963, outperforms the precision value of the LDCG model, which stands at 0.935, during the training phase.

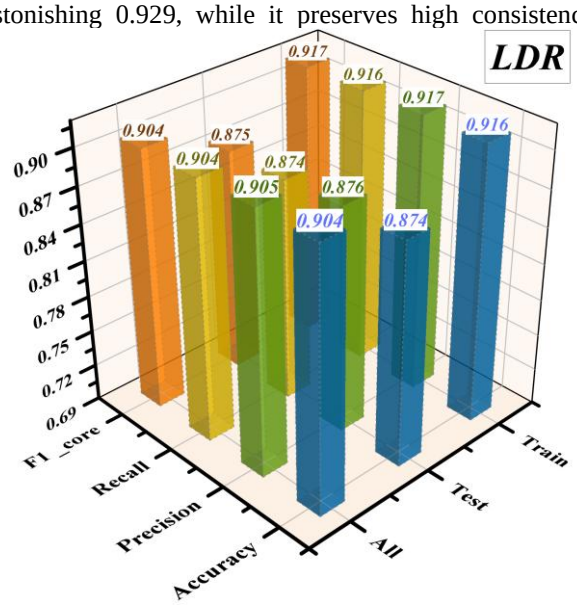
Upon comparing the outcomes of the schemes during the testing phase, it becomes apparent that the recall value of the hybrid forms of GPC schemes exceeds that of the hybrid form of the LDR model. Specifically, during the testing phase, it is evident that LDCG, with a recall value of 0.922, demonstrates weaker functionality than GPCG, which achieves a recall value of 0.957. However, following the LDCB model with a recall value of 0.961, the LDCG model boasts the highest value among its group members. Conversely, GPCG, with a recall value of 0.957, signifies that its performance surpasses that of the GPHG and GPC schemes, which have recall values of 0.935 and 0.909, in that order, although it does not outperform GPCB, with a recall value of 0.978, during the testing phase.

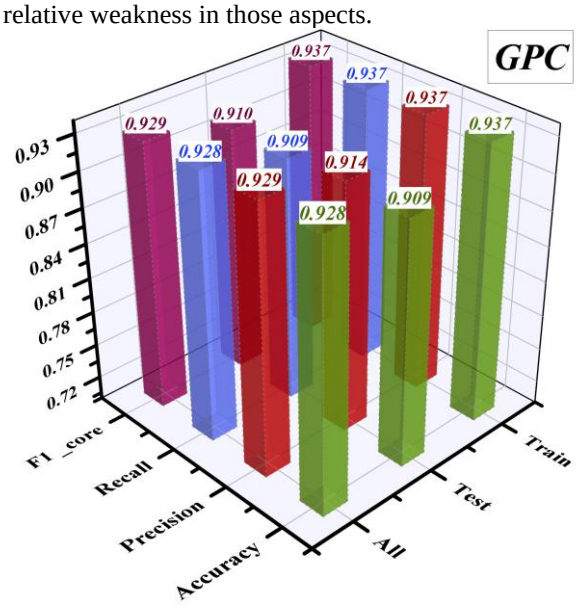
Table 1: The outcome of the showcased developed schemes

Section	Model	Metric values			
		Accuracy	Precision	Recall	F1-score
Train	LDR	0.916	0.917	0.916	0.917
	LDHG	0.926	0.925	0.926	0.925
	LDCG	0.935	0.935	0.935	0.935
	LDCB	0.972	0.972	0.972	0.972

	GPC	0.937	0.937	0.937	0.937
	GPHG	0.946	0.947	0.946	0.946
	GPCG	0.963	0.963	0.963	0.963
	GPCB	0.981	0.981	0.981	0.981
Test	LDR	0.874	0.876	0.874	0.875
	LDHG	0.913	0.913	0.913	0.913
	LDCG	0.922	0.921	0.922	0.921
	LDCB	0.961	0.961	0.961	0.961
	GPC	0.909	0.914	0.909	0.910
	GPHG	0.935	0.937	0.935	0.936
	GPCG	0.957	0.961	0.957	0.957
	GPCB	0.978	0.979	0.978	0.978
All	LDR	0.904	0.905	0.904	0.904
	LDHG	0.922	0.922	0.922	0.922
	LDCG	0.931	0.931	0.931	0.931
	LDCB	0.969	0.969	0.969	0.969
	GPC	0.928	0.929	0.928	0.929
	GPHG	0.943	0.944	0.943	0.943
	GPCG	0.961	0.962	0.961	0.961
	GPCB	0.980	0.981	0.980	0.980

The 3D wall plot of Fig. 4 visualizes model accuracy comparison across three different phases, namely Training, Testing, and All. By taking into account the performances for all the phases of three schemes, a number of thrilling trends can be found out. First and foremost, during the All phase, the LDR model performed best among them with a marvelous score of its precision metric 0.905, which really exhibits the competency of this model with a touch towards precision. With that said, GPC outcompetes all its contenders during the same stage with outstanding precision and F1 score records at an astonishing 0.929, while it preserves high consistency between its measures, which remain around 0.928 with regard to both accuracy and recall, demonstrating an overall robust behavior in performance. In sharp contrast, the LDHG model displays very consistent results in all four metrics, reaching a stable performance of 0.922 in all, reflecting a balanced performance considering different evaluation standards. In contrast, the GPHG model has strengths and weaknesses mixed up on the metrics. Although it has a very commendable score in the precision metric of 0.944, the value is low in other metrics, having 0.943 for accuracy, recall, and F1 score, showing its relative weakness in those aspects.





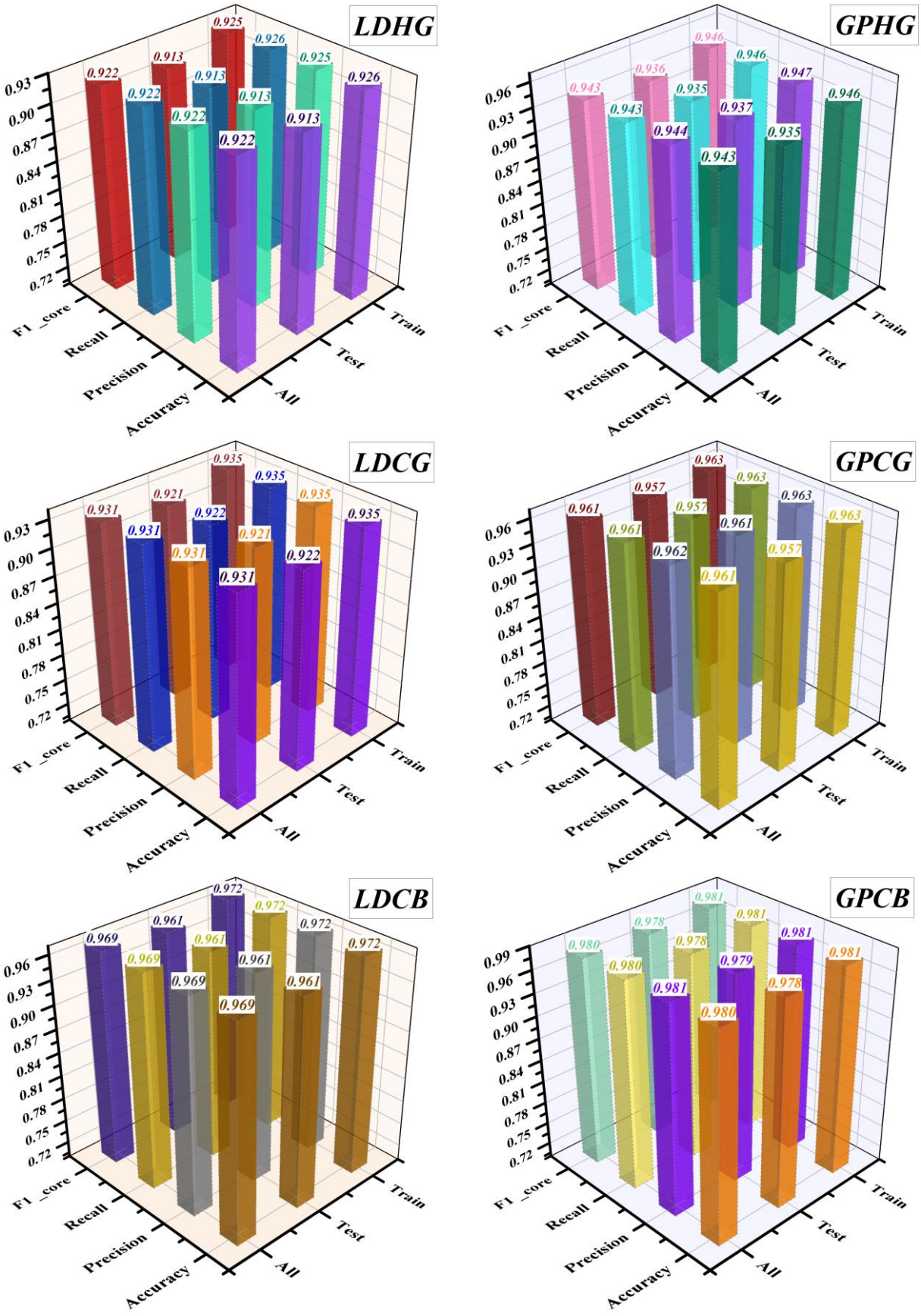


Figure 4: 3D Walls-plot for the performance of the schemes across phases

Table 2 presents a comparison of the functional performance of schemes under both healthy and diabetes conditions. For instance, the LDR model showcases an accuracy of 0.93 under healthy conditions, aligning with

the precision value of the LDHG model. However, the LDCB model emerges as the top performer with a precision value of 0.97, indicating its superiority over the LDCG model, which achieves a precision value of 0.94, as well as other preceding schemes. Among the hybrid versions of the GPC model, the GPCB and GPCG schemes emerge with the highest accuracy under healthy conditions, boasting precision values of 0.99 and 0.98, respectively. Following closely, the GPHG model achieves a precision value of 0.97, while the GPC model records a precision value of 0.95, indicating slightly weaker functionality compared to the former schemes. Nevertheless, the hybrid forms of the GPC model showcase superior functionality in contrast to the LDA scheme and its variants.

Furthermore, under diabetes conditions, the LDCB model exhibits a higher recall value of 0.95, surpassing the recall values of the LDCG, LDHG, and LDA schemes, which stand at 0.90 and 0.88, in that order. Moreover, the recall value of the LDCB model exceeds that of the GPC and GPHG schemes, which are 0.91 and 0.94, respectively. However, it falls short of surpassing the recall values of the GPCG and GPCB schemes, which are 0.96 and 0.98, respectively.

Table 2: Categorization of assessment criteria for the performance of the developed schemes

Metric values	Condition	Model							
		LDR	LDHG	LDCG	LDCB	GPC	GPHG	GPCG	GPCB
Precision	Healthy	0.93	0.93	0.94	0.97	0.95	0.97	0.98	0.99
	Diabetes	0.85	0.90	0.91	0.96	0.88	0.90	0.93	0.97
Recall	Healthy	0.92	0.95	0.95	0.98	0.94	0.94	0.96	0.98
	Diabetes	0.88	0.88	0.90	0.95	0.91	0.94	0.96	0.98
F1-score	Healthy	0.93	0.94	0.95	0.98	0.94	0.96	0.97	0.98
	Diabetes	0.86	0.89	0.90	0.95	0.90	0.92	0.95	0.97

The column line symbol plot in Fig. 5 provides a comparison between the values recorded in both healthy and diabetic situations and the values predicted by the schemes. Under the diabetes condition, it is evident that the LDCB model, with 254 out of 268 measured values, demonstrates higher accuracy than the LDCG model, which achieves 240 out of 267 measured values. Similarly, the base model, LDR, performs better with 236 out of 268 measured values compared to the LDHG model, which also achieves 236 out of 268 measured values. Conversely, under the healthy condition, both GPC and GPHG schemes achieve 468 and 471 out of 500 measured values, respectively, indicating lower accuracy compared to the GPCG and GPCB schemes, which achieve 480 and 491 out of 500 measured values, respectively. Besides, the GPCG and GPHG schemes attain values of 258/268 and 253/268, respectively, under the diabetes condition, which indicates moderate performance by the GPCB model, with attained values of 262/268, and the GPC model, at 245/268.

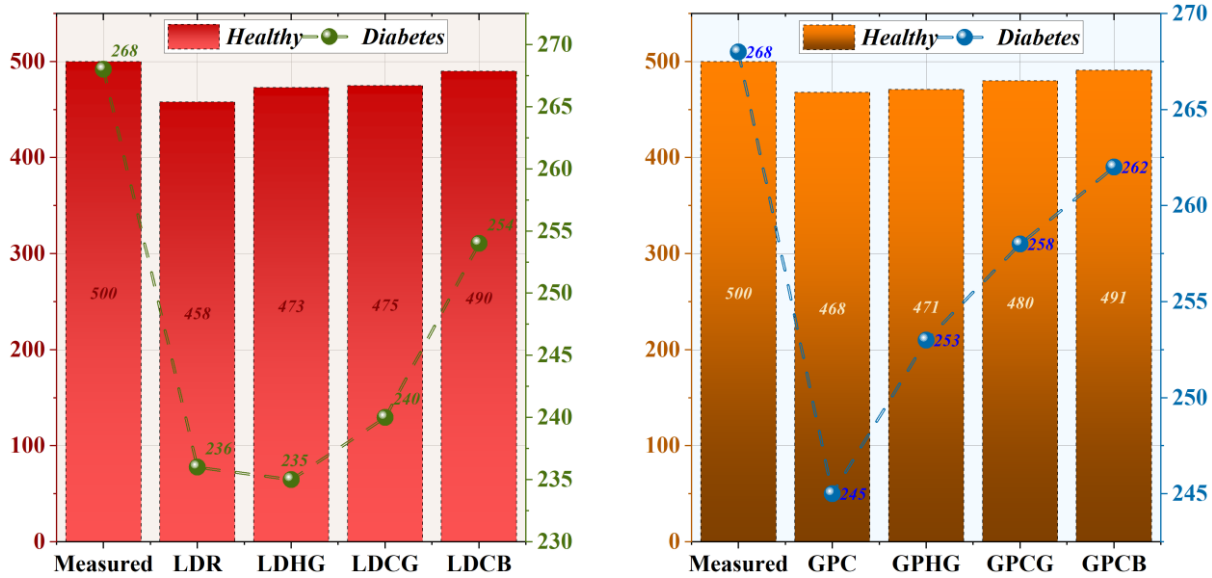


Figure 5: Column-line symbol plot to represent the difference among the schemes

To avoid overfitting, the model's performance was checked at three different phases: training, testing, and overall. Also, the fact that the training and testing measures show the same patterns means that the model is generalizing instead of overfitting. Even though there wasn't a formal validation set, the hybrid schemes' performance in all phases give us an idea of how strong they are. In the future, we will use cross-validation and explicit regularization approaches to better control overfitting and make the model more generalizable.

The ROC is a measure that fundamentally depends on how well binary classifiers work. It compares the false positive rate (1-specificity) to the true positive rate (sensitivity) at various thresholds. This graph conveys useful information about the capability of the classifier to differentiate classes in all possible threshold settings. The ROC is a tool that actually enables the researchers to study the compromise between true positives and false positives, thus giving a complete view of the efficiency of the classifier. Besides, the ROC's AUC gives a quantitative measure of the discriminatory power of a classifier, where larger AUC means better performance. Also, the ROC plot allows for better selection of the optimal cut-off value to classify the samples according to the needs of the specific application, considering sensitivity and specificity to get the same result desired. Therefore, the ROC curve displays a very important means for testing, comparing, and fine-tuning binary classification schemes, thus contributing to enhanced ML model predictive power in a slew of applications. Moreover, in Fig. 6, the outcomes of the suggested schemes are carefully analyzed with the help of the ROC curve, which is a perfect inseparable tool used to analyze the performance of the classifier. It is observed, upon detailed analysis, that GPCB and GPCG are ahead of their competitors in reaching a TPR value of 1.0 at an earlier stage and hence delivers exceptional performance in classification problems. After that, LDCB and GPHG come very close as the second and third schemes, reaching a TPR of 1.0 just a little later but with a sharp increase, further establishing their effectiveness. In sharp contrast, the LDR model lags far behind its counterparts since its vector has the gentlest slope among the compared schemes. Nevertheless, the LDR model eventually attains 1.0 TPR but takes its time in comparison with the others. The above analysis displays how different schemes may perform to the extent and also how often the ROC curve proves useful for making subtle choices regarding classifier behavior, which might not be immediately apparent in other forms, and helps drive better decisions for predictive modeling tasks.

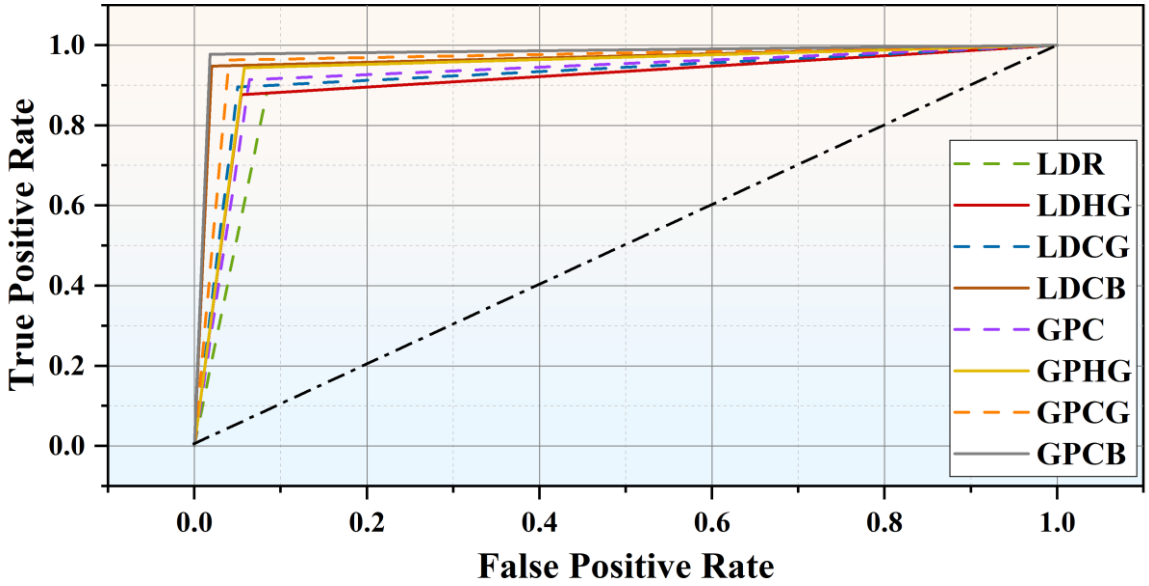


Figure 6: ROC curves depict the performance of the most efficient hybrid schemes

The SHAP additive explanations in Fig. 7 depict the effects of various factors such as glucose or BMI indicators that influence the possibility of diabetes. The following explanation succinctly defines the effects of such factors on the occurrence of diabetes.

- High levels of blood glucose, normally due to excessive consumption of sugar or reduced action of insulin, may eventually lead to the development of diabetes. Blood glucose that remains high over a continuous period places a load on the pancreas secreting insulin, and, with time, may make it lose its efficiency. This can result in insulin resistance-a condition whereby cells become unable to efficiently act in response to insulin signals, causing more accumulation of glucose. Besides, high levels of glucose can cause the damage of blood vessels and neurons, which raise the risk of complication development in diabetic patients. Hence, keeping blood glucose within the norm through proper nutrition, regular physical activity, and medication is considered a significant approach to diabetic prevention and management. BMI, which is determined using weight and height measures, is another widely accepted indicator of body fatness associated with the risk of developing diabetes.
- A high BMI means excess adipose tissue interferes with insulin action, apart from increasing the inflammatory component, leading to insulin resistance and impaired glucose tolerance. The underlying fat also secretes hormones and cytokines, further dampening metabolic processes and increasing diabetes risk. Also, a higher BMI is more often than not associated with other risk factors like sedentary lifestyle and lousy food, increasing the chances of diabetes. By enhancing insulin sensitivity and overall metabolic health, dietary and activity changes that control body mass index (BMI) can lower the risk of diabetes. Therefore, maintaining a

healthy BMI is crucial for both preventing and treating diabetes.

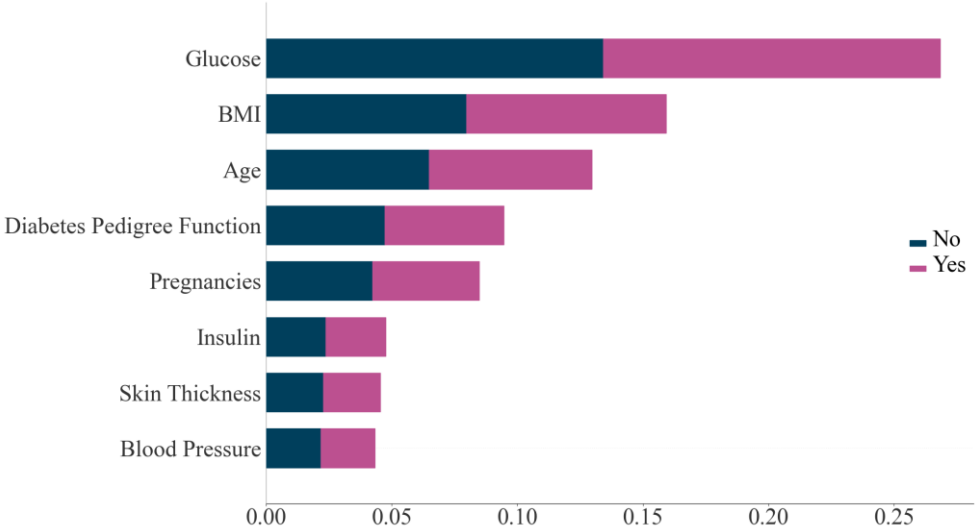


Figure 7: The sensitivity analysis results

Table 3 provides the results of a 5-fold cross-validation for the GPC and LDR models, assessing their stability and generalization across different subsets of the dataset. Each fold (K1 to K5) represents an independent split where the model was trained on 80% of the data and tested on the remaining 20%. The GPC model demonstrates consistently high performance across all folds, with accuracy values ranging from 0.916 to 0.928, indicating strong generalization and low variance. In contrast, the LDR model shows slightly lower accuracy across all folds, with values ranging from 0.887 to 0.904. The results clearly suggest that GPC outperforms LDR not only in individual experiments but also in terms of cross-validated reliability. These findings reinforce the robustness of GPC for diabetes prediction tasks under varying training-test partitions.

Table 3: K-fold cross validation.

Models	**K Fold Number**				
	K1	***K2***	***K3***	***K4***	***K5***
GPC	0.920	0.927	0.924	0.916	0.928
LDR	0.887	0.895	0.901	0.896	0.904

Table 4 presents the results of the Wilcoxon signed-rank test conducted to compare the performance differences between baseline classifiers and their hybrid optimized variants. The test evaluates whether observed differences in classification performance are statistically significant. A lower p-value (typically < 0.05) indicates a statistically meaningful improvement. Among the models, the GPCHG scheme achieved a p-value of 0.0348, indicating a statistically significant enhancement over the base GPC model. Similarly, GPCG produced a marginally significant result with a p-value of 0.0679, while others such as GPC-CBOA and LDR-based hybrids did not show statistically significant improvements, as their p-values exceeded 0.1. The stat column represents the test statistic for ranking the difference between paired models. These findings validate that only specific optimizer integrations particularly with GPC deliver meaningful predictive advantages, supporting the selective use of metaheuristics in medical classification contexts like Type 2 diabetes prediction.

Table 4: Wilcoxon test.

Models	stat	P value
GPC	644	2.25E-01
GPC Henry gas solubility optimization	338	3.48E-02
GPC chaos game Optimization	155	6.79E-02
GPC Chef-Based Optimization Algorithm	48	4.39E-01
LDR	1200	2.45E-01
LDR-Henry gas solubility optimization	824	4.39E-01
LDR-chaos game Optimization	675	6.80E-01
LDR-Chef-Based Optimization Algorithm	125	4.14E-01
GPC	644	2.25E-01
GPC-Henry gas solubility optimization	338	3.48E-02
GPC-chaos game Optimization	155	6.79E-02
GPC-Chef-Based Optimization Algorithm	48	4.39E-01

4 Conclusion

The various advantages of early detection of diabetes by using ML are: it enables early interference, thus preventing the development of complications such as cardiovascular diseases and neuropathy; ML algorithms sift through enormous volumes of data to spot patterns that are so subtle they could indicate diabetes risk, hence improving their accuracy. This will, therefore, be enabling personalized treatment plans for better patient care. Also, automating diagnostics cuts down the healthcare costs and workload for medical staff. In a nutshell, ML aims at early diabetes detection, providing an improvement for patient outcomes through easy healthcare access, thus adopting a proactive stance towards the disease's management.

However, this work aims to project diabetes using ML schemes comprising GPC and LDA, coupled with 3 optimizers: Henry Gass Solubility Optimization, Chef Base Enhancement Algorithm, and Chaos Game Optimization. With the view of improving the accuracy of the prediction, it was decided to couple the schemes with the optimizers. These results mean that the model GPC and its hybrid forms provide better performance than the LDA scheme and its hybrids. Comparing results in GPC, GPHG, GPCG, and GPCB, for instance, out of these, the best result was from the GPCB model in the "All" phase, with an accuracy value of 0.980. In that respect, the GPCG model stands out as the second-best model with an accuracy of 0.961, while the GPHG model gives medium performance in this comparison, with an accuracy of 0.943. In this comparison, the GPC model has the weakest functionality, with an accuracy of 0.928.

- **Limitations:**

There are several drawbacks to projection using ML techniques. The most critical problem of overfitting that most schemes biased the training data and gather noise rather than underlying patterns, which is poor in generalization in unknown data. When the schemes are relatively simple to represent the complexity of the data, underfitting happens with poor accuracy in the forecast. Biases in training data can persist in ML schemes, leading to biased forecasts, especially in sensitive domains like healthcare and criminal justice. Furthermore, ML algorithms need big, high-quality datasets for training, which are not always available, especially in specialist sectors or when dealing with sensitive data. The dynamic nature of real-world data makes it challenging to sustain model correctness over time; hence, regular monitoring and updating become necessary. To solve these limitations, several methods have been tried to reduce overfitting, such as regularization; feature engineering to make the schemes perform better; and algorithms that are fair-aware to reduce biases. All of the above can be further improved by enhancing openness and interpretability of schemes, thus building trust and enabling their adoption in applications of importance. This calls for more research and development on these issues so that the MLC forecasts become increasingly accurate and dependable.

References

[1] S. M. Haffner, "Epidemiology of type 2 diabetes: risk factors," *Diabetes Care*, vol. 21, no. Supplement_3, pp. C3–C6, 1998. Publisher: American Diabetes Association. https://doi.org/10.2337/diacare.21.3.C3.

[2] Y. Wu, Y. Ding, Y. Tanaka, and W. Zhang, "Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention," *Int J Med Sci*, vol. 11, no. 11, p. 1185, 2014. Publisher: National Library of Medicine. https://doi.org/10.7150/ijms.10001.

[3] E. Wilmot and I. Idris, "Early onset type 2 diabetes: risk factors, clinical impact and management," *Ther Adv Chronic Dis*, vol. 5, no. 6, pp. 234–244, 2014. Publisher: Sage Publications. https://doi.org/10.1177/2040622314548679.

[4] G. L. Robertson, "Diabetes insipidus," *Endocrinol Metab Clin North Am*, vol. 24, no. 3, pp. 549–572, 1995. Publisher: Elsevier. https://doi.org/10.1016/S0889-8529(18)30031-8.

[5] J. R. GREEN, G. C. BUCHAN, E. C. ALVORD JR, and A. G. SWANSON, "Hereditary and idiopathic types of diabetes insipidus," *Brain*, vol. 90, no. 3, pp. 707–714, 1967. Publisher: National Library of Medicine. https://doi.org/10.1093/brain/90.3.707.

[6] M. Babey, P. Kopp, and G. L. Robertson, "Familial forms of diabetes insipidus: clinical and molecular characteristics," *Nat Rev Endocrinol*, vol. 7, no. 12, pp. 701–714, 2011. Publisher: Nature. https://doi.org/10.1038/nrendo.2011.100.

[7] R. D. Lawrence, "Three types of human diabetes," *Ann Intern Med*, vol. 43, no. 6, pp. 1199–1208, 1955. Publisher: ACP. https://doi.org/10.7326/0003-4819-43-6-1199.

[8] R. D. Lawrence, "Types of human diabetes," *Br Med J*, vol. 1, no. 4703, p. 373, 1951. Publisher: National Library of Medicine. https://doi.org/10.1136/bmj.1.4703.373.

[9] A. Ota and N. P. Ulrih, "An overview of herbal products and secondary metabolites used for management of type two diabetes," *Front Pharmacol*, vol. 8, p. 224659, 2017. Publisher: frontiers. https://doi.org/10.3389/fphar.2017.00436.

[10] W. J. Pories, J. H. Mehaffey, and K. M. Staton, "The surgical treatment of type two diabetes mellitus," *Surgical Clinics*, vol. 91, no. 4, pp. 821–836, 2011. Publisher: Surgical Clinics.

[11] J. Zhang, Y. Deng, Y. Wan, J. Wang, and J. Xu, "Diabetes duration and types of diabetes treatment in data-driven clusters of patients with diabetes," *Front Endocrinol (Lausanne)*, vol. 13, p. 994836, 2022. Publisher: frontiers. https://doi.org/10.3389/fendo.2022.994836.

[12] Y. Yang and L. Chan, "Monogenic diabetes: what it teaches us on the common forms of type 1 and type 2 diabetes," *Endocr Rev*, vol. 37, no. 3, pp. 190–222, 2016. Publisher: Oxford Academic. https://doi.org/10.1210/er.2015-1116.

[13] L. S. Greci *et al.*, "Utility of HbA1c levels for diabetes case finding in hospitalized patients with hyperglycemia," *Diabetes Care*, vol. 26, no. 4, pp. 1064–1068, 2003. Publisher: American Diabetes Association. https://doi.org/10.2337/diacare.26.4.1064.

[14] K. Plis, R. Bunescu, C. Marling, J. Shubrook, and F. Schwartz, "A machine learning approach to predicting blood glucose levels for diabetes management," in *Workshops at the Twenty-Eighth AAAI conference on artificial intelligence*, 2014. Publisher: AAAI.

[15] R. E. Glasgow, "A practical model of diabetes management and education," *Diabetes Care*, vol. 18, no. 1, pp. 117–126, 1995. Publisher: American Diabetes Association. https://doi.org/10.2337/diacare.18.1.117.

[16] S. Nam, C. Chesla, N. A. Stotts, L. Kroon, and S. L. Janson, "Barriers to diabetes management: patient and provider factors," *Diabetes Res Clin Pract*, vol. 93, no. 1, pp. 1–9, 2011. Publisher: Elsevier. https://doi.org/10.1016/j.diabres.2011.02.002.

[17] P. J. Watkins, P. L. Drury, K. W. Taylor, and W. G. Oakley, *Diabetes and its management*. Wiley Online Library, 1990.

[18] S. H. Ley, O. Hamdy, V. Mohan, and F. B. Hu, "Prevention and management of type 2 diabetes: dietary components and nutritional strategies," *The Lancet*, vol. 383, no. 9933, pp. 1999–2007, 2014. Publisher: The Lancet.

[19] S. Vijan, "Type 2 diabetes," *Ann Intern Med*, vol. 152, no. 5, pp. ITC3-1, 2010. Publisher: ACP. https://doi.org/10.7326/0003-4819-152-5-201003020-01003.

[20] J. E. B. Reusch and J. E. Manson, "Management of type 2 diabetes in 2017: getting to goal," *JAMA*, vol. 317, no. 10, pp. 1015–1016, 2017. Publisher: Jama Network. DOI: 10.1001/jama.2017.0241.

[21] E. Ahmad, S. Lim, R. Lamptey, D. R. Webb, and M. J. Davies, "Type 2 diabetes," *The Lancet*, vol. 400, no. 10365, pp. 1803–1820, 2022. Publisher: The Lancet.

[22] R. A. DeFronzo *et al.*, "Type 2 diabetes mellitus," *Nat Rev Dis Primers*, vol. 1, no. 1, pp. 1–22, 2015. Publisher: Nature. https://doi.org/10.1038/nrdp.2015.19.

[23] D. H. Laursen, A. Frølich, and U. Christensen, "Patients' perception of disease and experience with type 2 diabetes patient education in D enmark," *Scand J Caring Sci*, vol. 31, no. 4, pp. 1039–1047, 2017. Publisher: Wiley Online Library. https://doi.org/10.1111/scs.12429.

[24] O. Peleg, E. Hadar, and A. Cohen, "Individuals with type 2 diabetes: an exploratory study of their experience of family relationships and coping with the illness," *Diabetes Educ*, vol. 46, no. 1, pp.

83–93, 2020. Publisher: Sage Publications. https://doi.org/10.1177/0145721719888625.

[25] E. A. Beverly, C. K. Miller, and L. A. Wray, "Spousal support and food-related behavior change in middle-aged and older adults living with type 2 diabetes," *Health Education & Behavior*, vol. 35, no. 5, pp. 707–720, 2008. Publisher: Sage Publications. https://doi.org/10.1177/1090198107299787.

[26] J. L. Browne, A. Ventura, K. Mosely, and J. Speight, "'I call it the blame and shame disease': a qualitative study about perceptions of social stigma surrounding type 2 diabetes," *BMJ Open*, vol. 3, no. 11, p. e003384, 2013. Publisher: BMJ Journals. https://doi.org/10.1136/bmjopen-2013-003384.

[27] A. Dagliati *et al.*, "Machine learning methods to predict diabetes complications," *J Diabetes Sci Technol*, vol. 12, no. 2, pp. 295–302, 2018. Publisher: Sage Publications. https://doi.org/10.1177/1932296817706375.

[28] M. Soni and S. Varma, "Diabetes prediction using machine learning techniques," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 9, pp. 921–925, 2020. Publisher: IJERT.

[29] L. Kopitar, P. Kocbek, L. Cilar, A. Sheikh, and G. Stiglic, "Early detection of type 2 diabetes mellitus using machine learning-based prediction models," *Sci Rep*, vol. 10, no. 1, p. 11981, 2020. Publisher: Nature. https://doi.org/10.1038/s41598-020-68771-x.

[30] H. Sifaou, A. Kammoun, and M.-S. Alouini, "High-dimensional linear discriminant analysis classifier for spiked covariance model," *Journal of Machine Learning Research*, vol. 21, no. 112, pp. 1–24, 2020. Publisher: JMLR. Available: https://www.jmlr.org/papers/v21/19-428.html.

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

[32] Y. Li, W. Zhang, L. Wang, F. Zhao, W. Han, and G. Chen, "Henry's law and accumulation of weak source for crust-derived helium: A case study of Weihe Basin, China," *Journal of Natural Gas Geoscience*, vol. 2, no. 5–6, pp. 333–339, 2017. Publisher: Elsevier. https://doi.org/10.1016/j.jnggs.2018.02.001.

[33] J. Staudinger and P. V Roberts, "A critical review of Henry's law constants for environmental applications," *Crit Rev Environ Sci Technol*, vol. 26, no. 3, pp. 205–297, 1996. Publisher: Taylor & Francis. https://doi.org/10.1080/10643389609388492

[34] E. Trojovská and M. Dehghani, "A new human-based metahurestic optimization method based on mimicking cooking training," *Sci Rep*, vol. 12, no. 1, p. 14861, 2022. Publisher: Nature. https://doi.org/10.1038/s41598-022-19313-2.