# Performance Optimization of Support Vector Machine with Adversarial Grasshopper Optimization for Heart Disease Diagnosis and Feature Selection

Nan Tang, Lele Wang, Kangming Li, Zhen Liu, Yanan Dai, Ji Hao, Qingdui Zhang, Huamei Sun\* , Chunmei Qi\* Department of Cardiology, The Second Affiliated Hospital of Xuzhou Medical University, Xuzhou 221000, Jiangsu, China

*Abstract***—The World Health Organization reports that cardiac disorders result in approximately 1.02 million deaths. Over the last years, heart disorders, also known as cardiovascular diseases, have significantly influenced the medical sector due to their immense global impact and high level of danger. Unfortunately, accurate prognosis of heart problems or CD, as well as continuous monitoring of the patient for 24 hours, is unattainable due to the extensive expertise and time required. The management and identification of cardiac disease pose significant challenges, particularly in impoverished or developing nations. Moreover, the absence of adequate medical attention or prompt disease management can result in the individual's demise. This study presents a novel optimization technique for diagnosing cardiac illness utilizing Support Vector Machine (SVM) and Grasshopper Optimization Algorithm (GOA). The primary objective of this approach is to identify the most impactful characteristics and enhance the efficiency of the SVM model. The GOA algorithm, which draws inspiration from the natural movements of grasshoppers, enhances the search for features in the data and effectively reduces the feature set while maintaining prediction accuracy. The initial stage involved pre-processing the ECG data, followed by its classification using several algorithms such as SVM and GOA. The findings demonstrated that the suggested approach has markedly enhanced the effectiveness and precision of heart disease diagnosis through meticulous feature selection and model optimization. This approach can serve as an efficient tool for early detection of heart disease by simplifying the process and enhancing its speed.**

#### *Keywords—Heart disease predictions; Support Vector Machine; Grasshopper Optimization Algorithm; feature selection*

#### I. INTRODUCTION

Hence, conducting early analysis of cardiovascular disease is crucial in order to mitigate its profound impact and enhance personal well-being [1,2]. An all-encompassing electronic gadget for monitoring the heart that is utilized for analyzing irregular heart rhythms. Th is equipment continually gathers human electrocardiogram (ECG) readings for a whole 24-hour period [3]. A cardiac monitoring model based on artificial intelligence (AI) was created to accurately categorize ECG signals as either regular or irregular patterns. This was achieved by training and testing the model using the standard MIT-BIH arrhythmia database [4,5], which is publically accessible on PhysioNet [6]. Electrocardiography (ECG) is a highly prevalent non-invasive diagnostic technique used for identifying various cardiac conditions, including myocardial infarction (MI) [7]. Early detection of this condition can halt its course and ultimately avert myocardial infarction. Consequently, the objective of numerous studies in this domain has consistently been to get an early diagnosis of this ailment. Furthermore, the utilization of ECG signals for diagnosis is highly significant owing to its accessibility and cost-effectiveness in comparison to the costly techniques of cardiac echocardiography and MRI [8,9]. Currently, there are various techniques for diagnosing myocardial infarction (MI), and we will provide a concise overview of a few of them. In 2019, Sugimoto et al. introduced a technique for identifying myocardial infarction (MI) using cannulation networks. The user's text is [10,11]. This approach utilizes the electrocardiogram (ECG) signal obtained from 12 leads [12,13,14]. The researchers initially constructed a convolution-based model specifically for normal ECG signals in that particular investigation. Subsequently, a computer-aided engineering model is constructed for every lead. If inputted, the model retrieves normal electrocardiogram (ECG) data. Otherwise, the output waveform will be distorted due to unsuitable data. Next, the healthy and MI data were classified by reconstructing model errors using the K-nearest neighbor (KNN) method [15]. Ultimately, the outcomes of this classification technique are documented to surpass those of other established procedures. Panagiotis Barampoti and others. In 2019, a method was proposed that utilizes ECG to identify MI using Grossman and Euclidean mapping [16,17].

Artificial intelligence, specifically deep learning, is a branch of machine learning that focuses on analyzing ECG signal structures over multiple hierarchical levels. Its goal is to address complex tasks that were challenging for standard neural network models [18]. Artificial intelligence or deep learning-based simulations for heart monitoring face difficulties in accurately classifying ECG heartbeats when they are overtrained due to the pseudo-periodic activity of the ECG signal [19,20,21]. Hence, it is imperative to utilize the quantity of samples saved in the ECG heart rate segment as a means to describe the input variables for escape training. In this research, a method for quantifying the number of peak ECG heartbeats is employed to encode the suggested input variables, as depicted in Fig. 2. The starting elements are modified to classify the ECG heartbeat into 16 disease categories, including 15 arrhythmias and 1 normal arrhythmia.

The GOA is employed in this study to enhance the efficiency and accuracy of heart disease diagnosis by selecting the most influential features for the SVM. GOA excels in feature selection by balancing exploration and exploitation, ensuring a comprehensive search of the feature space while avoiding local

minima. Inspired by the natural movements of grasshoppers, the algorithm mimics their slow, gradual movements and sudden leaps, facilitating both broad exploration and focused exploitation. This approach is particularly beneficial in handling high-dimensional data, common in heart disease diagnosis, as it effectively reduces the feature set without compromising predictive accuracy. By identifying the most significant features, GOA improves the training efficiency and performance of the SVM, resulting in faster and potentially more accurate diagnoses. The algorithm's versatility and robustness in optimization problems further validate its use, ensuring that the SVM operates with an optimized feature set, thus enhancing the overall diagnostic process.

The proposed method for heart disease diagnosis integrates the locust evolutionary algorithm with SVM, leveraging computational advancements to improve accuracy and efficiency in medical data mining. By employing the locust evolutionary algorithm for feature selection, the method optimizes the identification of relevant data attributes essential for precise diagnosis. This step is crucial in medical datasets where numerous features may be present but not all contribute significantly to diagnostic outcomes. Coupled with SVM, known for its robustness in handling complex datasets and highdimensional feature spaces, the method ensures that only the most informative features are utilized for classification. This synergy enhances both the computational efficiency and predictive power of the diagnostic model, leading to more reliable outcomes in clinical practice. The main contributions of the authors in this research are as follows:

- Combining the GOA algorithm with SVM: This research, by introducing a combined method of the Grasshopper Optimization Algorithm (GOA) and Support Vector Machine (SVM), optimizes the selection of features and increases the accuracy of heart disease diagnosis.
- Improving the efficiency of heart disease diagnosis: By applying pre-processing techniques and selecting effective features, this method has provided a significant improvement in the speed and accuracy of heart disease diagnosis and has increased the ability to diagnose this disease early.

The remainder of the paper is structured as follows. The second section contains a list of earlier works. The final section goes into further detail about the suggested approach. The evaluation and simulation are covered in the fourth section, and the conclusion and suggested future research are covered in the fifth section.

## II. RELATED WORKS

Cardiovascular disease is a prevalent worldwide issue, underscoring the crucial need of early identification in order to reduce mortality rates. Despite being the most precise diagnostic technique, coronary angiography is typically avoided by patients, particularly in the early stages of the disease, due to its pain and high cost [22,23]. Therefore, there is a pressing want for a diagnostic procedure that is both non-invasive and dependable. Machine learning has become pervasive in modern times, encompassing numerous facets of human existence and serving as a catalyst for transformative changes in the healthcare sector. Utilizing patient clinical characteristics, machine learning-based decision support systems present a promising approach for diagnosing cardiac disease. Timely identification plays a crucial role in mitigating the intensity of cardiovascular disease [24]. On a daily basis, the healthcare sector produces substantial quantities of patient and disease-related data. Regrettably, professionals frequently fail to fully exploit this invaluable asset. Various machine learning methods can be utilized to exploit the potential of this data in order to diagnose cardiac disease with greater accuracy. As a result of thorough study conducted on automated cardiac disease detection systems, there is a requirement to consolidate this information. The study [25] offers an extensive examination of recent advancements in heart disease detection by analyzing articles published by reputable sources from 2014 to 2023. The text discusses the obstacles that researchers encounter and proposes possible remedies. Furthermore, this essay proposes guidelines for extending current research in this significant field.

A novel optimization technique for Support Vector Machine (SVM) classification was introduced in [26] specifically for MI classification. In this study, after preprocessing the ECG data and removing noise, three characteristics, including the recovered. Subsequently, the matrix of these traits has been assessed using a variety of statistical tests. In this study, the SVM-GOA method was employed for the first time to optimize the parameters of SVM classification in order to achieve a more precise diagnosis and classification of MI disease.

Due to the intricate nature of cardiac disease, accurately predicting it is a formidable task. Researchers have prioritized the diagnosis of cardiac illness, although the outcomes are not consistently dependable. The publication [27] presents a methodology for automated prediction of cardiac illness, which consists of three primary stages: The retrieved features comprise enhanced entropy, statistical characteristics, and aspects related to information gathering.

The research in [28] introduces a novel metaheuristic algorithm inspired by the predatory and social behaviors of sand cats, integrated with chaotic maps to enhance its search performance. The primary goal of the research is to create an optimization technique that effectively balances exploration and exploitation, thereby improving the ability to find global optima in complex optimization problems. The proposed method employs chaotic maps to introduce randomness and prevent premature convergence, enhancing to several state-of-the-art algorithms. Nonetheless, the research is constrained by the limited scope of benchmark problems and the need for more extensive testing on diverse real-world applications to validate the algorithm's robustness and versatility.

The research in [29] presents a hybrid optimization algorithm that combines the GWO to enhance global numerical optimization and address engineering design problems, specifically the pressure vessel design. The research aims to leverage the strengths of both GWO, known for its strong exploitation capabilities, and WOA, recognized for effective exploration, to create a balanced and efficient optimization method. The hybrid approach integrates the social hierarchy and hunting strategies of grey wolves with the bubble-net attacking method of whales, aiming to improve convergence speed and solution accuracy. Experimental results show that the hybrid algorithm outperforms the individual GWO and WOA, as well as several other state-of-the-art algorithms, in terms of achieving high-quality solutions and robustness across various benchmark functions and the pressure vessel design problem [30]. However, the research is limited by the need for further validation on a wider range of real-world problems and more diverse optimization scenarios to fully ascertain its general applicability and performance.

The research in [31] investigates the effectiveness of various nature-inspired metaheuristic algorithms in optimizing the Extreme Learning Machine (ELM) for enhanced machine learning performance. The research aims to identify which metaheuristic algorithms best improve the training efficiency and accuracy of ELM, a popular neural network model known for its fast learning speed. The study evaluates multiple algorithms, including Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Differential Evolution (DE), and Ant Colony Optimization (ACO), among others, by integrating them with ELM and assessing their performance on several benchmark datasets. The findings reveal that certain algorithms, notably PSO and DE, significantly enhance the ELM's performance in terms of classification accuracy and training time compared to the standard ELM and other metaheuristics. However, the research is limited by the specific selection of benchmark datasets and the need for additional testing on more diverse and complex datasets to validate the generalizability and scalability of the optimized ELM models.

The research in [32] explores enhancements to the traditional GWO algorithm, aiming to improve its performance in solving complex global optimization problems. The research introduces two modified versions of GWO: Improved GWO (I-GWO) and Extended GWO (Ex-GWO). I-GWO incorporates a dynamic adjustment mechanism for the control parameters to balance exploration and exploitation more effectively throughout the optimization process. Ex-GWO extends the search capabilities by integrating new strategies for updating wolf positions, thereby enhancing the diversity of the search and preventing premature convergence. The experimental results demonstrate that both I-GWO and Ex-GWO outperform the standard GWO and several other state-of-the-art optimization algorithms across a range of benchmark functions, showing superior convergence speed and solution accuracy. Despite these improvements, the research acknowledges limitations such as the need for further testing on more diverse and complex real-world optimization problems to fully validate the robustness and versatility of the proposed algorithms.

#### III. PROPOSED METHOD

Fig. 1 illustrates the algorithm described in this paper. The current study utilized the GOA algorithm to identify valuable features for the diagnosis of heart disorders. Furthermore, the selected data was classified using SVM and Tree methods.



Fig. 1. Process flowchart of the proposed method.

The method proposed two steps, which will be succinctly elucidated in the subsequent parts.

*1) Step 1: Data loading, preprocessing, and normalization:* Currently, the diabetes dataset contained 768 records and 8 characteristics, which were eliminated using the outlier approach. Subsequently, the data was uploaded for the purpose of conducting pre-processing and normalization. Subsequently, the data underwent preprocessing and missing data elimination using the nearest neighbor technique, wherein Nan values were substituted with the nearest neighbor column values [33]. Normalization was used to eliminate duplicate records during data preprocessing. Data normalization can be achieved by several strategies, with MinMax being well recognized as one of the most commonly used methods. This approach allows for the conversion of any data to a certain range, while also performing normalization for each individual feature. Normalization is commonly characterized as the process of adjusting data to fit inside a specific range, such as - 1 to 1. The calculation of normalization is determined using the MinMax approach, which is based on the equation (1):

$$
Z = \frac{X - \min(x)}{(x) - \min(x)}\tag{1}
$$

Let X represent the number that has to be normalized, while  $min(x)$  and  $max(x)$  represent the smallest and greatest integers in the set, respectively.

*2) The second phase involves selecting features based on gene ontology annotation (GOA):* During this stage, the feature selection technique was employed to ascertain the crucial features that contribute to a specific outcome. In this work, feature selection was performed using GOA [34]. During the feature selection step, the utilization of this approach leads to premature convergence as a result of its inherent characteristics. Therefore, the technique has the potential to achieve maximum convergence in the final optimization step. Locusts are insects that are classified as pests because they cause significant harm to crops and agriculture.

$$
X_i = S_i + G_i + A_i \tag{2}
$$

Eq. (2) is employed to simulate the behavior of locusts [35]. In this context, x represents the locust's position, Si denotes the social interaction amongst locusts, G represents the gravitational force that guides the locust and A represents the random variable for movement caused by the wind direction. The final three show the specific location of the propeller. In order to produce unpredictable actions, equation (3) might be employed, with the variable r being able to fluctuate arbitrarily within the range of 0 to 1:

$$
S_i = S_i r_1 + G_i r_2 + A_i r_3 \tag{3}
$$

The value of  $S_i$ , which represents the target function, is determined by the rate of social contact. This rate is calculated using Eq. (4), where  $d_{ii}$  represents the distance between grasshopper *ith* and grasshopper *ith*.

$$
S_i = \sum_{i=1}^n s d_{ji} (\widehat{d_{ji}})
$$
 (4)

In Eq. (4), the variable  $d_{ii}$  represents the distance between grasshopper i and grasshopper j, and it is determined as the absolute value of the difference between the selected feature  $d_{ji} = |x_j - x_i|.$ 

$$
s(r) = f e^{\frac{-r}{l} - e^{-r}} \tag{5}
$$

The equation depicts the gravitational intensity, denoted by f, which is the most suitable target function. The length of the gravity scale is represented by l. The equation follows the general formula stated in Eq. (6).

$$
X_{i} = \sum_{i=1}^{n} s\left( |x_{j} - x_{i}| \right) \frac{x_{j} - x_{i}}{d_{ji}} \tag{6}
$$

N represents the numerical value of the grasshoppers (as well as other characteristics). Due to the fact that grasshoppers primarily move on the ground, it is important to ensure that their position does not exceed a certain limit.

#### *3) The third stage involves categorizing the chosen data.*

During this stage, the training data is subjected to training using S.V.M, G.O.A, and GOA TREE methods. Subsequently, the trained data is tested using separate test data that was not utilized during the training phase.

Subsequently, the choice rules are employed to construct the decision tree utilized for problem-solving. Ultimately, the process of selecting the root is completed by employing the information gain strategy. The Eq. (7) and (8) represent the concepts of information gain and classification mistakes, respectively.

$$
entropy (pc) = -\sum_{i=1}^{n} p\left(\frac{pc}{i}\right) log 2p\left(\frac{pc}{i}\right) \tag{7}
$$

$$
classification error (pc) = 1 - max_{ip} \left(\frac{pc}{i}\right) \tag{8}
$$

 $\left(\frac{pc}{t}\right)$  $\frac{\partial c}{\partial t}$  denotes the proportion of inputs in the diabetic disease dataset that are associated with a certain set of primary components. N represents the overall quantity of inputs inside the dataset.

#### *A. Support Vector Machine Algorithm*

SVM are a highly efficient approach for building a classifier. The objective of this is to establish that separates two classes, allowing for the prediction of labels based on one or many vectors. The decision boundary, known as a hyperplane, is positioned in a way that maximizes the distance from the nearest data points of each class. The support vectors refer to the nearest points. Given that we possess a dataset with labeled estimators:

$$
(x_1, y_1, \dots, (x_n, y_n), x_i \in R^d \text{ and } y_i \in (-1, +1) \tag{9}
$$

The feature vector is represented by  $x_i$  and the class label (either negative or positive) is represented by  $y_i$  in the estimator combination *i*. Therefore, the desired hyperplane is specified as follows:

$$
wx^T + b = 0 \tag{10}
$$

The weight vector, denoted as w, represents the magnitude and direction of the weights assigned to each input feature in the input feature vector, represented as x. The orientation, represented as b, refers to the bias term. The values of w and b

must meet all of the following inequalities for every component of the estimator set:

$$
wt_i^T + b > +1 \text{ if } y_i = 1wt_i^T + b < -1 \text{ if } y_i = -1 \tag{11}
$$

The goal of estimating an SVM model is to determine the values of w and b that allow for the separation of data points by a hyperplane, while also maximizing the boundary defined by 1/||w||2. Therefore, the xi vectors with an absolute value of  $|y_i|$  $1wt_i^T + b$  equal to 1 are referred to as support vectors.

*1) Enhancing the performance of the Support Vector Machine by optimizing the opposite locus for diagnosing heart disease and selecting features:* The heart rate classification model (HCM) for diagnosing heart illness from ECG signals utilizes a Support Vector Machine that is enhanced by the hybridization of the Grasshopper Optimization Algorithm (GOA) system. The model involves five crucial components. Firstly, implementing various signal preprocessing techniques can enhance the quality of the data and employing an appropriate segmentation method can effectively isolate the peaks from the ECG signals. In this case, discrete wavelet transform (DWT)-based smoothing is employed to enhance signal quality by reducing undesired noise data from ECG signals. If necessary, the normal support vector machine function can be used to provide input for training and testing the HCM. The disease classification was performed using the Support Vector Machine function as an example to obtain the projected output for the automatic intelligent HCM. The maternal ECG signal, which contains a significant amount of data, is displayed in Fig. 1 with a distinct and identifiable period. Fig. 2 displays a typical human ECG heartbeat, which is accompanied by signal interference. In order to proceed with the processing, it is necessary to decrease or minimize the degree of noise [36].



Fig. 2. Uploaded ECG signal in its original form.

Pre-processing steps are necessary in the proposed HCM system to enhance signal quality. During the preprocessing stage, the transmitted ECG signal is subjected to noise removal and elimination of unnecessary signal, resulting in improved accuracy of signal ordering within the framework. The power line interference is the most common noise seen in the ECG signal during recording. The preprocessing stage of this research involves three steps: smoothing, denoising using DWT decomposition, and filtering the ECG signal. The following are the preprocessing stages for the proposed HCM.

*2) Preprocessing:* Pre-processing techniques were employed to cleanse and eliminate any extraneous interference from the ECG signals in this investigation. Initially, various data from all ECG leads were processed and filters were implemented to eliminate any deviation from the baseline and remove power line noise from the signals. Next, the ECG signals undergo four smoothing filters, including moving average, Kaiser, Butterworth, and median filters, to provide a smoother result. The outcomes of this preprocessing are presented in the findings part of this article.

*3) The process of smoothing:* Signal smoothing is a frequently employed technique to decrease the level of noise in a signal, resulting in a noise-free ECG signal with a reduced bit value. In this proposed model, the process of refining the estimating method has been conducted. The Fig. 3 displays the ECG signal after being subjected to a smoothing process. The algorithm used for smoothing is presented below:



*4) Discrete Wavelet Transform (DWT):* After the signal is smoothed, we utilize the Discrete Wavelet Transform (DWT) to break down the signal into two components using different filters, such as a low-pass filter and a high-pass filter. The DWT calculates and provides the coefficients of the ECG signal details. The decomposition algorithm for the DWT is expressed as follows:

*Algorithm 2: Wavelet Decomposition of ECG Signal Input: Filtered ECG Signal (FECG), Decomposition Level (N), Wavelet Type (e.g., Haar) Output: Decomposition Results (Coeff, Lengths)*



*4. Output the processed ECG signal as Cleaned\_ECG\_Signal* 

*5. End*



Fig. 3. Flawless electrocardiogram (ECG) signal.

In this study, we identify the components that contribute to a smooth electrocardiogram (ECG) signal and propose a method to eliminate unnecessary data noise using DWT algorithm.

on the smooth electrocardiographic (ECG) data. The algorithm for filtering the electrocardiography (ECG) signal is expressed

in the following manner:

*6) Extraction of features:* Signal classification models were trained and designed using feature extraction. Three characteristics were derived from ECG signals in this work, namely the QRS-complex integral, T-integral, and Q-integral. Initially, the R wave of each electrocardiogram (ECG) cycle was computed from the ECG signal. Subsequently, the positive and negative peaks preceding and following the R peak are designated as the Q wave and S wave, respectively. Next, the integral (representing the area under the curve) was computed from the Q point to the S point in order to determine the integral characteristic of the QRS complex. Similarly, the T wave of the ECG signal was isolated and the T wave integral and Q wave integral were computed as two additional characteristics. Fig. 4 depicts the procedure of this feature selection algorithm, as described.



Fig. 4. The feature selection system of the suggested technique.

Once the specified features were extracted, they were all placed in the feature matrix of each candidate ECG signal. Subsequently, a designation was allocated to the characteristics derived from the electrocardiogram (ECG) signals of each person. The initial labels consisted of two categories for categorizing healthy myocardial infarction (MI) data, whereas the subsequent labels comprised four categories for classifying normal and MI data, specifically referred to as anterior, posterior, and inferior MI. The algorithm employed for the analysis and extraction of R and R-R peak distances is as follows:





*7) Locust optimization algorithm:* The selection of the SVM classifier parameter is a crucial factor that has a direct impact on the classification results. In this study, the GOA (Genetic Optimization Algorithm) proposed by ref [14] is utilized to choose the optimal parameters for various SVM (Support Vector Machine) classification kernels. Initially, we will provide a concise overview of GOA. Optimization refers to the process of determining the optimal values for the variables of a specific problem in order to minimize or maximize an objective function. The user's text is [8]. Optimization challenges exist throughout diverse academic disciplines. Nevertheless, there have been limited investigations into the simulation of locust swarming algorithms. Grasshoppers, despite being commonly observed as solitary insects in nature, are actually part of a vast category of organisms. Algorithms that draw inspiration from nature separate the user's text is [9]. During exploratory activities, search agents are incentivized to make sudden movements, while they typically travel within a limited area during exploitation. Locusts often carry out these two duties and focus their search in the GOA. Hence, if a mathematical model can be discovered to accurately represent this behavior, it is possible to create a novel algorithm that takes inspiration from nature. The user's text is [10]. In the subsequent text, we outline the algorithm as it was provided in the prior research [19].



*Input: Detected R-peaks & R-R intervals Output: Refined R-peaks & R-R intervals 1. Initialize key parameters:* 

- *– Maximum Iterations (MaxIter)*
- *– Population Size (PopSize)*
- *– Lower Search Bound (LBound)*
- *– Upper Search Bound (UBound)*
- *– Objective Function (ObjFunc)*



#### IV. IMPLEMENTATION SYSTEM

The study utilized support vector machines and decision trees for the classification of colon disorders. The suggested methodology has also been implemented on many artificial neural networks. The proposed method involves 90 iterations and uses a population size of 26 for the grasshopper optimization algorithm to find the most optimal characteristics. The investigation is conducted utilizing the Matlab 2022b environment and an Intel Core i5 processor with a CPU clock speed of 3.82 GHz.

#### *A. Evaluation of Metrics*

Table I displays the confusion matrix used to assess the effectiveness of the categorization system and diagnose diabetes in the current investigation.

TABLE I. DISPLAYS THE CONFUSION MATRIX

Projected values Real values	Unhealthy	Healthy
Healthy	FP (False positive)	TN (True negative)
Unhealthy	TP (True positive)	FN (False negative)

*1) Accuracy:* This criterion measures the overall precision of the classification. It not only assesses the probability of precise categorization in the diagnosis of a healthy individual or patient, but it also assigns each patient to an appropriate disease category [18].

$$
Accuracy = \frac{TP + TN}{TP + FP + FN + TN}
$$
 (12)

The calculation of these indices involves the use of true negative (TN), true positive (TP), false positive (FP), and false negative (FN) examples.

Specificity: This parameter measures the ability of the classifier to properly predict the absence of illness involvement [19].

$$
Specificity = \frac{TN}{FP + TN}
$$
 (13)

Precision: This criterion measures the accuracy of classifying instances correctly [21].

$$
Precision = \frac{\text{TP}}{\text{TP} + \text{FP}_i} \tag{14}
$$

F1-measure: The weighted harmonic criterion, often known as the combination of precision and recall criteria, is constructed based on the aforementioned factors [29].

$$
F1 = \frac{2*Precision*Recall}{Precision+Recall}
$$
 (15)

*2) Root mean square error:* Residual refers to the discrepancy between the anticipated value determined by a statistical model or estimator and the true value.

$$
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (value_{actual} - value_{predicted})^2}{n}}
$$
 (16)

Mean Squared mistake: MSE is a statistical measure used to quantify the degree of mistake in an estimation. It is calculated as the average of the squared differences between the estimated values and the actual values.

$$
MSE = \frac{\sum_{i=1}^{n} |value_{actual} - value_{predicted}|}{n}
$$
 (17)

#### *B. Analysis of the Objective Function*

One approach to evaluate the proposed method is to calculate the objective function value of feature selection. Fig. 5 show the ROC curve with feature selection. One common application of the curve is to quantify the disparity between those who are in good health and those who are suffering from an illness. Indeed, t curve is widely regarded as one of the foremost metrics for



evaluating classification performance. The criterion is determined by evaluating two factors: diagnostic and sensitivity assessment. Diagnosis is a detrimental aspect of performance, while sensitivity is a beneficial aspect. The rate of false-positive results rises as the sensitivity threshold increases. Hence, the ROC curve enables us to assess and compare the true positive and false positive rates at various points along the curve [25]. Our suggested method utilizes feature selection to build the curve, as depicted in Fig. 5.

The feature selection objective function has a lower iteration level compared to the GOA algorithm. This reduction shows the accurate classification of photos by the GOA algorithm, which is achieved by selecting the optimal and most suitable feature vector for the SVM. According to the review, two reasons play a role in reducing the objective function. The main reason is to reduce the dimensions of the feature. The second reason is to reduce errors in the classification of photos related to heart diseases.

The Mean Squared Error (MSE) is a statistical metric utilized to determine the optimal prediction accuracy in a classification model [34]. Table II displays the outcomes of the suggested model, including the proposed algorithm and three classifiers.

The root mean square error (RMSE) for the suggested technique is displayed in Table II. Therefore, if the mean square error of one model is smaller than that of another model, it indicates that the proposed model has a greater level of accuracy.





Fig. 5. Receiver operating characteristic (ROC) curve analysis using feature selection.

TABLE II. RESULTS FOR CRITERIA MEAN S[QUARE](https://www.sciencedirect.com/topics/engineering/mean-square-error) ERROR AND ROOT-MEAN-SQUARE ERROR

	<b>Mean Square Error</b>	Root-mean-square error		
N.B	5.236	3.495		
S.V.M	2.987	2.012		
Tree	4.421	3.654		

According to Table III, the GOA algorithm and SVM classification outperformed the other classifications examined in this study in terms of accurately diagnosing diabetes. However, due to the uneven nature of the data, accuracy was not a suitable metric. Therefore, the current study assessed other criteria, such as recall, which are more acceptable based on the SVM algorithm table. Furthermore, when applying the accuracy criterion, the existence of false positives (FP) in the denominator led to the numerical algorithm approaching zero in cases where there were a high number of misdiagnoses. This raised concerns about the effectiveness of the model. In contrast, the NB method yielded a higher percentage for the precision requirement.

In the current era of machine learning, the standards of comprehension and precision are commonly prioritized over the primary metric of accuracy. Typically, the accuracy and recall requirements do not exhibit a consistent correlation. Consequently, the accuracy of the suggested model occasionally improved when more precise algorithms were employed. Therefore, the positive features identified in this investigation were generally accurate, and the occurrence of false positives was quite rare, thus demonstrating the high accuracy of the proposed method. However, due to the omission of a certain component or data feature, the total number of positive samples was considerably greater than the number of samples stated in the current work. This discrepancy accounts for the remarkably low recall rate. Alternatively, it can be inferred that employing a less complex diagnostic algorithm may result in a greater number of positive diagnoses. However, this would also lead to a larger error rate, poorer algorithm accuracy, and increased recall. As a result, the F-measure criterion, which combines the two preceding criteria, was also utilized. Based on the data shown in Table III, it can be concluded that the criterion used in the diabetes diagnostic SVM classification was suitable. Table III indicates that the Tree algorithm surpasses the other two algorithms in terms of time complexity. Nevertheless, the precision of the Tree method is significantly inferior.

ECG Sig. No.	Accuracy $(\% )$		Time(S)		<b>Specificity</b>		Error $(\% )$		<b>Sensitivity</b>	
	Existing	This Work	Existing	This Work	Existing	This Work	Existing	This Work	Existing	This Work
	97.01	98.85	1.89	1.24	0.895	0.937	2.19	1.21	0.941	0.967
$\overline{2}$	98.14	99.13	2.14	1.88	0.921	0.950	0.94	0.87	0.924	0.943
3	97.32	98.90	1.97	1.15	0.897	0.903	1.56	0.98	0.925	0.972
4	98.27	98.97	3.22	2.55	0.879	0.889	1.27	0.68	0.981	0.960
5	97.26	99.21	2.26	1.39	0.932	0.947	0.98	0.79	0.937	0.948
6	98.22	99.86	2.84	1.87	0.881	0.892	1.25	0.96	0.915	0.936
$\mathcal{I}$	96.53	99.34	3.65	2.84	0.928	0.937	2.01	0.84	0.924	0.953

TABLE III. PERFORMANCE METRICS OF THE PROPOSED PROJECT



### *C. Comparison of the Proposed Method with Other Existing MH Baased Algorithms*

Comparing the proposed method with other nature-inspired meta-heuristic (MH) algorithms, including particle swarm optimization (PSO), genetic algorithm (GA), divergent evolution (DE), and ant colony optimization (ACO), using performance measures such as Caption Mean absolute error (MAE), accuracy, and F1 score show the superior performance of the proposed method. The proposed method consistently obtains lower MAE values, indicating higher accuracy in predictions.

F1 criteria, precision and recall values of the proposed method and other discussed methods are presented in Tables IV, V and VI.

Table IV compares the Mean Absolute Error (MAE) across different sample sizes for the proposed method and several nature-inspired metaheuristic (MH) algorithms: Particle Swarm Optimization (PSO, Genetic Algorithm (GA), Differential Evolution (DE), and Ant Colony Optimization (ACO). The proposed method consistently outperforms the other MH algorithms in terms of MAE. For instance, at a 10% sample size, the MAE for the proposed method is 0.0263, whereas PSO, GA, DE, and ACO have MAE values of 0.0350, 0.0400, 0.0380, and 0.0365, respectively. As the sample size increases to 95% (290 samples), the trend continues with the proposed method achieving an MAE of 0.0295 compared to higher values ranging from 0.0405 to 0.0430 for PSO, GA, DE, and ACO. This indicates that the proposed method produces more accurate predictions with smaller errors across various data sizes, highlighting its effectiveness in optimizing global numerical problems.

TABLE IV. COMPARISON OF THE PROPOSED METHOD WITH OTHER MH METHODS BASE MAE

Sample / Algorithm	<b>PSO</b>	GA	DE.	ACO	Proposed Method
10% (30)	0.035	0.04	0.038	0.0365	0.0263
20% (60)	0.045	0.0425	0.0405	0.041	0.0396
30% (90)	0.033	0.0355	0.0345	0.035	0.0317
40% (120)	0.0345	0.036	0.035	0.0348	0.0249
50% (150)	0.032	0.034	0.0335	0.0325	0.0256
60% (180)	0.0335	0.0365	0.0355	0.033	0.0286
70% (210)	0.035	0.038	0.036	0.0345	0.0284
80% (240)	0.0325	0.035	0.0335	0.033	0.029
90% (270)	0.038	0.041	0.039	0.0375	0.0295
95% (290)	0.0405	0.043	0.0415	0.04	0.0295

Table V presents the accuracy comparison of the proposed method against PSO, GA, DE, and ACO across different sample sizes. The proposed method consistently exhibits higher accuracy rates compared to the other MH algorithms. For example, at a 10% sample size, the proposed method achieves an accuracy of 0.8947, whereas PSO, GA, DE, and ACO achieve accuracies of 0.8450, 0.8350, 0.8400, and 0.8425, respectively. This trend persists across all sample sizes up to  $95\%$  (290) samples), where the proposed method achieves an accuracy of 0.9565, significantly surpassing the accuracy values ranging from 0.8750 to 0.8850 for PSO, GA, DE, and ACO. Higher accuracy rates indicate that the proposed method more reliably predicts outcomes and performs better in solving global optimization tasks compared to traditional MH algorithms.

TABLE V. COMPARISON OF THE PROPOSED METHOD WITH OTHER MH METHODS BASE ACCURACY

Sample / Algorithm	<b>PSO</b>	GA	DE.	<b>ACO</b>	Proposed Method
10% (30)	0.845	0.835	0.84	0.8425	0.8947
20% (60)	0.855	0.8455	0.85	0.852	0.9143
30% (90)	0.86	0.85	0.855	0.8575	0.9412
40% (120)	0.87	0.86	0.865	0.867	0.9394
50% (150)	0.875	0.865	0.87	0.8725	0.9518
60% (180)	0.86	0.85	0.855	0.8575	0.95
70% (210)	0.8755	0.8655	0.8705	0.872	0.9478
80% (240)	0.88	0.87	0.875	0.8775	0.9549
90% (270)	0.8705	0.8605	0.8655	0.867	0.9533
95% (290)	0.885	0.875	0.88	0.8825	0.9565

Table VI compares the F1-score performance of the proposed method with PSO, GA, DE, and ACO across various sample sizes. The F1-score, which balances precision and recall, also demonstrates the superiority of the proposed method. At a 10% sample size, the proposed method achieves an F1-score of 0.8637, whereas PSO, GA, DE, and ACO achieve scores of 0.8255, 0.8200, 0.8225, and 0.8235, respectively. Across all sample sizes up to 95% (290 samples), the proposed method consistently maintains higher F1-scores (up to 0.9146), compared to values ranging from 0.8600 to 0.8650 for PSO, GA, DE, and ACO. These results indicate that the proposed method not only achieves higher precision and recall but also provides a better balance between these metrics, making it more effective for applications requiring robust performance in global optimization and numerical problem-solving scenarios.

TABLE VI. COMPARISON OF THE PROPOSED METHOD WITH OTHER MH METHODS BASE F1-SCORE

Sample / Algorithm	<b>PSO</b>	GA	DE.	<b>ACO</b>	Proposed Method
10% (30)	0.8255	0.82	0.8225	0.8235	0.8637
20% (60)	0.8355	0.83	0.8325	0.8335	0.8744
30% (90)	0.8405	0.835	0.8375	0.8385	0.8889
40% (120)	0.85	0.845	0.8475	0.8485	0.891
50% (150)	0.855	0.85	0.8525	0.8535	0.8991



#### V. CONCLUSION

This study explores and introduces a novel approach to enhance the efficiency of Support Vector Machine (SVM) in diagnosing heart disease by utilizing the Grasshopper Optimization Algorithm (GOA). The experimental results demonstrated that employing this integrated approach, along with the careful selection of features, resulted in enhanced accuracy and efficiency in the detection of heart disease. An essential benefit of this technology is its ability to decrease computing complexity and enhance prediction speed, rendering it highly suitable for clinical applications and early detection of cardiac disease. Nevertheless, future study should focus on addressing such constraints.

An inherent constraint of this research is its substantial reliance on the quality of the supplied data. The accuracy of the model will diminish if the ECG data exhibits significant noise or low quality. Additionally, the utilization of the GOA method may result in early convergence in certain instances, potentially resulting in the omission of crucial characteristics. Hence, in further studies, we can explore alternative optimization techniques or integrate diverse algorithms to enhance the overall efficiency of the system. Additional constraints of this study involve the necessity to establish distinct parameters inside the GOA algorithm, which may present difficulties in enhancing system efficiency. Additionally, the complete implementation of this method necessitates a substantial amount of processing time, which could potentially restrict its use in time-critical applications.

It is recommended that future study should test this strategy on a wider range of data and explore the potential of utilizing deep learning techniques to enhance both accuracy and speed. Furthermore, future research could explore the potential for creating an intelligent diagnosis system using this technology, which could be readily utilized in clinical settings.

#### **REFERENCES**

- [1] J. Umamaheswari and G. Radhamani, "A hybrid approach for classification of dicom image," World of Computer Science and Information Technology Journal (WCSIT), vol. 1, no. 8, pp. 364–369, 2011.
- [2] Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2021). Prediction of heart disease using a combination of machine learning and deep learning. Computational intelligence and neuroscience, 2021(1), 8387680.
- [3] D. M. Gopal, R. Aditya, C. V. K. Reddy, S. Gautham, and N. Nagarathna, "A Survey on Data Mining Applications, Techniques and Challenges in Healthcare," International Journal of Emerging Technologies and Innovative Research, vol. 2, no. 4, 2015.
- [4] Sharma, S., & Parmar, M. (2020). Heart diseases prediction using deep learning neural network model. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(3), 2244-2248.
- [5] Vaziri, A. Y., Makkiabadi, B., & Samadzadehaghdam, N. (2023). EEGg: Generating Synthetic EEG Signals in Matlab Environment. Frontiers in Biomedical Technologies, 10(3), 370-381.
- [6] Vincent Paul, S. M., Balasubramaniam, S., Panchatcharam, P., Malarvizhi Kumar, P., & Mubarakali, A. (2022). Intelligent framework for prediction of heart disease using deep learning. Arabian Journal for Science and Engineering, 47(2), 2159-2169.
- Fansen Wei, Liang Zhang, Ben Niu, Guangdegn Zong. Adaptive decentralized fixed-time neural control for constrained strong interconnected nonlinear systems with input quantization. International Journal of Robust and Nonlinear Control, 2024, https://doi.org/10.1002/rnc.7497.
- [8] Vayadande, K., Golawar, R., Khairnar, S., Dhiwar, A., Wakchoure, S., Bhoite, S., & Khadke, D. (2022, May). Heart disease prediction using machine learning and deep learning algorithms. In 2022 international conference on computational intelligence and sustainable engineering solutions (CISES) (pp. 393-401). IEEE.
- [9] Haoyu Zhang, Quan Zou, Ying Ju, Chenggang Song, Dong Chen. Distance-based Support Vector Machine to Predict DNA N6-methyladine Modification. Current Bioinformatics. 2022, 17(5): 473-482.
- [10] Chen Cao, Jianhua Wang, Devin Kwok, Zilong Zhang, Feifei Cui, Da Zhao, Mulin Jun Li, Quan Zou. webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. Nucleic Acids Research.2022, 50(D1): D1123-D1130.
- [11] Trik, M., Pour Mozaffari, S., & Bidgoli, A. M. (2021). Providing an Adaptive Routing along with a Hybrid Selection Strategy to Increase Efficiency in NoC-Based Neuromorphic Systems. Computational Intelligence and Neuroscience, 2021(1), 8338903.
- [12] Sai Huang, Guangdeng Zong, Ben Niu, Ning Xu, Xudong Zhao, Dynamic Self-Triggered Fuzzy Bipartite Time-Varying Formation Tracking for Nonlinear Multi-Agent Systems With Deferred Asymmetric Output Constraints, IEEE Transactions on Fuzzy Systems, 32(5): 2700-2712, 2024.
- [13] Boyan Zhu,Ning Xu, Guangdeng Zong, Xudong Zhao.Adaptive optimized backstepping tracking control for full-state constrained nonlinear strict-feedback systems without using barrier Lyapunov function method. Optimal Control Applications and Methods, 2024, https://doi.org/10.1002/oca.3136.
- [14] Trik, M., Akhavan, H., Bidgoli, A. M., Molk, A. M. N. G., Vashani, H., & Mozaffari, S. P. (2023). A new adaptive selection strategy for reducing latency in networks on chip. Integration, 89, 9-24.
- [15] Wang, Z., Jin, Z., Yang, Z., Zhao, W., & Trik, M. (2023). Increasing efficiency for routing in internet of things using binary gray wolf optimization and fuzzy logic. Journal of King Saud University-Computer and Information Sciences, 35(9), 101732.
- [16] Khezri, E., Yahya, R. O., Hassanzadeh, H., Mohaidat, M., Ahmadi, S., & Trik, M. (2024). DLJSF: Data-Locality Aware Job Scheduling IoT tasks in fog-cloud computing environments. Results in Engineering, 21, 101780.
- [17] Fakhri, P. S., Asghari, O., Sarspy, S., Marand, M. B., Moshaver, P., & Trik, M. (2023). A fuzzy decision-making system for video tracking with multiple objects in non-stationary conditions. Heliyon, 9(11).
- [18] Saidabad, M. Y., Hassanzadeh, H., Ebrahimi, S. H. S., Khezri, E., Rahimi, M. R., & Trik, M. (2024). An efficient approach for multi-label classification based on Advanced Kernel-Based Learning System. Intelligent Systems with Applications, 21, 200332.
- [19] Khosravi, M., Trik, M., & Ansari, A. (2024). Diagnosis and classification of disturbances in the power distribution network by phasor measurement unit based on fuzzy intelligent system. The Journal of Engineering, 2024(1), e12322.
- [20] Jaferian, G., Ramezani, D., & Wagner, M. G. (2024). Blockchain Potentials for the Game Industry: A Review. Games and Culture, 15554120231222578.
- [21] Jaferian, G., Ramezani, D., Polyak, E., & Wagner, M. (2024). EXPLORING BLOCKCHAIN'S HORIZONS IN EDUCATIONAL GAMING. INTED2024 Proceedings, 5050-5058.
- [22] Sun, J., Zhang, Y., & Trik, M. (2024). PBPHS: a profile-based predictive handover strategy for 5G networks. Cybernetics and Systems, 55(5), 1041-1062.
- [23] Wang, G., Wu, J., & Trik, M. (2023). A novel approach to reduce video traffic based on understanding user demand and D2D communication in 5G networks. IETE Journal of Research, 1-17.
- [24] Li, Y., Wang, H., & Trik, M. (2024). Design and simulation of a new current mirror circuit with low power consumption and high performance and output impedance. Analog Integrated Circuits and Signal Processing, 119(1), 29-41.
- [25] Zhang, L., Hu, S., Trik, M., Liang, S., & Li, D. (2024). M2M communication performance for a noisy channel based on latency-aware source-based LTE network measurements. Alexandria Engineering Journal, 99, 47-63.
- [26] Birdawod, H. Q., Khudhur, A. M., Kadir, D. H., & Saleh, D. M. (2024). A Wavelet Shrinkage Mixed with a Single-level 2D Discrete Wavelet Transform for Image Denoising. Kurdistan Journal of Applied Research, 9(2), 1-12.
- [27] Ameen, A. K., Kadir, D. H., Abdullah, D. A., Maolood, I. Y., & Khidir, H. A. (2024). Assessing E-Government Effectiveness. ARO-THE SCIENTIFIC JOURNAL OF KOYA UNIVERSITY, 12(2), 52-60.
- [28] Mokhlesi Ghanevati, D., Khorami, E., Boukani, B., & Trik, M. (2020). Improve replica placement in content distribution networks with hybrid technique. Journal of Advances in Computer Research, 11(1), 87-99.
- [29] Liao, Y., Tang, Z., Gao, K., & Trik, M. (2024). Optimization of resources in intelligent electronic health systems based on Internet of Things to predict heart diseases via artificial neural network. Heliyon.
- [30] E. I. Elsedimy, S. M. M. AboHashish, and F. Algarni, "New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization," Multimed Tools Appl, vol. 83, no. 8, pp. 23901–23928, 2024.
- [31] Hassanzadeh, H., Qadir, J. A., Omer, S. M., Ahmed, M. H., & Khezri, E. (2024, June). Deep Learning for Speaker Recognition: A Comparative Analysis of 1D-CNN and LSTM Models Using Diverse Datasets. In 2024 4th Interdisciplinary Conference on Electrics and Computer (INTCEC) (pp. 1-8). IEEE.
- [32] Trik, M., Pour Mozafari, S., & Bidgoli, A. M. (2021). An adaptive routing strategy to reduce energy consumption in network on chip. Journal of Advances in Computer Research, 12(3), 13-26.
- [33] Kiani, Farzad, Sajjad Nematzadeh, Fateme Aysin Anka, and Mine Afacan Findikli. "Chaotic sand cat swarm optimization." Mathematics 11, no. 10: 2340, 2023.
- [34] Mohammed, Hardi, and Tarik Rashid. "A novel hybrid GWO with WOA for global numerical optimization and solving pressure vessel design." Neural Computing and Applications 32, no. 18: 14701-14718, 2020.
- [35] Struniawski, Karol, Ryszard Kozera, and Aleksandra Konopka. "Performance of Selected Nature-Inspired Metaheuristic Algorithms Used for Extreme Learning Machine." In International Conference on Computational Science, pp. 498-512. Cham: Springer Nature Switzerland, 2023.
- [36] Seyyedabbasi, Amir, and Farzad Kiani. "I-GWO and Ex-GWO: improved algorithms of the Grey Wolf Optimizer to solve global optimization problems." Engineering with Computers 37, no. 1 (2021): 509-532.