

Automatika

Journal for Control, Measurement, Electronics, Computing and Communications



ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/taut20

Grey wolf optimized stacked ensemble machine learning based model for enhanced efficiency and reliability of predicting early heart disease

Geetha Narasimhan & Akila Victor

To cite this article: Geetha Narasimhan & Akila Victor (2024) Grey wolf optimized stacked ensemble machine learning based model for enhanced efficiency and reliability of predicting early heart disease, *Automatika*, 65:3, 749-762, DOI: [10.1080/00051144.2024.2317098](https://doi.org/10.1080/00051144.2024.2317098)

To link to this article: <https://doi.org/10.1080/00051144.2024.2317098>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 26 Feb 2024.



Submit your article to this journal [↗](#)



Article views: 880



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



Grey wolf optimized stacked ensemble machine learning based model for enhanced efficiency and reliability of predicting early heart disease

Geetha Narasimhan and Akila Victor

School of Computer Science and Engineering, VIT, Vellore, Tamil Nadu, India

ABSTRACT

Heart disease is one of the foremost reasons for death globally. Machine learning (ML) can be used to predict heart diseases early, which can help improve patient outcomes. This research proposes a novel machine learning method for predicting heart disease using a combination of Grey Wolf Optimization (GWO) and stacked ensemble techniques. GWO is a metaheuristic algorithm that can be used to optimize the parameters of machine-learning models. The stacked ensemble technique is a combination of multiple machine learning models to improve the overall accuracy of the prediction. The model proposed was evaluated using a dataset of heart patients. The results showed that the model achieved a 93% accuracy, which was significantly higher compared to traditional machine learning methods. The proposed method also had a higher precision of 91%, sensitivity of 95.3%, F1 score of 92.9%, and Matthew coefficient of 0.83, less in Log_Loss 2.87 than the traditional methods. The results of this research suggest that the proposed model is a promising new approach for predicting heart diseases. This method is more accurate and reliable than traditional methods and has the potential to improve patient outcomes.

ARTICLE HISTORY

Received 26 December 2023
Accepted 6 February 2024

KEYWORDS

Heart disease; Grey Wolf Optimization; stacked ensemble; performance evaluation; machine learning

1. Introduction

Healthcare is a vast area of research that includes cardiovascular diseases, diabetes, drugs, and cancer. For these diseases, different factors and features are considered, and different datasets are available online. As described by the World Health Organisation (WHO), the leading cause of death globally is cardiovascular disease [1]. According to the WHO [1], every year around 17.9 million deaths occur due to heart disease. This indicates a global death count of 31%. In the CDC report [2], 80% of the deaths were due to heart attacks. Predicting patients with heart disease is important for reducing their risk. Therefore, it is necessary to identify patients with heart disease and treat them with the utmost care to decrease the risk of death. Apart from the traditional method of diagnosing the disease, other methods, such as machine learning (ML), help identify high-risk patients.

Patients with heart complications must be treated early to diagnose cardiovascular disease. This study mainly focused on applying ML to a heart disease dataset optimized by Grey Wolf Optimization (GWO) [3]. Among various diagnostic methods, this study specifically investigates the feasibility of accurate heart disease prediction through machine learning. To achieve accurate disease prediction, the system first employs the Grey Wolf Optimizer (GWO) algorithm

to carefully select the most relevant features, effectively eliminating those that are redundant or irrelevant. This refined set of features is then fed into a stacked machine learning (ML) classifier for robust prediction. GWO, inspired by the social hierarchy and hunting strategies of grey wolves, offers distinct advantages over traditional optimization algorithms like PSO and GA: Reduced complexity: GWO operates with fewer parameters, simplifying its implementation and comprehension. Straightforward principles: Its core concepts are easy to grasp, making it accessible to a wider range of users. Effortless implementation: It can be readily integrated into various ML frameworks without extensive configuration. This combination of GWO's efficient feature selection and the stacked ML classifier's predictive power fosters a powerful and streamlined approach to disease prediction [3].

Different machine-learning techniques are involved in the detection of heart diseases. It is essential to accurately detect diseases using ML techniques. Misdiagnosis may lead to an increased risk for heart patients. Physicians and radiologists stereotypically use physical tests and the medical history of the patients for diagnosing the disease, and later procedural and diagnostic tests will be carried out based on the symptoms. Artificial intelligence led to the development of ML models that are accurate in predicting heart disease.

Most models use existing statistical algorithms to examine large datasets of patients, which include all kinds of medical information about the patient. From supervised classification to unsupervised anomaly detection, researchers have embraced a diverse toolbox of ML techniques in their pursuit of improved heart disease identification. Several ML algorithms are trained for predicting disease, including Decision Tree (DT), Naïve Bayes (NB), Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). These models work on huge patient databases and are used to categorize risk features and improve a predictive model for identifying the risks.

SGD approximates gradient descent by estimating the model improvement direction using a single data point instead of the entire dataset. This leads to computational efficiency for large-scale learning but introduces inherent randomness [31]. While SGD sacrifices quick convergence for the optimal solution, its statistical properties and implicit bias often result in good predictions and competitive performance. However, additional computation during each iteration and potential accuracy loss due to gradient compression techniques are trade-offs to consider [31]. While SGD saves time in calculations, it comes with the cost of potentially taking longer to reach the best result. Additionally, using tricks like gradient compression for further efficiency might compromise the accuracy of the final model [32].

Robust optimization in the field of machine learning pertains to the capacity of a learning algorithm to exhibit high performance across diverse data types [33]. This entails striking a balance between performance and supplementary costs, such as an increased number of data samples, intricate objective functions, or protracted optimization iterations [34, 35]. The potential loss of interpretability arises from the intricate nature of robust optimization problems [36]. In comparison to standard approaches, the resulting solutions may exhibit reduced interpretability [37]. Consequently, comprehending the reasoning behind the decisions made by the model and identifying potential issues becomes more challenging [38]. The internal complexity of robust models further exacerbates the difficulty in debugging and enhancing their performance. Additionally, certain robust optimization techniques may necessitate specialized expertise or software tools. Moreover, specific implementations can be sensitive to hyperparameter settings, thereby increasing the difficulty in attaining optimal performance [39].

Various studies have been involved in developing ML-based models for predicting heart disease, where diagnostic tests are inaccessible. ML uses classification techniques to predict data. These classification techniques use training data to predict a developed model.

There are various classification techniques, such as traditional methods, hybrid methods, deep learning, and ensemble techniques. All these models aim to achieve maximum accuracy in healthcare.

The process begins with initializing parameters like the number of wolves and iterations. Base-level classifiers: A set of base-level classifiers, such as Random Forest (RF), Support Vector Machine (SVM), etc., are used to evaluate the initial population of features. The performance evaluation is analyzed. Among the ML methods, Random Forest produced 88% accuracy. GWO is a swarm intelligence technique inspired by the hunting behaviour of grey wolves in nature. It's used to optimize complex problems and can be applied to feature selection, which involves choosing a subset of relevant features from a larger dataset. Three coefficients, alpha, beta, and delta, are used to control the movement of the wolves in the GWO algorithm. The fitness of each wolf (feature subset) is calculated based on its classification performance using the base-level classifiers. The positions of the wolves are updated based on their position and the positions of the alpha, beta, and delta wolves. The continuous positions of the wolves are converted into discrete feature indices. The process stops if a stopping criterion, such as a maximum number of iterations or a desired fitness level, is met. If the stopping criteria are not met, alpha, beta, and gamma are updated, and the fitness is recalculated with the updated feature subset. This loop continues until the stopping criteria are satisfied, and the feature subset with the highest fitness is returned as the optimal feature subset. The feature indices are reranked based on their fitness values. The performance of the selected features is evaluated using meta-learners, which are models that combine the predictions of multiple base-level classifiers.

The contributions of the paper

- The proposed GWO was used to optimize the considered dataset.
- Ensemble machine-learning techniques were stacked. Approximately 18 ML models were stacked to achieve the desired accuracy.
- The proposed model performance and other state-of-the-art models are analyzed and compared.

The remaining section of this paper is organized as follows: Section 2 deals with related works that include various existing models. Section 3 explains the materials and methods involved in the study, including the ML techniques, GWO algorithm basics, performance metrics, and the dataset. Section 4 explains the experimental results, analysis, and discussion. To end, Section 5 completes the research.

2. Related work

2.1. Various situation

Various studies have explained the effective and real-world applications of ML in heart disease detection. The study involved a dataset from the UCI repository, which is available publicly to everyone [4]. ML techniques are very useful in the medical field for increasing accuracy and reducing computational costs. Consider Verma et al., who projected a model that used particle swarm optimization (PSO) [5] and ML techniques for heart disease prediction, which obtained 90.28% accuracy.

Al-Tashi et al. [6] and El Bakrawy [7] worked on GWO with single ML techniques and achieved 89.33% and 87.45%, respectively. Garavand et al. [8] worked with ML techniques and reached 85%. Dissanayake et al. [9] Sabab et al. [10] and Garavand et al. [8] worked with the single ML algorithm and feature selection method. They achieved 88.52%, 87.8%, and 84% accuracy, respectively. Saqlain et al. used the Fisher score algorithm [11] for selecting features and SVM for prediction. This study achieved 81.91% accuracy and 88.68% specificity. Latha and Jeeva developed a hybrid model that used four ML algorithms, namely NB, BN, RF, and MP, which achieved 85.45% accuracy [12]. Itoo and Garg [13] proposed a model that stacked ensemble models like LR, KNN, and NB for predicting heart disease. The model achieved 90% accuracy.

Liu et al. [14] applied 10 classifiers for the stacking ensemble model. The stacked model achieved 89.86% accuracy in predicting heart disease. PAL and GANGWAR [15] reached 82.95% accuracy in predicting heart disease by applying stacked models to the dataset. About nine base learners were involved in this study. Six base learners like Rf, Extra Tree Classifier, KNN, XGB, SGD, Adaboost, and MLP were applied by PA & PRIYA [16] and obtained 90.2% accuracy. Harika et al. [17] proposed an ensemble framework for rapid prediction of heart disease. The model obtained 87.05% for stacked ensemble, 84.74% for ANN, 81.35% for NB, and 79.66% for SVM. Karadeniz et al [18] adopted a data-driven approach, utilizing a Lasso graph for feature selection and Ledoit-Wolf shrinkage for improved predictive performance in predicting heart disease. The models yielded 88.7 and 88.8 accuracy, respectively.

Talukdar & Singh [40] addresses the rise in mortality rate, with cardiovascular disease being a significant contributor, and the need to predict and treat heart disease using medical data and analytical insights. The study introduces an artificial neural network methodology for identifying potential cardiovascular disease risk factors and generating a predicted list of risk features most likely to result in cardiovascular disease. The model achieved 81% accuracy which used a back-propagation algorithm along with MLP. Taylan et. al [41] proposed a methodology that combines machine

Table 1. Existing study compared to this study.

Existing study	Method	Accuracy (%)	Feature Selection
Verma et al. [5]	PSO + ML	90.28	No
Al-Tashi et al. [6]	GWO + Single ML	89.33	No
El Bakrawy [7]	GWO + Single ML	87.45	No
Garavand et al. [8]	ML	85	No
Dissanayake et al. [9]	Single ML + Feature selection	88.52	Yes
Sabab et al. [10]	Single ML + Feature selection	87.8	Yes
Garavand et al. [8]	Single ML + Feature selection	84	Yes
Saqlain et al. [11]	Fisher Score + SVM	81.91	Yes
Latha and Jeeva [12]	Hybrid ML (4 algorithms)	85.45	No
Itoo and Garg [13]	Stacked ensemble (LR, KNN, NB)	90	No
Liu et al. [14]	Stacked ensemble (10 classifiers)	89.86	No
Pal & Gangwar [15]	Stacked ensemble (9 base learners)	82.95	No
Pa & Priya [16]	Stacked ensemble (6 base learners)	90.2	No
Harika et al. [17]	Ensemble framework (stacked, ANN, NB, SVM)	87.05 (stacked)	No
Karadeniz et al [18]	Lasso + Ledoit-Wolf	88.7 & 88.8	Yes
Talukdar & Singh [40]	MLP with backpropagation	81	No
Taylan et al. [41]	ML + Neuro-fuzzy + Statistical methods	> 90.00	No
Hossain et al. [42]	Random Forest with feature selection	90	Yes
Jawalkar et al. [43]	Decision Tree-based Random Forest	86	No

learning, neuro-fuzzy, and statistical methods to predict cardiovascular diseases with high accuracy, exceeding 90%. Hossain et. al., [42] aimed to analyze patient data to accurately predict heart disease and identify the most significant attributes for prediction using the Correlation-based Feature Subset Selection Technique with Best First Search. Distinct artificial intelligence techniques are applied and compared, with random forest using selected features achieving the highest accuracy rate of 90% for heart disease prediction. Jawalkar et. al., [43] proposed an ML approach for heart disease prediction using a decision tree-based random forest classifier with loss optimization. The paper does not explicitly mention the specific traditional methods that were compared to the proposed approach. The model achieved 86% precision and 86% recall. The existing study in this section is tabulated in Table 1.

A thorough investigation of existing models for predicting heart disease shows that stacked ensemble models achieve an accuracy of up to 90%. However, the authors could have explored other ensemble methods, such as bagging and boosting. Additionally, different evaluation metrics, such as ROC and AUC, are missing. The existing models also do not analyze feature importance to understand which features are most predictive of heart disease, and they do not compare the proposed framework to various state-of-the-art methods.

From the overhead studies, it is evident that these models do not achieve greater accuracy, and feature optimization is not performed. Our study employed feature optimization and stacking, which will be explained in further sections. This study aimed to identify high-risk patients with heart disease. To address limitations identified in prior research, this study developed a model that optimizes stacked ensemble ML for specific applications.

3. Materials and methods

3.1. Proposed method

Figure 1 illustrates the framework of the proposed model. This model has two phases. In the first phase, data were obtained from the UCI repository and then preprocessed. The preprocessed data are used for GWO, which optimizes the dataset. In the second phase, the optimized data were applied to stacked

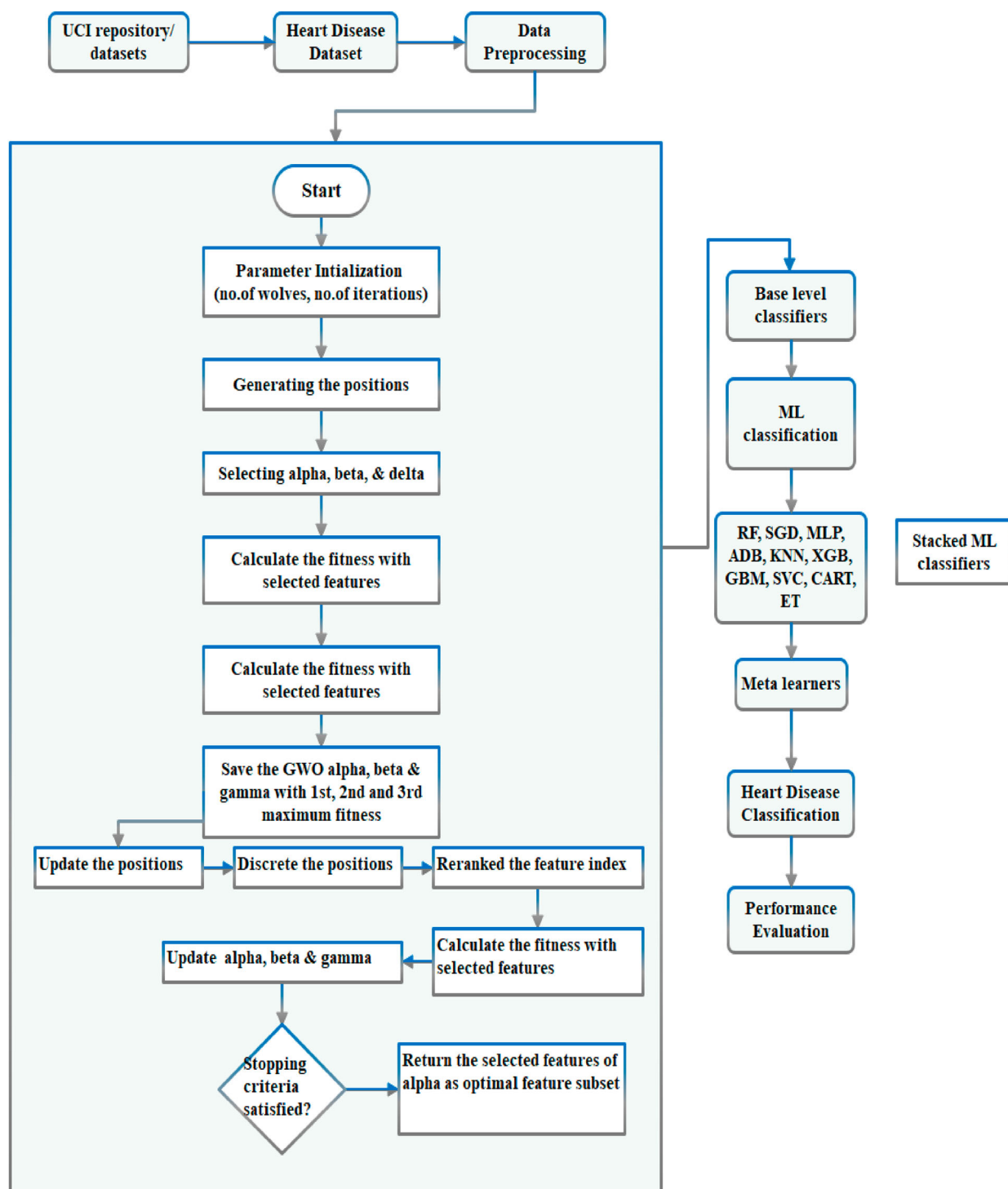


Figure 1. Process Flow of proposed model. The model uses a dataset from the UCI repository and is preprocessed. The preprocessed model is passed to the GWO algorithm, and features are optimized. The optimized features are then used by ML techniques and performance is evaluated.

ensemble ML techniques. About 18 ensemble techniques were stacked. The generation of the model, followed by a performance analysis of the results obtained from the model, was compared with the basic ML learners. Each step is elucidated in detail in the subsequent sections.

3.2. Dataset preparation

In the proposed model, the open-source heart disease dataset is retrieved from the UCI repository. The dataset included 1190 examples of records. The dataset consisted of 13 independent features and one dependent target class [4]. Table 2 lists the attributes of the datasets. This dataset was obtained by combining various clinical test outputs, namely, serum cholesterol, vessel count, thalassemia, and fasting blood sugar. From the electrocardiogram, ST depression and sloping ST segments were achieved.

3.3. Statistical analysis and preprocessing

The dataset was initially loaded and analyzed. In this study, outlier detection was performed as the first stage of data pre-processing. Z-score outlier detection is used to boost the efficiency of the model. According to the experimental rule, a data point in a collection with a z-score of more than 3 is regarded as an outlier. The dimension of a data point's deviation from the mean value is called the z-score, which is called the standard score. It displays the range of values for an attribute in the dataset Beunza et al., [19]

$$Z_{score} = \frac{x - \mu}{\sigma} \quad (1)$$

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

One-hot encoding was used for categorical features, such as the slope of the ST segment (slope), chest pain (cp), sex, and resting electrocardiogram (restecg). It

Table 2. List of attributes of the Heart dataset from UCI with feature information [4].

S. No	Attribute	Description	Type	Value Range
1	Age in years	Continuous value	Numeric	29 < age > 77
2	Sex	1 represents male 0 represents female	Nominal	1-male 0-female
3	CP-Chest pain	Chest pain type 0 represents typical angina 1 represents atypical angina 2 represents non-anginal pain 3 represents asymptomatic	Nominal	4-asymptomatic 3-non anginal pain 2-atypical angina 1-typical angina
4	trestbps	Resting blood pressure	Numeric	> 160: very high 140–160: High 120–140: Unusual 90–120: Normal
5	chol	Serum cholesterol in mg/dl	Numeric	> 250: very high 110–200: Normal 200–240: borderline high 240–250: High
6	Fasting Blood sugar	Fasting blood sugar > 120 mg/dl 1 represents true 0 represents false	Nominal	1-true 0-false
7	Restecg – resting electrocardiographic (ECG)	resting electrocardiographic results 0 represents normal 1 represents having ST-T wave abnormality 2 represents Possible or definite left ventricular hypertrophy	Nominal	0-normal 1-STTwave abnormality 2-showing probable
8	thalach	Maximum heart rate achieved	Numeric	60–100-normal > 100-tachycardia
9	Exangexercise-induced angina	Blood supply when you exercise	Nominal	1 0
10	Old peak	1 represent yes 0 represents no	Numeric	0-6.2
11	Sl: slope	ST depression induced by exercise relative to rest The slope of the peak exercise ST segment 0 represents upsloping 1 represents flat 2 represents down-sloping	Nominal	1-upsloping 2-flat 3-downsloping
12	Ca	number of vessels (0-3) coloured by fluoroscopy	Nominal	0 1 2 3
13	thal	Thallium stress test 1 represents normal; 2 represents fixed defect; 3 represents a reversible defect	Nominal	3-normal 6-fixed 7-reversible
14	target	The predicted attribute 0 represents no chances of heart failure 1 represents chances of heart failure	Class variable	0-no heart disease 1-heart disease

transforms an attribute into a form that can be understood by ML algorithms by turning it into a numerical format. Two sets of data were pre-processed for preliminary analysis before feature scaling. A standard scaler was used to standardize the attribute values for the initial set of data. The min-max scaler was used to normalize the values for the other set. The numbers in the min-max scale range from 0 to 1, where 0 represents the smallest value discovered and 1 represents the maximum. The remaining information consists of decimals between 0 and 1, as shown in Equation 1,2. The statistics for the numerical columns are presented in Figure 2.

Both cholesterol and resting blood pressure had outliers, as can be seen from the description above. While both variables had the lowest value of 0, cholesterol also had an outlier on the upper side, with 603 as the maximum value. The outliers were removed. Once the outliers are removed the shape of the dataset includes 1171 records. To select the best models to be utilized in level 0 of the stacked ensemble approach, we will develop various baseline models and perform 10-fold cross-validation at this stage. After comprehensive preprocessing, we tested the dataset on various established machine learning models: Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes (NB), Random Forest (RF), XGBoost (XGB), Decision Trees (DT), and Neural Networks (NN). Each model underwent classification,

and their performance was thoroughly evaluated using the metrics outlined in Table 3.

3.4. Stacked ML classifiers

Basic and popular ML techniques in the healthcare sector for predicting disease are highly capable and have greater heterogeneity Taha et al., [20]. Building upon the power of ensemble learning, this work explores a two-stage stacking technique, where multiple ML models are sequentially combined for enhanced robustness and accuracy. This study investigated the impact of feature optimization on stacked ensemble performance. Stage 1 utilized all features, while Stage 2 employed the GWO algorithm to identify the most informative features, refining the data input for the stacked ensembles in the final step. Tables 4 and 5 explains the results obtained in both stages. The ensemble techniques stacked are RF, multilayer perceptron (MLP), KNN, extra tree classifier, XGB, support vector classifier (SVC), Stochastic Gradient descent (SGD), Adaboost, classification and regression tree (CART), and Gradient Boosting Machine (GBM). These ensemble methods explain how a target value can be predicted based on other values.

RF and CART are supervised ML techniques that use decision trees as basic classifiers. Both methods produce many classifiers, and the results are accumulated using the maximum number of votes. SVC is

	age	sex	chest_pain_type	resting_blood_pressure	cholesterol	fasting_blood_sugar
count	1190.000000	1190.000000	1190.000000	1190.000000	1190.000000	1190.000000
mean	53.720168	0.763866	3.232773	132.153782	210.363866	0.213445
std	9.358203	0.424884	0.935480	18.368823	101.420489	0.409912
min	28.000000	0.000000	1.000000	0.000000	0.000000	0.000000
25%	47.000000	1.000000	3.000000	120.000000	188.000000	0.000000
50%	54.000000	1.000000	4.000000	130.000000	229.000000	0.000000
75%	60.000000	1.000000	4.000000	140.000000	269.750000	0.000000
max	77.000000	1.000000	4.000000	200.000000	603.000000	1.000000

	rest_ecg	max_heart_rate_achieved	exercise_induced_angina	st_depression	st_slope	target
1190.000000		1190.000000		1190.000000	1190.000000	1190.000000
0.698319		139.732773		0.387395	0.922773	1.624370
0.870359		25.517636		0.487360	1.086337	0.610459
0.000000		60.000000		0.000000	-2.600000	0.000000
0.000000		121.000000		0.000000	0.000000	1.000000
0.000000		140.500000		0.000000	0.600000	2.000000
2.000000		160.000000		1.000000	1.600000	2.000000
2.000000		202.000000		1.000000	6.200000	3.000000

Figure 2. Statistical analysis of dataset.

Table 3. Various evaluation criteria used for measuring performance. Jangle & Narayankar [27].

Measure	Formula	Description
Precision	$\frac{TP}{TP + FP}$	Total positives predicted to that of number of positives
Recall	$\frac{TP}{TP + FN}$	Prediction of total positives to that of samples total positives
Specificity	$\frac{TN}{TN + FP}$	Number of negative results from the total number of non-disease samples
Accuracy	$\frac{TP + FP + FN + TN}{TP + FP + FN + TN}$	Number of correct predictions from the entire predictions
F-score	$2 * \left(\frac{Precision * Recall}{Precision + Recall} \right)$	This F-score is mainly used to resolve classification problems when there is an imbalanced class.
Log_Loss	$-1/N \sum_{i=1}^N (\log(P_i))$	It deals with the quantifying difference between predicted probabilities and actual values.
MCC	$\frac{TPTN - FP * FN}{\sqrt{TP + FP * (TP + FN) * (TN + FP) * (TN + FN)}}$	A tool used to measure the model evaluation. It's a measure between predicted and actual values.

Table 4. Performance analysis before feature selection.

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity (%)	Classification error rate (%)
LR	85.25	86	88	87	81	14.75
KNN	68.85	70	76	73	59	31.15
SVM	70.49	70	82	76	55	29.51
NB	85.25	84	91	87	78	14.75
RF with Grid Search	86.89	88	88	88	81	13.11
RF	88.89	89	94	91	85	9.84
XGB	81.97	83	85	84	77	18.03
DT	83.61	88	82	85	85	16.39
NN	83.61	80	94	86	43	16.39

applied to obtain linear and non-linear data. A decision hyperplane is employed by SVC for class recognition. This method is robust, as it accurately discourses the bias and variance in the data. In KNN, data are clustered and neighbours are defined by K. Based on the similarity measure, new instances can be classified. The predictions from various decision trees combined for the final prediction were made using a gradient boosting machine. The eigenvalues were divided into K intervals, and the results were classified in this model. This increases the speed of prediction and decreases storage. AdaBoost iteratively corrects its mistakes by leveraging wrongly classified samples. It assigns higher weights to these samples, guiding subsequent classifiers to focus on areas where the previous one faltered. This process continues until a desired level of accuracy is achieved. Multi-Layer Perceptron (MLPs) excel at classifying data by mapping inputs to specific classes.

They apply interconnected layers of neurons to learn complex relationships within the data. The extra-tree classifier builds its predictive power by constructing numerous decision trees. Unlike traditional random forests, it samples data without replacement, ensuring each tree has a unique data sample. Extreme Gradient Boosting (XGB) tackles classification through an efficient boosting technique. It builds an ensemble of decision trees iteratively, focusing on improving predictive accuracy for previously misclassified samples. The performance is better than other state-of-the-art ML models Chiu et al. [21].

Combining various categories of classification techniques using a meta-classifier is known as stacking Verma & Pal [22]. The idea is to merge weak learners to attain robust generalization ability. During stacking, the results of the base learners are fused. The first-level learners are the base learners, and combinations

Table 5. Performance evaluation of the stacked model with other ML techniques.

Model	Accuracy	Precision	Sensitivity	Specificity	F1 Score	ROC	Log_Loss	mathew_corrcoef
Stacked Classifier	0.861017	0.857143	0.884615	0.834532	0.870662	0.859574	4.800367	0.720979
RF	0.864407	0.858025	0.891026	0.834532	0.874214	0.862779	4.683286	0.727878
MLP	0.772881	0.811189	0.74359	0.805755	0.77592	0.774673	7.844473	0.548688
KNN	0.783051	0.810811	0.769231	0.798561	0.789474	0.783896	7.493234	0.566852
Extra tree classifier	0.881356	0.88535	0.891026	0.870504	0.888179	0.880765	4.09787	0.761846
XGB	0.847458	0.84472	0.871795	0.820144	0.858044	0.845969	5.268695	0.693701
SVC	0.776271	0.796053	0.775641	0.776978	0.785714	0.77631	7.727403	0.551958
SGD	0.745763	0.712042	0.871795	0.604317	0.783862	0.738056	8.781194	0.497445
Adaboost	0.776271	0.796053	0.775641	0.776978	0.785714	0.77631	7.727403	0.551958
CART	0.827119	0.814371	0.871795	0.776978	0.842105	0.824387	5.971195	0.653431
GBM	0.816949	0.81875	0.839744	0.791367	0.829114	0.815555	6.322431	0.632336

of base learners are meta-learners, also called second-level learners. The base learners were trained using this dataset. The output from these base learners is passed on as input features for second-level learners. A new dataset with original labels for training meta-learners. The 13 individual models were trained using the available dataset.

3.5. GWO algorithm

Mirjalili et al. [23] established the GWO after being inspired by the behaviour of grey wolf packs in 2014. The interesting characteristic of the grey wolf is the exceptional hunting and looking for prey. The group of wolves, called canines, plays a different role and completes their tasks by cooperating with other wolves. The GWO comprises four levels in hierarchical order, as shown in Figure 3. The most important is wolf α , which decides on hunting activities. The second step is wolf β , which is subordinate to wolf α , and β is the best candidate for α in making decisions. The third rank is wolf δ , which is a subordinate of wolf α and β . The wolf δ is responsible for scouting and hunting. The last and fourth values are the wolf ω values. This wolf maintains its pack. Wolf hunting is categorized into tracking, chasing, and prey attacks.

Wolf α , β , δ are responsible for each reiteration, and the mathematical model is described as [23]

$$D = |C \cdot X_p(t) - X(t)| \omega \quad (3)$$

$$X(t+1) = X_p(t) - A \cdot D \quad (4)$$

Equation 3 explains the distance between the prey and the grey wolf. The current iterations are denoted as t . The location of the grey wolf iteration at t is denoted by $X_p(t)$ and position is denoted by $X(t)$. Prey location was updated using Equation 4. The A and C coefficient vectors were calculated using Equations 5 and 6, respectively. The components of a are reduced linearly from 2 to 0 after a few iterations. Random vectors are r_1 and

r_2 in $[0,1]$. The main role of this random vector is to increase the randomness

$$A = 2a \cdot r_1 - a \quad (5)$$

$$C = 2 \cdot r_2 \quad (6)$$

Finally, the trade-off between exploitation and exploration is achieved by updating a vector, which is given by Equation 7.

$$A = 2 - t \quad (7)$$

The random vectors r_1 and r_2 let the wolves reach any position. As a result, a grey wolf can update its position inside the space surrounding the prey in any random location based on Equations 3 and 4.

$$X_1 = X_\alpha - A_1 * (D_\alpha), X_2 = X_\beta - A_2 * (D_\beta),$$

$$X_3 = X_\delta - A_3 * (D_\delta) \quad (8)$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (9)$$

3.6. GWO Optimized Stacked ML classification

Feature selection is a method used to reduce the number of suitable features, which boosts the classification by obtaining the feature subset from the initial features. It ignores the less important features that help reduce computational and memory costs. This algorithm is a meta-heuristic that replicates the process of grey wolves that live in groups of five to twelve individuals. Emmanuel et al. [24] use search and optimization problems. Algorithm 1 operates in two phases: In phase 1 the dataset is given to the stacked ensemble techniques, and in phase 2 feature optimization through the GWO algorithm is applied, and the resulting streamlined set of features is utilized by the stacked ensemble models.

The code uses a UCI dataset, which is a common source of publicly available datasets for machine learning research. The dataset contains features (f_1 ,



Figure 3. Grey wolf optimization system.

f_2, \dots, f_n) representing various attributes related to heart disease. The dataset is divided into a training set (X) for model development and a test set (Y) for final validation. Feature Selection using GWO to identify the most relevant features for predicting heart disease, improving model accuracy and interpretability. GWO population (X_i) representing different feature combinations is created. Parameters a , A , and C , which control GWO's search behaviour, are initialized. Each feature combination's fitness (its ability to predict heart disease) is calculated. GWO's exploration and exploitation mechanisms are used to update feature combinations over multiple iterations. The best three feature combinations (X_α , X_β , X_δ) are tracked. The best feature combination (X_α) is returned after the iterations.

Stacked Ensemble Model Training and Testing: For each model in the ensemble (T_i), a subset of features (D_1) is selected from the optimal set identified by GWO. Each model (T_i) is trained and tested on each fold of the data using the corresponding feature subset (D_1). The final output is a summary of performance scores for each model in the ensemble, providing insights into their strengths and the overall ensemble's effectiveness. The step-by-step pseudocode is explained in Algorithm 1. This two-step approach aimed to combine the power of feature selection with the advantages of ensemble learning.

Algorithm 1. GWO Algorithm Optimized Stacked Ensemble ML

```

Let  $D = \{f_1, f_2, f_3, \dots, f_n\}$  given in UCI dataset
 $X =$  training dataset
 $Y =$  test dataset for final validation
 $X \in D$ 
 $Y \in D$ 
Input:  $X_t$  for training models
 $X_v$  for testing base models
 $T_{13} = \{LR, LDA, DT, NB, RF, MLP, KNN, ET, XGB, SVC, SGD, ADB, GBM\}$ 
Apply GWO for optimization
  Initialize the GWO population  $X_i$  ( $i = 1, 2, \dots, n$ )
  Initialize  $a$ ,  $A$  and  $C$ 
  Calculate the fitness of the search agent
  The best search agent is  $X_\alpha$ 
  The 2nd best search agent  $X_\beta$ 
  The 3rd best search agent  $X_\delta$ 
  While ( $t <$  maximum iterations)
    For each search agent
      Reset the current search agent position by equation (9)
    End for
    Update  $a$ ,  $A$  and  $C$ 
    Calculate the fitness of search agents
    Update  $X_\alpha$ ,  $X_\beta$ ,  $X_\delta$ 
     $T = t + 1$ 
  End while
  Return  $X_\alpha$ 
For every  $i$  in  $T_i$ 
  Let  $D_1 = \{f_1, f_2, f_3, \dots, f_i\}$  obtained from  $D$ 
  Input:  $X_t$  for training models  $\in D_1$ 
   $X_v$  for testing base models  $\in D_1$ 
   $T_9 = \{MLP, KNN, ET, XGB, SVC, SGD, ADB, CART, GBM\}$ 
  For  $j = 1, 2, 3, \dots, 10$ 
  Do
    Train and test every fold for every classifier  $T_i$ 
  end for
end for
output: model score summary

```

3.7. Performance measures

Data mining is evaluated through performance metrics. Once the classification is completed, the performance metrics are evaluated by determining the confusion matrix and the ROC curve. The recall was used to assess the completeness of the model. If recall is higher, then fewer false negatives (FN) are produced. The exactness was measured using precision. While accuracy contributes to overall performance, the F1 score offers a more nuanced view by balancing precision and recall. This makes it the most reliable indicator of a model's ability to accurately predict both positive and negative cases. For comprehensive model evaluation, the F1 score surpasses simple accuracy by considering both how well the model identifies true positives and avoids false positives. This makes it the preferred metric for determining a model's effectiveness in real-world applications.

The ROC curve is a graph that explains the model's classification performance. The ROC curve was plotted with true-positive and false-positive values. Table 3 presents the performance metrics obtained using their formulas. The accuracy metric measures the percentage of predictions for which the model is correct. The precision metric measures the percentage of predictions classified as positive or positive by the model. The recall metric measures the percentage of positive cases correctly identified by the model. The F1 score is a weighted average of precision and recall. The Log_Loss called logarithmic loss or cross-entropy loss is a measure of how well the model predicts the actual values against the predicted probabilities. A lower Log_Loss indicates a better model, Wang et al. [25]. The Mathew correlation coefficient (MCC) is also a statistical tool for model evaluation. It is an amount of the variation between actual and predicted values. This metric is more reliable and produces a high score in case of good results obtained during prediction in the four categories, namely TP, FN, TN, FP Chicco & Jurman [26]. This value is equal to the chi-square statistics.

4. Experiments results and analysis

4.1. Experimental setup

The performance of the suggested model is assessed using ML models. The proposed models begin with the necessary data collection and proceed with preprocessing to accommodate missing values. The following step is to choose crucial attributes from the provided dataset. The performance of the ML model will also be examined using specific features. A variety of evaluation criteria are used to analyze performance. The last decision is used to conclude. The Anaconda Jupyter Notebook 6.4.8, which features built-in packages for ML models, and an Intel(R) Core (TM) i7-7600U CPU running at 2.80 and 2.90 GHz, is used to experiment.

4.2. Result analysis

ML algorithms were trained on a dataset. The algorithm's performance was assessed using the metrics listed in Table 3. Table 4 shows the detailed performance analysis of each algorithm. RF outperformed with 88.89% accuracy. This means that 88.89% of the predictions made by the RF algorithm were correct. Among all tested algorithms, Random Forest (RF) stood out with the lowest classification error rate at 9.84%. This translates to only a small fraction (less than 10%) of the model's predictions being inaccurate, as visualized in the chart below.

4.3. Result of stacked ML classifiers

Stacking works by training a secondary model, or "meta-learner," on top of several pre-trained ML models. This meta-learner learns how to optimally combine the predictions of the underlying models, resulting in more accuracy. The algorithms used for stacking are RF, multi-layer perceptron, KNN, support vector classifier, extra tree classifier, DT, extended gradient boosting, stochastic gradient descent, AdaBoost, and gradient boosting. The performance metrics are evaluated and tabulated in Table 5.

All the 13 features are passed to the stacked model. The stacked model achieved an accuracy of 86%, which is lower than the individual ML techniques. The precision, recall, sensitivity, specificity, Log_Loss, F1 score and Matthew correlation coefficient were also lower for the stacked model, and the results were also not as good as the individual ML techniques. To overcome this issue, GWO algorithm optimization is performed on the dataset. The optimized features are then applied to the ML techniques for training and stacking. GWO algorithm is a technique that can be used for performance improvement of ML models. It works by randomly generating a population of solutions, and then iteratively evaluating and improving the solutions until a satisfactory solution is found. In this case, the GWO algorithm is used to optimize the features of the dataset, which results in a better-performing ML model.

4.4. Results of optimized stacked ML classifiers

The limitations inherent to individual machine learning algorithms may not be overcome by a single approach. Combining multiple models can potentially improve overall prediction accuracy. While individual machine learning algorithms can achieve satisfactory performance on certain datasets, their inherent biases and limitations can be mitigated by ensemble learning techniques, which combine the predictions of multiple models to achieve greater accuracy. The drawbacks of stacked ML methods are overcome with the proposed model of GWO-optimized stacked ensemble techniques. To improve model performance, feature optimization was performed on the 13-dimensional dataset using the GWO algorithm. The features are reduced from 13 to 9. After optimization, the attributes selected are chest pain type, age, fasting blood sugar, resting blood pressure, ST slope, cholesterol, resting ECG, ca, and thal stress rate. The less important features are removed. The optimized features were then trained and passed to the stacked ensemble methods. The results achieved by the model proposed were much better than the methods and stacked methods. The performance evaluation for the proposed model is specified in Table 6. The model proposed achieved an accuracy of 93%, precision of 91%, sensitivity of 95%, specificity of 87%, F1 score of 93%, ROC of 91%, Log_Loss of 2.87, and Matthew correlation coefficient of 83%. The proposed model outperformed all other state-of-the-art methods in terms of specificity, precision, accuracy, sensitivity, F1 score, ROC, Log_Loss, and Matthew correlation coefficient.

The bar graph in Figure 4 shows the performance of different ML models for predicting a certain outcome. The bar graph indicates that the suggested model outperforms all other models. The model has an accuracy of 93% compared to other models, which is visualized in Figure 4. The other models shown in the bar graph are all state-of-the-art ML models. The proposed GWO-optimized stacked ensemble model combines multiple machine-learning models that learn from the strengths of each model. The bar graph shows that the model proposed has an important improvement

Table 6. Performance evaluation of the proposed model with other ML techniques.

Model	Accuracy	Precision	Sensitivity	Specificity	F1 Score	ROC	Log_Loss	mathew_corrcoef
Proposed Model	0.93	0.91	0.95	0.87	0.93	0.91	2.87	0.83
MLP	0.82	0.80	0.90	0.71	0.85	0.80	6.33	0.63
KNN	0.83	0.82	0.90	0.75	0.86	0.82	5.73	0.66
Extra tree classifier	0.90	0.88	0.98	0.82	0.92	0.90	3.17	0.82
XGB	0.89	0.89	0.92	0.85	0.91	0.89	3.77	0.78
SVC	0.76	0.76	0.83	0.66	0.79	0.74	8.45	0.50
SGD	0.76	0.85	0.69	0.84	0.76	0.77	8.45	0.53
Adaboost	0.83	0.83	0.88	0.77	0.86	0.83	5.73	0.66
CART	0.85	0.86	0.88	0.81	0.87	0.85	5.13	0.70
GBM	0.86	0.85	0.93	0.78	0.89	0.85	4.68	0.72

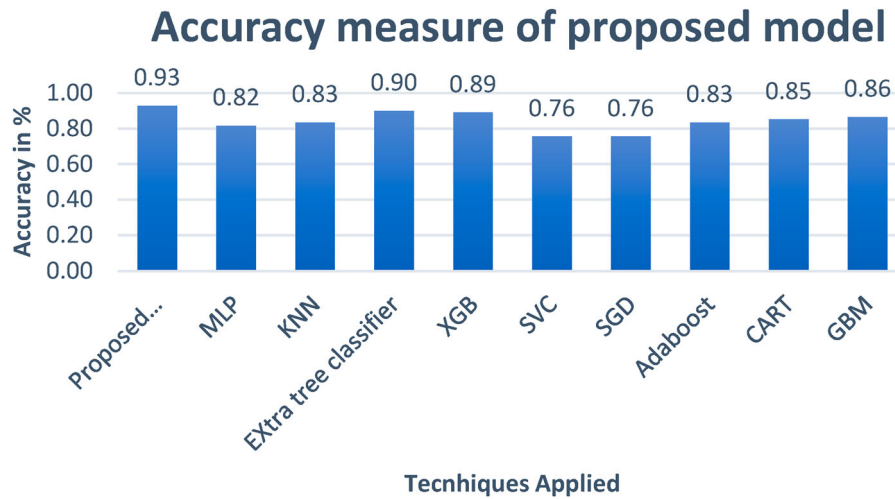


Figure 4. Accuracy for various models with the proposed model.

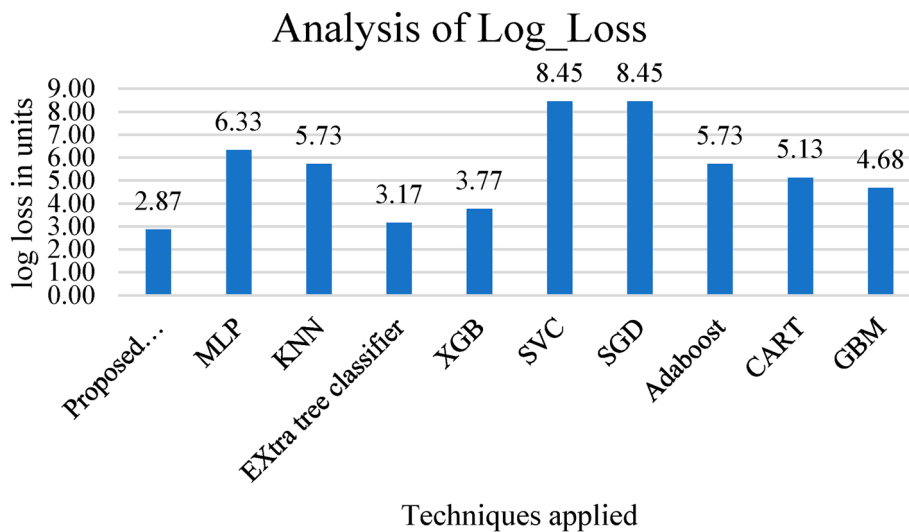


Figure 5. Log_Loss of various models with the proposed model.

over traditional ML models for predicting the outcome, as they are more accurate, precise, and reliable. This means the proposed model is more likely to make correct predictions.

By leveraging the power of ensemble learning, the proposed model exhibits remarkable adaptability. Figure 5 shows that incorporating more stacked models effectively reduces its Log_Loss, leading to superior heart disease prediction. The proposed model with 10 ensemble techniques has a Log_Loss of 2.87. This is a substantial enhancement over the Log_Loss of the traditional ML models, which ranged from 3.0 to 3.5. The proposed model with 17 models is also more accurate than the stacked ensemble model, which had a Log_Loss of 2.87. The bar graph in Figure 6 shows that the proposed model is more effective than the other stacked model and state-of-the-art ML methods without optimization. Focusing on heart disease prediction, Figures 4 and 5, and 6 reveal a substantial advantage for the proposed model compared to both traditional state-of-the-art ML models and stacked ensemble models.

4.5. Discussion

Table 6 reveals the proposed model's remarkable achievement of accuracy compared to every other model in Table 7. The proposed model of GWO-optimized stacked ensemble techniques is a significant improvement over traditional ML methods and stacked methods. By optimally combining and refining diverse ML models using the GWO algorithm, this model unlocks previously unattainable levels of performance and prediction accuracy. The proposed model is more accurate and reliable, with lower log loss. Figure 6 compares the projected model to the state-of-the-art methods and stacked ensemble ML techniques. The Y-axis of the graph displays the accuracy of each model, and the X-axis displays the different models being compared. The graph shows that the projected model beats other models on performance metrics. The proposed model has a higher accuracy, precision, recall, F1 score, Log_Loss, and Matthew correlation coefficient. The continual decrease of parameter space and

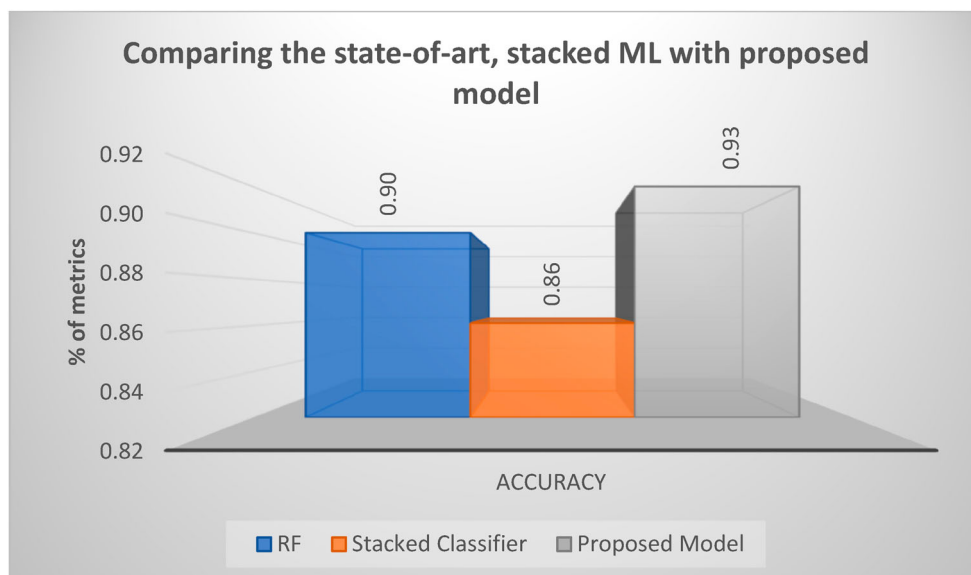


Figure 6. Performance metrics of the proposed model with stacked classifier.

Table 7. Comparative Analysis of the proposed model with the models available.

Models	Accuracy
Proposed Model	93%
GWO-SVM (Qasem Al-Tashi et al.,) [6]	89.83%
BGWOPSO (Al-Tashi et al.,) [28]	85%
ML (Garavand et al.,) [8]	87%
Backward feature selection + DT (Dissanayake & Johar) [9]	88.52%
FS + SMO (Sabab et al.,) [10]	87.8%
GWO + NB (El Bakrawy) [7]	87.45%
92HGDGWO (Kitonyi, & Segeera) [29]	80%
NB + FS (Nassif et al.,) [30]	84%
Fisher score algorithm + SVM (Saqlain et al.,) [11]	81.91%
Hybrid (NB, BN, RF, and MP) (Latha & Jeeva) [12]	85.45%
Modified GWO + LR (Mohiddin et al.,) [44]	86.91%
Modified GWO + KNN (Mohiddin et al.,) [44]	87.46%

reduced design alternatives, only two control parameters, along with its ability to avoid local minima, leads to faster convergence. These properties ensure the grey wolf metaheuristic algorithm is very robust and stable.

5. Conclusion

Heart disease is a leading cause of death globally, and early prediction is crucial for improving patient outcomes. The study proposed a stacked ensemble model optimized by GWO for predicting heart disease. GWO helps select the most important features from the dataset, while stacked ensemble learning combines multiple machine learning models to improve accuracy. The proposed model achieves an accuracy of 93%, significantly higher than traditional methods like logistic regression 85.25% and support vector machines 70.49%. The model also shows high precision 91%, recall 95.3%, F1-score 92.9%, Matthew coefficient 0.83, and a low Log_Loss 2.87. This study demonstrates the potential of GWO-optimized stacked ensemble models for improving the accuracy of heart disease prediction. This could lead to earlier diagnosis, better treatment

outcomes for patients, and a promising new approach for predicting heart disease with high accuracy.

Future scope: The study suggests exploring other ensemble techniques and different evaluation metrics, as well as analyzing feature importance to understand which features are most predictive of heart disease. They also recommend comparing the proposed model to various state-of-the-art methods for a more comprehensive evaluation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The data considered for the study is taken from the UCI repository Cleveland dataset and can be accessed from (<https://archive.ics.uci.edu/dataset/45/heart+disease>)

Authors' contributions

Geetha Narasimhan analyzed and interpreted the patient data regarding the heart disease and was a major contributor in writing the manuscript. Akila Victor has read and approved the final manuscript.

References

- [1] World Health Organization: WHO. (2018, August 27). *Cardiovascular diseases*. <https://www.who.int/westernpacific/health-topics/cardiovascular-diseases>.
- [2] Kochanek KD, Xu J, Arias E. (2022). *Mortality in the United States, 2021*. <https://doi.org/10.15620/cdc:122516>.
- [3] Dominic V, Gupta D, Khare S. An effective performance analysis of machine learning techniques for cardiovascular disease. *Appl Med Inform.* 2015;36(1):23–32.
- [4] Detrano J, Steinbrunn P, Schmid S, et al. *Heart Disease. UCI Machine Learning Repository. 1989*. Retrieved

- December 18, 2023, from <https://archive.ics.uci.edu/dataset/45/heart+disease>.
- [5] Verma L, Srivastava S, Negi PC. A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data. *J Med Sys*. 2016;40:1–7. doi:10.1007/s10916-016-0536-z
 - [6] Al-Tashi Q, Rais H, Jadid S. Feature selection method based on grey wolf optimization for coronary artery disease classification. In: *Recent trends in data science and soft computing: proceedings of the 3rd international conference of reliable information and communication technology (IRICT 2018)*. Springer International Publishing; 2019. p. 257–266.
 - [7] El Bakrawy LM. Grey wolf optimization and naive Bayes classifier incorporation for heart disease diagnosis. *Aust J Basic Appl Sci*. 2017;11(7):64–70.
 - [8] Garavand A, Salehnasab C, Behmanesh A, et al. Efficient model for coronary artery disease diagnosis: a comparative study of several machine learning algorithms. *J Healthc Eng*. 2022;2022. doi:10.1155/2022/5359540
 - [9] Dissanayake K, Md Johar MG. Comparative study on heart disease prediction using feature selection techniques on classification algorithms. *Appl Comput Intell Soft Comput*. 2021;2021:1–17.
 - [10] Sabab SA, Munshi MAR, Pritom AI. Cardiovascular disease prognosis using effective classification and feature selection technique. In: 2016 international conference on medical engineering, health informatics and technology (MediTec). IEEE; 2016, December. p. 1–6.
 - [11] Saqlain SM, Sher M, Shah FA, et al. Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines. *Knowl Inf Syst*. 2019;58:139–167. doi:10.1007/s10115-018-1185-y
 - [12] Latha CBC, Jeeva SC. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Inform Med Unlocked*. 2019;16:100203. doi:10.1016/j.imu.2019.100203
 - [13] Itoo NN, Garg VK. Heart disease prediction using a stacked ensemble of supervised machine learning classifiers. In: 2022 International mobile and embedded technology conference (MECON). IEEE; 2022, March. p. 599–604.
 - [14] Liu J, Dong X, Zhao H, et al. Predictive classifier for cardiovascular disease based on stacking model fusion. *Processes*. 2022;10(4):749. doi:10.3390/pr10040749
 - [15] Pal GK, Gangwar S. Heart disease prediction by stacking ensemble models on multiple classifiers by applying feature selection methods. *J Theor Appl Inf Technol*. 2022;100(23).
 - [16] Pa S, Priya SM. A novel approach based on voting ensemble and pca dimensionality reduction method for the prediction of heart disease. *J Theor Appl Inf Technol*. 2022;100(24).
 - [17] Harika N, Swamy SR, Nilima. Artificial intelligence-based ensemble model for rapid prediction of heart disease. *SN Comput Sci*. 2021;2(6):431. doi:10.1007/s42979-021-00829-9
 - [18] Karadeniz T, Tokdemir G, Maraş HH. Ensemble methods for heart disease prediction. *New Gener Comput*. 2021;39(3-4):569–581. doi:10.1007/s00354-021-00124-4
 - [19] Beunza JJ, Puertas E, García-Ovejero E, ... Landecho MF. Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease). *J Biomed Inform*. 2019;97:103257. doi:10.1016/j.jbi.2019.103257
 - [20] Taha K, Ross HJ, Peikari M, et al. An ensemble-based approach to the development of clinical prediction models for future-onset heart failure and coronary artery disease using machine learning. *J Am Coll Cardiol*. 2020;75(11_Supplement_1):2046–2046. doi:10.1016/S0735-1097(20)32673-5
 - [21] Chiu CC, Wu CM, Chien TN, et al. Applying an improved stacking ensemble model to predict the mortality of ICU patients with heart failure. *J Clin Med*. 2022;11(21):6460. doi:10.3390/jcm11216460
 - [22] Verma AK, Pal S. Prediction of skin disease with three different feature selection techniques using the stacking ensemble method. *Appl Biochem Biotechnol*. 2020;191(2):637–656. doi:10.1007/s12010-019-03222-8
 - [23] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. *Advances Eng Softw*. 2014;69:46–61. doi:10.1016/j.advengsoft.2013.12.007
 - [24] Emmanuel DADA, Joseph S, Oyewola D, et al. Application of grey wolf optimization algorithm: recent trends, issues, and possible horizons. *Gazi Univ J Sci*. 2021;35(2):485–504. doi:10.35378/gujs.820885
 - [25] Wang Q, Ma Y, Zhao K, et al. A comprehensive survey of loss functions in machine learning. *Ann Data Sci*. 2020; 1–26.
 - [26] Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over the F1 score and accuracy in binary classification evaluation. *BMC Genom*. 2020;21(1):1–13. doi:10.1186/s12864-019-6413-7
 - [27] Jangle P, Narayankar S. Alternating decision trees for early diagnosis of heart disease. *Int J Eng Comput Sci*. 2016;05(16070):16070–16072. doi:10.18535/ijecs/v5i4.02
 - [28] Al-Tashi Q, Kadir SJA, Rais HM, et al. Binary optimization using hybrid grey wolf optimization for feature selection. *IEEE Access*. 2019;7:39496–39508. doi:10.1109/ACCESS.2019.2906757
 - [29] Kitonyi PM, Segera DR. Hybrid gradient descent grey wolf optimizer for optimal feature selection. *BioMed Res Int*. 2021;2021. doi:10.1155/2021/2555622
 - [30] Nassif AB, Mahdi O, Nasir Q, et al. Machine learning classifications of coronary artery disease. In: 2018 International joint symposium on artificial intelligence and natural language processing (ISAI-NLP). IEEE; 2018, November. p. 1–6.
 - [31] Fjellström C, Nyström K. Deep learning, stochastic gradient descent and diffusion maps. *J Comput Math Data Sci*. 2022;4:100054. doi:10.1016/j.jcmds.2022.100054
 - [32] Sun S, Cao Z, Zhu H, et al. A survey of optimization methods from a machine learning perspective. *IEEE Trans Cybern*. 2019;50(8):3668–3681. doi:10.1109/TCYB.2019.2950779
 - [33] Gong S, Nong Q, Xiao H, et al. Uncertainty in study of social networks: robust optimization and machine learning. *Asia-Pac J Oper Res*. 2023;40(01):2340006. doi:10.1142/S0217595923400067
 - [34] Yan Y. Enhance robustness of machine learning with improved efficiency. In: Proceedings of the AAAI conference on artificial intelligence. 2023, June. Vol. 37, No. 13, p. 15461–15461.
 - [35] Zhang C, Wang Z, Wang X. Machine learning-based data-driven robust optimization approach under uncertainty. *J Process Control*. 2022;115:1–11. doi:10.1016/j.procont.2022.04.013

- [36] Hong LJ, Huang Z, Lam H. Learning-based robust optimization: procedures and statistical guarantees. *Manag Sci.* 2021;67(6):3447–3467. doi:10.1287/mnsc.2020.3640
- [37] Ben-Tal A, Nemirovski A. Lectures on stochastic programming. Princeton University Press; 2006.
- [38] Pilanci M, Wainwright MJ. The perils of optimization for robustness to outlier data. *IEEE Trans Inf Theory.* 2019;67(8):5209–5234.
- [39] Wang Z, Tu L. Theoretical foundations of robust optimization for machine learning. *J Mach Learn Res.* 2016;17(1):1–37.
- [40] Talukdar J, Singh TP. Early prediction of cardiovascular disease using artificial neural network. *Paladyn J Behav Robotics.* 2023;14(1):20220107. doi:10.1515/pjbr-2022-0107
- [41] Taylan O, Alkabaa AS, Alqabbaa HS, et al. Early prediction in classification of cardiovascular diseases with machine learning, neuro-fuzzy and statistical methods. *Biology (Basel).* 2023;12(1):117.
- [42] Hossain MI, Maruf MH, Khan MAR, et al. Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison. *Iran J Comput Sci.* 2023: 1–21.
- [43] Jawalkar AP, Swetcha P, Manasvi N, et al. Early prediction of heart disease with data analysis using supervised learning with stochastic gradient boosting. *J Eng Appl Sci.* 2023;70(1):122. doi:10.1186/s44147-023-00280-y
- [44] Mohiddin SK, Peteti S, Swathi T, et al. A Modified Grey Wolf Optimizer algorithm for feature selection to predict heart diseases. *IJFANS Int J Food Nutr Sci.* 2023;2319:1775. doi:10.48047/IJFANS/V11/I12/180