

An explainable analytical approach to heart attack detection using biomarkers and nature-inspired algorithms

Maithri Bairy^a, Krishnaraj Chadaga^a, Niranjana Sampathila^b, R. Vijaya Arjunan^{a,*}, G. Muralidhar Bairy^b

^a Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India

^b Department of Biomedical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India

ARTICLE INFO

Handling editor: Madijd Tavana

Keywords:

Heart attack prediction
Explainable machine learning
Feature selection analytics
Predictive healthcare modeling
Data-driven risk assessment
Computational clinical insights

ABSTRACT

Heart attacks are among the leading causes of death globally, and the earliest possible identification of at-risk patients is critical to lowering deaths. Advanced machine learning and deep learning algorithms have been effectively used to predict the presence of heart attack based on clinical and laboratory markers. This study used five explainable artificial intelligence techniques (XAI) to ensure that predictions made by the model are understandable and interpretable to facilitate clinical decisions. Fourteen nature-inspired feature selection algorithms were applied to identify the most informative markers while optimizing the predictive models for greater accuracy and reliability. Mutual information achieved a maximum testing accuracy of 90 % and highest precision of 94 %. The Whale Optimization Algorithm, Jaya Algorithm, Grey Wolf Optimizer and Sine Cosine Algorithm were the next best performing algorithms. The XAI results showed that the most important markers were ST slope, Oldpeak, exercise-induced angina, chest pain type, and fasting blood sugar. These models can be implemented in healthcare institutions to predict heart attack risks early, allowing timely interventions to reduce the likelihood of severe cardiovascular diseases. By supporting healthcare professionals with computer-aided diagnostic tools, these systems can enhance patient-specific decision-making while alleviating strain on healthcare resources.

1. Introduction

Cardiovascular diseases, particularly heart attacks, are one of the main health burdens worldwide and remain the leading cause of death [1]. Although medical science has shown significant progress in improving diagnostic and treatment protocols, the onset and progression of a heart attack cannot be predicted in patients with risk factors for hypertension, diabetes, or high cholesterol [2]. Thus, early warning of adverse cardiac events can allow early intervention in the hope that there will be better chances of patient survival and treatment [3]. In the field of healthcare, predictive analytics has shown great promise with machine learning and deep learning approaches for identifying patterns in highly complex datasets and producing predictive forecasting [4].

Explainable AI furthers this by providing transparency in the decision framework of the Machine learning (ML) model, which enables clinicians to better understand predictions and make informed decisions [5]. XAI techniques have been widely applied in healthcare to interpret

predictions for COVID-19, diabetes, cancer, and cardiovascular diseases [6–9]. It has recently been shown that ML models are applicable for predicting outcomes for heart diseases using markers that include cholesterol levels, resting blood pressure, and maximum heart rate [10, 11]. Prominent changes in these markers are often expressed early before the manifestation of clinical symptoms; hence, early detection is crucial.

Researchers have developed several innovative approaches for predicting coronary heart disease (CHD) using hybrid ML algorithms [12] along with deep learning models optimized with feature engineering techniques [13]. These models leverage a range of clinical data, including blood pressure, cholesterol levels, and age, to make accurate predictions. Feature selection methods have also been studied to enhance the performance of ML-based models for heart disease prediction [14]. Additionally, a deep learning-based clinical decision support system (CDSS) was designed, which is a much more efficient and reliable method for assisting healthcare providers [15]. Another notable

* Corresponding author.

E-mail address: vijay.arjun@manipal.edu (R.V. Arjunan).

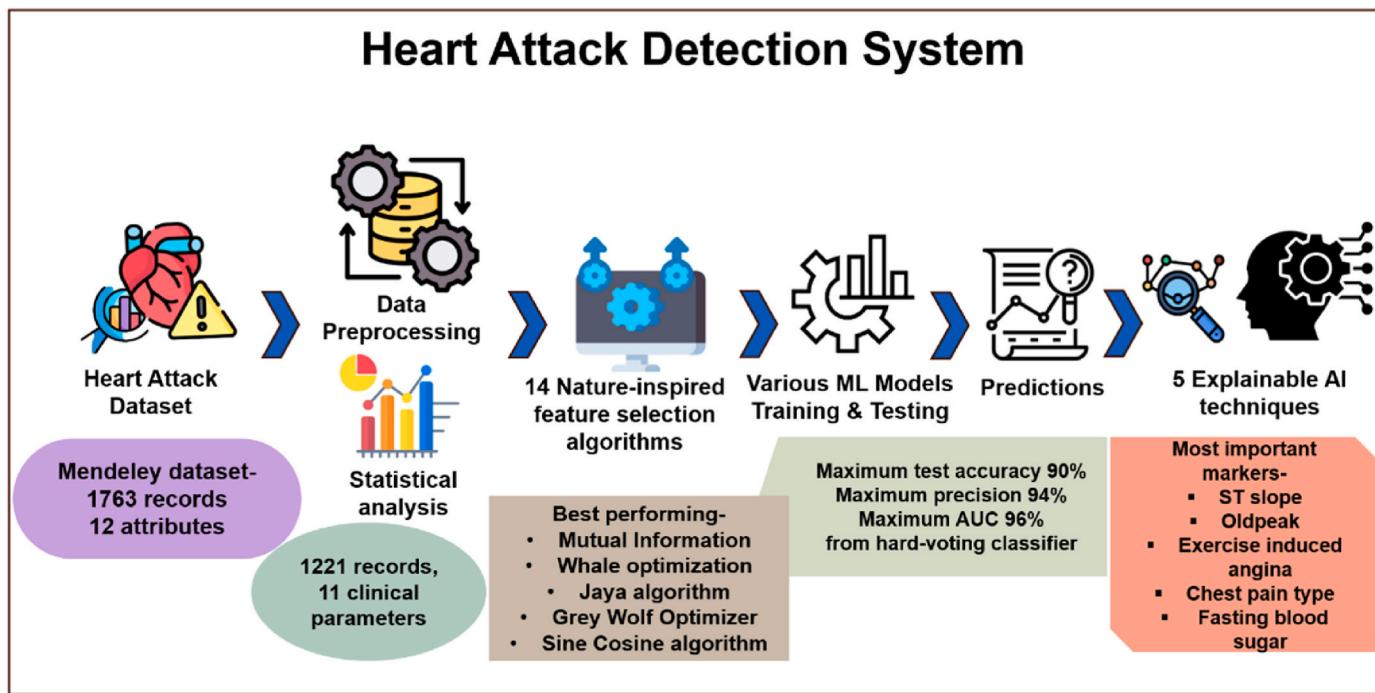


Fig. 1. Heart attack detection system.

Table 1
Dataset features.

Sl. No.	Variable	Description	Type
1	Gender	Gender of the patient	Categorical
2	CP (Chest Pain type)	Describes the type of chest pain experienced	Categorical
3	FBS (Fasting Blood Sugar)	1 if fasting blood sugar > 120 mg/dl, otherwise 0	Categorical
4	Rest ECG (Resting Electrocardiogram)	Resting electrocardiographic results	Categorical
5	Exng (Exercise Induced Angina)	1 if exercise-induced angina is present, otherwise 0	Categorical
6	ST Slope	Represents the slope of the peak exercise ST segment	Categorical
7	Label	Target variable: 1 indicates presence of heart disease, 0 indicates absence	Categorical
8	Age	Indicates the age of the patient	Continuous
9	TRTBPS (Resting Blood Pressure)	Resting blood pressure (in mm Hg) on admission to hospital	Continuous
10	Chol (Serum Cholesterol)	Serum cholesterol measured in mg/dl	Continuous
11	Thalachh (Maximum Heart Rate Achieved)	Maximum heart rate achieved	Continuous
12	Oldpeak	ST depression induced by exercise relative to rest	Continuous

approach includes crow-intelligence-optimization-based deep learning models for heart disease classification, which are optimized to improve prediction accuracy through process optimization [16]. A recent study by Alshraideh et al. utilized the Jordan University Hospital dataset and applied PSO-based feature selection with various ML models, achieving 94.3 % accuracy using SVM, though with limited focus on explainability frameworks [17]. Taken together, these approaches highlight the growing potential of ML and AI techniques to transform cardiovascular healthcare, providing more accurate, interpretable, and actionable insights for clinicians.

While much work has been done on heart attack prediction models in machine learning and deep learning, most of these studies have been

focused on maximizing predictive accuracy without much emphasis on interpretability and clinical applicability. Others, more or less from an explainability perspective, compromise on the predictive performance or rarely combine strong feature selection methods with their approaches. This separation between predictive modeling and explainability remains a major bottleneck in the widespread acceptance of predictive models as clinical tools, wherein accuracy and transparency are equally necessary for trust and decision-making.

To this end, our study presents a comprehensive set of solutions to produce a well-optimized feature set through multiple advanced ML algorithms, integrated with five XAI methods. This holistic approach, apart from increasing prediction accuracy, also provides transparent and interpretable insights into a model's decisions so that clinicians can understand and thus trust the outputs. By going beyond prior attempts to bridge high-performance prediction and explainability, our work enables proper clinical implementation of an early assessment methodology for heart attack risk. This convergence of prediction and interpretability will ultimately foster better clinical decisions and patient outcomes.

In this regard, an appropriate heart attack dataset containing patient attributes was preprocessed and then trained on various machine learning classifiers, optimizing them for enhanced predictive accuracy. The integration of explainable artificial intelligence (XAI) tools enables the interpretation of models' decision-making processes and identifies the key features that contribute to their predictions. This transparency allows clinicians to gain insight into the rationale behind risk stratification. By integrating ML predictions with clinical insights, this study provides a practical framework for the early identification of high-risk patients, which will help in timely interventions and personalized treatment strategies in cardiovascular care.

The significant contributions of this study are outlined below.

- Develop an explainable machine learning model for heart attack prediction.
- Optimize predictive accuracy using nature-inspired feature selection.
- Identify key clinical markers influencing heart attack risk.
- Achieve 90 % accuracy with mutual information for feature selection.

Table 2
Descriptive statistical parameters.

	Label	Number of instances	Marker mean	Marker median	Standard deviation	Variance	Inter quartile range	Range	Minimum value	Maximum value
Age	Absence	574	51.132	52	9.5	90.258	13.75	48	28	76
	Presence	647	56.056	57	8.563	73.325	11	46	31	77
TRTBPS (Resting Blood Pressure)	Absence	574	129.915	130	16.407	269.195	20	110	80	190
	Presence	647	134.267	132	19.592	383.856	25	200	0	200
Chol (Serum Cholesterol)	Absence	574	231.556	230.5	69.557	4838.16	66	564	0	564
	Presence	647	192.168	225	118.425	14024.52	127.5	603	0	603
Thalachh (Maximum Heart Rate Achieved)	Absence	574	151.073	154	22.657	513.349	31	133	69	202
	Presence	647	130.148	129	23.687	561.052	33.5	135	60	195
Oldpeak	Absence	574	0.459	0	0.728	0.53	0.8	5.3	-1.1	4.2
	Presence	647	1.339	1.2	1.191	1.419	1.9	8.8	-2.6	6.2

Table 3
Independent samples T-Tests results.

Independent Samples T-Test			
Marker	Statistic	df	P
age	Student's t	-9.52 ^a	1219
trestbps	Student's t	-4.18 ^a	1219
cholesterol	Student's t	6.97 ^a	1219
max heart rate	Student's t	15.72	<0.001
oldpeak	Student's t	-15.33 ^a	1219

Note. $H_0 \mu_0 \neq \mu_1$.

^a Levene's test is significant ($p < 0.05$), suggesting a violation of the assumption of equal variances.

Table 4
Chi-square test results.

χ^2 Tests			
χ^2 Value	df	p	Marker
112	1	<0.001	gender
301	4	<0.001	chest pain type
59	1	<0.001	fasting blood sugar
18.7	2	<0.001	resting ecg
284	1	<0.001	exercise angina
315	3	<0.001	ST slope

- Support healthcare professionals with interpretable predictive analytics.

The remainder of this paper is organized as follows. Section 2 provides the literature review and Section 3 explains the materials and methodology used. Section 4 provides a detailed analysis of the main findings. The interpretation and discussion of these results are presented in section 5, and the paper concludes with key takeaways in section 6.

The heart attack detection method used in this study is shown in Fig. 1.

2. Literature review

Nature-inspired algorithms are increasingly used for feature selection because of their ability to traverse complex search spaces and find informative features. Dyoub and Letteri showed methods of bio-inspired optimization on different chronic disease datasets and stressed the fact that with good selection of a feature subset, the system will work better [18]. Ahmad et al. further reiterated such importance in diagnosing heart disease and explained how a combination of methods from nature can accelerate the learning phase of an optimization model and yield optimized model performance [19].

More studies have emerged that apply nature-inspired algorithms to heart disease datasets. Natarajan et al. implemented the Firefly Algorithm for feature selection to achieve better classification accuracy

through a stacked ensemble model [20]. Dissanayake et al. proposed a meta-learning-based hybrid feature selection framework combining multiple algorithms to improve diagnostic accuracy and reduce feature dimensionality [21]. Similarly, Biswas et al. tried a variety of nature-inspired approaches for improving predictive modeling of early-stage heart disease using ML techniques [22].

The flexibility and adaptability of these algorithms make them well-suited to handle noisy and high-dimensional clinical data, a frequent challenge in heart disease prediction [23, 24]. Classic types of meta-heuristics that are used for feature selection include Genetic Algorithm, Particle Swarm Optimization, along with newly introduced algorithms of Reinforced Swarm Optimization and Frog-Snake Prey-Predation Optimization [25, 26]. Apart from heart diseases, nature-based algorithms have also been put into effective application in other domains of healthcare, such as in detecting COVID-19 from medical images, which attests to their versatility and efficiency in medical data analysis [27, 28]. These researches not only attest to the ability of nature-inspired algorithms to adapt to varied healthcare datasets but also to the improvement of early detection of diseases, a principle that directly applies to cardiovascular risk prediction.

3. Methods

3.1. Dataset description

The dataset was sourced from Mendeley and comprised 1763 patient records with 12 attributes related to heart disease, combining demographic, clinical, and diagnostic variables. It included age, sex, chest pain type, resting blood pressure, serum cholesterol levels, fasting blood sugar, and key diagnostic markers, such as maximum heart rate, exercise-induced angina, and ST depression. The target variable indicated the presence or absence of a heart disease. Their ages ranged from 29 to 77 years old. The mean age of the patients was 54 years, blood pressure at rest fell within 94–200 mmHg. The average serum cholesterol level was approximately 246 mg/dl. This dataset allowed for a mix of continuous and categorical variables, which reflected the diversity of different patient characteristics. It offers substantial potential to help develop predictive models, establish key risk markers, and improve diagnostic and preventive strategies for cardiovascular health. After preprocessing, the dataset included 574 records without heart attacks and 647 records with heart attacks. This study considered 11 clinical parameters, of which five were continuous and the rest were categorical. Table 1 lists the identified clinical markers.

3.2. Data preprocessing

The preprocessing of a dataset is an important step in machine learning. Duplicate records were detected and eliminated for data quality. The steps included in the preprocessing stage were encoding categorical attributes, scaling continuous variables, and data balancing. Missing values were not present in the dataset. Descriptive statistical

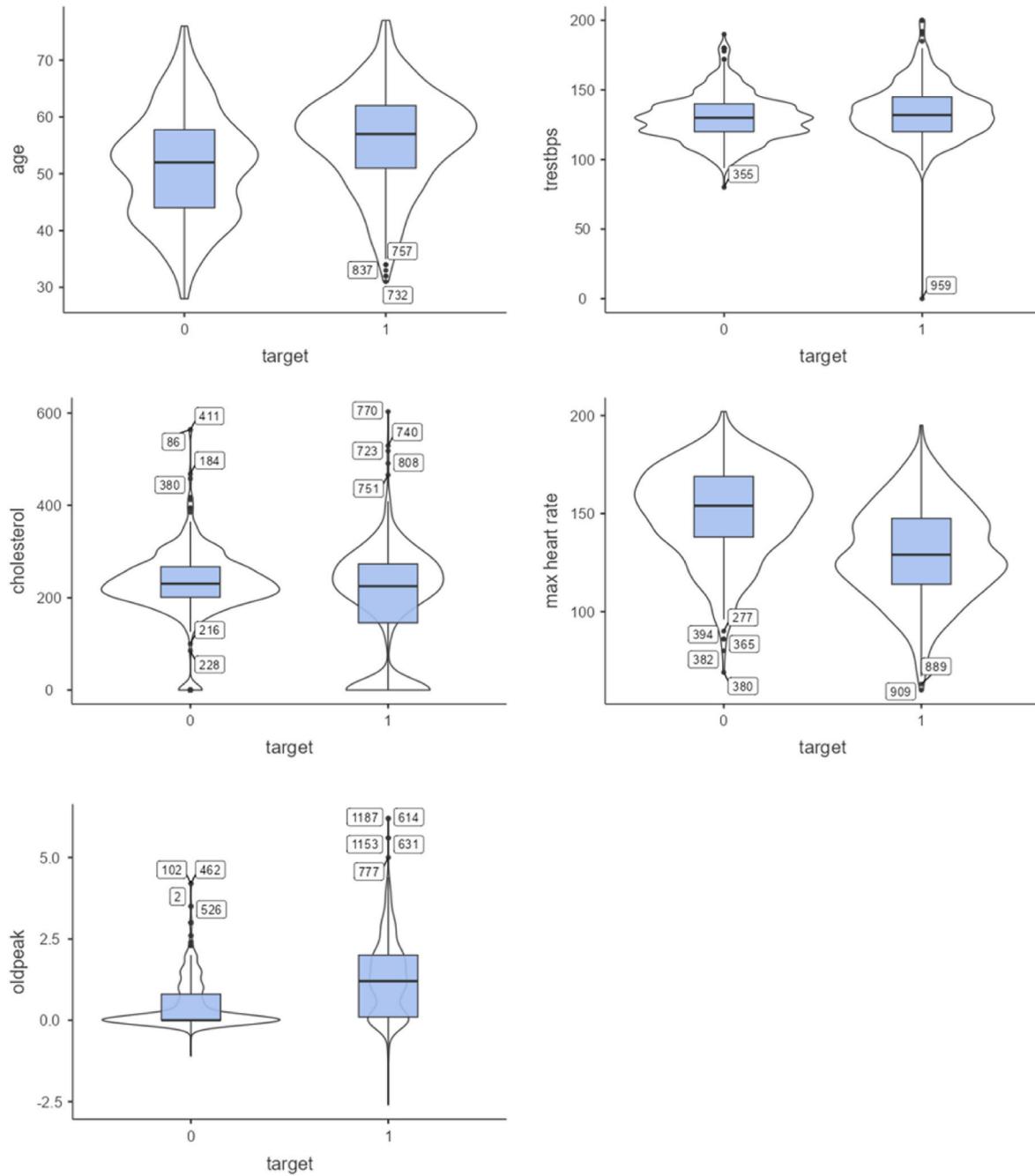


Fig. 2. Violin plots for continuous markers.

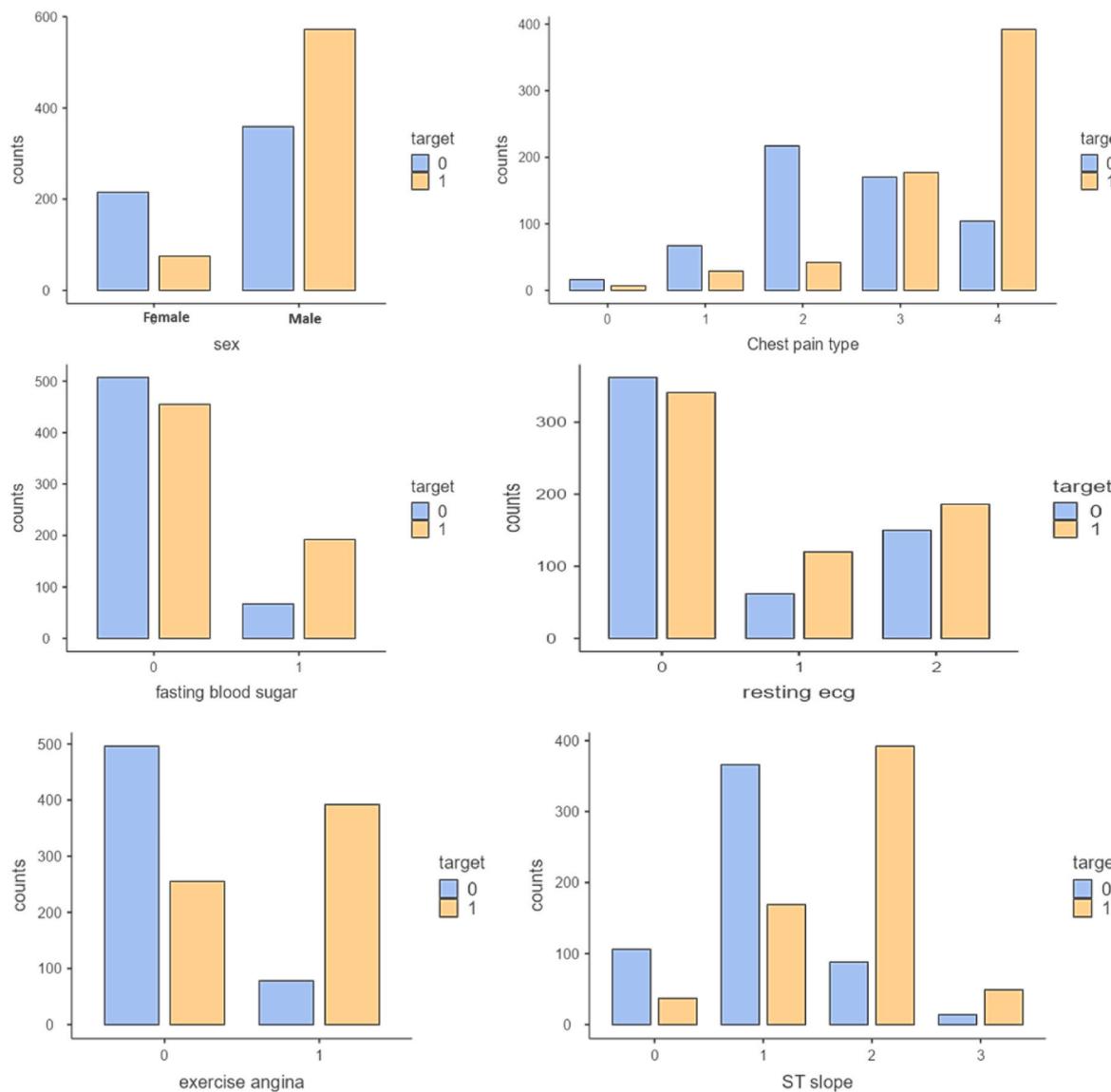


Fig. 3. Frequency distributions of the categorical attributes.

analysis was carried out using the open-source statistical software Jamovi, with the key statistical parameters summarized in Table 2. Independent Samples T-Tests and Chi-square tests were conducted to examine whether the differences in key clinical features between patient groups were statistically significant. The results show highly significant differences ($p < 0.001$). These p-values prove that the observed differences are unlikely due to chance and provide early evidence of strong feature-outcome associations. The p-value obtained by the t-tests are detailed in Table 3. The inferences obtained by the chi-square test are listed in Table 4. Violin plots were used to uncover notable patterns within the dataset, as illustrated in Fig. 2. The target value of 0 represents the absence of a heart attack, while a target value of 1 represents its presence. These visualizations revealed that the median age was higher among the patients who experienced a heart attack. Additionally, markers such as Resting Blood Pressure (TRTBPS) and Oldpeak were elevated in heart attack patients, highlighting their potential significance as predictors. Fig. 3 shows the counts of categorical variables denoting a heart attack or no heart attack in terms of bar plots. In total, 290 females and 931 males. The records without heart attacks included 215 females and 359 males, whereas the records with heart attacks included 75 females and 572 males.

Categorical values in the data must be encoded because classifiers

cannot work on text-based data. There are various methods of encoding [29]. In this study, one-hot encoding was utilized to encode the "Gender" attribute [30]. This approach prevents problems, such as ordinality, which may exist with categorical variables. The data were scaled using a standardization method. Standardization is crucial when the data values vary significantly, as it does not allow classifiers to bias toward parameters with larger values, irrespective of the units. Among the two widely used scaling methods, normalization and standardization [31], standardization was selected because it is efficient at handling outliers. The dataset was split at 70:30 between the training and testing sets. There was a slight imbalance in the dataset. This means that the output might be biased because the models favor the majority class. SMOTE [32] Borderline Synthetic Minority Oversampling Technique was used so that the models would generate synthetic samples using K-nearest neighbors algorithm in order to enhance the balancing of the training data. This also supports the borderline cases. Avoiding undersampling helped to avoid missing major trends and patterns. The testing dataset was kept unbalanced to maintain the data integrity.

Fourteen feature selection methods were used to select the most significant markers. A set of metaheuristic nature-inspired algorithms is used in this study. Such algorithms aim to optimize complex problems by simulating naturally observed processes, including natural selection,

Table 5
Features selected using the algorithms.

Sl. no	Feature selection method used	Number of features chosen	Features chosen
1	Whale Optimization (WO) [33]	6	Fasting Blood Sugar, Gender, Rest ECG, Oldpeak, Exercise Induced Angina, ST Slope
2	Sine Cosine Algorithm (SCA) [34]	4	Oldpeak, ST Slope, Fasting Blood Sugar, Exercise Induced Angina
3	Salp Swarm Optimization (SSO) [35]	6	Chest Pain type, Rest ECG, Exercise Induced Angina, Resting Blood Pressure, Fasting Blood Sugar, ST Slope
4	Particle Swarm Optimization (PSO) [36]	5	Gender, Oldpeak, Fasting Blood Sugar, Exercise Induced Angina, ST Slope
5	Jaya Algorithm (JA) [37]	7	Rest ECG, Exercise Induced Angina, Gender, Chest Pain type, Fasting Blood Sugar, Oldpeak, ST Slope
6	Harris Hawks Optimization (HHO) [38]	6	Gender, Exercise Induced Angina, Oldpeak, Chest Pain type, Rest ECG, ST Slope
7	Grey Wolf Optimizer (GWO) [39]	4	Fasting Blood Sugar, ST Slope, Exercise Induced Angina, Oldpeak
8	Genetic Algorithm (GA) [40]	5	Gender, Oldpeak, ST Slope, Fasting Blood Sugar, Exercise Induced Angina
9	Flower Pollination Algorithm (FPO) [41]	4	Gender, Exercise Induced Angina, Oldpeak, ST Slope
10	Firefly Algorithm (FA) [42]	5	Gender, Chest Pain type, Exercise Induced Angina, Oldpeak, ST Slope
11	Differential Evolution (DE) [43]	6	Gender, Chest Pain type, Fasting Blood Sugar, Exercise Induced Angina, Oldpeak, ST Slope
12	Cuckoo Search Algorithm (CSA) [44]	5	Gender, Fasting Blood Sugar, Exercise Induced Angina, Oldpeak, ST Slope
13	Bat Algorithm (BA) [45]	6	Gender, Chest Pain type, Resting Blood Pressure, Rest ECG, Exercise Induced Angina, ST Slope
14	Mutual Information (MI) [46]	11	Age, Gender, Rest ECG, Thalachh, Exercise Induced Angina, Oldpeak, Chest Pain type, Serum Cholesterol, Resting Blood Pressure, Fasting Blood Sugar, ST Slope

foraging behavior, and swarm intelligence. These are helpful when the problem space is complex or even nonlinear and fails to perform traditional optimizations. Feature selection is necessary because the performance of classifiers improves during the removal of redundant features. Table 5 lists the number of features selected by the different algorithms. Mutual information had the maximum number of features (11). The Jaya Algorithm selects seven features. The sine cosine algorithm, grey wolf optimizer, and flower pollination algorithm selected four features each. Harris Hawk's optimization, whale optimization, salp swarm optimization, differential evolution, and the bat algorithm selected six

features each. The markers chosen by the feature selection techniques are illustrated in Fig. 4. Exercise-induced Angina and ST slope were the most selected features because they were included in all 14 algorithms. The next three were Oldpeak, Gender and Fasting blood sugar, selected 13, 11, and 9 times, respectively.

3.3. Machine learning concepts

Machine learning (ML), which is a part of artificial intelligence, allows the prediction of outcomes from historical data. In this study, a

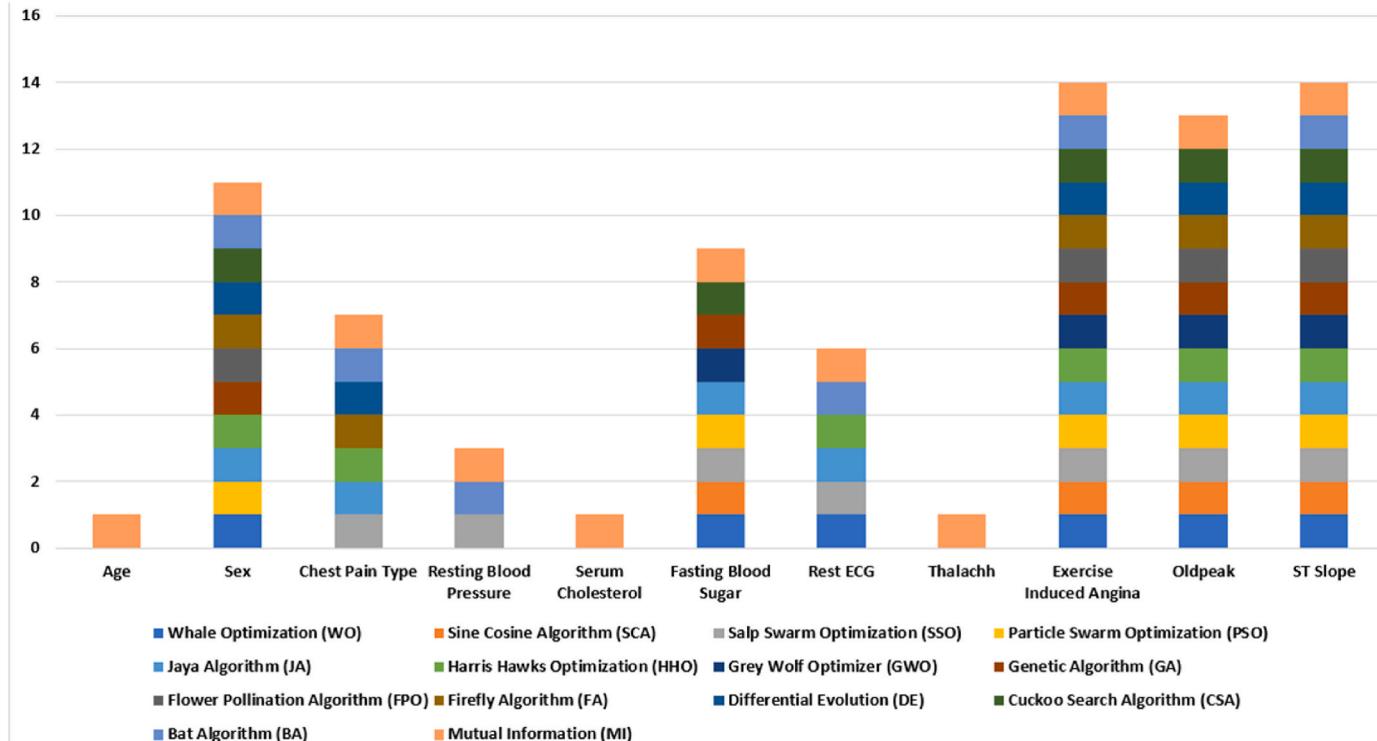


Fig. 4. Markers selected by the feature selection methods.

Table 6

Hyperparameters chosen.

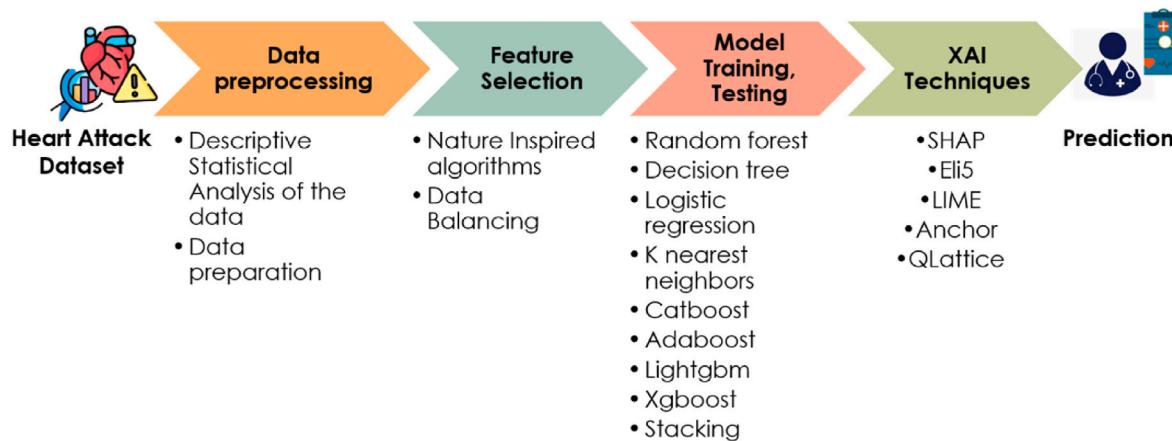
Mutual Information	
Algorithm	Hyperparameters
Random Forest	{'bootstrap': True, 'max_depth': 100, 'max_features': 2, 'min_samples_leaf': 3, 'min_samples_split': 8, 'n_estimators': 300}
Logistic Regression	{'C': 1, 'penalty': 'l2'}
Decision Tree	{'criterion': 'entropy', 'max_depth': 7, 'max_features': 'log2', 'min_samples_leaf': 5, 'min_samples_split': 30, 'splitter': 'best'}
KNN	{'n_neighbors': 3}
Adaboost	{'learning_rate': 0.1, 'n_estimators': 300}
Catboost	{'border_count': 32, 'depth': 3, 'iterations': 250, 'l2_leaf_reg': 1, 'learning_rate': 0.03}
Lightgbm	{'lambda_l1': 1, 'lambda_l2': 0, 'min_data_in_leaf': 30, 'num_leaves': 31, 'reg_alpha': 0.1}
Xgboost	{'colsample_bytree': 0.3, 'gamma': 0.2, 'learning_rate': 0.15, 'max_depth': 5, 'min_child_weight': 1}
Whale Optimization	
Random Forest	{'bootstrap': True, 'max_depth': 90, 'max_features': 3, 'min_samples_leaf': 3, 'min_samples_split': 8, 'n_estimators': 200}
Logistic Regression	{'C': 10, 'penalty': 'l2'}
Decision Tree	{'criterion': 'entropy', 'max_depth': 90, 'max_features': 'auto', 'min_samples_leaf': 8, 'min_samples_split': 10, 'splitter': 'best'}
KNN	{'n_neighbors': 8}
Adaboost	{'learning_rate': 0.1, 'n_estimators': 1000}
Catboost	{'border_count': 32, 'depth': 3, 'iterations': 250, 'l2_leaf_reg': 3, 'learning_rate': 0.03}
Lightgbm	{'lambda_l1': 1, 'lambda_l2': 0, 'min_data_in_leaf': 30, 'num_leaves': 31, 'reg_alpha': 0.1}
Xgboost	{'colsample_bytree': 0.3, 'gamma': 0.0, 'learning_rate': 0.05, 'max_depth': 3, 'min_child_weight': 3}
Jaya Algorithm	
Random Forest	{'bootstrap': True, 'max_depth': 80, 'max_features': 3, 'min_samples_leaf': 5, 'min_samples_split': 10, 'n_estimators': 300}
Logistic Regression	{'C': 10, 'penalty': 'l2'}
Decision Tree	{'criterion': 'gini', 'max_depth': 40, 'max_features': 'sqrt', 'min_samples_leaf': 1,

Table 6 (continued)

Mutual Information	
Algorithm	Hyperparameters
KNN	{'min_samples_split': 30, 'splitter': 'best'}
Adaboost	{'n_neighbors': 3}
Catboost	{'learning_rate': 0.01, 'n_estimators': 1000}
Lightgbm	{'border_count': 32, 'depth': 3, 'iterations': 250, 'l2_leaf_reg': 3, 'learning_rate': 0.03}
Xgboost	{'lambda_l1': 1, 'lambda_l2': 0, 'min_data_in_leaf': 50, 'num_leaves': 31, 'reg_alpha': 0.1}
Grey Wolf Optimization and Sine Cosine Algorithm	
Random Forest	{'bootstrap': True, 'max_depth': 110, 'max_features': 2, 'min_samples_leaf': 5, 'min_samples_split': 8, 'n_estimators': 300}
Logistic Regression	{'C': 1, 'penalty': 'l2'}
Decision Tree	{'criterion': 'entropy', 'max_depth': 7, 'max_features': 'sqrt', 'min_samples_leaf': 8, 'min_samples_split': 50, 'splitter': 'best'}
KNN	{'n_neighbors': 4}
Adaboost	{'learning_rate': 1.0, 'n_estimators': 100}
Catboost	{'border_count': 32, 'depth': 3, 'iterations': 250, 'l2_leaf_reg': 10, 'learning_rate': 0.03}
Lightgbm	{'lambda_l1': 0, 'lambda_l2': 0, 'min_data_in_leaf': 50, 'num_leaves': 31, 'reg_alpha': 0.1}
Xgboost	{'colsample_bytree': 0.3, 'gamma': 0.0, 'learning_rate': 0.05, 'max_depth': 3, 'min_child_weight': 7}

variety of ML classifiers were used to analyze and predict heart disease outcomes: Random Forest, Decision Tree, Logistic Regression, K-Nearest Neighbors, CatBoost, AdaBoost, XGBoost, Lightgbm, and ensemble methods, including Stacking and Voting algorithms. Stacking is a form of ensemble learning that uses a meta-learner to combine the predictions of multiple base models to improve overall model accuracy [47]. For this research, Logistic Regression was used as the meta-learner, and the aggregation of model predictions was performed based on their respective weights. The Voting Classifier is another ensemble technique that combines the predictions of several classifiers based on majority voting [48]. It can be applied in either of two ways: hard voting, where the class with the majority of votes wins outright, regardless of probability [49]. Soft voting involves choosing the class with the highest average probability across the classes [50]. Both these methods exploit the diversity of multiple classifiers to increase the likelihood of a correct prediction.

To evaluate the efficiency and robustness of the models, five-fold cross-validation was performed. This technique splits the dataset into five equal subsets, where four are used for training and one for testing. This process is repeated over all combinations to ensure full validation of

**Fig. 5.** Machine learning methodology used.**Table 7**

Precision values for various feature selection methods (In %).

Algorithm	BA	CSA	DE	FA	FPO	GA	GWO	HHO	JA	MI	PSO	SCA	SSO	WO
Adaboost	83	85	83	84	85	85	88	83	87	88	85	88	84	87
Catboost	85	85	84	84	85	85	88	86	88	90	85	88	84	86
Decision Tree	86	85	83	83	83	85	88	77	81	83	85	88	78	82
Hard-voting	84	85	84	84	84	85	86	84	82	94	85	86	80	83
KNN	87	83	86	84	85	83	85	81	79	84	83	85	81	80
Lightgbm	85	87	85	86	84	87	86	86	85	94	87	86	82	84
Logistic regression	86	81	83	85	86	81	84	86	86	89	81	84	83	85
Random forest	84	86	84	84	86	86	89	83	87	90	86	89	82	82
Soft-voting	79	83	79	80	84	83	82	77	80	92	83	82	81	80
Stacking	79	83	79	80	83	83	83	78	80	91	83	83	79	81
Xgboost	85	85	84	83	86	85	89	84	87	93	85	89	85	87

Table 8

Recall values for various feature selection methods (In %).

Algorithm	BA	CSA	DE	FA	FPO	GA	GWO	HHO	JA	MI	PSO	SCA	SSO	WO
Adaboost	82	78	73	75	79	78	70	77	77	80	78	70	74	85
Catboost	79	80	77	79	76	80	78	80	80	83	80	78	77	88
Decision Tree	70	77	67	72	75	77	71	72	73	76	77	71	76	85
Hard-voting	83	78	76	79	77	78	79	81	80	86	78	79	75	86
KNN	71	77	68	75	72	77	80	80	73	78	77	80	73	82
Lightgbm	81	77	77	78	78	77	80	79	81	86	77	80	75	84
Logistic regression	76	71	70	75	66	71	72	78	80	78	71	72	75	81
Random forest	81	79	73	79	77	79	78	80	79	84	79	78	75	86
Soft-voting	78	82	74	79	78	82	72	81	77	85	82	72	78	81
Stacking	78	81	76	77	78	81	71	77	78	84	81	71	77	84
Xgboost	80	80	77	80	78	80	74	80	82	84	80	74	77	88

the models. Hyperparameter tuning using grid search was used to identify optimal parameters, as the performance of the ML classifiers largely depends on the hyperparameters chosen. The hyperparameters chosen by the algorithms are summarized in [Table 6](#).

The models are compared using different classification and loss metrics. Emphasis was placed on precision and recall, as these metrics are crucial for minimizing false positives and negatives, particularly in critical healthcare applications. After training and testing the models, five XAI techniques were applied to interpret the model predictions. The results from these techniques were presented in graphical and tabular formats so that insights derived from the models could be understood and verified easily by healthcare professionals and researchers. [Fig. 5](#) illustrates the methodological process of this study.

4. Results

4.1. Model testing

In this study, various machine learning classifiers were trained and tested to predict heart attacks. [Table 7](#) presents the precision achieved by the models using various feature selection techniques. From the table below, it can be seen that the Hard-voting and Lightgbm classifiers attained the highest precision of 94 % with features selected using mutual information. Stacking and soft voting attained precisions of 91 % and 92 %, respectively. The Sine cosine and grey wolf optimizers also performed well. The random forest, stack, hard-voting, and soft-voting classifiers gave precisions of 89 %, 83 %, 86 %, and 82 %, respectively. The Cuckoo Search algorithm, genetic algorithm, and Particle swarm optimizer were also efficient. Stack, hard-voting, and soft-voting achieved precisions of 83 %, 86 %, and 82 %, respectively. The precision for stack, hard-voting, and soft-voting after using the Jaya algorithm

Table 9

Loss and classification metrics for the best five nature-inspired algorithms.

Algorithm	Accuracy (%)	F1-score (%)	AUC	AP	MCC	Log loss	Jaccard Score	Hamming loss
Mutual Information								
Adaboost	83	84	0.89	0.86	0.65	6.02	0.72	0.17
Catboost	85	86	0.93	0.94	0.7	5.08	0.76	0.15
Decision Tree	78	79	0.86	0.86	0.58	7.15	0.68	0.21
Hard-Voting	90	89	0.96	0.97	0.79	3.67	0.82	0.1
KNN	79	81	0.86	0.85	0.66	5.74	0.72	0.17
Lightgbm	89	90	0.95	0.96	0.77	3.95	0.81	0.11
Logistic regression	82	83	0.89	0.91	0.68	5.45	0.73	0.16
Random forest	86	87	0.94	0.95	0.77	3.95	0.8	0.12
Soft-Voting	88	89	0.95	0.96	0.78	3.86	0.81	0.11
Stacking	86	87	0.94	0.95	0.75	4.33	0.79	0.13
Xgboost	88	88	0.95	0.96	0.75	4.23	0.8	0.12
Whale Optimization								
Adaboost	86	86	0.9	0.9	0.63	6.4	0.7	0.19
Catboost	87	87	0.92	0.92	0.65	6.02	0.72	0.17
Decision Tree	82	83	0.88	0.88	0.58	7.34	0.66	0.21
Hard-Voting	83	84	0.9	0.9	0.65	6.11	0.71	0.18
KNN	80	81	0.87	0.84	0.62	6.68	0.68	0.19
Lightgbm	83	84	0.9	0.91	0.65	6.02	0.72	0.18
Logistic regression	82	83	0.9	0.91	0.54	7.9	0.64	0.23
Random forest	83	84	0.9	0.9	0.63	6.3	0.7	0.18
Soft-Voting	79	80	0.87	0.86	0.59	6.96	0.68	0.2
Stacking	81	82	0.87	0.86	0.59	6.96	0.68	0.2
Xgboost	86	87	0.91	0.92	0.64	6.21	0.71	0.18
Jaya Algorithm								
Adaboost	83	82	0.91	0.91	0.61	6.68	0.67	0.19
Catboost	85	84	0.92	0.94	0.68	5.55	0.72	0.16
Decision Tree	78	77	0.88	0.88	0.62	6.49	0.68	0.18
Hard-Voting	81	82	0.91	0.93	0.66	5.93	0.71	0.17
KNN	77	76	0.84	0.8	0.59	7.05	0.67	0.2
Lightgbm	84	83	0.92	0.93	0.67	5.55	0.73	0.16
Logistic regression	84	83	0.89	0.9	0.62	6.59	0.68	0.19
Random forest	83	82	0.92	0.93	0.65	5.92	0.71	0.17
Soft-Voting	79	79	0.89	0.9	0.64	6.21	0.7	0.18
Stacking	80	79	0.87	0.88	0.64	6.11	0.71	0.17
Xgboost	85	84	0.92	0.93	0.67	5.74	0.72	0.17
Grey Wolf Optimization and Sine Cosine Algorithm								
Adaboost	79	78	0.87	0.85	0.53	8.18	0.63	0.24
Catboost	83	83	0.88	0.88	0.58	7.25	0.67	0.21
Decision Tree	80	78	0.87	0.85	0.57	7.43	0.66	0.21
Hard-Voting	82	82	0.88	0.88	0.58	7.15	0.68	0.2
KNN	82	82	0.87	0.85	0.57	7.53	0.65	0.22
Lightgbm	83	83	0.88	0.87	0.58	7.25	0.68	0.21
Logistic regression	78	78	0.86	0.85	0.53	8.09	0.64	0.23
Random forest	83	83	0.89	0.89	0.58	7.15	0.68	0.2
Soft-Voting	77	77	0.84	0.85	0.54	7.99	0.64	0.23
Stacking	77	77	0.84	0.84	0.53	7.9	0.66	0.23
Xgboost	82	81	0.87	0.86	0.56	7.53	0.66	0.22

were 80 %, 82 %, and 80 %, respectively.

The recall values of the models for the different feature selection techniques are listed in [Table 8](#), with Whale Optimization identified as the top-performing method. Using this technique, the CatBoost, XGBoost and Hard-Voting classifiers achieved recalls of 88 %, 88 %, and 86 %, respectively. Mutual Information also performed well, yielding recalls of 84 %, 86 %, and 85 % for Stacking, Hard-Voting and Soft-Voting classifiers, respectively. The Bat Algorithm demonstrated good performance, with recalls of 83 %, 82 %, and 81 % for the Hard-Voting, AdaBoost and Random Forest classifiers, respectively. Using the Jaya algorithm, XGBoost, Lightgbm and Hard-Voting achieved recalls of 82 %, 81 %, and 80 %, respectively. The five best feature selection techniques were used for detailed analysis. They are mutual information, whale optimization, the Jaya algorithm, the Grey Wolf Optimizer and Sine Cosine algorithm.

The loss and classification metrics are listed in [Table 9](#). Among these methods, mutual information provided the best performance. The Stacking, Hard-Voting and Soft-Voting classifiers achieved accuracies of 86, 90, and 88 %, respectively. Whale Optimization also produced excellent results, with AdaBoost, CatBoost, Stacking, Hard-Voting, and Soft-Voting classifiers achieving accuracies of 86 %, 86 %, 81 %, 83 %,

and 79 %, respectively. The Jaya algorithm showed competitive performance, with accuracies of 85 %, 80 %, 81 %, and 79 % for the XGBoost, Stacking, Hard-Voting and Soft-Voting classifiers, respectively. The Grey Wolf Optimizer and Sine Cosine Algorithm achieved accuracies of 83 %, 83 %, and 82 % for the Random Forest, CatBoost and Hard-Voting classifiers, respectively.

The ROC curves of the stacking model in [Fig. 6](#) show that the model achieved the highest AUC, with a value of 0.96, during feature selection by mutual information. The precision-recall curves of the stacking classifier in [Fig. 7](#) show that the model achieved the highest average precision of 0.97, with the features selected by mutual information during training.

4.2. Explainable AI

In this study, five XAI methods were implemented to make the models more explainable: SHAP, LIME, QLattice, Eli5 and Anchor. The stacked model was used for interpretation because the results obtained were reliable. SHAP (SHapley Additive exPlanations) is used during the interpretation of machine learning models as it can quantify how relevant each feature is in yielding predictions from the model. Its use in

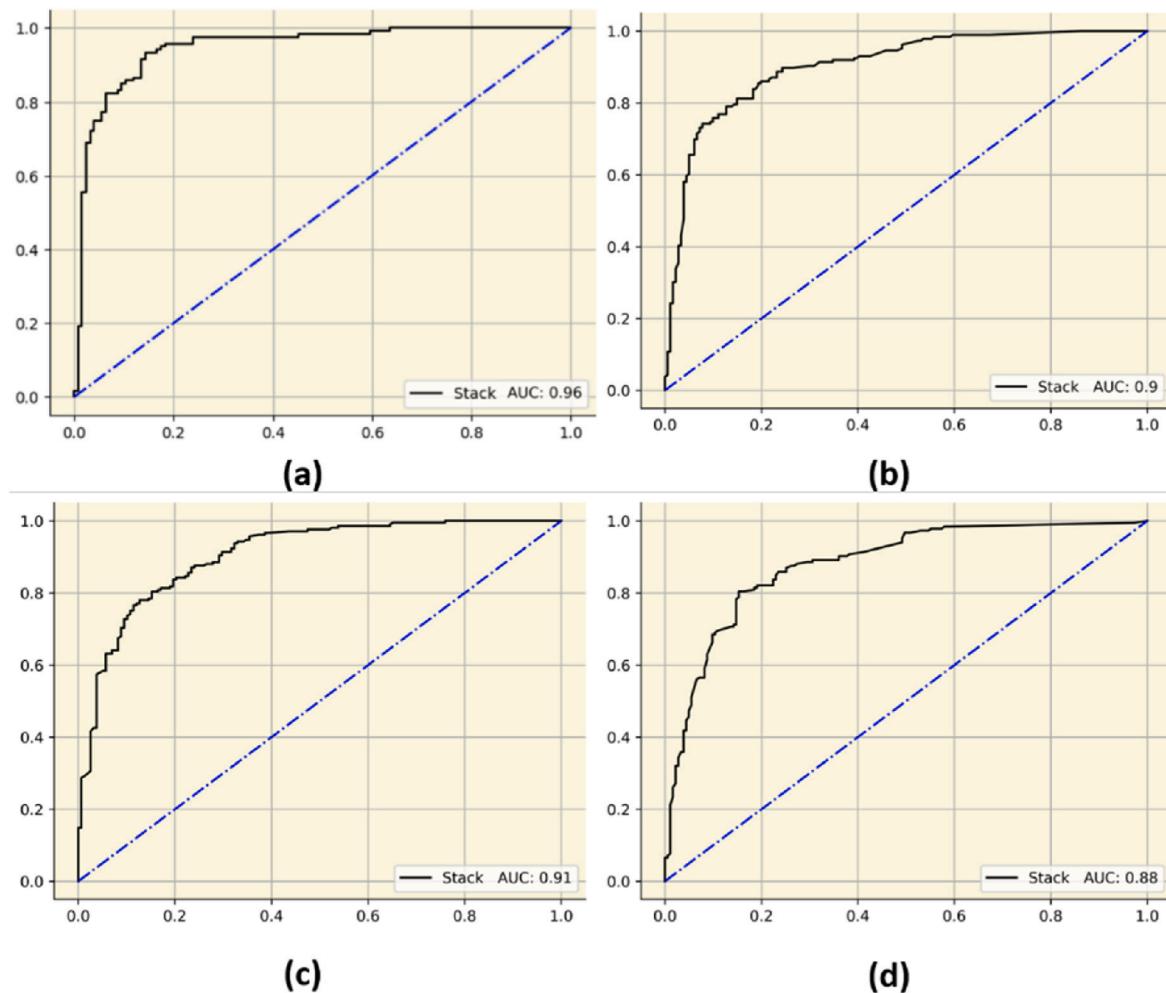


Fig. 6. ROC curves for the stacked models. (a) MI (b) WO (c) JA (d) GWO and SCA.

heart disease prediction provides clarity as it depicts how various clinical features impact the prediction of risk in individual patients. SHAP creates force plots that tell us about the direction and strength of the effect that each feature contributes to the outcome, so that healthcare experts can understand the model's decision-making. Fig. 8 depicts the global interpretation of the models using beeswarm plots. The figure determines that the most important markers are ST slope, Oldpeak, Exercise angina, chest pain type, and fasting blood sugar. The maximum heart rate levels decreased, and the Oldpeak levels increased in the heart attack records.

Local interpretations can be made using the SHAP force plot shown in Fig. 9. Heart attack prediction is shown in Fig. 9 a, b, and d. Markers such as ST slope, Oldpeak, Age, and Exercise angina significantly contribute to predictions indicating heart attack prediction. Non-heart attack prediction is shown in Fig. 9 c. Markers such as fasting blood sugar, sex, and chest pain type influence predictions toward a non-heart attack prediction.

LIME, or Local Interpretable Model-agnostic Explanations, is an effective tool for understanding the behavior of complex machine-learning models in their decision-making processes. This is achieved by generating locally interpretable explanations that highlight individual feature contributions to a given prediction. The LIME results are shown diagrammatically in Fig. 10. It can be inferred that the most important attributes are ST slope, Oldpeak, Exercise angina, Chest pain type and fasting blood sugar.

ELI5 stands for "Explain Like I'm 5" and is a library designed to make interpreting machine learning models easier. These tools enable

inspecting and debugging model behavior, providing feature importance visualizations in a way that is easily understood. ELI5 provides feature importance scores or decision explanations, which indicate the clinical markers that have the greatest influence on the output of the model. Fig. 11 displays the Eli5 predictions, and it can be inferred that the most significant features are Cholesterol, ST slope, Oldpeak and Chest pain type. This explanation accounts for the bias (error rate).

QLattice identifies and depicts the interaction of salient features that drive the prediction, which makes it easy for researchers to understand the feature relationships. The models constructed using QLattice are expressed as mathematical equations and graphs, making them interpretable. From the QGraphs presented in Fig. 12, it can be noted that the most important markers were ST slope, Oldpeak, Exercise angina, and chest pain type.

Anchors are conditions or subsets of feature values that guarantee consistent prediction results with high precision when met. Anchors work by identifying decision rules or "anchors" that capture the local behavior of a model around a particular prediction. The anchors for non-heart and heart attack cases are listed in Table 10. The most important markers were ST slope, Oldpeak, Exercise angina, chest pain type, and fasting blood sugar.

5. Discussion

This study used multiple machine learning algorithms to predict heart attacks well in advance so that specific treatments can be conducted on critical patients. Five heterogeneous XAI techniques are

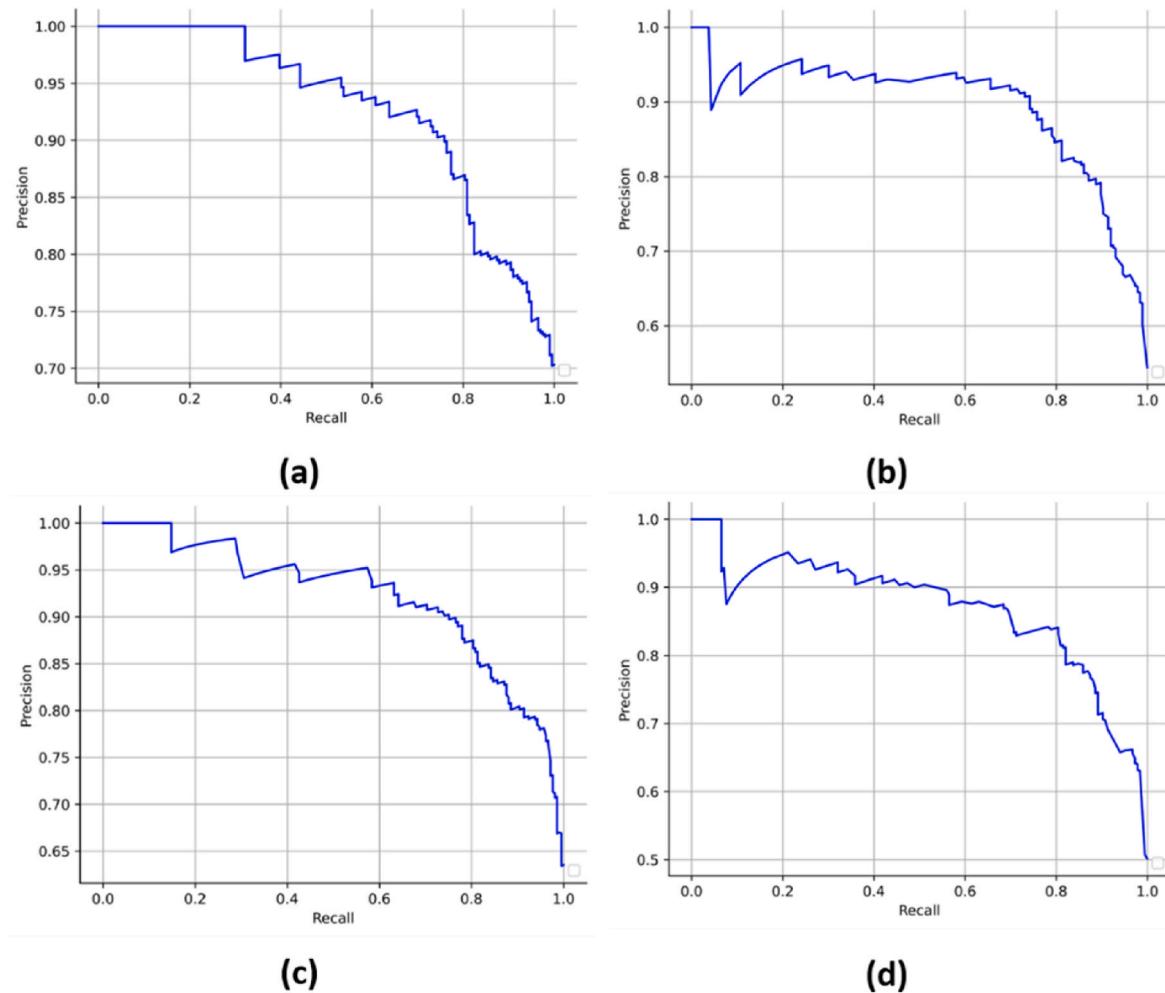


Fig. 7. PR curves for the stacked models. (a) MI (b) WO (c) JA (d) GWO and SCA.

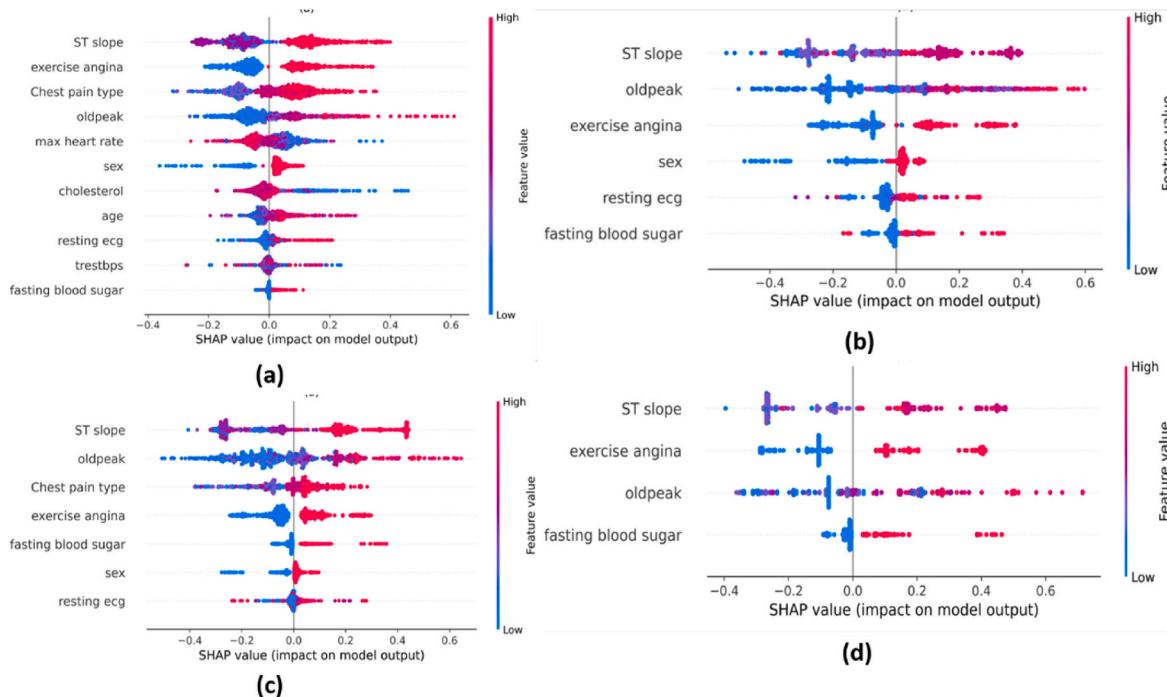


Fig. 8. Beeswarm plots for Global SHAP interpretation (a) MI (b) WO (c) JA (d) GWO and SCA.

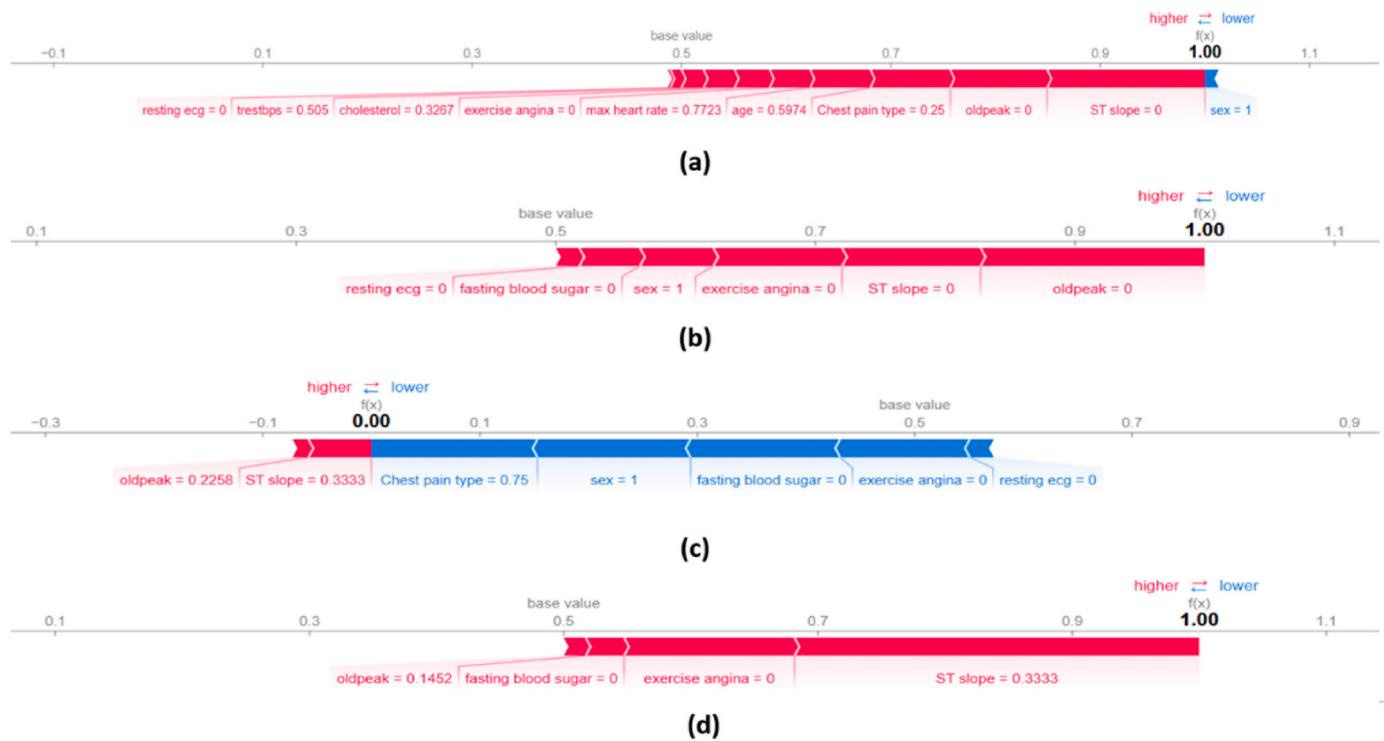


Fig. 9. Force plots for local SHAP interpretation (a) MI b) WO (c) JA (d) GWO and SCA.

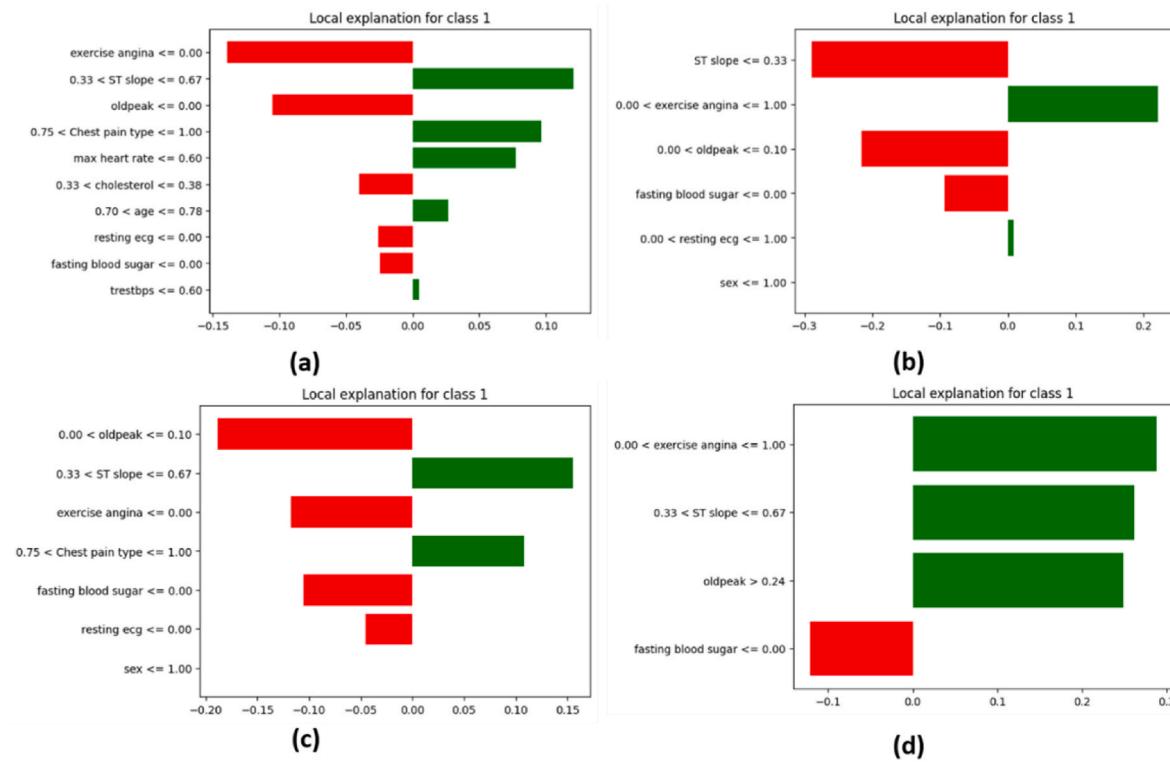


Fig. 10. LIME results (a) MI (b) WO (c) JA (d) GWO and SCA.

applied to demystify the predictions. The results of this study stress that ML models, combined with XAI techniques, will significantly improve the prediction of heart attack risks. By incorporating clinical markers such as ST slope, Oldpeak and Exercise angina into the models, we were able to gain clear insights into how these features influence the risk of a

heart attack.

Fourteen feature selection methods were used, and the best five were used for further analyses. They are Mutual Information, Whale optimization, the Jaya algorithm, the Grey Wolf Optimizer and Sine Cosine algorithm. A maximum accuracy of 90 % is obtained using the mutual

y=0.0 (probability 1.000) top features			y=1.0 (probability 1.000) top features		
Contribution?	Feature	Value	Contribution?	Feature	Value
+0.600	cholesterol	0.353	+0.500	<BIAS>	1.000
+0.500	<BIAS>	1.000	+0.383	oldpeak	0.242
+0.268	Chest pain type	0.000	+0.306	ST slope	0.667
+0.220	ST slope	0.000	+0.124	sex	1.000
-0.181	oldpeak	0.226	-0.156	fasting blood sugar	0.000
-0.408	max heart rate	0.619	-0.157	exercise angina	0.000

(a)			(b)		
y=0.0 (probability 1.000) top features			y=0.0 (probability 1.000) top features		
Contribution?	Feature	Value	Contribution?	Feature	Value
+0.500	<BIAS>	1.000	+0.500	<BIAS>	1.000
+0.244	ST slope	0.333	+0.375	oldpeak	0.113
+0.141	oldpeak	0.000	+0.211	ST slope	1.000
+0.038	resting ecg	0.000	+0.116	exercise angina	0.000
+0.033	exercise angina	0.000	-0.202	fasting blood sugar	1.000
+0.026	sex	0.000			
+0.017	Chest pain type	0.750			

(c)			(d)		
y=0.0 (probability 1.000) top features			y=0.0 (probability 1.000) top features		
Contribution?	Feature	Value	Contribution?	Feature	Value
+0.500	<BIAS>	1.000	+0.500	<BIAS>	1.000
+0.244	ST slope	0.333	+0.375	oldpeak	0.113
+0.141	oldpeak	0.000	+0.211	ST slope	1.000
+0.038	resting ecg	0.000	+0.116	exercise angina	0.000
+0.033	exercise angina	0.000	-0.202	fasting blood sugar	1.000
+0.026	sex	0.000			
+0.017	Chest pain type	0.750			

Fig. 11. Eli5 results (a) MI (b) WO (c) JA (d) GWO and SCA.

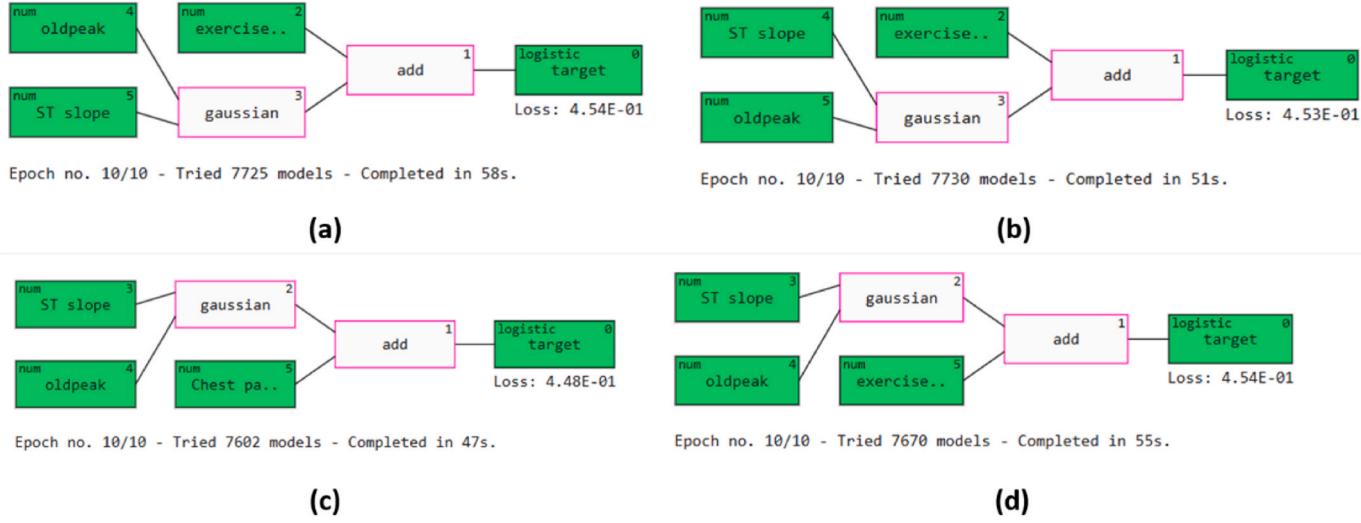


Fig. 12. QLattice results (a) MI (b) WO (c) JA (d) GWO and SCA.

information algorithm. The maximum F1-score, AUC, and AP were 89 %, 0.96, and 0.97, respectively. When the Whale optimizer was utilized, 87 % accuracy was obtained. The maximum F1-score, AUC, and AP were 87 %, 0.92, and 0.92, respectively. An accuracy of 85 % was obtained when the Jaya algorithm was used. The highest F1-score, AUC, and AP were 84 %, 0.92, and 0.94, respectively. When the grey wolf optimizer and sine-cosine algorithms were utilized, 83 % accuracy was obtained. The highest F1-score, AUC, and AP were 83 %, 0.89, and 0.89, respectively.

Other studies utilizing similar XAI techniques have also identified crucial predictors of cardiovascular diseases. Wesolowski et al. [51] used XAI methods to analyze electronic health records (EHRs) and focused on cardiovascular outcomes and comorbid conditions. Dritsas and Trigka [52] applied deep-learning models, including MLP and CNN, and used SHAP to enhance interpretability, with the hybrid model outperforming other methods. Guldogan et al. [53] implemented Explainable Boosting Machines (EBM) to predict angina, finding it to be

more accurate than other AI models, such as Logistic Regression and XGBoost. Deokar and Pradhan [54] applied LIME and SHAP for heart disease prediction, highlighting critical features, such as ST slope and cholesterol. Titti et al. [55] used boosting algorithms, XGBoost and CatBoost, alongside XAI methods to demonstrate the effectiveness of these techniques for heart disease prognosis. Ferdowsi et al. [56] proposed a responsible AI framework for cardiovascular disease detection, focusing on privacy-preserving and interpretable models, whereas Mienye and Jere [57] optimized ensemble models with SHAP for better heart disease prediction, achieving high sensitivity and specificity. Table 11 presents a comparison between this study and existing studies. Several studies have investigated machine learning for the prediction of heart disease. Chang et al. [58] used a random forest classifier with an accuracy of 83 %. Bukhari et al. [59] proposed a stacked CNN with the Levy Flight-based Grasshopper Optimization Algorithm with an accuracy of 99 % on various publicly available datasets. Ozcan and Peker [60] used a Classification and Regression Tree (CART) algorithm with an

Table 10
Anchor results.

Prediction	Anchor	Precision	Coverage
Mutual Information			
Absence of heart attack	oldpeak≤0.00 AND ST slope≤0.33	0.80	0.30
Absence of heart attack	ST slope≤0.33 AND Chest pain type≤0.50	0.81	0.25
Presence of heart attack	ST slope > 0.33 AND Chest pain type > 0.75	0.93	0.29
Presence of heart attack	ST slope > 0.33 AND oldpeak > 0.26	0.91	0.15
Whale Optimization			
Absence of heart attack	oldpeak≤0.13 AND ST slope≤0.33	0.88	0.38
Absence of heart attack	ST slope≤0.33 AND oldpeak≤0.00	0.92	0.28
Presence of heart attack	oldpeak > 0.26 AND 0.33 < ST slope≤0.67	0.92	0.14
Presence of heart attack	oldpeak > 0.26 AND exercise angina > 0.00	0.96	0.14
Jaya Algorithm			
Absence of heart attack	ST slope≤0.33 AND oldpeak≤0.00	0.86	0.29
Absence of heart attack	oldpeak≤0.10 AND ST slope≤0.33	0.84	0.32
Presence of heart attack	oldpeak > 0.11 AND Chest pain type > 0.75	0.85	0.25
Presence of heart attack	oldpeak > 0.24 AND Chest pain type > 0.75	0.92	0.13
Grey Wolf Optimizer and Sine Cosine Algorithm			
Absence of heart attack	ST slope≤0.33 AND oldpeak≤0.00	0.99	0.28
Absence of heart attack	ST slope≤0.33 AND fasting blood sugar≤0.00	0.94	0.30
Presence of heart attack	0.33 < ST slope≤0.67 AND fasting blood sugar > 0.00	0.91	0.13
Presence of heart attack	ST slope > 0.33 AND 0.00 < exercise angina≤1.00	0.93	0.27

accuracy of 87 %. These suggest that machine learning holds promise for cardiovascular risk prediction.

This study has some limitations that should be addressed in future research. The dataset was taken from a single publicly available source, which may lead to dataset bias and constrained representativeness of the population. This situation limits the models' generalization in terms of different demographic, geographic, and clinical settings. Thereby, future initiatives should consider multi-institutional and multi-ethnic datasets for the results to have robust external validity. The models have not undergone validation in real-time and in prospective clinical contexts. While the retrospective performance assessment sheds some light on how these models perform, testing them under real conditions is, in fact, essential to determine how they hold when they are time constraint-based decisions, come in with incomplete records, or when patient conditions start to change. The current model approach is based solely on supervised methods. It would be beneficial if this model learns from semi-supervised, unsupervised, or reinforcement learning; it would

be more flexible when labeled data is not easy to procure. A lot of hurdles wait to be addressed in integrating these predictions in clinical workflow, such as technical integration and interoperability with electronic health record (EHR), clinical acceptance, and regulatory compliance. Interdisciplinary collaborative efforts are needed to overcome these practical barriers and to assess the clinical utility, interpretability, and ethical consequences of deploying such models in everyday patient care.

Future studies could also be designed to incorporate more sophisticated algorithms, such as deep learning models and real-time data streams, to evaluate the dynamic nature of the risks associated with heart disease. Another potential extension of this study would be to expand it to other demographic groups, such as age, ethnicity, and comorbidities, as they can affect cardiovascular health and impact model performance. Such advancements will enable more accurate and clinically applicable predictions in AI-driven systems for heart disease.

6. Conclusions

In this study, we used several machine-learning models combined with explainable AI techniques to predict the risk of heart attack based on clinical markers. XAI techniques, such as SHAP, LIME, Eli5, QLattice, and Anchor, make it possible for us to understand the reasoning behind a model. In healthcare applications, understanding the rationale behind the prediction is as important as the prediction itself. The mutual information reached a maximum precision of 94 %, with a maximum testing accuracy of 90 %. The most important markers identified were ST slope, Oldpeak, exercise-induced angina, chest pain type, and fasting blood sugar level. These results suggest that ML models can be effectively integrated into clinical decision support systems to assist healthcare professionals in early risk stratification, ultimately leading to timely and individualized interventions.

The future of heart disease prediction using machine learning and explainable AI is promising, with several avenues for improvement and broader adoption. One major avenue is the integration of multimodal data including electronic health records, medical imaging, and genomic data. Combining these data sources can significantly increase the reliability of predictive models and further improve the scope of risk assessment. Furthermore, the development of real-time predictive systems powered by cloud computing could enable healthcare providers to make quicker decisions based on up-to-date patient information. Ensemble methods and federated learning, by which models can be learned over decentralized data sources without sharing sensitive information, can increase the robustness of models to privacy. There may be a combination of wearable devices with real-time patient monitoring that offers a continuous flow of data. By integrating these data into heart disease prediction models, clinicians can monitor the patient's condition in real time. Future studies should focus on increasing the interpretability of complex models, especially in high-stakes domains, such as healthcare. The integration of machine learning, XAI, and diverse healthcare data will transform the heart disease prediction and management, making it more personalized and effective.

Table 11
Comparison with existing studies.

Paper	Classifiers	Maximum Accuracy/AUC	Dataset	Explainable AI Techniques
[10]	XGBoost	97.5 % accuracy	Combined dataset	SHAP
[11]	LBGM	Avg Training accuracy 99.33 %	Heart Disease Classification, Kaggle	LIME
[52]	Hybrid model	91 % accuracy	Cleveland	SHAP
[55]	XGBoost, CatBoost	90 % accuracy	Heart failure Classification, Kaggle	LIME
[56]	Logistic Regression	80 % accuracy	IEEE DataPort	SHAP, LIME
[57]	Optimized XGBoost	98 %, 96 % accuracies	Cleveland and Framingham	SHAP
Our paper	Random Forest, AdaBoost, Lightgbm, CatBoost, XGBoost, Stacking Model	90 % accuracy	Prediction of Heart Attack, Mendeley	SHAP, LIME, ELI5, Anchor, QLattice

CRediT authorship contribution statement

Maithri Bairy: Writing – original draft, Data curation, Software. **Krishnaraj Chadaga:** Methodology, Supervision. **Niranjana Sampathila:** Writing – review & editing, Supervision. **R. Vijaya Arjunan:** Writing – review & editing, Supervision. **G. Muralidhar Bairy:** Writing – review & editing, Supervision, All authors have read and approved the final version of the manuscript.

Funding

This research received no external funding.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: R. Vijaya Arjunan reports was provided by Manipal Academy of Higher Education. All authors have no conflict of interest If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank the Manipal Academy of Higher Education for providing us with a platform to conduct this study.

Data availability

The data that support the findings of this study are openly available in Mendeley Data at <https://doi.org/10.17632/yrwd336rkz.2>

References

- [1] World Health Organization, *World Health Statistics 2024*, WHO, Geneva, Switzerland, 2024.
- [2] X. Yao, J. Zhang, X. Zhang, T. Jiang, Y. Zhang, F. Dai, H. Hu, Q. Zhang, Age at diagnosis, diabetes duration and the risk of cardiovascular disease in patients with diabetes mellitus: a cross-sectional study, *Front. Endocrinol.* (2023) 1131395, <https://doi.org/10.3389/fendo.2023.1131395>.
- [3] Z. Liu, D. Meng, G. Su, P. Hu, B. Song, Y. Wang, J. Wei, H. Yang, T. Yuan, B. Chen, T.H. Ou, Ultrafast early warning of heart attacks through plasmon-enhanced Raman spectroscopy using collapsible nanofingers and machine learning, *Small* 19 (2) (2023) 2204719, <https://doi.org/10.1002/smll.202204719>.
- [4] S. Shukla, Real-time monitoring and predictive analytics in healthcare: harnessing the power of data streaming, *Int. J. Comput. Appl.* 185 (8) (2023) 32–37.
- [5] B. Sahoh, A. Choksuriwong, The role of explainable artificial intelligence in high-stakes decision-making systems: a systematic review, *J. Ambient. Intell. Hum. Comput.* 14 (6) (2023) 7827–7843.
- [6] S. Kumar, A. Das, Peripheral blood mononuclear cell derived biomarker detection using eXplainable Artificial Intelligence (XAI) provides better diagnosis of breast cancer, *Comput. Biol. Chem.* (2023) 107867, <https://doi.org/10.1016/j.combiolchem.2023.107867>.
- [7] C.W. Oei, Y.M. Chan, X. Zhang, et al., Risk prediction of diabetic foot amputation using machine learning and explainable artificial intelligence, *J. Diabetes Sci. Technol.* (2024), <https://doi.org/10.1177/19322968241228606>.
- [8] B. Yagin, F.H. Yagin, C. Colak, et al., Cancer metastasis prediction and genomic biomarker identification through machine learning and eXplainable artificial intelligence in breast cancer research, *Diagnostics* 13 (2023) 3314, <https://doi.org/10.3390/diagnostics13213314>.
- [9] K. Chadaga, S. Prabhu, N. Sampathila, et al., Explainable artificial intelligence approaches for COVID-19 prognosis prediction using clinical markers, *Sci. Rep.* 14 (2024) 1783, <https://doi.org/10.1038/s41598-024-52428-2>.
- [10] H. El-Sofany, B. Bouallegue, Y.M.A. El-Latif, A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method, *Sci. Rep.* 14 (2024) 23277, <https://doi.org/10.1038/s41598-024-74656-2>.
- [11] P. Paudel, S.K. Karna, R. Saud, L. Regmi, T.B. Thapa, M. Bhandari, Unveiling key predictors for early heart attack detection using machine learning and explainable AI technique with LIME, in: Proc. 10th Int. Conf. Networking, Systems and Security, 2023, pp. 69–78.
- [12] K.B. Sk, D. Roja, S.S. Priya, L. Dalavi, S.S. Vellela, V. Reddy, Coronary heart disease prediction and classification using hybrid machine learning algorithms, in: 2023 Int. Conf. Innovative Data Commun. Technol. Appl. (ICIDCA), IEEE, 2023, pp. 1–7.
- [13] A.M. Qadri, A. Raza, K. Munir, M.S. Almutairi, Effective feature engineering technique for heart disease prediction with machine learning, *IEEE Access* 11 (2023) 56214–56224.
- [14] Z. Noroozi, A. Orooj, L. Erfannia, Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction, *Sci. Rep.* 13 (2023) 22588.
- [15] A.A. Almazroi, E.A. Aldhahri, S. Bashir, S. Ashfaq, A clinical decision support system for heart disease prediction using deep learning, *IEEE Access* 11 (2023) 61646–61659.
- [16] A.K. Dubey, A.K. Sinhal, R. Sharma, Heart disease classification through crow intelligence optimization-based deep learning approach, *Int. J. Inf. Technol.* 16 (2024) 1815–1830.
- [17] M. Alshraideh, N. Alshraideh, A. Alshraideh, Y. Alkayed, Y. Al Trabsheh, B. Alshraideh, Enhancing heart attack prediction with machine learning: a study at Jordan University Hospital, *Appl. Comput. Intell. Soft Comput.* 2024 (2024) 5080332, <https://doi.org/10.1155/2024/5080332>.
- [18] A. Dyoub, I. Letteri, Dataset optimization for chronic disease prediction with bio-inspired feature selection, *arXiv preprint arXiv:2401.05380* (2023), <https://doi.org/10.48550/arXiv.2401.05380>.
- [19] B. Ahmad, J. Chen, H. Chen, Feature selection strategies for optimized heart disease diagnosis using ML and DL models, *arXiv preprint arXiv:2503.16577* (2025), <https://doi.org/10.48550/arXiv.2503.16577>.
- [20] K. Natarajan, V. Vinoth Kumar, T.R. Mahesh, et al., Efficient heart disease classification through stacked ensemble with optimized firefly feature selection, *Int. J. Comput. Intell. Syst.* 17 (2024) 174, <https://doi.org/10.1007/s44196-024-00538-0>.
- [21] K. Dissanayake, M.G. Md Johar, Heart disease diagnostics using meta-learning-based hybrid feature selection, *Appl. Comput. Intell. Soft Comput.* 2024 (2024) 8800497, <https://doi.org/10.1155/2024/8800497>.
- [22] N. Biswas, M.M. Ali, M.A. Rahaman, M. Islam, M.R. Mia, S. Azam, K. Ahmed, F. M. Bui, F.A. Al-Zahrani, M.A. Moni, Machine learning-based model to predict heart disease in early stage employing different feature selection techniques, *BioMed Res. Int.* 2023 (2023) 6864343, <https://doi.org/10.1155/2023/6864343>.
- [23] R. Mukherjee, S. Sadhu, A. Kundu, Heart disease detection using feature selection based KNN classifier, in: D. Gupta, et al. (Eds.), *Data Analytics and Management*, Springer, Singapore, 2022, https://doi.org/10.1007/978-981-16-6289-8_48.
- [24] R. Spencer, F. Thabtah, N. Abdelhamid, M. Thompson, Exploring feature selection and classification methods for predicting heart disease, *Digit. Health* 6 (2020) 2055207620914777, <https://doi.org/10.1177/2055207620914777>.
- [25] H. Saitou, H. Haraguchi, Frog-Snake prey-predation Relationship Optimization (FSRO): a novel nature-inspired metaheuristic algorithm for feature selection, *arXiv preprint arXiv:2403.18835* (2024), <https://doi.org/10.48550/arXiv.2403.18835>.
- [26] H. Basak, M. Das, S. Modak, RSO: a novel reinforced swarm optimization algorithm for feature selection, *arXiv preprint arXiv:2107.14199* (2021), <https://doi.org/10.48550/arXiv.2107.14199>.
- [27] L. Singh, H. Monga, Nature-inspired algorithms-based optimal features selection strategy for COVID-19 detection using medical images, *New Generat. Comput.* 42 (2024) 761–824, <https://doi.org/10.1007/s00354-024-00255-4>.
- [28] M. Riaz, M. Bashir, I. Younas, Metaheuristics based COVID-19 detection using medical images: a review, *Comput. Biol. Med.* 144 (2022) 105344, <https://doi.org/10.1016/j.combiomed.2022.105344>.
- [29] F. Bolikulov, R. Nasimov, A. Rashidov, F. Akhmedov, C. Young-Im, Effective methods of categorical data encoding for artificial intelligence algorithms, *Mathematics* 12 (2024) 2553.
- [30] R. Karthiga, G. Usha, N. Raju, K. Narasimhan, Transfer learning-based breast cancer classification using one-hot encoding technique, in: Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), IEEE, 2021, pp. 115–120.
- [31] P.J. Ali, R.H. Faraj, E. Koya, Data normalization and standardization: a technical report, *Mach. Learn. Tech. Rep.* 1 (2014) 1–6.
- [32] H. Han, W.Y. Wang, B.H. Mao, Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning, in: *Advances in Intelligent Computing: International Conference on Intelligent Computing, ICIC 2005, Hefei, China, August 23–26, 2005, Proceedings, Part I*, Springer, Berlin Heidelberg, 2005, pp. 878–887.
- [33] S. Mirjalili, A. Lewis, The whale optimization algorithm, *Adv. Eng. Software* 95 (2016) 51–67, <https://doi.org/10.1016/j.advengsoft.2016.01.008>.
- [34] L. Abualigah, A. Diabat, Advances in sine cosine algorithm: a comprehensive survey, *Artif. Intell. Rev.* 54 (4) (2021) 2567–2608, <https://doi.org/10.1007/s10462-020-09909-3>.
- [35] M. Castelli, L. Manzoni, L. Mariotti, M.S. Nobile, A. Tangherloni, Salp swarm optimization: a critical review, *Expert Syst. Appl.* 189 (2022) 116029, <https://doi.org/10.1016/j.eswa.2021.116029>.
- [36] D. Wang, D. Tan, L. Liu, Particle swarm optimization algorithm: an overview, *Soft Comput.* 22 (2018) 387–408, <https://doi.org/10.1007/s00500-016-2474-6>.
- [37] R.A. Zitar, M.A. Al-Betar, M.A. Awadallah, I.A. Doush, K. Assaleh, An intensive and comprehensive overview of JAYA algorithm, its versions and applications, *Arch. Comput. Methods Eng.* 29 (2022) 763–792, <https://doi.org/10.1007/s11831-021-09585-8>.
- [38] A.A. Heidari, et al., Harris hawks optimization: algorithm and applications, *Future Gener. Comput. Syst.* 97 (2019) 849–872, <https://doi.org/10.1016/j.future.2019.02.028>.
- [39] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, *Adv. Eng. Software* 69 (2014) 46–61, <https://doi.org/10.1016/j.advengsoft.2013.12.007>.
- [40] S. Mirjalili, S. Mirjalili, Genetic algorithm, *Evol. Algor. Neural Netw. Theory Appl.* (2019), https://doi.org/10.1007/978-3-319-93025-1_4.

- [41] M. Abdel-Basset, L.A. Shawky, Flower pollination algorithm: a comprehensive review, *Artif. Intell. Rev.* 52 (2019) 2533–2557, <https://doi.org/10.1007/s10462-018-9624-4>.
- [42] V. Kumar, D. Kumar, A systematic review on firefly algorithm: past, present, and future, *Arch. Comput. Methods Eng.* 28 (2021) 3269–3291, <https://doi.org/10.1007/s11831-020-09498-y>.
- [43] M. Pant, H. Zaheer, L. Garcia-Hernandez, A. Abraham, Differential evolution: a review of more than two decades of research, *Eng. Appl. Artif. Intell.* 90 (2020) 103479, <https://doi.org/10.1016/j.engappai.2020.103479>.
- [44] M. Guerrero-Luis, F. Valdez, O. Castillo, A review on the cuckoo search algorithm, in: *Fuzzy Logic Hybrid Extension Neural Optimization Algorithms Theory and Applications*, 2021, https://doi.org/10.1007/978-3-030-68776-2_7.
- [45] T. Agarwal, V. Kumar, A systematic review on bat algorithm: theoretical foundation, variants, and applications, *Arch. Comput. Methods Eng.* (2021), <https://doi.org/10.1007/s11831-021-09673-9>.
- [46] J.R. Vergara, P.A. Estévez, A review of feature selection methods based on mutual information, *Neural Comput. Appl.* 24 (2014) 175–186, <https://doi.org/10.1007/s00521-008-0206-2>.
- [47] M. Gollapalli, et al., A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: pre-diabetes, T1DM, and T2DM, *Comput. Biol. Med.* 147 (2022) 105757, <https://doi.org/10.1016/j.combiomed.2022.105757>.
- [48] S. Kumari, D. Kumar, M. Mittal, An ensemble approach for classification and prediction of diabetes mellitus using soft-voting classifier, *Int. J. Cogn. Comput. Eng.* 1 (2021) 40–46, <https://doi.org/10.1016/j.ijcce.2021.01.001>.
- [49] R. Atallah, A. Al-Mousa, Heart disease detection using machine learning majority voting ensemble method, in: *2019 2nd International Conference on New Trends in Computing Sciences (ICTCS)*, IEEE, 2019, pp. 1–6.
- [50] E. Tasçi, C. Uluturk, A. Uğur, A voting-based ensemble deep learning method focusing on image augmentation and preprocessing variations for tuberculosis detection, *Neural Comput. Appl.* 33 (22) (2021) 15541–15555, <https://doi.org/10.1007/s00521-021-06177-2>.
- [51] S. Wesolowski, G. Lemmon, E.J. Hernandez, A. Henrie, T.A. Miller, D. Weyhrauch, M.D. Puchalski, B.E. Bray, R.U. Shah, V.G. Deshmukh, R. Delaney, An explainable artificial intelligence approach for predicting cardiovascular outcomes using electronic health records, *PLOS Digit. Health* 1 (1) (2022) e0000004, <https://doi.org/10.1371/journal.pdig.0000004>.
- [52] E. Dritsas, M. Trigka, Application of deep learning for heart attack prediction with explainable artificial intelligence, *Computers* 13 (10) (2024) 244, <https://doi.org/10.3390/computers13100244>.
- [53] E. Guldogan, F.H. Yagin, A. Pinar, C. Colak, S. Kadry, J. Kim, A proposed tree-based explainable artificial intelligence approach for the prediction of angina pectoris, *Sci. Rep.* 13 (1) (2023) 22189, <https://doi.org/10.1038/s41598-023-29394-1>.
- [54] A.S. Deokar, M.A. Pradhan, Prediction of cardiovascular diseases using explainable AI, in: *Artificial Intelligence, Blockchain, Computing and Security*, 1, CRC Press, 2023, pp. 201–206.
- [55] R.R. Titti, S. Pukkella, T.S. Radhika, Augmenting heart disease prediction with explainable AI: a study of classification models, *Comput. Math. Biophys.* 12 (1) (2024) 20240004, <https://doi.org/10.1515/cmb-2024-0004>.
- [56] M. Ferdowsi, M.M. Hasan, W. Habib, Responsible AI for cardiovascular disease detection: towards a privacy-preserving and interpretable model, *Comput. Methods Progr. Biomed.* 208 (2024) 108289, <https://doi.org/10.1016/j.cmpb.2024.108289>.
- [57] I.D. Mienye, N. Jere, Optimized ensemble learning approach with explainable AI for improved heart disease prediction, *Information* 15 (7) (2024) 394, <https://doi.org/10.3390/info15070394>.
- [58] V. Chang, V.R. Bhavani, A.Q. Xu, M.A. Hossain, An artificial intelligence model for heart disease detection using machine learning algorithms, *Healthcare Analytics* 2 (2022) 100016, <https://doi.org/10.1016/j.health.2022.100016>.
- [59] S.M.S. Bukhari, M.H. Zafar, S.K.R. Moosavi, M. Mansoor, F. Sanfilippo, An integrated stacked convolutional neural network and the Levy flight-based grasshopper optimization algorithm for predicting heart disease, *Healthc. Anal.* 7 (2025) 100374, <https://doi.org/10.1016/j.health.2024.100374>.
- [60] M. Ozcan, S. Peker, A classification and regression tree algorithm for heart disease modeling and prediction, *Healthc. Anal.* 3 (2023) 100130, <https://doi.org/10.1016/j.health.2022.100130>.