Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/120505

Effective Models for Predicting Heart Disease Using Machine Learning Techniques – A Comparative Study

D. Trabay[1,], W. Gharibi[2], and W. M. Abd-Elhafiez[3,4]*

[1]Information System Department, Obour Institutes for Obour High Institute for Management and computers and Information Systems, Kilo 21 Cairo / Belbeis, Obour City, Egypt
[2]Division of Computing, Analytics, and Mathematics, School of Computing and Engineering, University of Missouri-Kansas City (UMKC), USA
[3]Computer Science Department, Faculty of Computers and Artificial Intelligence, Sohag University, Sohag, Egypt
[4]College of Computer Science & Information Technology, Jazan University, Jazan, Kingdom of Saudi Arabia

Received: 21 Feb. 2023, Revised: 22 Mar. 2023, Accepted: 24 Mar. 2023.
Published online: 1 May 2023.

Abstract: Cardiovascular disease is one of the most important causes of death in the modern world, and a significant barrier to clinical information assessment may be the expectation of cardiovascular illness. Machine learning (ML) has proven helpful for forecasting and decision-making in the healthcare industry's large amount of data. Moreover, ML algorithms have been applied in many important fields such as the internet of things and others.

In this paper, we applied various ML methods to predict and classify heart patients' disease including K-Nearest Neighbor Algorithm (KNN), Naïve Bayes (NB), Neural Network (NN), Decision trees (DT), Support Vector Machine (SVM), Random-Forest (RF), Logistic Regression (LR), Gradient-Boosting GB), Stochastic Gradient Descent (SGD), and Ada-Boost. All models were evaluated, and the most accurate predictive model was chosen to increase the accuracy of heart attack prediction. Compared to other models, our results are efficient, and adequate and could help to predict heart disease more effectively and precisely.

Keywords: Algorithms classification, Machine learning techniques, Heart disease, Prediction model.

1 Introduction

A wide range of illnesses impairing heart function is called "heart disorders." The principal arteries and veins leading to and from the heart may be affected, as well as the heart's valves, surrounding membrane, and muscle. The number of deaths from cardiovascular illnesses outnumbers all other causes of mortality, making them the top cause of death worldwide. Heart diseases were responsible for 17.7 million deaths in 2015, or the proportion of all fatalities was 31%. Of these deaths, 7.4 million were caused by coronary heart disease and 6.7 million by strokes [1,2,3].

Heart disease begins as bouts of severe pain caused by blockages in the arteries that carry blood and oxygen to the heart. As a result, the amount of oxygen reaching the heart either slows down or stops completely, leading to heart attacks, angina, and other chronic diseases that may endanger the patient's life. For those with or at high risk of cardiovascular disease, early detection and control are essential due to one or more risk factors, such as excessively high blood pressure, diabetes mellitus, or hyperlipidemia [3,4,5,6].

Large data sets are analyzed through ML to retrieve vital information for future studies from previous collections [7,8]. There is a wealth of patient data in the medical field and various ML algorithms exploit this data to make an effective diagnostic choice with the help of clinical analysis and healthcare experts.

The subject of ML is incredibly diverse and broad, and its application and breadth are expanding daily. To predict and assess the correctness of the input dataset, ML employs a wide range of classifiers and different techniques [9,10,11].

Disease severity is classified by multiple algorithms such as KNN, DT, Genetic Algorithm (GA), NB, etc. [10,11]. Due to the complex nature of heart disease, the condition requires careful management, and failure to do so may accuse damage to the heart or lead to premature death.

Our paper focuses on predicting heart disease by different key features and using ML techniques represented by KNN, NB, ANN, DT, SVM, RF, LR, GB, SGD, and Ada-Boost [11,12,13]. The first three models are classified as having the highest accuracy in detecting heart disease based on the predefined features that are determined by evaluating and

*Corresponding author e-mail: doaatrabay@oi.edu.eg

comparing different mentioned techniques and then the maximum accurate models are predicted.

The paper is organized as follows: Section 2 presents the relevant works. Section 3 presents classification models. Our proposed techniques are presented in section 4. Results and analysis are highlighted in section 5. Section 6 includes prediction models, and we conclude our paper in section 7.

2 Literature review

Many researchers presented various techniques from previous studies to predict heart disease, such as the authors of Ref. [14] who introduced an approach that focuses on identifying key details using ML techniques, resulting in improvements in the accuracy of cardiovascular disease prediction up to 88.7%. Ref. [15] used ML techniques to improve the precision inside the expectation of clutter by combining many known classification procedures. Whereas Ref. [16] suggested a model that intends to identify relevant traits by utilizing ML techniques that increase the accuracy of disorder prediction where specific feature combinations and well-known are used to introduce the prediction model with an accuracy of 88.7%. Authors of Ref. [17] used a special computation based on a swarm algorithm and ensemble Deep Learning (DL) that provides greater accuracy—up to 95.7% than evaluating using several approaches. Ref. [18] developed a heart disease prediction model using a hybrid RF and a linear model, achieving a more desirable level with a 92% accuracy rate. Ref. [19] studied various predictors of heart disease with forecasting algorithms.

Authors of Ref. [20] presented an ensemble classifier for predicting cardiovascular illness using SVM, KNN, and weighted KNN classifier to evaluate performance. The authors of Ref. [21] surveyed several methods of data mining and ML techniques for diagnosing and predicting cardiac problems to improve the prediction performance methods and accuracy. Ref. [22] used the comparative study of ML algorithms such as NB, SVM, RF, and supervised learning models and achieved an accuracy of 95.08% with Random Forest. Using hierarchical RF formation and a nonlinear regression model, Ref. [23] proposed an approach to find similarities in predicting cardiovascular disease, their method obtained an accuracy of 90.3%. To highlight the key characteristics that can lead to heart disease, the authors of Ref. [24] constructed a heart disease prediction model based on RF, LSTM, KNN, and DNN algorithms to determine the improvement of the prediction accuracy. Ref. [25] recommended comparing different heart disease prediction systems based on their accuracy and reliability using MLP, SVM, and ANN models. For the prediction of mortality in patients with congestive heart failure, Ref. [26] suggested an improved strategy using the trait selection and classification method. The proposed method is compared with older classification algorithms, which include SVM, NB, KNN, and DT and according to an experiment using NetBeans IDE, the model predicts outcomes with an accuracy of 99.2%.

Scanning many references and studying the mentioned previous works, we identified that some of them used even three- or four-ML techniques to predict heart disease, but we think that this is not adequate to obtain an optimal model for predicting heart diseases as there are so many other good techniques, requirements and key features that should be applied to get more efficient results.

3 Data classification and modeling

A dataset can be classified into several categories using the following classification methods:

- KNN Technique - a non-parametric, supervised learning classifier that relies on closeness to produce classifications or predictions about the grouping of a single data point [27].
- DT Technique- the trees are constructed using high entropy inputs for training data samples. Trees are fast and easily constructed by the top-down recursive divide and conquer method [28].
- SVM Technique- a supervised ML technique applied to classification or regression problems where each data point is represented as a point in n-dimensional features space [29].
- Random Forest (RF) Technique - an ensemble classifier that builds numerous decision trees and incorporates them to produce the best results [30].
- ANN Technique - a network of connected units called artificial neurons just like synapses in the human brain which can receive signals, process, and transmit them to the surrounding neurons [31].
- NB Technique - a supervised learning classifier that uses a high-dimensional training dataset to handle classification issues based on text classification [32].
- LR Technique- a predictive analysis method based on probability; it serves as a cover binary response from a binary predictor. It is one of the most often utilized tools for applied statistics and discrete data analysis [33].
- Gradient-Boosting (GB)Technique- a strong ML technique that combines Gradient descent with Boosting. It

comprises an additive model, a loss function, and a weak learner [34].

- Ada-Boost Technique- adaptation parameters to the data based on the actual outcomes in the current iteration. This technique can be used to improve the performance of decision trees and was mainly created for binary classification issues [35].
- SGD Technique – one of the most widely used algorithms for optimization and especially for Neural Networks [36].

4 Our proposed technique based on comparative algorithms

In this paper, we focus on predicting heart disease by using key features and ML techniques implemented by KNN, Tree, SVM, SGD, RF, NN, NB, LR, GB, and Ada-boost. We use a data source that contains the medical history of 918 patients of different ages [37]. The patient's medical parameters, such as age, sex, resting blood pressure, exercise-induced angina, serum cholesterol, etc., provided by this dataset allow us to identify diagnosed patients with heart disease. The dataset enables us to categorize and identify people who are at risk for developing cardiac conditions. The steps applied to the mentioned techniques are as follows (See Fig.1).

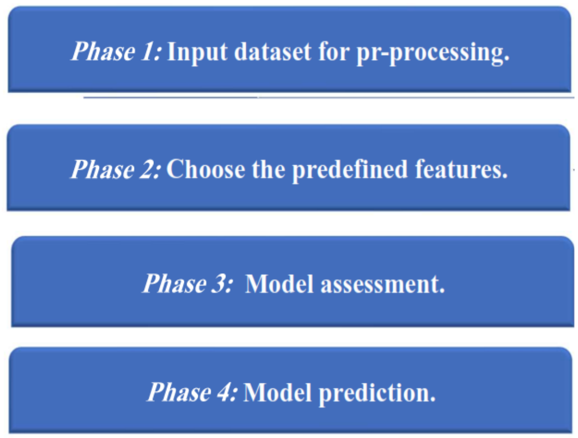


Fig.1. Applied phases.

Different ML techniques are used to classify and predict heart diseases. We evaluate various ML techniques by calculating the confusion matrix; accuracy, F1, recall, precision, an area under the curve, log-loss, and specificity results.

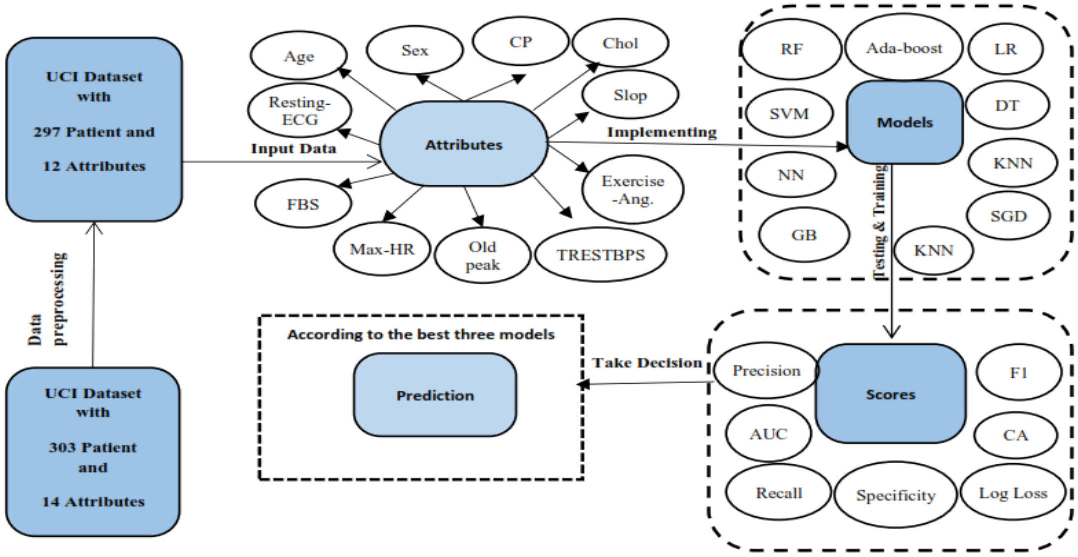


Fig. 2. Proposed model for predicting heart disease.

We choose the most accurate prediction model for heart disease. The presented model is adequate compared with other models, and it could predict the presence of heart disease in particular individuals (See Fig. 2).

4.1 Data pre-processing

All losing records were removed from the dataset in the data pre-processing and data normalization. It utilized twelve standard attributes such as age, sex, resting blood pressure, exercise-induced angina, serum cholesterol, maximum heart rate, old peak, highest exercise slope, chest pain, fasting blood sugar, resting electrocardiography, and heart disease (Table 1). Our target is predicting heart disease, which indicates the level of disease severity in patients where zero indicates no heart disease, while different critical levels of the other values indicate the presence of heart disease. In our paper, critical significance levels range from one to four and are normalized to one level in order to predict whether heart disease is present or not.

Table 1: Description of features with twelve attributes

Attributes	**Description**	**Value**
Age	Patient age	28:77
Sex	Sex of Subject	0= Female, =1= Male
CP	Chest Pain Type	0= standard angina 1= atypical angina 2= non-anginal pain 3= Asymptomatic
TRESTBPS	Resting blood pressure	80:180
Chol	Serum Cholesterol in milligrams per deciliter	0:600
FBS	Fasting blood sugar	0 = false; 1 = true
Resting-ECG	findings of resting electrocardiography	0=Normal, 1= having an aberrant ST-T wave (T wave inversions or ST elevation or depression of more than 0.05 mV 2= potential or actual left ventricular hypertrophy
Max-HR	maximum heart rate achieved	60:202
Exercise-Ang.	Exercise-induced angina	0 = false; 1 = true
Old peak	depression brought on by exercise	0:5
ST-Slope	highest exercise slope	0= upsloping

		1= Flat 2= Sloping down
Heart Disease	heart disease in the patient.	0 = No Disease; 1 = Disease

4.2 Features selection

The primary feature selection is to find the best features that help to create applicable heart disease models. We use Filter Method [38] by choosing the intrinsic houses of the aspects measured via univariate measurements alternatively than cross-validation execution. And to distinguish the most imperative features, we perform a set of tests using 12 features with the classification models: KNN, NB, DT, SVM, NN, RF, LR, GB, SGD, and Ada-Boost.

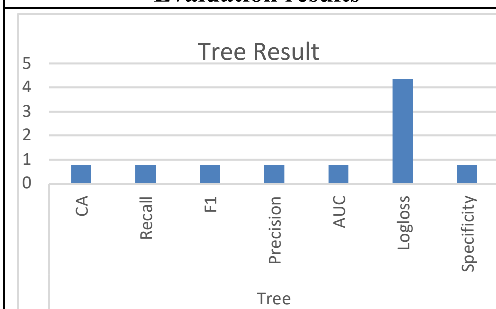
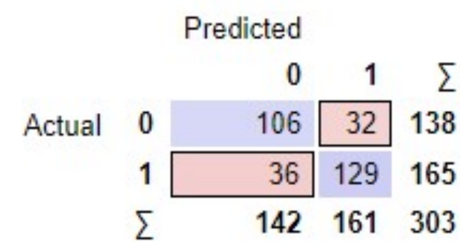
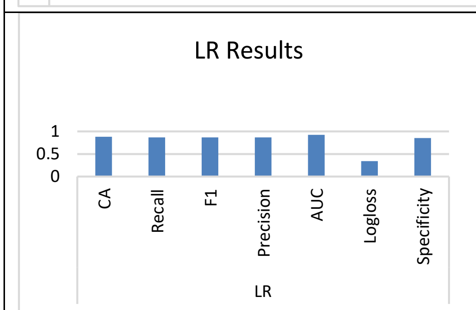
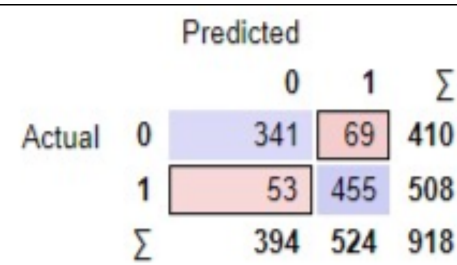
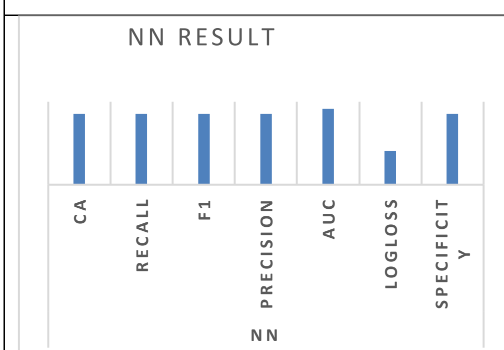
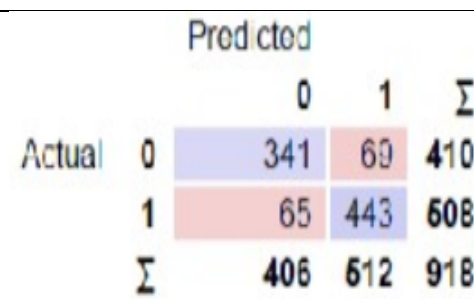
5 Results- evaluation and discussion

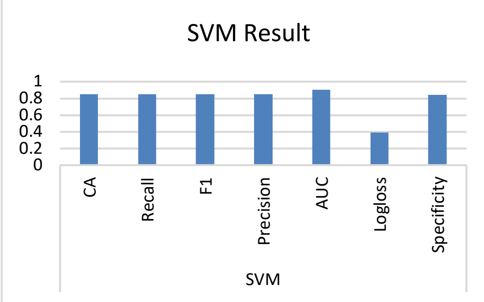
In our research, we focus on the previous ten ML models and implement the following techniques:

5.1 Model prediction

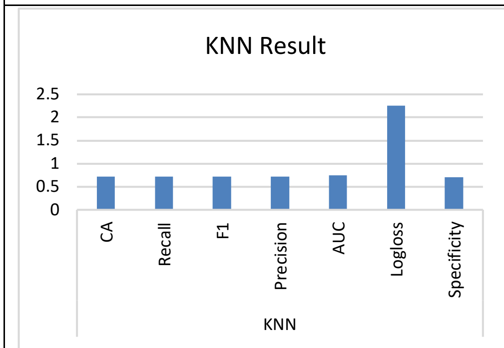
Table 2 depicts our results implementing the Confusion Metrix (CM), Accuracy (CA), F1 score, Recall, Precision, Area Under the Curve (AUC), log Loss, and Specificity:

Table 2: Evaluation results

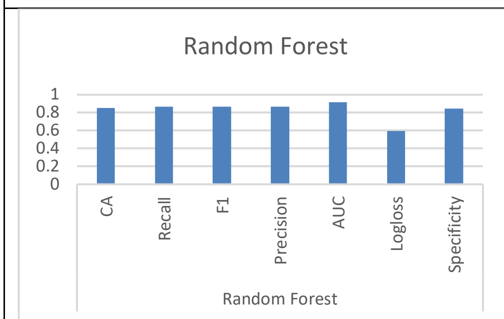
Evaluation results	**CM**
	
	
	



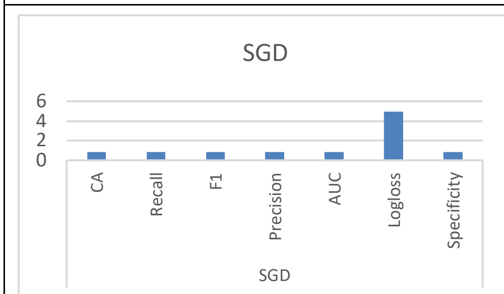
		Predicted		
		0	1	∑
Actual	0	323	87	410
	1	49	459	508
	∑	372	546	918



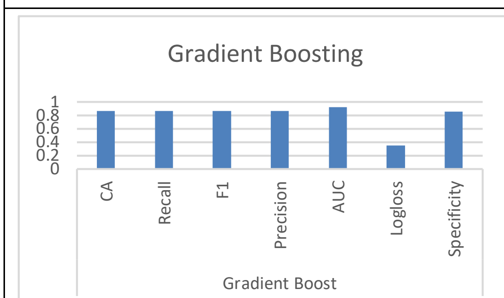
		Predicted		
		0	1	∑
Actual	0	270	140	410
	1	121	387	508
	∑	391	527	918



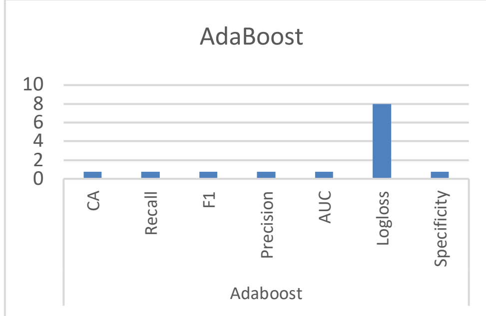
		Predicted		
		0	1	∑
Actual	0	329	81	410
	1	59	449	508
	∑	388	530	918



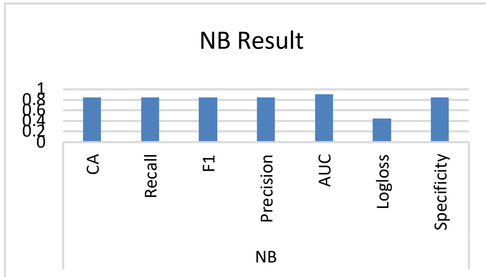
		Predicted		
		0	1	∑
Actual	0	341	69	410
	1	55	453	508
	∑	396	522	918



		Predicted		
		0	1	∑
Actual	0	336	74	410
	1	51	457	508
	∑	387	531	918



		Predicted		
		0	1	∑
Actual	0	313	97	410
	1	115	393	508
	∑	428	490	918



		Predicted		
		0	1	∑
Actual	0	341	69	410
	1	66	442	508
	∑	407	511	918

According to the confusion matrix, we calculate the accuracy; the ratio of correct predictions to the total predictions made, and we summarize our results as follows:

- We obtained the accuracy of LR Model according to the performance of the classification model as the best model that achieved 96%.
- The second-best model is GB Model which achieves 95% accuracy compared to other models.
- By evaluating the performance of the classification models, the RF Model obtained 94% higher accuracy for early heart disease prediction.

5.2 Results of patients' distribution

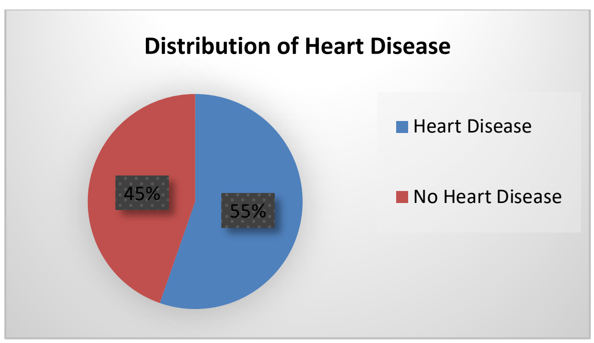


Fig. 3. Heart Disease Distribution

By processing our collected data, it was observed that heart disease occurred in 55.34% of patients, while 44.66% did not suffer from heart disease (See Fig. 3).

5.3 Evaluation of significant features

From our results, we deduced that the features ST-slop, CP, and Exercise-Angina significantly influence our accuracy results. Fig. 4,5, and 6 show the box plot for each feature.

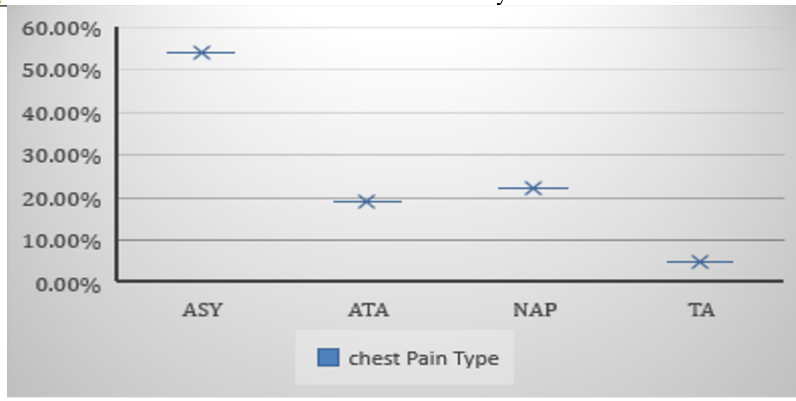


Fig. 4. Box plot CP

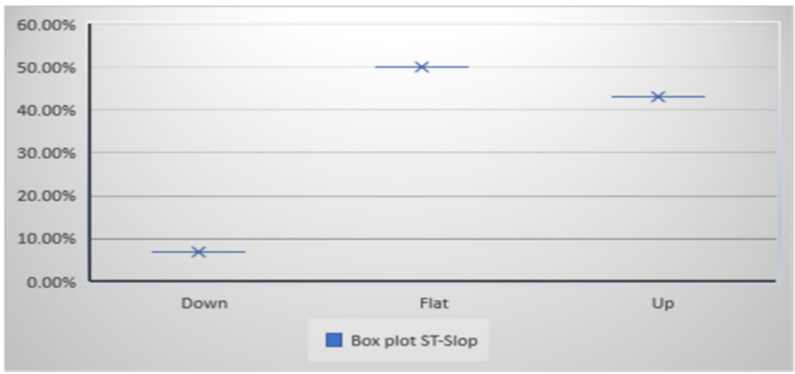


Fig. 5. Box plot ST-slop

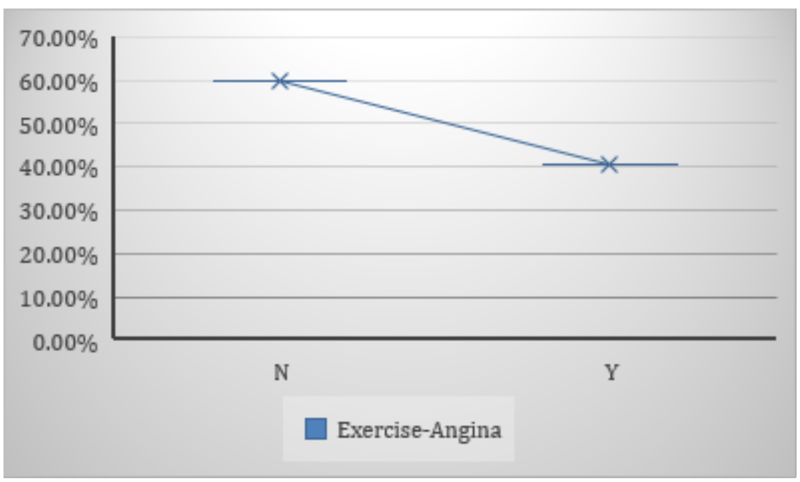
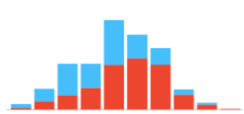


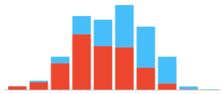
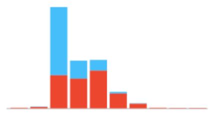
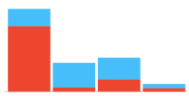
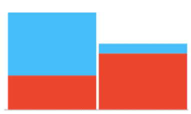
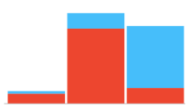
Fig. 6. Box plot Exercise-Angina

5.4 Distribution statistics and measures of dispersion

The data is dispersed over a specified range in each data set. Distribution can be used to gauge the data's dispersion or proximity to the Mean; As a result, Table (3) measures each property with its Mean, Median, Dispersion, Optimal values, and Distribution Statistics.

Table 3: Distribution statistics and measures of dispersion

Data types	Name	Distribution	Mean	Median	Dispersion	Min	Max
Numeric	Age		53.51	54	0.18	28	77

Data types	Name	Distribution	Mean	Median	Dispersion	Min	Max
	Resting-ECG		132.40	130	0.14	0	200
	Chol		198.80	223	0.55	0	603
	Max. HR		136.81	138	0.19	60	202
	Old peak		0.887	0.6	1.201	-2.6	6.2
Categorial	Sex			M	0.514	-	-
	CP			ASY	1.13	-	-
	FBS			0	0.543	-	-
	Restringing			Normal	0.949	-	-
	Exercise Angina			N	0.675	-	-
	ST. Slope			Flat	0.893	-	-

6 Prediction for three models

We predict the highest three models that give us the highest accuracy for the 12 previous models: (LR, GB, and RF). We choose a data set sample with 450 instances, 11 variables, and 12 features (7 categorical and five numeric) (See Table 4).

Table 4: Prediction for best models

	LR	Gradient boosting	Random Forest	Age	Sex	CPT	Resting-BP	Chol	FBS	Rest-ECG	Max-HR	Exercise Angina	Old peak	Slope
1	0	0	1	40	M	ATA	140	289	0	Normal	172	N	0.0	Up
2	0	0	1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat
3	0	0	1	37	M	ATA	130	283	0	ST	98	N	0.0	Up
4	1	1	1	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat

5	1	0	1	54	M	NAP	150	195	0	Normal	122	N	0.0	Up
6	0	0	1	39	M	NAP	120	339	0	Normal	170	N	0.0	Up
7	0	0	0	45	F	ATA	130	237	0	Normal	170	N	0.0	Up
8	0	0	1	54	M	ATA	110	208	0	Normal	142	N	0.0	Up
9	1	1	1	37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat
10	0	0	0	48	F	ATA	120	284	0	Normal	120	N	0.0	Up
11	0	0	1	37	F	NAP	130	211	0	Normal	142	N	0.0	Up
12	1	1	1	58	M	ATA	136	164	0	ST	99	Y	2.0	Flat
13	0	0	1	39	M	ATA	120	204	0	Normal	145	N	0.0	Up
14	1	1	1	49	M	ASY	140	234	0	Normal	140	Y	1.0	Flat
15	0	0	0	42	F	NAP	115	211	0	ST	137	N	0.0	Up

After analyzing the prediction result of three models, we observed that GB and LR Models are more accurate than the RF Model by 99%.

7 Conclusion

Heart disease is one of the most prevalent diseases, and early diagnosis is essential for saving lives. In this paper, we introduced ML techniques for predicting heart disease: KNN, NP, NN, SVM, RF, LR, GB, SGD, Ada-Boost, and DT. We considered 297 patients with twelve attributes, identified, and concluded that LR, GB, and RF Models are more accurate than other algorithms. According to the distribution data, we found that 55.34% of patients had heart disease. Based on different variances, we selected significant factors critical to doctors' staff, performed Distribution Statistics, and extracted the Mean, Median, and both minimum and maximum values. The three most reliable methods, which help doctors diagnose diseases before they appear, were predicted based on our used features. The GB and LR Models are 99% more accurate than the RF model.

Finally, it is essential to note that while machine learning can be a useful tool for prediction and decision making in healthcare, it should not be relied upon solely and should be used in conjunction with other methods and clinical expertise. It is also essential to consider ethical issues and ensure that the use of ML in healthcare is transparent, fair, and respects patient privacy.

Funding Statement:

The authors received no specific funding for this study.

Conflict of interest

The authors declare that there is no conflict regarding the publication of this paper.

References

[1] Kasper, D., Fauci, A., Hauser, S., Longo, D., Jameson, J., & Loscalzo, J. (2015). Harrison's principles of internal medicine, 19e, **l(2)**, *New York, NY, USA:: Mcgraw-hill*, (2015).

[2] Muller, J. E., Tofler, G. H., & Stone, P. H. (1989). Circadian variation and triggers of onset of acute cardiovascular disease. *Circulation*, **79(4)**, 733-743 (1989).

[3] Pepine, C. J., & Nichols, W. W. (2007). The pathophysiology of chronic ischemic heart disease. Clinical Cardiology: *An International Indexed and Peer-Reviewed Journal for Advances in the Treatment of Cardiovascular Disease*, **30(S1)**, I-4 (2020).

[4] Touray, M., & Touray, A. (2021). Clinical Work and General Management of a Standard Minimal-Resource Facility. *Springer International Publishing.*

[5] Khalil, H., & Zeltser, R. (2022). Antihypertensive medications. In StatPearls [Internet]. *StatPearls Publishing*.

[6] Khan, N., Akbar, A., Fahad, S., Faisal, S., & Naushad, M. (2020). Analysis of Heart Treatment and Its Impact on Socioeconomic Conditions on the World Community. *Available at SSRN 3727588* (2020).

[7] Gharibi, G., Walunj, V., Alanazi, R., Rella, S., & Lee, Y. (2019, June). Automated management of deep learning experiments. *In Proceedings of the 3rd International Workshop on Data Management for End-to-End Machine Learning*, 1-4 (2019).

[8] Gharibi, G., Walunj, V., Nekadi, R., Marri, R., & Lee, Y. (2021). Automated end-to-end management of the modeling lifecycle in deep learning. *Empirical Software Engineering*, **26(2)**, 1-33 (2021).

[9] Almahshi, H. M., Almasri, E. A., Alquran, H., Mustafa, W. A., & Alkhayyat, A. (2022, May). Hypothyroidism Prediction and Detection Using Machine Learning. *In 2022 5th International Conference on Engineering Technology and its Applications (IICETA) IEEE*, 159-163 (2022).

[10] Ayon, S. I., Islam, M. M., & Hossain, M. R. (2020). Coronary artery heart disease prediction: a comparative study of computational intelligence techniques. *IETE Journal of Research*, 1-20 (2020).

[11] Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & education*, 143, 103676 (2020).

[12] Pujari, A. K. (2001). Data mining techniques. *Universities press*.

[13] Katarya, R., & Meena, S. K. (2021). Machine learning techniques for heart disease prediction: a comparative study and analysis. *Health and Technology*, **11(1)**, 87-97 (2021).

[14] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, **7**, 81542-81554 (2019).

[15] Faieq, A. K., & Mijwil, M. M. (2022). Prediction of Heart Diseases Utilising Support Vector Machine and Artificial Neural Network. *Indonesian Journal of Electrical Engineering and Computer Science*, **26(1)**, 374-380.

[16] Nirmala, S., Veena, K., Indu, B., & Kalshetty, J. N. (2022, October). Heart Disease Prediction Using Artificial Intelligence Ensemble Network. *In 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)*, 1-6 (2022).

[17] Wankhede, J., Sambandam, P., & Kumar, M. (2021). Effective prediction of heart disease using hybrid ensemble deep learning and tunicate swarm algorithm. *Journal of Biomolecular Structure and Dynamics*, 1-12 (2021).

[18] Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. *SN Computer Science*, **1(6)**, 1-6 (2020).

[19] Nyamathulla, S., & Varikuti, R. (2022). Predicting heart disease with hybrid machine learning algorithms. *Available at SSRN 4121828*, (2022)

[20] Sharma, R., & Singh, S. N. (2022). An Efficient Hybrid Classifier for Prognosing Cardiac Disease. *Webology*, **19(1)**, 5028-5046 (2022).

[21] Thilagamani, S. (2021). A Survey on Efficient Heart Disease Prediction Technique. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, **12(9)**, 130-136 (2021).

[22] Yadav, D. C., & Pal, S. A. U. R. A. B. H. (2020). Prediction of heart disease using feature selection and random forest ensemble method. *International Journal of Pharmaceutical Research*, **12(4)**, 56-66 (2020).

[23] Aslam, M., & Jaisharma, K. (2021, January). Hierarchical Random Forest Formation with Nonlinear Regression Model for Cardiovascular Diseases Prediction. *In 2021 International Conference on Computer Communication and Informatics (ICCCI), IEEE*, 1-5 (2021).

[24] Liu, Y., Zhang, M., Fan, Z., & Chen, Y. (2020, December). Heart disease prediction based on random forest and LSTM. *In 2020 2nd International Conference on Information Technology and Computer Application (ITCA) IEEE*. 630-635 (2020).

[25] Shetty, S. (2022). Heart Disease Prediction Systems and a comparison of methods to improve Public healthcare. (2022).

[26] Sowmiya, C., & Sumitra, P. (2021). A hybrid approach for mortality prediction for heart patients using ACO-HKNN. *Journal of Ambient Intelligence and Humanized Computing*, **12(5)**, 5405-5412 (2021).

[27] Bailey, T., & AK, J. (1978). A note on distance-weighted k-nearest neighbor rules. (1978).

[28] Quinlan, J. R. (1987). Simplifying decision trees. *International journal of man-machine studies*, **27(3)**, 221-234 (1987).

[29] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, **20(3)**, 273-297 (1995).

[30] Ho, T. K. (1995, August). Random decision forests. *In Proceedings of 3rd international conference on document analysis and recognition, IEEE*, **1**, 278-282 (1995).

[31] Hand, D. J., & Yu, K. (2001). Idiot's Bayes—not so stupid after all?. *International statistical review*, **69(3)**, 385-398 (2001).

[32] McCallum, A., & Nigam, K. (1998, July). A comparison of event models for naive bayes text classification. *In AAAI-98 workshop on learning for text categorization*, **752(1)**, 41-48 (1998).

[33] Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The journal of educational research*, **96(1)**, 3-14 (2002).

[34] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232 (2001).

[35] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119-139 (1997).

[36] Ma, S., Bassily, R., & Belkin, M. (2018, July). The power of interpolation: Understanding the effectiveness of SGD in modern over-parametrized learning. *In International Conference on Machine Learning*. 3325-3334 (2018). PMLR.

[37] Siddhartha, M. (2020). Heart disease dataset (comprehensive). *IEEE Dataport*, (2020)

[38] Sánchez-Maroño, N., Alonso-Betanzos, A., & Tombilla-Sanromán, M. (2007, December). Filter methods for feature selection–a comparative study. *In International Conference on Intelligent Data Engineering and Automated Learning*. 178-187 (2007). Springer, Berlin, Heidelberg.