Comparison of Cardiac Stroke Prediction and Classification Using Machine Learning Algorithms

S. Tamil Selvan, R. Rajkumar, P. Chandrasekar, A. Poonguzhali, and Karthick Balasubaramaniam

1 Introduction

Cardiac disorder [1] usually refers to extraordinary functioning of the heart. This usually occurs in the elderly; however, these days it is not unusual among people of all ages. Newborn babies are also affected by this disorder, which is known to be hereditary. One of the most important organs in the human body is the heart. The functioning of our heart leads to the functioning of our life. The impact of the non-proper functioning of the heart may lead to the failure of other parts of the body. If the heart does not function properly, it will have an impact on other human body parts such as the brain and kidneys. The heart/cardiovascular system merely comprise a pump that circulates blood throughout the body. If the body's blood supply is insufficient, numerous organs, including the brain, suffer, and if the heart stops working, the internal mind dies. Life throws a lot of obstacles in the way of the

S. Tamil Selvan
Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India

R. Rajkumar (✉)
Department of ECE, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India

P. Chandrasekar
Department of ECE, Siddartha Institute of Science and Technology, Puttur, India

A. Poonguzhali
Department of ECE, Sri Sairam College of Engineering, Bangalore, Karnataka, India

K. Balasubaramaniam
Universität Duisburg-Essen, Duisburg, Germany

C. Ram Kumar, S. Karthik (eds.), *Translating Healthcare Through Intelligent Computational Methods*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-031-27700-9_10

ability of the heart's to function properly. The term heart disease refers to a disease that affects the cardiovascular system. Examine your family's history of heart disease for the following habits or symptoms: smoking, poor habits, unusual pulse, cholesterol, high blood cholesterol, obesity, etc.

The signs and symptoms of the condition differ from one person to the next. Mostly, there are no early signs or symptoms and the condition is only detected later on. Heart disease [2] manifests in a variety of ways.

- Pain in the chest (angina pectoris).
- A pain in the regions of neck, chest, and shoulders such as tightness or burning.
- Chest problems.
- Symptoms to watch for include sweating, light-headedness, dizziness, and breathing issues.
- Growth of pain during exercising that stretches between the chest to the arm and neck.
- A cough.
- Abeyance of fluids.

Although the causes of heart disease remain unknown, age, sex, case records and ethnic background have all been suggested as possible reasons in numerous studies. The major factors that are associated with the escalation of developing heart disease include irritating behaviors, hypertension with varying stress and strain levels, deep fried oily foods, lack of exercise, saturated fat abnormality in the body, environmental contamination, overweight of the body, abnormal glucose levels. Most illness occurrence is documented in adults between the ages of 50 and 60 years, according to cardiac studies.

Owing to the development of the latest technologies, new algorithms and methods have been developed for the prediction and diagnosis of the disease. One such efficient method of the predications and classification is based on machine-learning techniques. Machine learning (ML) [3], is a substantial multidisciplinary correction that draws on values from applied technological know-how, facts, scientific subjects, engineering, optimization theory, and a variety of other mathematics and science fields. Various gadget learning applications are available, but records management is the most important. The two major classifications of machine-learning algorithms are unsupervised learning and supervised learning.

Machine learning [4–8] without supervision will not be able to make inferences from datasets that comprise the documents without label, or in other words, unsupervised learning does not provide preferred output. Supervised device learning challenge approaches to understanding the relationship between input attributes (unbiased variables) and a given aim (structured variables). Methods that are supervised can be divided into two categories: category and reversal. The output variable in reversal accepts continuous values, whereas the output variable in category takes class labels. Machine-learning techniques are a type of programming that mimics many aspects of human thought, allowing us to quickly solve extremely tough issues. As a result, gadget learning holds much promise for enhancing the performance and accuracy of smart PC software. Concept mastering and category learning are two aspects of machine-learning awareness. Classification is a commonly used machine-learning awareness of an approach that entails separating data into discrete, non-overlapping pieces. As a result, classification is the process of identifying a set of ways to define and distinguish the data object. Mobile gadgets, such as smart phones, smartcards, and sensors, as well as hand-held and automotive computer systems, can all benefit from machine learning.

Mobile terminals (such as Laptops, Tablets, PCs, and other digital assistants) and mobile networks (such as a global system for mobile communications, 3G+, wireless networks, and Bluetooth) have supported progress. Cell devices benefit from machine-learning approaches such as C4.5, Naive Bayes (NB), and decision tree (DT). Classification is a data-processing (machine-learning) technique for predicting group membership in data examples. Although there are a multitude of strategies for system mastering, there are a few that are most commonly employed. Classification is a difficult task in system awareness, especially in future planning and knowledge discovery.

Researchers in the fields of machine learning and information processing aid the classification of a collection of fascinating investigated problems. Classification is a well-known strategy for learning about gadgets, but it has drawbacks, such as dealing with a lack of facts. Missing values in a data set can cause issues at both the educational and the classification level. Non-access of report acknowledgments owing to erroneous impressions, records diagnosed irrelevant at the time of entry, records removed owing to discrepancies with other documented records, and gadget failure are just a few of the possible explanations for missing facts.

Data miners can delete missing records, replace all missing values with a private global standard, replace an omitted price with its characteristic recommendation for the current class, manually review samples with missing values, and insert a possible or likely price. We will concentrate our attention on a few main classification strategies in this diagram. The following are the types of strategies [9] that a machine can learn:

Supervised Learning: A classified dataset is used to train the model developed. The document is entered, as well as the results. Data are categorized and separated into two datasets: schooling and checking. The accuracy of the learning model is ensured by the testing data set attributes and the basic training performed by the training dataset. The output of the algorithms will be based on the training data, which will be an example of classification and regression models.

Unsupervised Learning: The dataset contains no classification or labeling of the data utilized to produce it. The objective of the unsupervised models is recognizing the patterns hidden in the data based on the conditions provided. For any data set given in the input, this learning model makes decisions on the patterns hidden in the input and explores the data. The collection technique is one of the examples and this unsupervised method has no effects on the data set provided in the input.

Reinforcement Learning: This learning method learns the information based on the conditions provided and no connected information about the data set is known to the learning. Using this strategy, the description enhances its presentation by tying it to a specific area and identifies an optimized output by evaluating and experimenting with variable inputs.

It is hard to detect heart disease owing to a number of causal threat elements similar to diabetes, excessive pressure, excessive cholesterol, unusual pulse, and plenty of different elements. A type of strategy in information processing and neural networks is working to seek not in the harshness of heart disease in humans. The severity of the disease is identified using several approaches such as DT, K-Nearest Neighbor Algorithm (KNN), NB and Genetic Algorithm. People with heart ailments complications should be dealt with proper analysis and predictions. The attitude of medicinal technology is to discover and treat different types of metabolic syndromes. The machine learning and information processing plays a vital role in heart disease identification research.

The major purpose of these investigations is to improve the detection accuracy of cardiac problems. Several investigations have been conducted that have resulted in mixture of algorithmic techniques with performancer. In an assessment, the hybrid random forest with linear model (HRFLM) technique uses all functions with feature choice. This behavior experiment will use a hybrid technique to find a system's structures by observing a set of rules. In comparison with earlier methodologies, the findings of this research indicate that our suggested hybrid technique has a higher potential for predicting heart disease. HRFLM data pre-processing experimentally followed by, classification modeling, and performance measurement.

The paper is organized as Sects. 1 and 2 presentation with the target of the work and the survey of the framework. Section 3 examines the proposed framework of the proposed algorithm and examined with models in Sect. 4. Section 5 discuss about the execution and results of the proposed method with the examination results. Finally, Sect. 6 concludes the paper.

2 Literature Review

Yan et al. [10] in their research fostered a framework for diagnosing innate heart problems. The framework also utilizes back-propagation neural networks, which depend on data and heart disease signs and side effects. The innovation has a 90% precision rate. Newman et al. [11] made a framework that utilizes a fake neural network, which is normally used for prediction in the clinical field. This study looks at the promotion advantages and disadvantages of artificial intelligence (AI) calculations such as support vector machines (SVM), NB, and neural networks.

Lu et al. [12] fostered a technique for distinguishing educational quality subsets utilizing correlation feature selection methodology, which used as a heuristic pursuit strategy to consider the space factors, and the subset weight was determined utilizing these estimations. The exactness of the SVM approach was 76.33% on data obtained for 52 patients out of 4726 cases. Priyanka and Kumar [13] laid out a strategy for foreseeing heart issues utilizing data-mining procedures, DT and NB, and showed that the DT were more precise than NB for a dataset applicable to heart infections gathered from the University of California, Irvine.

Jaidhan et al. [14] fostered a strategy for identifying false Mastercard exchanges utilizing an AI procedure called the random forest calculation, which displayed a 0.267% increment in effectiveness over standard models. Utilizing data mining techniques, Palaniappan and Awang [15] fabricated a model called the Intelligent Heart Disease Prediction System. The examples and cooperations between clinical boundaries related to coronary illness are utilized to anticipate coronary illness. Utilizing data mining methods, for example, NB, DT calculation, KNN, and neural networks, Thomas and Princy [16] recommended a framework for anticipating cardiovascular issues. This examination shows that having a larger number of characteristics prompts a more significant level of exactness.

The initial segment included making a dataset with 13 credits, which was then used to run arrangement calculations utilizing DT and random forest algorithms. Finally, the exactness of the two is not set in stone. Subsequently, it tends to be shown that in the prediction of heart illnesses, random forests outperform DTs.

Gavhane et al. [17] made an application that can foresee the weakness of a cardiac infection in view of essential side effects such as age, sex, pulse rate, and different variables. The AI calculation of neural networks has been demonstrated to be the most exact and solid method. Esfahani et al. [18] utilized crude data on cardiovascular patients from the University of California, Irvine. Design acknowledgment procedures, for example, DTs, neural networks, rough sets, SVMs, and NB are tried in the research center for precision and prediction.

Gandhi and Singh [19] featured various ways of dealing with information by utilizing data-mining techniques, which are currently being utilized in heart disease prediction research. Data-mining approaches such as NB, neural networks, and DT calculations are analyzed using calculations on clinical data sets. Nahar et al. [20] utilize the UCI Cleveland dataset, an organic database, and the three rule age calculations – a priori, predictive a priori, and tertius – to reveal these causes utilizing affiliation data mining, a computational insight strategy. Women are accepted to have a lower chance of coronary issues and heart disease than men, in light of data accessible on debilitated and healthy people and involving certainty as a pointer.

There were various elements that highlighted both solid and perilous conditions. Asymptomatic chest problems and exercise-instigated angina are believed to predict the presence of heart disease in all kinds of people. A typical or high resting ECG and a level slant (dextrocardia) are possible high-risk pointers for women. Just a single rule, communicating a high resting, not set in stone to be determinant in men. This suggests that the resting ECG condition is a vital differentiator in foreseeing heart disease. If the slant is up, the quantity of shaded vessels is zero, and the old pinnacle is not exactly or equivalent to 0.56 while contrasting the soundness of people it is more predicted towards the heart disease.

3 Proposed Method

The architecture of the future version for calculating cardio/coronary heart disease is shown in Fig. 1. Two of the 13 features in the facts set, relating to age and communication, are used to calculate the patient's non-public indicators. The final features are kept for critical consideration because they contain essential clinical information. Clinical data are essential for predicting and understanding the severity of an aerobic/heart problem. It gathers information and implements taxonomy procedures, including the HRFLM algorithm. Later on, the variety of the final results may be predicted, and accuracy will be taken into account.

3.1 Data Flow Diagram

A bubble chart is another name for a data flow diagram. It is a genuine graphical formalism for addressing a machine concerning an information report to the gadget, different handling allotted in these insights, and consequently the result records are created by this technique. One of the main demonstrating devices is the data flow diagram, as seen in Fig. 2. Mimicking the framework additives is standard.

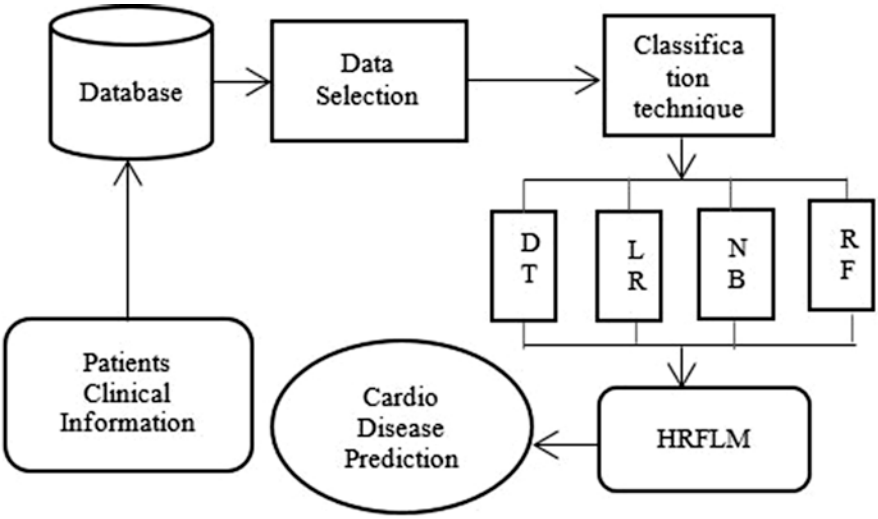


Fig. 1 System architecture

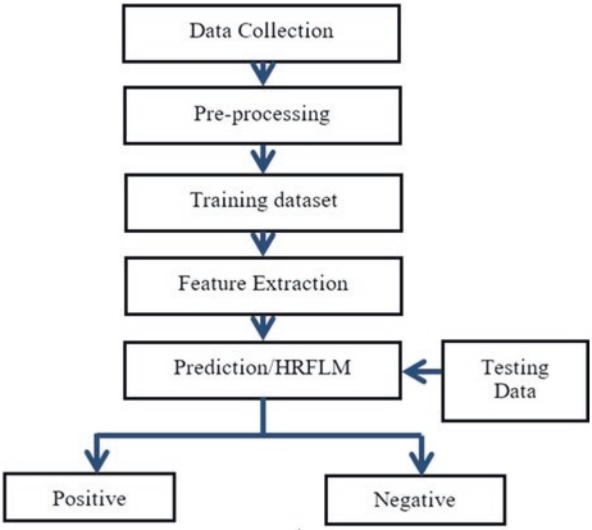


Fig. 2 Data flow diagram

3.2 UML Diagrams

In the domain of computer programming, the unified modeling language (UML) is a predictable, broadly useful display language. The UML is a basic part of article, situated for improvement in the product progression process. To address the plan of programming projects, the UML utilizes graphic documentation.

The following are the primary goals [21] of the UML design:

- To provide customers with an easy-to-use, expressive visual modeling language so that they can expand and trade major trends.
- To provide methods for extendibility and specialization.
- To be self-contained in terms of programming languages and the development process.
- To provide a solid foundation for understanding how to use the modeling language.
- To encourage higher-level enhancement concepts such as partnerships, frameworks, styles, and components.
- To incorporate best practices.

3.3 Sequence Diagram

In the UML, a sequence diagram (displayed in Fig. 3) is a type of cooperation diagram that shows how cycles associate with each other. It comprises a message sequence diagram. Occasion diagrams, occasion situations, and timing diagrams are terms used to describe arrangement diagrams [22].

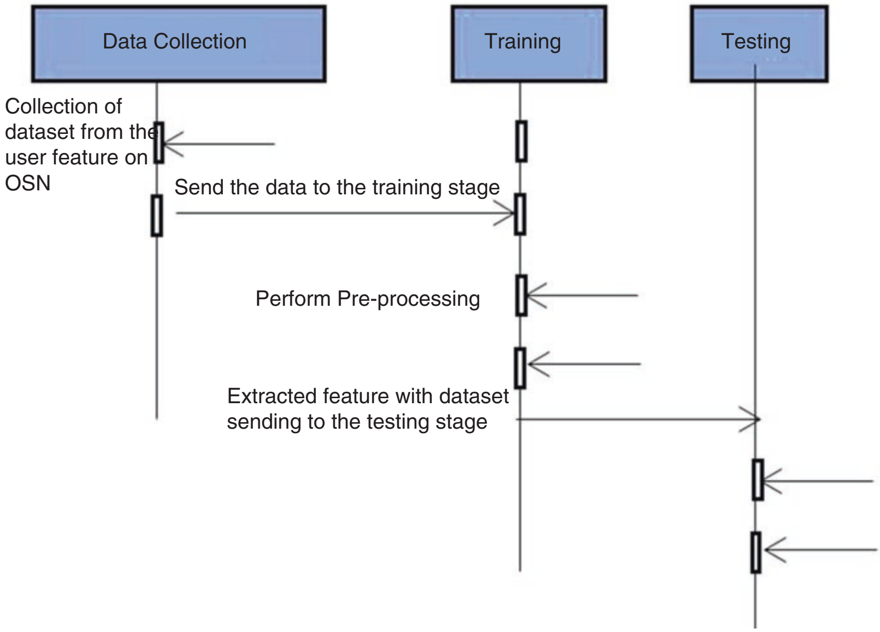


Fig. 3 Sequence diagram

4 System Implementation

The following modules are present in the proposed method:

- Data collection
- Data preparation
- Model selection
- Analysis and prediction

4.1 Data Collection

This is the primary step towards the improvement accumulating records. This is a crucial step to cascade and relate to how the original model will be. If the records obtained are better and high, the output will be accurate. There are several strategies like net scraping and guide interventions etc., for acquiring the statistical data. The Cleveland Heart Disease dataset from the UCI repository will be used in this project. The dataset is made up of 303 unique facts. The dataset [23] contains 13 rows, which are shown in Table 1.

Table 1 Data set

Facts	Description
Sex	The gender is displayed using the following format: 1 denotes a male and 0 denotes a female
Chest-pain type	Displays the individual's type of chest discomfort in the following format: 1 = typical angina; 2 = atypical angina; 3 = non-anginal pain; asymptotic = 4
Resting blood pressure	Displays the gender using the following format: 1 = male; 0 = female
Serum cholesterol	The serum cholesterol level is displayed in milligrams per deciliter (mg/dl) (unit)
Fasting blood sugar	The fasting blood glucose cost of a person with 120 mg/dl is compared. If fasting blood sugar >120 mg/dl then: 1 (real) else: zero (fake)
Resting ECG	Shows resting electrocardiographic outcomes 0 = regular; 1 = having ST-T wave abnormality; 2 = left ventricular hypertrophy
Max cardio/heart rate achieved	Displays the maximum cardio/heart rate achieved by a person
Exercise induced angina	1 = yes, 0 = no
Peak exercise ST segment	1 = up-sloping, 2 = flat, 3 = down-sloping
ST depression induced by exercise relative to rest	Displays the value, which is an integer or a float
Number of major vessels (0–3) color by fluoroscopy	Displays the value as an integer or a float
Thalassemia	Displays the thalassemia: 3 = normal; 6 = fixed defect; 7 = reversible defect
Diagnosis of cardio/heart disease	Displays whether or not the individual has cardio/heart disease: 0 = absent; 1 = present

4.2 Data Preparation

Data Must be clean and no redundant when it is used for any analysis. Randomization of records, eradicate the consequences of the genuine request inside, which is accumulated as well as in some other case. Picture records are splitted into tutoring and evaluation units.

4.3 Model Selection

Random forests are one the most famous system learning algorithm. They are so successful because they provide predictive performance, low over-fitting, better interpretability etc. This interpretability is given by using the fact that it is simple to derive the importance of every variable on the tree selection. In other words, it is straightforward to compute how much each variable is contributing to the selection.

Feature choice using random forests comes under the category of embedded strategies. They are carried out using algorithms that have their own built-in feature choice methods.

4.4 Collaboration Diagram

Collaboration diagrams describe interactions among classes and associations. Here, as shown in Fig. 4, all through this venture the collaboration diagram incorporates the flow that gathers the information set from the user function on the output sequence number and sends the information to the education level, performs preprocessing, and teaches the dataset and extract feature by sending the dataset to the test stage. Subsequently, admin views the information and upload documents as per that requirement.

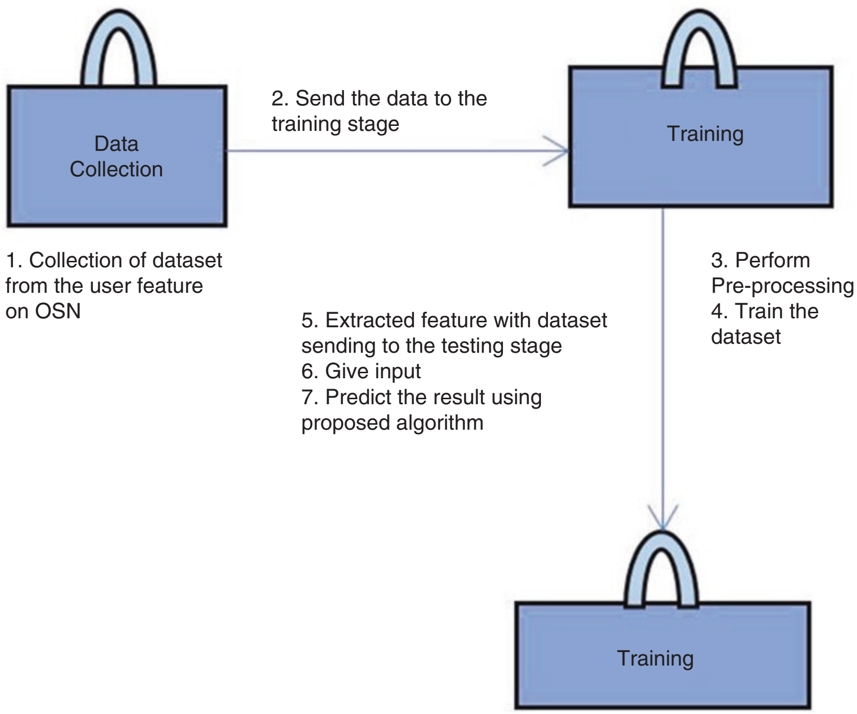


Fig. 4 Collaboration diagram

5 Analysis and Prediction Results of HRFLM

5.1 Parameters used for analysis

The parameters used for the analysis and prediction of heart disease [24, 25] are as follows.

True positives: people who self-identify as having a condition truly have it; in other words, the true positive represents the number of people who are unwell and self-identify as such.

$$\text{True Positive} = \frac{\text{Number of cases correctly identified as positive for disease}}{\text{Total number of observations}} \tag{1}$$

True negatives: people who say that they do not have the condition are actually discovered to be free of it; in other words, the true negative represents the number of people who are healthy and say that they are healthy.

$$\text{True Negative} = \frac{\text{Number of cases correctly identified as negative for disease}}{\text{Total number of observations}} \tag{2}$$

False positives: people who claim to be afflicted with the sickness are actually discovered not to be afflicted. In other words, the number of persons who are healthy but are incorrectly diagnosed as being ill are represented by the false positive.

$$\text{False Positive} = \frac{\text{Number of cases incorrectly identified as positive for disease}}{\text{Total number of observations}} \tag{3}$$

False negatives: people who are afflicted with the disease are anticipated not to be afflicted with the condition. The false negative, on the other hand, shows the number of persons who are ill but have been mislabeled as healthy.

$$\text{False Negative} = \frac{\text{Number of cases incorrectly identified as negative for disease}}{\text{Total number of observations}} \tag{4}$$

Accuracy: accuracy is a great measure while the goal variable classes in the data are almost balanced. Accuracy is a relevant measure for a binary classifier. For a binary classifier that classifies instances into positive (1) and negative (0) times, any single prediction can fall into one of four terms below:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}. \tag{5}$$

Sensitivity: the number of true-positive instances identified as such is referred to as sensitivity (or true positive). Sensitivity is also known as recall. It indicates that the number of actual good circumstances that are incorrectly categorized as bad will rise (and, thus, could also be termed false negative). It can also be described as a false negative rate. 1 would be the sensitivity-to-false-negative-rate ratio. Let us look at the process for determining whether or not someone has the disease. Sensitivity refers to the percentage of patients who were accurately diagnosed with the condition. In another sense, the sick individual was expected to get sick. Sensitivity is calculated using the following formula:

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}. \tag{6}$$

The true-positive value is higher and the false-negative value is lower as sensitivity increases. The true-positive value is lower and the false-negative value is higher as the sensitivity decreases. In the health care and banking industries, very sensitive models will be desired.

Specificity: specificity is characterized as the quantity of true negatives that were recognized as negatives. Thus, more true negatives will be seen as positives, bringing about false positives. This proportion is otherwise called the false-positive rate. The amount of particularity and false-positive rate is generally 1 in this situation. How about we investigate the system for deciding if somebody has the condition? The negligible proportion of people who are not impacted by a precisely anticipated sickness is known as specificity. Specificity is likewise useful:

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}. \tag{7}$$

The higher the specificity, the lower the false-positive rate and the higher the true-negative rate. The higher the specificity, the lower the true-negative value and the greater the false-positive value.

The specificity of a test is characterized in various ways. For example, specificity being the capacity of a screening test to recognize a true negative, being founded on the true-negative rate, appropriately distinguishing individuals who do not have a condition/sickness, or on the other hand, if 100%, distinguishing all the patients who do not have the infection from those individuals testing negative. The prediction models are created using 13 features, and the various parameters for the various techniques are calculated and displayed in Table 2. The table compares parameters such as precision, sensitivity, F-measure, and accuracy. With the suggested prediction approach, the parameters are compared with several current methods. In comparison with existing approaches, the suggested HRFLM classification method achieves the maximum accuracy, sensitivity, and F-measure, as shown in the comparison graph.

Table 2 Comparison of various algorithms against different parameters

Algorithms	Accuracy	Sensitivity	F-measure	Precision
Logical regression	84.8%	89.92%	85.4%	81.24%
Decision tree	85.4%	84.9%	84.6%	83.1%
Random forest	86.9%	84.9%	86.1%	86.4%
Naïve Bayes	74.8%	78.9%	77.8%	80.3%
Support vector machine	75.4%	82.4%	80.3%	82.1%
Ensemble classifier	89.1%	88.3%	86.8%	89.1%
Hybrid random forest with linear model (proposed)	89.9%	90.02%	87.1%	88.7%

The UCI dataset is additionally grouped into eight styles of datasets upheld by arrangement rules. Each dataset is additionally grouped and handled by Rattle in RStudio. The outcomes are produced by applying the order rule for the dataset.

Combining the properties of random forest and linear method, the proposed hybrid HRFLM technique is applied. As a result, it was found to be quite accurate in predicting heart disease. The accuracy rate for a dataset is expressed as a percentage of correct predictions. According to this method, if we have a machine learning model with a 90% accuracy rate, we can anticipate having 90 accurate predictions out of every 100. Compared with existing models, machine learning approaches focus on the best performing model. The HRFLM is the model that predicts heart/cardio disease with high accuracy and low classification error, which is shown in Fig. 5.

5.2 Simulation Model for Prediction of the Disorder

The home web page shown in Fig. 6 includes the identity of the challenge with branding records and the brand of the venture. The intention of the homepage is not to be a library of textual content and content material but alternatively to characterize the manual in the direction of the pages that have the desired statistics.

Admin login may be a set of credentials accustomed to authenticating a consumer, as shown in Fig. 7. Most frequently, those contain a username and password. They are a protection measure designed to stop unauthorized access to personal facts. When a login fails (i.e., the username and password aggregate does not fit a consumer account), the user refused access. The Uploading and transmission of a file from one computing system to a one that has an upload option, as shown in Fig. 8, generally a large, computerized statistics processing device to add a document is to send it to a different PC.

A statistics view might be a gadget or visible representation of facts that differs from physical facts. The sample statistical data view of the uploaded image is shown in Fig. 9. Views are frequently created to form records that are more applicable, readable, and thrilling for human consumption. The shape or the visualization of information always differs from an information repository.

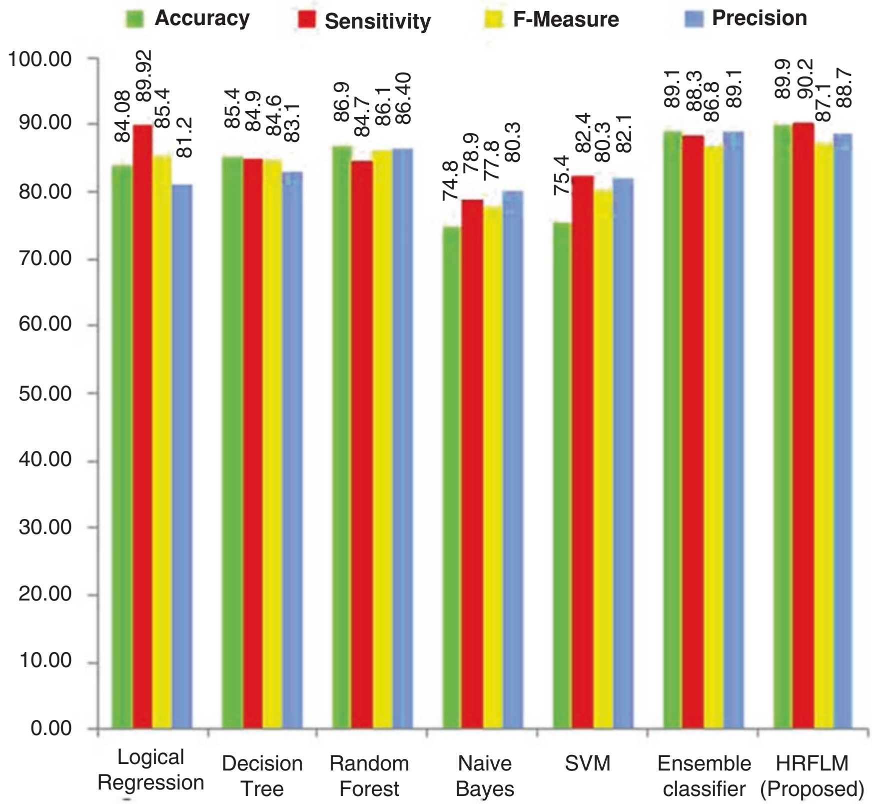


Fig. 5 Comparison of algorithms versus parameters

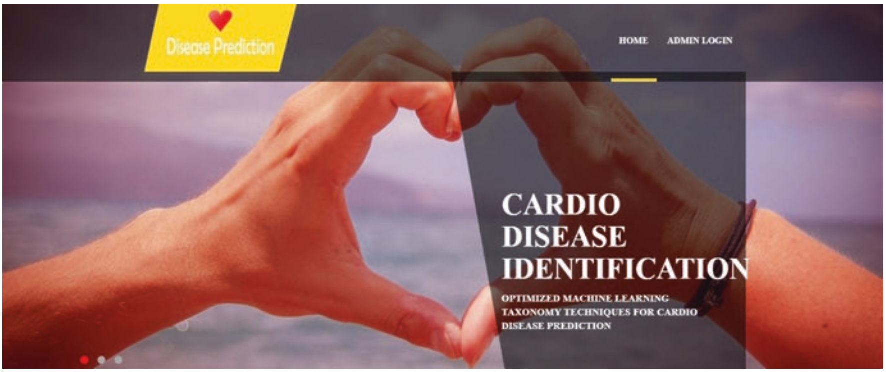


Fig. 6 Home page

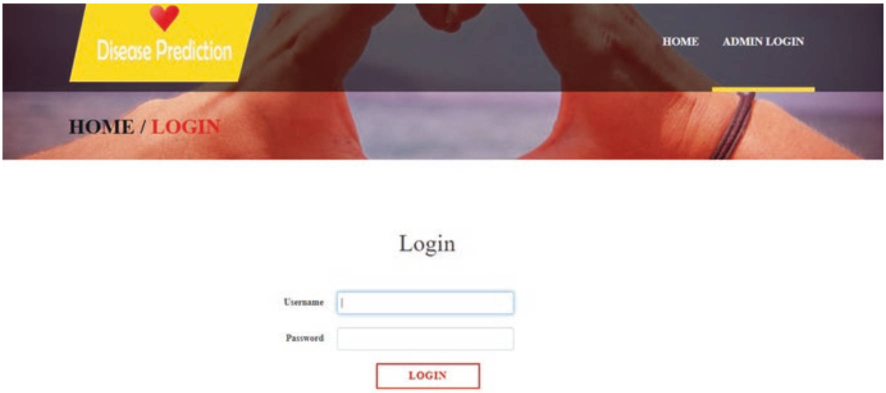


Fig. 7 Admin login page

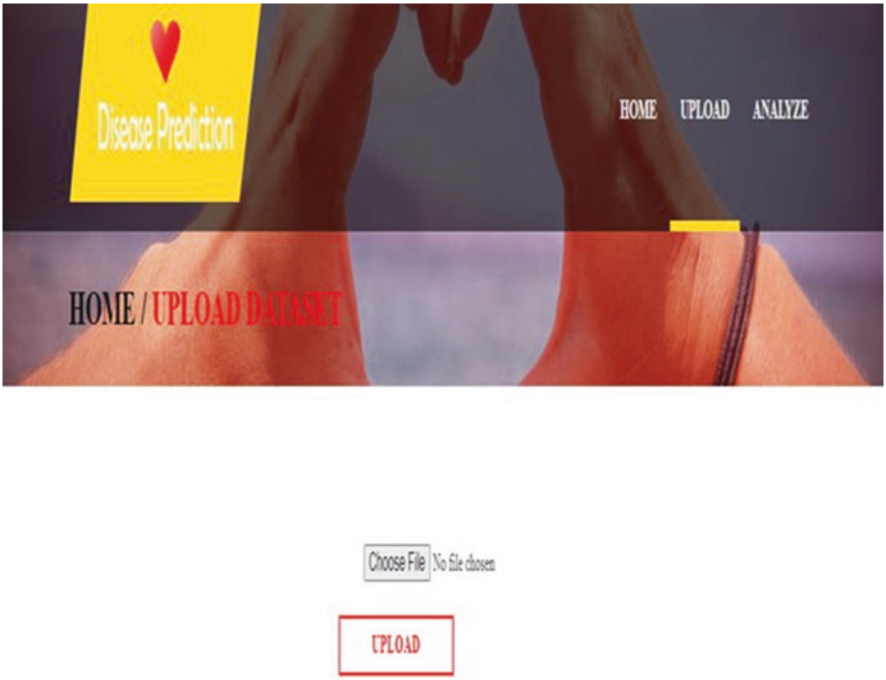
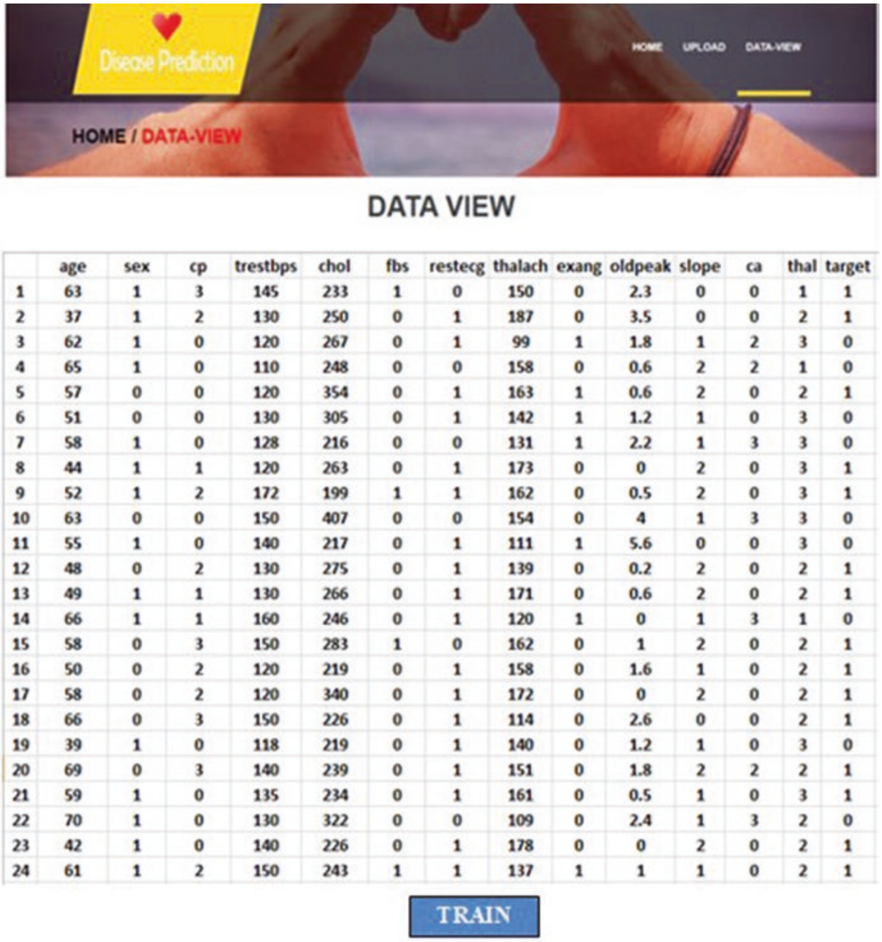


Fig. 8 Upload file page



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
1	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
2	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
3	62	1	0	120	267	0	1	99	1	1.8	1	2	3	0
4	65	1	0	110	248	0	0	158	0	0.6	2	2	1	0
5	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
6	51	0	0	130	305	0	1	142	1	1.2	1	0	3	0
7	58	1	0	128	216	0	0	131	1	2.2	1	3	3	0
8	44	1	1	120	263	0	1	173	0	0	2	0	3	1
9	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
10	63	0	0	150	407	0	0	154	0	4	1	3	3	0
11	55	1	0	140	217	0	1	111	1	5.6	0	0	3	0
12	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
13	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
14	66	1	1	160	246	0	1	120	1	0	1	3	1	0
15	58	0	3	150	283	1	0	162	0	1	2	0	2	1
16	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
17	58	0	2	120	340	0	1	172	0	0	2	0	2	1
18	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
19	39	1	0	118	219	0	1	140	0	1.2	1	0	3	0
20	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
21	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
22	70	1	0	130	322	0	0	109	0	2.4	1	3	2	0
23	42	1	0	140	226	0	1	178	0	0	2	0	2	1
24	61	1	2	150	243	1	1	137	1	1	1	0	2	1

Fig. 9 Statistical data view

Figure 10 shows data analyzed from the statistical information. The data evaluation is the process of accumulating and organizing records so as to draw helpful conclusions from them. The approach to fact evaluation makes use of analytical and logical reasoning to realize statistics.

The positive predictive value is the chance that patients with a positive screening test genuinely have the disease. The positive predictive value definition is similar to the sensitivity. However, positive predictive value and sensitivity are more beneficial to the doctor. Positive prediction will also provide the probability of someone having a disease, as shown in Fig. 11.

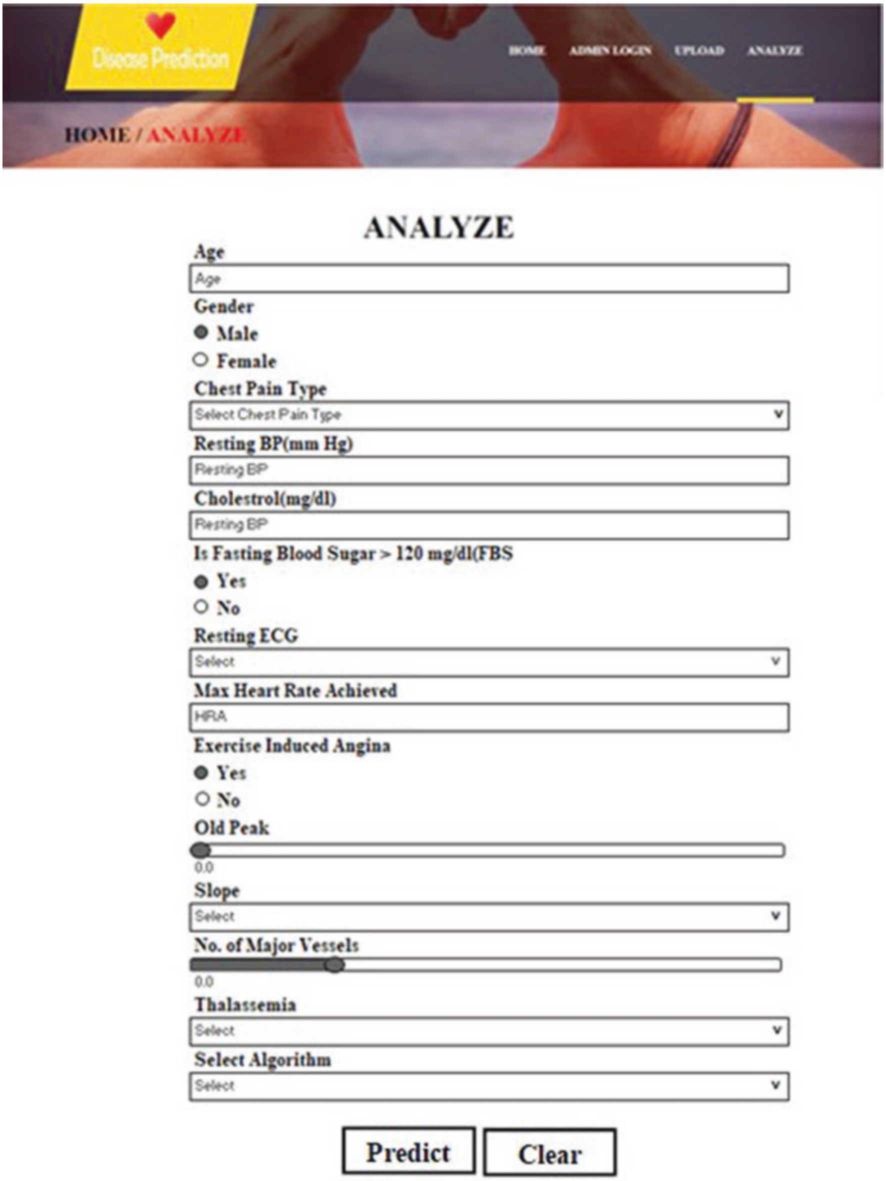


Fig. 10 Data analysis

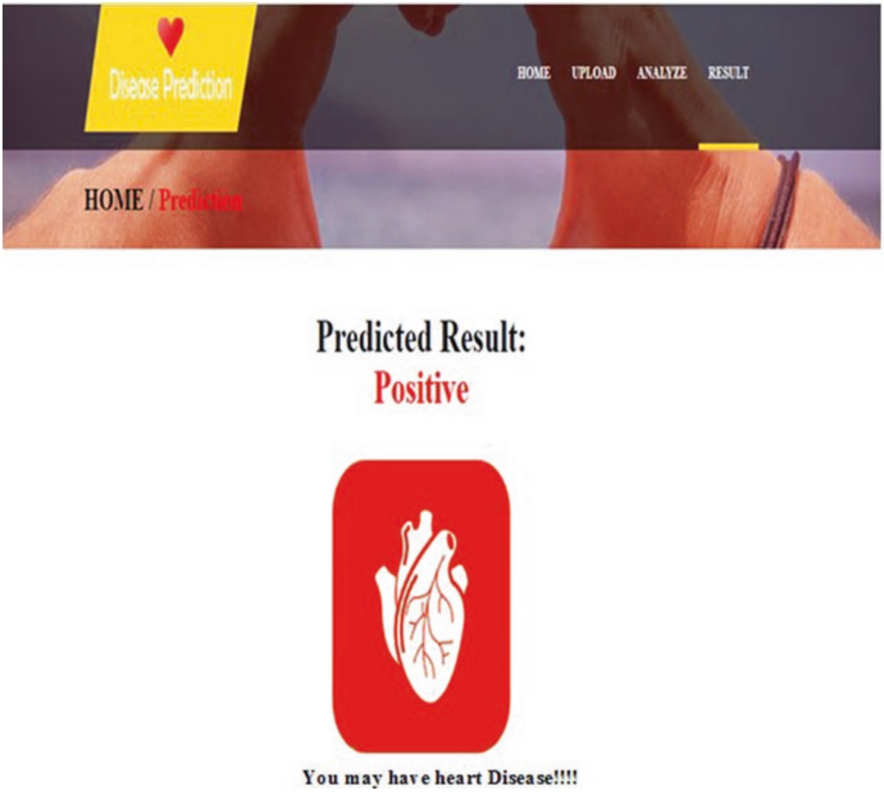


Fig. 11 Positive predictive value result

Figure 12 shows the negative predictive value, which represents the chance that an individual would not have a disease or circumstance, i.e., the negative predictive value represents the percentage of people with a negative test who are efficiently identified or recognized.

6 Conclusion

The prognosis for cardiovascular diseases using data mining methodologies when contrasted with past methodologies, HRFLM gives further developed exactness and improvement in all boundaries. The discoveries of the correlation show that when contrasted with the other individual calculations, the random forests strategy delivers improved results. As far as foreseeing cardio/heart disease, the HRFLM was demonstrated to be genuinely precise. The recommended approach joins random forests and with a direct model, and developed execution. The underlying algorithm utilizes the attributes klike age, orientation, CP, Tresbps, Chol, fbs, Restecg, Thalach, Exang, oldpeak Slope, Ca, Thal Target etc., for result prediction.

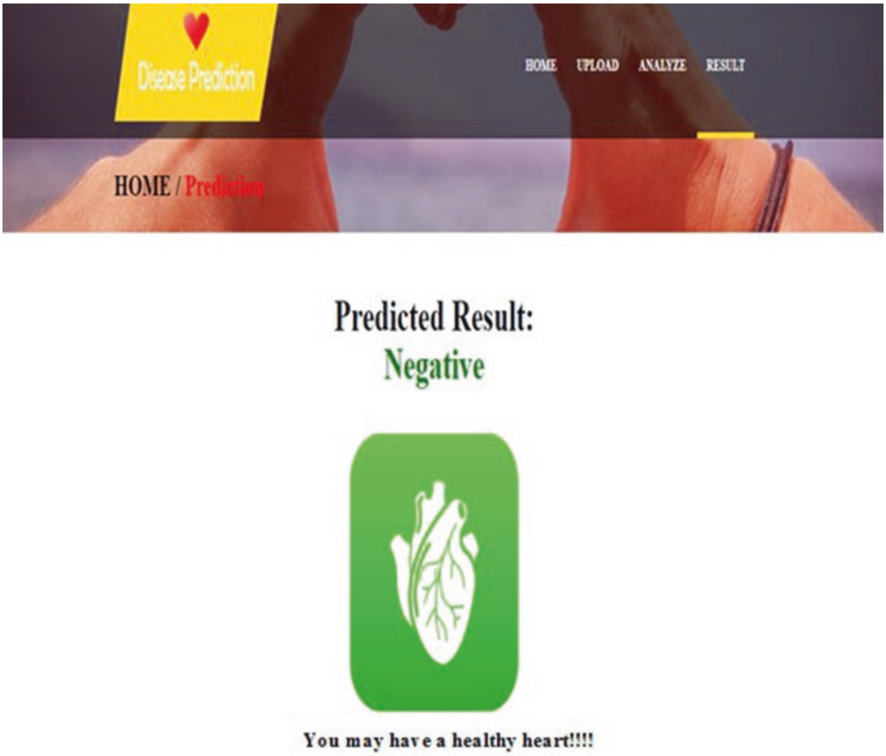


Fig. 12 Negative predictive value result

The future objective of this examination is to further develop prediction calculations by utilizing different combinations of machine learning algorithms. Later on, this exploration can be done utilizing different blends of AI calculations to further develop prediction strategies. Moreover, new component selection methods could be created to improve the heart disease prediction.

References

1. Winham SJ, de Andrade M, Miller VM. Genetics of cardiovascular disease: importance of sex and ethnicity. Atherosclerosis. 2015;241(1):219–28.
2. Everson-Rose SA, Lewis TT. Psychosocial factors and cardiovascular diseases. Annu Rev Public Health. 2005;26:469–500.
3. Sarkar D, Bali R, Sharma T. Practical machine learning with python. A problem-solvers guide to building real-world intelligent systems. Berkely: Apress; 2018.
4. Qiu Z, Liu D, Cui W. Cardiac diseases classification based on scalable pattern representations. In: 2016 3rd International Conference on Systems and Informatics (ICSAI). Piscataway: IEEE; 2016. p. 1034–9.
5. Ray S. A quick review of machine learning algorithms. In: 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon). Piscataway: IEEE; 2019. p. 35–9.

6. Angra S, Ahuja S. Machine learning and its applications: a review. In: 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC). Andhra Pradesh, India: IEEE; 2017. pp. 57–60.
7. Ekız S, Erdoğmuş P. Comparative study of heart disease classification. In: 2017 Electric Electronics, Computer Science, Biomedical Engineering' Meeting (EBBT). Piscataway: IEEE; 2017. p. 1–4.
8. Hamdy A, El-Bendary N, Khodeir A, Fouad MM, Hassanien AE, Hefny H. Cardiac disorders detection approach based on local transfer function classifier. In: 2013 federated conference on computer science and information systems. Piscataway: IEEE; 2013. p. 55–61.
9. Kotsiantis SB, Zaharakis ID, Pintelas PE. Machine learning: a review of classification and combining techniques. Artif Intell Rev. 2006;26(3):159–90.
10. Yan H, Jiang Y, Zheng J, Peng C, Li Q. A multilayer perceptron-based medical decision support system for heart disease diagnosis. Expert Syst Appl. 2006;30(2):272–81.
11. Newman DJ, Hettich SC, Blake CL, Merz CJ. UCI repository of machine learning databases. UCI Machine Learning Repository 1998.
12. Lu X, Peng X, Liu P, Deng Y, Feng B, Liao B. A novel feature selection method based on CFS in cancer recognition. In: 2012 IEEE 6th International Conference on Systems Biology (ISB). Piscataway: IEEE; 2012. p. 226–31.
13. Priyanka N, Kumar PR. Usage of data mining techniques in predicting the heart diseases – Naïve Bayes & decision tree. In: 2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT). Piscataway: IEEE; 2017. p. 1–7.
14. Jaidhan BJ, Madhuri BD, Pushpa K, Devi BL. Application of big data analytics and pattern recognition aggregated with random forest for detecting fraudulent credit card transactions (CCFD-BPRRF). Int J Recent Technol Eng. 2019;7(6):1082–7.
15. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. In: 2008 IEEE/ACS international conference on computer systems and applications. Doha, Qatar: IEEE; 2008. pp. 108–15.
16. Thomas J, Princy RT. Human heart disease prediction system using data mining techniques. In: 2016 international conference on circuit, power and computing technologies (ICCPCT). Piscataway: IEEE; 2016. p. 1–5.
17. Gavhane A, Kokkula G, Pandya I, Devadkar K. Prediction of heart disease using machine learning. In: 2018 second international conference on electronics, communication and aerospace technology (ICECA). Piscataway: IEEE; 2018. p. 1275–8.
18. Esfahani HA, Ghazanfari M. Cardiovascular disease detection using a new ensemble classifier. In: 2017 IEEE 4th international conference on knowledge-based engineering and innovation (KBEI). Piscataway: IEEE; 2017. p. 1011–4.
19. Gandhi M, Singh SN. Predictions in heart disease using techniques of data mining. In: 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE). Piscataway: IEEE; 2015. p. 520–5.
20. Nahar J, Imam T, Tickle KS, Chen YP. Association rule mining to detect factors which contribute to heart disease in males and females. Expert Syst Appl. 2013;40(4):1086–93.
21. Sun D, Wong K. On evaluating the layout of UML class diagrams for program comprehension. In: 13th International Workshop on Program Comprehension (IWPC'05). Los Alamitos: IEEE; 2005. p. 317–26.
22. Souri A, Ali Shariflo M, Norouzi M. Formalizing class diagram in UML. In: 2011 IEEE 2nd international conference on software engineering and service science. Piscataway: IEEE; 2011. p. 524–7.
23. Rajkumar A, Reena GS. Diagnosis of heart disease using datamining algorithm. Global J Comp Sci Technol. 2010;10(10):38–43.
24. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access. 2019;19(7):81542–54.
25. Raja MS, Anurag M, Reddy CP, Sirisala NR. Machine learning based heart disease prediction system. In: 2021 International Conference on Computer Communication and Informatics (ICCCI). Piscataway: IEEE; 2021. p. 1–5.