2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)
Bangalore, India. Oct 7-9, 2022

Comprehensive Analysis of Heart Disease Prediction: Machine Learning Approach

Swetha Sivakumar
Department of Computer Science and Engineering
Siddaganga Institute of Technology
Tumkur,India
swetshiv@gmail.com

Pramod T.C
Department of Computer Science and Engineering
Siddaganga Institute of Technology
Tumkur, India
tcpramodhere@gmail.com

2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT) | 978-1-6654-6855-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/GCAT55367.2022.9972035

***Abstract*— cardiovascular disease remains the major cause of fatality for both men and women worldwide. Heart disease is on the rise in both old and the young of males and females in today's society. As a result, developing and implementing comprehensive health-tracking rules should be spotlight in order to tackle the epidemic of heart-associated illnesses. As a result, early detection and treatment, using both traditional and novel techniques, must be prioritized. The primary goal of this study is to determine the best classifying approach for heart disease-related health data and the factors that impact it. This comprehensive work is based on the performance of systems that have been evaluated and described using various models presented in various research papers, and it provides a complete review of those research papers in order to set up the heart disease prognostication model and its performance.**

Keywords—Heart Disease, Predictive Analysis, Machine Learning, Heart Attack.

I. Introduction

Heart Pathology, often known as cardiac disease, is characterized by decreased heart or blood vessel function. It can increase the risk of heart stroke, cardiogenic shock, sudden cardiac death, myocardial infarction, and cardiac rhythm difficulties, all of which can lead to a much worse quality of life and a shorter life expectancy. Skeletal injury, infection, eczema, the surroundings, lifestyle, and hereditary are all factors that play a role in cardiac disease. People should adopt a healthy lifestyle that includes avoiding smoking, greasy foods, and stress to prevent cardiovascular disease. Many countries throughout the world are concerned about myocardial infarct. Every 33 seconds, somebody from the U. S. dies from myocardial infarct. According to WHO statistics, 17.5 million people worldwide die a year as a consequence of cardiac disease.

Healthcare organizations are storing large volumes of data in databases that are exceedingly complicated and difficult to evaluate as digital technology advances. Heart disease diagnosis must be precise and exact. A medical practitioner is generally the one to diagnose it. Adopting a process that is incorporated into the medical information system is advantageous and less expensive. After evaluating the applicability of various data mining methodologies, this may be done. To identify hidden patterns in massive datasets, data mining incorporates analysis of statistical data, machine learning procedures, and database management systems.

Researchers have demonstrated that machine learning algorithms respond very well when processing health records data sets. These data sets will be instantly put into machine learning algorithms, which will work as expected and bring results. The Cleveland Dataset from UCI, Kaggle, and the Long Beach VA heart disease database from UCI are the most commonly utilized datasets in many papers. The commonly used attributes in the dataset for prediction of heart disease are as follows:

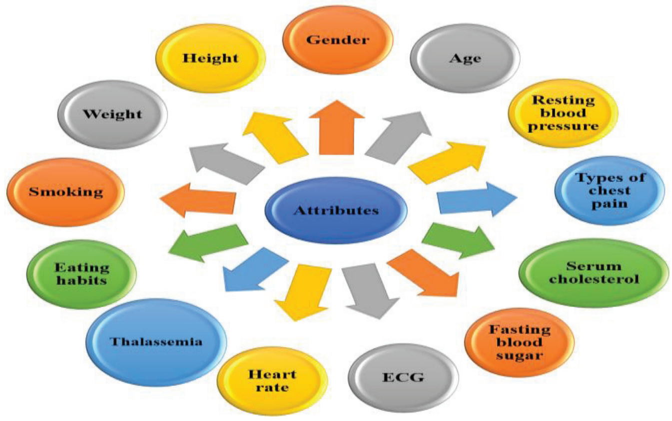

Fig. 1. Attributes used to predict heart diseases

A. Gender

They are categorized into three groups: Male\s Female\s and Neutral. Men are the most common group to be impacted by heart disease. Heart disease affects 255-525 males per 100,000 people and 225-299 women per 100,000 people.[16]

B. Age

People who are 60 or older are more prominently observed to have heart diseases [17].

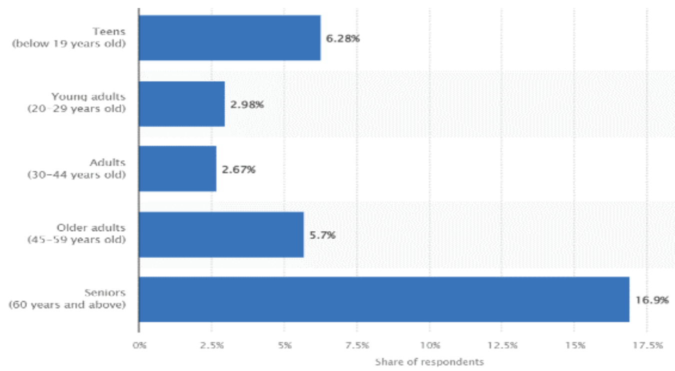

Fig. 2. Statistic of heart disease diagnosed in different age group.

C. Resting Blood Pressure

Ideal blood pressure is usually stated as 120 mm Hg systolic (pressure measured when your heart is beating) and 80 mm Hg diastolic (pressure measured when your heart is at rest). The ideal for your diastolic heart rate is between 60 and 100 beats per minute (BPM).[18].

D. Types of Chest Pain

There are various types of Chest pain some of which are listed below:[19]

1) Left Chest Pain: Pain on the left side of the chest might be very significant. It might be a heart attack or another medical ailment such as lung troubles or heart inflammation

2) Angina: Angina is a condition in which there is a lack of oxygen-rich blood going to a specific area of the heart. Because of fatty deposits in the arterial walls, the arteries of the heart narrow. Angina is given rise to a reduction in blood flow to the heart due to artery constriction

3) Pericarditis: Pericarditis is a condition in which the pericardium becomes inflamed (the fibrous sac surrounding the heart). The pericardium allows the heart to work efficiently by preventing the heart from over-expanding as blood volume grows. Pericarditis causes significant chest pain and discomfort while breathing heavily.

4) Pleuritic chest pain: Pleura or pleuritis refers to the delicate tissues that border your lungs and chest wall. Pleura becomes irritated and bloated when infected or inflamed, producing intense chest discomfort while breathing, coughing, or sneezing. Pleurisy or pleuritis is the medical term for this illness

5) Pulmonary embolism: A pulmonary embolism occurs when the lung pulmonary arteries can get clogged. Blood clots that travel from veins in the legs and other parts of the body to the lungs are the most frequent cause of pulmonary embolism

E. Serum Cholesterol

A person's serum cholesterol level is used to determine how much total cholesterol is present in the blood. Serum cholesterol is the total amount of high-density lipoprotein (HDL), low-density lipoprotein (LDL), and triglycerides in a person's blood. Most healthy people should have total cholesterol less than 200.0 mg/dL, LDL less than 100.0mg/dL, and HDL greater than 40 mg/dL. (20 and above). Children aged below 20 years should have total cholesterol levels less than 170.0mg/dL, LDL levels less than 110.0mg/dL, and HDL levels greater than 45.0mg/dL.[20]

F. Fasting Blood Sugar

The blood sugar level measured after an individual has fasted overnight is known as fasting blood sugar. The test is performed first thing in the morning, before the individual has eaten anything. Normal blood glucose levels are between 70 and 100 mg/dl. Long-term blood sugar levels exceeding 99 mg/dl have been linked to a higher risk of heart disease and stroke as per studies.

G. ECG

An ECG, or electrocardiogram, is an instrument that assesses the rhythm and electrical activity of your heart's conduction system. Because it is this electrical activity that causes your heart to contract, any abnormalities with your heart's pace or rhythm may be detected by detecting it. The normal ranges from 60.0-99.0 bpm. [21]

H. Heart Rate

The count of how regularly one's heart beats in one minute is termed as one's heart rate or pulse. Adults should have a diastolic heart rate of 60 to 100 beats per minute.

I. Thalassemia

The occurrence of an aberrant kind of hemoglobin is the signature of this genetic blood condition. It causes stomach enlargement, exhaustion, a yellowish complexion, black urine, and anomalies of the facial bones. Severe thalassemia can cause congestive heart failure and irregular cardiac rhythms.

J. Eating Habits

Excessive usage of Caffeine and alcohol, eating packaged, canned, and processed foods, as well as Maida, and Deep-fried and salty foods may cause harm to heart.

K. Smoking

The primary risk factor for heart disease is smoking. Immediately after the smoke, the chemicals harm the heart and blood vessels, increasing the risk factor of atherosclerosis, or plaque accumulation in arteries.

L. Weight

BMI (Body Mass Index) is a metric that may be used to determine body weight. BMI ranges are classified as follows by the National Institutes of Health (NIH): Underweight >19.0, Normal Weight: 18.6 – 24.9, Overweight: 25 -30 and Grade_1 Obesity: 31-35, Grade_2 Obesity: 36-40 and Grade_3 Obesity< 41.0.

M. Height

Height is a metric of vertical distance, either in terms of vertical stretch or vertical position. Many variables influence height, including inherited factors, physical exercise, and so on.

Rest of the paper is organized as follows: Section 2 discusses techniques for heart disease prediction, Section 3 gives the details of different classification of Machine Learning Algorithm, Section 4 describes about the general architecture for predicting heart diseases, Section 5 gives insights about existing work on heart disease prediction, Section 6 gives the comparative analysis of papers on heart prediction and followed by conclusion in Section 7.

II. TECHNIQUES FOR HEART DISEASE PREDICTION

The various techniques with which heart disease can be predicted is given in Figure 3 and is discussed below in brief.

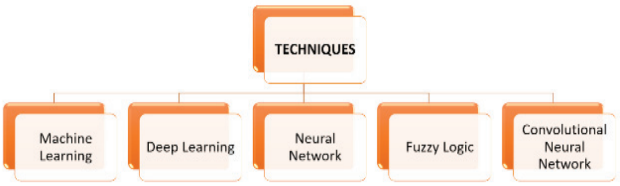

Fig. 3. Techniques used for heart disease prediction

A. Machine Learning

Machine learning is an Artificial Intelligence (AI) implementation that enables an architecture to learn and enhance from knowledge without having to be explicitly designed. Machine learning's goal is to create an Artificial Intelligence (AI) platform that is superior to the human intellect. Several AI researchers believe that this goal may be accomplished using machine learning algorithms that imitate the learning phase of the human brain.

B. Deep Learning

Deep Learning is a technology that resembles the human brain in that it is made up of numerous neurons with various layers, similar to the human brain. The network consists of a data entry layer, a data exit layer, and one or more hidden layers. The network produces predictions after attempting to learn from the data it has received.

C. Neural Network

Artificial neural networks, often known as ANNs, are software programs that are based on biological neural networks and are capable of performing a wide range of tasks while utilising a wide range of data. Several algorithms are used to grasp the relationships in a given set of data in order to acquire the best results from changing inputs. Given the data, several models are used to predict future outcomes, and the network is trained to provide the necessary outputs. The way the nodes are connected allows it to operate similarly to how the human brain does.

Perceptron’s, a kind of machine learning unit, are used by an artificial neural network known as a convolutional neural network (CNN) to analyse input. Image processing, natural language processing, and other cognitive tasks may benefit from CNNs. ConvNet might be another title for a convolutional neural network.

D. Fuzzy Logic

The term "fuzzy" refers to things that are unclear or ambiguous. Sometimes, in real life, we are unable to determine whether a given problem or statement is true or false. This idea provided a variety of values between true and false at the time, allowing for flexibility in problem-solving. Fuzzy logic is a growing field in which we can create hybrid systems for heart disease detection and prediction. It's difficult to diagnose the disease without laboratory tests because multiple diseases have similar symptoms. But from the fuzzy software, similar types of expert systems have been developed in almost every field of medicine.

III. Classification Of Machine Learning Algorithms

The concept "supervised learning" refers to learning that takes place in the existence of an overseer or an instructor. It indicates that for certain data sets, a group of data sets that is labelled that exists earlier with the intended output. The computer is then provided with new data records to evaluate and create the proper output based on the training data sets.

There isn't an instructor or supervisor present in Unsupervised Learning. The data isn't categorized or classified in this case, and the network has no previous instruction. Without any prior training, the machine must organize the data records presented in accordance with motifs, affinities, and distinctions.

In reinforcement learning, the input-output mapping is done by interacting with the conditions continuously for the sake of reducing the scalar performance index. A critic, rather than a teacher, turns the main the reinforcement signal—a scalar input derived from the environment—into a heuristic reinforcement signal, which is a higher-quality reinforcement signal. Reduce the cost to go function, which is the estimated total cost of the activities carried out throughout a number of phases, is the aim of this training.

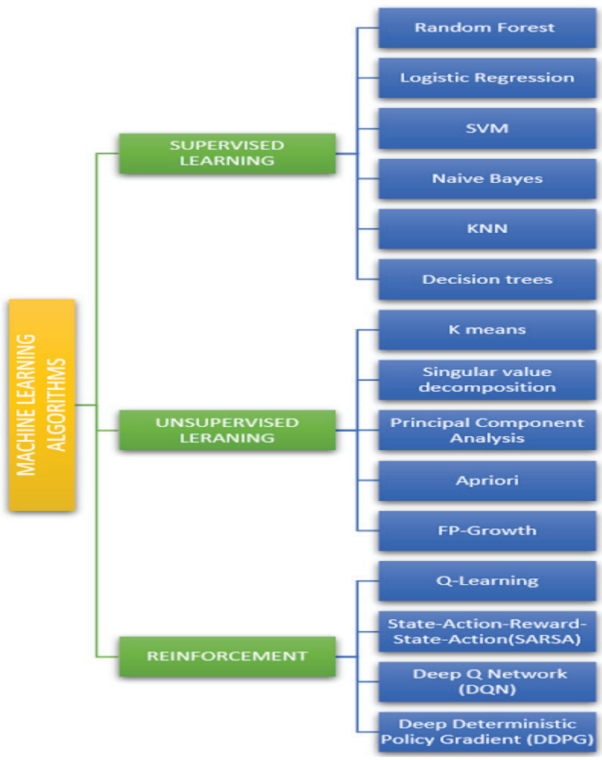

Fig 4: Classification of Machine Learning Algorithms

IV. General Architecture for Prediction Of Heart Diseases

The General Architecture for prediction of heart disease is shown in Figure 5.

Step 1: The appropriate data record is found on the University of California, Irvine dataset page.

Step 2: Data cleansing, addressing missing values, estimating some missing values, and data transformation are all examples of data processing.

Step 3: Some features or parameters from the dataset may be chosen or eliminated to improve the model's performance and give better results.

Step 4: On the dataset, a suitable model or combination of models is used to test the models' performance.

Step 5: The resulting outcome reflects the effectiveness of the experimenter's or researcher's working model.

Fig 5: General Architecture for Heart disease Prediction

V. EXISTING WORK ON HEART DISEASE PREDICTION

A. Javeed, [1] introduced a novel framework where two algorithms are combined in the learning system. The very first algorithm is a random search algorithm for locating a feature subset that offer further information on heart failure. The model used in this study, random forest, is used to predict cardiac failure based on a selection of parameter sets. The suggested RSA-RF learning system has been shown to improve the random forest framework's performance by 3.3 percent.

Dinesh Kumar.G,[2] This research adds to the correlative implementation and analyzation of several machine learning methods in R software, allowing users to use machine learning techniques in R application for cardiovascular disease prediction in real time. This non-ethical study attempts to deploy accessible machine learning methods using R software.

Dolatabadi,[3] When compared to earlier research, the given technique in this study uses fewer features to obtain a maximum accuracy of 99.2 percent, which is a significant addition to their work. This study used 10 data points for training, giving it a credible alternative for diagnosing circumstances when the amount of data points available is low. As a consequence, the approach proposed in this study is suitable for practitioners and might be used in hospitals to detect CAD using HR signals. The effectiveness of classifiers depends on the value of the parameters used to develop these functions. This emphasizes how the parameters of the SVM algorithm in this article must be changed.

Mirmozaffari,[4] A novel model with several filters and assessment techniques has been developed. To identify the best algorithm and more accurate clinical decision support systems for ailment detection, researchers used a multilayer filtering preprocess as well as several assessment methodologies. The accuracies, error functions, and construction times of classifiers are compared. Within each level, high-performance algorithms were introduced. The experiment might be used by doctors as a tool to better foresee and counsel patients in unclear situations.

M.Sultana,[5] This study examined a number of data mining techniques to see which one was the most effective at predicting heart disease. Two data sets (manually collected dataset and from the standard UCI database) are used individually for each data mining technique in this study. The Bayes Net and SMO classifiers outperform the other five classifiers examined for prediction of heart disease: Bayes Net, SMO, KStar, MLP, and J48, according to their findings.

Reddy,[6] The Heart Failure database, which is distinct from the UCI Heart Disease dataset, served as the foundation for our study. The DT algorithm can help predict cardiovascular disease since it is based on rules. It creates a prediction while analyzing a set of rules, and then decides. The Naive Bayes algorithm is a better option for predicting sickness in the medical field while keeping in mind the premise of independence.

Rabbi,[7] As a result, a computer-based, pretty accurate prediction system might be a viable option for detecting and treating cardiac disease. As a result, three commonly used data mining classification approaches, SVM model, KNN model, and Artificial Neural Network model, were explored and assessed using a typical Cleveland cardiac disease dataset in this research. On the basis of classification rate, RBF kernel derived from SVM model can outshine KNN model and Artificial Neural Network model, while KNN model can also outperform Artificial Neural Network model. In order to give further therapy and avoid early-stage heart failure, this study also recommends using the highly appraised classifier for real-time prediction of heart disease patients and for detecting the risk factor for heart failure.

Patra,[8] This paper outlines a research need in the area of heart disease prediction using the Python Anaconda explorer, spider, and Weka tool, on which we have spent a lot of effort. Feature selection, NumPy library, panda's library, decision tree classification, Classification algorithm, entropy, gini-index, and confusion matrix are a few approaches and methodologies that have been utilized to train the model of cardiac datasets. The decision tree classifier is the best and most appropriate classifier for forecasting the UCI Cleveland heart dataset, according to the results. The results highlight the value of the decision tree classification technique in the study of medical data, particularly when it comes to categorizing cardiac issues.

Mohan,[9] This research improved the precision of cardiovascular disease prediction by introducing a novel method for detecting critical traits using machine learning techniques. The prediction model is introduced using a variety of distinctive combinations and various well-known classification techniques. They have enhanced and attained a higher performance level of accuracy using the prediction model for heart disease that combines a hybrid random forest with a linear model.

Repaka,[10] Data is gathered from a variety of sources that are key causes of heart disease, and the database is built utilizing a framework. The study has developed SHDP (Smart Heart Disease Prediction), which employs the NB (Naive Bayesian) classification technique and the AES (Advanced Encryption Standard) algorithm to solve the problem of heart disease prediction. It was discovered that, despite limiting the characteristics, the current approach outperforms the Naive Bayes in terms of accuracy.

Thomas,[11] This paper investigated several categorization methods for estimating each person's risk level based on factors including age, gender, blood pressure, cholesterol, and pulse rate. Data mining classification methods like Naive Bayes, KNN, Decision Tree Algorithm, Neural Network, etc. are used to categorize the patient risk level.

Bhatla,[12] The objective of the research is to examine the various data mining methods that have been developed recently for the prediction of heart disease. The observations show that 15-attribute neural networks beat all other data mining strategies. Another finding from the investigation is that using a genetic algorithm and feature subset selection, decision trees have also demonstrated high accuracy.

Wu,[13] The motive of this research is to discuss how to make use of different data mining approaches to create prediction models for heart disease survival rates. In this study, the impact of various classifiers and data processing techniques is examined along with the best classifier for diagnosing patients with cardiac disease and to get an insight of the importance of how data processing may be used to increase accuracy level.

Pavithra,[14] In this paper, data mining and artificial intelligence approaches are used for gathering important data and forecasting results. Data mining technology is used to collect and categorize necessary patient data, such as age, blood sugar level, and kind of chest pain, so that the condition may be quickly recognized.

Nayak,[15] This study outlines the essential strategies for identifying and preventing heart-related ill health utilizing a variety of classification techniques that are applied using the R analytical tool. When compared to other methods, the accuracy of Naive Bayes is superior for the early prediction of heart-related ill health. According to research, the ability to predict cardiac disease is not always accurate and depends on the platform. When utilizing the Naive Bayes classifier in R environment using its data analytics aid to predict heart disease with attribute filtering, the accuracy and area under the curve are excellent.

VI. Comparative Analysis

The comparison of articles on heart disease prediction is shown in Table 1. The dataset column includes the dataset used in the published papers. The accuracy column shows how accurate the general framework that was used was. The models that were utilized and their corresponding accuracies are listed in the model and accuracy column. The highlight column provides information about a prominent feature of the paper.

TABLE I. Comparison of Heart Disease Prediction Models

Ref. No	Dataset	Accuracy	Model and Accuracy	Highlights
[1]	Cleveland Dataset	93.3%	Random search algorithm (RSA), Random Forest. 93.33%	It drew attention to the issue of overfitting in publicly available heart defect prediction approaches and offered a novel learning methodology to improve diagnosis. Reduces the number of features in machine learning models, lowering the temporal complexity of the models.
[2]	NA	86%80%84 %84%79%	Logistic regression 86%, Random Forest 80% Naïve Bayes 84% Gradient boosting 84% SVM 79%	Data cleaning and data transforming employs techniques including reducing noisy data, replacing missing data with default values, and grouping features for measures to be taken at different phases. The diagnostic model's performance has been evaluated using methods such as classification, accuracy, sensitivity, and specificity studies.
[3]	Kaggle Dataset	99.2%	Optimized SVM 99.2%	Optimizing the parameters of the SVM to get better results
[4]	Collected from a hospital in Iran, under the supervision of National Health Ministry	97.6077%	Random tree ,Random committee (Meta),Randomizable Filtered Classifier (Meta)	The accuracy, error functions, and steps were taken to ensure that classifiers are compared. Within each iteration, high-performance algorithms were introduced.
[5]	UCI Machine Learning Repository	75%, 86%, 89%, 87%, 86%	KStar model 75.0%,J48 model 86.0%,SMO 89.0%,Bayes Net 87.0% Multilayer Perceptron 86.0%	To create a more reliable method for predicting heart disease, researchers used a set of data mining approaches. For each data mining approach in this research, two data sets (collected and University of California, Irvine standard) are autonomously utilized.
[6]	UCI Heart disease dataset	86%,82%	Gaussian Naïve Bayes 86.0%,Decision tree 82.0%	The conditional independence assumption may not hold true for certain characteristics, Naive Bayes delivers superior results for datasets that are substantially larger, So, using a varied mix of algorithms can yield higher accuracy.
[7]	Cleveland standard heart disease dataset	85%,83%,7 3%	SVM 85.182%,KNN 82.963%(k=4) ANN 73.333%	The highly evaluated classifier may be used to anticipate heart disease patients in real time, as well as identify the risk factor for heart failure, in order to provide additional treatment and prevent early-stage heart failure.
[8]	UCI repository Cleveland heart dataset.	87%,62%,6 4%,65,93% ,76%,63%	J48 87.12% KNN(WEKA) 62.4% RBF 63.69% Naive Bayes 65.01%,Decision Tree 93.4% KNN(Python) 76.3% SVM 63.15%	Weka and Python are used to process the selected features using the information gain approach for selecting the best characteristic.
[9]	Cleveland dataset UCI	88.7%	Hybrid Random Forest with a	Developed Hybrid Random Forest with a Linear Model (HRFLM), a model for predicting heart disease with high precision and minimal

Ref. No	Dataset	Accuracy	Model and Accuracy	Highlights
	repository		Linear Model (HRFLM) 88.7%	classification error.
[10]	UCI dataset	84.07%,81.11%,77.4%,89.77%	Sequential Minimal Optimization (SMO) 84.07%,Bayes Net (BN) 81.11%,Multilayer Perception (MLP) 77.4%,Navies Bayesian (NB) 89.77%,AES98.2%,Parallel_Homomorphic_Encryption_Algorithm(PHEA) 92.21%	SHDP (Smart Heart Disease Prediction) uses Navies Bayesian to forecast main consideration for cardiovascular disease with greater accuracy.
[11]	UCI Dataset	80.6%	KNN and ID3 80.6%	The risk percentage of cardiovascular disease was identified applying KNN and the ID3 algorithm, and the accuracy level was also supplied for various numbers of characteristics.
[12]	Cleveland database Statlog database from UCI Dataset	86.53%89% 85.53%	Naive Bayes 86.53%,Decision Trees 89%,ANN 85.53%	A neural network with 15 characteristics has so far demonstrated the maximum accuracy, decision tree, which uses 15 characteristics, has also done well. Using data mining techniques, this study has created a prototype IHDPS that's executed on .NET environment.
[13]	UCI Dataset	70% 84% 68% 55% 69% 68%	Decision Tree 70%,Random Forest 84%,Logistic Regression 68%,SVM 55%,MLP Classifier 69%,Naïve Bayes 68%	Decision Tree and Random Forest perform finer on low dimensional data sets than Logistic Regression and Naive Bayes when used on big dimensional datasets. Random Forest is an efficient learning algorithm, it performs more accurately than Decision Tree Classifier.
[14]	Kaggle Dataset	89%	C4.5 89%	To prevent overfitting, the algorithm automatically uses the Single Pass Pruning Process. It is suitable for partial data and can function effectively with both discrete and continuous data.
[15]	UCI Dataset	69.8%81% 89%72%68%	Decision Tree model 69.8%,Support Vector Machine 81%,Naive Bayes 89%,k-NN(ROC) 72%	Frequent item mining is used in this case for attribute filtering, and a combination of different mining classification techniques, such as Decision Tree classification, Naïve Bayes classification, Support Vector Machine model of classification, and k-NN classification model of, are used for identifying and preventing diseases.

VII. Conclusion

In this work, the various heart disease prediction algorithms are reviewed and assessed. This report describes the data mining approaches used to forecast cardiac disease. By definition, heart disease is a fatal disease. This research intends to learn more about the algorithms that have the ability to improve healthcare systems, the vast majority of which are already in use. There is a high possibility of human error due to the dependence on machine learning, artificial intelligence, and deep learning. The medical industry recognizes the value of data mining. Different measures are taken to implement pertinent disease prediction algorithms. This analysis looks at how different researchers use successful ways to conduct their research. It delivers great accuracy by examining various research articles.

References

[1] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor and R. Nour, "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection," in IEEE Access, vol. 7, pp. 180235-180243, 2019, doi: 10.1109/ACCESS.2019.2952107.

[2] Dinesh Kumar G., Santhosh Kumar D., Arumugaraj K., & Mareeswari, V. (2018). “Prediction of cardiovascular disease using machine learning algorithms”. In Proceedings of 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India (pp. 978-1).

[3] Dolatabadi, Azam Davari, Siamak Esmael Zadeh Khadem, and Babak Mohammadzadeh Asl. "Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM." Computer methods and programs in biomedicine 138 (2017): 117-126.

[4] Mirmozaffari, Mirpouya, Alireza Alinezhad, and Azadeh Gilanpour. "Data Mining Classification Algorithms for Heart Disease Prediction." Int'l Journal of Computing, Communications & Instrumentation Engg 4.1 (2017): 11-15.

[5] Sultana, Marjia, Afrin Haider, and Mohammad Shorif Uddin. "Analysis of data mining techniques for heart disease prediction." 2016 3rd international conference on electrical engineering and information communication technology (ICEEICT). IEEE, 2016.

[6] Reddy, V. Sai Krishna, et al. "Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers." Journal of Physics: Conference Series. Vol. 2161. No. 1. IOP Publishing, 2022.

[7] Rabbi, M. F., Uddin, M. P., Ali, M. A., Kibria, M. F., Afjal, M. I., Islam, M. S., & Nitu, A. M. (2018). “Performance evaluation of data mining classification techniques for heart disease prediction”. American Journal of Engineering Research, 7(2), 278-283.

[8] Patra Radhanath, and Bonomali Khuntia. "Predictive analysis of rapid spread of heart disease with data mining." 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT). IEEE, 2019.

[9] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[10] Repaka, Anjan Nikhil, Sai Deepak Ravikanti, and Ramya G. Franklin. "Design and implementing heart disease prediction using naives Bayesian." 2019 3rd International conference on trends in electronics and informatics (ICOEI). IEEE, 2019.

[11] Thomas, J., and R. Theresa Princy. "Human heart disease prediction system using data mining techniques." 2016 international conference on circuit, power and computing technologies (ICCPCT). IEEE, 2016.

[12] Bhatla, Nidhi, and Kiran Jyoti. "An analysis of heart disease prediction using different data mining techniques." International Journal of Engineering 1.8 (2012): 1-4.

[13] Wu, Ching-seh Mike, Mustafa Badshah, and Vishwa Bhagwat. "Heart disease prediction using data mining techniques." Proceedings of the 2019 2nd international conference on data science and information technology. 2019.

[14] Pavithra, M., et al. "Effective Heart Disease Prediction Systems Using Data Mining Techniques." Annals of the Romanian Society for Cell Biology (2021): 6566-6571.

[15] Nayak, Sinkon, et al. "Prediction of heart disease by mining frequent items and classification techniques." 2019 International Conference on Intelligent Computing and Control Systems (ICCS). IEEE, 2019.

[16] Prabhakaran, Dorairaj, Panniyammakal Jeemon, and Ambuj Roy. "Cardiovascular diseases in India: current epidemiology and future directions." Circulation 133.16 (2016): 1605-1620.

[17] https://www.statista.com/statistics/1123509/india-share-ofrespondentswithheartissuesbyagegroup/#:~:text=As%20per%20the%20results%20of%20a%20large%20scale,for%20three%20percent%20of%20the%20disease%20that%20year.

[18] https://www.healthcheckup.com/general/normal-blood-pressure-range/#:~:text=Normal%20BP%20Range%20For%20Adults%20By%20Age%20,%20%2082%20%205%20more%20rows%20.

[19] https://www.medicoverhospitals.in/symptoms/chest-pain.

[20] https://www.verywellhealth.com/cholesterol-levels-by-age-chart-5190176.

[21] The Normal ECG – The Student Physiologist (thephysiologist.org).A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.