Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Analyzing Machine Learning Algorithms for Predicting Heart Disease

[1]Aakriti Malik, Pranav Goswami, [2]Tanvi Kaushik, Rishabh Gupta, [3]Mr. Md. Shahid
Department of CSE, Meerut Institute of Engineering & Technology, Meerut 250001, India

Analyzing Machine Learning Algorithms for Predicting Heart Disease

***Abstract* :** In recent years, cardiovascular diseases have become a leading cause of global mortality, attributed to factors such as shifts in lifestyle, dietary patterns, and work cultures. This pervasive issue affects individuals worldwide, spanning developed, underdeveloped, and developing nations. Detecting early signs of cardiovascular diseases and providing consistent medical monitoring are crucial in mitigating the escalating number of patients and lowering mortality rates. However, challenges arise due to limited medical resources and a shortage of specialist doctors, hindering continuous patient monitoring and consultations. To address this, technological interventions are essential to facilitate remote patient monitoring and treatment. Leveraging healthcare data generated from various medical procedures and continuous monitoring can lead to the development of effective prediction models for cardiovascular diseases. Early prognosis enables timely interventions, guiding lifestyle changes in high-risk individuals and ultimately reducing complications, marking a significant milestone in the field of medicine. This paper explores commonly used machine learning algorithms for predicting heart diseases by utilizing medical data and historical information. The study delves into techniques such as KNN, Decision Tree, Gaussian Naïve Bayes, Logistic Regression, and Random Forest, offering a comparative analysis of their effectiveness. Additionally, the report discusses the advantages and disadvantages of employing these techniques in developing prediction models.

Keywords- Machine Learning, Prediction of Cardiovascular Diseases, Logistic Regression, Decision Tree, Random Forest, Gaussian Naïve Bayes, K-Nearest Neighbors (KNN), Cross-Validation

1. Introduction

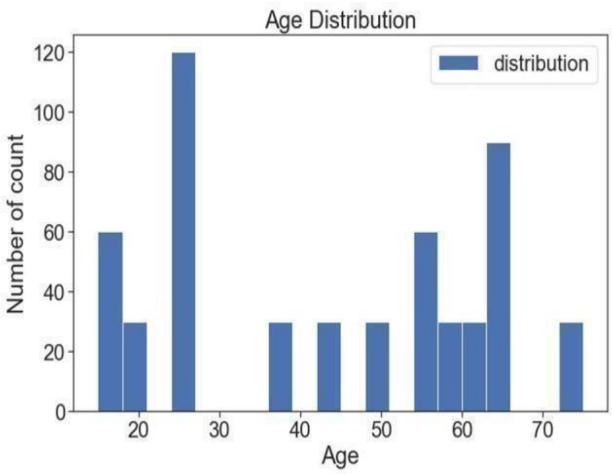
In the last ten years, coronary heart disease has grown

to be a major source of worry and difficulty. Accurately identifying symptoms and providing precise diagnoses

for illnesses, particularly in the case of heart disease, poses a significant challenge. The methods used to forecast cardiac disease during the past ten years have not been adequately effective or productive [1]. There are many clinical tools available for predicting coronary sickness, but they have two major drawbacks: they are expensive and ineffective in correctly diagnosing heart disease in humans. According to the latest WHO survey, clinical experts could successfully predict only 67% of coronary illness cases [2], highlighting the extensive scope for researching predicting heart disease in humans. There exist numerous types of heart diseases, each characterized by distinct symptoms and requiring specific treatments. Some individuals may experience a significant improvement in their well-being through lifestyle changes and medication, while others may necessitate surgical intervention [3]. The impact of heart diseases on the global population is substantial. The Global Burden of Disease (GBD) 2019, a collaborative worldwide research initiative, continuously assesses the burden of illness for every country. This ongoing efforts updated annually, aiming to facilitate consistent comparisons over time from 1990 to 2019, considering factors such as age, sex, and geographical regions. The study generates standard epidemiological estimates such as frequency, prevalence, and death rates. Additionally, it offers summary health metrics such as Disability-Adjusted Life Years (DALYs). DALYs quantify the premature loss of life and years lived with disability. These measures can be evaluated from life tables, prevalence estimates, and disability weights, and can be expressed either as counts or rates [4].

Significant advancements in software engineering have created substantial opportunities across various domains, with Clinical Science standing out as a field where these tools can be effectively utilized. Clinical research has also adopted several of the main software engineering tools that are now in use. Over thepast decade, artificial intelligence has gained prominence, reaching a significant milestone due to advancements in computational power.

Machine Learning serves as a versatile tool widely adopted across various domains due to its capacity to handle diverse datasets without requiring specialized calculations. The reprogrammable functionalities of Artificial Intelligence (AI) provide substantial strength, creating fresh opportunities, particularly in domains such as clinical science. In the field of clinical science, anticipating coronary illness presents a considerable challenge given the numerous variables and intricate details involved. Artificial Intelligence (AI) is clearly the better option when it comes to obtaining high precision in not only forecasting cardiac illness but also other diseases. This transformational program predicts cardiac illness using feature vectors and diverse types of data under different circumstances. A variety of algorithms, each with a specialty area, are used to evaluate the risk of heart disease, including Naïve Bayes, Decision Tree, KNN, and Neural Network. For instance, Neural Network reduces prediction mistakes, Decision Tree offers an organized report, and Naïve Bayes uses probability to forecast cardiac sickness. The utilization of these algorithms leverages the strengths of AI to predict the risk of heart diseases with a high level of accuracy and efficiency. All of these procedures utilize historical patient records to make predictions about new patients. This predictive system for coronary illness proves invaluable in assisting specialists to foresee the disease in its early stages, ultimately resulting in the potential to save a significant number of lives [5].



CAD, or coronary artery disease	The condition happens when atherosclerosis, or the accumulation of cholesterol and other chemicals, narrows or prevents blood from entering the heart muscle through the coronary arteries.
Cardiac Valvular Disease	The valves in the heart that control blood flow within the heart, are malfunctioning in this case. Stenosis and regurgitation of the valves are among the conditions.
Stroke	Disruption in blood delivery can lead to damage in the brain.

Fig1. Age-wise distribution.

In scientific centers, the application of data mining techniques and machine learning algorithms plays a crucial role in data analysis. These methods and algorithms are directly applied to datasets to generate models, draw meaningful conclusions, and make inferences. Factors that are frequently used in the evaluation of heart disease include age, gender, fasting blood pressure, and the nature of chest pain. The resting electrocardiogram (ECG), the number of major arteries visible on a fluoroscopy scan, the resting blood pressure (a sign of hypertension), the serum cholesterol level (a factor in determining the risk of heart disease), the maximum heart rate reached (Thalach), the ST depression (an abnormality in the ST phase of the electrocardiogram), and the Pain loc (a chest pain location where substernal=1 and otherwise=0), are all measures of heart electrical activity. Blood sugar levels during fasting, Ex(angina brought on by exercise), dietary habits, smoking, hypertension, height, and obesity[6].

Table 1 enumerates the most prevalent kind of cardiac disease.

Table 1. Many heart disease forms [7]

Heart Arrhythmia	An abnormal, too slow, or too rapid coronary heart rhythm is indicative Of an underlying issue or dysfunction.
Cardiac arrest	Breathing, awareness, and cardiac function all abruptly stop.
Congestion heart failure	Chronic heart failure arises when the heart is not pumping blood as efficiently as it should.
Congenital coronary heart disease	A pre-existing abnormality of the heart.

1.1 STRUCTURE

There are five sections to the article. An summary of previous studies and trials on machine learning-based heart disease prediction is given in the second part..In the third part , the paper delves into specifics concerning commonly utilized machine learning algorithms, specifically those applied in the prediction of heart attacks and other cardiac conditions. Section four delineates the examination of the employed approaches and their resultant outcomes, with the conclusion being articulated in section five.

2. Related Work

The growing focus on healthcare research, coupled with advancements in machine learning, has led to numerous experiments and studies in recent years. These efforts have provided valuable knowledge about how modern technologies can be used in healthcare. Marjia Sultana and associates propose employing WEKA software to predict coronary heart disease using various methods, including Multilayer Perceptron, Bayes Net, J48, SMO, and K star. Using k-fold cross-validation, the findings show that SMO with 89% and Bayes with 87% accuracy outperformed Kstar, Multilayer Perceptron, and J48. These algorithms haven't been able to deliver performance outcomes that are satisfactory, though. Enhancing the accuracy performance may help improve the diagnostic decision-making for heart disease. S. Musfiq Ali et al. assessed four alternative methods using the 303 occurrences in the Cleveland dataset for heart disorders. The researchers took into account 13 factors and employed 10-fold Cross Validation. The results demonstrated that Random Forest and Gaussian Naïve Bayes, with an accuracy of 91.2%, had the greatest results. Four models were tested using the Framingham, Massachusetts dataset, and the maximum accuracies obtained during training and testing were as follows: Random Forest Classifier (84%), Decision Tree Classifier (79%) , k-Neighbors Classifier (87%) , and Support Vector Classifier (83%). [10]. Abdullah and Rajalaxmi developed a data mining model that makes use of the Random Forest Classifier (CHD) in order to improve analysis and accuracy in the detection of coronary heart disease[11].

The research showcased the effectiveness of employing an ensemble approach to the Random Forest classification algorithm for predictive modeling of coronary heart disease (CHD). At the University of Southern Queensland, Lafta et al. developed an intelligent prediction approach using a novel time series prediction algorithm. [12]. Utilizing the medical information of each patient, the system offered decision support for healthcare professionals.

Improvements in the algorithm's capability could lead to improvements in accuracy and efficiency. A set of 303 records, each including 76 medical characteristics included in the prediction model, was utilized by Sonam Nikhar et al. Only 19 features were included in the prediction model, though, by the Naïve Bayes Classifier and Decision Tree model techniques. Moreover, their results showed that the Decision Tree approach was more accurate than the Naïve Bayes Classifier [13]. [14] is used a machine learning technique that included a variety of algorithms, such as Decision Tree, Naïve Bayes, Neural Network, Deep Learning, and SVM, during the Heart Disease Prediction Survey. Utilizing The CART, ID3, CYT, C%.0, and J48 were used to generate the decision tree result [14]. The dataset utilized by Devansh Shah and colleagues comprised 76 attributes and 303 occurrences. Supervised learning methods to leverage 14 characteristics, such as Random Forest, K-NN, Decision Tree, and Naïve Bayes. [15]. The results indicate that K-NN achieved the best accuracy. Using a dataset comprising 14 characteristics, Archana Singh et al. trained and tested the models, producing the following peak accuracies: K-NN (87%),78% for Linear Regression, 79% for Decision Tree,and Support 83% for Vector Machine .K-NN had the best accuracy, according to the results [16]. A study comparing several classification algorithms to improve performance in the prognosis of coronary heart disease using WEKA was carried out by Jaymin Patel et al. [17]. Salman Nasir, Mustafa Jan, and others. Utilized five classifier model approaches in their ensemble learning research to predict and diagnose the recurrence of cardiovascular disease: Support Vector Machines, Random Forests, Naïve Bayesian, Artificial Neural Networks, and Regression Analysis. Fourteen attributes were used in the dataset, which came from the VCI data repository to train and evaluate the models. The Random Forest algorithm achieved the highest accuracy at 98.17% [18]. Asif Khan, Jianping Li, et al. used a variety of classification methods in their analysis, such as SVM, Decision Trees, KNN, Artificial Neural Networks, Logistic Regression, and Naïve Bayes. The model's 92.37% accuracy is higher than previous models. [19]. Utilizing machine learning techniques including naïve Bayes, neural networks, KNN, SVM, random forests, decision trees, and language models Senthil Kumar Mohan et al. attempted to identify ample features. They combined features from random forests and linear approaches using the suggested hybrid HRFLM method. This model's accuracy was 88.4%. [20]. Chala Beyene and associates. Recommend assessing and estimating the prevalence of employing data mining technologies to study cardiac disease. Predicting heart disease is the primary objective so that an automated diagnosis may be made early and quickly.

In healthcare settings where specialists may have limited experience and expertise, the recommended approach is pivotal. This strategy employs diverse medical parameters, such as age and gender, blood sugar levels, and heart rates, among other factors, to ascertain whether an individual has heart disease. Dataset analysis is done using the WEKA program [21]. A survey was done by M. Siddapa and Kavitha B. S. to predict cardiac disease The Cleveland repository provided most of the data used in this work, which used a variety of ML classifiers to create a model for heart disease prediction. The Random Forest (RF) algorithm outperformed other models in terms of accuracy, according to the survey results [22]. Various machine learning algorithms were employed by Komal, Sarika, and associates to predict cardiovascular illness. According to the proposed model, Random Forests outperformed other classification algorithms with an accuracy of 85.71% [23]. Aditi Gavhane et al. conducted a study in which they developed a Model for predicting heart disease utilizing the Multilayer Perceptron (MLP) machine learning algorithm. The prediction results conveyed information about the user's condition, indicating a potential link to coronary artery disease (CAD). [24]

Table 2. Analysis of Performance Based on the Work

Writers	**Approaches employed**	**Precision**
Nikhar Sonam et.al[13]	Decision tree of Naïve Bayes	78%
Archana Singh et.al[16]	KNN	87%

	Decision Tree	79%
Pushkala V et.al[4]	Naïve BayesDecision tree	92%
Mustafa and associates [18]	Random Forest	99.19%
Senthil Mohan Kumar and others [20]	Random Forest, Linear Method (HRFLM)	88.4%
Thankgod Obasi et.al[28]	Random Forest Logistic Regression Naïve Bayes	92.44% 59.7% 61.96%

Supervised machine learning techniques are used in the study by S. Kathiresan to forecast cardiac issues. The algorithm for learning works in real-time on a four- stage cloud system based on many factors. It is predicted that productivity will increase tenfold, and cloud technology will be used to diagnose and predict cardiac disease [5]. In their talk, Himanshu Sharma and

M. A. Rizvi emphasized the most modern techniques and resources for predicting illness. The relatively new artificial intelligence discipline of deep learning has shown promise in a number of areas, including including the capacity to diagnose patients with remarkable accuracy[25].

In medical data analysis, one of the most crucial tasks is the prediction of cardiovascular illnesses. It has been demonstrated that machine learning (ML) can help make predictions and decisions from the massive amounts of data that the healthcare sector generates. Recent developments in a few Internets of Things (IoT) domains have also made use of machine learning (ML) techniques. A few publications offer a brief overview of the use of ML techniques in the determination of heart disease. Research studies showed that, processing unprocessed medical data for cardiac records can help spot irregularities in cardiac conditions early on and potentially save lives in the long run. To process the raw data and offer fresh and extra insights into coronary disease, artificial intelligence techniques were used in this study. In the medical field, coronary disease prognosis is difficult but important. However, if the disease is detected early and preventive actions are implemented as soon as feasible, the death rate can be considerably reduced.

An evaluation of the effectiveness of several machine learning algorithms and how accurate they are at predicting heart disease are shown in Table 2.

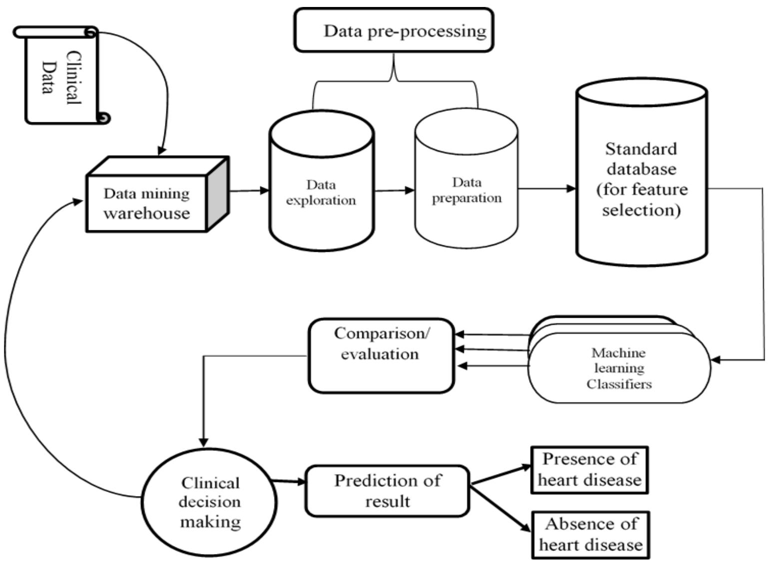


Fig 2. The procedure of employing data mining techniques to forecast heart disease [18].

3. Techniques for Prediction

Figure 2 shows how data mining and machine learning approaches can be used to predict cardiac issues.. The processes in the process that are highlighted are gathering data, preprocessing, classifying, and determining whether person has a heart disease or not. In order to identify the most accurate machine learning algorithm, the study investigates a number of classification algorithms, including KNN, Decision Tree classifier, Bayes algorithm, Random Forest , and Logistic Regression.

3.1 Logistic Regression

In the context of supervised classification, one sort of technique used for outcome prediction and probability analysis is called logistic regression. It functions by calculating probabilities according to the correlation between independent variables (like risk factors) and dependent variables (like Ten year CHD). This relationship is modeled and predicted using the logistic equation, which uses a sigmoid function:

$$q=1/(1+e^{-x}) \quad (1)$$

Logistic regression coefficient such as (c0, c1, c2….cn) are determined using an algorithm for each instance's independent variables (x1, x2, x3….. xn) during this training phase. These coefficient values are then estimated and adjusted using stochastic gradient descent.

$$x= c0x0+c1x1+c2x2+\ldots\ldots.cnxn \quad (2)$$

$$\text{where } c=c+l *(y-q)*(1-q)*q*x \quad (3)$$

Every instruction instance starts with a zero coefficient and predicts the output value (y). The learning rate (l) is applied to update the coefficients during training. For the biased input (c0), x is always set to 1. Logistic Regression depends heavily on how data is presented. To enhance the model's effectiveness, significant features from the dataset are chosen using methods like backward elimination and recursive elimination [26]

3.2 Method of Reverse Elimination

In the process of designing a machine learning model, it is essential to select features that significantly impact the target variable. The determination of a significance level, or P-value, is the first step in the feature selection procedure's backward elimination phase. A P-value of 0.05, or a significance level of 5%, was selected for this model. The largest P-value is identified using RFCEV and if its P-value is more than the chosen significance level, it is removed from the dataset. After that, the model is fitted to a new dataset, and this step will continue until all of the dataset remaining features exhibit significance levels that are below the predefined cutoff. The important variables selected following the backward elimination, for this model procedure was applied are: male, age, cigarettes per day, prevalence of stroke, diabetes, and 4osit [27].

3.3 Cross-validation in Recursive Feature Elimination

To find the most useful subset of features, the greedy optimization approach RFECV (Recursive Feature Elimination with Cross-Validation) is used. Fitting a model repeatedly and gradually removing the least significant feature at each iteration until the required number of features is reached is known as recursive feature elimination, or RFE Several feature subsets are scored during the Recursive Feature Elimination (RFE) procedure, and the set of features with the highest score is chosen. When employing the Recursive Feature Elimination with Cross-Validation (RFECV) variant, it identifies the optimal number of features for achieving the highest performance. The main drawback of this algorithm is its potential implementation cost. Consequently, before using this approach, it is best to preprocess and minimize the number of characteristics and can handle it, a matrix with correlated features is generated, and variables with high correlation are omitted. The parameters for the RFECV instance include:

- estimator : The type of model being used, which is RandomForestClassifier.
- step : How many features are taken out in each round of the process.

- cv : cv using Stratified Fold.
- scoring : The measure used for evaluation, in this case, it's accuracy.

The least significant features can be extracted and removed from the dataset after RFECV has ran and completed its execution. The following list shows the top 10 features in our model, sorted from least to most important by the RFECV technique [26].

- Blood pressure medications (BPMeds)
- current Smoker
- Prevalent hypertension
- male
- Number of cigarettes per day
- heart rate
- glucose

3.4 Decision Tree

This algorithm use tree like structure which helps to support in decisions for problems involving regression and classification in the field of supervised learning. Using basic decision rules learned from data features, this non-parametric approach builds a predictive model to forecast the value of a target variable. The most important dataset attribute is5ositionned at the tree's base during the decision tree model's creation. Multiple subsets in the training dataset are then formed based on this attribute. This process iterates until all leaf nodes are identified, each representing the desired outcomes. Decision rules and conditions are incorporated into the decision tree along its journey, forming a "if-else" structure that directs the algorithm's conclusions based on particular situations. The ultimate outcome is determined by the leaf nodes. Notably, there are minimal hyperparameters that require configuration. It's important to note that as the decision tree becomes more complex to minimize training set errors, there might be an increase in errors when applied to the test set.

The challenge of overfitting is common when constructing a decision tree model. Decision trees are beneficial for visualizing logical processes, illustrating all potential decision outcomes. Prioritizing the attribute with the highest accuracy is the main objective of the decision tree method. It uses data taken from the dataset to create a model that predicts the variable "num" [4].

3.5 Random Forests

Using a collection of decision tree classifiers, Random Forests creates a forest in an unpredictable manner. The random forest approach relies on Bootstrap Aggregation, also known as Bagging, to improve prediction accuracy by combining the predictions of several decision trees into a forest [4]. Similar to classification and regression trees (CART), Random Forest uses a decision tree classifier to fit a subsample of the dataset. In this supervised learning procedure, the dataset is divided into numerous random subsamples. A CART model is then trained on each subsample, and based on test results, each model average prediction is calculated. These trees show a high variation and minimal bias. This strategy reduces the chance of overfitting, even if preparation time may be greater. Unlike CART,which can be self-centered, Random Forest modifies the algorithm to yield sub-trees with lower correlation, addressing this issue. The Out-Of-Bag samples, leftover from the bootstrap samples, are used to estimate accuracy based on every model prediction on its remainingsamples. Feature significance, or selectively choosing important features for prediction, can be implemented in Random Forest, enabling the exclusion of less important features from the dataset [4].

3.6 Gaussian Naïves Bayes

By assuming a Gaussian distribution, the Naive Bayes classifier can be extended to real-valued characteristics. It is created using the Bayes Theorem. Gaussian Naive Bayes is one variation that can be used for binary and multiclass classification tasks. Often referred to as Idiot Bayes, this algorithm is favored for its tractable estimation of hypothesis probabilities. Optimal coefficient fitting is unnecessary, contributing to the ease of training data. The primary assumption made during prediction using this classifier is the independence of attributes in the dataset. Calculating various statistical parameters for each input variables. Hence this process includes computing the probabilities for each class by using the formula [4].

$.P(X/Y) = [P(Y/X) \times P(X)]/P(Y)$

where P(X/Y) is posterior probability and P(X) represents prior probability; P(Y) denotes predictor prior probability; and P(Y/X) represents the likelihood of the predictor. The complex and non-linear Naive Bayes technique is used to categorize data. Its reliance on assumptions and the presumption of class conditional independence may result in a decrease in accuracy even if it is easy to use, efficient, and simple to apply [15].

3.7 KNN(K- Nearest Neighbor)

The K-nearest neighbor approach is a well-liked case-based learning method that is frequently applied in practical settings. It may be used to resolve problems with both regression and classification. This algorithm is also called lazy learning which trains the model, tests it, and preprocesses the dataset. An important part in the algorithmic process is cleaning and eliminating outlier and erroneous values from the dataset, which is commonly done during the preprocessing phase. Prior to running algorithmic tests, it's critical to verify dataset accuracy, paying close attention to outliers and missing numbers. The most recent test results are graphed using a curve by the K-nearest neighbor approach, where "K" represents the number of neighbors taken into account during the classification process—typically a single digit number.

Table 3. Machine Learning Technique Comparisons [27]

Algorithm	Main Issue	Collinearity	Execution	Conclusion	Scaling	Outliers
KNN	multiclass or binary	yes	simple to execute	good	yes	yes
Logistic regression	binary	yes	simple to execute	good	no	yes
Naïve-Bayes	multiclass or binary	yes	Average to execute	good	NA	yes
Decision-Tree	multiclass or binary	no	hard to execute	good	no	yes
Random forest	multiclass or binary	no	hard to execute	good	yes	yes

4. In Summary

In the healthcare sector, solutions based on machine learning play a vital role by providing tools for the analysis of patient data, forecasting diseases, and recommending potential treatments. In critical domains like healthcare, identifying the most effective and precise machine learning method is crucial. This comparison study focuses on different machine learning methods employed to predict heart disease. The paper examines five widely used techniques: CNN, Random Forest, Naive Bayes, Decision Trees, and Random Forest. The performance of each technique is then examined and discussed to calculate best classifier for the prediction of heart disease. The analysis includes a performance evaluation of multiple machine learning algorithms and an assessment of their accuracy based on a review of previous studies and research as shown in Table

2.According to the findings, machine learning-based techniques have shown a great deal of potential for transforming the healthcare industry by improving disease prediction and therapy prescription.

5. References

[1] M. A. Jabbar, P. Chandra, and B. L. Deekshatulu, "Prediction of risk score for heart disease using associative classification and hybrid feature subset selection," Int. Conf. Intell. Syst. Des. Appl. ISDA,pp. **628–634**, (2012).

[2] V. Kirubha and S. M. Priya, "Survey on Data Mining Algorithms in Disease Prediction," vol. **38**, **no. 3, pp. 124–128**, (2016).

[3] https://www.webmd.com/heart-disease/heart- disease-types-causes-symptoms (23 October 2020)

[4] PushkalaV, A. T., & Angayarkanni, S. A. (2019). is plotted, the distance formula is used to calculate the K closest neighbors. [4].

[5] Comparative Study of Heart Disease Prediction Using Machine Learning Algorithms. International Journal of Innovations in Engineering and Technology (IJIET) http://dx. doi. org/10.21172/ijiet. 124, 10**,2319-1058.**

[6] h t t p : / / i c t a c t j o u r n a l s . i n / p a p e r / IJDSML_Vol_2_Issue_2_Paper_7_153_156.p df (25 October 2020)

[7] T.Mythili, Dev Mukherji, Nikita Padaila and Abhiram Naidu, "A Heart Disease Prediction Model using SVM- Decision Trees- Logistic Regression (SDL)", International Journal of Computer Applications, **vol. 68**, (16 April 2013).https://www.medicalnewstoday.com/articles/257484.php.(28 October 2020)

[8] A. H. M. S. U. Marjia Sultana, "Analysis of Data Mining Techniques for Heart Disease Prediction," (2018).

[9] M. I. K. A. I. S. Musfiq Ali, "Heart Disease using Machine Learning Algorithms.

[10] Abdullah AS, Rajalaxmi RR. A data mining model for predicting the coronary heart disease using random forest classifier. International C o n f e r e n c e o n R e c e n t T r e n d s i n Computational Methods, Communication and Controls (ICON3C 2012) Proceedings published in International Journal of Computer Applications® (IJCA); (2012).

[11] Lafta R, YanLi, Tseng VS. An Intelligent Recommender System based on Short Term Risk Prediction for Heart Disease patients. IEEE/WIC/ ACM International Conference on Web Intelligence and Intelligent Agent Technology. Singapore: IEEE; (2015).

[12] SonamNikhar, A.M. Karandikar, "Prediction of Heart Disease Using Machine Learning Algorithms", International Journal of Advanced Engineering, Mana