Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction

Ruby Hasan[1,*]

[1]Department of Computer Science and Engineering, Jamia Hamdard, New Delhi, India

Abstract. In the last few years, cardiovascular diseases have emerged as one of the most common causes of deaths worldwide. The lifestyle changes, eating habits, working cultures etc, has significantly contributed to this alarming issue across the globe including the developed, underdeveloped and developing nations. Early detection of the initial signs of cardiovascular diseases and the continuous medical supervision can help in reducing rising number of patients and eventually the mortality rate. However with limited medical facilities and specialist doctors, it is difficult to continuously monitor the patients and provide consultations. Technological interventions are required to facilitate the patient monitoring and treatment. The healthcare data generated through various medical procedures and continuous patient monitoring can be utilized to develop efficient prediction models for cardiovascular diseases. The early prognosis of cardiovascular illnesses can aid in making decisions on life-style changes in high hazard sufferers and in turn lessen the complications, which may be an outstanding milestone inside the field of medicine. This paper studies some of the most widely used machine learning algorithms for heart disease prediction by using the medical data and historical information. The various techniques are discussed and a comparative analysis of the same is presented. This report compares five common strategies for predicting the chance of heart attack that have been published in the literature. KNN, Decision Tree, Gaussian Naive Bayes, Logistic Regression, and Random Forest are some of the approaches used. Further, the paper also highlights the advantages and disadvantages of using the various techniques for developing the prediction models.

Keywords- *Machine Learning, Heart disease prediction, Logistic regression, Decision Tree, Random Forest, Gaussian Naïve Bayes, KNN, Cross-Validation.*

1. INTRODUCTION

Coronary illness has caused serious concerns over the last decade. One of the most significant difficulties in coronary illness is recognizing the symptoms and the correct diagnosis of the illness. The early procedures were not enough productive and efficient in predicting the heart disease [1]. There are different clinical instruments accessible on the lookout for anticipating coronary illness. Such instruments have two significant issues in them, the first is that they are a lot costly and second one is that they are not efficient to correctly compute the coronary illness in human. As indicated by most recent review directed by WHO, the clinical experts are ready to effectively anticipated just 67% of coronary illness [2], so there is a tremendous extent of exploration in the space of predicting coronary illness in humans. There are numerous kinds of coronary illness, and everyone has its own side effects and treatment. For a few, way of life changes and medication can have a gigantic effect in improving individual wellbeing. For others, they may require a medical procedure to make it work [3]. Due to the heart diseases there is a lot of impact on population worldwide. GBD (Global Burden of Disease) 2019 is a worldwide cooperative examination study that gauges illness trouble for each country on the planet. The investigation is a continuous exertion, refreshed every year, and is intended to consider predictable correlation after some time from 1990 to 2019, by age and sex, and across areas. The examination produces standard ep id em io lo g i cal e s t im a tes l ik e f r equ en cy, pervasiveness, and demise rates just as synopsis proportions of wellbeing, like DALYs. DALYs address the amount of long periods of everyday routine lost rashly and years experienced with incapacity; can be assessed from life tables, evaluations of commonness,

and handicap loads; and might be communicated as checks or rates [4]. Headway in software engineering has brought tremendous openings in various territories. Clinical Science is one of the fields where the instrument of software engineering can be utilized. Clinical science additionally utilized a portion of the major accessible devices in software engineering. In a decade computerized reasoning has acquired its second in view of progression in calculation power.

Machine Learning is one such instrument which is broadly used in various spaces since it doesn't need extraordinary calculation for various dataset. Reprogrammable limits of AI bring a ton of solidarity and opens new entry ways of chances for region like clinical science. In clinical science coronary illness is one of the significant difficulties; in light of the fact that a great deal of boundaries and detail is include for precisely predicating this infection. AI could be a superior decision for accomplishing high exactness for predicating coronary illness as well as another infections since this change instrument uses include vector and its different information types under different condition for predicating the coronary illness, calculations like Innocent Bayes, Choice Tree, KNN, Neural Organization, are utilized to predicate hazard of heart sicknesses every calculation has its claim to fame, for example, Innocent Bayes utilized likelihood for predicating coronary illness, though choice tree is utilized to give arranged report to the coronary illness, while the Neural Organization gives freedoms to limit the blunder in predication of coronary illness. Every one of these procedures are utilizing old patient record for getting predication about new tolerant. This predication framework for coronary illness assists specialists with anticipating coronary illness in the beginning phase of infection bringing about saving large number of life [5].
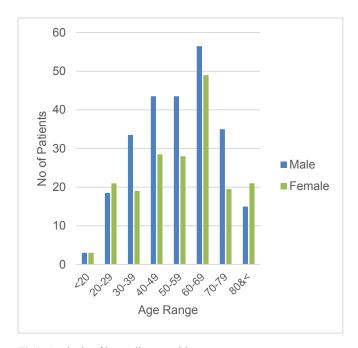
Fig1. Analysis of heart disease with respect to age group (2009),(https://www.researchgate.net/figure/Bar-chart-of-age- years-sex-distribution-of-patients-with-heart-failure_fig1_26831468.)

In scientific centers, records mining strategies and system getting to know algorithms are crucial within the evaluation of records. The strategies and algorithms can be carried out directly to a dataset to create models or draw crucial conclusions and inferences. Commonplace attributes used for coronary heart disease are Age, sex, Fasting Blood stress, Chest pain type, Resting ECG(test that measures the electrical hobby of the heart), number of foremost vessels colored with the aid of fluoroscopy, Thrust Blood strain (excessive blood pressure), Serum cholesterol (decide the chance for growing heart ailment), Thalach (most heart fee executed), ST despair (finding on an electrocardiogram, hint in the ST phase is abnormally low under the baseline), painloc (chest ache vicinity (substernal=1, otherwise=0)), Fasting blood sugar, Exang (workout included angina), smoke, high blood pressure, meals conduct, weight, peak and obesity[6]. Table 1 summarizes the most common types of the coronary heart disorder.

Table 1. Different types of heart disease [7]

Arrhythmia	If the coronary heart rhythm is abnormal, too sluggish, or too rapid, it is a sign that something is incorrect.
Cardiac arrest	Heart function, consciousness, and breathing all stop working unexpectedly.
Congestive heart failure	Chronic heart failure takes place whilst the coronary heart does not pump blood as successfully as it can.
Congenital coronary heart disease	An abnormality of the heart that appears before start.

Coronary artery disease	The main blood vessels of the heart can be damaged, and any disease that affects the blood vessels can be fatal.
Excessive Blood stress	It has a disorder in which the blood stress in opposition to the artery walls is excessive.
Peripheral artery disease	The circulatory disorder is characterized by narrowed blood vessels that restrict blood flow in the limbs.
Stroke	Interruption of blood deliver occur damage to the mind.

1.1 PAPER ORGANIZATION

The paper is divided into five sections. Section two presents an overview of the past researches and studies conducted on the on machine learning based heart disease prediction. Section three provides information about the widely used machine learning algorithms used for prediction especially for heart attacks and other heart diseases. The analysis of the approaches used and their results is presented in section four. Section five presents the conclusion.

2. RELATED WORK

With increasing research in the field of healthcare along with advanced machine learning, various experiments and researches were carried out in the last few years that provides significant information about the potential of modern day technologies in the healthcare sector. Marjia Sultana et.al, endorse coronary heart ailment prediction the usage of KStar, J48, SMO and Bayes internet and Multilayer perceptron using WEKA software [8]. Relying on the performance of various thing SMO (89% of accuracy) and Bayes internet (87% of accuracy) generate most beneficial performance than KStar, Multilayer perceptron and J48 techniques the use of k- fold go validation. These algorithms has not been able to generate satisfactory results of performance. If one can improve the performance of accuracy then it can help in better decision making to diagnosis heart disease. S. Musfiq Ali et.al, research has been conducted using Cleveland dataset for heart diseases which involves 303 instances and used 10-fold Cross Validation, mentioning 13 attributes, putting 4 different algorithms, they resulted that Gaussian Naïve Bayes and Random Forest has given the maximum accuracy of 91.2% [9]. The usage of the same dataset of Framingham, Massachusetts, the experiments had been executed the usage of 4 fashions and had been skilled and tested with most accuracy ok- neighbors Classifier: 87%, support Vector Classifier: 83%, selection Tree Classifier: 79% and Random wooded area Classifier: 84% [10]. To improve accuracy and analysis to coronary heart disease, Abdullah and Rajalaxmi introduced a DM model using the RF Classifier (CHD) [11]. Within the studies, more than one CHD incidents have been tested, which includes angina and acute myocardial infraction (AMI) and skip graft surgical treatment. The study showed that in predictive mode for CHD, an ensemble approach to the

RF classification algorithm can be very effective .Lafta et.al, based on the groundbreaking time series prediction algorithm, developed an intelligent prediction method at the University of Southern Queensland [12]. The scheme offered decision support for medical practitioners based on the medical details of each patient. By improving the capacity of the algorithm, the method has the potential to improve accuracy and efficiency. A total of 303 records with 76 medical attributes were used for the prediction model by Sonam Nikhar et al. for the prediction model, only 19 attributes were used for Naive Bayes Classifier, Decision Tree Classifier algorithm. They also studied that, compared to Naïve Bayes Classifier, Decision Tree has better accuracy [13]. For the Heart Disease Prediction Survey, Ravindra Yadav et.al, used a Machine Learning methodology with Decision Tree algorithm, Naïve Bayes, Neural Network, Deep Learning and SVM. CART, ID3, CYT, C%.0 and J48 are used to generatethe outcome of the decision tree [14]. Devansh Shah et.al, used the dataset which includes 303 instances and 76 attributes, of these14 attributes were used for supervised learningalgorithm which were Decision Tree, Naïve Bayes, K- NN and Random forest [15]. The results portray that the highest accuracy has achieved by K-NN.Archana Singh et.al, used dataset of 14 attributes for the models to trained and tested to get the maximum accuracy Linear Regression: 78%, decision Tree: 79%, Support Vector Machine: 83% and K-NN : 87%. The results showed that the maximum accuracy has with K- NN [16]. PushkalaV et.al, possess dataset of 75 attributes and 303 instances, out of which only 14 attributes are used for Decision Tree, Random Forest, SVM, K- Nearest Neighbors Classifier, Gaussian Naïve Bayes. The Naïve Bayes gives the highest accuracy of 91% [4]. Jaymin Patel et.al, this research paper compares extraordinary algorithms of class in search of better performance in coronary heart disease prognosis using WEKA [17]. Mustafa Jan, Salman Nasir et.al, in this observe, ensemble mastering combines five classifiers version approaches which includes assist vector system, synthetic neural network, naïve Bayesian, regression analysis and random forests to expect and diagnose the recurrence of cardiovascular disease. The dataset was taken from VCI data repository, 14 attributes were taken to train and test the models. The highest accuracy is 98.17% for the random forest algorithm [18]. Jianping Li,Asif Khan et.al, In this paper the classification algorithms were based on SVM, Logistic Regression, Artificial neural networks, KNN, Naïve Bayes and Decision tree. The model achieved accuracy of 92.37%, as compared to previous models [19]. Senthil Kumar Mohan et.al the primary aim of this paper was to find sufficient features by applying machine learning techniques like decision trees, language model, SVM, random forests, naïve Bayes, neural networks, KNN. The proposed hybrid HRFLM method became used for combining the characteristics of random forests and linear technique. This model achieved the accuracy of 88.4% [20]. Prediction and evaluation of the prevalence of heart sickness the use of data Mining strategies turned into counseled by Chala Beyene et al. the key intention is to are expecting the occurrence of heart sickness in an effort to make an early computerized analysis of the sickness with a short end result. In healthcare establishments with specialists

who lack enjoy and ability, the cautioned method is also crucial. It uses a ramification of medical traits, which includes blood sugar and heart price, as well as age and sex, to determine whether or not someone has heart disease. WEKA software is used to compute dataset analyses [21]. Kavitha B S, M.Siddapa carried out a survey for the prediction of heart disease.In this study it was found that most of the data was taken from Cleveland repository various machine learning classifiers are used to build heart disease prediction model. Based on the survey, the RF algorithms showed highest accuracy as compared to other models [22]. N. Komal, G. Sarika et.al .In this the various machine learning techniques were used for the prediction of cardio vascular disease.The proposed model showed that random forests having highest accuracy achieved of 85.71% as compared to other classifier techniques [23]. Aditi Gavhane et.al In this paper the model was built for the prediction of heart disease using a machine learning algorithm MLP (Multilayer Perceptron) which provide prediction results that gives the state of a user leading CAD to its users [24].

Table 2. Performance Analysis Chart based on the Reviewed work

Authors	Techniques used	Accuracy
Sonam Nikhar et.al[13]	Naïve Bayes Decision tree	78%
Archana Singh et.al[16]	KNN Decision Tree	87% 79%
Pushkala V et.al[4]	Naïve Bayes	91%
Mustafa et.al[18]	Random Forest	98.17%
Senthil Kumar Mohan et.al[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%

S.Kathiresan in this article, supervised algorithms of machine learning are used to forecast heart disease. The learning algorithm is implemented on a variety of parameters in a real-time four-stage cloud system. The productivity is estimated to be over ten times higher, and cloud technology for forecasting and diagnosing heart disease has been integrated [5]. Himanshu Sharma, M A Rizvi .This paper summarizes current state-of-the-art strategies and available tools for disease prediction. Deep learning, a new branch of artificial intelligence, has shown promise in other fields, such as high-accuracy medical diagnosis [25].

Coronary illness is perhaps the main sources of mortality on the planet today. Expectation of cardiovascular contamination is a primary check inside

the area of scientific records research. AI (ML) has been established to be powerful in supporting with settling on alternatives and expectancies from the great quantity of data created by the hospital therapy enterprise. We've additionally seen ML tactics being utilized in late enhancements in numerous spaces of the web of things (IoT). Unique examinations provide only a short investigate foreseeing heart infection with ML techniques. From the above noted paper recognizing the preparing of crude hospital therapy records of coronary heart records will help in the drawn out saving of human lives and early region of irregularities in coronary heart situations. AI techniques were applied in this work to deal with crude facts and deliver every other and novel insight in the direction of coronary infection. Coronary infection expectation is trying out also, crucial in the medical discipline. Anyways, the mortality fee may be appreciably managed if the infection is distinguished at the start phases and guard measures are embraced as earlier than long as might be anticipated. Table 2 presents the performance analysis of various machine learning algorithms along with their accuracies used for prediction of heart diseases.

3. METHODS FOR PREDICTION

Figure 2 shows the process of predicting heart diseases using data mining and machine learning techniques. The process of data collection, preprocessing, classification and prediction for predicting of results, whether the person has presence or absence of heart disease is highlighted. Numerous classification techniques, such as KNN, Decision Tree, Gaussian naïve Bayes, Random Forest, and Logistic Regression, are studied in this paper to determine the best suited machine learning algorithm for accuracy, as well as many feature selection strategies, such as backward elimination and recursive feature elimination are also reviewed.

3.1 Logistic Regression

Logistic Regression is a supervised classification algorithm. It's a probabilistic analysis algorithm that predicts outcomes. By estimating probabilities using the underlying logistic equation, it assists in measuring the relationship between the dependent variable (TenyearCHD) and one or more independent variables (risk factors) (sigmoid function).

The following logistic function is used in the logistic regression algorithm:

$$p = 1 \backslash (1 + e^{\wedge} - x) \qquad (1)$$

The logistic coefficients for each instance $x1, x2, x3.....$ xn will be $b0, b1, b2..... bn$ during the training stage. Stochastic gradient descent is used to estimate and update the coefficient values.

$$\text{Values} = b0x0 + b1x2 + ...bnxn \qquad (2)$$

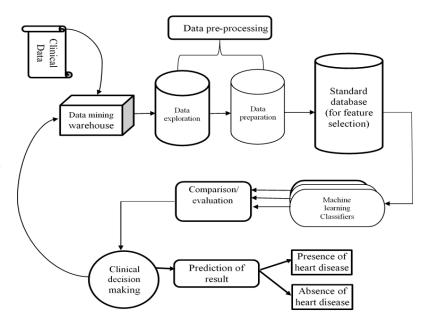$$b = b + l*(y-p)*(1-p)*p*x \qquad (3)$$

Fig 2. Data mining methodology and heart disease prediction process [18]

Each training case's output value is represented by y, and all of the coefficients are initially set to 0. For b0's biassed input, l is the learning rate, and x is always 1. It changes the coefficient values at the training level until it predicts the correct performance.

Logistic Regression relies highly on the proper presentation of data. So, to make the model more powerful, important features from the available data set are selected using backward elimination and recursive elimination techniques [26].

3.2 Backward Elimination Method

Only the features that have a major impact on the goal variable should be chosen when designing a machine learning model. The first step in the backward elimination process for feature selection is to choose a significance level or P-value. We chose a 5% significance rating, or a P-value of 0.05, for our model. The function with the highest P-value is marked, and it is excluded from the dataset if its P-value is greater than the significance level. The model is fit to a new dataset, and the process is repeated until all remaining features in the dataset have a significance level less than the significance level. Male, age, cigsPerDay, prevalent Stroke, diabetes, and sysBP were chosen as significant factors in this model after using the backward elimination algorithm [26].

3.3 Recursive Feature Elimination Using Cross Validation

The RFECV algorithm is a greedy optimization algorithm that seeks out the best performing function subset. Recursive Feature Elimination (RFE) suits a model several times, removing the weakest feature each time until the desired number of features is reached. RFE scores various feature subsets and selects the highest scoring range of features, which is RFECV, using the optimal number of features. The key disadvantage of this algorithm is that it can be costly to implement. As a result, it is preferable to reduce the number of features in advance. Correlated features can be removed before

RFECV because they provide the same details. The correlation matrix is plotted to solve this, and the correlated features are omitted. The arguments for instance of RFECV are:

- estimator - model instance (RandomForestClassifier)
- step - number of features removed on each iteration
- cv – Cross-Validation (Stratified Fold)
- scoring – scoring metric (accuracy)

Once RFECV is run and execution is finished, the features that are least important can be extracted and dropped from the dataset. Top 10 features ranked by the RFECV technique in our model listed below from least importance to highest importance [26].

- prevalent Stroke
- diabetes
- BPMeds
- current Smoker
- prevalentHyp
- male
- cigsPerDay
- heart rate
- glucose
- diaBP

3.4 Decision Trees

A decision tree is a tree-like structure or a flowchart-like structure that is used as a decision-support tool in both classification and regression problems to help build automated predictive models. It's a non-parametric supervised learning algorithm. It creates a model that predicts the cost of a target variable via mastering simple selection regulations inferred from facts capabilities. The decision tree model is generated by putting the best attribute of the dataset at the root of the tree. There are several subsets in the training package. Each subset contains information for the same attribute. The algorithm is repeated until all of the leaf nodes have been discovered, i.e., the desired result has been discovered. There are almost no hyper-parameters that need to be set. The decision tree is associated with decision rules and conditions along the path form and if else rule to allow the algorithm to decide on the basis of the situation, where the outcome is the contents of the leaf node. As the algorithm becomes more complex in order to reduce the error of the training set, the test set's error will increase.

This over fitting problem will occur frequently when building a decision tree model. The use of a decision tree aids in the visualization of logic. It generates all possible decision outcomes. The decision tree algorithm's main goal is to prioritize the attribute that can provide the highest level of accuracy. It creates a model to predict the variable "num" using the inferred dataset [4].

3.5 Random Forests

Random Forest produces a forest in an ad hoc manner. The ensemble of decision tree classifiers is used to build the forest. Bootstrap Aggregation or Bagging, which combines the predictions of multiple decision trees and brings them together to construct a forest to make more precise predictions, is the

foundation of the random forest algorithm. To fit a sub-sample of the dataset, it employs a decision tree classifier. It's a supervised learning process, and the algorithm is similar to classification and regression trees (CART). They respond to the data gathered during the training. The CART algorithm bagging generates a number of random sub-samples from the dataset, each of which is used to train the CART model, and the average prediction from each model is then determined using the test results. Low bias and high variance characterize these trees. The planning for this process will take a long time, but the problem of overfitting will not arise. CART's problem is that they are self-centered. Random Forest redesigned the CART algorithm in such a way that the sub-trees are known. The resulting prediction from these sub-trees has a lower correlation. Out-Of-Bag samples are the samples left over from the bootstrap samples taken from the training results. The estimated accuracy is determined by the output of each model on its left out samples. It is possible to selectively select the features needed for the prediction process from the dataset, a process known as feature significance. The random forest can be built without the less important feature that contributes less to the prediction process [4].

3.6 Gaussian Naïve Bayes

The Naive Bayes classifier is built using the Bayes Theorem. By assuming a Gaussian distribution, it can be extended to real-valued attributes. A version of Naive Bayes is Gaussian Naive Bayes. It's a supervised classification learning algorithm for binary and multi-class classification problems. This algorithm is also known as Idiot Bayes because the probabilities of hypothesis estimation are tractable. There is no need for optimum coefficient fitting because data training is easy. When using this classifier to make a prediction, the main assumption is that the attributes of the data set are independent. The probabilities in each class, the mean and standard deviations for each input variable in each class are determined [4].

$$P(X/Y) = [P(Y/X)\times P(X)]/P(Y) \qquad (4)$$

P(X/Y) is the posterior probability, P(X) is the class prior probability, P(Y) is the predictor prior probability, P(Y/X) is the likelihood, probability of predictor.

Nave Bayes is a non-linear, difficult data categorization algorithm that is simple, straightforward to implement, and efficient. However, because it is reliant on assumption and class conditional independence, there is a loss of accuracy [15].

3.7 KNN (K- Nearest Neighbor)

The K nearest neighbor algorithm is an example-based learning algorithm that is widely used in real-life scenarios. The K Nearest Neighbors algorithm can be used to solve both classification and regression problems. The K Nearest Neighbor algorithm is another name for lazy learning. The K Nearest Neighbor algorithm involves preprocessing the dataset, training the model, and testing the model. Cleaning and removing erroneous and outlier values from a dataset is usually part of the preprocessing phase. In the algorithmic process, this is the most important step. It's also crucial to check the dataset's accuracy before running algorithmic tests on it. It is important to pay attention to missing values and

outliers. The K Nearest Neighbor uses the curve to plot the latest test results. The K element is the number of neighbors that the classification takes into account. In most cases, the K value should be a single digit number. After the test data point is plotted, the distance formula is used to calculate the K closest neighbors. [4].

Table 3. Comparisons of Machine Learning Techniques [27]

Technique	Primary Problem	Predictors	Raw Implementation	Interpretability	Regression Also	Normalization
KNN	Multiclass or binary	Numeric	Easy	Good	No	Required
Logistic regression	Binary	Numeric	Easy	Good	No	No
Naïve Bayes	Multiclass or binary	Categorical	Medium	Good	No	Required
Decision Tree	Multiclass or binary	Numeric or Categorical	Difficult	Good	Yes	No
Random forest	Multiclass or binary	Numeric or Categorical	Difficult	Good	Yes	No

4. CONCLUSION

Machine learning based solutions are widely used in healthcare sector for analyzing patients' data, predicting diseases and suggesting possible treatments. With a number of machine learning techniques available today, it is important to identify the most efficient and accurate technique especially in critical domains like healthcare. A comparative analysis of the various Machine learning algorithms used in the heart disease prediction is presented. Five of the most widely used techniques, KNN, Logistic regression, Naive Bayes, Decision tree and random forest are discussed and compared to identify the best suited classifier for heart disease prediction. Many previous researches and studies related to heart disease prediction were identified and analyzed. Based on the identified researches and studies a performance analysis of various machine learning algorithms along with their accuracies used for prediction of heart diseases was done as presented in table 2. The findings shows that in most cases machine learning based approaches have shown significant potential to transform the healthcare sector and improve the entire process of disease predictions and suggesting treatments.

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). 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.**
5. http://ictactjournals.in/paper/ IJDSML_Vol_2_Issue_2_Paper_7_153_156.p df (25 October 2020)
6. 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).
7. 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. K. Bhanot, "towarddatascience.com," (13 Feb 2019 .) [Online]. Available: https:/ / towardsdatascience.com/predicting-presence- of - heart-diseases-using-machine -learning-36f00f3edb2c. (2 November 2020).
11. Abdullah AS, Rajalaxmi RR. A data mining model for predicting the coronary heart disease using random forest classifier. International Conference on Recent Trends in Computational Methods, Communication and Controls (ICON3C 2012) Proceedings published in *International Journal of Computer Applications® (IJCA); (2012).*
12. 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).*
13. SonamNikhar, A.M. Karandikar, "Prediction of Heart Disease Using Machine Learning Algorithms", International Journal of Advanced Engineering, Management and

Science (IJAEMS) [Vol-2, Issue-6, June- 2 0 1 6] *I n f o g a i n P u b l i c a t i o n* (Infogainpublication.com) ISSN: **2454-1311**, (2016).

14. https://www.ijcrt.org/papers/ IJCRT1807425.pdf (13 November 2020)

15. Devansh Shah, Samir Patel, Santosh Kumar Bharti, "Heart Disease Prediction using Machine Learning Techniques", *Springer Nature Singapore Pte Ltd (2020)*.

16. Archana Singh, Rakesh Kumar, "Heart Disease Prediction using Machine Learning Algorithms", *International Conference on E l e c t r i c a l & E l e c t r o n i c s Engineering(ICE3-2020)* : **452-457**.

17. Patel, J., TejalUpadhyay, D., & Patel, S. (2015). Heart disease prediction using machine learning and data mining technique. *Heart Disease*, *7*(1), **129-137.**

18. Jan, M., Awan, A. A., Khalid, M. S., & Nisar,S. (2018). Ensemble approach for developing a smart heart disease prediction system using classification algorithms. *Research Reports in Clinical Cardiology*, *9*, **33-45**.

19. Li, J. P., Haq, A. U., Din, S. U., Khan, J.,Khan, A., & Saboor, A. (2020). Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare. *IEEE Access*, *8*, **107562-107582**.

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

21. Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of H e a r t D i s e a s e U s i n g D a t a M i n i n g Techniques", *International Journal of Pureand Applied Mathematics*, (2018).

22. Kavitha, B. S., & Siddappa, M. A Survey on Machine Learning Techniques to Predict Heartt Disease.

23. Kumar, N. K., Sindhu, G. S., Prashanthi, D. K., & Sulthana, A. S. (2020, March). Analysis and prediction of cardio vascular disease using machine learning classifiers. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* **(pp. 15-21).** *IEEE.*

24. Gavhane, A., Kokkula, G., Pandya, I., & Devadkar, K. (2018, March). Prediction ofheart disease using machine learning. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)* **(pp. 1275-1278).** IEEE.

25. Sharma, H., & Rizvi, M. A. (2017). Prediction of heart disease using machine learning algorithms: A survey. *International Journal on Recent and Innovation Trendsin Computing and Communication*, *5***(8), 99-104**.

26. https:// www. academia. edu/ 42249626 / Mini_Project_Report_On_Heart_Disease_Pred iction_ (27 November 2020)

27. https://jamesmccaffrey.files.wordpress.com/ 2 0 1 8 / 1 0 / ml_classification_algorithms_comparison.jpg(16 December 2020)

28. O b a s i, T. , & O m a i r S h a f i q , M .(2019). Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attack and Diseases. *2019 IEEE International C o n f e r e n c e o n B i g D a t a (B i g D a t a)*.

*Coresspanding author: ruby.hasan92@gmail.com