

Early Detection of Cardiovascular Disease using Machine learning Techniques an Experimental Study



Najmu Nissa, Sanjay Jamwal, Shahid Mohammad

Abstract: Human body prioritizes the heart as the second most important organ after the brain. Any disruption in the heart ultimately leads to disruption of the entire body. Being the members of modern era, enormous changes are happening to us on a daily basis that impact our lives in one way or the other. A major disease among top five fatal diseases includes the heart disease which has been consuming lives worldwide. Therefore, the prediction of this disease is of prime importance as it will enable one to take a proper and needful approach at a proper time. Data mining and machine learning are taking out and refining of useful information from a massive amount of data. It is a basic and primary process in defining and discovering useful information and hidden patterns from databases. The flexibility and adaptability of optimization algorithms find its use in dealing with complex non-linear problems. Machine Learning techniques find its use in medical sciences in solving real health-related issues by early prediction and treatment of various diseases. In this paper, six machine learning algorithms are used and then compared accordingly based on the evaluation of performance. Among all classifiers, decision tree outperforms over the other algorithms with a testing accuracy of 97.29%.

Keywords: Heart Disease, Machine Learning Models, Python, Spyder.

I. INTRODUCTION

Heart disease is considered as one of the hazardous diseases that affect human body at random. It is a fatal disease since hundreds of people die daily on an average due to this disease and other such kinds of similar diseases [1-3]. According to a study conducted by the Indian Heart Association, more than 17 lakh people die annually due to heart disease and by 2030 the score is expected to increase with 2.3 crore deaths [4]. The variety of conditions that affect the functioning of heart allude to heart diseases. The various

types of heart disease include:

- Atherosclerotic is a type of heart disease which affects the arteries of heart.
- Valvular heart is a disease that alters the functioning of valves to regulate blood in and out of the heart.
- Cardiomyopathy affects how the muscles of a heart clutch.
- Arrhythmias disturbs the electric conduction of the heart.
- Heart infections may develop before birth and heart suffer from structural problems.
- Coronary arteries supply blood to the heart and the presence of lipids in the blood in higher quantity forms a cholesterol plaque inside the walls of the artery, which later on block the arteries and lower the flow of blood through them causing the coronary artery disease. The most common type of CVD includes heart attack that occurs when the cholesterol plaque bursts and forms a clot in the arteries causing blockage of blood.

The personal and professional behavior of a person is completely dependent on the health of a human heart. Due to rapid increase in cardiovascular disease, it is thus necessary to build a model including symptoms and diet habits that can predict the heart disease without putting so many efforts. Individuals at risk of cardiovascular disease may demonstrate the following:

- More Cholesterol Level.
- Raised Blood Pressure.
- Smoking
- Lipids in high quantity.
- Being overweight and obese.
- Family History of Heart Illness.

Cardiovascular diseases can be predicted and its diagnosis can be done by taking various attributes like age, sex, cholesterol, resting blood pressure etc., and analyzing them in a better way that allow experts to make more better and more accurate knowledge-based results. Due to the rapid increase in the amount of data, Machine Learning is an emerging field which makes it possible to gather knowledge from a vast amount of data. The role of data mining in healthcare is to solve real-life health problems and their diagnosis and treatment. The objective of this paper is, to compare different machine learning algorithms using different tools and techniques and their results in terms of accuracy.

Revised Manuscript received on August 01, 2020.

Revised Manuscript received on August 05, 2020.

Manuscript published on September 30, 2020.

* Correspondence Author

Najmu Nissa*, Department of Computer Sciences, BGSBU University Rajouri, J&K, India. Email: najmunissa@bgsbu.ac.in

Sanjay Jamwal, Sr. Assistant Professor, Department of Computer Sciences BGSBU Rajouri, J&K, India. Email: sanjayjamwal@bgsbu.ac.in

Shahid Mohammad, Department of Computer Sciences, BGSBU University Rajouri, J&K, India. Email: shahidmohammad@bgsbu.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

It also highlights the future scope of a prediction model in healthcare using Data mining and vast analysis of machine learning algorithm. According to findings of the survey by IPSOS [5], the average number of deaths in percentile due to heart diseases in comparison with deaths due to other diseases in 2019 is shown below in figure 1.

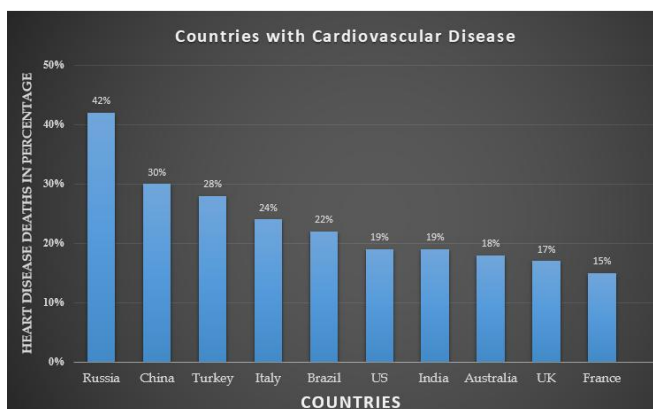


Figure 1: Top Ten Countries with Cardiovascular Disease. [5]

II. LITERATURE REVIEW

Copious work has been done for heart Disease prediction using Machine Learning, Deep Learning, Data mining tools and techniques. Different Datasets, Algorithms and methods used by the researchers and observed results alongside the future work is carried out in finding efficient methods of medical diagnosis for Cardiovascular disease. Design of prediction models for CVD diagnosis has remained an active research area for past one decade. The automatic diagnosis of CVD is a crucial real-world medical problem. Detection of heart disease in its early stages is crucial for better treatment. It can be achieved only by Machine Learning and Deep Learning, the core components of predictive analytics. Researchers' all over the world have used various approaches for early detection and prediction of heart disease. Some of the main approaches are: Jaymin Patel, et.al (2015): This study aims at creation of model for prediction of heart diseases and at achievement of marginal success. The experiment was carried in WEKA on Cleave Land UCI dataset and during training the data undergoes the following: preprocessing classification, regression, clustering, association and visualization. For testing classification tools and Explorer mode of WEKA is used. Decision tree classifiers like J48, Logistic Model Tree Algorithm and Random Forest were used for the analysis with cross-validation 10-fold and reduced error pruning were employed. Among all, j48 with reduced error pruning achieved the highest accuracy. The accuracy could be enhanced if the use of different discretization techniques, multiple classifiers, voting techniques and other decision tree Algorithms (Gain ratio, Gini index) would have been entertained. There is a lack of proper combination and more complex models so that the more accuracy in terms of sensitivity, specificity and accuracy could be achieved. The aforementioned algorithms in this study can be used in future for the development of a system to be used for checking the imbalance with other data mining models [6].

Sonam Nikahr et.al (2016): This study focuses on increasing the performance of Naïve Bayes classifier by excluding form the dataset, the unwanted and irrelevant attributes and selecting only the informative ones. A dataset with 303 records and 76 attributes were collected from cleave Land Heart Disease Database among which only 19 attributes were selected for analysis. In order to achieve this, Selective Naïve Bayes classifier is used in which c4.5 trees are constructed. In this research work, the application of Naïve Bayes and Decision Tree with Information Gain calculations provide better results as compared to other classifiers. The algorithm for the induction of the decision tree is the greedy algorithm and it has been found that the decision tree has more accuracy than Naïve Bayes according to results [7].

SyedahaminPouriyeh et.al (2017): The principal goal of this paper is to make a comparison of different machine learning techniques on a small dataset. The dataset in this research has been taken from cleave land database containing 303 instances with a total no. of 76 attributes among which a maximum of 14 attributes was used for the study. Due to lower variance in comparison with other estimators like single-fold approach, 10- fold cross-validation has been used for data portioning. The different machine learning classifiers used include: decision tree, Naïve Bayes, Multi-Layer Perceptron, K Nearest Neighbor, Single Conjunctive Rule, Radial Basis Function and Support Vector Machine. Bagging, Boosting and Stacking have been employed in combination with the aforementioned machine learning techniques. Precision, Recall, F-measure, and ROC are the main performance measures in this study. The experiment was conducted in two instances, in the first instance the whole dataset was used and the machine learning algorithms were employed using 10-Fold cross-validation. The results show that SVM achieved a higher accuracy of 89.12% among all. In the second instance, an experiment was conducted using Bagging, Boosting and stacking. Bagging improved the accuracy of DT from 77.55% to 78.54%. Boosting improved the accuracy of DT from 77.55% to 82.17%. Although, by combining SVM using stacking, the highest accuracy of 84.15% have been obtained. Youness Khourdifi et.al (2018): Proposed heart disease prediction has been made using Machine Learning Algorithms optimized by Particle Swarm Optimization and Ant-Colony Optimization. To increase the quality of heart disease classification by avoiding redundant features, a method called Fast Correlation Based Feature Selection (FCFS) has been used. The classification is done using various algorithms like KNN, SVM, NB, RF, MLP (ANN) along with combination of approaches like Particle Swarm Optimization with Ant-Colony Optimization (ACO). The dataset has been collected from the UCI machine learning repository. Binary type classification problem prepares training dataset and the new features are classified using WEKA. The dataset is operated on the proposed hybrid method. The optimized model proposed by FCBF, PSO and ACO achieved the highest accuracy of 99.65%. Hence the performance of the said hybrid model proved excellent than the other mentioned classification techniques. The author's knowledge base, the tools and the period available are the main limitations of this study [9].

Mamta Alex et.al (2019): This study aims at organization and rapid detection technique early stage and prediction of disease at right time. This paper highlights the limitations of data mining algorithms i.e., while diagnosis, the prediction is less accurate and takes more time. The mode of data collection is real time data that have been collected from “jubilee Mission College And Research Thrissur” by interacting with the patients one by one and discharge details of heart disease patients.

A total of 2200 records containing 20 attributes have been collected followed by sorting and arranging in an ordered manner. The attributes were fed into SVM, RF, KNN, ANN and among all, KNN achieved the best accuracy of 92.21% [10].

III. MACHINE LEARNING TOOLKIT

The processing of the whole experimental work has been done using open source Anaconda 2020[11]. Also, for programming tasks and measurements Python (3.7.6) along with Spyder as an integrated development environment is used. For importing databases, data preprocessing, machine learning algorithms, rich standard packages are employed also with pandas for reading and analyzing the dataset.

IV. DESCRIPTION OF THE DATASET

The dataset has been obtained from the application of UCI machine learning repository. Two datasets have been obtained one containing 1026 instances and 14 attributes and the second one containing 303 instances and 14 attributes. On combining the resulting dataset contains 1329 instances and 14 attributes. The description of the attributes is being described in Table 1.2

S. No	Attribute Name	Description
01	Age	Age in years
02	Sex	Male/female
03	Cp	Constructive pericarditis
04	Trestbps	Resting blood pressure in mmHg on admission to hospital
05	Chol	Serum cholesterol in mg/dl
06	Fbs	Fasting blood sugar (greater than 120mg/dl). values:1=true, 0=false.
07	Restecg	Resting electrocardiographic results. values:0=normal, 1=having ST-T wave abnormality.
08	Thalch	Maximum heart rate achieved.
09	Exang	Exercise including angina.value:1=yes,0=no
10	Oldpeak	St depression induced by exercise relative to rest
11	Slope	The slope of peak exercise ST segment.value:1=up sloping, 2=flat, 3=down sloping.

12	Ca	No. of major vessels (0-3) colored by fluoroscopy
13	Thal	Inherited blood disorder that causes your body to have lesser HB than normal. Values:3=normal, 6=fixed defect, 7=reversible defect.

Table 1: Description of Attributes of Heart Diseases

V. MOTIVATION FOR THE STUDY

The main motivation for this research work is to develop a heart disease prediction model for the prediction of the presence of heart diseases. Furthermore, this research work aims at identifying the best classification algorithm in terms of the highest accuracy for the prediction of the aforementioned disease. This work is justified by performing an analysis comparatively using six algorithms viz., Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, Support Vector Machine and Artificial Neural Network. At different levels of analysis and analysis strategies, these algorithms have been used and analyzed based on their performance. This study will enhance research to identify a better solution and best method for prediction of heart disease.

VI. PROPOSED METHODOLOGY

Figure 2 below depicts the proposed methodology for the whole experiment starting from the collection of data up to the generation of results. The process starts with the data collected from sources (as mentioned above) followed by the data pre-processing. Data pre-processing is done to avoid the inconsistency, noise and to remove the bias. Following the data pre-processing step, the dataset is divided into training and testing sets respectively. Furthermore, machine learning algorithms are applied to train and test the data separately. The process completes with the generation of results in terms of accuracy which are compared among the different machine learning algorithms.

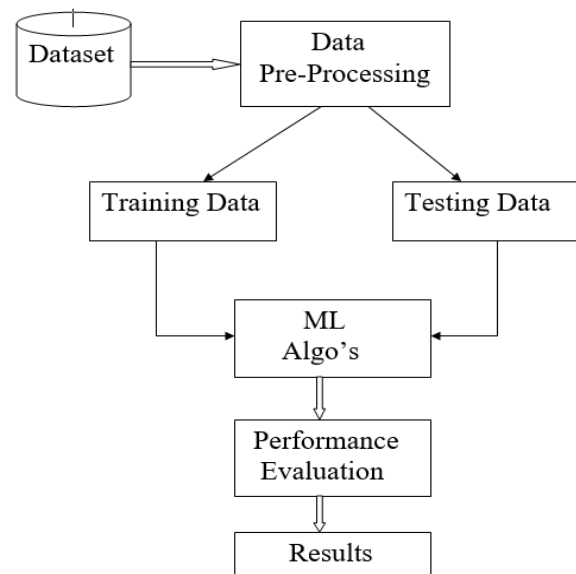


Figure 2: The Workflow of the Experimental Study



VII. MACHINE LEARNING ALGORITHM

The Machine Learning Algorithms in healthcare have a great potential beyond the human capacity to process a vast volume of heterogeneous datasets from varied sources. The conversion of data into clinical insights that aid physicians in planning and providing care ultimately leads to better outcomes, lower costs of care, and increased patient satisfaction.

The below-mentioned algorithms have been used for the experiment.

- 1) **Logistic Regression:** [12] It is a kind of statistics regression analysis, particularly used for prediction and calculation of the probability of success. The results of the prediction of the categorical dependent variable which is in binary form from a set of independent variables can be obtained. It involves fitting the following equation in the form of data:

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n \quad (1)$$

- 2) **Decision Tree:** Decision tree is a flowchart-like figure containing leaf nodes as branches [13]. Due to its property of not requiring any particular domain knowledge, it has become famous for the actual discovery of knowledge. For classification of data, decision tree uses j-48 type which could generate pruned and un-pruned trees. This algorithm acquires the property of handling both continuous and continual attributes [14]. The simple decision tree is shown below in figure 3[15].

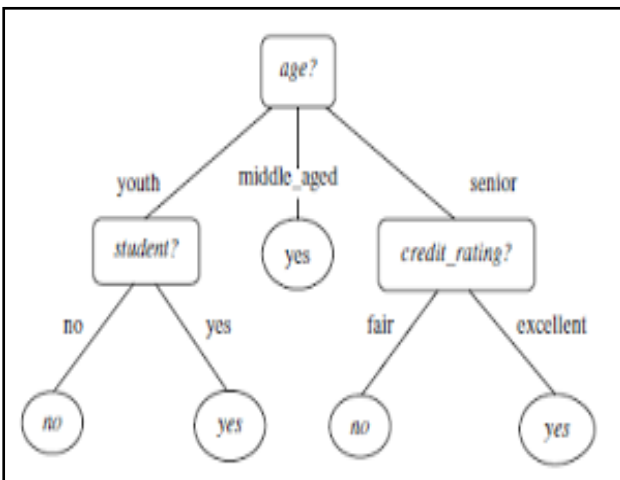


Figure 3: The Decision Tree

- 3) **Artificial Neural Networks:** It is a mathematical model based on the simple biological neuron. ANN is the network type figure consisting of nodes, that are connected through directional lines where each node represents a processing unit and directional lines represent the links between processing units. Neural networks are composed of many layers starting with the input layer, where each of them corresponds to predicate variable [16]. Nodes of the input layer are connected to nodes of a hidden layer and the nodes of hidden layer are connected to the nodes of the output layer. The simple ANN IS shown in figure 4[17].

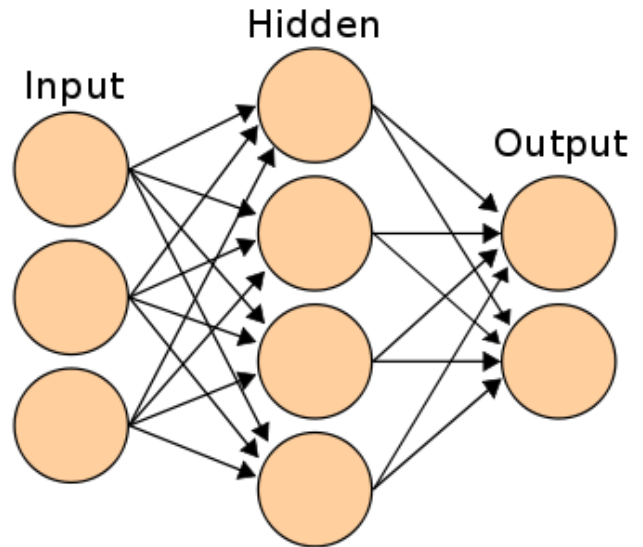


Figure 4: The Artificial Neural Network

- 4) **Naïve Bayes Classifier:** Naïve Bayes classifier is based on Bayes theorem [18] and conditional probability rule. It is easy to interpret and use all attributes in a given dataset and analyze them individually [19]. A conditional probability is the likelihood of some conclusion C, given some evidence or observation E, where dependency relation exists between C and E as follows [20]:

$$P(H|E) = \frac{P(E|H).P(H)}{P(E)}$$

- 5) **Random Forest:** It is a supervised machine learning algorithm composed of various decision trees used for prediction and due to its higher number of trees that result in an improved accuracy compared to the decision tree. The trees are trained independently and their predicted values are combined on an average. The generalization error in a forest tree classifier depends on the strength of an individual tree in the forest and the correlation between them [21]. A simple random forest is shown in figure 5 below [22].

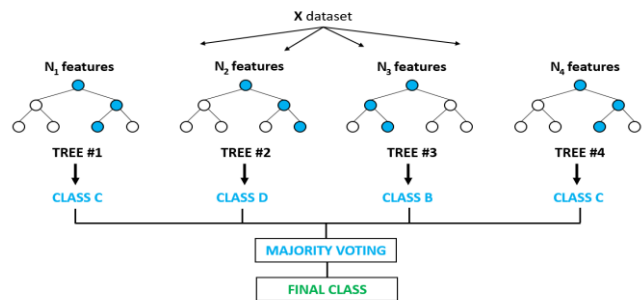


Figure 5: The Artificial Neural Network

- 6) **Support Vector Machine:** SVM is a supervised machine learning algorithm which classifies and separates the heart disease dataset by a hyper plane [23]. SVM is very hard to interrupt and tune due to is a feature of memory intensive.



The basic function of this supervised classifier is to segregate or classify the given data points in the best possible and appropriate way in multidimensional space. To work in high and complex dimensions, the SVM classifier uses different versions of the kernels like linear kernel, polynomial kernel and radial basis function kernel. The equations for various kernels are [24]:

7) Linear Kernel Equation

$$8) (X) = (0) + \text{Sum}(ai * (x,i)) \quad (3)$$

9) Polynomial Kernel Equation

$$10) K(X1, X2) = (a + X1TX2) b \quad (4)$$

11) Where b = degree of kernel & a = constant term

12) Radial Basis Function Kernel Equation

$$13) (X1, X2) = \text{ponent}(-\gamma \|X1, X2\|) 2 \quad (5)$$

14) Where $\|X1, X2\|$ = Euclidean distance between (X1 & X2)

VIII. MEASUREMENT AND COMPARATIVE ANALYSIS

In this research work, we established an experimental and analytical setup and accordingly evaluated six machine learning algorithms for effective prediction of heart disease. The resulting histogram is shown in figure 6 below indicating the plots among which each attribute of the dataset is distributed.

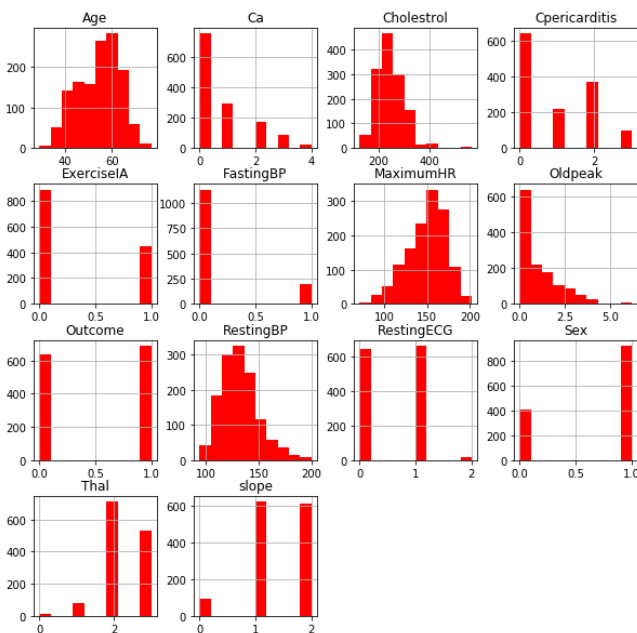


Figure 6: Histogram for each attribute in the dataset.

Figure 7 below represents the correlation among the selected parameters that have been used for prediction of heart disease in the form of Heat Map.

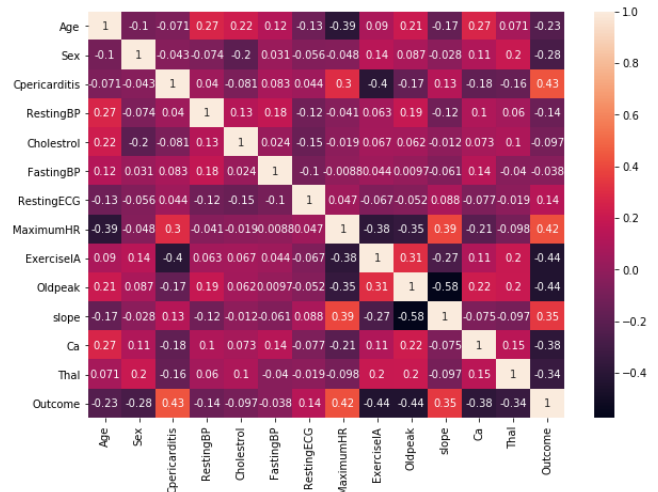


Figure 7: The Heatmap of parameters.

The next step is to find the statistical results of the six classification algorithms followed by the implementation of these Algorithms. The confusion matrix is used to describe the performance of these models. The following measures have been calculated to find the final accuracy of each model as follows:

1) **ACCURACY:** Accuracy is interrupted from given formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \quad (6)$$

In the equation TP, TN stand for True Positive and True Negative and FP, FN stand for False Positive and False Negative respectively. TP+TN signify the percentage of correctly classified instances and TP+TN+FN+FP signify a total of correctly and incorrectly classified instances.

2) **PRECISION:** Precision is a part of significant instances involving the retrieved instances. The equation for precision is given below:

$$\text{Precision} = \frac{TP}{TP + FN} \quad (7)$$

3) **RECALL:** Recall is small portion of appropriate instances, which is retrieved over the total quantity of relevant instances. The equation for the Recall is given below:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

4) **SPECIFICITY:** The following equation which satisfies the definition of specificity is given below:

$$\text{Specificity} = \frac{N}{TN + FP} \quad (9)$$

5) **F-MEASURE:** F-measure is based on double the precision times Recall divided by the sum of precision and recall. The equation for F-Measure is given below:

$$F - \text{Measure} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (10)$$

Similarly, the equations for False-Positive Rate, False-Negative Rate and Negative Predicted Values are given as below:

$$FalsePositiveRate = \frac{FP}{FP + TN} \quad (11)$$

$$FalseNegativeRate = \frac{FN}{FN + TP} \quad (12)$$

$$NegativePredictedValue = \frac{TN}{TN + FN} \quad (13)$$

Table 2 below shows the various measures of Accuracy, Recall, Specificity, F1-score, False-Positive Rate, False Negative Rate and Negative Predicted Value.

Classifier	Accuracy	Precision	Recall	Specificity	F1-Score	FPR	FNR
LR	85.87%	0.77	0.91	0.81	0.84	0.18	0.83
NB	83.80%	0.81	0.83	0.83	0.82	0.16	0.16
SVM	87.12%	0.76	0.93	0.80	0.86	0.18	1.00
DT	94.25%	1.00	0.96	0.95	0.93	0.02	0.03
RF	97.29%	1.00	0.98	1.00	0.96	0.00	0.15
ANN	81.70%	0.83	0.85	0.85	0.80	0.14	0.14

Table 2: The measurement evaluation of ML models.

Table 3 below shows the performance of six Machine Learning Algorithms that have been used for the prediction of heart disease. The random forest outperforms over the other algorithms with a testing accuracy of 97.29% followed by a decision tree with a testing accuracy of 94.25%.

Algorithm	TrainingAccuracy	TestingAccuracy
Logistic Regression	85.55%	85.87%
Naive Bayes	83.29%	83.80%
Support Vector Machine	83.91%	87.12%
Decision Tree	100.00%	94.25%
Random Forest	100.00%	97.29%
Artificial Neural Network	85.58%	81.70%

Table 3: The Accuracy of ML models.

Figure 8 below shows the graphical representation of accuracy (Training and Testing) of six machine learning algorithms.

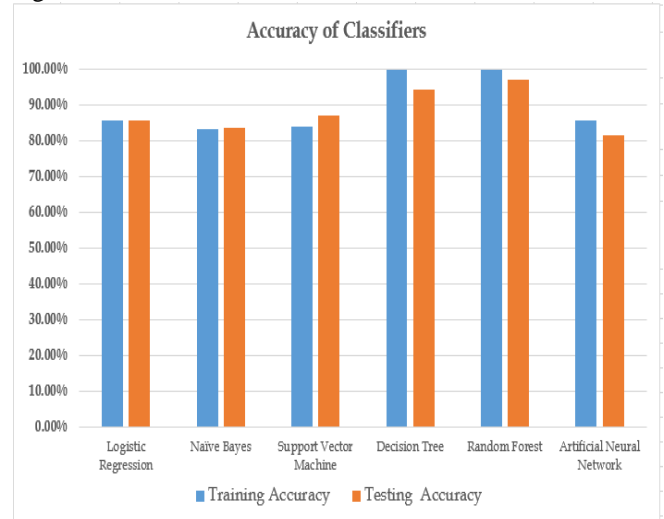


Figure 8: Graphical representation of Accuracy.

The ROC (receiver operative curve) is shown in figure9 below. ROC exhibits the performance of a classification model at all classification thresholds. The curve shows two parameters; True Positive RATE and False Positive Rate. Lowering the classification threshold classifies more items as Positive thus increases both False Positives and True Negatives.

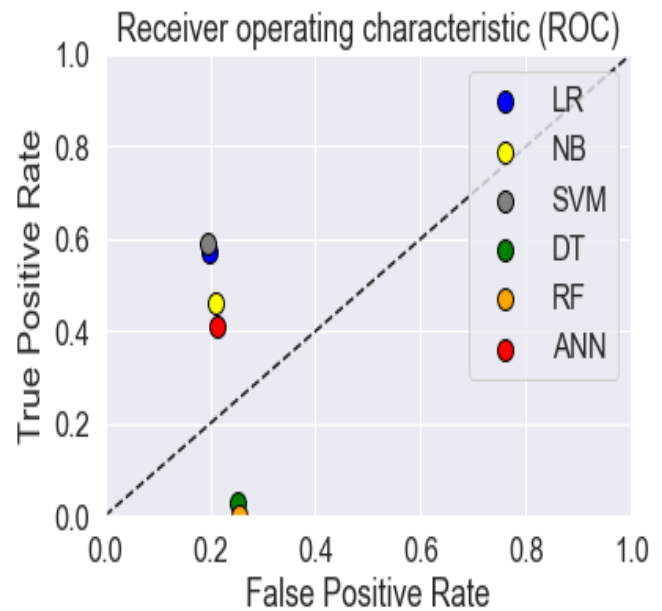


Figure 9: The measurement evaluation of ML models

IX. CONCLUSION AND FUTURE SCOPE

The Study mainly focused rather say attempted to give one an insight into prediction of heart disease(s) on its early stage and as such a comparison of various ML Algorithms have been drawn to meet this purpose. In a bid to improve the quality of data-set, pre-processing techniques where removing outliers and to handle corrupted and missing values remained the main concern. Nonetheless, we also proposed a block diagram for developing machine learning models.



These models can classify a potential patient's inclination towards (heart) disease based on contributing parameters of dataset propounded. We executed the six different machine learning algorithms to predict the disease and the results obtained were compared with various statistical measures upon which experimental results reflected accuracy of Random Forest for our dataset is (100% for training set) and (97.29% for the testing set) which is supposed to be the highest among all classifiers. 10-fold cross validation technique was employed for validation of robustness of various machine learning algorithms (LR, NB, SVM DT, RF, ANN). The RF model has taken help of also produced good results for other statistical parameters like precision, recall, specificity, F1-score, False-Positive Rate, False-Negative Rate and Negative Predicted Value. One feels that the ML classifiers need to be trained and tested with large data-sets in future to establish its scope for a better prediction of this disease. To improve the quality of research, this Study still holds scope for further research from the data acquisition up to the visualization of the results. In order to save the human lives at early stages, the current research hybridization or ensemble approaches may be used for better results to establish efficiency, reliability and validity.

REFERENCES

1. Babu, Sarath, "Heart disease diagnosing using data mining technique." Electronics Communication and Aerospace Technology (ICECA), 2017 International conference of vol.1.IEEE, 2017
2. MissChaitrali S. Dangare, Dr.Mrs. SulabhaS.Apte, A data mining approach for prediction of heart disease using neural networks, international journal of computer engineering and technology, 2012.
3. <https://www.statista.com/statistics/1108824/cardiovascular-as-cause-of-death-estimate-and-actual-worldwide/>
4. Jayumin Patel, Prof.TejalUpadhyay, Dr.SamirPatel,"Heart Disease Prediction Using Machine Learning and Data Mining Techniques ", IJCS, vol 7.No 1-sept 2015-Mar 2016.
5. SonamNikahr, A. M.Karandikhar, "Prediction Of Heart Disease Using Machine Learning Algorithms", International Journal Of Advanced Engineering, Management and Science, vol-2, June-2016.
6. SeyedaminPouyiech, Sara Vahid, Giovanna Sannino,"A Comprehensive Investigation and Comparison Of Machine Learning Techniques In The Domain Of Heart Disease", 22nd IEEE symposium on computers and communication (ISCC 2017):workshop-ICTS4EHealth 2017.
7. YounessKhourdifi, Mohamed Bahaj, " Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized By Particle Swarm Optimization and Ant Colony Optimization", international journal of intelligent engineering and systems, vol-12, No-1, 2019.
8. Mamta Alex P and Shaicy P Shaji,"Prediction and diagnosis of heart disease patients using data mining techniques", International Conference on Communication and Signal Processing, April 4-6, 2019.
9. Anaconda Inc., "Anaconda Distribution," Anaconda, 2019, [Online]. Available: <https://www.anaconda.com/distribution/>
10. Mythili T., Dev Mukherji, Nikita Padalia, and Abhiram Naidu," A Heart Disease Prediction Model using SVM-Decision Trees-Logistic Regression (SDL)", International Journal of Computer Applications (0975 – 8887) Volume 68– No.16, April 2013.
11. Larose,D.," Discovering knowledge in data: an introduction to data mining". New jersey: john wiley&sons, inc,2005.
12. Umair Shafique, Faiz Majeed, Haseeb Qaiser, and Irfan Ul Mustafa,"Data Mining in Healthcare for Heart Disease", internationaljournal of innovation and applied sciences, vol.10 No.4 Mar, 2015, pp.1312-1322
13. https://encryptedtbn0.gstatic.com/images?q=tbn%3AANd9GcRb0oO WXHEXhRAdMH_YVV8nP6_HJEIda-gwYQ&usqp=CAU
14. Umair Shafique, Faiz Majeed, Haseeb Qaiser, and Irfan Ul Mustafa,"Data Mining in Healthcare for Heart Disease", internationaljournal of innovation and applied sciences, vol.10 No.4 Mar 2015, pp.1312-1322.

15. <https://www.analyticsvidhya.com/blog/wp-content/uploads/2014/10/ANN.png>
16. GeorgeJonh and Pat Langley. "Estimating continuousdistributions in Bayesian classifiers" in proceedings of the Eleventh Conference On Uncertainty in Artificial Intelligence, Morgan Kaufman, pages 338-345,1995.
17. Umair Shafique, Faiz Majeed, Haseeb Qaiser, and Irfan Ul Mustafa,"Data Mining in Healthcare for Heart Disease", internationaljournal of innovation and applied sciences, vol.10 No.4 Mar 2015, pp.1312-1322.
18. Sonam Nikhar and A.M. Karandikar,"Prediction Of Heart Disease Using Machine Learning Algorithms", International Journal Of Advanced Engineering, Management and Science, vol-2, issue-6, June-2016.
19. H.Benjamin Fedrick David and S. Antony Belcy, "Heart Disease Prediction Using Data Mining Techniques"
20. https://miro.medium.com/max/1170/1*58f1CZ8M4i10OZYg2oRN4w.png
21. D.K. Srivastava and L. Bhambhu, "Data classification using support vector machine," J.Theor.Appl.Info.Technol., 2009
22. A. A. Abdilllah and Suwarno, "Diagnosis of diabetes using support vector machines with radial basis function kernels," Int. J. Technol., vol. 7, no. 5, pp. 849–858, 2016, DOI: 10.14716/ijtech.v7i5.1370.

AUTHORS PROFILE



Najmu Nissa received The Author has done Bachelors in Science from University of Kashmir and MCA degree in Computer Sciences from Baba Ghulam Shah Badshah University, Rajouri, J&K, India. She is currently pursuing M.Phil. degree from Department of Computer Sciences, Baba Ghulam Shah Badshah University. Her research interests are Machine Learning in Healthcare.



Sanjay Jamwal is selection grade Assistant Professor, Department of Computer Sciences BGSBU Rajouri, India. Who has guided many PhD and MPhil scholars. He has more than 30 international and 30 national publications to his credit. He has organized more than 30 workshops and national conferences. He has presented 55 papers in national and international conferences. He is permanent member of Indian Science Congress, IETE and Computer Society of India.



Shahid Mohammad received B.Sc. (Mathematics) degree from University of Kashmir, J&K, India and MCA degree from Baba Ghulam Shah Badshah University, Rajouri, J&K, India. He is currently pursuing Ph.D. degree from Department of Computer Sciences, Baba Ghulam Shah Badshah University, Rajouri, J&K, India. His current research lies Big Data Analytics in Healthcare, Artificial Intelligence and Machine Learning.