IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# A Novel Framework For Heart Disease Detection Using Machine Learning

Dr. Mahesh Kotha, Aluri Gopi, Sathini Santhosh Kumar

Associate professor, Department of CSE (AI &ML), CMR Technical Campus, Hyderabad.
Assistant Professor, Department of Electronic and Communication Engineering, CMR Engineering College, Hyderabad.
Assistant Professor, Department of Computer Science and Engineering, Vignana's Institute of Management and Technology for Women, Ghatkesar, Hyderabad

**Abstract:** According to the World Health Organization, coronary heart disease and all related diseases account for 18.6 million deaths worldwide each year. The disease's early detection and study could be important, and they might even hold the key to its ultimate cure. Because the main goal is to identify the illness at an early stage, the majority of scientists and academics concentrate on machine learning techniques that can accurately identify illnesses from large and complex data sets. These techniques then offer medicinal assistance. In order to identify cardiac illnesses early on and prevent outcomes, this research employs a variety of machine learning algorithms, including KNN Decision Tree (DT), Logistic Regression, SVM, Random Forest (RF), and Nave Bayes (NB). The article's main goal is to create a system that is entirely artificial intelligence-based and uses machine learning to identify heart diseases. We outline a technique for anticipating the progression of cardiac disease using device learning. This service, which is crucial given its estimated 88% accuracy rate over educational statistics, requires data analysis.

**Keywords:** ML, AI, classification algorithms, hear issues, Decision Tree, Heart Disease Prediction.

## I.INTRODUCTION

Heart conditions frequently take the position of circulatory diseases. These diseases focus particularly on conditions in which blood vessels become blocked or constricted, which can result in a heart attack, angina, or a stroke. Disorders of the coronary heart fall under a broader heading called "coronary heart disorders," which also encompasses conditions that affect the heart's muscle, valve, or rhythm. On the other hand, figuring out if everyone has had a cardiac illness requires knowledge of gadgetry. In either case, if these are expected beforehand, doctors may find it much easier to gather the information required for patient diagnosis and treatment. Coronary artery disease is frequently confused with heart disease. It is one of the safest computer languages, with many programmes used in the therapeutic area, according to a study by Loku et al. With projects ranging from AI-based software programmes to numerous other web programmes, it has also developed into a popular and extensively used computer language. According to Mathur's theory [2], the Python framework makes it simple to build computer or internet-based programmes. According to Guleria and Sood [3], Python-based scalable and dynamic programmes used in the healthcare industry can offer patients better and enhanced results, especially for the early detection of cardiac illnesses.

This study specifically examines a range of machine learning methods for forecasting the course of ischemic heart disease. Because the arterial heart controls the body's blood flow and is responsible for doing so, any abnormalities in it can harm other parts of the body. Heart disease has more causes today, such as unhealthy lifestyle decisions, smoking, etc. One of the major contributors to ischemic heart disease is alcohol consumption. Healthy exercise habits and early diagnosis are the best ways to avoid cardiac disease. Machine learning is a subfield of artificial intelligence that analyses massive databases and forecasts information about new or unknown data based entirely on its own learning processes. Cardiovascular diseases come in a variety of forms, and each type's symptoms include: 1 - a cardiac condition marked by an irregular heartbeat, anxiety, and stomach discomfort. 2. A condition of the blood vessels that causes chest discomfort and shortness of breath. There are many causes of coronary heart disease, such as high blood pressure, hypertension, and medicines. Coronary heart infections, coronary heart failure, cardiac arrest, hypertension, slow heartbeat, and stroke are a few examples of heart diseases. Age, a family history of the illness, blood pressure, and lipid levels are some risk factors for ischemic heart disease.

## II.RELATED WORK

Data are presently being generated from many different facilities and people in the field of fitness care. By successfully utilising this knowledge, doctors are better able to anticipate cutting-edge therapeutic approaches and improve the entire healthcare industry's distribution system [4]. One of the Python framework's most important applications is its ability to guide and motivate analytical centres in the retrieval of critical insights from data pertaining to the health care and exercise sectors. One of the most well-known computer languages in the entire world is regarded as being Python. 32% of individuals in the UK believe that this computer programme is secure for developing healthcare applications [5]. Heart attacks and coronary artery disease are frequently brought on by high levels of LDL cholesterol, also known as "bad" cholesterol (CAD). Inside the patient's coronary vessels, an obstruction has grown. Early CAD doesn't show any symptoms. Patients may experience signs like exhaustion, breathlessness, and chest pain as the plaque enlarges to the point where it restricts blood flow.

Over the past few years, many academics have been working on device learning techniques that support the healthcare industry. and specialists in spotting illnesses linked to the heart. A few instances include Nave Bayes, Logistic Regression (LR), K-Nearest Neighbor (KNN), Random Forest, Decision Tree, and SVM. With the use of unique samples and machine learning techniques, the current system has experienced extensive development to attain exceptional precision.

Senthil Kumar et AL coronary .'s illness prognosis using composite device learning methods, which include a method that looks for significant effects through the use of device learning, will lead to an improvement in forecast accuracy for cardiovascular disease.

Using deep learning, Abhay Kishore et al. [2] improved forecasts of coronary heart attacks. The essay goes into great detail about the top modules of the framework and the Recurrent Neural Network, which is a progressive characterization technique that uses the Deep Learning strategy in Artificial Neural Networks. The reason this approach is recommended is that it combines deep learning and information extraction to deliver the most accurate results with the fewest errors. Based on and making reference to this piece of art, a unique heart attack forecast tool has been created.

Senthil Kumar et al, [1] advanced prediction ofcardiovascular sickness the usage of composite device getting to knowstrategies that consists of a method which seeks tolocate giant implication thru the making use of devicegetting to know, main to more advantageous accuracy inside predictionof cardiovascular illness.

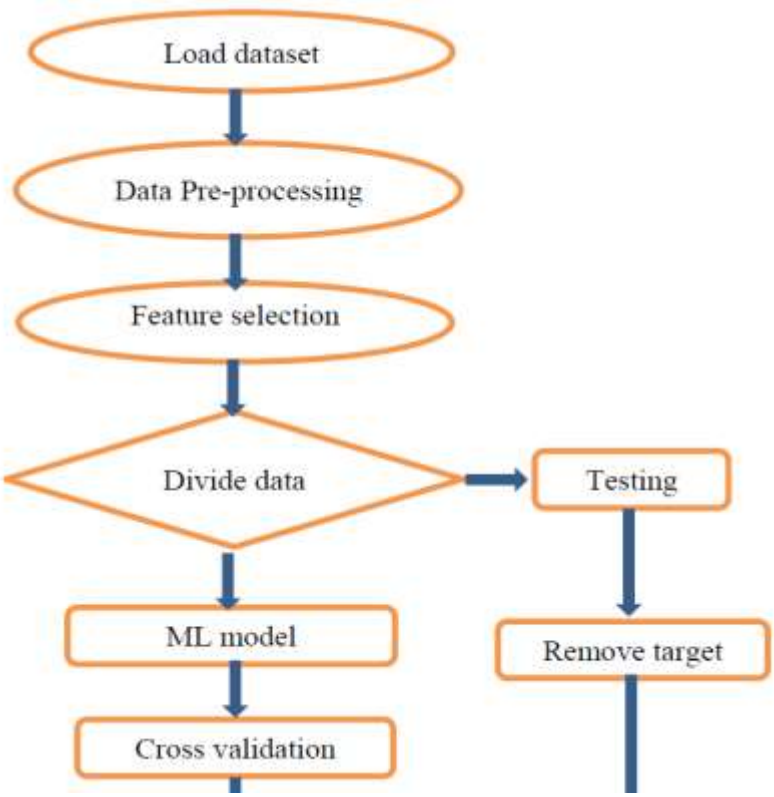
Abhay Kishore et al,[2] used Deep Learning tobroaden coronary heart assault predictions Recurrent NeuralNetwork that is a progressive characterization technique that uses the Deep Learning techniquein Artificial Neural Network, The paper is going into greatelement at the framework's number one modules, and additionally therelated assumption. This recommended method usesdeep getting to know with information mining to gain the mostcorrect effects with the bottom failures. This paintingspresents a basis and reference factor for the buildingof a exclusive sort of coronary heart assault prediction platform.

Vembandasamy et al,[3] paintings changed into preformed via way of means ofthe usage of Naïve Bayes Algo that is a powerfulindependence assumption, the information changed into acquired fromdiabatic studies institute and it includes 500 patientsfile and Naïve Bayes Algorithm gives 86.919% ofaccuracy.

Mr. Santhana Krishnan. J and Dr. Geetha. S, [4] usedcategory strategies to expect cardiac sickness inmale patients. This paintings presents complete informationconcerning Heart Diseases, overlaying Background,Prevalent Type, and Factors Associated. All 3 weakinterfaces are used there; the important thing information miningmethodologies are Naive Bayes, Artificial Neural
Networks, and Decision Trees, and those strategies areused to expect coronary heart sickness.

## III.PROPOSED APPROACH

The initial steps that were taken for each system mastering model are shown in the accompanying block diagram: selecting key functions and data cleaning, which converts raw data into a form that can be used because it cannot be used directly. These steps are then applied to the predictions from each system mastery model.



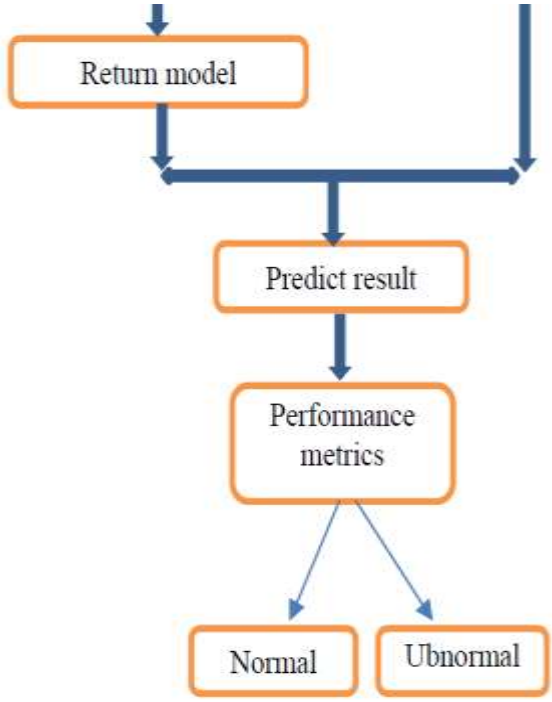



Fig 1. Complete approach

Artificial intelligence and machine learning have subsets called guided and autonomous learning. These methods support the educational process. In this instance, training on the recognised and unorganised dataset was specifically used. A machine learning prediction and classification method for ischemic heart disease is provided by this study [9]. Figure 1 depicts the suggested ML-based variant's structure;

Machine learning (ML) is crucial for determining whether various conditions, such as locomotor disorders, coronary heart illnesses, and cardiac rhythms, are present or absent. It was expected to provide medical professionals with profound insights that would allow them to individually tailor the prognosis and course of treatment for each patient. The method used in this attempt to find coronary heart disease using Python is based on the random forest collection of principles for creating coronary heart disease detection.

| S. No. | Parameter | Description | Values |
|---|---|---|---|
| I. | Age | Age in years | Numeric values |
| II. | Sex/Gender | Gender type, i.e., Male or Female | 1: Male, 2:Female |
| III. | CP | Chest Pain Level | Four types of Chest pain (0,1,2,3) |
| IV. | Trestbps | Blood pressure vale at the time of rest | < or > 120 Mg/DL |
| V. | Chol | Represents the level of Serum Cholesterol | Numeric values |
| VI. | FBS | Represents the level of sugar in fasting blood | Numeric values |
| VII. | Restecg | Represents the level of Resting electrocardiographic | Five types of Values (0,1,2,3,4) |
| VIII. | Thalach | Maximum heart rate level | Numeric values |
| IX. | Exang | Exercise enduced level | Yes / No |
| X. | Oldpeak | ST level during the workout, compared with the results of rest taken | Numeric values |
| XI. | Slope | level of peak exercise in ST-segment | Three values (0:up, 1:flat, 2:down) |
| XII. | CA | Reprenets the number of flourosopy vessels | Four values (o to 3) |
| XIII. | Thal | Used for Defect classes (4 classes) Normal; fixed; reversible; Non-reversible | Four values (0 to 3) |
| XIV. | Class | Representing the Target | Two classes (0,1) |

Table 1. Attributes in Heart disease dataset (UCI)

***Pre-Processing of Heart Disease Dataset:***

The crowd is not blocked by any apparent obstacles, and it is no longer randomly dispersed. In the compilation, there are a lot of noisy and missing numbers. These data have been pre-processed to account for price-deficit issues and produce accurate predictions. Pre-processing steps for statistics include cleaning, change (normalisation and aggregation), integration, and reduction [39]. The proposed computer uses methods like normalisation and aggregation for data pre-processing.

The results obtained by first using Exclusion Forest for feature extraction and abnormality elimination and then resolving overfitting issues with a widely used dataset are, however, very intriguing. Using a variety of plotting techniques, the skewness of the data, anomaly identification, and statistics distributions are examined [3]. The correlation between number characteristics and cardiac conditions is shown in Figure 2. This matrix gives a general overview of the complete dataset as a basis for a more in-depth analysis and as a potential area for further study.

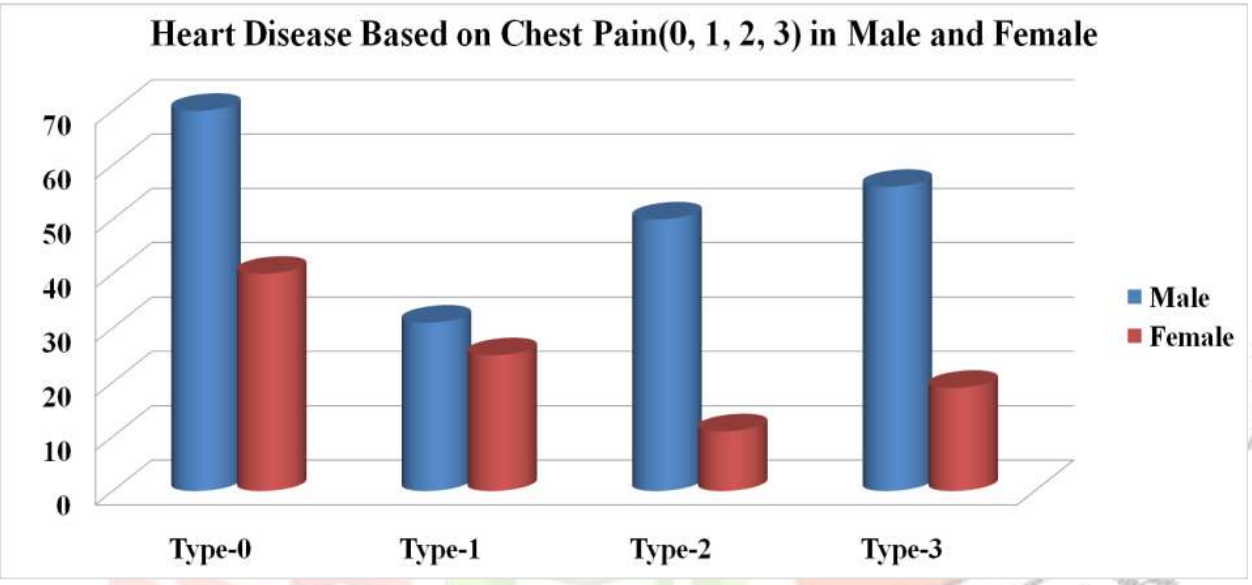



Fig. 2. Heart disease ratio based on Chest pain level

IV.ALGORITHMEMIC APPROACH

***Naïve Bayes***

A simple and effective set of instructions, the Naive Bayes set of rules for guided learning. The Bayes hypothesis is the basis for it. Chance and opportunity, along with a few minimum educational requirements, form the basis of New Brunswick. It is essential for classification that the lifestyle characteristic can be categorised separately from other teachings. The Naive Bayes set of rules and datasets with reliable measurement instructions are typically used in simple predictive modelling. The equation has already shown that P(c)(A) is assigned a few early possibilities of occurrence.

$$p\left(\frac{A}{c}\right) = \frac{p\left(\frac{A}{c}\right)p(c)}{p(A)}$$

$$p(\frac{c}{A}) = p(\frac{A_1}{c})p(\frac{A_2}{c}) * \ldots * P(\frac{A_n}{c})\, p(c)$$

***Decision Tree***

Using a low chart or tree-like structure, the Decision Tree collection of principles is displayed. A group of supervised learning ideas are used to address type problems. While the inner subset (node) shows the dataset characteristics, the outer subset (branch) represents the result. This division of the data into smaller groups for easier management is known as subset analysis. Entropy is used in this set of guidelines to determine the design uniformity and statistics gain, and the feature with the greatest statistics gain is then selected. Fast, reliable, and user-friendly describe Decision Tree. [9]

Swap choice trees (DT) were used in conjunction with PCA by M.A. Jabbar et al. [7] to achieve 93.5% accuracy.Outstanding results were obtained in Kamran Farooq et al's experiments [8,] which combined a decision tree-based predictor with forward selection. and achieved with a 79.46% accuracy rate.

***Logistic Regression***

The logistic regression variant is one of the best statistical models for estimating the probability of a specific event or accomplishment, including success or failure. Numerous anticipated variables, whether fantastic or virtual, are used in logistical regression. Logit and the overall image of entropy are two additional names for logistic regression. Regression for transportation is a component of supervised machine learning methods for "type" tasks. For binary and linear type problems that work well with linear separation levels, logistics regression is a fast and more environmentally responsible solution.The logistics regression is used to forecast the rate of information based on the historical observation of a data gathering. This is done by examining the relationship between a feature of based information and one or more characteristics of unbiased information.

***Support Vector Machine***

The Support Vector Machine is a well-known collection of supervised learning ideas used in regression and classification methods (SVM). It is considered to be the most successful method for gaining computer management. To deal with excessive complexity, some SVMs employ hyper planes, which provide the greatest separation between two commands. The range of entry variables with this technique is noticeably greater than the range of data.

SVM can be utilised as a forecast in addition to classifying the directions utilising the hyperplane. The statistics separation technique determines the great hyperplane, so changing or removing its area necessitates resetting every other additional degree. The statistical element that is closest to the hyperplane is called a support vector, and support vectors play a key role in the statistics-based approach. The larger the margin, or the separation between the hyperplane and the nearest statistical set, the higher the likelihood that newly discovered statistics will be correctly classified.

The margin is essential for the hyperplane to accomplish its exceptional performance. In datasets from People's Hospital, SVM accuracy was 98.9% [6] and SVM with boosting method showed 84.81% [7], according to Shan Xu et al.

***Random Forest***

The Random Forest set of principles, which is applied to both classification and regression, is one of the best categorization methods [8]. The Random Forest technique, as its name suggests, comprises of numerous independent choice trees that operate as a single entity. A particular character tree can be located among all the limbs with the same distribution. Utilizing bootstrap aggregation and random function selections, the forecast is added up.

$$GiniImpurity = 1 - Gini$$

$$Gini = P12+P22*P32 \ldots\ldots..+Pn2$$

***K-Nearest Neighbor***

The K-Nearest Neighbor strategy, a category tactic, is simple but effective. It is typically used to categorise issues when there is little prior knowledge regarding the spread of the data and when there are no obvious presumptions. This method entails selecting the acceptable facts factors that are most similar to the fact component for which a target fee is missing in the education set, and then giving those facts factors the average fee of the collected facts factors.

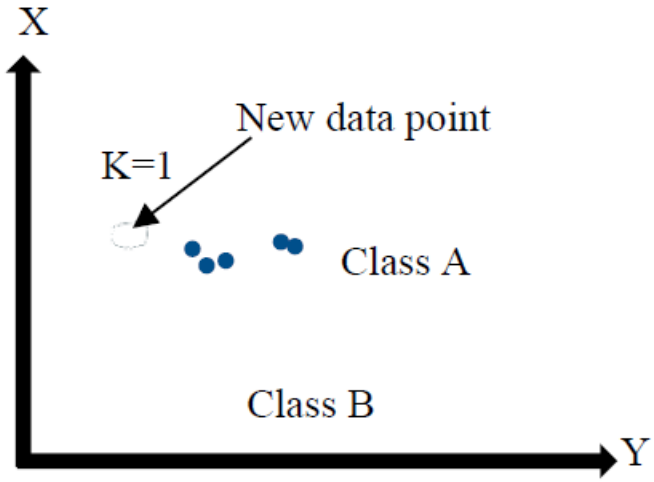



Fig 3. KNN algorithm chart

When using the 10-pass validation technique and an okay fee of 9, KNN has an accuracy of 84.5%. In [8,] KNN with Ant Colony Optimization outperforms other methods with an accuracy of 72.56% and an error rate of 0.526.

V.F1 SCORE VALUES

***Accuracy:*** The percentage of incidents that were correctly forecasted across the entire dataset is as follows:

$$Accuracy=TP+TN(TI)$$

***Sensitivity/Recall:*** the percentage of positive data set events that were accurately anticipated [9]. The memory equation is presented in equation [10[.

. Sensitivity or Recall= $TP$ $(TP+FP)$

• ***Precision*:**It gauges the proportion of correctly foreseen good outcomes. As a consequence, for this small group, accuracy depends on precision. It is calculated by subtracting the total number of anticipated sample instances from the total number of instances in which affirmative predictions were effective [3].

$$Precision=[TPTP+FP]$$

| ML Method | Sensitivity / Re-call | Precision | F1-Score | AUC (Area under curve) |
|---|---|---|---|---|
| Decision Tree (DT) Method | 0.7548 | 0.778 | 0.864 | 0.812 |
| Naive Bayes (NB) Method | 0.804 | 0.854 | 0.865 | 0.845 |
| Random Forest (RF) Method | 0.832 | 0.887 | 0.879 | 0.898 |
| k-Nearest Neighbor (KNN) Method | 0.948 | 0.917 | 0.908 | 0.799 |
| SVM(Linear Kernel) Method | 0.878 | 0.907 | 0.885 | 0.809 |
| Logistic Regression (LR) Method | 0.861 | 0.881 | 0.865 | 0.813 |

Table 2. Experimental Results for various ML Methods

| Method Name | TP | FP | FN | TN |
|---|---|---|---|---|
| LR Method | 80 | 17 | 11 | 104 |
| SVM(Linear Kernel) Method | 89 | 8 | 6 | 109 |
| KNN Method | 82 | 15 | 13 | 102 |
| RF Method | 97 | 0 | 0 | 115 |
| DT Classifier Method | 97 | 0 | 0 | 115 |
| NB Method | 97 | 0 | 0 | 115 |

Table 3. Confusion Matrix Results for Training (Heart Disease Dataset)

• ***F1-Score:***The F1-Score, a compound harmonic mean, combines memory and accuracy (mean of reciprocals). The model's precision or ability to identify only each pertinent data source was first evaluated by researchers [2].

$$F1Score=2\times[precision\times recall(precision+recall)$$

| Method Name | TP | FP | FN | TN |
|---|---|---|---|---|
| LR Method | 34 | 7 | 5 | 45 |
| SVM(Linear Kernel) Method | 36 | 5 | 6 | 44 |
| KNN Method | 35 | 6 | 6 | 44 |
| RF Method | 33 | 8 | 8 | 42 |
| DT Classifier Method | 34 | 7 | 13 | 37 |
| NB Method | 33 | 8 | 8 | 42 |

Table 4. Confusion Matrix Results for Testing (Heart Disease Dataset)

• ***UC (Area under the curve):***UC (Area under the curve): A structure that favours a randomly selected positive case over a specifically selected negative case. The complete area under the curve receiver required to implement characteristic curves is converted to graphic components in order to evaluate the effectiveness of the models.

$$y = f(x)\ betweenx = a\ ,\ \&x = b$$

| | male | age | cigsPerDay | prevalentStroke | diabetes | sysBP |
|---|---|---|---|---|---|---|
| **0** | 1.153113 | -1.234283 | -0.758062 | -0.077014 | -0.162437 | -1.196267 |
| **1** | -0.867217 | -0.417664 | -0.758062 | -0.077014 | -0.162437 | -0.515399 |
| **2** | 1.153113 | -0.184345 | 0.925410 | -0.077014 | -0.162437 | -0.220356 |
| **3** | -0.867217 | 1.332233 | 1.767146 | -0.077014 | -0.162437 | 0.800946 |
| **4** | -0.867217 | -0.417664 | 1.177931 | -0.077014 | -0.162437 | -0.106878 |
| **...** | ... | ... | ... | ... | ... | ... |
| **4235** | -0.867217 | -0.184345 | 0.925410 | -0.077014 | -0.162437 | -0.061487 |
| **4236** | -0.867217 | -0.650984 | 0.504542 | -0.077014 | -0.162437 | -0.265747 |
| **4237** | -0.867217 | 0.282295 | -0.758062 | -0.077014 | -0.162437 | 0.051991 |
| **4238** | 1.153113 | -1.117623 | -0.758062 | -0.077014 | -0.162437 | 0.392425 |
| **4239** | -0.867217 | -1.234283 | 1.767146 | -0.077014 | -0.162437 | 0.029296 |

4240 rows × 6 columns

Figure 2. Dataset After Dropping Columns after Feature Selection

| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---|---|---|---|---|---|---|---|---|
| 1.000 | 1.000 | 0.651 | 1.000 | 0.789 | ? | 0.497 | 0.650 | tested_negative |
| 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.497 | 0.348 | tested_positive |
| 0.651 | 0.651 | ? | 0.651 | ? | ? | 0.497 | 0.544 | |

Fig 3. F1-scores

## VI.CONCLUSION

This research predicts and diagnoses coronary heart disease, which is typically very important. It accomplishes this by analysing various machine learning methods. Based on the aforementioned research, one can infer that machine learning algorithms have a simply enormous capacity for forecasting and analysing cardiovascular diseases or any heart-related conditions. With a wide variety of datasets, the Decision Tree method works poorly. Because it employs a variety of methods to deal with the overfitting issue, Random Forest generated remarkably good results. various choice trees The Nave Bayes classifier performed well and was very fast in terms of calculation. SVM increases effectiveness in a significant portion of instances.

## VII.REFERENCES

[1] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, “Heart disease identification method using Machine Learning classification in e-healthcare,” IEEE Access, Vol. 8, 2020, pp. 107562–107582.

[2] M. Mullen, A. Zhang, G. K. Lui, A. W. Romfh, J. W. Rhee, and J. C. Wu, “Race and Genetics in Congenital Heart Disease: Application of iPSCs, Omics, and Machine Learning Technologies,” Frontiers in Cardiovascular Medicine, Vol. 8, 2021, pp. 37–51.

[3] S. Mohan, C. Thirumalai, and G. Srivastava, “Effective heart disease prediction using hybrid Machine Learning techniques,” IEEE Access, Vol. 7, 2019, pp. 81542–81554.

[4] N. K. Trivedi, S. Simaiya, U. K. Lilhore, and S. K. Sharma, “An efficient credit card fraud detection model based on Machine Learning methods,” International Journal of Advanced Science and Technology, Vol. 29, No. 5, 2020, pp. 3414–3424.

[5] Abhay Kishore1, Ajay Kumar2, Karan Singh3,Maninder Punia4,Yogita Hambir5,"Heart Attack Prediction Using Deep Learning",International Research Journal of Engineering and Technology(IRJET), Volume: 05 Issue: 04 | Apr-2018.

[6] Vembandasamy, K., Sasipriya, R. and Deepa, E. (2015) HeartDiseases Detection Using Naive Bayes Algorithm. IJISETInternationalJournal of Innovative Science, Engineering &Technology, 2, 441-444.

[7] Mr. Santhana Krishnan.J, Dr.Geetha.S," Prediction of Heart DiseaseUsing Machine Learning Algorithms",2019 1st InternationalConference on Innovations in Information and Communication Technology(ICIICT),doi:10.1109/ICIICT1.2019.8741465.

[8] Ravindra Changala, "Challenges and Solutions for the Semantic Web and Future of Document Management in Enterprises " in Journal of innovations in computer science and engineering (JICSE), Volume 6, Issue 1, Pages 10-13, September 2016.ISSN: 2455-3506.

[9] Shan Xu ,Tiangang Zhu, Zhen Zang, Daoxian Wang, Junfeng Hu andXiaohui Duan et al. "Cardiovascular Risk Prediction Method Basedon CFS Subset Evaluation and Random Forest Classification Framework", 2017 IEEE 2nd International Conference on Big DataAnalysis.

[10] Shan Xu ,Tiangang Zhu, Zhen Zang, Daoxian Wang, Junfeng Hu andXiaohui Duan et al. "Cardiovascular Risk Prediction Method Basedon CFS Subset Evaluation and Random Forest ClassificationFramework", 2017 IEEE 2nd International Conference on Big DataAnalysis.

[11] Ravindra Changala, "Automated Health Care Management System Using Big Data Technology", at Journal of Network Communications and Emerging Technologies (JNCET), Volume 6, Issue 4, April (2016), 2016, pp.37-40,ISSN: 2395-5317, ©EverScience Publications.

[12] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe DePietro, Hamid Arabnia, Juan Gutierrez et al. " A ComprehensiveInvestigation and [8] Houda Mezrigui, FouedTheljani and Kaouther Laabidi et al. "Decision Support System forMedical Diagnosis Using a Kernel-Based Approach", ICCAD'17,Hammamet - Tunisia, January 19- 21, 2017.

[13] Ravindra Changala, "Retrieval of Valid Information from Clustered and Distributed Databases" in Journal of innovations in computer science and engineering (JICSE), Volume 6, Issue 1, Pages 21-25, September 2016.ISSN: 2455-3506.

[14] Amir Hussain, Peipei Yang, Mufti Mahmud and Jan Karasek et al."A Novel Cardiovascular Decision Support Framework foreffectiveclinical Risk Assessment.", 978-1-4799-4527- 6/14/$31.00 ©2014IEEE.

[15] D. Krishnani, A. Kumari, A. Dewangan, A. Singh, N.S. Naik,Prediction of Coronary Heart Disease using Supervised MachineLearning Algorithms, TENCON 2019 – 2019 IEEE Reg. 10 Conf.,2019, pp. 367–372.

[16] Ravindra Changala, A Dominant Feature Selection Method for Deep Learning Based Traffic Classification Using a Genetic Algorithm, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, ISSN : 2456-3307, Volume 8, Issue 6, November-December-2022, Page Number : 173-181.

[17] Puneet Bansal and Ridhi Saini et al. "Classification of heart diseases from ECG signals using wavelet transform and KNN classifier", International Conference on Computing, Communication and Automation (ICCCA2015).

[18] K. Bhanot, "towarddatascience.com," 13 Feb 2019. [Online]. Available: https://towardsdatascience.com/predicting-presence-of-heart-diseases-using-machinelearning-36f00f3edb2c. [Accessed 2 March 2020].

[19] Ravindra Changala, Development Of Predictive Model For Medical Domains To Predict Chronic Diseases (Diabetes) Using Machine Learning Algorithms And Classification Techniques, ARPN Journal of Engineering and Applied Sciences, Volume 14, Issue6.

[20] Ravindra Changala ,"Development of Predictive Model for Medical Domains to Predict Chronic Diseases (Diabetes) Using Machine Learning Algorithms And Classification Techniques", ARPN Journal of Engineering and Applied Sciences, Volume 14, Issue 6, 2019.

[21] Ravindra Changala, "Evaluation and Analysis of Discovered Patterns Using Pattern Classification Methods in Text Mining" in ARPN Journal of Engineering and Applied Sciences, Volume 13, Issue 11, Pages 3706-3717 with ISSN:1819-6608 in June 2018.

[22] [Online]. Available: https://www.kaggle.com/ronitf/heart-disease-uci#heart.csv.. [Accessed 05 December 2019].

[23] M. A. K. S. H. K. M. a. V. P. M Marimuthu, "A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach".