# Analyzing of Heart Disease Based on Various Machine Learning Techniques

**Shinde Divya.1Jadhav Pankaj.2 Nalawade Sanchit.3 Pawar Mahesh.4 Dr.Khatal.S.S5**

1,2,3,4BE Student 5Head of Department

1,2,3,4,5Department of Computer Engineering

1,2,3,4,5Sharadchandra Pawar College of Engineering, Otur, Maharashtra, India

*Abstract*— HEART disease is one in every most common disease in now a days, and for Who those provide health care, and save their life the heart disease dataset to classify it properly to predict heart disease cases with large amount data thought of this case a cardiovascular disease. Prediction system is developed exploitation supply regression nearest neighbor, call tree, random forest the heart malady risk level. Support vector machine obtained the best accuracy result of heart disease people and it is obtained by supply regression, KNN classifier and call tree severally. It is impractical for a common man to frequently system in place which is handy and the same time reliable, in predicting the chances of heart disease thus proposed system. Heart attack disease is one of the leading cause of the death worldwide. Now days machine learning when implemented in health care is capable of early and accurate detection of disease .the dataset are been proceed in python using machine learning algorithm. i.e Random forest algorithm. This technique uses for old patient record. and preventing the loss of lives this work reliable, heart disease prediction system is implemented using strong machine learning algorithm, which is the random forest algorithm.

*Keywords:* Machine Learning, Logistic Regression, Heart Disease, Support Vector Machine, Accuracy

## I. INTRODUCTION

Data mining is that the method by that we are able to realize typically unknown scriptures, patterns, and current trends in databases and it uses that piece of knowledge to structure prognostic models. data processing technology combines analysis supported statistics, machine learning algorithmic rule, and information technology management system to come up with disclosed patterns and establish relationships from immense databases. The World Health statistics 2012 highlights the problem that every one in 3 adult age bracket showed susceptible to high blood pressure- a state of affairs that leads to half the deaths from heart problems and strokes. Disease-related to the guts, also known as disorder (CVD), discusses numerous conditions that have an effect on the guts not simply the illness. This juncture proven fatal for one person in each thirty four seconds within the United States. Heart disease of the coronary arteries, cardiopathy, and vessel health problems assure subdivisions where the blood is tense and its circulation is formed throughout the body. diagnosing is a vital task that has got to be performed with efficiency. this can be chiefly done underneath a doctor's guidance. This causes disappointing results & excessive medical prices of treatments provided to patients. So, we conclude that an automatic diagnosis and prediction system would prove very favourable.

## II. LITERATURE REVIEW:

Numerous studies are done that have targeted on the diagnosing of cardiopathy. they need applied completely different data mining techniques for diagnosing & achieved completely different probabilities for various ways. This system evaluates those parameters victimisation the information mining classification technique. The datasets ar evaluated in python victimisation 2 main Machine Learning Algorithms: the choice Tree algorithmic rule and therefore the Naive Bayes algorithmic rule that shows the most effective algorithmic rule between these 2 in terms of the accuracy level of cadiopathy[1].

Aditi Gavhane et al.foreseen coronary failure for early diagnosis to scale back the count of deaths. For this downside machine learning plays a significant role during this paper. This prediction takes individuals from the zone of their life. In this paper, we have a tendency to use the KNN algorithmic rule and random forest algorithmic rule to predict the center attack earlier[2].

Senthil kumar et al. introduced a prediction model with different combos of options, and several other celebrated classification techniques. It made associate degree increased performance level with associate degree accuracy level of eighty eight. 7% Hybrid Random Forest with Linear Model (HRFM) [3].

Himanshu Sharma et al. declared and well-tried that machine learning opportunities for precise prediction of a coronary failure. Paper provides heaps of knowledge regarding state of art An analytical comparison has been provided to assist new researchers operating during this field [4].

M. Nikhil Kumar et al. worked with eight algorithms including call Tree, J48 algorithmic rule, provision model tree algorithmic rule, Random Forest algorithmic rule, Naïve Bayes, KNN, Support Vector Machine, Nearest Neighbor to predict heart diseases. The accuracy of the prediction level is high once victimisation additional attributes [5].

Amandeep Kaur et al. declared that data processing is associate degree important stage of the KDD method that may be used for disease management, diagnosis, and prediction in healthcare organizations. This paper discusses reviews on different ways and approaches in data processing that have been wont to predict cardiopathy [6].

Pahulpreet Singh Kohli developed associate degree increased New Dynamic processing (ENDDP) algorithmic rule to predict the early stages of cardiopathy. The results prove the performance of the projected system.

## III. FUTURE WORK:

To initiate with the work we've got started collection data in every and each facet towards the goal of the system. In the

first place, the analysis was within the direction of the most causes or the factors that have sturdy influence on the center health. Some factors square measure unmodifiable like age, sex and family background however there square measure some parameters like pressure level, heart rate etc. which may be unbroken au fait by following sure measures [4]. several doctorsstay the center healthy. Following square measure the parameters which square measure thought of for the study in planning the system which have major risk share with reference to CAD [5] 1. Age 2. Sex 3. pressure level 4. Heart Rate 5. Diabetes 6. Hyper steroid alcohol

## IV. DATA SET INFORMATION:

The name of the dataset is heart.csv. There are 303 instances in this dataset, where the cases are either people having heart disease or they're healthy.

Among 303, 165 (54.45%) cases are people with cardiovascular disease and 138 (45.54%) are people without cardiopathy. the quantity of attributes is 14. There are no missing values within the data set nor any null values cholesterol, blood glucose level, ECG result, maximum heart rate achieved, exercise- induced angina, ST depression, the,. bar chart (Fig.1) showing the positive and negative cases (1=positive, 0=negative) Scatter plot (Fig.2)showing the positive and negative cases looking on age.
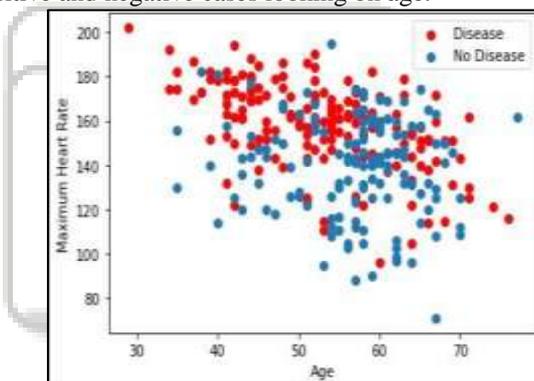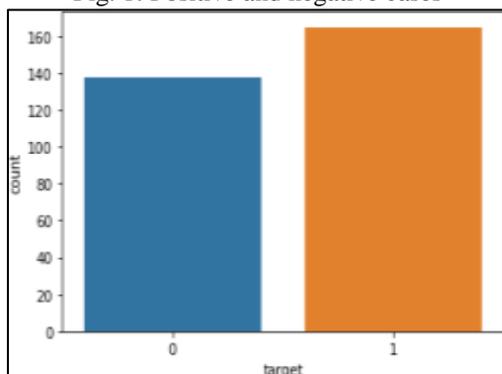


Fig. 1: Positive and negative cases



Fig. 2: Positive and negative cases depending on age

## V. METHODOLOGY:

### A. Data Set Information:

The main objective of this analysis is to develop a cardiopathy prediction system. The system will discover and extract hidden data related to diseases from a historical heart knowledge set cardiopathy prediction system aims to use data processing techniques on medical knowledge set to help within the prediction of the guts diseases

### B. Training and Testing:

The coaching section extracts the options (independent variables) from the dataset and therefore the testing section (containing dependent variables) is employed to work out however the acceptable model behaves for prediction. We've got divided the dataset into 2 sections. These are the coaching and testing phases. We've got split the dataset into ninetieth coaching and 100 percent testing section. And that we have taken the random state as one. For initializing the fastened internal random variety generator, we have a tendency to use the random state parameter which can decide the cacophonic of knowledge into train and check indices. Setting a random state can guarantee a hard and fast price that constant sequence of random numbers are generated on every occasion the code is being run. Setting random state, a hard and fast price can guarantee that constant sequence of random numbers is generated on every occasion we have a tendency to run the code. Then we have a tendency to scaled information |the info| exploitation normal scattered and fitted the coaching and testing data exploitation 'fit. transform'

### C. Classification Used:

1) Logistic regression: Logistic regression is the classification algorithm. Logistic regression transform its output using the logistic sigmoid function and give probability value . the hypothesis of logistic regression tends to limit the cost function in between the 0 and 1 the sigmoid function maps any real value to the another value between 0 and 1 . in this scenario we use sigmoid function to make prediction probabilities.

2) KNN CLASSIFIER: k- nearest neighbour is one of the simplest algorithm. KNN algorithm for regression as well as for classification but most probably it was used for classification problem. KNN algorithm consider the similarly between the new data and available the data and put this new data into the category which is most similar to category. When new data appear then it can be easily classified into a well suitable category by using KNN algorithm. KNN algorithm at learning phase store the dataset and when it gets new data set then knn algorithm classifies these data into a category which is similar to new data .

3) SUPPORT VECTOR MACHINE: support vector machine associated with learning algorithm such as supervised learning model. support vector machine can be used for classification for analyzing data. We plot each data item as a point in n dimensional space with the value of each future being of the svm is available in sickie library and these is widely used library for implementation of machine learning algorithm. The main aim of support vector machine algo is to determine a hyper plane in a n-dimension space which uniquely identify the data point. For separating the two classes of data points there are many hyper plans there are would be which will be used for find out maximum distance between data point and both classes.

4) RANDOM FOREST ALGORITHM: It is supervised learning technique of machine learning. It also used for

both classification and regression problem. It is the process of combining multiple classifier to solve a complex problem and improve the performance of model. It contains the number of decision trees on various subset of given dataset which improve the accuracy of dataset. the greater number of trees in the forest leads to the highest accuracy present the problem of overwriting and random forest combine multiple trees.to predict the class of dataset. Random forest take less time as compared to other algorithm

## VI. DECISION TREE:

It is type of machine learning. the leaves are the dicision or the final outcome it is a tree structure classifier, where internal node shows the feature of data set branches of represent decision run the decision are performed on the basis of future of dataset. branches represent the decision if condition is true and then goes to the next node attached to that decision .

## VII. RESULT:

| Sl no | Algorithm | Testing accuracy |
|-------|-----------|------------------|
| 1 | Logistic regression | 87.09% |
| 2 | K nearest neighbour | 70.96% |
| 3 | Random forest classifier | 90.32% |
| 4 | Support vector machine | 90.32% |
| 5 | Decision tree | 83.87% |

Table 1: Comparison of performances difference classifier

## VIII. CONCLUSION:

Heart disease prediction is the mejor challenge in today world .random forest algorithm provide an essential way an predicting heart disease which use N of decision trees and output the class that is the average of all decision tree output with this application it become easier for patient who will be away from hospital they can immediately we these application in prediction of heart disease just by entering the report value. The rate of people dying due to heart disease will reduce by using these proposed system.

## REFERENCE:

[1] "Prediction of Hear Disease using Machine Learning Algorithms" Krishnan J Santhana and S Geetha ICIICT |Year:2019| Conference Paper | Publisher: IEEE.
[2] "Prediction of Heart Disease using Machine Learning". Aditi Gavhane, Gouthami Kokkula, Isha Panday and Kailash Devadkar, Proceedings of the 2nd International conference on Electronics Communication and Aerospace Technology(ICECA) |Year :2018| Conference Paper | Publisher: IEEE.
[3] "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" Senthil Kumar, Mohan Chandrasegar Thirumalai and Gautam Srivastva |Year:2019| Conference Paper | Publisher: IEEE.
[4] "Prediction of Heart Disease using Machine Learning Algorithms: A Survey" International Journal on Recent and Innovation Trends in Computing and Communication vol. 5 no. 8, Himanshu Sharma and M A Rizvi |Year:2019| Conference Paper | Publisher: IEEE
[5] "Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools" International Journal of Scientific Research in Computer Science Engineering and Information Technology IJSRCSEIT, M. Nikhil Kumar K. V. S. Koushik and K. Deepak | Year: 2019| Conference Paper | Publisher: IEEE.
[6] "Heart Diseases Prediction using Data