iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 8 **Issue:** VI **Month of publication:** June 2020

**DOI:** http://doi.org/10.22214/ijraset.2020.6376

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 8 Issue VI June 2020- Available at www.ijraset.com*

# Cardiovascular Diseases (CVDs) Detection using Machine Learning Algorithms

Eqbal Ahmad[1], Anamika Tiwari[2], Arpan Kumar[3]
[1, 2, 3]*Department of CSE,* [1,2]*Bhabha Institute of technology,Kanpur,*[3]*Himalayan Institute of Technical, Lucknow*

***Abstract:*** ***Cardiovascular disease are group of diseases which is caused due to the dysfunction of heart and blood artery and it incorporate coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic heart disease, and deep vein thrombosis and pulmonary embolism. This paper represents a model for detecting cardiovascular diseases using machine learning algorithm. The methodology used in this research is agile methodology, in which during the stages of production process, planning, requirements analysis, designing, coding, testing and documentation is also done in parallel. In this paper patient dataset is used to train the model using four different machine learning algorithms (Support Vector Classifier, K-Nearest Neighbors Classifier, Random Forest Classifier, and Decision Tree Classifier). The predictions will done through algorithm which give the most precise result. This model was implement to the web using a python framework flask, the web has 13 inputs which was filled by the user in order to make prediction. The machine learning algorithms and the model itself implemented using python programming language and flask. After taking in to consideration all four machine learning algorithms we uses a K-Nearest Neighbors Classifier algorithm and the results obtained from the prediction have an accuracy of about 85.83%, which is pretty good for any model.***

***Keywords*****:** ***Cardiovascular Disease, Machine Learning, K-Nearest Neighbors, Support Vector machine, Decision Tree, Random Forest.***

## I. INTRODUCTION

2016, speaking to 31% of every single worldwide passing, of these passings, 85% are because of coronary failure and stroke, more than seventy five percent of CVD passings occur in low-and center income countries, out of the 17 million unexpected losses (younger than 70) because of no communicable sicknesses in 2015, 82% are in low and center pay nations, and 37% are caused. Raised circulatory strain is the main hazard factor for cardiovascular illness.

The worldwide commonness of raised circulatory strain (characterized as systolic as well as diastolic pulse more than or equivalent to 140/90 mmHg) in grown-ups matured 18 years and over was around 24.1% in men and 20.1% in ladies in 2015.The number of adults with raised blood pressure increased from 594 million in 1975 to 1.13 billion in 2015. [1] As we can see from the following figure how dangerous CVDs are one of four people die of CVDs, analyze by villa medica by 2030 CVDs has surpassed infectious disease in developing as the leading cause of death.

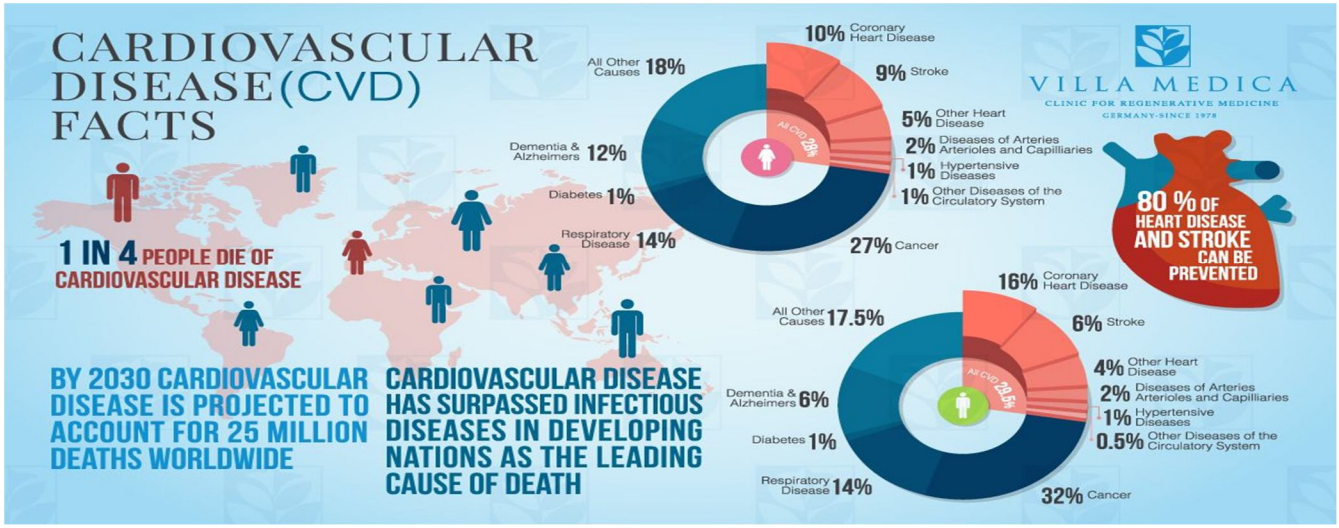



Figure 1: Cardiovascular disease facts and globel condition [6].

Most cardiovascular illnesses can be forestalled by tending to conduct chance factors, for example, tobacco use, undesirable eating routine and corpulence, physical idleness and hurtful utilization of liquor utilizing populace wide procedures, individuals with cardiovascular sickness or who are at high cardiovascular hazard (because of the nearness of at least one hazard factors, for example, hypertension, diabetes, hyperlipidaemia or effectively settled malady) need early location and the executives utilizing guiding and medications, as fitting.

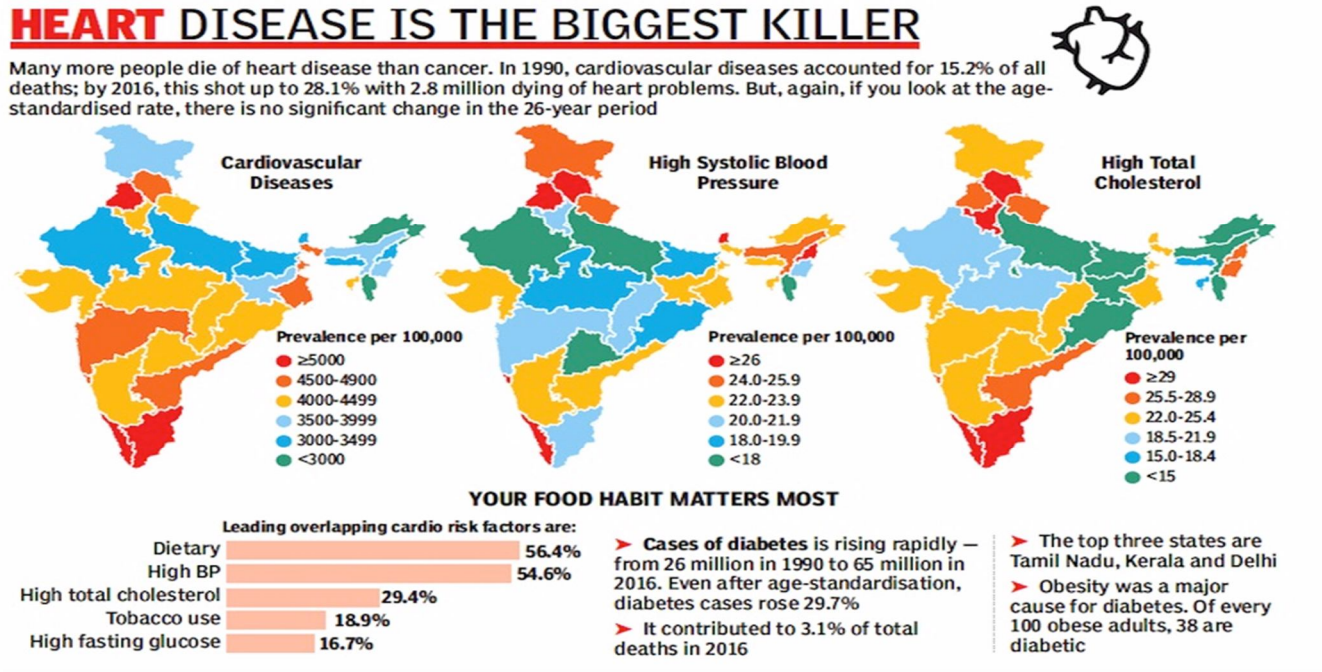



Figure 2: Cardiovascular disease condition in India and comparison with other dangerous

As we talk about CVD it is a group of disease which can be caused due to heart and blood artery and this group include many diseases like coronary heart diseases : a disease of the blood artery supplying the heart muscle, cerebrovascular disease: a disease of the blood artery supplying the brain, peripheral arterial disease: disease of blood artery supplying the arms and legs, and rheumatic heart disease: damage to the heart thews and heart valves from rheumatic fever, caused by streptococcal bacteria.

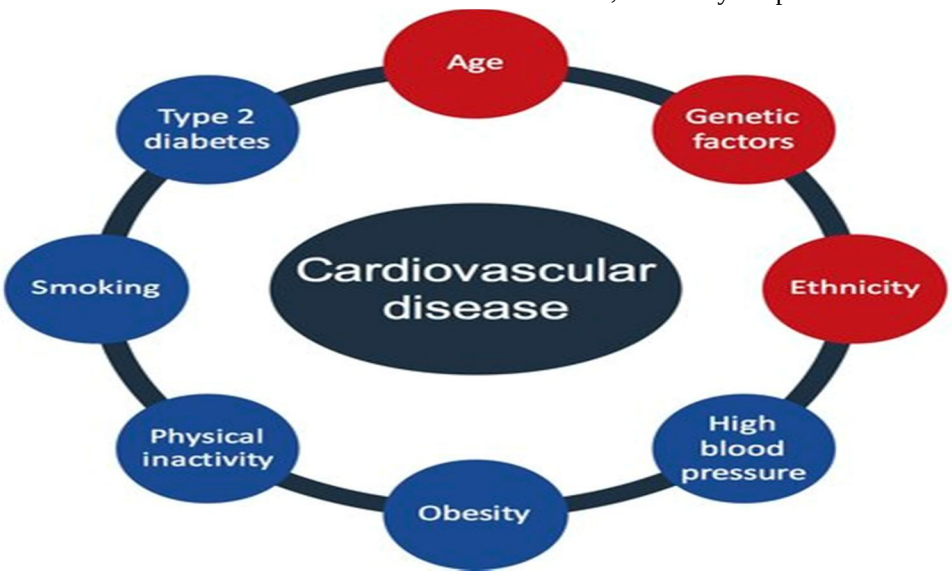



Figure 3: Cardiovascular disease controllable and uncontrolable risk factors

This paper presents a model for detecting cardiovascular disease using different machine learning algorithms and concludes to a result which algorithm is better for making such model using python programming.

## II. LITERATURE REVIEW

In the year 2012 research conducted by Nidhi Bhatla and Kiran Jyoti is "An Analysis of Heart Disease Prediction using Different Data Mining Techniques" in this paper author aims at analyzing the various data mining techniques introduced in recent years for heart disease prediction. Various techniques and data mining classifiers are defined in this work which has emerged in recent years for efficient and effective heart disease diagnosis. The investigation shows that Neural Network with 15 characteristics has demonstrated the most noteworthy exactness for example; Decision Tree has additionally performed well with 99.62% precision by utilizing 15 characteristics. Moreover, in combination with Genetic Algorithm and 6 attributes, Decision Tree has shown 99.2% efficiency [2].

In the paper "Prediction of heart disease using machine learning algorithms" authors founds that during small datasets in some other cases most of time decision trees direct them to a solution which is not accurate, but when they look at Naïve Bayes results they are getting more accurate results with probabilities of all other possibilities but due to guidance to only one solution decision trees may miss lead. Finally they can say by this experiment that Naïve Bayes is more accurate if the input data is cleaned and well maintained even though ID3 can clean itself, it can't give exact outcomes without fail, and in this equivalent way Naïve Bayes likewise won't give precise outcomes each time we have to think about after effects of various calculations what's more, by the entirety of its outcomes if a forecast is made it will be exact. But they can use Naïve Bayes consider variables as individual they can use combination of algorithms like Naïve Bayes and K-means to get accuracy [3].

In the paper "A Model to Detect Heart Disease using Machine Learning Algorithm" author make a machine learning model was trained in order to determine if a person has a heart disease or not. This machine model uses a dataset which have 13 test outcomes directed on various people. Author uses four algorithms where used in checking the percentage of accurate results using different numbers of n values. For K Neighbors, the most noteworthy exact outcome is 97.47% roughly when n =1, for Support Vector Machine, the most elevated precise outcome is 98.83% when number of estimator =10, for Decision Tree, the most elevated precise outcome is 98.83% when number of n =1, for Random Forest, the most noteworthy exact outcome is 98.83% when number of estimator =10. After the testing of accuracy, they used Decision Tree Classifier which has one of the highest accurate results in making prediction [4].

In the paper 'Prediction of Heart Diseases uses Support Vector Machine', she uses the SVM Machine learning algorithm using R language to make a model for heart detection technique. According to the author support vector machine algorithm is best algorithm to predict the heart disease based on the given factors like sex, age, pulse rate etc, as the author is using R language to develop the model R is a bit difficult to pick up, especially since it doesn't follow the basic conventions which are followed by other common programming languages, where as python is simple and very interactive programming language to implement the model [5].

## III. PROPOSED WORK

Machine learning is a process of teaching machines, how to do any difficult or easy task by themselves. As we know that machine learning is the process of making machine think and perceive the things like human.

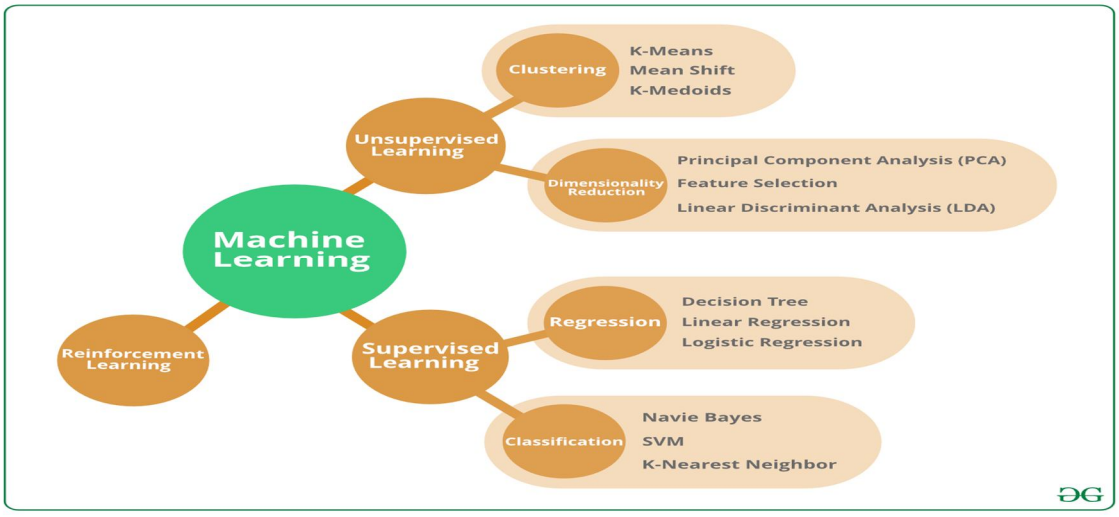



Figure 4: Different types of machine learning algorithms

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 8 Issue VI June 2020- Available at www.ijraset.com*

As we can see from above figure machine learning is mainly divided into three main categories these are:

1) *Supervised Learning*: Supervised learning is the process of teaching the machines using labels and tags, it is the process in which direct supervision is involved just like in a class of students teacher is directly teaches them on the basis of previous existing data. Supervised learning involve two process that are :

a) *Regression*: It is the process of recognizing the patterns and computing the guesses of continues results. The system has to understand the number, their values, grouping etc.

b) *Classification*: It is the process where input data is tagged based on past data specimen and manually instruct the algorithm to perceive specific types of objects and classify them appropriately.

2) *Unsupervised Learning*: Unsupervised learning is the process that does not involve direct supervision of the developer. Unlike supervised learning in which results are known, unsupervised machine learning algorithms does not know the desired results. And unlike supervised learning unsupervised learning is work on unlabeled data.

In this paper we are using four different supervised learning algorithms and check which algorithm gives the best result to detect whether a person have disease or not and then we make the model using that algorithm.

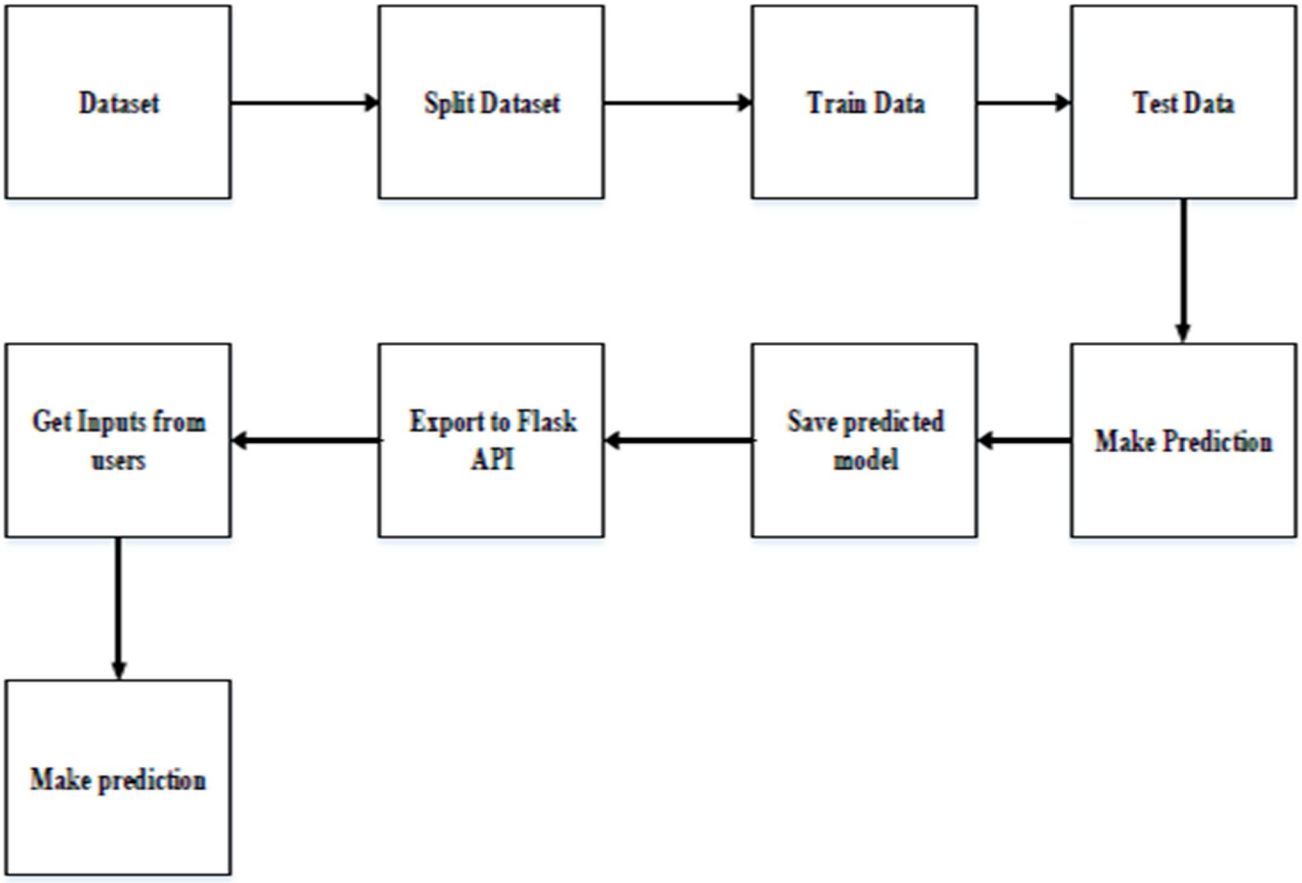



Figure 5: The architecture of the model for prediction

This system uses a dataset called patient dataset which is obtained from kaggle.com. This dataset contains fourteen different attributes of test results which are performed on 600 persons. The dataset is divided into two sets that are train set (80%) and a test sets (20%). After that, we trained our model using four different machine learning algorithms that are: K-Nearest Neighbours Support Vector Classifier, Decision Tree Classifier and Random Forest Classifier. Now after training, testing and checking the results of predictions and on the basis of the accuracy of the four different algorithms used in training the model, we will come to the result that by using which algorithm we will make the model to makes prediction. This trained model is saved and then loaded into web using flask, which is an Application Programming Interface (AP1). Flask loads this trained model and asks for the test results from the users. These results are being passed to the trained model to process the results and make prediction of a patient having disease or not.

## IV. DATA ANALYSIS AND RESULT

In this paper, an AI model was being prepared in order to decide whether a client has a coronary illness or not this model uses a set of data with thirteen different pathological test results conducted on different persons. This dataset is very clean and there are no null values present in the dataset. As we are using python programming we import train_test_split from sklearn.model_selection to split our data into x and y variables. This dataset was part into x and y factors. Where the x variable contains the 13 qualities which are the diverse test outcomes and the y variable contains the yield.

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| 1 | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| 2 | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| 3 | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| 4 | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |

Figure 6: Discription of the dataset with its all 14 attributes

When we analyze the data set we get the following results according to the different attributes like below picture shows the probability of happening of disease or not:

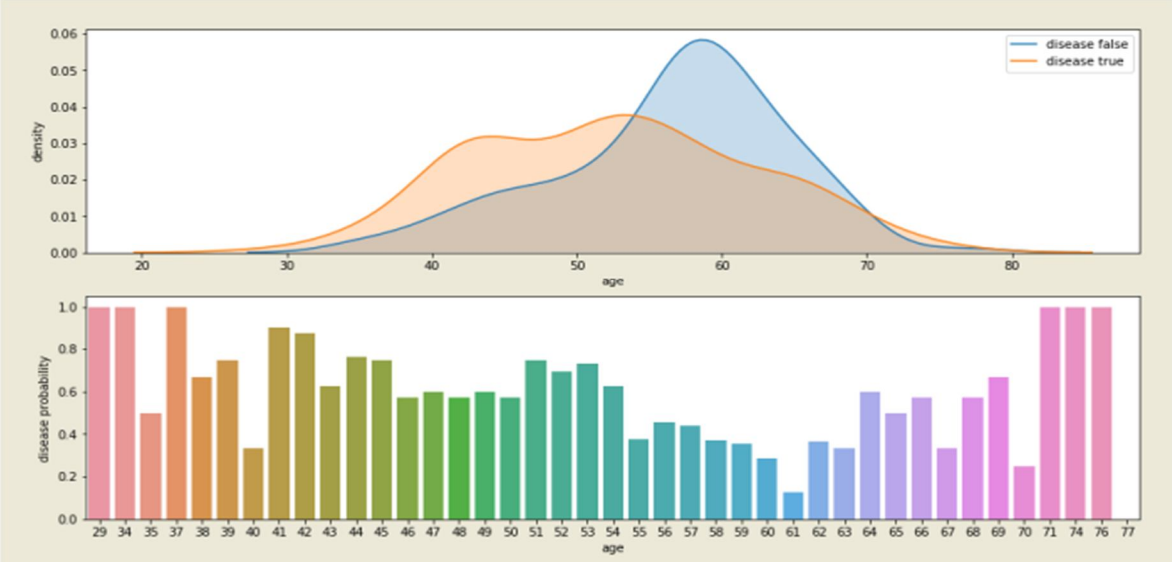



Figure 7: The probability of happening of disease according to the ages of different peoples

Similarly when we talk with respect to the sex and fasting blood sugar then we get the following results:

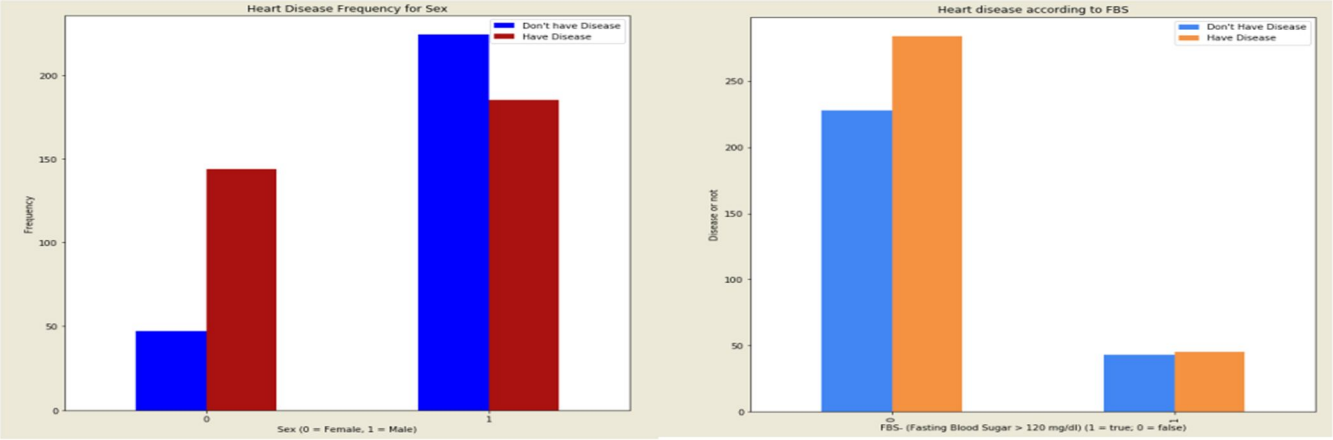



Figure 8: Disease according to the sex and fasting blood sugar of different peoples

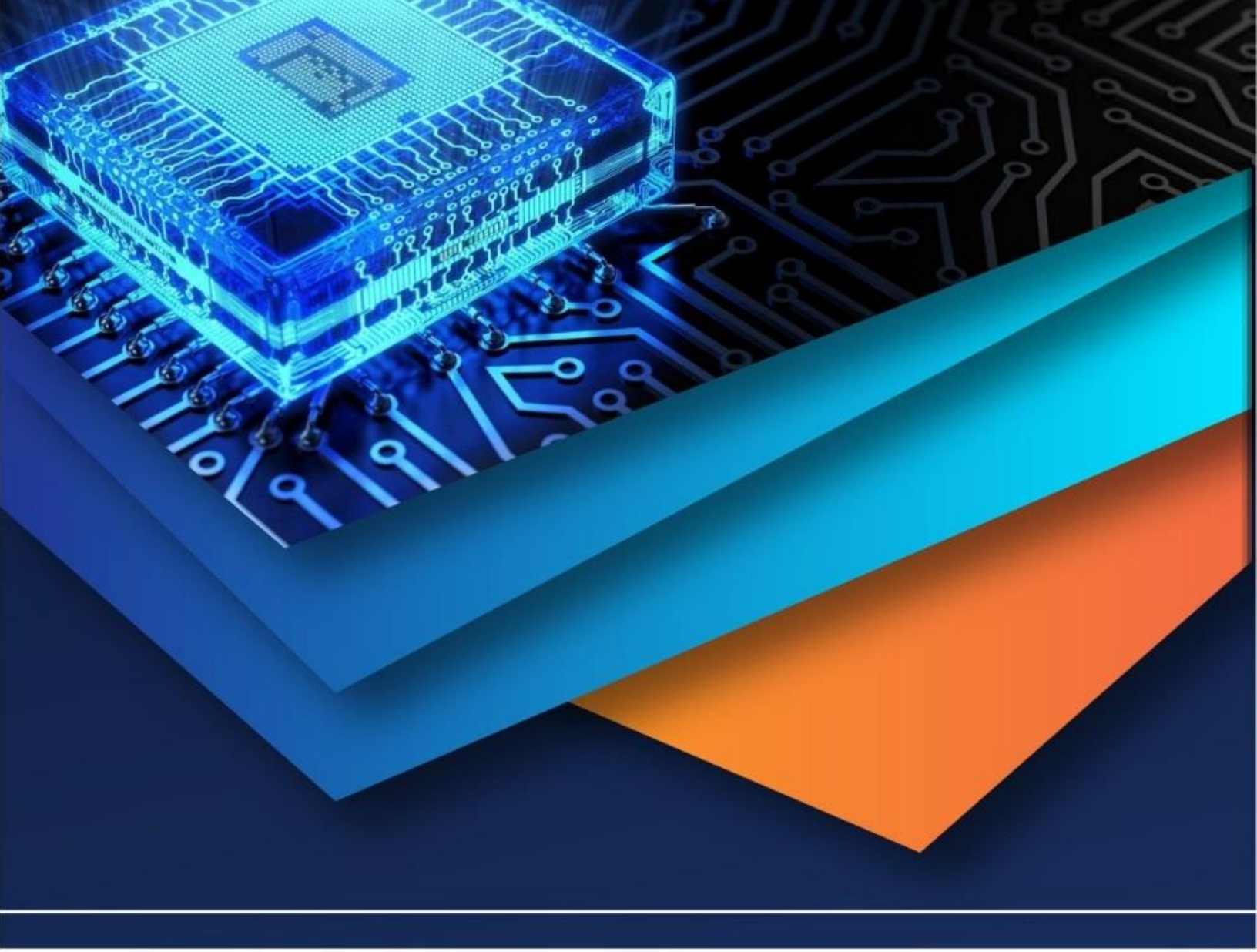
The x variable was being scaled utilizing StandardScaler. The x variable and y variable were additionally partitioned into x_train, x_test, y_train and y_test. These x_train and y_train where being fitted or prepared utilizing four machine calculations which are K-Nearest Neighbors, Support Vector Machine, Decision Tree and Random Forest. The four calculations where utilized in checking the level of precise outcomes utilizing various quantities of n esteems. For K Neighbors, the most noteworthy exact outcome is 84.49% roughly when k =12, for Support Vector Machine, the most noteworthy precise outcome is 85.32% when number of estimator =10, for Decision Tree, the most elevated exact outcome is 99.83% when number of n =10, for Random Forest, the most noteworthy exact outcome is 99.83% when number of estimator =20. After the testing of precision, we utilized Random forest which has one of the most elevated exact outcomes in making expectation. The Random forest was being spared and stacked into the web utilizing an Application Programming Interface called Flask. Utilizing Flask, we made a HTML page containing 13 contributions of which the clients will enter their diverse test outcome and pass the contributions to the model to identify in the event that they have a Heart Disease or not.

## V. CONCLUSION AND FUTURE SCOPE

This paper which disturbed on proposing "A model to Detect Cardiovascular Disease Using AI Algorithm" was created by utilizing an AI approach. In this AI approach four calculations were utilized to prepare and investigate the dataset which contains the test consequences of various patients and these calculations were likewise tried for exactness plotting a diagram utilizing matplotlib. Subsequent to testing for precision, Decision Tree and Random Forest have the most elevated exact outcome which is about 99.83% roughly while K Nearest Neighbors and Support Vector Machine have 84.49% and 85.32% respectively. Choice Tree model was likewise incorporated in the web through an API called Flask, and it anticipated great outcomes when tried multiple times on the web without a mistake. This exploration can be reached out to a constant framework utilizing Deep Learning approach, where clients can transfer their test results as picture.

## REFERENCES

[1] https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
[2] Nidhi Bhatla, Kiran Jyoti *An Analysis of Heart Disease Prediction using Different Data Mining Techniques* International Journal of Engineering Research & Technology (IJERT)(ISSN:2278-0181) Vol. 1 Issue 8, October – 2012
[3] Rajesh N, T Maneesha, Shaik Hafeez, Hari Krishna , Prediction *of heart disease using machine learning algorithms* International Journal of Computer Sciences and Engineering & Technology, 7 (2.32)(2018) 363-36
[4] O.E. Taylor, P. S. Ezekiel, F.B. Deedam-Okuchaba , *A Model to Detect Heart Disease using Machine Learning Algorithm* International Journal of Computer Sciences and Engineering (ISSN: 2347-2693) Volume 7 Issue 11 Nov 2019.
[5] Yamla Sandhya , "Prediction of Heart Diseases using Support Vector Machine" , International Journal for Research in Applied Science & Engineering Technology (IJRASET**)** (ISSN: 2321-9653) Volume 8 Issue II Feb 2020.
[6] https://villa-medica.com/wp-content/uploads/2017/07/villa-medica-cardiovascular-infographic-2-compressor.jpg



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY