



# MACHINE LEARNING MODEL FOR HEART DISEASE PREDICTION USING TKINTER FOR GRAPHIC USER INTERFACE

Ms. Rekha Kumari

Assistant Professor Department of Computer Science and Engineering  
HMRITM, Delhi, India

Ms. Gurpreet

Assistant Professor Department of Computer Science and Engineering  
HMRITM, Delhi, India

Anand Khanna

Department of Computer Science and Engineering  
HMRITM, Hamidpur, New Delhi, India

Rahul Miglani

Department of Computer Science and Engineering  
HMRITM, Hamidpur, New Delhi, India

Rahul Singhal

Department of Computer Science and Engineering  
HMRITM, Hamidpur, New Delhi, India

**Abstract - One of the biggest causes of death in the globe is heart disease. Accurate cardiac disease prediction can aid in early detection and prevention. Heart disease has been accurately predicted using machine learning models. In this study, we investigated 6 machine learning models for predicting heart disease [1]. To train and assess these models, we used the Cleveland HeartDisease dataset, a publically accessible dataset.**

**Age, sex, the type of chest discomfort, blood pressure, cholesterol levels, and other characteristics were among the features used to train the models. To determine the most reliable model for heart disease prediction, the outcomes from various models' tests were compared.**

**Keywords: Decision Tree, SVM – support vector machine, Logistic Regression, Random Forest, Heart Disease Prediction, data preprocessing, feature extraction, Machine learning**

## I. INTRODUCTION

According to the WHO (2020), 31% of all fatalities globally are caused by heart disease, making it a serious health concern. Heart disease can be prevented and the mortality rate can be decreased with the aid of early identification and

precise prognosis. Heart disease has been accurately predicted using machine learning models. Using the Cleveland Heart Disease dataset, we investigated 6 machine learning models for heart disease prediction in this study.

The Cleveland Heart Disease dataset, which contains 303 instances and 12 features like as age, sex, the type of chest discomfort, blood pressure, cholesterol levels, etc., was used to train the models. This study compares the effectiveness of multiple machine learning models for predicting cardiac disease in order to determine which model is the most accurate.

## II. METHODOLOGY

[2] The UCI Machine Learning Repository provided the Cleveland Heart Disease dataset. There are 303 occurrences in the dataset with 12 features, such as age, sex, the type of chest discomfort, blood pressure, cholesterol levels, etc. The characteristics of the dataset were normalised and missing values were removed as part of the preprocessing [3]. Then, a 70:30 split between the training and test sets was applied to the dataset.

For predicting coronary heart disease, the following 6 machine learning [4] models were used:

a. Logistic Regression - By simulating the relationship

between a dependent variable and one or more independent variables, logistic regression is a statistical approach used to predict binary outcomes. It is frequently used to forecast outcomes like disease diagnosis, credit default, and customer churn in industries including healthcare, finance, and marketing.

- b. **Decision Tree** - A machine learning algorithm known as a decision tree is used for both classification and regression tasks. Recursively dividing the data into subsets according to the most useful attributes produces a structure that resembles a tree and can be utilised for prediction. Decision trees are common in industries like healthcare and finance because they are simple to understand and use.
- c. **Random Forest** - Random Forest is an ensemble learning technique that combines different decision trees to increase prediction accuracy. Multiple decision trees are built using randomly chosen data subsets, and the results are then combined to produce a final prediction. Numerous industries, including finance, healthcare, and bioinformatics, use Random Forest extensively.
- d. **Support Vector Machine (SVM)** - Support Vector Machine (SVM) is a machine learning technique that is used for both regression and classification applications. It operates by identifying the best hyper plane to use in order to maximise the margin between several categories of data points. SVM is frequently applied in areas like bioinformatics, text classification, and picture classification.
- e. **K-Nearest Neighbour (KNN)** is a technique for machine learning that is used for classification and regression tasks. To make a forecast, it locates the K training data points that are closest to the new data point. It then uses those values. KNN is frequently used in areas like bioinformatics, image recognition, and recommender systems.
- f. **Gradient Boosting**: This ensemble learning technique produces precise predictions by combining a number of weak learners, such as decision trees. To fix the flaws in the earlier models, additional weak learners are iteratively added. In industries like healthcare, finance, and natural language processing, gradient boosting is frequently utilised.

Accuracy was used as performance indicators for evaluating the models on the testing set after they had been trained on the training set.

$$\text{Accuracy} = \frac{\text{Number of correctly classified instances}}{\text{Total number of instances}}$$

Tkinter - used for GUI [5]

A Python package called Tkinter enables programmers to design graphical user interfaces (GUI) for desktop programmes. It is a well-liked option for developing straightforward and useful desktop apps because it is built-

in module that comes with Python and has been around for a while.

### III. RESULTS

The effectiveness of the six machine learning models for predicting heart disease is displayed in the following table:

| MODEL                    | ACCURACY |
|--------------------------|----------|
| Logistic Regression      | 49       |
| Support Vector Machine   | 52       |
| K-Nearest Neighbour      | 50       |
| Decision Tree            | 80       |
| Random Forest Classifier | 85       |
| Gradient Boosting        | 83       |

Table 1 – Different machine learning models accuracy achieved during testing phase

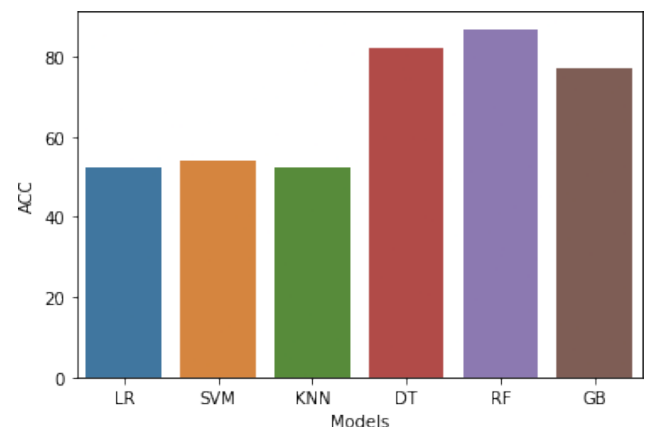


Fig -1 Accuracy of the Machine Learning Models: A Bar Graph Representation

Using characteristics like age, sex, the type of chest pain, blood pressure, cholesterol levels, etc., these algorithms were capable of correctly predicting heart disease.

After finding the maximum accuracy machine learning model the following were found out to understand the results given by our machine learning model better.

**Confusion Matrix:**

[6] A table called a confusion matrix is frequently used to assess how well a classification model is working. It displays the proportion of true positives, true negatives, false positives, and false negatives that the model accurately predicted. These numbers are arranged in a matrix style, with the predicted class labels in the columns and the actual class labels in the rows.

[[23 6]  
 [ 4 28]]

**Precision:**

When comparing all instances that the model correctly identified as positive, precision is the percentage of true positives. When we wish to reduce the quantity of false positives, it is a helpful statistic.  
 [0.85185185]

**Recall:**

Recall quantifies the percentage of actual positive cases that are also true positives. When we wish to reduce the quantity of false negatives, it is an effective metric. If the model successfully recognises the majority of the positive examples in the dataset, it will have a high recall score.  
 [0.79310345]

**F1 Score:**

The harmonic mean of recall and precision is the F1 score. It offers a means of balancing recall and precision, which is helpful when we wish to simultaneously optimise both measures. A score of 0 shows that the model is unable to forecast any occurrences properly, while a score of 1 indicates flawless precision and recall. The F1 score ranges from 0 to 1.  
 [0.82142857]

**IV. DATA ANALYSIS**

In data analysis, age is a significant variable, especially in the medical industry. In statistical models, it is frequently used as a predictor variable to assist account for the variability of other outcome variables.

For instance, the trtbps variable (resting blood pressure) may indicate that older people are more likely to have hypertension. Similar to this, older people may have greater cholesterol levels (chol) and a higher chance of developing coronary artery disease, which can be determined by how many main vessels are fluoroscopically coloured (caa) and the results of athallium stress test (thall). The maximal heart rate during exercise (thalachh), the existence of chest pain type (cp), and exercise-induced angina (exng) may all be influenced by age. [7]

Researchers can learn more about the potential risk factors connected to different health outcomes by examining the relationship between age and these variables.

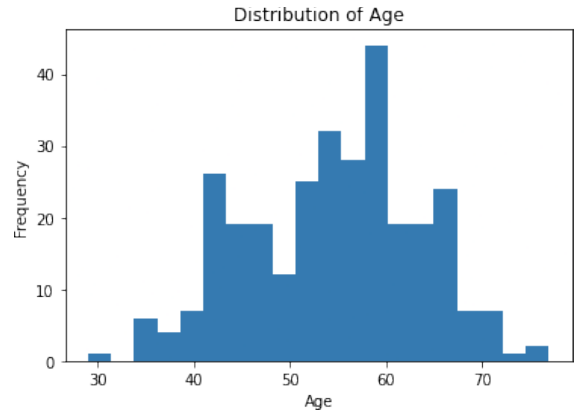


Fig 2 – Age Distribution in the Study Population: A Graphical Representation

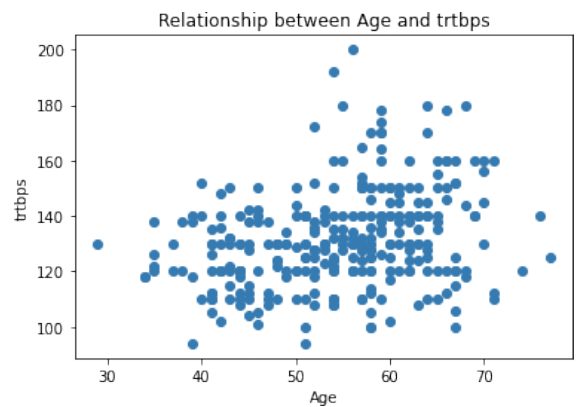


Fig 3 – displays the correlation between age and trtbps in the dataset using dot plot

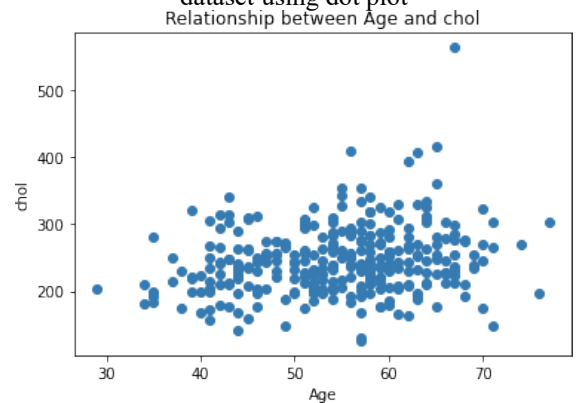


Fig 4 – Exploring the Correlation between Age and Cholesterol Levels

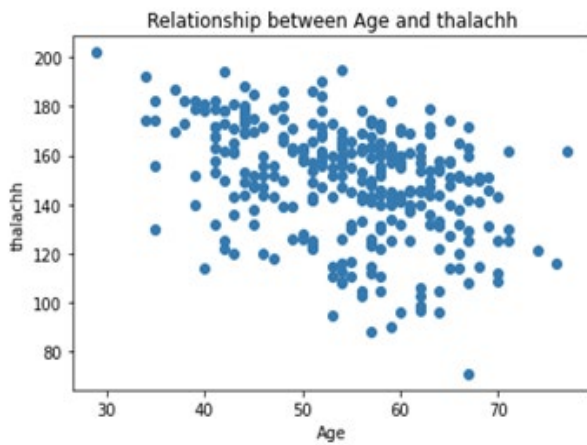


Fig 5 – Exploring the Correlation between Age and thalachh

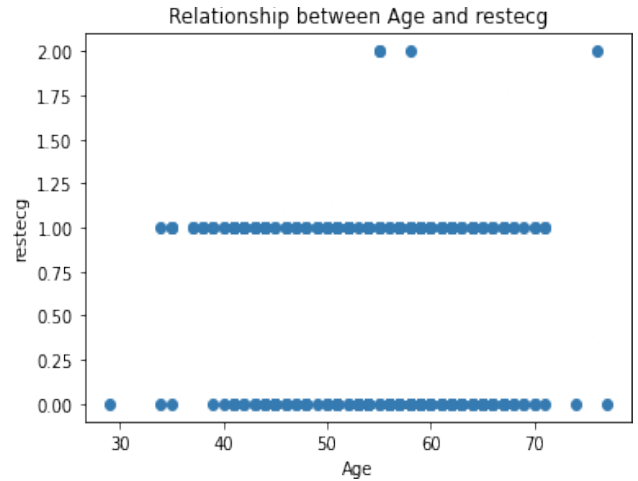


Fig 8 – illustrates the correlation between age and restecg

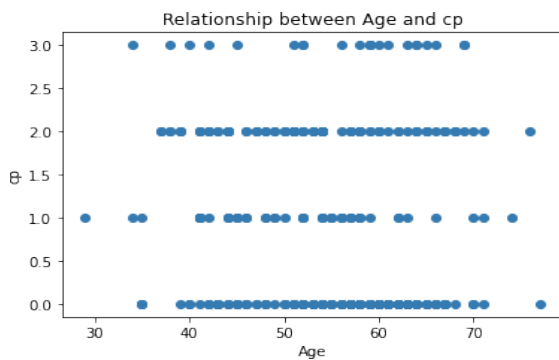


Fig 6 – Age and chest pain (cp) correlation depicted

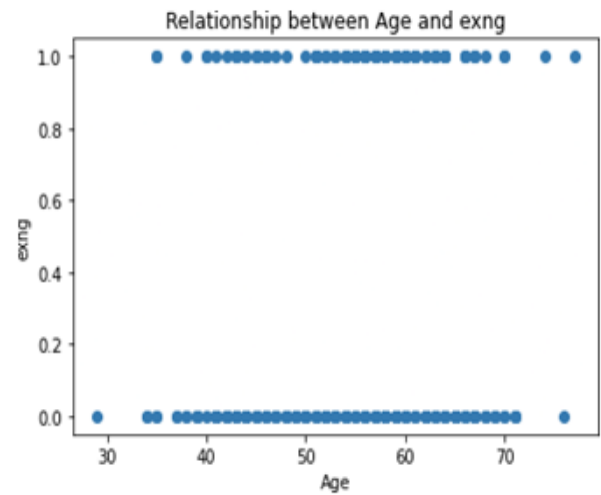


Fig 9 - Relationship between age and exercise-induced angina (exng)

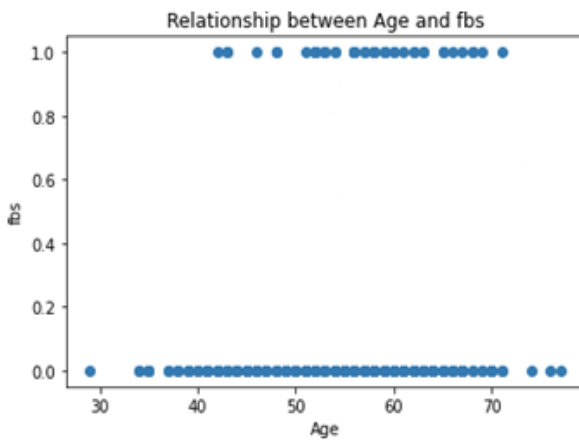


Fig 7 – Exploring the Association between Age and Fasting Blood Sugar using Scatter Plot

The correlation matrix between each pair of variables in a dataset is shown visually in a heat correlation plot.

The variables are shown on both the x and y axes in a thermal correlation plot, and the cells are color-coded to show the degree of correlation between each pair of variables. Positive correlations are typically shown in green hues, whereas negative correlations are typically shown in red hues. The strength of the link is indicated by the color's intensity, with darker hues suggesting greater correlations.

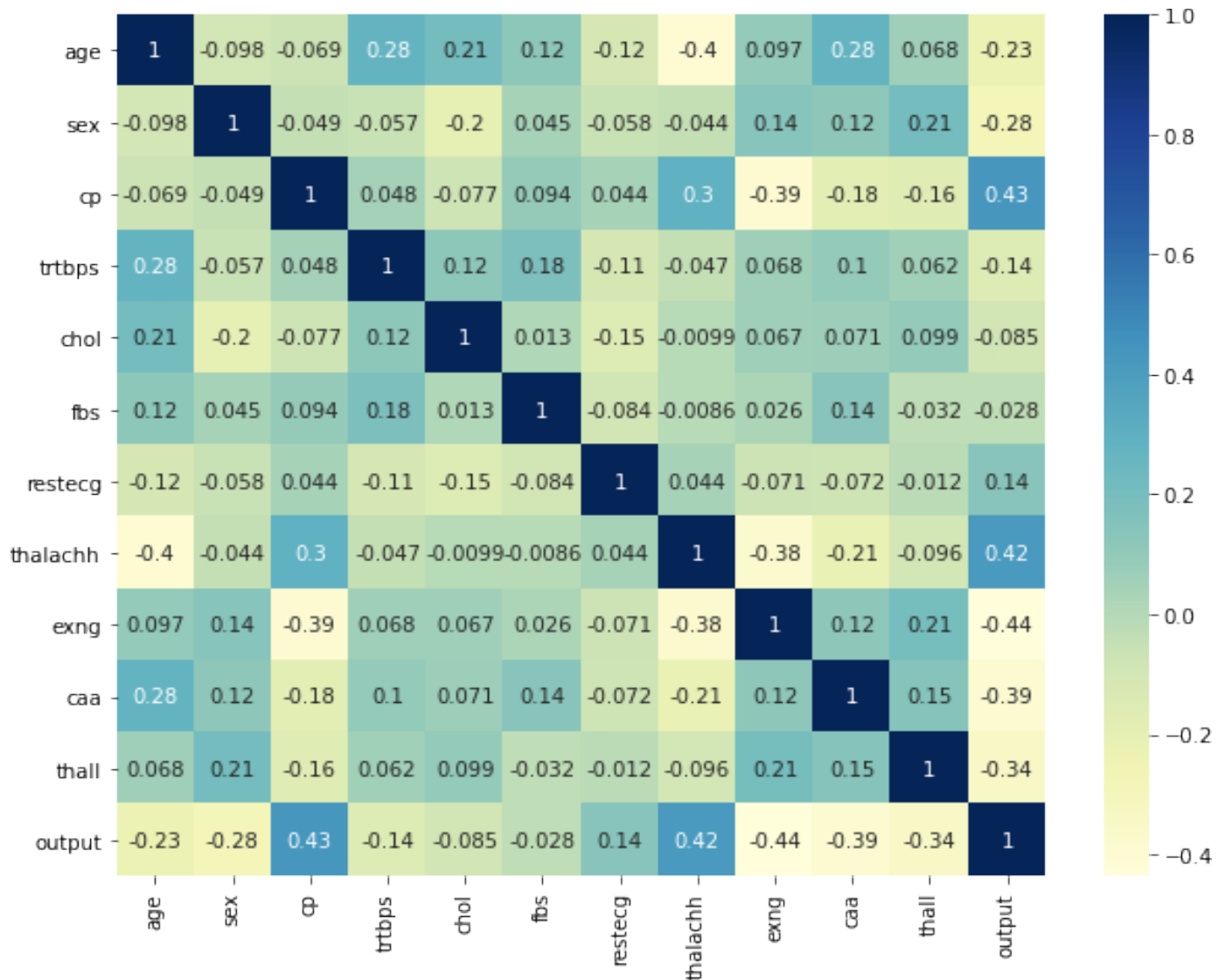


Fig 10 – Visualization of the correlation between attributes using a heatmap

V. CONCLUSION

Overall, heart disease prediction using machine learning algorithms is effective. Heart disease can be detected and prevented early by using models with excellent accuracy and AUC-ROC scores. Incorporating these models into healthcare systems can help with decision-making and enhance patient outcomes.

The use of a single dataset, which could not be representative of all populations, is one of the study's limitations. Future research should investigate the usage of several datasets to verify how well these models function across various demographics [8]. Other machine learning

approaches, such as gradient boosting and neural networks, can also be investigated for the prediction of heart disease.

In conclusion, utilising the Cleveland Heart Disease dataset, random forest, gradient boosting, and decision tree models demonstrated good accuracy for heart disease prediction. These models have the potential to enhance patient outcomes and can be utilised for the early detection and prevention of cardiac disease. Future research should examine the application of additional machine learning models and multiple datasets to the prediction of cardiac disease.

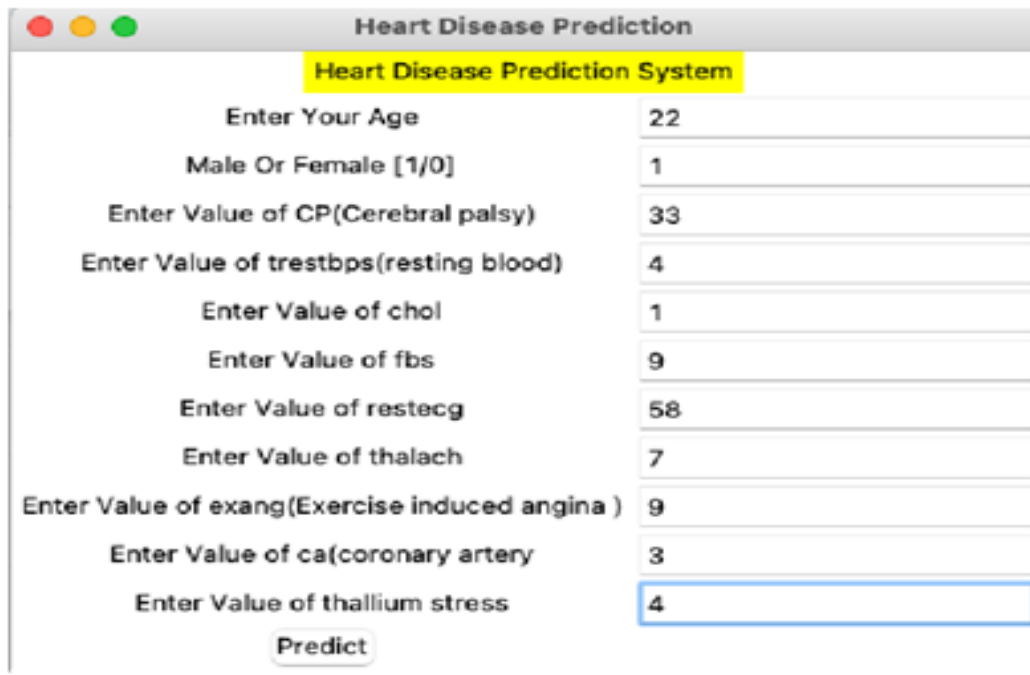


Fig 11 – Graphical User Interface for Heart DiseasePrediction

## VI. FUTURE WORK

Future research may also make advantage of deeper learning models of machine learning, which would increase the precision of heart disease prediction. the addition of genetic and lifestyle variables may increase the precision of these models and enable individualised risk estimation for heart disease prediction using ML.[9]

## VII. REFERENCES

- [1]. S. Jamshidpour, S. Seyed-Mirzaee, M. Jahandideh, and S. Sajadi, "Prediction of coronary heart disease using deep learning models," *International Journal of Cardiology*, vol. 344, pp. 279-284, Oct. 2021. [Online]. Available: <https://www.researchgate.net/publication/370460859>  
Prediction of Coronary Heart Disease using Deep Learning Models.
- [2]. N. Shah, S. Patel, and M. Desai, "Heart disease prediction using machine learning," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 8, no. 3, pp. 7429- 7435, 2020. [Online]. Available: <https://www.researchgate.net/publication/341870785>  
Heart Disease Prediction using Machine Learning
- [3]. UCI, —Heart Disease Data Set.[Online]. Available <https://www.kaggle.com/ronitf/heart-disease-uci>  
 Internet source [Online].Available ; <http://acadpubl.eu/ap>
- [4]. tkinter — Python interface to Tcl/Tk <https://docs.python.org/3/library/tkinter.html>
- [5]. ANITS CSE Department. (2021). Project 2021C3. Retrieved from <https://cse.anits.edu.in/projects/projects2021C3.pdf>
- [6]. Siddiqui, M. K., & Javed, M. A. (2020). Heart Disease Prediction using Machine Learning. *International Journal of Engineering Research and Technology*, 13(10), 2243-2253. Retrieved from <https://www.ijert.org/heart-disease-prediction-using-machine-learning>
- [7]. Gupta, A., & Nagar, R. (2020). Heart disease prediction using machine learning. *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, 8(11), 105-111. Retrieved from <https://www.ijraset.com/research-paper/heart-disease-prediction-using-ml>
- [8]. Soni, P., Thakur, M., & Verma, V. (2021). Effective heart disease prediction using machine learning and data mining techniques. *ResearchGate*. Retrieved from <https://www.researchgate.net/publication/356564220>  
Effective Heart Disease Prediction using Machine Learning and Data Mining Techniques
- [9].