



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 12, December 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 7.569

 9940 572 462

 6381 907 438

 [ijirset@gmail.com](mailto:ijirset@gmail.com)

 [www.ijirset.com](http://www.ijirset.com)



# Analysis of Heart Disease Prediction using Machine Learning Techniques

Tulika Lodh<sup>1</sup>, Anirban Dey<sup>2</sup>, Naorem Rinita<sup>3</sup>, Sunil Kumar<sup>4</sup>, Subodh Kumar<sup>5</sup>

Assistant Professor, Department of Computer Science, Biswanath College, Assam, India<sup>1</sup>

UG Student, Department of ECE, NERIST, Arunachal Pradesh, India<sup>2</sup>

UG Student, Department of ECE, NERIST, Arunachal Pradesh, India<sup>3</sup>

PG Student, Department of SC&SS, JNU, New Delhi, India<sup>4</sup>

PG Student, Department SC&SS, JNU, New Delhi, India<sup>5</sup>

**ABSTRACT:** Heart, the muscular organ that pumps blood in the entire body of any living organism needs to be immune from diseases. Heart disease is the condition in which the structure or the function of the heart is affected. It has become the number one killer of men and women in the U.S. In this project, we have designed a machine learning model for predicting the accuracy of heart diseases of a group of individuals depending on various parameters. This model is designed using different machine learning algorithm such as Support Vector Machine, Logistic Regression, Random Forest, Decision Tree, K-neighbour.

**KEYWORDS:** Classification, Heart Disease, Logistic Regression, Decision Trees, SVM, KNN, Random Forest.

## I. INTRODUCTION

Heart, being the core of the human life, we should as well take proper care of it so as to keep it healthy. Coronary heart disease, one of the most common heart diseases, occurs to any individual when the arteries that supplies blood to the heart narrows or hardens due to the build-up of plaque. Heart diseases has become a burning topic these days as due to a number of different factors, affecting the heart of living organism. Strict actions should be taken in order to overcome such problems and also maintain a good diet.

According to the Centers for Disease Control and Prevention (CDC), heart disease has been the leading cause of death in the United States. Around 1 in 4 deaths trusted Source in the U.S. occur due to heart disease, and the condition affects all genders as well as all racial and ethnic groups[6]. Heart disease has also become so common in India as well. 52 percent of all cardiovascular deaths in India occur before the age of 70, and according to the Times of India, heart attacks kill 1 person every 33 seconds in the country. This growing epidemic has even earned India the reputation of being the coronary capital of the world. It's time we take our heart health into our own hands and progress to become a healthier people[7].

Thus, in this project, a comparison of how much accurate a machine learning algorithm can be, is depicted using a dataset on the basis of various parameters such as sex-male, age, current smoker cigarettes per day, Blood Pressure Medicines, prevalent stroke, prevalent hyp, diabetes, total cholesterol, system blood pressure, diaBP, BMI, HeartRate, glucose, ten year Coronary Heart Disease. The different machine learning algorithms that are used in this model in order to find the prediction of a patient that whether that patient will have 10 years risks of future coronary heart disease are support vector machine, linear regression, logistic regression, decision tree, K-neighbour.

## II. LITERATURE SURVEY

Dhai Eddine SalhiUsing, Abdelkamel A Kamel Tari,[1] "Machine Learning for Heart Disease Prediction" in this paper, researcher have proposed Machine Learning Mode such as Neural Network, SVM, KNN where they have found accuracy of 93% in Neural Network.

Keshav Srivastava, Dilip Kumar Choubey,[2] "Heart Disease Prediction using Machine Learning and Data Mining" in this paper, author obtained from UCI repository and applied pre-processing to drop the data with missing values and



applied some algorithm like Decision tree, K-Nearest Neighbour, Support Vector Machines and, Random Forest of which Decision Tree gives an accuracy of 79%.

Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, Parneet Singh, [3]”Prediction of heart disease using a combination of machine learning and deep learning”, in this project, authors developed a model which uses combination machine learning as well as deep learning algorithms in order to predict heart diseases so that it can be easily prevented and be prepared as well.

Harshit Jindal, Sarthak Agarwal, Rishabh Khara, Rachna Jain, Preeti Nagrath, [4]”Heart Disease prediction using machine learning algorithm” this project mainly focuses on the prediction of people going to have heart diseases in the coming time. This is done by using different machine learning algorithms such as Logistic regression, KNN, etc. it predicts and classifies the patients with heart diseases.

### III. METHODOLOGY

#### A. Data Collection

Data was extracted from Kaggle. The data set contains set of people who have test and analysis to detect heart disease.

#### B. Manual Exploration

In this step where we analysis the dataset to each person as sick or not. Here in this algorithm the training dataset, we form the dataset.

#### C. Data Pre-processing

In this section we have use correlation matrix to detect the attributes and those attributes are highly dependent on each other in machine learning. After that for training we have taken 80% of the dataset while the remaining dataset is used for testing.

#### D. Logistic regression

Logistic regression is a type of regression analysis in statistics used for prediction of outcome of a categorical dependent variable from a set of predictor or independent variables. In logistic regression the dependent variable is always binary. Logistic regression is mainly used to for prediction and also calculating the probability of success.

Mathematically,  
$$P = \frac{e^{\beta_0 + \beta_1 X_1}}{1 + e^{\beta_0 + \beta_1 X_1}}$$

When all features plugged in:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 * \text{Sexmale} + \beta_2 * \text{age} + \beta_3 * \text{cigsPerDay} + \beta_4 * \text{totChol} + \beta_5 * \text{sysBP} + \beta_6 * \text{glucose}$$

#### E. KNN

In this case Euclidian function  $\text{dis}(x_i, x_j)$  is used to extract the information out of the samples by using the majority vote of the k-nearest neighbours.  $\text{Dis}(x_i, x_j) = \sqrt{(x_{i,1} - x_{j,1})^2 + \dots + (x_{i,m} - x_{j,m})^2}$ .

#### F. SVM

Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection. We have used linear kernel function i.e  $(x, x')$ . Linear kernel is used when the data is linearly separable.

#### G. Decision Tree

The decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller subsets with an increase in depth of the tree. The final result is a tree with decision nodes and leaf nodes. The entropy can be calculated by using:

#### H. Random Forest

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.



**I. Modelling**

In this phase where we apply and test the chosen algorithm (Linear regression, SVM, KNN, Random Forest, Decision Tree) to find the best among them. Our approach is based on the application of 5 algorithm on the dataset to find the accuracy of each one and which one is more stable and gives highly accurate.

**IV. EXPERIMENTAL RESULTS**

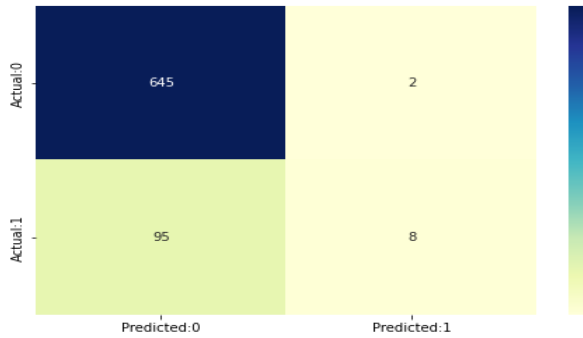


Fig1: Confusion Matrix

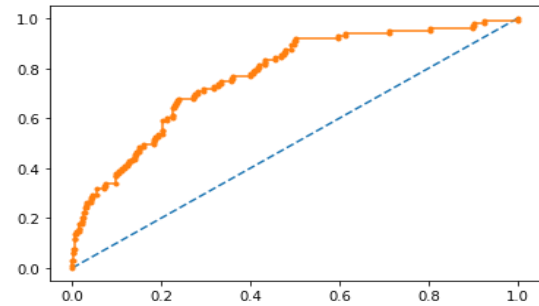


Fig2: ROC of Logistic Regression

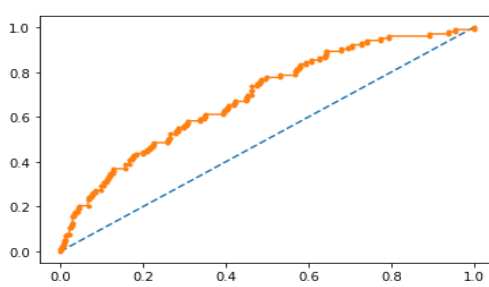


Fig 3: ROC of SVM

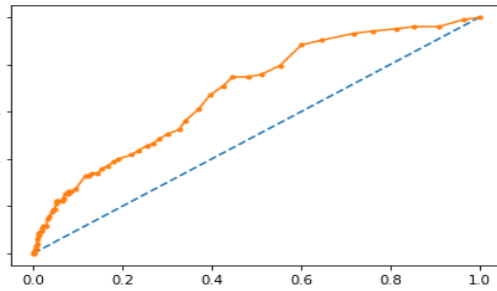


Fig4: ROC of Random Forest

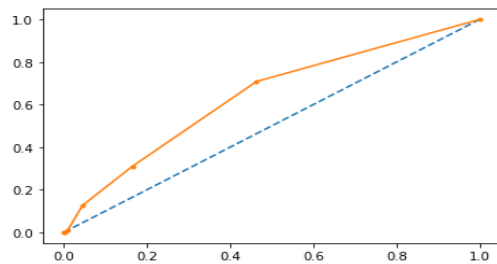


Fig5: ROC of KNN

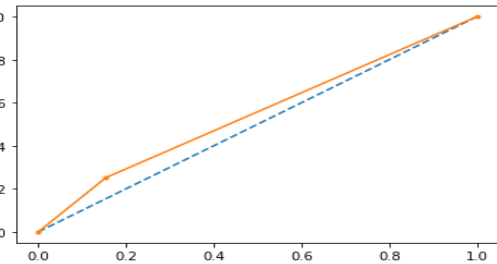


Fig6: ROC of Decision Tree

The confusion matrix shows 645+4 = 649 correct predictions and 95+1= 96 incorrect ones. As shown in Fig1

**True Positives: 4**

**True Negatives: 645**

**False Positives: 1 (Type I error)**

**False Negatives: 95 (Type II error)**

The ROC curve for Linear Regression, SVM, Random Forest, KNN, Decision Tree gives an AUC of 77.3%, 69.1%, 68.8%, 63.9% and 54.9% as shown in Fig. 2, 3, 4, 5 and 6 respectively.

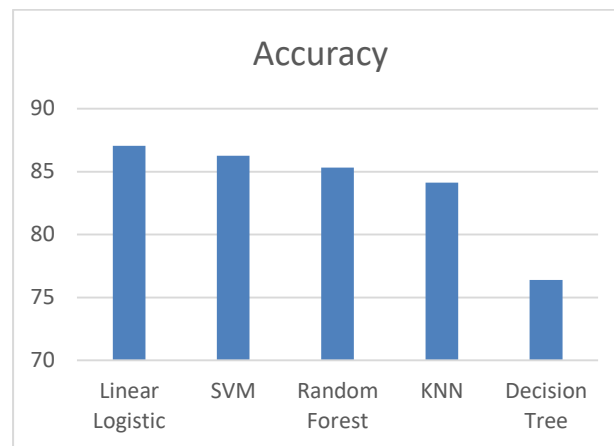


Fig7: Accuracy of different algorithm

The final result that we have obtained after applying different machine learning algorithms for predicting the heart diseases of a group of individuals shows different accuracy as shown in Fig7. Logistic shows accuracy of 87.06% which is highest out of all the algorithms used, followed by SVM showing an accuracy of 86.26% while Random Forest shows an accuracy of 85.13% and KNN shows an accuracy of 84.13%. The Decision Tree shows the lowest accuracy of all the algorithms used so far in this project giving an accuracy of 76.4%.

## V. CONCLUSION

Thus, from the above discussion, it can be concluded that out of the five different machine learning algorithms that we have used so far, logistic regression gives the highest accuracy followed by support vector machine in finding the prediction of heart diseases of the individuals and a comparison is made using different graphs such as correlation, area under a curve, bar diagram in order to find the algorithm which shows most accurate result.

## REFERENCES

- [1] Dhai Eddine SalhiUsing, Abdelkamel A Kamel Tari “Machine Learning for Heart Disease Prediction” ResearchGate, February 2021, In book: Advances in Computing Systems and Applications (pp.70-81).
- [2] Keshav Srivastava, Dilip Kumar Choubey “Heart Disease Prediction using Machine Learning and Data Mining”, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-9 Issue-1, May 2020.
- [3] Rohit Bharti , Aditya Khamparia , Mohammad Shabaz , Gaurav Dhiman , Sagar Pande ,and Parneet Singh, “Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning”, Hindawi Computational Intelligence and Neuroscience Volume 2021, Article ID 8387680, 11 pages.
- [4] Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain and Preeti Nagrath,” Heart disease prediction using machine learning algorithms”, IOP Conf. Series: Materials Science and Engineering, ICCRDA 2020, OP Publishing.
- [5] Dataset: <https://www.kaggle.com/amanajmera1/framingham-heart-study-dataset/data>
- [6] <https://www.medicalnewstoday.com/articles/237191>
- [7] <https://www.abbott.in/corpnewsroom/healthy-heart/indian-heart-disease-101.html>





**Impact Factor:  
7.569**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details