**Heart Safety using Supervised Learning Algorithm**

Dr.V. Pandimurugan 1 Kushagra Bajpai *2* Rashmi Rawat 3 Aryan Dagore 4 Devansh Rana

*1, Assistant Professor, Department of Computing Science and engineering, VIT Bhopal University, Madhya Pradesh*

*2,3, Students, Department of Computing Science and engineering, VIT Bhopal University, Madhya Pradesh*

**Abstract**

The exponential growth of data from different sources such as wearable sensor devices used in the Internet of Things health monitoring, streaming system, and others have been generating an enormous amount of data continuously. The collaboration of data interpretation and machine learning can have an important influence on the healthcare field, especially the early detection of heart disease. CVD prediction model researches a specific lifestyle and genetic background. Elements that form the bedrock of the dataset include age, height, gender, blood pressure, took to smoke, or perform a simple workout. Dataset brought about by putting data together from diagnosing and monitoring an additional fifty thousand (50,000) individuals.

1. **Introduction**

Cardiovascular disease (CVD) is the leading cause of death in India, both across the rural and urban populations. Centre for Disease Control and Prevention identifies heart disease, chronic obstructive pulmonary disease, and stroke as the top three killers [4]. To diagnose heart-related problems, a General Practioner (GP) asks about medical and family history, checks blood pressure, and a blood test to assess cholesterol levels. Before having the cholesterol test, you may be asked not to eat for 12 hours, so no food in your body that could affect the result. Other tests include X-rays, electrocardiogram (ECD), MRI scans, and CT scans. These tests are not cost-efficient, and the results take some time to produce. Also, carrying these tests such as X-rays is unhealthy [5].

The goal of our project is to bring out the odds of contracting CVD with a harmless and cost-effective process. Applying the CVD prediction model, we can obtain an anticipated percentage of contracting Cardiovascular disease. The outcome differs upon the difference in the lifestyle of the person [6].

1. **Literature Review**

Throughout the year's heart disease prediction model has been a difficult challenge to humans. Many peoples have been working on this situation. Various data mining techniques have been used for diagnosis and have achieved different success results for different methods. Different methods have been applied due to the selection of different p7arameters this further results in indifference accuracies. Paper [7][8][9] produced by the fellow author has continuously supported the working generation with further research and helping us to understand certain aspects of the prediction model. Few works process is described below:

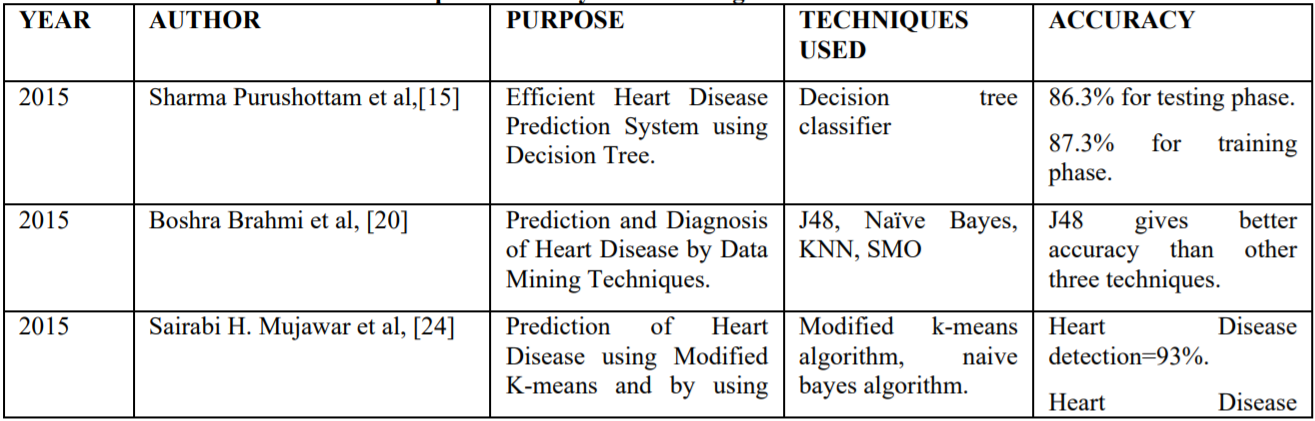


Fig 1: Previous Work by some Author

1. **Proposed Framework**

With the help of Machine Learning, we can predict whether the person is suffering from Cardiovascular(heart) disease or not [2][3].

* 1. **Gathering data**

The patient database is collected through various websites such as Kaggle, UCI Repository. Our database consists of 13 attributes that are: age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, target.

Data Contains;

* age – Age in years
* sex – (1= male; 0 = female)
* cp – chest pain type
* trestbps – resting bood pressure (in mmHg on admission to the hospital)
* chol – serum cholestoral in mg/dl
* fbs – (fasting blood sugar > 120 mg/dl) (1= true; 0 = false)
* restecg – resting electrocardiographic results
* thalach – maximm heart rate achieved
* exang – exercise induced angina (1 = yes; 0 = no)
* oldpeak – ST depression induced by exercise relative to rest
* slope – the slope of the peak exercise ST segment
* ca – number of major vessels (0 - 3) colores by flourosopy
* thal – 3 = normal; 6 – fixed defect; 7 = reversable defect
* target – have diseases or not (1 = yes, 0 = no)
  1. **Data Pre-processing**

Data preparation, where we load our data into an appropriate place and prepare it for use in our machine learning training. It includes Data Visualization through certain graphs and plots and splitting our data into testing as well as in the training model. In this phase, we transform data through missing fields, normalization of data.

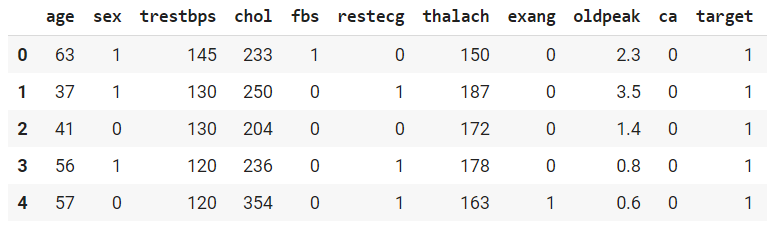


Fig 2: Data Distribution

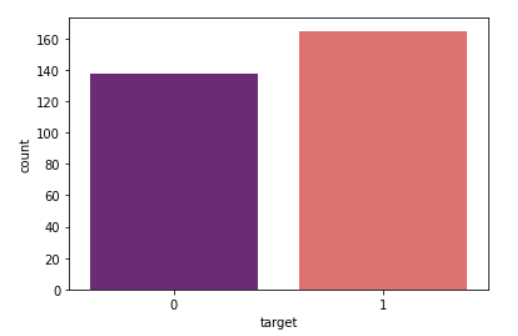


Fig 3: Target Variable

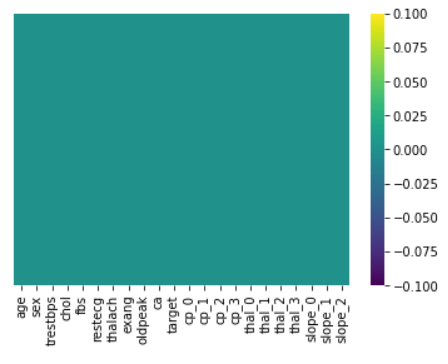


Fig 4: Data Wrangling

* 1. **Training**

In this step, we will use our data to improve our model’s ability to predict whether a person is suffering from Heart disease or not. We have used the fit () method for the training of our model. It is done through the following steps by finding consequential attributes by computing their discrete P-values.

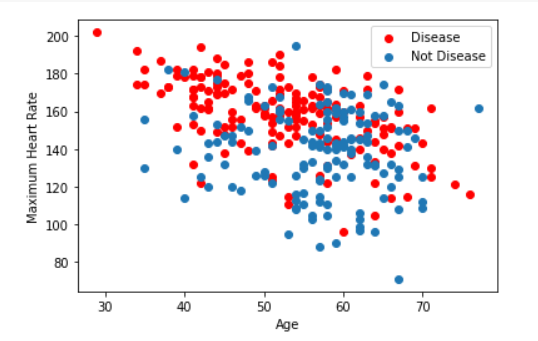
****

Fig 5: Checking Diseases

* 1. **Evaluation**

Evaluation [1] support us to examine our model against data that has never been utilized for training. This is how we will represent the working of our model in the real world. Through logistic regression, we predict the results of the categorical dependent variable. The dependent variable is always in two categories. This regression analysis is used for the prediction of success probability.

* 1. **Hyperparameter Tuning**

After the evaluation of our model, we can improve our model's performance by tuning the parameters. We have used grid-search and Random-search parameters tuning for accessing more good parameters from the training dataset which will increase the accuracy of our predictions.

* 1. **Prediction**

At last, after hyperparameter tuning, we have made our predictions on testing data by using method predict (). From predictions, we can examine whether a person is suffering from Cardiovascular (Heart) disease or not.

Final Result-

And finally, the accuracy we have achieved

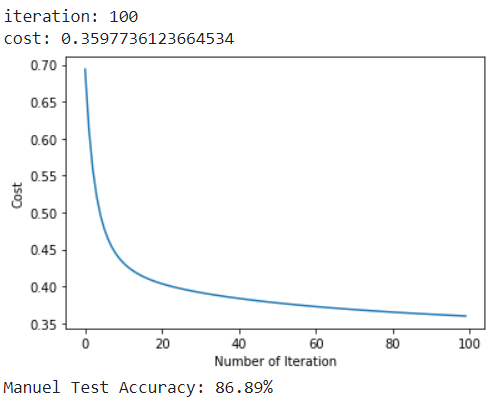


Fig 6: Accuracy of the model

This accuracy shows that e.g. out of 100 predictions we made, 85 0f them are suffering from Cardiovascular (Heart) disease.

1. **Conclusion**

In conclusion, we can say that deep learning methods produce better results as compared to other supervised learning methods [10]. Though classification is also an important issue. A combination of different prediction models might be more accurate in the prediction of the early symptoms of cardiovascular diseases. Cardio diseases are complex and death due to these diseases increases every year. The fundamental motive of these predicting models is to achieve a high accuracy rate in heart disease prediction. The future prediction algorithm model should be based on the basis of less time complexity.

1. **References**

[1] T. M., Mukherji D., Padalia N., Naidu A., “A Heart Disease Prediction Model using SVM-Decision Trees-Logistic Regression (SDL)”, International Journal of Computer Applications, Vol 68, Issue 16, 2013.

[2] Saw M., Saxena T., Kaithwas S., Yadav R., Lal N., “Estimation of prediction for getting heart disease using logistic regression model of machine learning”, International Conference on Computer Communication and Informatics, 2020.

[3] Prasad R., Anjali P., Adil S., Deepa N., “Heart disease prediction using logistic regression algorithm using machine learning”, International Journal of Engineering and Advanced Technology, Vol 8, Issue 3S, 2019.

[4] Avinash Golande, Pavan Kumar T., “Heart disease prediction using effective machine learning techniques”, International Journal of Recent Technology and Engineering, Vol 8, Issue 1S4, 2019.

[5] V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja, “Heart disease prediction using machine learning techniques: a survey”, International Journal of Engineering & Technology, 2018.

[6] Ashwini Shetty A, Chandra Naik, “Different Data Mining Approaches for Predicting Heart Disease”, International Journal of Innovative Research in Science, Engineering and Technology, Vol 5, Spe Issue 9, 2016.

[7] Purushottam, Saxena k, Sharma R, “Efficient heart disease prediction system using decision tree”, International Conference on Computing, Communication and Automation, 2015.

[8] Bahrami, Boshra, Bahrami, Boshra, “Prediction and Diagnosis of Heart Disease by Data Mining Techniques”, Journal of Multidisciplinary Engineering Science and Technology, Vol 2, Issue 2, 2015.

[9] Mujawar, Sairabi, Devale, H. P. R., “Prediction of Heart Disease using Modified K-means and by using Naive Bayes”, International Journal of Innovative Research in Computer and Communication Engineering, Vol 3, Issue 11, 2015.

[10] Dr. K. Usha Rani, “Analysis Of Heart Diseases Dataset Using Neural Network Approach”, International Journal of Data Mining & Knowledge Management Process, Vol 1, Issue 5, 2011.