*Research Result*

# Machine Learning for Heart Disease Prediction

**Sonal Shakya[1], Prof. Jayesh Jain[2]**

[1,2]*Global Nature Care Sangathans Group of Institutions*

**ABSTRACT**

*In this work, machine learning and Python programming are used to explore the prediction of heart disease. Heart disease has become a widespread and dangerous condition as a result of fat deficiency throughout time. The body's high pressure causes this sickness. Several dataset features can be used to forecast cardiac disease. To assess patient performance, we used a dataset with 12 parameters and 70000 unique data values. To increase the accuracy of heart disease diagnosis, the main goal of this research is to develop algorithms that can tell whether or not a person has the condition.*

**KEYWORDS**

*Heart Disease, symptoms, Machine Learning, Prediction, Health*

## 1. INTRODUCTION

The major topic is prediction using machine learning. These days, e-commerce is just one of the numerous business applications where machine learning is frequently used. Our research focuses on the prediction of heart disease using patient datasets and a dataset of patients for whom we must make predictions about the risk that heart disease may develop. One application of this machine learning is in prediction. In this project, the most widely used programming language, Python, is used together with a machine learning model and features a number of libraries. The machine learning subset model of an artificial intelligence network uses deep learning neural networks and complex algorithms. The human body is made up of numerous different organs, each of which has a distinct function. If one of these organs, the heart, fails to pump blood throughout the body, dangerous conditions may arise in the human body. One of the leading causes of death nowadays is heart disease. Because of this, it is crucial to preserve the health of our cardiovascular system and all other bodily systems. Sadly, cardiovascular issues are a problem for people all around the world. Any technology that can detect these ailments early on and minimize damage would prove useful in preserving people's finances and, ultimately, their lives. Data mining approaches can be useful for predicting heart problems. Analytical models can be produced by exploring databases for previously unrecognized patterns and trends and using the generated data. Before serious harm is done, technology like machine learning can help in the identification of cardiac disease. Machine learning, a relatively new field of science and technology, may determine whether or not a person is at risk for heart disease.

## 2. LITERATURE REVIEW

- The Naive Bayes Approach, Artificial Intelligence (AI) networks, Supports Vector Machine (SVM), Random Forest technique, and Simple Regression Method (Shadman Nashif) for the prediction of heart disease.
- For the Heart Disease Prediction System-HDPS, the paper presented and it is equated with KNN, SVM, Random classifier, and decision tree classifier had delivered an exact result. By using this, decision tree machine learning approach enhanced prediction accuracy to 98.85% when compared to other methods (O.E. Taylor).
- More accurate data mining methods for forecasting cardiac illness (Animesh Hazra). The accuracy of heart disease is increased in this work by comparing c4.5, Kmeans, decision trees, SVM, naive Bayes, and all other machine learning techniques.
- Instead, Praveen Kumar Reddy, 2019, will use a decision tree algorithm to try and lower the incidence of heart disease. In this, the decision tree is applied using the Gini index approach, where the highest gain of the attributes produces an enhanced representation of the decision tree algorithm, and the Support Vector Machine algorithm categorises the data values using hyperplane.

 

- Disease Predicting system utilising data mining approaches was the title of a study presented by M.A.Nishara Banu and B.Gomathy. MAFIA (Maximum Frequent Itemset Algorithm) and K-Means clustering were studied in this work. Because classification is important for disease prediction. The classification is based on the imprecision of the MAFIA and K-Means values.
- A study titled The Intelligence System for the Diagnosis Level of Coronary Heart Disease with K-Star Algorithm had been presented by both Wiharto and Hari Kusnanto. In this study, they use the Learning Vector Quantification Neural System Calculation to establish the expectation for heart infection. This structure's neural network recognises 13 clinical with several execution measurements, information, it forecasts whether the patient has near- or absent-absence of heart disease.

## 3. METHODOLOGY AND ANALYSIS

- **Data collection**

The general process of predicting heart disease conveys the following method:

1) Cardiovascular Disease Dataset by Svetlana Ulianova, 2019. The dataset contains 70,000 records of patient data made up of 11 features and the dataset is the information or tool that is important for any type of study or project.

- **Data Preprocessing**

1) Separation of target data and feature data as training and test data.

Scrambling the values in the data to be valued between 0 and 1 and scale all the values before training the Machine Learning models.

- **Applying Algorithms**

1) Equating 4-machine learning algorithms such as SVM, Decision tree, Random Forest classifier, and K-nearest neighbor to get the improved accurateness to which highest parameter may cause disease.

2) For every algorithm, there is a pseudo code helpful to grow any kind of programming language. In python, there is a simple way to create any kind of algorithm in which simple and shortcode is easier to predict accuracy.

## 4. MACHINE LEARNING ALGORITHM

The algorithms used in this project are highly helpful to predict the correct result to detect heart disease in which factors that cause disease can be detected. The following algorithms have been assembled in this project.

1) K-Nearest Neighbor Algorithm:

KNN is a managed classifier that carries out an observation from within a test set to predict classification labels. KNN is one of the classification methods used whenever there is a classification. It has a several assumptions includes the dataset has slight noise, is labeled and it should covers relevant features. By applying KNN in huge datasets takes a long time to process. The accuracy grown with this algorithm is 63.4%.

2) Random Forest Classier:

A random forest classifier is a powerful tool in the machine learning library. With this classifier, we will be able to increase accuracy, and training time should be a smaller amount. Primarily, we have to build a model and by splitting variables into training and test set. After that, train the dependent variables and predict the answer. By using the random forest classifier, the accuracy predicted result is approximately 71% but actually it is 71.4%.

3) Decision tree classier:

In this algorithm, pre-processing is initially created by dividing the data into training and test data. Feature scaling can be done by normalizing the values before the forecast. Import decision tree classification into training sets of determined and independent variables using Guinea-index criteria to predict the accuracy or response of a test set. The accuracy obtained with this algorithm is 68.4%.

4) Support Vector Machine (SVM):

SVM is also one of the classification algorithms in machine learning in which improved accuracy can be predicted. As compare to other algorithms, it is much better for expectedly predicting accuracy.

In our prediction, the predicted highest accuracy is 72.6% using linear SVM kernel.

In our prediction, the predicted highest accuracy is 86.2% using the Gaussian SVM kernel.

## 5. PROPOSED SYSTEM

By this experiment whatever we say is as Naïve Bayes results and decision tree results may change so for each prediction, we do not have a comparison of both the algorithms to get precise results and in the same way if we use only a single algorithm which cannot pre-process data, we even can't get good accuracy so it's better to have a combination of algorithms like k-means, ID3 and k-means, and Naïve Bayes.

After confirming that the data is balanced, the connection between the data is found out and is plotted as a heat map using the Seaborn library.

The heat map clearly displays that the attributes like cp

(chest pain) and thalack (maximum heart rate achieved) have a positive connection with the target attribute. Now that the connection has been checked, we need to convert categorical variables like sex, cp, FBS, restecg, exang, slope, ca, and thal into dummy variables. This can be done by using the get_dummies method of the Pandas library. After creating dummy variables, the data in columns like age, trestbps, chol, thali, and old peak needs to be standard scaled, because they have much varied quantities and units. This can be done using the Scikit-learn library in Python.

## 6. CONCLUSIONS

We found that decision trees usually lead us to the wrong conclusion when working with little datasets in other contexts. We are able to get more precise findings with probability for all other options when working with Naive Bayes results, though. Decision trees can occasionally fall short, though, because they frequently point us in the direction of a single solution. Finally, this experiment shows that Nave Bayes is more accurate even with clean and well-maintained input data, even if both ID3 and Nave Bayes are accurate and can clean themselves after each successful result. won't result in anything. We must always consider the results of different algorithms, and it would be more precise if these results could be predicted. To attain accuracy, we can use Naive Bayes to consider variables as individuals or mix Naive Bayes and K-means. After experimenting with various algorithms, it can be said that machine learning has been shown to be very helpful in anticipating heart disease, which is one of modern society's biggest problems. As machine learning research develops, fresh strategies for maximizing its utility in the healthcare sector may soon be created. The algorithms used in this experiment have shown good performance when using the specified attributes. Finally, it may be said that machine learning could decrease the harm done to a person's physical and mental health by predicting heart disease.

## REFERENCES

[1] Patel, J., Tejal Upadhyay, D., and Patel, S. (2015). Predicting heart condition using machine learning and data mining techniques. Heart disease, 7 (1), 129-137.

[2] Ramalingam, V. V., Dandapath, A., & Raja, M. K. (2018). Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology, 7(2.8), 684- 687.

[3] Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Intelligent and effective heart disease prediction system using weighted associative classifiers. International Journal on Computer Science and Engineering, 3(6), 2385-2392.

[4] Amudhavel, J., Padmapriya, S., Nandhini, R., Kavipriya, G., Dhavachelvan, P., Venkatachalapathy, V.S.K., "Recursive ant colony optimization routing in wireless mesh network", (2016) Advances in Intelligent Systems and Computing, 381, pp. 341-351.

[5] Alapatt, B.P., Kavitha, A., Amudhavel, J., "A novel encryption algorithm for an end-to-end secured fiber optic communication", (2017) International Journal of Pure and Applied Mathematics, 117 (19 Special Issue), pp. 269-275.

[6] Amudhavel, J., Inbavalli, P., Bhuvaneswari, B., Anandaraj, B., Vengattaraman, T., Premkumar, K., "An effective analysis on harmony search optimization approaches", (2015) International Journal of Applied Engineering Research, 10 (3), pp. 2035-2038.

[7] Amudhavel, J., Kathavate, P., Reddy, L.S.S., Bhuvaneswari Aadharshini, A., "Assessment on authentication mechanisms in a distributed system: A case study", (2017) Journal of Advanced Research in Dynamical and Control Systems, 9 (Special Issue 12), pp. 1437-1448.

[8] Amudhavel, J., Kodeeshwari, C., Premkumar, K., Jaiganesh, S., Rajaguru, D., Vengattatraman, T., Haripriya, R., "Comprehensive analysis on information dissemination protocols in vehicular ad hoc networks", (2015) International Journal of Applied Engineering Research, 10 (3), pp. 2058-2061.

[9] Amudhavel, J., Kathavate, P., Reddy, L.S.S., Satyanarayana, K.V.V., "Effects, challenges, opportunities and analysis on security-based cloud resource virtualization", (2017) Journal of Advanced Research in Dynamical and Control Systems, 9 (Special Issue 12), pp. 1458-1463.

[10] Svetlana Ulianova, Cardiovascular Disease dataset. The dataset consists of 70 000 records of patient data, 11 features + target.

[11] Kaur, B., & Singh, W. (2014). Review on heart disease prediction system using data mining techniques. International journal on recent and innovation trends in computing and communication, 2(10), 3003-3008.