

Heart Disease Prediction: A Machine Learning Approach for Risk Assessment

Mr. Karan Kale, Mr. Rohit More, Mr. Anant Sadawarte, Mr. Sanket Ghogare,
Mr. Shushant Sulakhe, Prof. Rajni Patel

SVRI's College of Engineering, Pandharpur, Maharashtra, India

Abstract: Objective: Heart disease is the leading cause of mortality globally. A variety of modern technologies are utilized to treat cardiac disease. As one of the most significant areas for prediction, the clinical data analysis covers a highly important condition, namely cardiovascular disease. It is the most prevalent problem in medical centers since many medical personnel do not have equal knowledge and skill to treat their patients, therefore they make their own decisions, which result in bad outcomes and occasionally death. To address these issues, heart disease prediction is being performed utilizing machine learning methods like logistic regression.

Cases of heart illness are developing at an alarming rate, and it is critical and important to predict any such ailments in advance. This is a challenging process that must be completed accurately and effectively. The study report primarily examines which patients are more likely to develop heart disease based on several medical characteristics. Using the patient's medical history, we developed a heart disease prediction system that predicts whether the patient is likely to be diagnosed with a heart disease or not. To predict and identify the patient with heart disease, we applied several machine learning methods such as logistic regression. To limit how the model may be utilized, a very helpful technique was adopted.

Some data mining and machine learning approaches used to anticipate cardiac illness (SVM) include Artificial Neural Network (ANN), Random Forest, and Support Vector Machine. Heart disease prediction and diagnosis has become a daunting undertaking for doctors and hospitals both in India and overseas. To reduce the staggering number of deaths from heart disorders, a rapid and effective detection method must be developed. In this discipline, data mining techniques and machine learning algorithms are particularly important. The researchers are stepping up their attempts to develop software that uses machine learning algorithms to help doctors forecast and diagnose heart illness. The major purpose of this study is to predict a patient's heart disease utilizing.

Keywords: Logistic Regression, Python Programming, Correlation matrix, Stream lit, Classification, Regression

I. INTRODUCTION

The Heart disease is a prevalent and life-threatening condition that demands accurate and efficient diagnostic methods. Predictive models based on machine learning algorithms, such as logistic regression, have gained significant attention for their potential in early detection and risk assessment. Logistic regression offers a transparent and interpretable model that can effectively predict the likelihood of heart disease based on various clinical and physiological factors. This study aims to explore the effectiveness of logistic regression as a predictive model for heart disease detection. By analyzing a comprehensive dataset encompassing relevant risk factors, including age, gender, blood pressure, cholesterol levels, and other medical indicators, we investigate the performance and interpretability of logistic regression in accurately identifying individuals at risk of heart disease. The insights gained from this research will contribute to improving diagnostic accuracy, enabling timely interventions, and ultimately reducing the burden of heart disease on global health.

1.1 What is Machine Learning?

Machine learning is an artificial intelligence application that employs statistical approaches to allow computers to learn and make judgements without being explicitly programmed. It is based on the idea that computers can learn from data, recognize patterns, and make decisions with little help from humans.

1.2 Approaches to Machine Learning:

1. Feature Selection Techniques: Researchers have used a variety of feature selection strategies to improve the accuracy of heart disease prediction models. These strategies seek to discover the best predictive factors.
2. Support Vector Machines (SVM): Due to its capacity to handle high-dimensional data and nonlinear correlations, SVM has gained favor in heart disease prediction.
3. Integration of Multi-Modal Data: Efforts have been made to integrate multi-modal data for cardiac disease prediction, including clinical, genetic, and imaging data.
4. Obstacles and Future Directions: Despite advances in cardiac disease prediction, obstacles remain. The absence of standardized datasets, the requirement for interpretability in complicated models, and the integration of predictive models into clinical processes are all topics that need to be investigated further. Future research should concentrate on overcoming these obstacles in order to promote the wider use of heart disease prediction algorithms in clinical practice.

1.3 Advantages of Machine Learning

1. Automation and Efficiency: Machine learning enables the automation of operations and processes that would otherwise require substantial manual effort. It can swiftly and correctly analyse large volumes of data and execute complicated computations, resulting in greater efficiency and production.
2. Pattern Recognition and Insights: Machine learning algorithms can detect patterns, trends, and correlations in massive datasets that people may miss. This capacity to glean insights from data may assist organizations and researchers in making educated decisions, uncovering hidden patterns, and gaining a greater knowledge of complicated events.

II. OBJECTIVE AND SCOPE

The primary goal of this study is to create a heart prediction system. A historical cardiac data set may be used to detect and extract hidden illness knowledge. The goal of the heart disease prediction system is to use data mining techniques on medical data sets to aid in the prediction of heart illnesses.

The project's scope is that integrating clinical decision assistance with computer-based patient records might reduce medical mistakes, improve patient safety, eliminate unnecessary practice variation, and improve patient outcomes.

This notion is intriguing since data modelling and analysis technologies, such as data mining, have the potential to create a knowledge-rich environment that can considerably enhance the quality of healthcare judgements.

The main objective of developing this project are:

- Using Logistic Regression, create a machine learning model to predict the likelihood of future heart disease.2.
- To identify major risk factors for heart disease based on a medical dataset.3.
- Analyze feature selection techniques and comprehend their operation

III. LITERATURE SURVEY

There In medical centers, multiple efforts have been done on illness prediction systems employing various data mining approaches and machine learning algorithms.

Prediction of Heart Disease Using Logistics Regression is acceptable for forecasting the likelihood of heart disease. The task is carried out using the previously specified training data set, which consists of 303 cases with 14 distinct properties. The data set is separated into two sections, with 70% used for training and 30% used fortesting. Based on the data, it is obvious that the Logistic Regression method outperforms the other algorithms in terms of classification accuracy.

Data mining techniques are being used to forecast heart disease. Data mining is now used to forecast a variety of diseases. The number of tests can be lowered by applying data mining techniques. This research focuses on predicting heart disease, diabetes, and breast cancer, among other things.

Due to the rising costs of heart disease diagnostics, there was a need to design a new system that could anticipate heart illness. The prediction model is used to anticipate the patient's condition after evaluating numerous factors such as heart rate, blood pressure, cholesterol, and so on.

Traditional Risk Factors

Several research have focused on traditional risk factors related with the prediction of heart disease. Age, gender, hypertension, diabetes, smoking status, family history, and cholesterol levels are among these variables. To analyse the influence of these variables on heart disease risk, researchers used logistic regression models and statistical methodologies

Support Vector Machines (SVM):

Due to its capacity to handle high-dimensional data and nonlinear correlations, SVM has gained favour in heart disease prediction. Based on data derived from medical records, SVM models were used to categories individuals with and without cardiac disease.

IV. METHOD

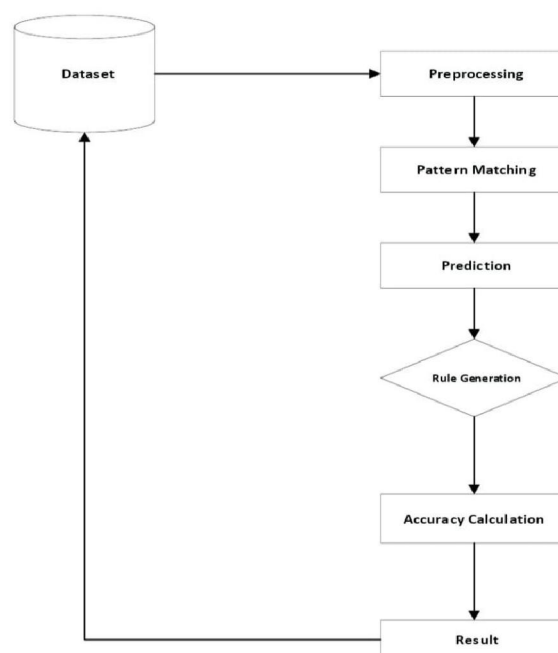
Logistic Regression

A supervised classification algorithm is logistic Regression. It is a predictive analytic method based on the probability notion. It estimates probabilities using an underlying logistic function (sigmoid function) to assess the association between the dependent variable (Tenured) and one or more independent variables .

Logistic Regression is heavily reliant on data presentation. As a result, in order to make the model more powerful, significant characteristics from the existing data set are chosen utilizing Backward elimination and recursive elimination procedures.

Typical Heart Disease Prediction Workflow

Data Flow Diagram



V. CONCLUSION

Heart disease is becoming increasingly common among individuals, including in our nation. Predicting the disease before being infected thereby reduces the chance of mortality. This forecast has received a lot of attention.

Our study is part of a larger investigation into the detection and prediction of heart disease.

It is based on the use of Machine Learning methods, Logistic Regression, and a data collection of experienced persons, where we achieved 93% accuracy with Neural Network. The strength of our study was that we examined the stability of the algorithm on different sizes of our data set and discovered that Neural Network produces the best results. In addition, we conducted research on feature selection.

REFERENCES

- [1]. SePatel S & Chauhan Y (2014). Heart attack detection and medical attention using motion sensing device - Kinect. International Journal of Scientific and Research Publications, 4(1).
- [2]. Worthen W J, Evans S M, Winter S C & Balding D (2002). U.S. Patent No. 6,432, 124. Washington, DC: U.S. Patent and Trademark
- [3]. Soni J, Ansari U, Sharma D & Soni S (2011). Predictive data mining for medical diagnosis: an overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-8
- [4]. Parthiban, Latha and R Subramanian. "Intelligent heart disease prediction system using CANFIS and genetic algorithm." International Journal of Biological, Biomedical and Medical Sciences 3.3 (2008).
- [5]. K. Bhanot, "towar ddatascience.com," 13 Feb 2019. [Online]. Available: <https://towar ddatascience.com/predicting-presence-of-heart-diseases-using-machinelearning-36f00f3edb2c>. [Accessed 2 March 2020].
- [6]. M. A. K. S. H. K. M. a. V. P. M Marimuthu, "A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach".
- [7]. A. H. M. S. U. Marjia Sultana, "Analysis of Data Mining Techniques for Heart Disease Prediction," 2018