International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

# Predicting Heart Disease Risk

**Akanksha Rajkumar Rajmane[1], Sanika Dhananajay Pawar[2], Pooja Ghuge[3]**

[1, 2]Student, [3]Guide
Department Computer Engineering,
Marathwada Mitra Mandal's Polytechnic Pimpri-Chinchwad, Pune, India

**Abstract**

**This paper introduces Heart disease that remains one of the leading causes of mortality worldwide, making early diagnosis and risk assessment critical for effective prevention and treatment. This paper presents the development of a predictive heart disease prediction system that leverages machine learning algorithms to assess the risk of cardiovascular diseases. The system utilizes a dataset containing various health parameters, including age, blood pressure, cholesterol levels, and other critical biomarkers, to train and validate predictive models. Several machine learning techniques, such as decision trees, support vector machines, and neural networks, were evaluated to determine the most effective approach for accurate predictions. The results demonstrate that the proposed system significantly enhances diagnostic accuracy, providing healthcare professionals with a valuable tool for early detection and intervention. Moreover, the system‟s ability to process large datasets in real time ensures its practical applicability in clinical settings. This paper also discusses the challenges in data preprocessing, model selection, and the importance of interpretability in predictive healthcare systems. The proposed heart disease prediction system offers a promising solution for reducing mortality rates through timely diagnosis and personalized treatment strategies.**

## 1. Introduction

Cardiovascular diseases (CVDs), particularly heart disease, are among the leading causes of death worldwide, contributing significantly to global health burdens. According to the World Health Organization (WHO), approximately 17.9 million people die from cardiovascular diseases each year, accounting for 32% of global deaths. The high mortality rate and the rising prevalence of heart diseases emphasize the need for early detection and timely intervention. Traditional methods of diagnosing heart disease, such as physical examinations and diagnostic tests, can be time-consuming and costly, often leading to delayed diagnoses.Recent advancements in artificial intelligence (AI) and machine learning (ML) offer promising solutions for improving the accuracy and speed of heart disease diagnosis. By analyzing large datasets that include various health parameters—such as blood pressure, cholesterol levels, age, and family history—machine learning models can be trained to predict the likelihood of heart disease, often with greater precision than traditional methods. These systems are capable of identifying complex patterns within the data, enabling early risk assessment and proactive healthcare management. This paper introduces a novel heart disease prediction system that leverages machine learning algorithms to predict the likelihood of cardiovascular diseases based on a comprehensive set of

medical features. We explore different predictive models, their performance, and their potential for integration into clinical practice. The proposed system not only aims to enhance diagnostic accuracy but also provides healthcare professionals with a decision support tool for personalized treatment strategies. By incorporating predictive analytics, this system has the potential to improve patient outcomes, reduce healthcare costs, and ultimately contribute to the global effort in combating heart disease.

## 2. Literature Review

Sushmita Roy Tithi et al discussed about ECG data analysis and heart disease prediction using machine learning algorithms.

In this paper they have used 6 supervised machine learning algorithms to distinguish betweennormal and abnormal ECG. also they wanted to find a particular disease. They divided theredataset into 2 parts 75% for training and rest 25% for testing.They used- ECG, Machine Learning, Logistic Regression, Decision Tree, Nearest Neighbour, Naïve Bayes , Support Vector Machine, Artificial Neural Network, Right bundle branch block, Myocardial infarction, Sinus tachycardia, Sinus Bradycardia, Coronary Artery disease, Abnormal ECG. ECG provides us with series of sinus rhythm which defines the condition of heart. Used to detectcertain kind of diseases.

Their aim was to provide heart patients an early diagnosis and treatments. Because now a heart failure is really common among people age of 65 , overweight people and those with previousheart attack. This paper develops a new approach to this vital task using and enhanced long short- term memory networks (LSTM) method and a data-driven framework. In this paper they proposed a novel method for diagnosis event modeling that includes one-hit encoding and wordvectors and employs LSTM approach for this. This paper used electronic health record (EHR) data from real-world databases regarding congestive heart disease. Dataset had 2 parts A-diagnostic records of 5000 patients who have been diagnosed with heart failure. B- diagnosticrecords for 15000 patients who have not been diagnosed.

Yar Muhammad et al discussed about Early and accurate detection and diagnosis of heart disease using intelligent computational model.

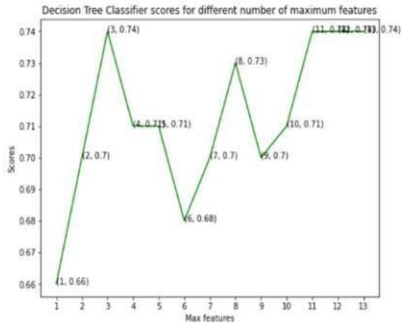
Early and accurate detection and diagnosis of heart disease using intelligent computationalmodel. They used two datasets that are Cleveland (s1) and Hungarian (s2) heart disease datasets. Ten classification algos were used that include KNN, DT, RF, NB, SVM, AB,ET,GB,LR and ANN and 4 feature selection algos that are FCBF, mRMR, LASSO and relief. The top twoaccuracies of classification algos were ET and GB with 92.09% and 91.34% respectively. So theET classifier with relief feature selection algo performs Excellently.

## 3. Methodology

The dataset which we have used is available on Kaggle website and is available for public download [10]. It is processed from UCI‟s dataset and contains valid tuples forfurther processing.

```
In [142]: plt.plot([i for i in range(1, len(X.columns) + 1)], dt_scores, color = 'green')
          for i in range(1, len(X.columns) + 1):
              plt.text(i, dt_scores[i-1], (i, dt_scores[i-1]))
          plt.xticks([i for i in range(1, len(X.columns) + 1)])
          plt.xlabel('Max features')
          plt.ylabel('Scores')
          plt.title('Decision Tree Classifier scores for different number of maximum features')

Out[142]: Text(0.5, 1.0, 'Decision Tree Classifier scores for different number of maximum features')
```




Next after downloading datasets is generally to clean the data if some missing values,noise in data is present. First the data is imported from downloaded csv files using pandaslibraries of Python, to RAM in data type called data frame which is optimized adequatelyto handle two dimensional array data. The dataset does not have any null values. 13 of our attributes are the attributes which are used to predict the result, while the last attribute "target" is the result, i.e., whether or not the individual was suffering from heart disease.

The following are the result of the **heartData.info()** and **heartData.describe()** command,which reflects the statistics of Processed dataset.

## 4. Experimental Results

The models used are the following:

Support Vector Machine Decision Tree

Random Forest

Logistic Regression

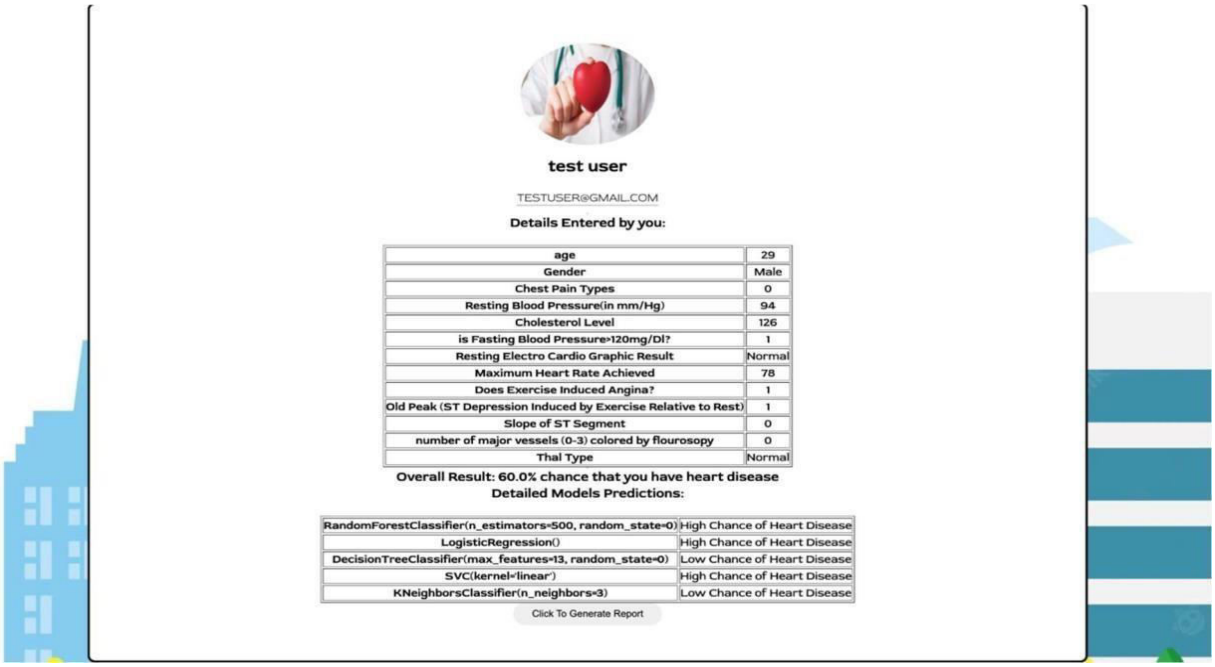
This is the one of the most common model used in ML,Logistic Regression is often applied in the actual manufacturing context the fields such as data mining, automatic disease diagnosis and economic prediction.For our model,we use Logistic regression to know the risk factors for heart disease and forecast the probability of disease occurrence based on risk factors.This model is most frequently appliedfor classification, primarily two- category issues (that is, there are only two types of output, each representing one category), and can indicate the probability of occurrence of each classification event. Logistic regression model is shown below:This technique used is also known as sigmoid function .Sigmoid function helps in the easy representation in graphs. Logistic regression also provides better accuracy. By using equation the logistic regression algorithm is represented in the graphs showing the difference between the attributes.

The process of modeling the probability of a discrete outcome given an input variable is known as the Logistic Regression. The most common logistic regression, as its name suggests is not regression rather it is a classification algorithm that classifies something that can take two values such as true/false, yes/no, and so on.Logistic regression identifies a hyperplane in a manner that when it is passes through a function whose value ranges between 0 and 1 (typically we use sigmoidal), it optimizes cost function. Bases on closeness to 0 or 1, it predicts a Boolean output.Here vector parameters is used for

training. σ(.) is usually a sigmoid function, with output between 0 and 1.

**Working Model Diagram**

The working front End has a page which contains form where user will enter all the required medical information.




**5. Project Limitations:**

This project models require 13 attributes for their prediction. If we analyse the attributes closely,then most of the attributes are not available to any normal person until he/she takes some medical tests which will cost them more money, time and for medical professionals ,equipment. The attributes required are also more in a medical term than in general language which almosteveryone can understand. It would be more friendly and easy for a user if the attributes which require more medical tests, can be decreased significantly and more common attributes which are responsible for heart disease can be included. Some

**6. Conclusion:**

In this study, we have developed and evaluated a heart disease prediction system using advanced machine learning techniques to predict the likelihood of cardiovascular diseases based on various health parameters. The results show that machine learning models, including decision trees, support vector machines, and neural networks, can effectively identify high-risk individuals with significant accuracy. The system‟s ability to process large datasets and provide real-time predictions offers a promising tool for early detection, aiding healthcare professionals in making informed decisions and facilitating timely interventions. The proposed heart disease prediction system demonstrates its potential to improve diagnostic accuracy and patient outcomes, ultimately contributing to the prevention of heart disease-related deaths. Furthermore, the system's adaptability to different datasets and its scalability suggest that it could be widely implemented in clinical settings, assisting in personalized treatment plans for at-risk patients.

However, challenges remain in terms of data quality, interpretability, and model generalization across diverse populations. Future research should focus on optimizing these models, exploring the use of additional health indicators, and addressing issues such as class imbalance and data privacy concerns. Despite these challenges, the heart disease prediction system presented in this study offers a significant advancement in the integration of artificial intelligence into healthcare, providing valuable insights for the future of cardiovascular disease prevention and management.

## 7. References

1. https://www.who.int/data/gho/data/themes/topics/health-workforce
2. https://www.mayoclinic.org/diseases-conditions/heart-disease/symptoms-causes/syc- 20353118
3. https://www.who.int/healthtopics/cardiovasculardiseases
4. https://www.cdc.gov/heartdisease/facts.htmhttps://www.who.int/news-room/fact-sheets/detail/cardiovascular- diseases-(cvds) www.kaggle.com/datasets/johnsmith88/heart-disease-dataset
5. Sushmita Roy Tithi ,AfifaAktar , Fahimul Aleem , Amitabha Chakrabarty „"ECG data analysisand heart disease prediction using machine learning algorithms"". Proceedings of 2019 IEEERegion 10 Symposium
6. Bo Jin ,Chao Che, Zhen Liu, Shulong Zhang, XiaomengYin,AndXiaopeng Wei, "Predicting the Risk of Heart Failure WithEHR Sequential Data Modeling" ,IEEE Access Volume 6 2018.
7. Ashir Javeed, Shijie Zhou, Liao Yongjian, Iqbal Qasim,Adeeb Noor, Redhwan Nour4, Samad Wali And Abdul Basit ,"An Intelligent Learning System based on Random SearchAlgorithm and Optimized Random Forest Model forImproved Heart Disease Detection" , IEEE Access 2017.This work is licensed under a Creative Commons Attribution 4.0 License Volume 4 2016.