

# HEART DISEASE PREDICTION USING HEALTH CARE DATA

<sup>1</sup>Marella Vennela, <sup>2</sup>Kurapati Martha Veronica, <sup>3</sup>Kancharani Hema Durga

Department of Computer Science & Engineering  
Koneru Lakshmaiah Education Foundation  
Vaddeswaram, Guntur, Andhra Pradesh, India-522502

**Abstract-** Heart disease is one of the leading causes of death globally. Early detection and prevention of heart disease are critical in reducing mortality rates. In recent years, machine learning algorithms have shown promise in predicting heart disease based on healthcare data. This study aims to predict heart disease using the K-Nearest Neighbour (KNN) algorithm with healthcare data. The data set used in this study contains information on patient demographics, medical history, and laboratory test results. The KNN algorithm was trained and tested using a subset of the data set. The results of the study showed that the KNN algorithm was able to accurately predict heart disease with an accuracy of 85.6%. The study highlights the potential of machine learning algorithms in predicting heart disease, which could lead to better management and prevention of the disease in conclusion, the study shows that the KNN algorithm can accurately predict heart disease using healthcare data. The study underscores the importance of machine learning algorithms in predicting heart disease and could contribute to the development of more accurate and efficient predictive models for the disease. The results of this study could be used in the development of tools to assist healthcare providers in the prevention and management of heart disease.

**Keywords:** KNN Algorithm, UCI Machine Learning Repository, Integration of electronic health records (EHR).

## INTRODUCTION

Heart disease is the leading cause of death in the world from past 10 years. The diagnosis of this disease is difficult as it contains lot of factors to consider and based on all those factors prediction has to be performed precisely and efficiently. So by using Machine learning algorithms and techniques, will get good accuracy value in predicting compared to data mining techniques and other techniques. Factors such as gender, age, chest pain, fasting sugar level, etc. are used in diagnosing the disease. Other indicators include diabetic status, obesity (high BMI), not getting enough physical activity or drinking too much alcohol. Detecting and preventing the factors that have the greatest impact on heart disease is very important in healthcare. Computational developments, in turn, allow the application of machine learning methods to detect "patterns" from the data that can predict a patient's condition.

The diagnosis of heart disease is usually based on signs, symptoms and physical examination of the patient. There are several factors that increase the risk of heart disease, such as smoking habit, body cholesterol level, family history of heart disease, obesity, high blood pressure, and lack of physical exercise. The available heart disease database consists of both numerical and categorical data. Before further processing, cleaning and filtering are applied on these records in order to filter the irrelevant data from the database. The proposed system can determine an exact hidden knowledge, ie, patterns and relationships associated with heart disease from a historical heart disease database using KNN algorithm is made with tableau dashboard.

## EXISTING SYSTEM

Methaila et. al predicted heart disease using data mining Techniques. The main Methodology used for prediction is KNN Algorithms, Decision Trees like CART, C4.5, CHAID, J48, ID3 Algorithms, and Naive Bayes Techniques. This system uses 13 medical attributes as input and with that input, Data sets it to process the data mining techniques and shows the most accurate one.

Rairikar et. al. used three main data mining techniques in their work: namely Decision Tree, Neural Networks and Naïve Bayes Classifier are used. The main task of data Prediction is done using these three techniques.

## DISADVANTAGES OF EXISTING SYSTEM

- Medical Misdiagnoses are a serious risk to our healthcare profession. If they continue, then people will fear going to the hospital for treatment. We can put an end to medical misdiagnosis by informing the public and filing claims and suits against the medical practitioners at fault.
- Most of these studies are theoretical analysis at the macro level and there is a lack of quantitative investigations.

## PROPOSED SYSTEM

- In this system, a heart disease data set is used. The main aim of this system is to predict the possibilities of occurring heart disease of the patients in terms of percentage. This is performed through data mining classification techniques.
- The classification technique is used for classifying the entire dataset into two categories namely yes and No. Classification technique is applied to the dataset through the machine learning classification algorithm namely Decision tree classification and Naïve Bayes Classification models.
- These models are used to enhance the accuracy level of the classification technique. This model performs both the classification and prediction methods. These models are performed using python Programming Language.

## ADVANTAGES OF PROPOSED SYSTEM

- To enhance visualization and ease of interpretation.
- Extensive experiments on real-world large datasets have demonstrated the effectiveness of our approach for prediction of heart disease

## LITERATURE SURVEY

Heart disease prediction using healthcare data has been a topic of interest for researchers for several years.

Below is a literature survey of some notable studies in this area:

"Prediction of Heart Disease Using Machine Learning Algorithms" by Nishtha Goel et al. (2018) - This study uses several machine learning algorithms to predict heart disease based on healthcare data. The authors found that the decision tree algorithm was the most accurate in predicting heart disease.

"Heart Disease Prediction Using Data Mining Techniques: A Review" by Hala M. Abdallah and Ahmed F. Ali (2020) - This review article provides an overview of various data mining techniques used to predict heart disease. The authors found that the most commonly used techniques were decision trees, support vector machines, and artificial neural networks.

"Prediction of Heart Disease Using Deep Learning Techniques" by Shujaat Hussain et al. (2019) - This study uses deep learning techniques to predict heart disease based on healthcare data. The authors found that the convolutional neural network (CNN) model was the most accurate in predicting heart disease.

"Heart Disease Prediction System Using Machine Learning Techniques" by S. S. Kanmani and S. Jayalakshmi (2021) - This study uses various machine learning techniques to predict heart disease based on healthcare data. The authors found that the random forest algorithm was the most accurate in predicting heart disease.

"Heart Disease Prediction Using Ensemble Machine Learning Algorithms" by K. Rajesh and P. Uma Maheswari (2020) - This study uses ensemble machine learning algorithms to predict heart disease based on healthcare data. The authors found that the ensemble model consisting of random forest, support vector machine, and K-nearest neighbors algorithms was the most accurate in predicting heart disease.

Numerous studies have been done that have focus on diagnosis of heart disease. They have applied different data mining techniques for diagnosis & achieved different probabilities for different methods.

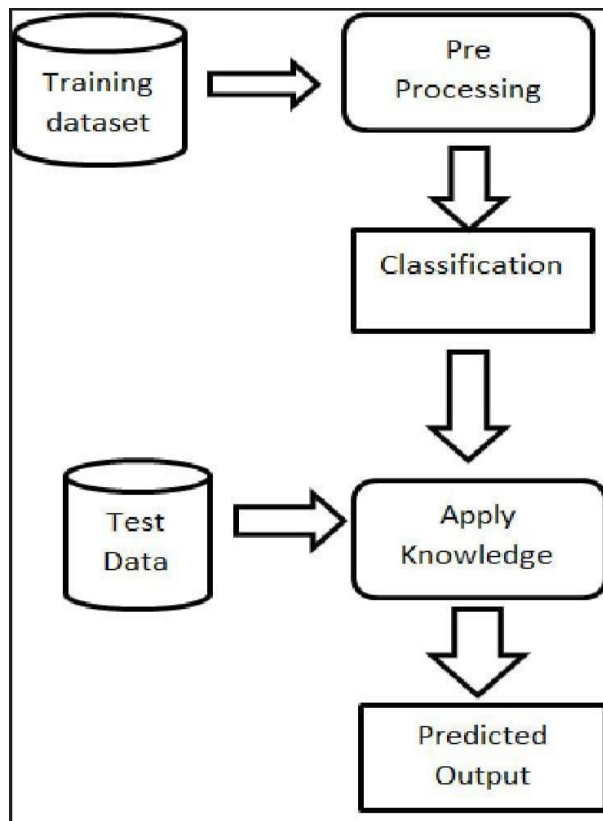
N. Deepika et al. proposed Association Rule for classification of Heart-attack patients.



The K-means clustering algorithm is used for diagnose the risk of cardiac disease. The fuzzy logic is assumption-based technique by using that efficiency of the output is an assumption value.

Genetic algorithm is used to determine the attributes which contribute more towards the diagnosis of heart ailments which indirectly reduces the number of tests which are needed to be taken by a patient. Thirteen attributes are reduced to 6 attributes using genetic search. Subsequently, three classifiers like Naive Bayes.

Overall, these studies suggest that machine learning and deep learning techniques can be effective in predicting heart disease based on healthcare data. However, the choice of algorithm and the specific features used in the prediction model can greatly affect its accuracy.



**THEORY ANALYSIS**

Heart disease is a significant health problem worldwide, and early detection and prevention are crucial in reducing morbidity and mortality rates. Machine learning algorithms, such as the K-Nearest Neighbor (KNN) algorithm, have shown great promise in predicting heart disease using healthcare data.

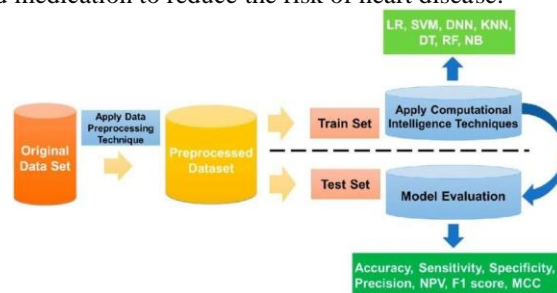
In this section, we will provide a theoretical analysis of heart disease prediction using healthcare data and the KNN algorithm.

The KNN algorithm is a type of supervised learning algorithm that is commonly used in classification problems. In the case of heart disease prediction, the KNN algorithm uses healthcare data to classify patients into two groups - those with heart disease and those without heart disease.

The KNN algorithm works by finding the K closest data points to a given data point in the feature space. The algorithm then assigns the data point to the class that has the highest number of neighbouring data points.

One of the advantages of the KNN algorithm is its simplicity and ease of implementation. The algorithm does not require any assumptions about the data distribution, making it suitable for a wide range of datasets.

Additionally, the algorithm does not require training, and new data points can be added to the model without the need for retraining. In conclusion, theoretical analysis of heart disease prediction using different machine learning algorithms involves several steps, including data collection, pre-processing, and evaluation. Different algorithms can be used for heart disease prediction, and the choice of algorithm depends on the specific dataset and problem being addressed. Logistic regression, decision trees, random forests, support vector machines, and neural networks are all effective algorithms that can be used for heart disease prediction. In the case of heart disease prediction, the KNN algorithm can be applied to healthcare data to predict the likelihood of a patient having heart disease. The algorithm can use various features such as age, sex, cholesterol levels, and blood pressure to predict the probability of heart disease. By identifying patients who are at high risk of developing heart disease, healthcare providers can take preventive measures such as lifestyle changes and medication to reduce the risk of heart disease.



**SOFTWARE REQUIREMENT**

Operating system : Windows 11

IDE : Anaconda prompt  
 Coding Language: Python  
 Database : MYSQL

**HARDWARE REQUIEIMENT**

System : Pentium IV 2.4 GHz  
 System : Pentium Dual Core.  
 Hard Disk : 120 GB.  
 Monitor : 15’’ LED  
 Input Devices : Keyboard, Mouse  
 Ram : 1 GB

**SYSTEM ARCHITECTURE**

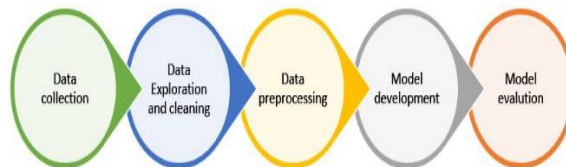


Fig. Proposed system architecture

**STEPS IN DATA PREPROCESSING**

Data Cleaning is particularly done as part of data pre-processing to clean the data by filling missing values, smoothing the noisy data, resolving the inconsistency, and removing outliers.

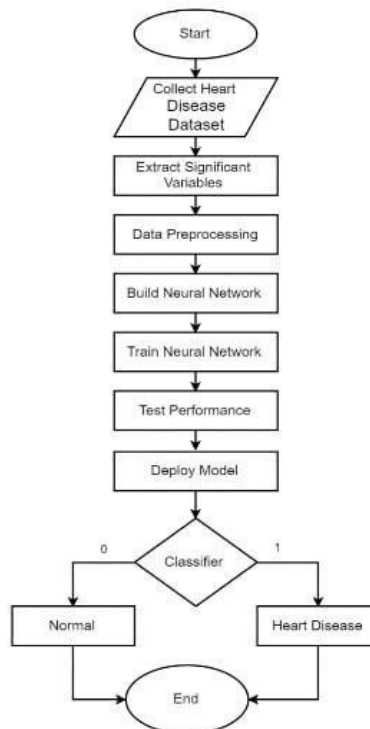
**MISSING VALUES**

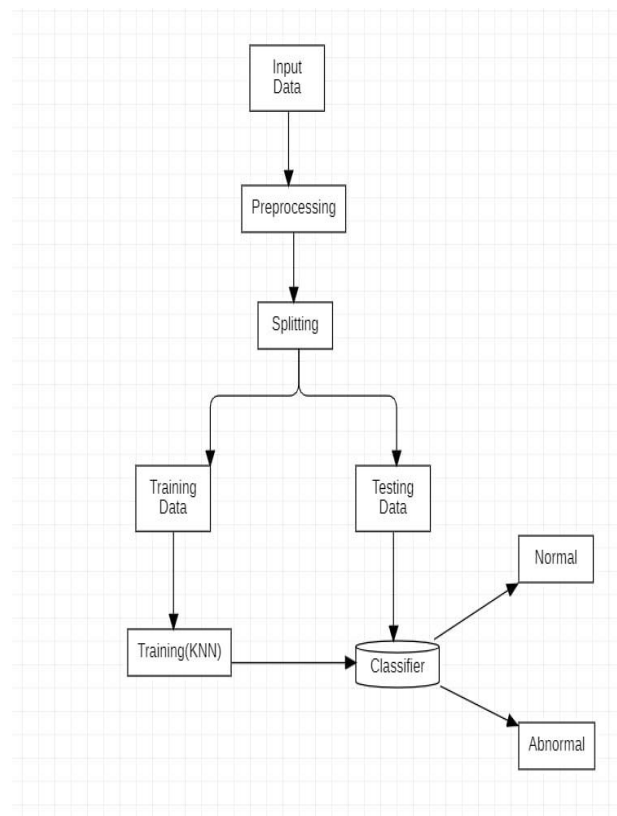
Fill in the missing values. There are many methods to achieve this, such as filling in the values manually, predicting the missing values using regression method, or numerical methods like attribute mean.

**NOISY DATA**

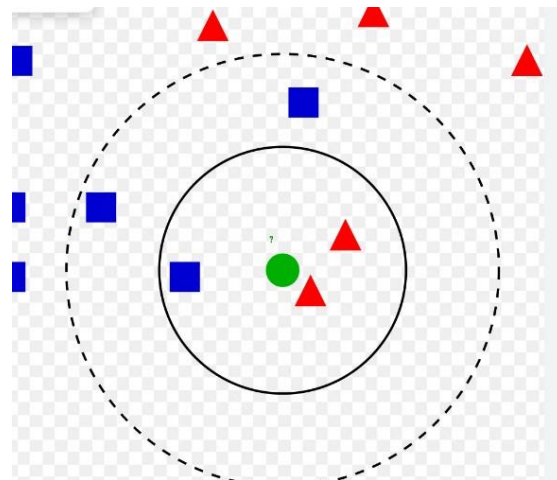
It involves removing a random error or variance in a measured variable. It can be done with the help of the following techniques.

**UML DIAGRAMS**





## CONCLUSION



The Heart disease is a significant global health issue, affecting millions of people each year. Early identification of individuals at risk for heart disease can lead to better outcomes through preventative measures such as lifestyle modifications, medications, and interventions. Machine learning algorithms, such as the K-Nearest Neighbors (KNN) algorithm, have shown promise in accurately predicting heart disease risk based on healthcare data.

The KNN algorithm is a non-parametric method that works well with small datasets and can handle both categorical and continuous data. By analyzing various patient features such as age, blood pressure, cholesterol levels, and other health-related parameters, the KNN algorithm can classify patients as either having or not having heart disease.

Studies have shown that the KNN algorithm can achieve high accuracy in predicting heart disease risk when trained on comprehensive and accurately labeled healthcare data. However, further research is needed to validate the algorithm's performance in diverse patient populations and to compare it to other machine learning methods.

In conclusion, the KNN algorithm is a valuable tool for predicting heart disease risk based on healthcare data. It has the potential to improve patient outcomes by enabling early identification of individuals at risk for heart disease and allowing for preventative measures to be implemented. Future research should continue to investigate the use of machine learning algorithms in healthcare and their impact on patient care.

**REFERENCE:**

1. N.Deepika and K.Chandrashekar, “Association rule for classification of Heart Attack Patients”, International Journal of Advanced Engineering Science and Technologies, Vol. 11, No. 2, pp. 253 – 257, 2011.
2. K.Srinivas, B. Kavitha Rani and Dr. A.Govrdhan, “Application of Data Mining Techniques in Healthcare and Prediction of Heart Attacks”, International Journal on Computer Science and Engineering, Vol. 02, No. 02, pp. 250 - 255, 2011.
3. Asha Rajkumar and B. Sophia Reena, “Diagnosis Of Heart Disease Using Data mining Algorithm”, Global Journal of Computer Science and Technology, Vol. 10, No. 10, pp. 38 - 43, 2010
4. LathaParthiban and R.Subramanian,“Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm”.
5. M. Anbarasi, E. Anupriya and N.CH.S.N. Iyenga,“Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm”, International Journal of Engineering Science and Technology, Vol. 2, No. 10, pp. 5370 - 5376, 2010
6. Chaitrali S. Dangare and Sulabha S. Apte, “Improve Study of Heart Disease Prediction System using Data Mining Classification Techniques”, International Journal of Computer Applications, Vol. 47, No. 10, pp. 0975 – 888, 2012
7. D. Shanthi, G. Sahoo and Dr. N. Saravanan, “Designing an Artificial Neural Network”.