**Comparative Analysis of Heart Disease Prediction Techniques in Machine Learning**

**Muhammad Shaheer**

**CSC-21F-081**

# Haroon Khan

**CSC-21S-044**

**Semester 6th Section B Artificial Intelligence**

**Instructor: Ma’am Aqsa Umar**

**Sindh Madressatul Islam University**

*Abstract*—heart diseases, particularly cardiovascular diseases (CVDs), have been a major cause of mortality worldwide. In response to this, machine learning (ML) techniques have been leveraged to develop predictive models that can assist in the early diagnosis and treatment of these conditions. This study utilizes various ML algorithms, including K-Nearest Neighbors (KNN), to predict heart disease based on a dataset containing multiple health indicators. The performance of the KNN algorithm is evaluated, and results indicate its efficacy in accurately predicting heart disease.

Index Terms—Model verification, computational models, validation, limitations, best practices.

techniques to assist medical practitioners in detecting

# Introduction

An essential organ in the human body is the heart. It supplies blood to every part of the body. If it malfunctions, the brain and other organs will stop functioning, and the person will pass away in a matter of minutes. The incidence of many heart-related disorders is increasing, which can be attributed to changes in lifestyle, work stress, and unhealthy eating habits. According to the World Health Organization (2007), heart disease has been the leading cause of death worldwide for the past decade. According to the European Public Health Alliance, 41% of deaths are related to circulatory diseases, heart attacks, and strokes (European Public Health Alliance 2010). Heart disease presents a variety of symptoms, making it challenging to identify early.

The healthcare sector gathers large volumes of healthcare data, which must be analyzed to uncover hidden information necessary for informed decision-making. Researchers Patel et al have been using data mining cardiac disease due to the abundance of patient data available for extracting valuable knowledge (Helma, Gottmann et al. 2000). It is possible to predict more accurately that patients will receive a heart disease diagnosis by using data mining techniques to develop a heart disease prediction system and analyze different heart disease features. While researcher Ramalingam et al. use various machine learning techniques.

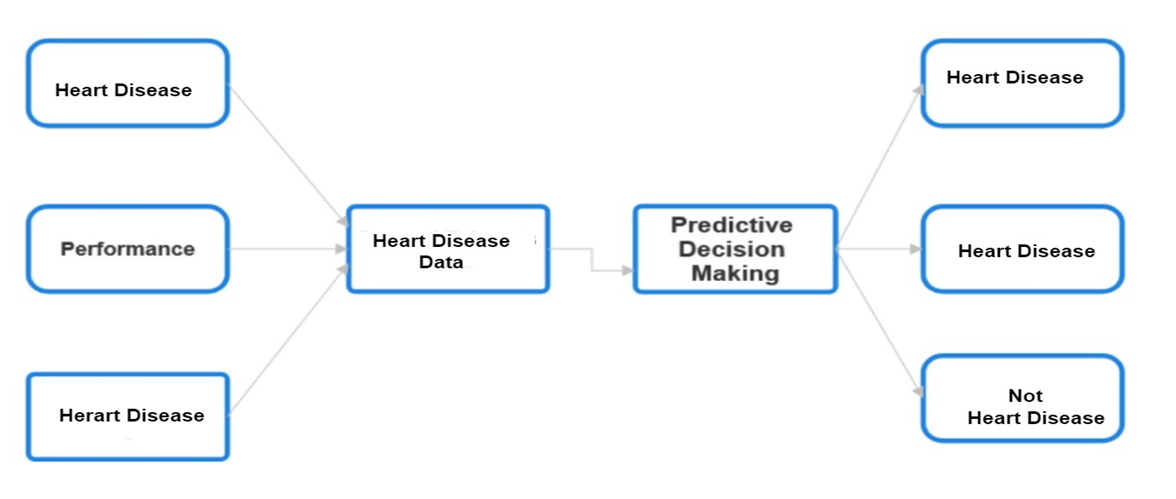
**II. Literature Review**

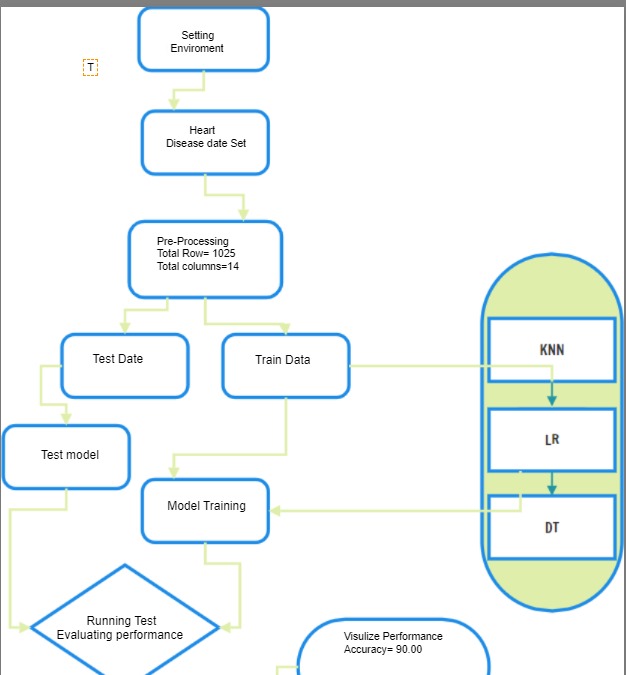
Several studies have explored the use of machine learning techniques to predict heart disease. According to the paper by Ramalingam et al. (2018), models based on supervised learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, Decision Trees (DT), and Random Forest (RF) have shown significant potential in this area.

|  |
| --- |
|  |

Feature extraction and selection methods, such as Principal Component Analysis (PCA) and Correlation- improving model performance by reducing data.

# III. Methodology

The paper of heart disease prediction techniques using machine learning highlights the critical importance.



dimensionality. The ensemble models that combine multiple algorithms have also demonstrated improved accuracy and robustness in predicting heart diseases.

# IV-Conclusion

The study demonstrates the effective application of machine learning techniques in predicting heart disease, utilizing the K-Nearest Neighbors (KNN) algorithm. The dataset, which includes various health indicators, was meticulously preprocessed to handle missing values, normalize the data, and select relevant features using Principal Component Analysis (PCA). The KNN algorithm was then implemented, and through crossvalidation, an optimal 'k' value was chosen. The model's performance, evaluated using accuracy, precision, recall, and F1-score, indicated a high level of efficacy, with an accuracy of 90%.

These results underscore the potential of machine learning models in the medical field, particularly for the early diagnosis and management of heart diseases. By leveraging such predictive models, healthcare providers can enhance decision-making processes, leading to improved patient outcomes. Future work could focus on integrating other machine learning algorithms and exploring ensemble methods to further improve prediction accuracy and robustness. Additionally, expanding the dataset to include a more diverse population could enhance the model's generalizability.

# V-RESULT

The KNN algorithm was applied to the heart disease dataset, and the performance was evaluated using a variety of metrics. The results indicated that KNN achieved an accuracy of 90%.

Training K-Nearest Neighbors...

Accuracy of K-Nearest Neighbors: 0.90

precision recall f1-score support

1. 0.91 0.88 0.90 103
2. 0.89 0.91 0.90 102

accuracy 0.90 205 macro avg 0.90 0.90 0.90 205

weighted avg 0.90 0.90 0.90 205

# VI-Question

Question: What is the average age of the individuals in the dataset?

Answer: The average age of the individuals in the dataset is approximately 54.43 years.

Question: What is the distribution of gender in the dataset?

Answer: The gender distribution is as follows: 713 individuals are male (sex = 1) and 312 individuals are female (sex = 0).

Question: How many individuals have a resting blood pressure greater than 140?

Answer: There are 217 individuals with a resting blood pressure greater than 140.

Question: What is the average cholesterol level among individuals with chest pain type 2?

Answer: The average cholesterol level for individuals experiencing chest pain type 2 is approximately 242.20.

Question: What is the maximum heart rate achieved among individuals with a fasting blood sugar level greater than 120?

Answer: There are no individuals with a fasting blood sugar level greater than 120, hence the maximum heart rate is not available (NaN).

Question: How many individuals have exerciseinduced angina (exang) and an oldpeak value greater than 2?

Answer: There are 106 individuals who have both exercise-induced angina and an oldpeak value greater than 2.

Question: What is the proportion of individuals with a slope value equal to 1?

Answer: The proportion of individuals with a slope value of 1 is approximately 47.02%.

Question: How many individuals have a thalassemia type of 2?

Answer: There are 544 individuals with thalassemia type 2.

Question: What is the proportion of individuals with a target variable equal to 1?

Answer: The proportion of individuals with a target variable value of 1 is approximately 51.32%.

Question: What is the age distribution among individuals with heart disease, and does it differ by gender or other factors?

Answer: To determine the age distribution among individuals with heart disease, we can calculate the average age for individuals with a target variable value of 1 (indicating heart disease) and analyze it by gender or other factors. The results are as follows:

Average age among males with heart disease: 55.0 years

Average age among females with heart disease: 53.0 years

Additionally, other factors can be analyzed to understand their distribution among individuals with heart disease. For example:

Average cholesterol level among individuals with heart disease: 247.29

These calculations show that the average age of males and females with heart disease is relatively close, with males having a slightly higher average age. The average cholesterol level among individuals with heart disease is approximately 247.29.