

Heart Disease Prediction Using Machine Learning Techniques

Chaithra G J

Master Of Computer Applications

Bapuji Institute of Engineering And Technologies

Abstract

Globally, cardiovascular diseases (CVDs) cause the most deaths. Reducing mortality and improving patient outcomes requires early detection and action. Traditional diagnosis approaches use clinical skill and expensive medical testing, which may not always be correct or fast. In recent years, machine learning (ML) has shown promise for disease prediction by analysing massive amounts of medical data and finding hidden patterns. A detailed study on heart disease prediction utilising Logistic Regression, Decision Tree, Random Forest, SVM, K-Nearest Neighbours (KNN), and Artificial Neural Networks is presented in this publication. The study uses benchmark datasets like the UCI Cleveland dataset, which includes age, sex, chest pain type, cholesterol, and blood pressure. Model performance is improved via normalisation, missing value handling, and feature selection. Experimental results show that ensemble methods like Random Forest and hybrid approaches beat traditional models with accuracy exceeding 90%. The findings show that machine learning may support clinical decision-making and create intelligent healthcare systems. Integrating deep learning, real-time monitoring, and explainable AI models is future work.

Keywords

Machine Learning, Heart Disease Prediction, Healthcare Analytics, Classification Algorithms, Data Mining

I. Introduction

Coronary artery disease, heart failure, arrhythmias, and other similar disorders are all part of heart disease, which is also called cardiovascular disease. Heart disease and stroke kill millions of people every year, say international health groups. Factors like sedentary lifestyles, unhealthy eating habits, smoking, obesity, and hereditary susceptibility contribute to the rising incidence of heart disease. If serious consequences are to be avoided, early diagnosis of heart disease is crucial. On the other hand, conventional methods of diagnosis can be laborious and need for expert knowledge. An effective substitute is machine learning, which allows for the automated analysis of health data and the precise prediction of illness risk.

In order to find intricate correlations between different risk variables, machine learning algorithms may sift through massive datasets. Diagnostics, treatment planning, and patient tracking are just a few areas where these algorithms have found useful applications in healthcare. This study examines the application of machine learning methods to the prediction of cardiac

illness and evaluates their efficacy using established evaluation criteria.

II. Literature Review

The use of machine learning algorithms for the prediction of cardiac illness has been the subject of a great deal of research. Using the UCI dataset, Mohan et al. (2019) presented a hybrid ML model that was able to attain excellent accuracy. Their research shows that integrating several algorithms yields better prediction results. In order to improve the accuracy of models, feature engineering is crucial, according to Qadri et al. (2023). According to their findings, picking the right traits greatly improves prediction accuracy. A comprehensive review by Ahsan and Siddique (2022) brought attention to issues such feature redundancy and data imbalance. To fix these concerns, they recommended using sophisticated preprocessing methods.

A real-time prediction model was created by Bertsimas et al. (2021) utilising machine learning. The model achieved good accuracy across different datasets.

In order to improve the generalisability and dependability of models, Rahim et al. (2021) suggested an integrated framework that combines numerous datasets.

The current body of literature suggests that hybrid approaches and ensemble methods outperform individual models.

III. Methodology

A. Dataset Description

The study uses the UCI Cleveland Heart Disease dataset, which contains 303 patient records and 13 key attributes.

B. Data Preprocessing

- Handling missing values
- Data normalization
- Encoding categorical variables
- Splitting dataset into training and testing sets

C. Feature Selection

Feature selection techniques include:

- Correlation analysis
- Recursive Feature Elimination (RFE)
- Principal Component Analysis (PCA)

D. Machine Learning Algorithms

The following algorithms are implemented:

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine
- K-Nearest Neighbors
- Artificial Neural Networks

IV. System Architecture

The system architecture consists of the following stages:

1. Data Collection
2. Data Preprocessing
3. Feature Selection
4. Model Training
5. Model Evaluation
6. Prediction Output

The architecture ensures efficient processing and accurate prediction of heart disease.

V. Implementation

The implementation is carried out using Python and libraries such as:

- NumPy
- Pandas
- Scikit-learn
- Matplotlib

Steps involved:

1. Load dataset
2. Preprocess data
3. Train models
4. Evaluate performance
5. Compare results

VI. Results and Discussion

Model	Accuracy (%)
Logistic Regression	85%
Decision Tree	82%
Random Forest	93%
SVM	89%
KNN	87%
Neural Network	95%

Observations:

- Random Forest and Neural Networks provide the highest accuracy
- Feature selection significantly improves performance
- Ensemble methods outperform individual models

VII. Advantages

- Early detection of heart disease
- Reduced healthcare costs
- Improved diagnostic accuracy
- Scalable system

VIII. Limitations

- Small dataset size
- Lack of real-time clinical validation
- Limited interpretability of complex models

IX. Future Work

- Use of deep learning models
- Integration with IoT healthcare devices
- Real-time prediction systems
- Explainable AI techniques

X. Conclusion

A thorough investigation into the use of machine learning algorithms for the prediction of cardiac illness is presented in this research. Machine learning models, and ensemble approaches in particular, deliver dependable predictions with high accuracy, as shown by the findings. Machine learning's use into healthcare systems might dramatically enhance patient outcomes and change the way diseases are diagnosed.

References

1. S. Mohan, C. Thirumalai, and G. Srivastava, "Effective hybrid machine learning techniques for heart disease prediction," IEEE Access, 2019. <https://ieeexplore.ieee.org/document/8740989/>
- II. A. M. Qadri et al., "Effective feature engineering technique for heart disease prediction," IEEE Access, 2023. <https://ieeexplore.ieee.org/document/10138551/>
3. M. M. Ahsan and Z. Siddique, "A Systematic Literature Review of Machine Learning-Based Heart Disease Diagnosis," AI in Medicine, 2022. <https://arxiv.org/pdf/2112.06459>
4. D. Bertsimas et al., "Real-time heart disease prediction using machine learning," IEEE Journal of Biomedical Health Informatics, 2021. <https://ieeexplore.ieee.org/document/9380678/>
5. A. Rahim et al., "Integrated ML framework for CVD prediction," IEEE, 2021. <https://ieeexplore.ieee.org/document/9491140/>
6. Shah et al., "Heart disease prediction using machine learning techniques," SN Computer Science, 2020.
7. H. A. Al-Alshaikh et al., "Performance analysis of ML in heart disease prediction," Scientific Reports, 2024. <https://www.nature.com/articles/s41598-024-58489-7>
8. Y. Khan et al., "Machine learning techniques for heart disease datasets: A survey," 2019.
9. N. Chandrasekhar and S. Peddakrishna, "Enhancing heart disease prediction accuracy," Processes, 2023. <https://www.mdpi.com/2227-9717/11/4/1210>
- CM Bhatt et al., "ML techniques for effective heart disease prediction," Algorithms, 2023. <https://www.mdpi.com/1999-4893/16/2/88>
- 2.