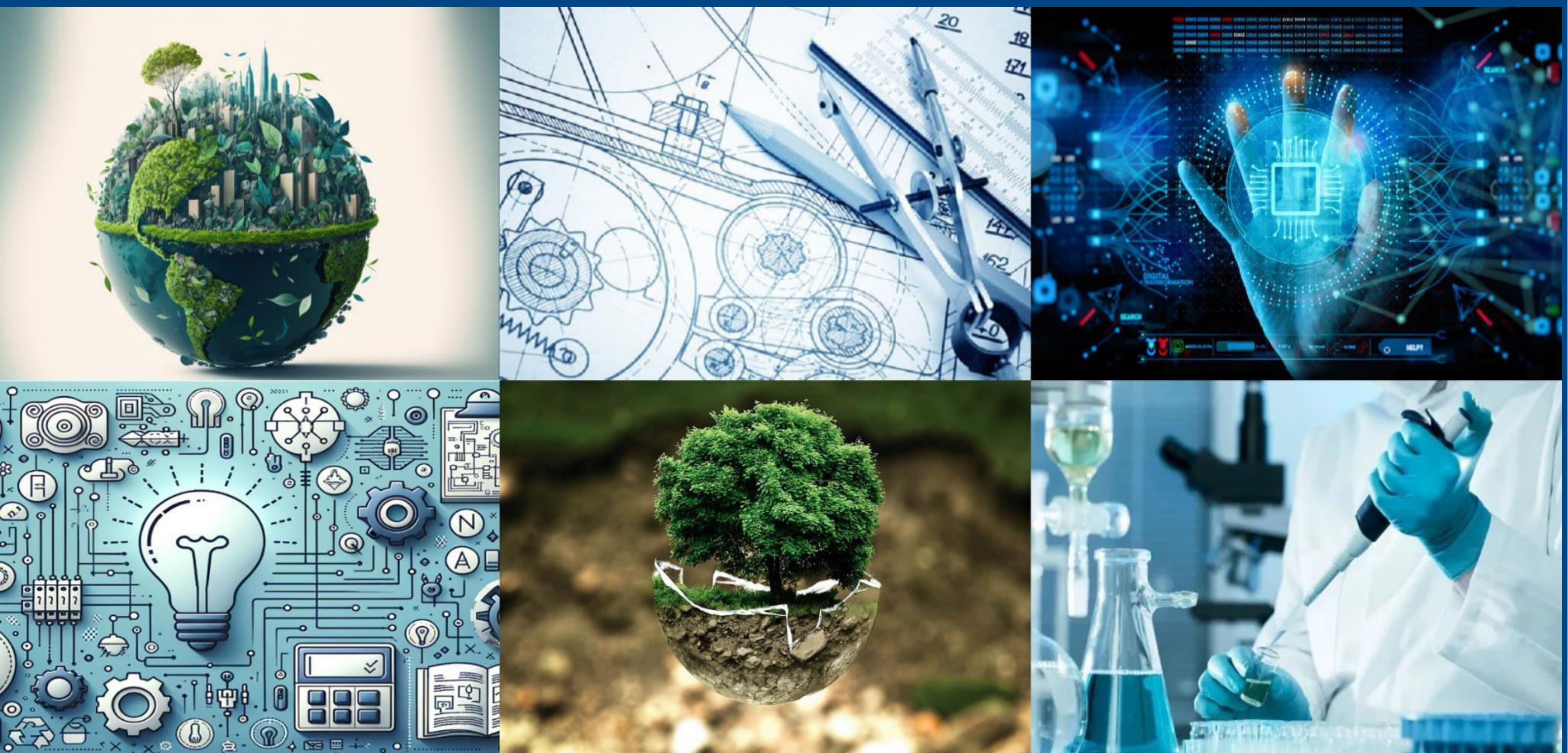
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 7, July 2025

© 2025 IJMRSET | Volume 8, Issue 7, July 2025| DOI:10.15680/IJMRSET.2025.0807074

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Heart Disease Prediction using Machine Learning Techniques

Bindushree T. N, Chethan M. S

M. Tech Student, Dept. of CSE, SIET, Tumkur, India

Assistant Professor, Dept. of CSE, SIET, Tumkur, India

ABSTRACT: Cardiovascular diseases are a foremost cause of morbidity and mortality around the world. Accurate tools for early detection are essential to reduce the risk and improve patient outcomes. The advances in machine learning (ML) have enabled the development of predictive models capable of analyzing medical data to forecast heart disease risk. This paper provides a comprehensive review of recent ML algorithms applied to heart disease prediction, commonly used datasets, relevant medical features, comparative analysis of models, and remaining challenges in practical application.

KEYWORDS: heart disease prediction, machine learning, artificial intelligence, data mining, cardiovascular risk.

I. INTRODUCTION

Cardiovascular disorders remain a primary threat to global health, causing millions of deaths annually. Early and precise prediction is crucial for effective treatment and management. Traditional diagnostic processes rely on a clinician's evaluation and often involve several tests that may yield inaccurate results due to human error or data complexity. With the exponential growth of clinical data, machine learning methods allow for deeper mining of this data, identifying patterns that might not be immediately apparent to clinicians, and paving the way for more reliable heart disease prediction.

II. BACKGROUND AND MOTIVATION

The classical approach to heart disease risk assessment focuses on scoring systems that combine various patient factors, such as cholesterol levels, age, blood pressure, and lifestyle choices. Despite their widespread use, these approaches may overlook nonlinear dependencies among diverse risk elements. ML methods, in contrast, leverage computational techniques to model these complex relationships and can improve the precision of predictions, thus supporting personalized medicine initiatives.

III. DATASETS USED FOR HEART DISEASE PREDICTION

Several well-known datasets are commonly used to train and evaluate predictive models:

• UCI Cleveland Heart Disease Dataset: This is a benchmark dataset with 303 patient records and 13 clinical markers.
• Statlog Heart Dataset: Contains 270 samples and shares many features with the Cleveland set.
• Institutional/Regional Data: Some studies utilize data collected from healthcare institutions or regional sources, contributing to dataset diversity

Key features across these datasets typically include: Age, Sex, Resting Blood Pressure, Serum Cholesterol, Fasting Blood Sugar, Resting ECG Results, Maximum Heart Rate, Chest Pain Type, Exercise-Induced Angina, ST Depression, Number of Major Vessels, and Thalassemia.
Results–Popular Machine Learning Methods

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. FEATURE ENGINEERING AND DATA PREPARATION

• Feature Selection: Techniques like principal component analysis (PCA) or the chi-squared test are employed to identify the most informative variables.
• Class Imbalance Handling: Strategies such as SMOTE (Synthetic Minority Over-sampling Technique) are used to ensure balanced datasets for training.
• Data Cleaning: Methods include imputing missing values, normalizing scales, and encoding categorical data into numerical formats.

V. COMPARATIVE RESULTS AND INSIGHTS

Recent literature reports that ensemble algorithms, including Random Forest and XGBoost, typically achieve the best accuracy, with rates between 91% and 97.5% on widely-used public datasets. The incorporation of deep learning has pushed reported accuracy in select studies to above 99%. The use of feature selection further refines model accuracy and interpretability. Moreover, several teams have successfully implemented these models in web and mobile applications, making them accessible for real-time clinical use.

VI. CHALLENGES AND OPEN PROBLEMS

• Generalization: Models may not perform consistently across populations with differing demographics due to the limitations of available datasets.
• Model Interpretability: While deep learning models are powerful, their "black box" nature raises concerns about clinical trust and adoption.
• Data Quality and Scope: Most models are trained on relatively small, homogeneous datasets, limiting their application in varied clinical environments.
• Integration: Deploying ML predictions effectively in medical practice necessitates regulatory validation, user training, and robust software infrastructures.

VII. CONCLUSION

ML-based prediction aids have shown high promise in detecting heart disease risk at an early stage. Continued research in dataset expansion, transparency, and validation in diverse settings is fundamental to bridging the gap between experimental results and practical, deployable solutions in clinical care.

Algorithm	**Features**	**Typical Accuracy Range**
Logistic Regression	Simple, interpretable	85–92%
Support Vector Machine	Effective in high dimensions	89–92%
Decision Trees	Transparent, can overfit	88–90%
Random Forest	Ensemble, robust flexibility	91–97%
XGBoost, AdaBoost	Boosting, strong with tabular data	Up to 97%
Naive Bayes	Fast, works with independence assumption	88–94%
Deep Neural Networks	Captures non-linear associations	94–99%

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

1. World Health Organization, "Cardiovascular diseases (CVDs)", 2023.
2. K. Srinivas, B.K. Rani, and A. Govrdhan, "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", 2011, IEEE.
3. D. Detrano et al., "International application of a new probability algorithm for the diagnosis of coronary artery disease", American Journal of Cardiology, vol. 64, pp. 304–310, 1989.
4. J. A. Khan, A. Hussain, and A. Iqbal, "Prediction of Heart Disease using Machine Learning Algorithms," Procedia Computer Science, vol. 173, pp. 221-228, 2020.
5. S. Choudhury and P. Gupta, "A Web Application for Heart Disease Prediction using Machine Learning," in Proc. IEEE, 2021.
6. Lighthouse Documentation. Accessed 2024. https://developers.google.com/web/tools/lighthouse
7. MongoDB Atlas Documentation. Accessed 2024. https://www.mongodb.com/docs/atlas/
8. Docker Documentation. Accessed 2024. https://docs.docker.com/
9. Kubernetes Documentation. Accessed 2024. https://kubernetes.io/docs/

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY