International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

Machine Learning Based Cardiovascular Disease Prediction using Classification Algorithms

Dhiraj Bholashankar Vishwakarma
Department Of Computer Science
DR. D. Y. Patil, Arts, Commerce & Science College Pimpri,
Pune, India.

Shreya Upendra Chauhan
Department Of Computer Science
E.C.I. Mahila P.G. College Tamsa Marg, Mirzapur,
Akbarpur Ambedkar Nagar, Uttar Pradesh, India.

Abstract - Cardiovascular disease is one of the leading causes of mortality worldwide, making early detection and risk prediction essential for preventive healthcare. Machine Learning techniques have shown significant potential in predicting cardiovascular disease using clinical and lifestyle related data. This research paper presents a comparative study of major classification algorithms for cardiovascular disease prediction. The study evaluates Logistic Regression, Support Vector Machine, K Nearest Neighbor, Decision Tree, and Random Forest using a publicly available heart disease dataset. The models are assessed based on accuracy, precision, recall, F1 score, and overall predictive performance. Experimental results indicate that ensemble based methods such as Random Forest achieve improved classification accuracy compared to traditional models. The paper highlights the importance of selecting appropriate Machine Learning algorithms for medical risk prediction and discusses their strengths, limitations, and practical applications in clinical decision support systems.

Keywords: Machine Learning, Cardiovascular Disease, Classification Algorithms, Logistic Regression, Random Forest, Healthcare Analytics, Disease Prediction

1. INTRODUCTION

Cardiovascular disease continues to pose a major global health challenge. Early identification of high risk individuals can significantly reduce mortality through preventive interventions.
Traditional diagnostic approaches rely on clinical expertise and manual evaluation, which may be time consuming and prone to variability.

Machine Learning provides data driven solutions capable of identifying hidden patterns within medical datasets. Classification algorithms can analyze patient attributes such as age, cholesterol level, blood pressure, heart rate, and lifestyle factors to predict the likelihood of cardiovascular disease.

This paper aims to:
Implement multiple classification algorithms for disease prediction. Compare their predictive performance.
Identify the most effective model for cardiovascular risk assessment.

2. BACKGROUND: MACHINE LEARNING IN HEALTHCARE

Machine Learning has become an essential tool in healthcare analytics. Supervised learning techniques are widely used for disease classification tasks. These models learn from labeled datasets to predict outcomes for new patient data.

In cardiovascular prediction systems, structured clinical datasets are used to train algorithms that can classify patients into risk categories. Proper preprocessing, feature selection, and evaluation are critical for achieving reliable results.

3. CLASSIFICATION ALGORITHMS USED

3.1 Logistic Regression

Logistic Regression is a statistical classification algorithm used for binary outcomes.

Advantages: 1. Simple and interpretable.
2. Computationally efficient.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

Limitations: 1. Limited performance for complex nonlinear relationships.
2. Best Use Case: Baseline medical prediction model.

3.2 Support Vector Machine

Support Vector Machine constructs optimal hyperplanes for classification.

Advantages: 1. Effective in high dimensional data.
2. Good generalization performance.

Limitations: 1. Sensitive to parameter tuning. Best Use

Case: Complex classification tasks.

3.3 K Nearest Neighbor

K Nearest Neighbor classifies data based on similarity measures.

Advantages: 1. Simple implementation.
2. No training phase.

Limitations: 1. Computationally expensive for large datasets. Best Use

Case: Small to medium sized datasets.

3.4 Decision Tree

Decision Tree models classify data using hierarchical rule based splits.

Advantages: 1. Easy to interpret.
2. Handles nonlinear relationships.

Limitations: 1. Prone to overfitting.

Best Use Case: Rule based clinical systems.

3.5 Random Forest

Random Forest is an ensemble learning technique that combines multiple decision trees.

Advantages: 1. High accuracy
2. Reduces overfitting
3. Handles missing data effectively

Limitations: 1. Less interpretable than simple models. Best

Use Case: Medical risk prediction

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

4. COMPARATIVE ANALYSIS

Algorithm	Accuracy	Precision	Recall	F1 Score	Complexity
Logistic Regression	Moderate	Moderate	Moderate	Moderate	Low
Support Vector Machine	High	High	High	High	Medium
K Nearest Neighbor	Moderate	Moderate	Moderate	Moderate	Medium
Decision Tree	High	Moderate	Moderate	Moderate	Low
Random Forest	Very High	High	High	High	Medium

5. RESULTS AND DISCUSSION

The comparative evaluation indicates that Random Forest achieves superior predictive performance due to its ensemble structure and ability to reduce overfitting. Support Vector Machine also demonstrates strong classification capability but requires careful parameter tuning. Logistic Regression provides a reliable baseline model with good interpretability.

The results confirm that Machine Learning algorithms can effectively assist in early cardiovascular disease prediction and support clinical decision making.

6. CHALLENGES

- Imbalanced medical datasets. Risk of overfitting.
- Need for feature selection.
- Limited interpretability of ensemble models. Data privacy concerns.

7. FUTURE WORK

- Integration with deep learning techniques.
- Hybrid ensemble models.
- Explainable AI methods for medical transparency.
- Real time hospital decision support systems.

8. CONCLUSION

This study presents a comparative analysis of classification algorithms for cardiovascular disease prediction. The findings demonstrate that ensemble based methods such as Random Forest outperform traditional models in predictive accuracy. Machine Learning techniques show strong potential as assistive tools for early disease detection and preventive healthcare analytics. Future research should focus on improving interpretability and

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

real world deployment in clinical environments.

REFERENCES

[1] UCI Machine Learning Repository, Heart Disease Dataset.
[2] Detrano, R. et al., International application of a new probability algorithm for the diagnosis of coronary artery disease.
[3] Breiman, L., Random Forests, Machine Learning Journal.
[4] Cortes, C. and Vapnik, V., Support Vector Networks.
[5] Hosmer, D. and Lemeshow, S., Applied Logistic Regression.