ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 4| April 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

PREDICTION OF TEN-YEAR CORONARY HEART DISEASE RISK USING DATA ANALYSIS AND MACHINE LEARNING

Sandhya Ombase[1], Namrata Nawale[2], Vaishnavi Jagadale[3]

[1]*Department of Computer Science, Dr. D. Y. Patil Arts, Commerce & Science College, Pimpri Pune, Maharashtra, India*

[1]*Department of Computer Science, Dr. D. Y. Patil Arts, Commerce & Science College, Pimpri Pune, Maharashtra, India*

[1]*Department of Computer Science, Dr. D. Y. Patil Arts, Commerce & Science College, Pimpri Pune, Maharashtra, India*

ABSTRACT

Coronary Heart Disease (CHD) is one of the leading causes of death worldwide. Early prediction of CHD risk plays a vital role in preventing severe health complications. This paper presents a method for predicting the ten-year risk of coronary heart disease using data analysis and machine learning techniques.

Healthcare datasets containing patient information such as age, cholesterol level, blood pressure, smoking habits, and diabetes are used. Various machine learning models like Logistic Regression, Decision Tree, and Random Forest are applied to predict the risk. The results show that machine learning techniques can provide accurate and efficient predictions, helping healthcare professionals in early diagnosis and treatment planning.

KEYWORDS: *Coronary Heart Disease, Machine Learning, Data Analysis, Predictive Analytics, Healthcare, Risk Prediction*

1. INTRODUCTION

Coronary Heart Disease (CHD) occurs due to the narrowing or blockage of coronary arteries, reducing blood flow to the heart. It is one of the most common causes of death globally. Early detection of CHD risk is essential to reduce mortality rates and improve patient outcomes.

With the advancement of technology, large volumes of healthcare data are being generated. Traditional methods of analysis are not sufficient to handle such data efficiently. Machine learning provides powerful tools to analyse complex datasets and identify hidden patterns.

This study aims to predict the ten-year risk of CHD using machine learning algorithms, enabling early intervention and better healthcare decision-making.

2. LITERATURE REVIEW

Several studies have explored the use of machine learning in heart disease prediction. Earlier methods relied on statistical techniques, which were limited in handling large datasets.

Recent research shows that machine learning models such as Logistic Regression, Support Vector Machines, and Random Forest can significantly improve prediction accuracy. These models can process multiple risk factors simultaneously and detect complex relationships in data.

However, challenges such as missing data, feature selection, and model interpretability still need to be addressed for better performance.

3. METHODOLOGY

3.1 Data Collection

The dataset used for this study includes patient health records with the following attributes:

- Age
- Gender
- Blood Pressure
- Cholesterol Level
- Smoking Status
- Diabetes

3.2 Data Preprocessing

Data preprocessing is an important step to improve model performance:

- Handling missing values
- Removing duplicate records
- Normalizing data
- Selecting relevant features

3.3 Machine Learning Models

The following algorithms are used:

- **Logistic Regression** – for binary classification
- **Decision Tree** – for rule-based prediction
- **Random Forest** – for higher accuracy using multiple trees

3.4 Model Evaluation

The models are evaluated using:

- Accuracy
- Precision
- Recall
- F1-Score

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 4| April 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

4. RESULTS AND DISCUSSION

The models were trained and tested on the dataset to evaluate their performance.

4.1 Model Performance

- Logistic Regression provided consistent results
- Decision Tree was easy to interpret
- Random Forest achieved the highest accuracy

4.2 Key Findings

- Age and cholesterol levels are major risk factors
- Smoking significantly increases CHD risk
- Machine learning improves prediction efficiency

4.3 Challenges

- Imbalanced dataset
- Missing values
- Need for large and high-quality data

4.4 Future Scope

- Use of deep learning models
- Integration with wearable health devices
- Real-time monitoring systems

5. CONCLUSION

This study shows that machine learning techniques are effective in predicting the ten-year risk of coronary heart disease. By analyse healthcare data, important risk factors can be identified, allowing early diagnosis and prevention.

The use of predictive models can assist doctors in making better decisions and improving patient care. Despite some challenges, machine learning has great potential in the healthcare sector.

6. REFERENCES

1. *World Health Organization – Cardiovascular Diseases https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)*
2. *Framingham Heart Study Dataset https://www.kaggle.com/datasets/amanajmera1/framingham-heart-study-dataset*
3. *IBM – Machine Learning in Healthcare https://www.ibm.com/topics/machine-learning*
4. *Microsoft – Healthcare AI https://azure.microsoft.com/en-in/solutions/healthcare/*
5. *Machine Learning Research Papers (IEEE Xplore) https://ieeexplore.ieee.org/*