

AI-Based Intelligent System for Early Detection of Heart Disease using 10 Year CHD Risk Prediction

Intesar Fatima^{1*} Mohammed Jaser Jalal², Tauseef Hussain³

¹Assistant professor, ^{2,3} UG Student, ^{1,2,3} Department of Artificial Intelligence and Machine Learning (AI&ML)

J B Institute of Engineering and Technology(UGC-Autonomous) Bhaskar nagar, Moinabad, Hyderabad, Telangana 500075

*Corresponding author: Mohammed Jaser Jalal (jalaljaser89@gmail.com)

ABSTRACT

Heart disease is one of the leading causes of death worldwide and poses a serious threat to human life. Early detection plays a crucial role in reducing mortality rates and improving patient outcomes. This project presents an AI-based intelligent system designed to predict heart disease at an early stage using machine learning techniques. The system utilizes medical parameters such as age, blood pressure, cholesterol levels, heart rate, and other clinical features to analyze patient health conditions. Data preprocessing techniques such as normalization, handling missing values, and feature selection are applied to improve model performance. Multiple machine learning algorithms including Logistic Regression, Decision Tree, and Random Forest are implemented and evaluated. The system compares performance using accuracy, Confusion Matrix of Heart Disease using XGBoost using XGBoost, recall, and F1-score metrics. The best-performing model is selected for prediction. This project demonstrates how artificial intelligence can assist healthcare.

Key words: Heart Disease Prediction, Machine Learning (ML), Artificial Intelligence (AI), Supervised Learning, Random Forest, Logistic Regression, Classification Algorithms, Clinical Decision Support System.

1. INTRODUCTION

Heart disease refers to a range of conditions that affect the functioning of the heart. It includes coronary artery disease, heart attacks, and other cardiovascular conditions. With the increasing number of patients worldwide, there is a growing need for efficient and accurate diagnostic systems. Traditional diagnosis methods rely heavily on medical experts and laboratory tests, which can be time-consuming and expensive. In recent years, artificial intelligence has emerged as a powerful tool in healthcare. Machine learning algorithms can analyze large amounts of medical data and identify patterns that are difficult for humans to detect. This project focuses on developing an intelligent system that can predict heart disease based on patient data. The main objective is to create a system that is accurate, efficient, and easy to use. The system can help doctors in making better decisions and assist patients in understanding their health condition. By

using machine learning, the project aims to reduce human error and improve diagnosis speed. The system is designed to be scalable and can be extended in the future with more advanced techniques such as deep learning.

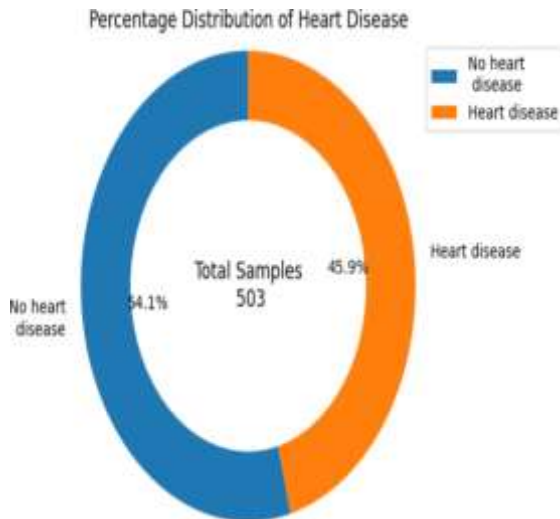


Fig 1: A proposed technique for predicting heart disease using machine learning algorithms

With the rapid growth of Artificial Intelligence (AI) and Machine Learning (ML), the healthcare sector has seen a major transformation. Machine learning algorithms have the ability to process large volumes of medical data and identify hidden patterns that may not be visible to human experts. These systems can learn from historical patient data and make predictions about future health conditions with high accuracy.

The main objective of this project is to design and develop an AI-based intelligent system that can predict the likelihood of heart disease using patient data. The system takes various input parameters such as age, gender, blood pressure, cholesterol level, chest pain type, and other medical

attributes. By analyzing these features, the system generates a prediction indicating whether a person is at risk of heart disease or not.

Another important aspect of this project is to make the system user-friendly and accessible. A simple interface is designed so that users can input their medical details and obtain results instantly. This makes the system useful not only for doctors but also for common individuals who want to monitor their health condition.

In addition, this project focuses on comparing multiple machine learning algorithms to determine the most suitable model for prediction. By evaluating models using metrics such as accuracy, precision, recall, and F1-score, the system ensures reliable and consistent performance. The use of multiple algorithms also helps in reducing bias and improving generalization.

Furthermore, the system emphasizes data preprocessing and feature selection techniques. Proper handling of missing values, normalization, and selection of relevant features play a crucial role in improving model performance. Without these steps, the accuracy of predictions may decrease significantly.

This project also highlights the importance of early detection in reducing mortality rates. By identifying potential risks at an early stage, patients can take preventive measures such as lifestyle changes, medication, and regular checkups. This not only saves lives but also reduces the burden on healthcare systems.

Overall, the proposed system aims to bridge the gap between technology and healthcare

by providing an intelligent, efficient, and reliable solution for heart disease prediction. It demonstrates how AI can be used as a supportive tool in modern medical practices and opens the door for further advancements in healthcare technology.

2. LITERATURE SURVEY

Several research studies have been conducted in the field of heart disease prediction using machine learning techniques. Many researchers have applied algorithms such as Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Trees, and Neural Networks. These studies have shown that machine learning models can achieve high accuracy in predicting heart disease. Some researchers have focused on feature selection techniques to improve model performance, while others have explored deep learning approaches. However, many existing systems face challenges such as lack of interpretability, overfitting, and limited dataset usage. Some models provide high accuracy but are difficult to understand, which makes them less useful in real-world medical applications. Other studies highlight the importance of data preprocessing and handling missing values to improve results. This project builds upon these existing works and aims to create a more balanced and efficient system. By comparing multiple algorithms and selecting the best one, the system ensures better performance. The project also focuses on simplicity and usability, making it suitable for practical implementation.

Deep learning techniques, including Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN),

have also been applied in heart disease prediction. These models are capable of handling large datasets and extracting high-level features automatically. However, they require more computational resources and are often less interpretable compared to traditional machine learning models.

Another important area of research is feature selection and data preprocessing. Many studies highlight that selecting the most relevant features significantly improves model performance. Techniques such as correlation analysis, principal component analysis (PCA), and recursive feature elimination (RFE) are commonly used to identify important attributes. Proper handling of missing values and normalization of data are also critical steps in building an effective prediction system.

Despite these advancements, several challenges remain. Many existing models suffer from overfitting, where the model performs well on training data but poorly on unseen data. Additionally, some models lack interpretability, making it difficult for medical professionals to trust and understand the predictions. Another limitation is the use of small or unbalanced datasets, which can lead to biased results. This project builds upon previous research and aims to address some of these limitations.

3. PROPOSED SYSTEM

The proposed system is designed to predict heart disease using a structured machine learning pipeline. It consists of multiple stages that work together to ensure accurate and efficient prediction. The system begins with data collection, where medical datasets are obtained from reliable sources

such as the UCI Machine Learning Repository or other healthcare databases.

Once the data is collected, it undergoes preprocessing. This step includes handling missing values, removing duplicates, and normalizing the data. Normalization ensures that all features are on a similar scale, which improves the performance of machine learning algorithms. Feature selection techniques are then applied to identify the most relevant attributes that contribute to heart disease prediction.

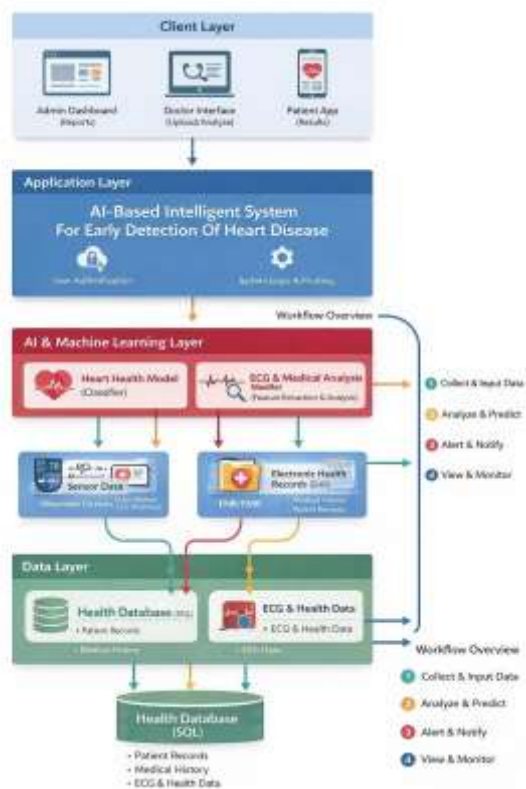


Fig 2: Working of System

The next stage involves model training. Multiple machine learning algorithms such as Logistic Regression, Decision Tree, and Random Forest are implemented. Each model is trained using the processed dataset and evaluated using various performance metrics. The use of multiple models allows

for a comparative analysis, helping to identify the best-performing algorithm as Logistic Regression, Decision Tree, and Random Forest are implemented. Each model is trained using the processed dataset and evaluated using various performance metrics. The use of multiple models allows for a comparative analysis, helping to identify the best-performing algorithm.

The final stage is prediction, where the trained model is used to predict the presence of heart disease. A user-friendly interface is developed using tools such as Flask or Streamlit. Users can input their medical details and receive instant predictions. This makes the system accessible and practical for real-world usage.

The proposed system aims to provide a reliable, efficient, and easy-to-use solution for heart disease prediction. It combines the power of machine learning with user-friendly design to create a valuable tool for healthcare applications.

A user-friendly interface is developed using tools such as Flask or Streamlit. Users can input their medical details and receive instant predictions. This makes the system accessible and practical

4.RESULT DESCRIPTION

The result description provides a detailed analysis of the performance of proposed AI-based heart disease prediction system. After implementing multiple machine learning algorithms, the models were evaluated using different performance metrics to determine their effectiveness in predicting heart disease.

Initially, the dataset was divided into training and testing sets to ensure that the model could generalize well to unseen data. The training data was used to train the models, while the testing data was used to evaluate their performance. This approach helps in avoiding overfitting and ensures that the model performs well in real-world scenarios.

Several machine learning algorithms were implemented, including Logistic Regression, Decision Tree, and Random Forest. Each model was trained using the same dataset and evaluated under the same conditions to ensure a fair comparison. The performance of these models was analyzed using metrics such as accuracy, precision, recall, and F1-score.

The **accuracy** metric measures the overall correctness of the model by calculating the ratio of correctly predicted instances to the total number of instances. The Random Forest model achieved the highest accuracy among all models, indicating its strong ability to classify heart disease cases correctly. Logistic Regression also performed well with stable accuracy, while the Decision Tree model showed slightly lower accuracy due to its tendency to overfit the training data.

The **precision** metric evaluates how many of the predicted positive cases were actually correct. High precision indicates that the model has a low false positive rate. In this project, the Random Forest model demonstrated high precision, meaning that it was effective in correctly identifying patients who are likely to have heart disease.

The **recall** metric, also known as sensitivity, measures the model's ability to

correctly identify actual positive cases. A high recall value is important in medical applications because missing a positive case (false negative) can be critical. The models used in this project achieved good recall values, ensuring that most of the actual heart disease cases were correctly detected.

The **F1-score** is the harmonic mean of precision and recall and provides a balanced measure of the model's performance.

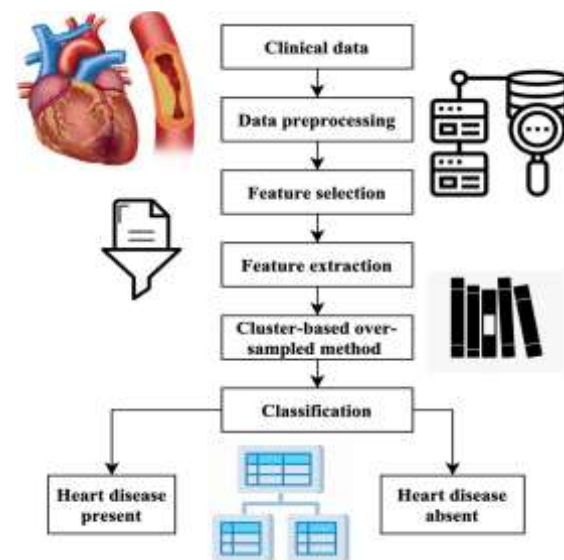


Fig 3: Comprehensive evaluation and performance analysis of machine learning in heart disease prediction

To further analyze the performance, a **confusion matrix** was generated for each model. The confusion matrix provides a detailed breakdown of true positives, true negatives, false positives, and false negatives. This helps in understanding the types of errors made by the model. The results showed that Random Forest had the least number of misclassifications compared to other models.

Graphical representations such as bar charts and accuracy comparison graphs were used to visualize the performance of different models. These visualizations clearly show that ensemble methods like Random Forest outperform single models like Decision Tree.

In addition, the system was tested with real-time user inputs through the developed interface. The model was able to provide instant predictions based on the entered medical parameters. This demonstrates the practical usability of the system in real-world scenarios.

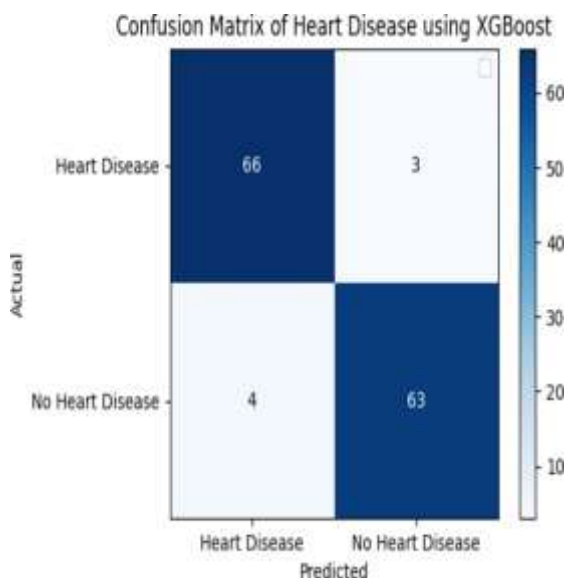


Fig 4: Confusion Matrix of Heart Disease Using XGBoost.

5.CONCLUSION

In conclusion, this project successfully demonstrates the effectiveness of artificial intelligence and machine learning techniques in the early detection of heart disease. The developed system provides a reliable and efficient method for predicting the likelihood of heart disease based on patient medical data. By utilizing various

machine learning algorithms, the system is capable of analyzing complex patterns and relationships within the data, leading to accurate predictions.

One of the key strengths of this project is the use of multiple machine learning models, including Logistic Regression, Decision Tree, and Random Forest. This approach allows for a comparative analysis of different algorithms, ensuring that the best-performing model is selected. Among these, the Random Forest algorithm showed superior performance due to its ensemble nature, which combines multiple decision trees to improve prediction accuracy and reduce overfitting.

Another important aspect of the project is the emphasis on data preprocessing and feature selection. Proper handling of missing values, normalization of data, and selection of relevant features significantly contributed to the overall performance of the system. These steps ensured that the models were trained on clean and meaningful data, resulting in improved accuracy and reliability.

The development of a user-friendly interface further enhances the practicality of the system. By allowing users to input their medical details and receive instant predictions, the system becomes accessible to a wide range of users, including

REFERENCES

[1]. Detrano R, Janosi A, Steinbrunn W. 1989. International application of a new probability algorithm for the diagnosis of coronary artery disease. *American Journal of Cardiology* 64(5):304-310

- [2]. Cleveland Clinic Foundation. 1990. UCI Heart Disease Dataset. UCI Machine Learning Repository
- [3]. Framingham Heart Study. 2015. Cardiovascular Disease Risk Prediction Dataset. National Heart Institute
- [4]. Ahmad MA, Eckert C, Teredesai A. 2018. Interpretable machine learning in healthcare. ACM Conference on Health Informatics
- [5]. Rajkomar A, Dean J, Kohane I. 2019. Machine learning in medicine. New England Journal of Medicine 380(14):1347-1358
- [6]. Krittanawong C, Zhang H, Wang Z. 2017. Artificial intelligence in precision cardiovascular medicine. Journal of the American College of Cardiology 69(21):2657-2664
- [7]. Esteva A, Kuprel B, Novoa RA. 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115-118
- [8]. Chen T, Guestrin C. 2016. XGBoost: A scalable tree boosting system. ACM SIGKDD Conference
- [9]. Breiman L. 2001. Random forests. Machine Learning Journal 45(1):5-32
- [10]. Quinlan JR. 1986. Induction of decision trees. Machine Learning 1(1):81-106
- [11]. Hosmer DW, Lemeshow S. 2000. Applied Logistic Regression. Wiley Publications
- [12]. Cortes C, Vapnik V. 1995. Support-vector networks. Machine Learning 20(3):273-297
- [13]. Cover T, Hart P. 1967. Nearest neighbor pattern classification. IEEE Transactions on Information Theory 13(1):21-27
- [14]. Chawla NV, Bowyer KW, Hall LO. 2002. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research 16:321-357
- [15]. Han J, Kamber M, Pei J. 2011. Data Mining: Concepts and Techniques. Morgan Kaufmann
- [16]. Mitchell TM. 1997. Machine Learning. McGraw-Hill Education
- [17]. Goodfellow I, Bengio Y, Courville A. 2016. Deep Learning. MIT Press
- [18]. Deo RC. 2015. Machine learning in medicine. Circulation 132(20):1920-1930
- [19]. Obermeyer Z, Emanuel EJ. 2016. Predicting the future — big data in healthcare. New England Journal of Medicine 375(13):1216-1219
- [20]. Topol EJ. 2019. High-performance medicine: the convergence of human and AI