



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



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

⊕ www.ijirset.com ⊠ ijirset@gmail.com 🖄 +91-9940572462 🕓 +91 63819 07438





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404261|

Real-Time Heart Disease Prediction

Dr. Thirupurasundari D.R, T. Raghul Balaji, H.Pranav Chandran, P. Naga Guravaiah,

Raushan Kumar

Associate Professor, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai,

Tamil Nadu, India

B. Tech Students, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India

ABSTRACT: Heart disease is a leading cause of death worldwide, emphasizing the need for early detection to improve patient care. This study explores the use of machine learning (ML) techniques to develop a predictive model for heart disease. Feature selection was performed using statistical methods to identify the most relevant attributes. Four ML classifiers—K-Nearest Neighbors (KNN), XGBoost, Decision Tree, and Random Forest were evaluated. The Synthetic Minority Oversampling Technique (SMOTE) was applied to address class imbalance and improve model performance. The models were trained and tested using cross-validation techniques. Performance was measured based on accuracy and precision to ensure reliability. Comparative analysis was conducted to assess the effectiveness of different classifiers. The findings demonstrate the ability of ML models to assist in early diagnosis. This approach can support healthcare professionals in making informed decisions. The study highlights the importance of data-driven methods in Medicare search. Machine learning techniques can enhance disease prediction and risk assessment. The results indicate potential improvements in early intervention strategies.

KEYWORDS: Heart disease, machine learning (ML), K-Nearest Neighbors, XGBoost, Decision Tree, Random Forest, Synthetic Minority Oversampling Technique

I. INTRODUCTION

Heart disease, including conditions such as coronary artery disease, heart failure, and arrhythmias, is a leading cause of morbidity and mortality worldwide. Current models often provide general predictions but struggle to accurately categorize specific heart diseases. Accurate classification is essential for timely diagnosis and effective treatment. This research proposes a custom machine learning model to classify multiple heart disease categories with higher precision than existing models, enabling better diagnosis and more personalized treatment plans.Heart disease encompasses a wide range of cardiovascular conditions, including coronary artery disease (CAD), heart failure, arrhythmias, valvular heart disease, and cardiomyopathy, making it one of the leading causes of morbidity and mortality worldwide. According to the World Health Organization (WHO), cardiovascular diseases account for approximately 32% of global deaths, with a significant proportion attributed to heart-related conditions. Early diagnosis and accurate classification of heart disease are crucial for improving patient outcomes, reducing mortality rates, and enabling timely and effective treatment interventions

II. LITERATURE SURVEY

Traditional machine learning models such as **Decision Trees (DT)**, K-Nearest Neighbors (KNN), Random Forests (RF), and XGBoost have been extensively used to predict heart disease based on patient data, including age, cholesterol levels, blood pressure, heart rate, and other risk factors. The study highlights that machine learning models can accurately predict the likelihood of heart disease, enabling early diagnosis and preventive intervention, thereby reducing the risk of life-threatening complications. The findings emphasize the potential of automating diagnostic procedures to assist healthcare providers and improve patient outcomes.

1) A Comprehensive Review of Machine Learning Models for Heart Disease Prediction

This study examines the performance of various machine learning algorithms, including K-Nearest Neighbors (KNN), Random Forests (RF), Decision Trees (DT), and Support Vector Machines (SVM), in predicting heart disease. The models were trained on datasets comprising patient attributes such as age, cholesterol, resting blood pressure, and





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404261|

exercise-induced angina. The results demonstrated that Random Forest and SVM achieved higher accuracy compared to other models. The study concludes that **ensemble methods** enhance classification accuracy and improve prediction reliability in clinical settings.

2) Enhancing Heart Disease Diagnosis with Feature Selection

The study highlights that **optimized models with relevant features** show improved accuracy and precision, making them more reliable for early diagnosis and classification of heart disease.

3) Application of XGBoost for Predictive Analysis in Heart Disease Diagnosis

This study explores the use of **XGBoost (Extreme Gradient Boosting)** to develop a predictive model for heart disease diagnosis. XGBoost, known for its efficiency and accuracy in handling structured data, was applied to classify heart disease cases based on multiple clinical features. The results demonstrated that XGBoost outperformed traditional models such as **KNN and Decision Tree**, achieving superior accuracy and reduced overfitting. The study highlights the potential of gradient boosting techniques for enhancing the predictive power of heart disease models.

4) Addressing Class Imbalance in Heart Disease Prediction Using SMOTE

This research focuses on addressing the issue of **class imbalance** in heart disease datasets by applying the **Synthetic Minority Oversampling Technique (SMOTE).** Class imbalance, where positive cases (patients with heart disease) are significantly fewer than negative cases, can lead to biased predictions. The application of SMOTE effectively balanced the dataset by generating synthetic samples, improving the model's ability to correctly predict minority class outcomes. The study demonstrates that incorporating SMOTE enhances the **sensitivity and recall** of machine learning models, making them more effective in identifying heart disease.

5) Comparative Analysis of Machine Learning Algorithms for Heart Disease Classification

This study conducts a **comparative analysis** of machine learning classifiers to evaluate their effectiveness in heart disease classification. Models such as **K-Nearest Neighbors (KNN)**, **Decision Trees (DT)**, **Random Forests (RF)**, **and XGBoost** were evaluated based on metrics such as **accuracy**, **precision**, **recall**, **and F1-score**. The results highlighted that **Random Forest and XGBoost consistently outperformed other models**, offering higher accuracy and better generalization. The research emphasizes the importance of selecting appropriate classification algorithms to ensure reliable predictions in real-world clinical applications.

6) Deep Learning for Heart Disease Prediction: A Convolutional Neural Network Approach

This work explores the application of **Convolutional Neural Networks (CNNs)** for heart disease prediction by analyzing ECG images and other visual data. CNNs were trained to automatically extract important features and distinguish between normal and abnormal heart activity. The results demonstrated that CNN models achieved high accuracy, making them suitable for detecting patterns that are often missed by traditional models. The study highlights the potential of **deep learning models** to enhance diagnostic capabilities by incorporating medical imaging into predictive systems.

7) Evaluation of Cross-Validation Techniques in Heart Disease Prediction

This study evaluates the impact of different **cross-validation techniques**, such as **k-fold cross-validation and stratified cross-validation**, on the performance of heart disease prediction models. Cross-validation ensures that models are evaluated on multiple subsets of the data, preventing overfitting and improving model generalization. The study concludes that **stratified k-fold cross-validation** provides more reliable performance estimates, leading to more robust and dependable models for heart disease classification.

8) Role of Ensemble Methods in Improving Heart Disease Prediction

This research explores the impact of **ensemble methods** such as **Bagging**, **Boosting**, **and Stacking** in improving the accuracy and robustness of heart disease prediction models. By combining multiple base models, ensemble methods reduce variability and enhance prediction accuracy. The study demonstrates that **stacked ensemble models** combining Decision Trees, Random Forests, and XGBoost achieved the highest classification accuracy, highlighting the effectiveness of ensemble learning in complex medical datasets.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404261|

9) Automated Diagnosis of Heart Disease Using Streamlit-Based Web Application

This work highlights the development of a **Streamlit-based web application** that allows users to input their clinical data and receive real-time predictions for heart disease risk. The web application integrates the machine learning model with an intuitive interface, making it accessible to healthcare professionals and patients. The system provides detailed predictions and visualization of results, ensuring a user-friendly experience and facilitating **early intervention and informed decision-making**.

III. METHODOLOGY

Data Collection

The first step is to gather a comprehensive dataset containing clinical and diagnostic data related to heart disease. The dataset should be diverse, including information on multiple risk factors, demographic details, and medical history. The most commonly used dataset for this task is the UCI Heart Disease Dataset, which contains records with attributes such as:

- Age, Gender, and Smoking History
- Chest Pain Type (CP)
- Resting Blood Pressure (trestbps)
- Serum Cholesterol Level (chol)
- Fasting Blood Sugar (fbs)
- Resting Electrocardiographic Results (restecg)
- Maximum Heart Rate Achieved (thalach)
- Exercise-Induced Angina (exang)
- Oldpeak (ST Depression Induced by Exercise)
- Number of Major Vessels Colored by Fluoroscopy (ca)
- Thalassemia Indicator (thal)

Data Preprocessing

To prepare the data for machine learning models, several preprocessing techniques are applied to improve data quality and make it suitable for analysis. The steps include:

• Handling Missing Values: Fill or drop missing data to ensure data integrity.

• Feature Encoding: Convert categorical variables into numerical formats using encoding techniques like One-Hot Encoding or Label Encoding.

• Data Normalization/Standardization: Scale numerical features to ensure that they contribute equally to model performance.

• Splitting Data: Split the dataset into training and testing sets, ensuring that the model learns from one part and evaluates on unseen data.

Feature Selection and Extraction

Feature selection plays a crucial role in the success of heart disease prediction models. The system extracts the most relevant features from the dataset that contribute to the classification of heart disease. Key features include:

- Age and Gender: Basic demographic information contributing to risk factors.
- Clinical Parameters: Blood pressure, cholesterol levels, and maximum heart rate.

Machine Learning Model Training

The extracted features are used to train multiple machine learning models to recognize patterns and relationships between the data and heart disease presence. The goal is to identify the most effective model for predicting heart disease accurately.

Popular Machine Learning Models for Heart Disease Prediction:

- K-Nearest Neighbors (KNN): Classifies data based on the similarity to the nearest neighbors.
- Decision Tree (DT): Constructs decision paths to classify patients based on feature values.
 - Random Forest (RF): An ensemble technique that aggregates multiple decision trees for robust classification.

• XGBoost (Extreme Gradient Boosting): Optimizes model performance by minimizing classification errors and handling # Example of Model Training





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404261|

Model Evaluation and Performance Testing

After training, the models are tested on a new set of data that was not used during the training phase. This ensures that the model generalizes well to unseen data.

Evaluation Metrics:

- Accuracy: Measures the overall correctness of predictions.
- Precision: Identifies how many of the predicted positives are actually positive.
- Recall (Sensitivity): Determines how many actual positives were correctly predicted.
- F1-Score: Provides a balance between precision and recall, especially for imbalanced datasets.
- ROC-AUC Curve: Assesses the model's ability to distinguish between positive and negative cases.

Deployment of the Web Application

Once the model achieves a satisfactory performance level, it can be deployed as a web application for real-time predictions. The Streamlit framework is used to develop a user-friendly interface that allows users to enter patient data and receive predictions.

Deployment Process:

- Model Serialization: Save the trained model using pickle or joblib for future use.
- Web Interface: Design an interactive web application using Streamlit with input forms and result sections.

• Backend Integration: Connect the machine learning model with the web application to receive user inputs and generate predictions dynamically.

System Working Process:

The system becomes familiar with the "patterns" of heart disease through feature analysis. By comparing the input patient data with learned patterns, the model can determine the likelihood of heart disease.

IV. MODELING AND ANALYSIS

The general system architecture of the Heart Disease Prediction Model outlines the process of collecting patient data, preprocessing it, applying machine learning algorithms, and providing accurate predictions regarding heart disease. By leveraging techniques such as data normalization, feature extraction, and model training, the system ensures high accuracy and reliability. The architecture follows a well-defined workflow where clinical parameters are analyzed and classified to predict the presence or absence of heart disease.







www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404261|

Data Collection and Preprocessing:

The process begins with acquiring clinical and demographic data from patients, including parameters data from patients, including parameters such as Age, Gender, Chest Pain Type, Blood Pressure, Serum Cholesterol Level, Fasting Blood Sugar, Maximum Heart Rate Achieved, Exercise Induced Angina, ST Depression Induced by Exercise, Number of Major Vessels Colored by Fluoroscopy.

Data Preprocessing is an essential step in preparing raw data for machine learning. First, we handle missing values by either removing or filling them in. Next, outliers are identified and dealt with to prevent them from affecting the model. The data is then normalized or scaled to ensure that all features have the same range. We perform feature selection to eliminate unnecessary or redundant features. If there's class imbalance, we may balance the classes by oversampling or undersampling. The dataset is then split into training and testing set to assess model performance. Finally, The data is ready for use in training the machine learning model.

Model Training and Evaluation:

In this stage, the preprocessed data is used to train multiple machine learning models, including, K-Nearest Neighbors, Decision Tree, Random Forest, XGBoost. The models learn to recognize patterns and relationships in the data to distinguish between patients with and without heart disease. Cross validation techniques are applied to ensure that the models generalize well to unseen data, minimizing the risk of overfitting or underfitting.

Web Application for Real-Tine Prediction Hand:

The trained model is integrated into a Streamlit-powered web application, allowing users to enter patient data and receive predictions in real-time.

The web interface ensures that predictions are accessible to healthcare professionals and patients, enabling easy interpretation of results.

Prediction and Classification of Heart Disease:

The trained model analyze new patient data due to classify individuals as:

0:"No Heart Disease", 1:"Mild Heart Disease", 2:"Moderate Heart Disease", 3:"High Heart Disease", 4:"Very High Heart Disease".

V. RESULTS AND DISCUSSION

The real-time heart disease prediction system, developed using machine learning models such as K-Nearest Neighbors, Random Forest, Decision Tree, and XGBoost, demonstrates a high level of accuracy in predicting the presence of heart disease.

The system effectively analyzes clinical and diagnostic data to generate reliable predictions, supporting timely medical intervention.

The heart disease prediction system achieves high accuracy, real-time prediction, data reliability, and seamless user interaction through advanced machine learning techniques. The web application using Streamlit.







Deplay 1

www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404261|

+ 0 set	* *	-	$\tau_{-} \propto$
e 9 0	Incalment Field	* 0 0	0 :

Heart Disease Prediction

Orag and drop file here	Groww New
and the second second second second	

Fig.2.web user interface (stresmlit app)

VI. CONCLUSION

The custom machine learning model for multiclass heart disease classification offers significant improvements over existing models. It accurately classifies heart disease categories with an accuracy, surpassing K-Nearest Neighbors, Decision Tree, and Random Forest. By incorporating data preprocessing, feature selection, and model fine-tuning, this system ensures reliable and precise predictions. It has the potential to revolutionize healthcare by enabling early detection and timely intervention for heart diseases. Data preprocessing techniques, including handling missing values, normalization, and class balancing using SMOTE, have ensured that the input data is clean, consistent, and well-structured. Feature selection further optimized the performance by identifying the most relevant attributes, minimizing noise, and focusing on features that have higher correlation with heart disease outcomes. The system's scalability and adaptability ensure that it can easily integrated with electrocardiogram (ECG) which can help diagnose heart conditions like arrhythmias, heart attacks, and coronary artery narrowing, this opens the door for continuous monitoring of high-risk patients and timely alerts for healthcare providers, paving the way for preventive care and long-term patient management.

REFERENCES

- F.A.Latifah, I. Slamet and Sugiyanto, "Comparision of heart disease classification with logistic regression algorithm and random forest algorithm," AIP Conference Proceedings 2296, 0200021 (2020), Nov.2020, Volume 2296, Issuw 1, doi; 10.1063/5.0030579.
- M. G. El-Shafiey, A. Hagag, E. -S. A. El-Dahshan, and M. A. Ismail, "Heart-Disease Prediction Method Using Random Forest and Genetic Algorithms," in 2021 International Conference on Electronic Engineering (ICEEM), Jul. 2021, pp. 1–6. doi: 10.1109/ICEEM52022.2021.9480625.
- L. KN, N. R, N. K, R. Kumari, S. N, and V. K, "Heart Disease Detection using Machine Learning Technique," in 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Aug. 2021, pp. 1738–1743. doi: 10.1109/ICESC51422.2021.9532705.
- K. Battula, R. Durgadinesh, K. Suryapratap, and G. Vinaykumar, "Use of Machine Learning Techniques in the Prediction of Heart Disease," in 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Oct. 2021, pp. 1–5. doi: 10.1109/ICECCME52200.2021.9591026.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404261|

- S. Kusuma and K. R. Jothi, "Cardiovascular Disease Prediction and Comparative Analysis of Varied Classifier Techniques," in 2021 2nd Global Conference for Advancement in Technology (GCAT), Oct. 2021, pp. 1–7. doi: 10.1109/GCAT52182.2021.9587734.
- Sharma, R. Kumar, and V. Jaiswal, "Classification of Heart Disease from MRI Images Using Convolutional Neural Network," in 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), Oct. 2021, pp. 358–363. doi: 10.1109/ISPCC53510.2021.9609408.
- 7. M. Fradi, L. Khriji, and M. Machhout, "Real-time arrhythmia heart disease detection system using CNN architecture based various optimizers-networks," Multimed. Tools Appl., vol. 81, no. 29, pp. 41711–41732, Dec. 2022, doi: 10.1007/s11042-021-11268-2.
- 8. **Y. Zhang, L. Diao, and L. Ma,** "Logistic Regression Models in Predicting Heart Disease," J. Phys. Conf. Ser., vol. 1769, no. 1, p. 012024, Jan. 2021, doi: 10.1088/1742-6596/1769/1/012024.









INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com