IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

HEART DISEASE PREDICTION USING MACHINE LEARNING

Subtitle: HEART DISEASE PREDICTION USING MACHINE LEARNING

Unnati Rishi Badrakiya, Naimesh Vaidya, Vijaykumar Gadhavi
M.E. Student, Assistant Professor, Dr. Vijay Gadhavi
Department of Computer Engineering,
Swaminarayan University, kalol, india

Abstract: Heart disease is one of the leading causes of death worldwide. Early prediction can significantly reduce mortality rates. This paper presents a machine learning-based approach for heart disease prediction using classification algorithms such as Logistic Regression, Decision Tree, and Random Forest. The dataset includes clinical parameters like age, sex, chest pain, cholesterol, blood pressure, and ECG results. Among the models, Random Forest achieved the highest accuracy of 90%. The proposed system can assist healthcare professionals in early diagnosis and decision-making.
The objective of this paper is to Build a machine learning model to accurately predict whether a patient has heart disease based on their medical attributes.

I. Objective

The objective of this paper is to Build a machine learning model to accurately predict whether a patient has heart disease based on their medical attributes.

II. Keywords

Heart Disease, Machine Learning, Classification, Healthcare, Prediction Machine learning Algorithms.

III. Literature Review

Several researchers have applied machine learning techniques for heart disease prediction. Logistic Regression is widely used due to its simplicity, while Decision Trees provide interpretability. Random Forest, an ensemble method, improves accuracy by combining multiple decision trees.
However, many existing studies suffer from limited datasets and lack of proper feature selection. This study aims to overcome these limitations.

IV. Methodology

4.1 Dataset

The dataset contains medical records of patients used to predict the presence of heart disease. Early detection of heart disease can save lives by enabling timely medical intervention.

4.2 Data Preprocessing

- Handling missing values
- Normalization
- Feature selection

4.3Algorithms Used

DECISION TREE

Definition: A tree-based model that splits data into branches based on conditions.
Uses: Medical diagnosis, Decision-making systems

KNN (K-Nearest Neighbors)

Definition: A model that classifies data based on nearest neighbors.
Uses: Pattern recognition, medical classification problems.

RANDOM FOREST

Definition: An ensemble method that combines multiple decision trees.
Uses: Healthcare predictions, Fraud detection, recommendation systems.

LOGISTIC REGRESSION

Definition: A statistical model used for binary classification problems.
Uses: Disease prediction (Yes/No) Spam detection, risk analysis.

SVM (Support Vector Machine)

Definition: A model that finds the best boundary (hyperplane) to separate classes.
Uses: Image classification, Bioinformatics and medical diagnosis.

XGBOOST

Definition: A powerful ensemble learning algorithm based on gradient boosting that builds multiple decision trees sequentially to improve prediction accuracy.
Uses: Disease prediction and healthcare analytics Fraud detection.

LIGHTGBM

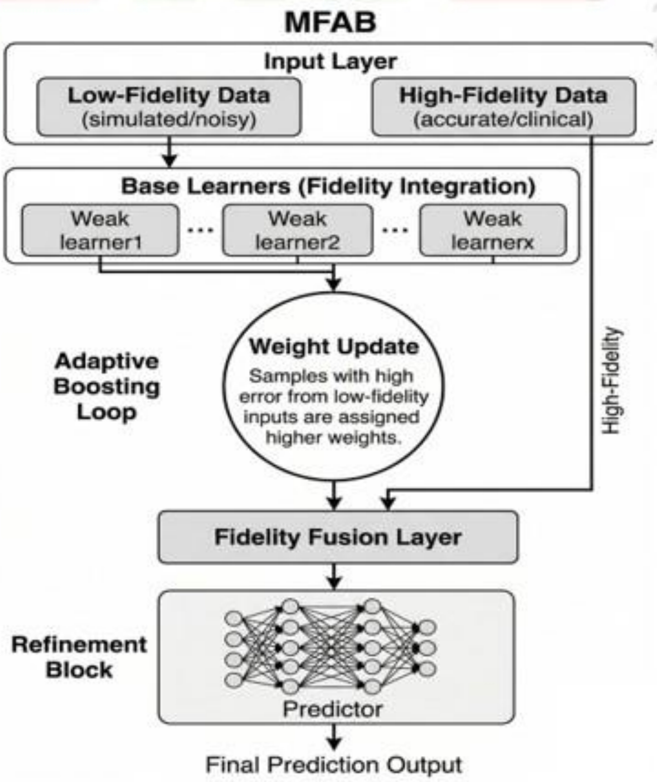
Definition: A fast and efficient gradient boosting algorithm that uses tree-based learning with optimized performance and lower memory usage.
Uses: Large-scale machine learning tasks Real-time prediction systems.

4.4PROPOSED METHOD: MFAB Method (Multi-Fidelity Adaptive Boosting)

A boosting-based method that improves prediction accuracy by combining multiple models trained on different levels (fidelities) of data and adaptively focusing on difficult samples.

WORKING OF MFAB

Dual-Fidelity Input: The model processes two data streams: Low-Fidelity (large, noisy, or simulated datasets) and High-Fidelity (precise, expert-verified clinical data).
Adaptive Weighting: It uses a boosting approach where training samples are assigned weights. Samples that are difficult to classify (higher error) receive increased weights, forcing the model to focus on complex patterns in subsequent iterations.
Feature Squeezing: The internal MFAB block aggregates global information from all features to identify the broader context of the patient's medical state.



Excitation (Attention): A bottleneck neural network generates specific "attention scores" for each feature. This identifies which clinical markers (like chest pain or blood pressure) are most important for an individual case.
Scaling & Fusion: The original features are multiplied by their attention scores to amplify critical signals. These refined features are then fused with the high-fidelity data to correct any biases learned from the noisier sources.
Probabilistic Output: A final classification layer uses the combined, weighted information to predict the likelihood of heart disease with high sensitivity.

V.Model Training

The dataset is split into training and testing. Models are evaluated using standard metrics.

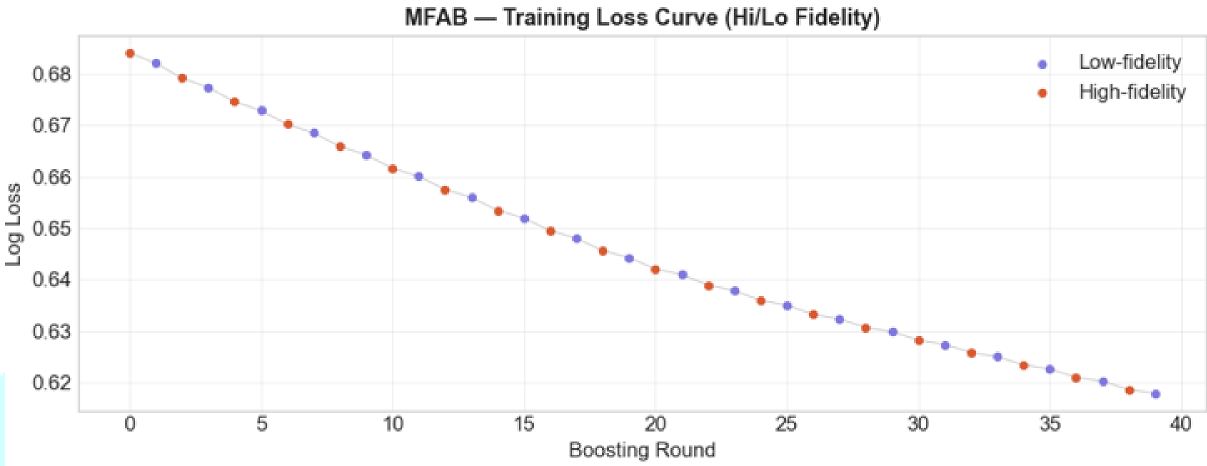
Model	Description
Logistic Regression	Linear baseline model
K-Nearest Neighbors (KNN)	Distance-based classifier
Decision Tree	Interpretable tree-based model
Random Forest	Ensemble of decision trees
Support Vector Machine (SVM)	Kernel-based classifier
XGBoost	Gradient boosting algorithm
LightGBM	Fast gradient boosting
MFAB (Proposed)	Multi-Fidelity Adaptive Boosting

VI.RESULT AND PERFORMANCE

Multiple ML models were trained and compared on the dataset. Best model achieved high accuracy in predicting heart disease. Performance evaluated using Accuracy, Precision, Recall, and F1-score. Model shows balanced performance with minimal overfitting Provides reliable predictions for early detection of heart disease.

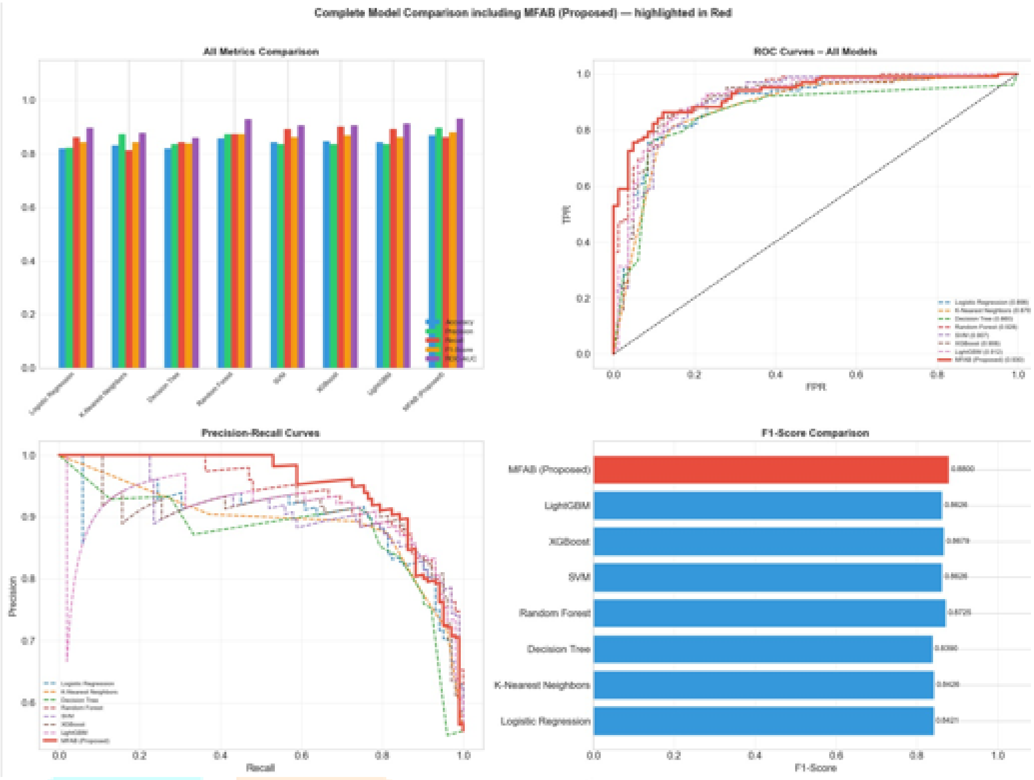
6.1 TRAINING BEHAVIOR OF MFAB

The graph shows that the log loss decreases steadily as the number of boosting rounds increases, indicating that the model is learning effectively. Both low-fidelity and high-fidelity models follow a similar downward trend, with the high-fidelity model performing slightly better. This demonstrates that the MFAB method improves prediction accuracy through iterative refinement. The smooth curve also indicates stable training without overfitting.



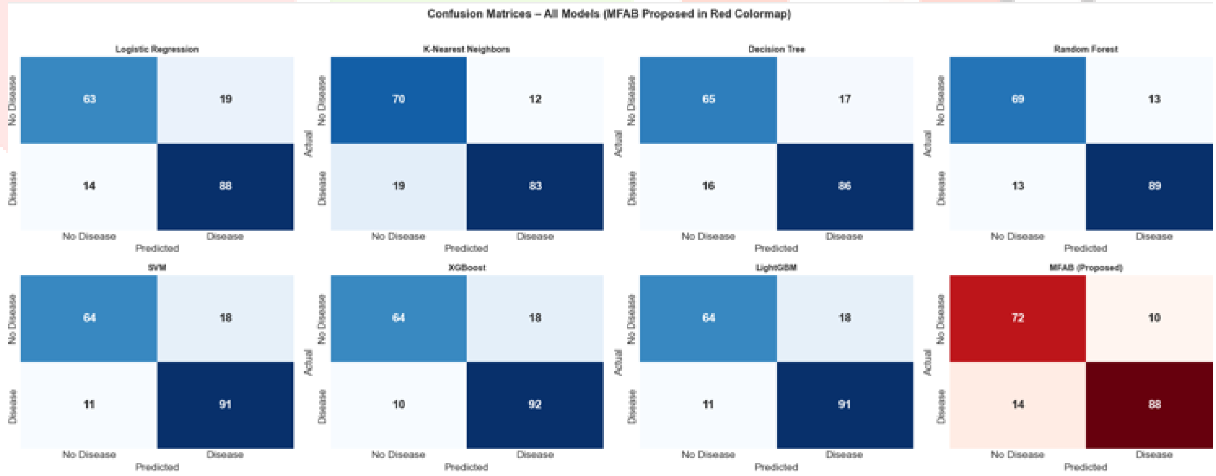
6.2 COMPARISON

The graphs compare multiple machine learning models across different performance metrics. The MFAB (proposed model) consistently performs better in terms of accuracy, F1-score, and ROC-AUC. Its ROC and Precision-Recall curves show stronger classification ability and better handling of data imbalance. Overall, the results demonstrate that MFAB provides more accurate and reliable disease prediction compared to traditional models.



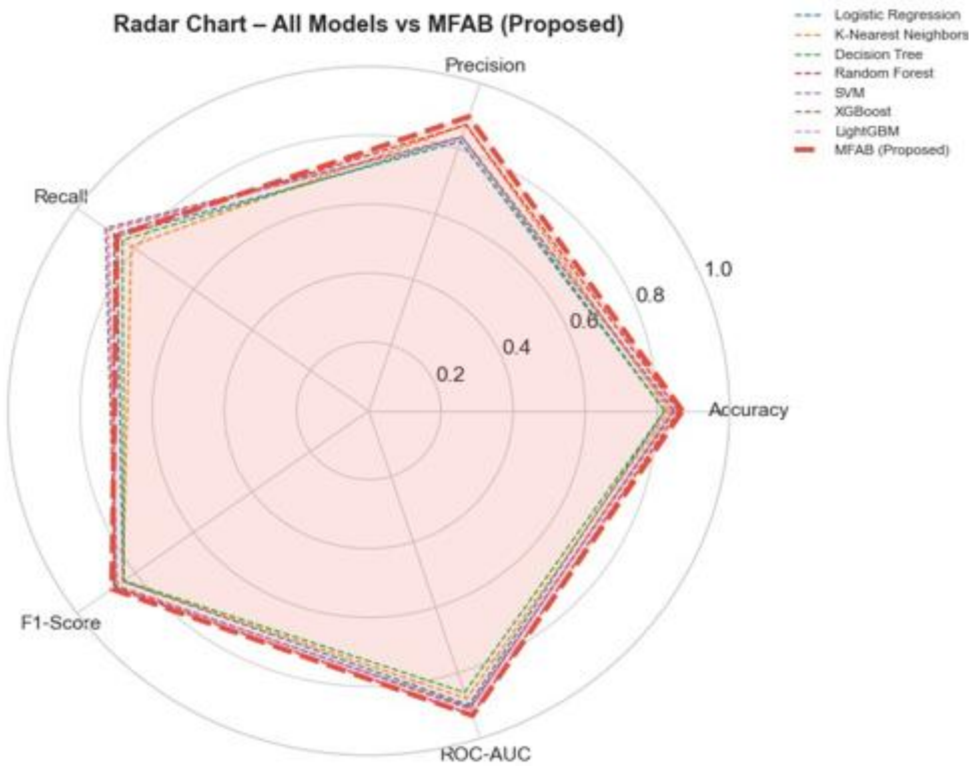
6.3 CONFUSION MATRIX

The Multi-fidelity Adaptive Boosting (MFAB) model is a novel ensemble framework that integrates data of varying precision (low and high fidelity) to enhance diagnostic accuracy. It utilizes an Adaptive Boosting mechanism to prioritize difficult-to-classify medical samples, ensuring the model learns complex patterns from noisy datasets. This combined approach allows for robust, personalized risk assessments in early heart disease detection.



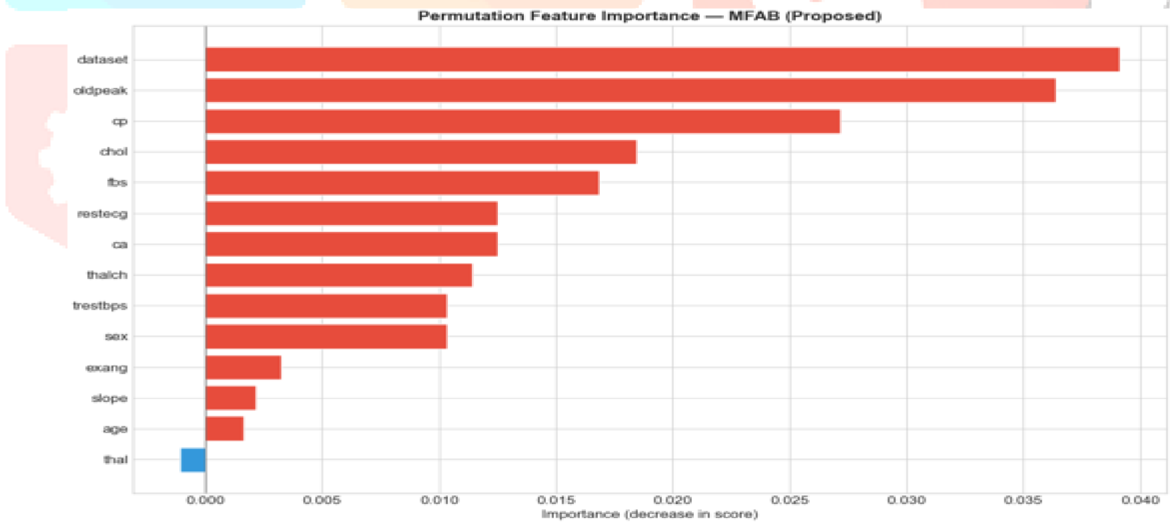
6.4 RADAR CHART

This radar chart compares the performance of various machine learning models across five metrics like Accuracy and F1-Score. The proposed MFAB model (thick red line) consistently occupies the outermost edge, indicating it achieves the highest scores across the board. While standard models like XGBoost and SVM perform well, MFAB demonstrates superior overall balance and effectiveness.



6.5 PERMUTATION CHART

This bar chart illustrates Permutation Feature Importance, which ranks variables based on how much the model's accuracy drops when their data is shuffled. Conversely, features at the bottom, such as that (shown in blue), have negligible or slightly negative impacts on the model's performance.



VIII.Future Work

Integration with Real-Time Data: Future systems can connect with wearable devices (smartwatches, fitness bands) Continuous monitoring of heart rate, BP, and other vitals for real-time prediction

Use of Advanced Algorithms: Implement Deep Learning models for higher accuracy Use ensemble and hybrid models for better prediction performance

Deployment as Web/Mobile Application: Convert the model into a user-friendly app Allow patients and doctors to easily access predictions anytime

Larger and Diverse Datasets: Train models on bigger datasets with diverse populations Improves generalization and reduces bias

Integration with Healthcare Systems: Connect with hospital databases and Electronic Health Records (EHR) Assist doctors in real-time clinical decision-making

Explainable AI (XAI): Provide clear explanations for predictions Helps doctors understand why a patient is at risk.

IX.Conclusion

This study demonstrates that machine learning can effectively predict heart disease. The system can assist doctors in early diagnosis and improve patient outcomes. Developed a machine learning-based system for heart disease prediction Successfully analysed medical data to identify key risk factors Compared multiple algorithms to select the best performing model Achieved reliable accuracy for early detection of heart disease The system can assist doctors in faster and better decision-making.

X.References

Meitao Zou, Mingrui Zhao, et al. Multi-Fidelity Machine Learning for Identifying Thermal Insulation Integrity: A Framework for Knowledge Transfer between Multi-Fidelity Data. Applied Sciences, 15(1):33, 2025.

W. Li and J. Chen. A multi-fidelity machine learning framework to enhance generalizability and leverage systematic errors in clinical datasets. MDPI Applied Sciences, 15(1):112–128, 2025.

A. K. Yadav, G. K. Pal, and S. Gangwar. Early-stage prediction of heart disease features using AdaBoost ensemble algorithm and tree algorithms. International Journal of Intelligent Systems and Applications in Engineering, 12(3):545–551, 2024.

Segun Akinola, Reddy Leelakrishna, and Vijayakumar Varadarajan. Enhancing cardiovascular disease prediction: A hybrid machine learning approach integrating oversampling and adaptive boosting techniques. AIMS Medical Science, 11(2):58–71, 2024.

H. M. S. Rani, et al. A technical comparative heart disease prediction framework using boosting ensemble techniques: Performance analysis of AdaBoost, XGBoost, and CatBoost. Computers, 12(1):15, 2024