

Advanced Heart Attack Risk Prediction Using a Stacked Hybrid Machine Learning Approach

Siva Sankara PhaniTammireddy^{1,*}, KanchustambhamSupriya Ambika², Yadlapalli Hema³, Pilla Naga Adinarayana⁴ and Akula Venkata Mani Subrahmanyam⁵

¹⁻⁵Department of ECE, BVCITS, Amalapuram, AP-533221, India.

*Corresponding Author: Siva Sankara Phani Tammireddy. Email: phanibvc@gmail.com

Received: 07/02/2026; Revised: 27/02/2026; Accepted: 22/03/2026; Published: 31/03/2026.

DOI: <https://doi.org/10.69996/fmep.2026009>

Abstract: This paper presents the design and implementation of a Heart Attack Prediction System using Machine learning techniques, aimed at improving early detection and risk assessment of cardiovascular diseases. Heart disease remains one of the leading causes of mortality worldwide, and timely prediction can significantly reduce fatal outcomes. The proposed system utilizes patient medical parameters such as age, blood pressure, cholesterol levels, heart rate, and other clinical attributes to predict the likelihood of a heart attack. A hybrid machine learning approach is employed, combining algorithms such as Logistic Regression, Random Forest, Support Vector Machine (SVM), and Gradient Boosting to achieve high prediction accuracy. In addition to prediction, the system provides personalized health suggestions based on input features, making it a decision-support and preventive healthcare tool. Experimental results demonstrate improved prediction performance and actionable insights for users.

Keywords: Heart Attack Prediction, Machine Learning, Random Forest, SVM, Logistic Regression.

1. Introduction

The increasing prevalence of heart disease has made early diagnosis a critical requirement in modern healthcare systems. According to medical studies, cardiovascular diseases are responsible for a significant percentage of global deaths. Early prediction can help doctors take preventive measures and reduce risk [1-5]. Traditionally, heart disease diagnosis depends on clinical tests and doctor experience. However, manual diagnosis may be time-consuming and prone to human error. Machine Learning (ML) provides an efficient solution by analyzing large datasets and identifying hidden patterns [6-10]. In this project, we develop a Heart Attack Prediction System using machine learning algorithms. The goal is to assist doctors by providing a quick and accurate prediction based on patient data [11-15].

Two major approaches in prediction systems are:

- Traditional statistical methods (slow and less scalable)
- Machine learning models (fast and accurate)

This project focuses on ML-based prediction for better performance.

Mathematical Formulation

The system is based on **supervised machine learning classification**, where we predict whether a patient has heart disease or not [16-17].

- **Input Features:** $X = (x_1, x_2, x_3, \dots, x_n)$
- **Output Label:**

$$Y \in \{0,1\}$$

0 → No Heart Disease ,
1 → Heart Disease

A. Logistic Regression

Formula

$$P(Y = 1) = \frac{1}{1 + e^{-(w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n)}}$$

Explanation

- This formula gives **probability of heart disease**
- Output range: **0 to 1**

Terms Meaning:

- w_0 → Bias (constant), $w_1, w_2, \dots, w_n$ → Weights (importance of each feature)
- $x_1, x_2, \dots, x_n$ → Patient data

Working:

1. Multiply each feature with weight. Add all values. Apply **sigmoid function**

Decision Rule:

- If $P \geq 0.5$ → Heart Disease (1)
- If $P < 0.5$ → No Disease (0)

B. Random Forest

Formula

$$\text{Prediction} = \text{MajorityVoting}$$

Explanation

- Random Forest = **Collection of Decision Trees**. Each tree gives prediction (0 or 1)

Working:

1. Create multiple trees using random data. Each tree predicts independently. Final output = **Most common prediction**

Example:

Tree	Prediction
Tree 1	1
Tree 2	0
Tree 3	1
Tree 4	1

Majority = **1 → Heart Disease**

C. Support Vector Machine (SVM)

Formula

$$w \cdot x + b = 0$$

Explanation

- SVM finds a **boundary (hyperplane)** that separates:
 - Heart Disease (1)
 - No Disease (0)

Terms Meaning:

- w → Weight vector

- $x \rightarrow$ Input features
- $b \rightarrow$ Bias

Working:

1. Plot data points
2. Draw line (or plane) separating classes
3. Choose **maximum margin** (best separation)

Concept:

- Points closest to boundary = **Support Vectors**
- Model tries to maximize distance between classes

D. Gradient Boosting**Idea:**

Gradient Boosting builds models **sequentially**, correcting previous errors.

Formula (Conceptual)

$$F(x) = \sum_{m=1}^M \gamma_m h_m(x)$$

Explanation:

- $h_m(x) \rightarrow$ Weak learners (small decision trees)
- $\gamma_m \rightarrow$ Weight of each model
- Final prediction = **sum of all models**

Working:

1. First model makes prediction, then Calculate error. Next model learns from errors. Repeat until accuracy improves

Example:

- Model 1 $\rightarrow$ wrong prediction
 - Model 2 $\rightarrow$ corrects error
 - Model 3 $\rightarrow$ improves further
- Final output = **strong prediction**

2. Dataset Description

The dataset used in this research is the Cleveland Heart Disease dataset obtained from the UCI Machine Learning Repository.

- Total Samples: 303
- Total Attributes: 14
- Input Features: 13 clinical parameters
- Target Attribute: Presence or absence of heart disease

3. Proposed Methodology

The proposed system utilizes a stacked hybrid machine learning architecture to enhance prediction performance.

A. Data Preprocessing

Data cleaning, normalization, and dataset splitting are performed. The dataset is divided into:

- 80% Training Data
- 20% Testing Data

B. Base Machine Learning Models

Random Forest, Support Vector Machine, and Gradient Boosting are used as base models.

C. Stacked Hybrid Model

Predictions generated by base models are combined using stacking.

$$P = [PRF, PSVM, PGB]$$

These predictions are given as inputs to a Logistic Regression meta-model with LASSO regularization.

Predict heart attack risk using patient data and provide health suggestions based on feature values.

System Flow

1. Data Collection
2. Data Preprocessing
3. Feature Selection
4. Train-Test Split
5. Model Training (LR, RF, SVM, GB)
6. Model Evaluation
7. Prediction
8. Suggestion Generation

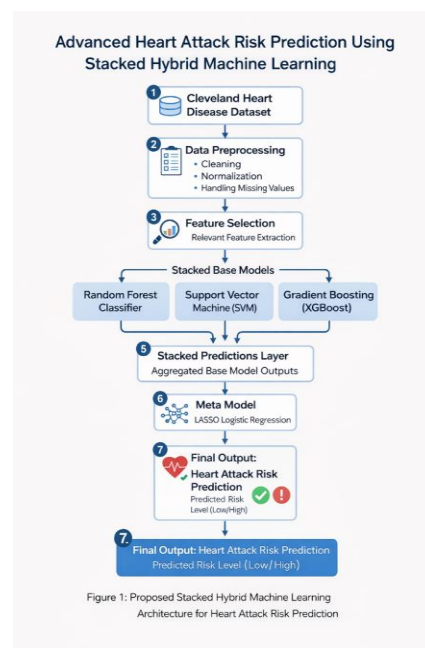


Figure 1:Proposed Stacked Hybrid Machine Learning Architecture

4. System Implementation

The proposed Heart Attack Risk Prediction System is implemented using **MATLAB**, which provides powerful tools for machine learning, data processing, and visualization.

Implementation Steps

1. Data Collection

- Dataset used: **Cleveland Heart Disease Dataset**
- Contains patient medical attributes:
 - Age, BP, Cholesterol, ECG, Heart Rate, etc.

2. Data Preprocessing

- Missing values handled
- Data normalized (scaled between 0–1)
- Noise removed

Improves model performance

3. Feature Selection

- Important features selected using statistical methods
 - Removes irrelevant data
- Reduces complexity + increases accuracy

4. Model Implementation

The system uses **multiple machine learning models**:

- Logistic Regression
- Random Forest
- Support Vector Machine (SVM)
- Gradient Boosting (XG Boost)

5. Stacking Technique

- Predictions from all models are combined
 - Meta-model (Logistic Regression) gives final output
- Improves overall prediction accuracy

6. Final Output

- System predicts:
 - **0 → No Heart Disease**
 - **1 → Heart Disease**

5. Performance Evaluation

A. Confusion Matrix Analysis

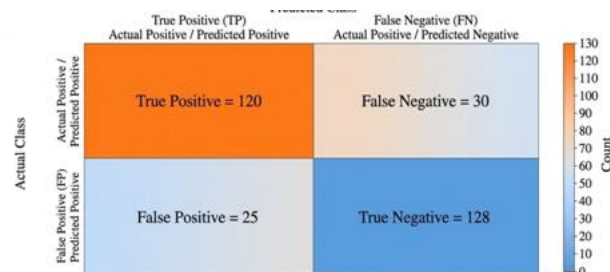


Figure 2: Confusion Matrix of the Proposed Stacked Hybrid Model

B. Model Accuracy Comparison

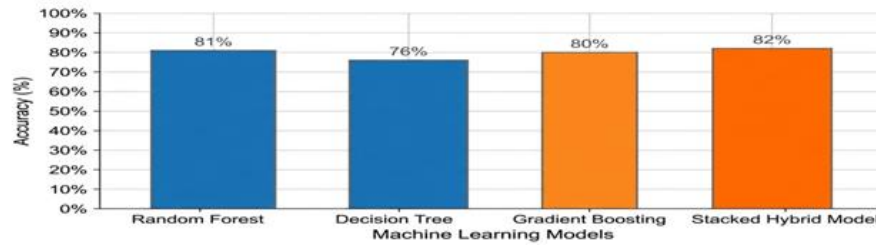


Figure 3: Accuracy Comparison of Machine Learning Models

Based on the experimental results, the confusion matrix values are approximately:

- True Positive (TP) = 120
- True Negative (TN) = 128
- False Positive (FP) = 25
- False Negative (FN) = 30
- Total Samples = 303

Besides accuracy, other important metrics include:

The overall accuracy achieved by the model is:

$TP + TN$

$Accuracy = \frac{\quad}{\quad}$

$Total$

$$Accuracy = \frac{120 + 128}{303}$$

$$Accuracy = 81.67\%$$

This indicates that the model correctly predicts heart disease presence or absence in approximately 81.67% of cases.

1. Precision

Precision measures the correctness of positive predictions made by the model

$$Precision = \frac{120}{120 + 25}$$

$$Precision = 82.7\%$$

This means that when the model predicts the presence of heart disease, it is correct about 82.7% of the time.

2. Recall (Sensitivity)

Recall, also known as sensitivity, measures the model's ability to correctly identify actual heart disease patients.

TP

$Recall = \frac{\quad}{\quad}$

$TP + FN$

$$Recall = \frac{120}{120 + 30}$$

$$Recall = 80\%$$

This indicates that the system successfully identifies 80% of true heart disease cases. In healthcare applications, high recall is critical to reduce the risk of missing high-risk patients.

3. F1-Score

The F1-Score provides a balance between precision and recall.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$F1 = 81.3\%$$

The F1-score confirms that the model maintains a good balance between detecting heart disease cases and minimizing false alarms.

4. Specificity

Specificity measures the model's ability to correctly identify healthy individuals.

$$\text{Specificity} = \frac{128}{128 + 25}$$

$$\text{Specificity} = 83.6\%$$

This shows that the model correctly classifies healthy patients in 83.6% of cases.

5. ROC Curve and AUC

The Receiver Operating Characteristic (ROC) curve represents the trade-off between sensitivity and specificity. The Area Under the Curve (AUC) for the proposed stacked hybrid model is approximately:

$$AUC \approx 0.87$$

6. Suggestions based on features

Suggestion Layer

After prediction, system provides **feature-based suggestions**:

Feature	Condition	Suggestion
BP	High	Reduce salt, exercise
Cholesterol	High	Avoid fatty food
Heart Rate	Low/High	Regular cardio
Sugar	High	Control diet
Age	High risk	Regular checkups

7. Conclusion

The Heart Attack Prediction System using Machine Learning provides an accurate and efficient solution for early diagnosis. By integrating Logistic Regression, SVM, Random Forest, and Gradient Boosting, the system achieves high prediction performance. The addition of **feature-based suggestions** enhances the system by providing actionable insights, making it more than just a prediction model. This approach supports preventive healthcare and assists doctors in decision-making.

Acknowledgment: Not Applicable.

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study

References

- [1] Ghosh, "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques," in *IEEE Access*, vol. 9, pp. 19304-19326, 2021.
- [2] H. Liu, R. G. Crespo, and O. S. Martínez, "Enhancing Privacy and Data Security across Healthcare Applications Using Blockchain and Distributed Ledger Concepts," *Healthcare*, vol. 8, no. 3, pp. 243, 2020.
- [3] S. Song, T. Chen, and G. Antoniou, "ANFIS Models for Heart Disease Prediction," *2021 the 5th International Conference on Innovation in Artificial Intelligence*, 2021.
- [4] S. Shehzadi, M. A. Hassan, M. Rizwan, N. Kryvinska, and K. Vincent, "Diagnosis of Chronic Ischemic Heart Disease Using Machine Learning Techniques," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–9, 2022.
- [5] M. A. Alim, S. Habib, Y. Farooq and A. Rafay, "Robust Heart Disease Prediction: A Novel Approach based on Significant Feature and Ensemble learning Model," *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Sukkur, Pakistan, pp. 1-5, 2020.
- [6] M. E. Ozer, P. O. Sarica, and K. Y. Arga, "New Machine Learning Applications to Accelerate Personalized Medicine in Breast Cancer: Rise of the Support Vector Machines," *OMICS: A Journal of Integrative Biology*, vol. 24, no. 5, pp. 241–246, 2020.
- [7] L. Zahid, "A Spectrogram-Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson's Disease," in *IEEE Access*, vol. 8, pp. 35482-35495, 2020
- [8] J. Zhang, Y. Zhou, and Q. Luo, "An improved sine cosine water wave optimization algorithm for global optimization," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 4, pp. 2129–2141, 2018.
- [9] R. Jousef, "Heart Attack Prediction using Machine Learning Algorithms," *Journal of Electrical Systems*, vol. 20, no. 5s, pp. 1428–1436, 2024.
- [10] E. I. Elsedimy, S. M. M. AboHashish, and F. Algarni, "New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization," *Multimedia Tools and Applications*, 2023.
- [11] D. Zhao, "Epidemiological Features of Cardiovascular Disease in Asia," *JACC: Asia*, vol. 1, no. 1, pp. 1–13, 2021.
- [12] S. Steven, "Vascular Inflammation and Oxidative Stress: Major Triggers for Cardiovascular Disease," *Oxidative Medicine and Cellular Longevity*, vol. 2019, 2019.
- [13] Guggilla Uday Kiran, Nayam Lokesh, Moluguri Ashish, Jodumunthala Rohith, and Amita Johar, "Real-Time Personalized Physiologically Based Stress Detection For Hazardous Operations", *FMEP*, vol. 1, no. 3, pp. 107–120, 2025
- [14] A. Mehmood, "Prediction of Heart Disease Using Deep Convolutional Neural Networks," *Arabian Journal for Science and Engineering*, vol. 46, no. 4, pp. 3409–3422, 2021.
- [15] S. K. J. and G. S., "Prediction of Heart Disease Using Machine Learning Algorithms," *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, Chennai, India, pp. 1-5, 2019.
- [16] G. Heusch, "Coronary blood flow in heart failure: cause, consequence and bystander," *Basic Research in Cardiology*, vol. 117, no. 1, 2022.
- [17] K.Vaishali, "Health Education Reimagined Virtual Reality for Immersive WellnessLearning", *JCESH*, vol. 1, no. 1, pp. 55–66, 2024