Enhancing Heart Attack Prediction with Machine Learning

Vikrant Singh Rajpoot
(Computer Science Engineering)
Shri Ram Institute Of Technology
Jabalpur, INDIA

Sneha Vishwakarma
(Computer Science Engineering)
Shri Ram Institute Of Technology
Jabalpur, INDIA

Shrishti Sahu
(Computer Science Engineering)
Shri Ram Institute Of Technology
Jabalpur, INDIA

Yash Kesharwani
(Computer Science Engineering)
Shri Ram Institute Of Technology
Jabalpur, INDIA

Abstract

Heart disease is a significant global health issue, with diagnosis often complicated by symptoms that overlap with other conditions. This study aims to address the challenges in early detection and diagnosis of heart disease by integrating various artificial intelligence techniques. The research utilizes machine learning classifiers such as random forest, SVM, decision tree, naive Bayes, and k-nearest neighbors (KNN), enhanced with particle swarm optimization (PSO) for feature selection. The goal is to improve the efficiency and accuracy of heart disease risk prediction. By incorporating real-time ECG data, the system offers a comprehensive view of the patient's cardiac health, enabling more informed medical decisions. This integration marks a pivotal step in cardiac healthcare, offering a robust tool for early detection and improved patient care.

Keywords— ECG(Electro Cardio Graph), PSO(Particle Swarm Optimization), KNN(K Nearest Neighbour), SVM(Support Vector Machine), HIPAA(Health Insurance Portability and Accountability Act)

I. Introduction

The heart attack detector is an innovative diagnostic tool that addresses the global challenge of heart disease, which remains a leading cause of mortality. It aims to overcome the diagnostic difficulties posed by the overlapping symptoms of various cardiac conditions.
Factors such as smoking, poor nutrition, high blood pressure, and sedentary lifestyles contribute significantly to heart health issues. Traditional diagnostic methods, including blood tests, X-rays, and echocardiography, are often time-consuming and costly. The heart attack detector circumvents these limitations by integrating multiple AI techniques, such as random forest, SVM, decision tree, naive Bayes, and KNN, with particle swarm optimization (PSO) for feature selection. This approach enhances the prediction accuracy of heart disease presence, utilizing a dataset from Jordan University Hospital comprising 486 patients.

This study represents a leap forward in medical technology, offering a promising solution to the urgent need for innovative and effective heart disease diagnostic tools.

II. Ease of Use

1. User-Friendly Interface: Design an intuitive and user-friendly interface for healthcare professionals to interact with the Heart Attack Detector. Include features like easy navigation, clear data visualization, and interactive elements for inputting patient data.

2. Automated Data Input: Implement features that allow for seamless integration of ECG reports into the system. This could involve automated data extraction from digital ECG systems or easy upload options for scanned ECG reports.

3. Real-Time Analysis: Enable real-time analysis of ECG data so that healthcare providers can receive immediate insights and recommendations. This can be particularly useful in emergency situations where quick decision-making is crucial.

4. Customizable Alerts and Notifications: Incorporate customizable alert systems that notify healthcare professionals about critical findings or potential heart attack risks based on ECG analysis. These alerts can be tailored to different severity levels for personalized patient care.

5. Interpretation Assistance: Provide features that assist in interpreting ECG findings, such as highlighting key abnormalities or providing context-specific explanations for complex medical terminology.

6. Security and Compliance: Implement robust security measures to protect patient data and ensure compliance with healthcare regulations such as HIPAA (Health Insurance Portability and Accountability Act). Prioritize data encryption, access control, and regular security audits.

III. RESEARCH METHODOLOGY

Let's delve into the literature view of the research paper, with a specific focus on advancements related to ECG (electrocardiogram) parameters.

1. Background and Related Works:
The research study is based on a heart disease dataset obtained from Jordan University Hospital (JUH) in Amman, Jordan.
The literature review section provides context by discussing terminology and related works.
Notable references include:
https://www.hindawi.com/journals/acisc/2024/5080332/
Received 22 September 2023; Revised 16 February 2024; Accepted 22 February 2024; Published 7 March 2024

Machine Learning Approaches to Heart Attack Risk Detection and Classification:
https://www.ijnrd.org/papers/IJNRD2403074.pdf

2. Advancements in ECG Report Parameters:

The research paper identifies several advancements related to ECG parameters:
Particle Swarm Optimization (PSO): By applying PSO algorithms to the dataset features, the study identified 19 features associated with heart disease. These features include:
Age
Sex

Family history of heart disease
Coronary diseases type
Patient's basic information
Patient's medical history
Reported symptoms
ECG findings

Future Directions:

Ensemble Methods: Combining multiple classifiers (e.g., SVM, random forest) could enhance accuracy and robustness against data fluctuations.
Multimodal Data Fusion: Integrating genetic information, wearable device data, and life style factors can provide a holistic risk profile.
Explainable AI (XAI): Interpretable models are crucial for clinical adoption. Techniques like feature importance analysis and saliency maps bridge the gap between AI and healthcare professionals.
Real-World Validation: Prospective studies in clinical settings validate model performance.
The research paper explores heart disease prediction using machine learning techniques, emphasizing the role of ECG parameters and potential advancements in this field.

IV. HYPOTHESIS

"Integrating real-time ECG data into an advanced Heart Attack Detector significantly enhances its accuracy in predicting and diagnosing heart attacks. By incorporating complex ECG signals using state-of-the-art machine learning techniques. This integration represents a pivotal step in cardiac healthcare, offering a robust tool for early detection and improved patient care."

The integration of multiple artificial intelligence techniques, including support vector machines (SVM), decision trees, naive Bayes, and k-nearest neighbors (KNN), can significantly enhance the accuracy and efficiency of heart disease diagnosis compared to traditional methods. Here's a detailed breakdown:

1. Support Vector Machines (SVM): SVM is a powerful classification algorithm that aims to find the optimal hyperplane to separate data points into different classes. By incorporating SVM, the heart attack detector can effectively handle complex feature spaces, making it robust in distinguishing between healthy and at-risk patients. SVM's ability to handle non-linear relationships in data contributes to improved heart attack prediction.

2. Decision Trees: Decision trees are intuitive and interpretable models that partition data based on feature values. By constructing decision trees, the detector can identify critical features (such as ECG abnormalities) that directly impact heart disease prediction. Decision trees also allow for feature importance ranking, aiding in clinical decision-making.

3. Naive Bayes: Naive Bayes is a probabilistic classifier based on Bayes' theorem. Incorporating Naive Bayes enables the detector to handle missing data and noisy features. It assumes independence between features, which simplifies computation and enhances efficiency.

4. K-Nearest Neighbors (KNN): KNN classifies data points based on their proximity to other data points. By considering the nearest neighbors, the detector can capture local patterns in ECG data. KNN's simplicity and adaptability make it valuable for heart attack prediction.

A. *Units*

1. ECG Signal Quality Indicator: Develop a unit that quantifies the quality of the ECG signal obtained from the patient. This unit could provide a score or rating based on factors like signal strength, noise level, and artifact presence. A higher score would indicate a clearer and more reliable ECG reading.

2. Risk Assessment Index: Create a unit that calculates a comprehensive risk assessment index based on various patient factors, including age, gender, medical history, reported symptoms, and ECG findings. This index could provide healthcare professionals with a quick overview of the patient's cardiac health status and potential risk of a heart attack.

3. Real-Time Monitoring Score: Implement a unit that generates a real-time monitoring score during ECG data analysis. This score could indicate the stability or instability of the patient's cardiac rhythm, helping clinicians identify any immediate concerns or abnormalities.

4. Treatment Recommendation Level: Design a unit that suggests appropriate treatment recommendations based on the heart attack risk assessment and ECG analysis. This unit could categorize treatment options into levels (e.g., low, moderate, high) based on the severity of the detected cardiac issues.

5. Patient Engagement Score: Develop a unit that evaluates the patient's engagement with their cardiac health management plan. This score could consider factors like medication adherence, lifestyle changes, follow-up appointments, and participation in cardiac rehabilitation programs.

These units would add value to the Heart Attack Detector system by providing actionable insights, simplifying decision-making processes, and promoting effective communication between healthcare providers and patients.

B. *Equations*

The confusion matrix can be used to calculate several different metrics that can be used to evaluate the performance of a classification model. Some of the most common metrics include the following.
1.Accuracy: The percentage of all correct predictions quantifies the ratio of correctly classified patients (TP + TN) to the total number of patients (TP + TN + FP + FN).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2.Precision: The percentage of correct optimistic predictions determines whether the model is reliable.

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. Recall: It attempts to answer the following question: What proportion of actual positives was identified correctly?

$$\text{Recall} = \frac{TP}{TP + FN}$$

4.F1-score: A harmonic means of precision and recall. It reaches a maximum when precision equals recall.

$$F1 - \text{Score} = \frac{1}{(1/\text{Recall}) + (1/\text{Precision})}.$$

C. *DATA INTERPRETATION*

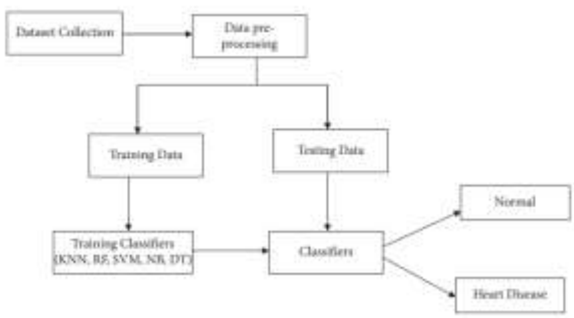

FIGURE 1: Proposed model for predictive analysis.

1. Load the Dataset
2. Perform data preprocessing
3. select the features based on values obtained after applying PSO
4. Split the data into training and testing sets
5. Define the machine learning models: SVM, Random Forest, Decision Tree, Naive Bayes, KNN
6. For each model:
   a. Perform cross-validation and hyperparameter optimization.
   b. Evaluate the model on the test set.
   c. Record the evaluation metrics (e.g., accuracy, Precision, recall, F1-score)
7. Compare the performance of the models based on the evaluation metrics
8. Select the best-performing model
9. Use the best-performing model for classification tasks on new, unseen data

FIGURE 2: Pseudocode representation of the proposed algorithm. A structured outline detailing the logic and steps of the algorithm designed to address a specific problem or task.

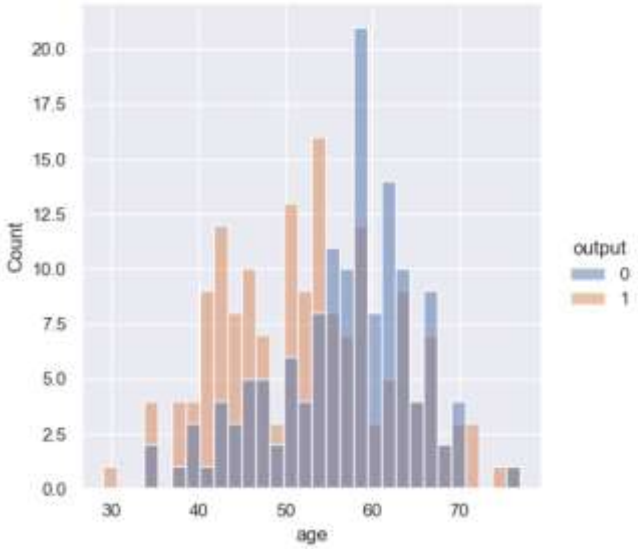

FIGURE 3: distribution of a dataset across different age groups.

The graph you provided visually represents the distribution of a dataset across different age groups, with two distinct outputs labelled as 0 and 1. Let's analyze and interpret this graph:

The x-axis represents age, ranging from 30 to 70 years. The y-axis represents the count of individuals in each age group. The dataset appears to be skewed toward older age groups. The two outputs (0 and 1) likely correspond to binary classification results (e.g., healthy vs. heart disease). The blue bars represent output 0, while the orange bars represent output 1. Around age 60, there is a significant peak in the blue bars (output 0). This peak suggests that a substantial number of individuals around age 60 are classified as healthy (output 0). Clinicians can pay special attention to patients around age 60, considering both preventive measures and early detection. ECG advancements may play a crucial role in understanding age-related cardiac health variations.

V. RESULT AND DISCUSSION

FIGURE 4: Heat map showing correlation.

These statistical visualizations confirm the diversity and variety in the dataset. Such a diverse array of values not only highlights the heterogeneous nature of the data but also emphasizes the complexity and multidimensionality of factors influencing our classification. This depth of variation serves as a testament to the dataset's capacity to identify trends and intricate patterns, laying the foundation for robust analysis and insightful interpretations during the exploration.

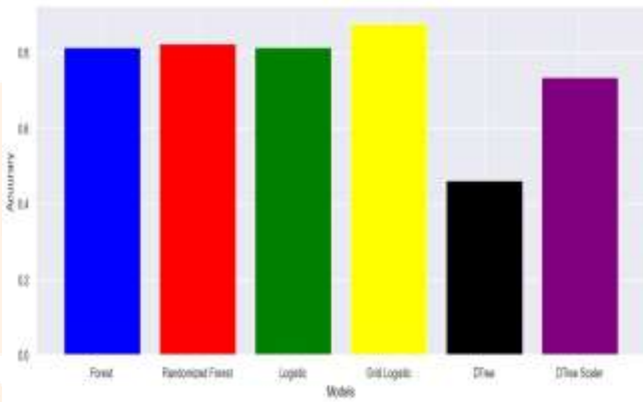

FIGURE 5: accuracy of different machine learning models.

The image you uploaded is a bar graph representing the accuracy of different machine learning models. Here are the details:

The graph displays the accuracy of various machine learning models. The y-axis represents accuracy, ranging from 0.0 to 0.8. The x-axis lists six different models like Forest, Randomized Forest, Logistic, Grid Logistic, D Tree, D Tree Scaler

Model accuracies:

Forest: Approximately 0.8 (indicated by a blue bar). Randomized Forest: Approximately 0.8 (indicated by a red bar). Logistic: Approximately 0.8 (indicated by a green bar). Grid Logistic: Approximately 0.8 (indicated by a yellow bar).| D Tree: Approximately 0.4 (indicated by a black bar). D Tree Scaler: Approximately 0.6 (indicated by a purple bar).

ACKNOWLEDGMENT

The completion of this research would not have been possible without the support and contributions of several individuals and organizations. We would like to express our sincere gratitude to **Shri Ram Group, Jabalpur** for providing access to the necessary resources and facilities essential for conducting experiments and analysis.

We extend our heartfelt appreciation to **Mr. Rajendra Arkh** for their invaluable guidance, encouragement, and expertise throughout the duration of this project. Their insightful feedback and constructive criticism significantly enriched the quality of this research.

Furthermore, we acknowledge the contributions of our colleagues and peers who provided valuable insights and suggestions during discussions and presentations.

Finally, we would like to express our gratitude to our families and friends for their unwavering support and understanding during the challenging phases of this endeavor…

References

[1] J. Ma, D. Lei, Z. Ren, C. Tan, D. Xia, and H. Guo, "Automated machine learning-based landslide susceptibility mapping for the three Gorges reservoir area, China," Mathematical Geo sciences, vol. 2, 2023.

[2] Z. Liu, J. Ma, D. Xia et al., "Toward the reliable prediction of reservoir landslide displacement using earthworm optimi zation algorithm-optimized support vector regression (EOA□SVR)," Natural Hazards, vol. 14, 2023.

[3] Z. Jia, J. Peng, Q. Lu et al., "A comprehensive method for the risk assessment of ground fssures: case study of the eastern weihe basin," Journal of Earth Sciences, vol. 34, no. 6, pp. 1892–1907, 2023.

[4] Y. Long, W. Li, R. Huang, Q. Xu, B. Yu, and G. Liu, "A comparative study of supervised classifcation methods for investigating landslide evolution in the Mianyuan River Basin, China," Journal of Earth Sciences, vol. 34, no. 2, pp. 316–329,2023.

[5] M. Rajabi, H. Golshan, and R. P. Hasanzadeh, "Non-local adaptive hysteresis despeckling approach for medical ultra sound images," Biomedical Signal Processing and Control, vol. 85, Article ID 105042, 2023.

[6] M. Rajabi and R. P. R. Hasanzadeh, "A modifed adaptive hysteresis smoothing approach for image denoising based on spatial domain redundancy," Sens Imaging, vol. 22, no. 1, p. 42, 2021.

[7] A. Borji, A. Seif, and T. H. Hejazi, "An efcient method for detection of Alzheimer's disease using high-dimensional PET scan images," pp. 729–749, 2023.