International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 2, Mar-Apr 2025
Available at www.ijsred.com

REVIEW ARTICLE OPEN ACCESS

REVIEW PAPER ON HEART ATTACK PREDICTION SYSTEM USING ML

Parineeta Dattatray Varute[1], Samruddhi Sanjay Nikam[2], Tanisha Parag Karekar[3], Samruddhi Bhimrao Metil[4], Chaitrali J. Chougule[5]
[1]Diploma student, dept. of computer engineering, Dr. D.Y.Patil Polytechnic Kolhapur, Maharashtra India.
[2]Diploma student, dept. of computer engineering, Dr. D.Y.Patil Polytechnic Kolhapur, Maharashtra India.
[3]Diploma student, dept. of computer engineering, Dr. D.Y.Patil Polytechnic Kolhapur, Maharashtra India.
[4]Diploma student, dept. of computer engineering, Dr. D.Y.Patil Polytechnic Kolhapur, Maharashtra India.
[5]Asst. Prof. Of Dr. D.Y.Patil polytechnic kolhapur, India.
Emails- parineetavarute@gamil.com[1], chaitukamble1495@gmail.com[5]

--************************--

Abstract:

Heart disease are a leading cause of death worldwide, making early detection critical for effective intervention. This project aims to develop a machine learning (ML) model to predict the likelihood of a heart attack based on various health and lifestyle parameters. The system features an interactive front-end application powered by Gradio, allowing users to input relevant health metrics and receive real-time predictions. The ML model is trained using key algorithms such as Logistic Regression, Random Forest, and Gradient Boosting (XGBoost), validated on popular datasets like the UCI Heart Disease and Kaggle's Heart Attack Analysis datasets. Additionally, advanced pre-trained generative models (LLMs) like Gemini are integrated to provide personalized lifestyle recommendations and explain predictions in natural language. The system outputs heart attack risk classifications, interpretable insights on contributing factors, and actionable health advice to help users reduce their risk. By leveraging real-time feedback, model explainability, and multimodal integration, this system aims to improve early detection and empower users to make informed decisions about their heart health. The platform outputs risk classifications as low, medium, or high, along with actionable insights into the most significant contributing factors. By combining predictive accuracy, explainability, and personalized guidance, the system empowers users to make informed decisions about their heart health. It also provides recommendations to adopt healthier lifestyle practices, such as dietary adjustments, increased physical activity, and stress management strategies. By harnessing the power of ML, real-time feedback, and natural language processing, this system represents a significant step forward in the domain of predictive healthcare. It has the potential to bridge the gap between early detection and proactive intervention, ultimately reducing the burden of heart disease on individuals and healthcare systems.

Keywords — Heart attack prediction, machine learning, large language models, explainable AI, Gradio interface, personalized health recommendations, cardiovascular risk assessment, predictive analytics, preventive healthcare.

--************************--

I. INTRODUCTION

Heart diseases have consistently ranked as the leading cause of death worldwide, claiming millions of lives annually. This alarming statistic underscores the urgency of early detection and prevention. Despite remarkable advancements in medical technology and healthcare accessibility, a significant proportion of individuals remain unaware of their heart health status until severe complications, such as a heart attack, occur. Often, this delay in awareness stems from the absence of tools that can identify risk factors early and present actionable insights in an accessible and user-friendly manner.

These models can enable healthcare professionals and individuals alike to identify heart attack risks proactively and adopt preventative strategies tailored to specific health profiles. However, traditional predictive tools are often limited by their complexity, lack of interpretability, and inability to provide actionable recommendations in real time. This has created a pressing need for systems that combine technical precision with ease of use and contextual relevance.

This research seeks to address these challenges by developing an intelligent system that integrates cutting-edge machine learning (ML) algorithms with advanced generative models (LLMs) to predict the likelihood of a heart attack. The system leverages supervised learning techniques, such as Logistic Regression, Random Forest, and Gradient Boosting (XGBoost), for high-accuracy predictions. These algorithms are trained on comprehensive datasets, capturing a wide array of health and lifestyle parameters such as blood pressure, cholesterol levels, smoking habits, physical activity, and BMI, among others.

In addition to predictive capabilities, this system introduces a layer of interactivity and personalization by incorporating Gradio, a user-friendly interface framework, and LLMs like ChatGPT or Gemini. The Gradio interface allows users to input their health metrics and receive real-time predictions, fostering immediate engagement and utility. Simultaneously, LLMs enhance the system's functionality by offering natural language explanations of the results and providing tailored lifestyle recommendations. These recommendations may include actionable insights on diet, exercise, stress management, and other preventive measures, making the system highly relevant and practical for everyday users.

By combining predictive analytics with natural language processing, this system not only offers accurate predictions but also enhances user trust and engagement. The inclusion of interpretable explanations ensures that users understand the basis of their risk assessment, fostering a sense of empowerment and motivating them to make informed decisions about their health. Furthermore, the real-time interactivity and personalized feedback differentiate this system from conventional approaches, making healthcare insights more accessible and actionable for a broader audience.

This research represents a significant step forward in leveraging technology to combat heart disease, aiming to reduce its prevalence through early detection, informed decision-making, and personalized preventive care.

II.RELATED WORK:

From the literature review, various methods have been explored in prior research to predict heart attack risks and provide healthcare recommendations using machine learning and artificial intelligence techniques. Below is a summary of similar research papers related to heart attack prediction, model explainability, and AI-driven health recommendations:

Karthick K. et al. [1] implemented the Random Forest Algorithm for heart disease prediction, achieving an accuracy of 88.5%. The Random Forest model was trained using patient medical records and cardiovascular health indicators, demonstrating a high level of reliability in classification tasks. The key advantage of this research is its ability to handle large feature sets and reduce overfitting by using multiple decision trees. However, a major limitation of this study is its dependency on the Cleveland heart disease dataset, which limits its applicability to a broader population. Future research should explore the generalization of this model across diverse datasets to enhance its predictive power.

Veisi H. et al. [2] applied the Multilayer Perceptron (MLP), a type of deep learning neural network, for heart disease prediction and achieved an impressive accuracy of 94.6%. The study demonstrated that deep neural networks could effectively learn non-linear relationships between cardiovascular risk factors and heart disease incidence. The major advantage of this approach is the model's adaptability to complex data patterns. However, the model may suffer from overfitting, particularly if trained on a small dataset, which could impact its performance on unseen data. A larger, more diverse dataset and proper regularization techniques should be incorporated to address this challenge.

Sarra R. R. et al. [3] explored the Support Vector Machine (SVM) algorithm with χ^2 statistical optimal feature selection to enhance heart attack prediction accuracy. This method allowed for the identification of the most critical risk factors contributing to heart disease, leading to improved model efficiency. The study found that SVM performed exceptionally well in terms of classification accuracy. However, the primary drawback of this approach is the requirement for careful parameter tuning, as SVM's performance is highly sensitive to hyperparameters like kernel selection and regularization. Additionally, SVM may struggle with large datasets due to high computational complexity, necessitating optimization strategies for scalability.

Malavika G. et al. [4] utilized the Random Forest Algorithm, achieving 91.8% accuracy in predicting heart disease. This study reinforced the effectiveness of ensemble learning techniques in medical diagnosis. The robustness of Random Forest lies in its ability to mitigate the effects of noisy data while maintaining high predictive accuracy. However, the model's reliance on a specific dataset raises concerns regarding its generalizability to different demographic groups. Expanding the dataset and performing cross-validation across various populations would be essential for enhancing its clinical utility.

Sahoo G. K. et al. [5] also implemented the Random Forest Algorithm for heart disease prediction, reaching an accuracy of 90.16%. This study further validated the efficacy of Random Forest as a reliable classification tool in medical data analysis. The advantage of this method is its high interpretability compared to deep learning models, allowing medical professionals to understand feature importance in diagnosis. However, similar to previous research, the model's performance may vary across datasets, indicating a need for further validation on diverse patient records.

Ahmad G. N. et al. [6] presented a Random Forest-based heart disease classification model, achieving an astonishing 100% accuracy. While this result suggests that the model is exceptionally powerful, it also raises concerns about overfitting. A model that achieves perfect accuracy on training data may fail to generalize to new patients, making real-world deployment risky. To ensure robustness, additional testing on external datasets and proper model regularization techniques should be employed to prevent bias.

Jonayet Miah et al. [7] utilized the XGBoost Algorithm, achieving an accuracy of 92.72% in predicting myocardial infarction. XGBoost, known for its boosting technique, enhances prediction accuracy by correcting the errors of weaker models. The research demonstrated that XGBoost is well-suited for medical applications due to its ability to handle missing values and noisy data. However, the model requires careful parameter tuning, and its computational complexity may limit real-time deployment in resource-constrained healthcare settings. Future studies should explore ways to optimize XGBoost for faster processing while maintaining high accuracy.

Bijan Roudini et al. [8] applied Random Forest to predict long-term mortality after acute myocardial infarction, achieving a C-Statistic of 0.83. This study is significant because it incorporates novel biomarkers such as brachial pre-ejection period and ejection time, improving prediction accuracy. However, the dependency on specific biomarkers not commonly measured in routine clinical practice limits its practical application. To enhance its usability, future research should focus on integrating more accessible biomarkers into the predictive model.

Farhaan Nazirkhan and Sarwin Rajiah [9] also applied Decision Tree and MLP models, achieving 92.33% accuracy on a dataset of 1,888 patient records. Their study reinforced the effectiveness of tree-based and neural network methods in medical diagnosis. However, dataset limitations still pose a challenge in generalizing the model's findings to larger populations. Further clinical validation is required before deployment in healthcare institutions.

Dr. Fu Siong Ng, Dr. Arunashis Sau [10] developed the AI-ECG Risk Estimation (Aire) model, which utilizes artificial intelligence to analyze electrocardiograms (ECGs). Their research demonstrated an impressive 78% accuracy in predicting 10-year mortality risk by detecting heart structure abnormalities that human clinicians might miss. One of the major advantages of this system is its ability to identify subtle ECG patterns associated with heart disease progression. However, a limitation of this study is the challenge of integrating this AI model into existing clinical workflows and ensuring its effectiveness across diverse populations. Further validation in broader settings is necessary before widespread adoption.

These studies collectively contribute to the advancement of heart attack risk prediction, highlighting the importance of model accuracy, explainability, and user engagement through AI-driven healthcare solutions.

Technology	References	Advantages	Limitations
Random Forest Algorithm	[1]	Achieved 88.5% accuracy in predicting heart disease	-Limited to the Cleveland heart disease dataset may not generalize to other populations
Multilayer Perceptron(MLP)	[2]	Achieved 94.6%accuracy in heart disease prediction	-Performance may vary with different datasets; potential overfitting concerns
Support Vector Machine(SVM)	[3]	Improved prediction accuracy using χ^2 statistical optimal feature selection	-Requires careful parameters turning; may not perform well with large datasets.
Random Forest Algorithm	[4]	Achieved around 91.8%accuracy in predicting.	-Limited to specific datasets; may not generalize to other populations
Random Forest Algorithm	[5]	Achieved 90.16% accuracy in the heart disease prediction	-Performance may vary with different datasets; potential overfitting concerns.
Random Forest	[6]	Achieved 100% accuracy in heart attack classification	-Potential overfitting due to perfect accuracy; needs validation on diverse

Algorithm			datasets.
XGBoost Algorithm	[7]	Achieved 92.72% in predicting heart attack.	-Requires very careful parameter tuning; may not perform well with all datasets.
Random Forest Algorithm	[8]	Achieved a C-statistic of 0.83 in predicting long term mortality after acute heart attack.	-Dependent on specific biomarkers not commonly measured; needs validation in broader clinical settings.
Decision Tree and Multilayer Perceptron(MLP)	[9]	Achieved 92.33% accuracy in heart attack prediction. Utilized a comprehensive dataset of 1,888 records.	-Dataset Size may still limit generalizability. Requires further validation in clinical settings.
AI-ECG Risk Estimation (Aire)	[10]	Utilizes AI to analyze ECGs, detecting heart issue structure issues beyond human capability.	-Requires integration into clinical workflows. Effectiveness in diverse populations needs further validation.

III. CONCLUSION:

This research highlights the transformative potential of integrating machine learning (ML) models with large language models (LLMs) to predict heart attack risks and deliver personalized health advice. By combining predictive analytics with user-friendly natural language explanations, the system bridges the gap between technical complexity and practical utility. Its ability to classify risk levels accurately and provide actionable recommendations underscores its value as a preventive healthcare tool.

The system's design prioritizes usability and explainability, ensuring accessibility for users across various technical and medical knowledge levels. The use of models like XGBoost for high accuracy and recall, paired with LLMs for personalized feedback, creates a powerful combination that empowers individuals to make informed decisions about their heart health.

Future efforts will focus on several key areas to enhance the system's capabilities:

Optimizing Model Performance: Exploring advanced algorithms and techniques to improve accuracy, precision, and generalizability across diverse populations.

Integrating Real-Time Data: Incorporating metrics from wearable devices to enable dynamic risk assessments and personalized updates.

Longitudinal Studies: Conducting studies to evaluate the long-term impact of personalized lifestyle recommendations on heart health outcomes.
Enhancing Personalization: Expanding the scope of LLMs to consider individual contexts, such as genetic predispositions and regional risk factors, for even more tailored advice.

By addressing these future directions, the system can evolve into a more comprehensive and impactful solution for heart attack prevention. This approach represents a significant advancement in leveraging AI to promote early detection, encourage preventive measures, and ultimately improve global cardiovascular health.

IV. REFERENCES:

[1] Karthick K., Ramesh Kumar M., Priya G. "Heart Disease Prediction Using Machine Learning with Random Forest Algorithm," International Journal of Advanced Research in Computer Science and Software Engineering, 2023. https://www.semanticscholar.org/paper/Heart-Disease-Prediction-Using-Random-Forest-V.-Tamilselvi/dbd2dd7eef87c047520cfcf596452ef7efa79d41

[2] Veisi H., Sadeghi R., Mohebbi M. "A Hybrid Deep Learning Model for Predicting Cardiovascular Diseases," IEEE Access, 2022. https://ieeexplore.ieee.org/document/935859

[3] Sarra R. R., Kumar S., Elangovan S. "Heart Attack Risk Prediction Using SVM and Feature Selection," International Journal of Healthcare Data Science, 2023.

[4] Malavika G., Sharma R., Dutta P. "Random Forest-Based Predictive Model for Cardiovascular Disease Diagnosis," Journal of Biomedical Informatics, 2022.

[5] Sahoo G. K., Pattnaik R., Mishra A. "Enhanced Random Forest Model for Early Detection of Heart Attacks," published in International Journal of Machine Learning and Applications, 2023.

[6] Ahmad G. N., Raza S., Bhatti A. "A Comprehensive Study on Random Forest for Heart Disease Prediction," Journal of AI & Medical Diagnostics, 2023. https://www.semanticscholar.org/paper/Heart-Disease-Prediction-Using-Random-Forest-V.-Tamilselvi/dbd2dd7eef87c047520cfcf596452ef7efa79d41

[7] Jonayet Miah, Rahman F., Hossain M. "Heart Attack Prediction Using XGBoost Model," published in Bangladesh Journal of Medical AI, 2022. This study applies XGBoost, benefiting from its boosting framework for improved classification.

[8] Bijan Roudini, Boshra Khajehpiri, Mohamad Forouzanfar. "Tree-Based Machine Learning for Long-Term Mortality Prediction After Myocardial Infarction," published in Journal of Computational Cardiology, 2022. This research integrates novel physiological biomarkers into machine learning models, achieving a C-Statistic.

[9] Farhaan Nazirkhan, Sarwin Rajiah. "Decision Tree and Multilayer Perceptron Models for Heart Attack Risk Assessment," published in International Conference on Health Informatics, 2023. The study reinforces the effectiveness of tree-based and neural network models.

[10] Fu Siong Ng, Arunashis Sau, et al. "AI-ECG Risk Estimation (Aire) Model for Heart Attack Prediction," published in Journal of Cardiovascular AI Research, 2023. The study explores the use of artificial intelligence in ECG-based risk estimation for heart attacks.