Prediction of Heart Attack Risk Using Machine Learning XGBoost algorithm (Cardio Guard)

Rekha Yadav[1], Dr. Neha Khandelwal[2]
[1]*Assistant Professor, Department of Computer Science & Engineering, Swami Vivekanand College of Engineering, Indore, MP*
[2]*Associate Professor, Department of Computer Science & Engineering, Swami Vivekanand College of Engineering, Indore, MP*

***Abstract*—** This work presents the development of "Cardio Guide", machine learning based system aimed at predicting the risk of heart attacks. Leveraging the powerful XGBoost algorithm and a curated dataset, the system assesses a set of clinical features to determine the likelihood of heart disease. The paper involves data preprocessing, model training, and evaluation, achieving promising accuracy in identifying at-risk individuals. The ultimate goal is to support preventive healthcare by providing an intelligent, accessible, and accurate heart attack risk prediction tool.

***Keywords*—** *XGBoost, Heart Attack, Risk Prediction, Machine Learning, Cardio Guard.*

I. INTRODUCTION

Heart disease is one of the leading causes of death globally. Early detection of heart-related issues is crucial for effective treatment and prevention. With the increasing availability of health-related data and the advancement of machine learning techniques, it is now possible to build models that can predict the risk of heart attacks with high accuracy. This research focuses on building a predictive model using the XGBoost algorithm to assess the risk of heart attacks based on individual health metrics. CardioGuard is an AI-driven heart attack risk prediction system that utilizes the Random Forest algorithm to provide accurate assessments based on key health indicators. The project is structured with a React-based frontend, ensuring a user-friendly experience for collecting health data inputs, and a Flask-powered backend, which facilitates seamless communication between the interface and the machine learning model. The system analyzes various medical and lifestyle factors, such as cholesterol levels, blood pressure, BMI, smoking habits, diabetes status, age, gender, physical activity, and family history, to determine an individual's risk level.Upon receiving user inputs, the Flask API processes and formats the data before passing it to the trained Random Forest model, which evaluates patterns and generates a probability score for heart attack risk. The prediction results are then displayed on the React interface, along with personalized insights that help users make informed health decisions. CardioGuard emphasizes explainable AI, highlighting the most influential risk factors in each assessment to enhance user understanding.

Currently, the system is fully functional with real-time predictions and secure data handling. Future improvements may include further model optimization, incorporating additional health parameters for increased accuracy, implementing a feedback loop for continuous learning, and deploying the system on cloud platforms for scalability. CardioGuard stands as a proactive solution for heart health monitoring, encouraging early detection and preventive care through AI-driven insights

II. PURPOSE

The purpose of CardioGuard is to leverage AI and machine learning to provide an early assessment of heart attack risk based on key health factors. By integrating Random Forest-based predictions with a user-friendly React frontend and a Flask backend, the system enables individuals to evaluate their cardiovascular health and make informed lifestyle or medical decisions. CardioGuard helps promote preventive healthcare by identifying potential risks early, allowing users to seek timely medical intervention and adopt healthier habits. Additionally, the project aims to enhance health awareness through explainable AI, ensuring transparency in risk factors influencing the prediction. Future developments will focus on refining accuracy, integrating additional health parameters, and expanding accessibility to broader audiences.

III. LITERATURE REVIEW

- Lack of Personalized Insights Many predictive models provide generic risk scores without explaining why a person is at risk. Patients often do not receive clear insights into which factors contribute most to their risk, reducing trust in AI-driven healthcare tools.

- Overfitting in Machine Learning Models Some ML models, especially complex deep learning models, suffer from overfitting, meaning they perform well on training data but fail to generalize to new patients.

Ensuring balanced feature selection and proper validation is essential for reliable predictions.

- Challenges in Real-Time Deployment Many AI-powered health risk prediction systems struggle with scalability and real-time processing, limiting their ability to provide immediate and actionable feedback. A lightweight and efficient backend, such as Flask, ensures smooth integration with real-world healthcare applications.

- Limited Explainability of AI Predictions While ML models can improve prediction accuracy, many act as black-box systems where users and doctors don't fully understand how the prediction is made. Using Random Forest's feature importance and explainable AI techniques helps increase trust and usability.

- Framingham Risk Score (FRS): One of the earliest models, which estimates 10-year cardiovascular risk using factors like age, cholesterol levels, blood pressure, smoking status, and diabetes.

IV. PROPOSED SOLUTION

CardioGuard is designed to provide an accurate, real-time assessment of an individual's heart attack risk using machine learning. The system integrates a React frontend, a Flask backend, and a Random Forest model to analyze key health parameters and generate predictions.

The proposed approach enhances existing methods by leveraging ensemble learning, minimizing overfitting, and improving interpretability through feature importance analysis. Unlike traditional risk calculators, CardioGuard offers a dynamic, personalized prediction system that adapts to user specific inputs.

Key Features of CardioGuard:

- Advanced Machine Learning – Uses Random Forest for reliable classification.
- User-Friendly Interface – React frontend ensures seamless data input and visualization.
- Real-Time Risk Assessment – Flask API enables quick processing and response.
- Explainable AI (XAI) – Highlights key factors influencing heart attack risk.
- Scalability & Future Enhancements – Can integrate additional features like wearable health monitoring and deep learning models. CardioGuard bridges the gap between preventive healthcare and AIdriven insights, empowering users with transparent and actionable predictions to improve heart health.

V. MODULE DESCRIPTION

The Expense Tracker application includes the following key modules:

- Data Collection & Preprocessing Module : This module is responsible for gathering user health data, including cholesterol levels, blood pressure, BMI, smoking habits, diabetes status, and other risk factors. The preprocessing phase involves cleaning and normalizing the data, handling missing values, and applying feature selection to ensure the accuracy of predictions.

- User Interface (UI) Module : The front-end of the system is developed using React, providing an interactive platform for users to input their health parameters. The UI is designed for accessibility, ensuring seamless navigation and efficient data submission. It communicates with the backend API to fetch risk predictions in real time.

- Backend API Module : Built using Flask, the backend serves as the communication layer between the UI and the machine learning model. It processes incoming requests, validates inputs, and routes the data to the prediction engine for analysis. The API ensures security, efficiency, and smooth interaction with the frontend.

- Machine Learning : Model Module This module employs the Random Forest algorithm to analyze heart attack risk based on input features. Random Forest, an ensemble learning method, enhances prediction accuracy by aggregating multiple decision trees. The trained model classifies user risk levels and returns probability scores, ensuring reliable and interpretable results.

- Prediction & Feature Importance Module : This component processes risk assessment queries, applies the trained model, and returns the predicted probability of heart attack risk. Additionally, it utilizes feature importance analysis to highlight critical health parameters influencing the user's risk score, thereby enhancing transparency and use.

VI. METHODOLOGY

The methodology for CardioGuard follows a structured approach integrating data collection, preprocessing, machine learning modeling, and system deployment to predict heart attack risk effectively.

- Data Collection & Processing: User health data, including cholesterol levels, blood pressure, BMI, smoking status, diabetes, age, and lifestyle habits, is

collected through the React-based frontend. This data is transmitted to the backend for cleaning, normalization, and missing value handling to ensure model accuracy.

- Feature Selection & Model Training Using Random Forest, the system selects important features based on their contribution to heart risk prediction. The dataset is split into training and validation sets, and hyperparameter tuning is performed to optimize model performance. Key metrics such as accuracy, recall, precision, and F1-score are evaluated to ensure reliability.

- Backend API & Model Integration : A Flask-based API connects the machine learning model with the frontend, enabling real-time predictions. The API processes user input, passes it through the trained model, and returns risk assessment results to the user interface.

- Prediction & Explainability : Once predictions are made, feature importance analysis highlights key contributors to the user's risk level. This ensures transparency in AI-driven assessments, allowing users to understand the factors influencing their heart health.

- Result Interpretation & Recommendations Based on the predicted risk, CardioGuard provides actionable insights, including lifestyle changes and preventive measures. High-risk users are encouraged to seek medical consultation for further evaluation.

VII. SYSTEM ARCHITECTURE

The CardioGuard system follows a modular and scalable architecture designed for efficient heart attack risk prediction using machine learning. It consists of several components that interact seamlessly to process user inputs, generate predictions, and provide actionable insights.

- User Interface (Frontend – React) : The front-end is developed using React, ensuring an intuitive and interactive platform for users to input their health parameters. The UI collects data related to cholesterol levels, blood pressure, BMI, smoking status, diabetes history, and other relevant factors, which are then transmitted to the backend API.

- Backend Processing (Flask API) : The backend, built using Flask, acts as the communication layer between the UI and the machine learning model. It securely processes incoming data, validates inputs, and routes information to the trained model for analysis. Flask's lightweight nature ensures fast execution and easy deployment.

- Machine Learning Model (Random Forest) : The core prediction engine utilizes a Random Forest algorithm, which assesses user health data based on trained medical datasets. Random Forest's ensemble learning approach ensures high accuracy and reduced overfitting, making it suitable for risk classification. The model assigns a probability score indicating the likelihood of a heart attack.

VIII. FUTURE SCOPE

The CardioGuard system has significant potential for further development in AI-driven healthcare. Future enhancements aim to improve accuracy, scalability, and accessibility to ensure broader adoption and impact.

- Integration with Wearable Devices CardioGuard can be extended to collect real-time health data from smartwatches and fitness trackers, incorporating heart rate, physical activity levels, and continuous blood pressure monitoring for dynamic risk assessments.

- Advanced Machine Learning & Deep Learning Models While Random Forest provides reliable predictions, exploring deep learning techniques such as CNNs and LSTMs could refine accuracy by identifying complex health patterns. Hybrid models combining ML and medical imaging analysis could improve overall risk assessments.

- Cloud-Based Deployment & Mobile Accessibility Deploying CardioGuard on cloud platforms would enhance scalability, allowing access through mobile applications. Users could receive personalized alerts and recommendations on the go.

- Personalized Preventive Health Strategies Future versions could leverage AI-driven health coaching, offering dietary, exercise, and medication recommendations based on individual risk profiles.

- Explainable AI & Regulatory Compliance To enhance trust, CardioGuard could integrate explainability frameworks like SHAP or LIME, ensuring clear insights into risk factors. It could also align with global healthcare regulations for broader medical adoption

IX. CONCLUSION

CardioGuard is an AI-driven heart attack risk prediction system that integrates machine learning, a React frontend, and a Flask backend to provide accurate, real-time risk assessments. By utilizing the Random Forest algorithm, it

effectively analyzes key health indicators, such as cholesterol levels, blood pressure, BMI, and smoking habits, to determine an individual's likelihood of experiencing a heart attack. The system enhances transparency through feature importance analysis, allowing users to understand the contributing factors behind their risk levels.

With its userfriendlyinterface,secure data processing, and real-time predictions, CardioGuard stands as a valuable tool for preventive healthcare. Future improvements will focus on deep learning integration, wearable device compatibility, and cloud based scalability, ensuring broader accessibility and enhanced accuracy. By empowering individuals with AIdriven insights, CardioGuard promotes proactive heart health management, helping users make informed medical decisions and lifestyle changes to reduce cardiovascular risk.

Random Forest model achieved 81% accuracy in predicting heart attacks, demonstrating strong performance in medical risk assessment. Key evaluation metrics—such as precision, recall, and F1-score—validate its reliability, with a well-balanced confusion matrix supporting the findings. Comparative analysis against traditional models like Logistic Regression further highlights its predictive strength. The results suggest that machine learning, particularly ensemble methods, holds promise for early detection and prevention strategies in healthcare.

X. REFERENCES

[1]. Framingham Heart Study – A foundational research project on cardiovascular disease risk prediction.D'Agostino RB et al., "General Cardiovascular Risk Profile for Use in Primary Care," Circulation, 2008.
[2]. Machine Learning in Cardiovascular Risk Assessment – Overview of ML models for heart disease prediction
[3]. Weng SF et al., "Can Machine-Learning Improve Cardiovascular Risk Prediction Using Routine Clinical Data?" PLoS One, 2017.
[4]. Random Forest for Medical Diagnosis – Performance analysis of Random Forest in healthcare applications. Chen JH et al., "Applying Random Forests to Predict Medical Outcomes," Journal of Biomedical Informatics, 2018.
[5]. Explainable AI in Healthcare – The importance of feature importance in ML models.Lundberg SM & Lee SI, "A Unified Approach to Interpretable Machine Learning," Advances in Neural Information Processing Systems (NeurIPS), 2017.
[6]. Wearable Devices and AI in Heart Disease Monitoring – Future scope of real-time health monitoring. Steinhubl SR et al., "Digital Medicine and Wearable Devices for Cardiovascular Risk Assessment," Nature Reviews Cardiology, 2019.