International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

HEART ATTACK PREDICTION SYSTEM USING MACHINE LEARNING TECHNIQUE

1. **Varun M**
PG Student, Department of MCA, Dayananda Sagar College of Engineering, Bangalore
Varun.mithyantha@gmail.com

2. **Smitha G V**
Assistant professor, Department of MCA, Dayananda Sagar College of Engineering, Bangalore
Smitha-mca@dayanandasagar.edu

Abstract:

Cardiovascular disease is currently the leading cause of death in high-income countries and is projected to become the top global cause of death by 2030. In the United States, heart disease, including conditions like coronary heart disease, hypertension, and stroke, remains the primary cause of death. According to the American Heart Association's 2019 update, approximately 46% of US adults, equivalent to 116.4 million people, are estimated to have hypertension based on the new 2017 Hypertension Clinical Practice Guidelines.

The impact of cardiovascular disease is profound, with a person dying from CVD every 38 seconds, resulting in approximately 2,303 CVD-related deaths per day based on 2016 data. Stroke, as a critical cardiovascular event, accounts for about 389.4 deaths daily, with a stroke-related death occurring every 3.70 minutes, according to the same data.

The heart's essential function of pumping blood throughout the body makes any malfunction potentially fatal, leading to organ failure, including the brain, within minutes. Contributing factors to the rising incidence of heart-related diseases include unhealthy lifestyle choices, work-related stress, and poor dietary habits, posing a significant global health concern. According to the World Health Organization, heart-related diseases cause approximately 17.7 million deaths annually, representing 31% of all global deaths. This worrisome trend is also apparent in India, where heart-related diseases have become the leading cause of mortality.

To combat the increasing rates of heart-related diseases, particularly among younger individuals, it is vital to develop an early detection system for heart stroke symptoms to prevent fatalities. This project aims to analyze heart disease datasets and utilize feature selection techniques to construct a high-accuracy model using the best-performing machine learning algorithms. By deriving valuable insights from the data, the objective is to contribute to the prevention and improved management of heart-related conditions.

I. INTRODUCTION

The main intention of this project is to make an ML oriented system for predicting heart attacks. The platform utilizes a web-based interface to gather essential data from healthcare professionals and users. After considering other machine learning models, Random Forest is selected for its superior accuracy in predicting the probability of a heart attack. The system facilitates patient interaction with heart specialists and enables them to offer feedback. Comparing the Random Forest algorithm with alternative models allows for an assessment of the heart attack rate prediction accuracy.

The subscription-based Heart Disease Predicting system aims to assist users who lack familiarity with selecting appropriate options. However, it currently falls short in terms of delivering timely and efficient results, validating accuracy, and providing comprehensive suggestions that patients can easily understand. The existing system has several drawbacks, including its high cost, low accuracy, inefficiency, and inability to monitor patients ona daily basis.

Our proposed system focuses on accurately predicting the risk of heart attacks by leveraging the Random Forest algorithm. Simultaneously, the system aims to enhance accessibility and deliver accurate results for all users. The system incorporates machine learning methodologies and various algorithms, with Random Forest chosen for its accuracy with combined datasets. The advantages of the proposed system include high accuracy, efficient storage and access, usability for normal users to predict their own diagnosed data, guidance for new doctors and users through instructions and a user-friendly GUI, reduced time consumption, and ease of use.

II. LITERATURE SURVEY

[1] "Investigating Machine Learning Methods for Predicting Coronary Heart Disease" CHD plays a crucial role in mortality rates. To tackle this problem, the study explored machine learning classifiers. The findings demonstrate that the Support Vector Machine (SVM) model demonstrated superior accuracy in predicting Coronary Heart Disease (CHD). Future research aims to apply these techniques to various datasets and develop web-based services for real-world medical assessments.

[2] "Developing an Effective Heart Disease Prediction System with the Decision Tree Algorithm" The paper discusses the substantial influence of Cardiovascular disease (CVD) on illness and death rates. It introduces an efficient system utilizing Decision Tree to identify risk levels in patients based on their health parameters. The evaluation of the system emphasizes its classification accuracy, indicating its potential for precise prediction of heart disease risk levels.

[3] "Identifying Classification Technique for Medical Diagnosis" It gives a thorough analysis of diverse application of ML approach in the field of health condition identification. The study investigated six machine learning classification algorithms, along with two decision tree-based algorithms, using three distinct UCI datasets. The results demonstrated that the SVM algorithm produced the most favorable outcomes across all three medical datasets, highlighting its promising potential in the field of health assessment.

[4] Development of a Heart Disease Prediction System Based on Hidden Naïve Bayes Classifier" Through experiments, it is observed that the HNB model surpasses other approaches, achieving an impressive accuracy of 100%. These findings indicate that the HNB classifier holds potential for medical datasets, specifically for diagnosing diseases with interdependent attributes like heart disease.

[5] "Utilizing Machine Learning Techniques for Predicting Heart Disease" It investigates various attributes associated with heart disease and applies supervised learning algorithms. The results suggest that the K-nearest neighbor algorithm attains the highest accuracy score.

[6] "A Novel Approach for Predicting Heart Attacks in Stroke Patients with Imbalanced Data" Implementing data analysis and ML models for early detection of heart disease in stroke patients is a novel strategy to decrease mortality rates associated with heart attacks. The research employs the UCO algorithm to train datasets and achieves a prediction accuracy of around 70% using stroke patient data.

[7] "Enhancing Heart Disease Prediction with Hybrid Machine Learning Approaches" The authors of this study aimed to address the rising mortality rate associated with heart disease by developing a predictive system based on clinical data from patients. They employed a hybrid approach combining Random Forest and Linear models to accurately predict heart disease. The proposed model known as HRFLM, demonstrated an enhanced performance level with an impressive accuracy of 88.7% in heart disease prediction.

[8] "An Innovative System for Wireless Heart Rate Monitoring and Vigilance" This paper proposes a novel approach to prevent heart attacks by continuously monitoring patients' heart rates and promptly alerting doctors when necessary. The main objective is to develop a cost-effective and efficient system for heart rate monitoring. The system ensures continuous monitoring, furnishing real-time updates on the patient's heart rate status and displaying the information on the screen. To enhance the detection rate, the authors recommend utilizing long-range RF modules.

III. METHODOLOGY

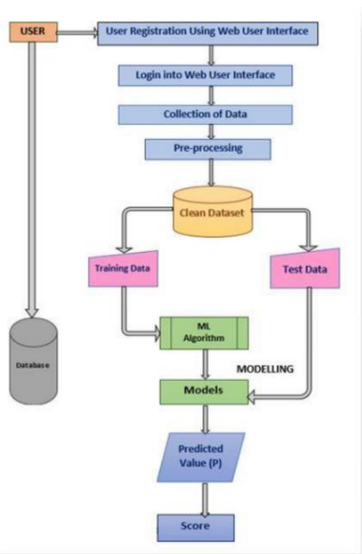

Figure (1) Flow diagram of EHR sharing system

A. Implementation Steps

Step 1: Collected 2 different heart related datasets from Heart Attack Analysis and Prediction dataset UCI-ML-(303) and from Kaggle. More information is given below [2]

1. Age: The age of the patients represented in years (Numeric).

2. Sex: Gender of the patient, with Male represented as 0 and Female as 1 (Nominal).

3. Chest Pain Type: This feature indicates the type of chest pain experienced by the patients and is categorized as follows:

a. Typical angina: This type of chest pain occurs due to reduced blood flow to the heart, leading toa sensation of heaviness in the central chest area.

b. Atypical angina: This type of chest pain can be caused by various factors such as respiratory issues, heavy exercise, gastrointestinal problems, or other non-cardiac reasons.

c. Non-anginal pain: Patients experiencing this type of chest pain may feel a squeezing or pressure-like sensation behind the breastbone, resembling heart-related pain, even though they do not have heart disease.

d. Asymptomatic: This category refers to silent chest pain that can lead to a heart attack, often occurring during sleep without any noticeable symptoms [14].

4. Resting bps: Blood pressure level at rest, measured in mm/HG [10].

5. Cholesterol: Serum cholesterol levels, measured in mg/dl [1].

6. Fasting blood sugar: Binary representation, where 1 indicates blood sugar levels greater than 120 mg/dl, and 0 represents levels equal to or below 120 mg/dl, after fasting.

7. Resting ECG: Result of the resting electrocardiogram, categorized as follows:

a. 0 (Normal): Indicates a normal resting ECG result.

b. 1 (Abnormality in ST-T wave): Signifies an abnormality in the ST-T wave of the resting ECG.

c. 2 (Left ventricular hypertrophy): Indicates left ventricular hypertrophy observed in the resting ECG.

8. Max heart rate: Maximum achieved heart rate during exercise, measured in beats per minute (bpm) [10].

9. Exercise angina: Binary representation, where 0 stands for "No" and 1 indicates "Yes" for angina induced by exercise.

10. Oldpeak: Exercise-induced ST-depression compared to rest.

11. ST slope: ST segment slope during peak exercise, categorized as follows:

a. 0 (Normal): Represents a normal ST segment slope during peak exercise.

b. 1 (Upsloping): Indicates an upsloping ST segment during peak exercise [7].

Step 2: Combining two datasets which gave 606 instances and attributes are selected based on the common feature selection with the help of a Weka tool which uses the best first search algorithm [14]. Selected 7 features /columns out of 11 attributes from dataset is shown below in figure(1).

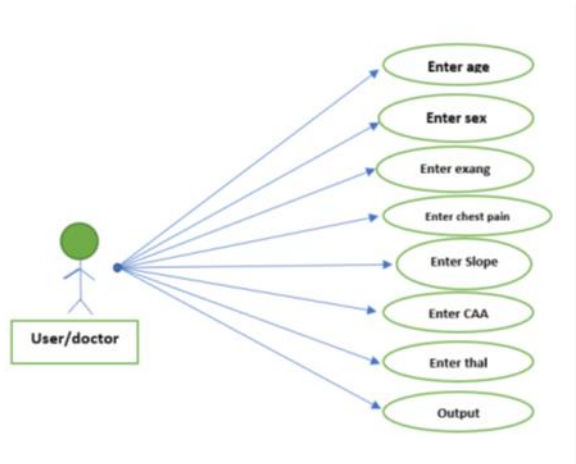

Figure (2) Feature Extracted attributes of dataset using Weka-Tool

Step 3: Importing required libraries and data visualization is done such as CSV, NumPy, Pickle, Flask, Flask-Mail, Pandas, Sklearn and Seaborn.

Step 4: The dataset is divided into 80% for training the machine learning algorithms and fitting the model, while the remaining 20% is allocated as testing data.

Step 5: The heart attack prediction models employed in this study include Naive Bayes, SVM, and Random Forest algorithms etc. [4]

Step 6: Performance measures in classification models evaluate the effectiveness of the models in a given context. Above mentioned metrics offer valuable insights to the model's performance and its capability to make accurate predictions in novel situations [5].

Because understanding the strengths and limitations of these models is essential for effectively utilizing them in various applications.[1]

• **The accuracy score** is a metric that quantifies the frequency of correct predictions made by a machine learning model out of the total number of predictions. However, it's important to understand that accuracy alone does not capture information about potential errors that the model may make when faced with new and unseen data. Additionally, it's worth noting that users or patients have the opportunity to provide feedback to their doctors.

• **The Precision Score** is a measure that quantifies the ratio of accurately predicted positive observations to the total predicted positive observations. It has a range of 0 to 1, where a value of 1 represents the highest performance, while 0 indicates the poorest performance.

• **The recall score** measures the model's ability to correctly predict positive instances out of all the actual positive instances. Recall focuses on the model's proficiency in identifying both positive and negative examples. A higher recall score signifies improved performance of the machine learning model in recognizing positive and negative instances.

• **The F1 score** is a performance metric for machine learning models that integrates both precision and recall scores into a unified measure. It assigns equal significance to both precision and recall when evaluating the model's accuracy. It is calculated as the harmonic mean of precision and recall, the F1 score offers a valuable alternative to accuracy metrics for assessing the model's performance.

Step 7: Comparing All Accuracy Score Values [9]

Table 4.1 Comparing various ML Algorithms.

Table				
Combined Dataset	Accuracy %	Precision Score%	Recall Score%	F1 Score%
KNN	75.40	70.88	88.88	78.87
LR	82.78	81.81	88.88	78.87
ensemble	86.88	83.09	93.65	88.05
SG Booster	90.98	88.23	95.23	91.60
DT	93.44	91.04	96.82	93.84
Naïve Bayes	80.33	80.00	82.54	81.25
SVM	82.79	80.88	87.30	83.97
Random Forest	95.08	93.85	96.83	95.31

Step 8: Selection of Random Forest over other algorithms based on the above noted metrics[15].

B. Classification Algorithm:

The classification algorithm is a supervised learning technique utilized to classify novel observations using training data. It acquires knowledge from a provided dataset and assigns new observations to distinct classes or groups, such as Yes or No, 0 or 1, and more. These classes can be referred to as targets, labels, or categories.

To assess the performance of a classification model, it is evaluated based on its ability to accurately predict classes. This evaluation involves categorizing the model's predictions into four categories:

True positives (TP) occur when the model accurately predicts instances of the positive class.

True negatives (TN) happen when the model correctly identifies instances of the negative class.

False positives (FP) are cases where the model incorrectly predicts the positive class.

False negatives (FN) are instances where the model incorrectly predicts the negative class [3].

C. Confusion Matrix:

The confusion matrix has a major role while measuring the performance of classification models on a specific dataset, serving as a crucial tool to assess their effectiveness. It requires knowledge of the true values of the test data. The matrix provides a structured representation of the model's errors, often referred to as an error matrix. The matrix's dimensions vary based on the number of prediction classes considered. It comprises two aspects: predicted values and actual values, along with the total number of model predictions. The predicted values indicate the model's observations, while the actual values represent the true observations [3].

D. Random forest algorithm

It combines multiple trees to form a forest. The class prediction of each individual tree is combined through majority voting to determine the final prediction. Increasing the number of trees generally improves accuracy. The algorithm handles missing values effectively. However, it can be slow for predictions due to large datasets and more trees. The results of the random forest are not easily interpretable. It is an ensemble algorithm that utilizes randomized decision trees. The process involves creating bootstrap datasets, randomly selecting features, and constructing decision trees. During prediction on test data, the algorithm aggregates insights from various decision trees and uses them to determine the final target value [12].

E. Analysis

In the final phase, we will test our classification model on the combined dataset and measure its performance. In order to assess the efficacy of our classification approach and ensure its comparability to established methods, we employ different metrics. Understanding the model's ability to predict outcomes for new instances is of utmost importance. After developing a predictive model using historical data, it becomes important to assess its performance on unseen data. Multiple model types may be experimented with for the same prediction problem to determine the most suitable model for real-world decision-making. We accomplish this by comparing their prediction performance, including accuracy, and actively choosing the model with the most favorable outcomes[16].

IV. CONCLUSION

Heart attack risk assessment is a crucial step in identifying potential health issues that require immediate attention, allowing individuals to take necessary precautionary measures. This project aims to provide a user-friendly AI-driven health risk assessment platform for users to perform basic health risk evaluations, with a specific focus on heart attack risk assessment.

The AI-driven health risk assessment application primarily focuses on predicting the risk of heart attacks based on user input. By utilizing machine learning-based health predictors [7], the application generates comprehensive heart attack analysis reports, providing valuable insights to users about their cardiovascular health [11].

To develop this platform, eight popular machine learning algorithms were employed for predictive analysis. These algorithms include Logistic Regression, K-Nearest Neighbors (KNN) [6], Decision Tree, SG Booster, Voting Ensemble (Logistic Regression, KNN, Decision Tree), Naïve Bayes, Support Vector Machine (SVM), and Random Forest. The feature extraction process utilizes the efficient Best First Search (BFS) algorithm to identify essential risk factors contributing to heart attack prediction [8].

To ensure accuracy and effectiveness, a dataset containing risk factors data from 606 patients was collected. This data forms the foundation of the model, enabling it to recognize and quantify the impact of various risk factors on heart attack risk. The machine learning models were extensively trained and tested using this dataset, with the Random Forest algorithm emerging as the most accurate, achieving an impressive accuracy rate of 95.04%. Additionally, the Precision value, measuring the model's ability to correctly identify positive cases, was 93.85%. The recall score, indicating the model's capability to detect all positive cases, reached 96.83%, and the F1 score, a harmonic mean of Precision and recall, stood at 95.31% [13].

One of the main advantages of this AI-driven health risk assessment application is its ability to predict heart attack risk without the need for costly medical tests. By utilizing identified critical risk factors, the platform can provide reliable predictions, saving both time and resources required for extensive medical checkups. Moreover, this empowers individuals to actively monitor their health and take preventive measures at the early stages of potential diseases, enhancing overall healthcare management and potentially reducing the occurrence and severity of heart-related conditions[7].

Expanding the scope of this project, future developments could involve incorporating risk assessments for various other health conditions beyond heart attacks. By leveraging machine learning algorithms and continuously updating the system with new research findings, this platform could serve as a comprehensive health risk assessment tool, empowering individuals to make informed decisions about their well-being. Moreover, integration with wearable health devices could enable real-time monitoring, providing timely alerts and personalized health recommendations to further improve preventive healthcare practices.

In conclusion, the development of a user-friendly AI-driven health risk assessment application focused on heart attack risk is a significant step towards promoting proactive healthcare management. By harnessing the power of machine learning algorithms and risk factor analysis, this platform can potentially revolutionize the way individuals approach their cardiovascular health, leading to better health outcomes and reduced medical burdens for individuals and healthcare systems alike.

V. REFERENCES

[1] Hisham Khdair ,Naga M Dasari International Institute of Business and Information Technology, Federation University Associate Adelaide, Australia 'Investigating Machine Learning Methods for Predicting Coronary Heart-related health issues' IJACSA Vol. 12, No. 5, 2021.

[2] Purushottam Research Scholar (PhD.) R.G.T.U. Bhopal (M.P), India. puru.mit2002@gmail.com, R.G.T.U. Bhopal (M.P), India. puru.mit2002@gmail.com , Richa Sharma Assistant professor Amity University Uttar Pradesh Noida, India s.richa.sharma@gmail.com 'Developing an Effective Heart Disease Prediction System with the Decision Tree Algorithm' ICCCA, ISBN:978-1-4799-8890-7/15/$31.00 ©2015.

[3] Potnuru Sai Nishant, Shashi Mehrotra, B. Gopesh Krishna Mohan and Gantakora Devaraju 'Identifying Classification Technique for Medical Diagnosis' published on 12 October 2020.

[4] M.A. Jabbar, Professor Vardhaman College of Engineering, Shirina Samreen, Professor Anurag Group of Institutions Hyderabad, Telangana, INDIA 'Heart disease prediction system based on hidden naïve bayes classifier'.

[5] Devansh Shah, Samir Patel, Santosh Kumar Bharti 'Utilizing Machine Learning Techniques for Predicting Heart Disease'. The paper, published by Springer Nature Singapore Pte Ltd in October 2020, indicates that the study was received on September 27, 2020, and accepted on October 2, 2020. It was made available online on October 16, 2020.

[6] Meng Wang , Xinghua Yao , And Yixiang Chen 'An Imbalanced-Data Processing Algorithm for the Prediction of Heart Attack in Stroke Patients ' Received December 8, 2020, accepted January 14, 2021, date of publication February 8, 2021, date of current version February 16, 2021.

[7] Gnaneswar B and Ebenezar Jebarani M.R. Dept. of ECE Sathyabama University Chennai, India 'A review on prediction and diagnosis of heart failure'

[8] Reddy, K.V.V.; Elamvazuthi, I.; Aziz, A.A.; Paramasivam, S.; Chua, H.N.; Pranavanand, S. Heart Disease Risk Prediction Using Machine Learning Classifiers with Attribute Evaluators 'Heart disease risk prediction using machine learning classifiers with attribute evaluators' published on 9 September 2021 https:// doi.org/10.3390/app11188352

[9] Honey Pandey and S. Prabha Department of ECE Hindustan Institute of Technology and Science 'Smart health monitoring system using IOT and machine learning Techniques'

[10] HDPS: Heart Disease Prediction System AH Chen, SY Huang, PS Hong, CH Cheng, EJ Lin Department of Medical Informatics, Tzu Chi University, Hualien City, Taiwan

[11] Abhiyash Hodge, Hrishikesh Humnabadkar, Akshay Bidwai 'Wireless Heart Rate Monitoring and Vigilant System' 2018 3rd International Conference for Convergence in Technology (I2CT) The Gateway Hotel, XION Complex, Wakad Road, Pune, India. Apr 06-08, 2018

[12] Norma Latif Fitriyani, Muhammad Syafrudin, Ganjar Alfian, (Member, IEEE), And Jongtae Rhee 'HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support ' Received July 7, 2020, accepted July 12, 2020, date of publication July 20, 2020, date of current version July 30, 2020

[13] Senthilkumar mohan , Chandrasegar Thirumalai and Gautam Srivastava, (member, IEEE) 'Effective heart disease prediction using hybrid machine learning techniques' Received May 13, 2019, accepted June 9, 2019, date of publication June 19, 2019, date of current version July 3, 2019.

[14] Anjan Nikhil Repaka Computer Science and Engineering Sathyabama Institute of Science and Technology Chennai, India, Sai Deepak Ravikanti Computer Science and Engineering Sathyabama Institute of Science and Technology Chennai, India, Ramya G Franklin Computer Science and Engineering Sathyabama Institute of Science and Technology Chennai, India 'Design And Implementing Heart Disease Prediction Using Naives Bayesian' Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8

[15] Sushmita Manikandan, Department of Computer Science and Engineering PES Institute of Technology, Bangalore South Campus Bengaluru, India 'Heart attack prediction system' International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017) Dataset obtained from https://archive.ics.uci.edu/ml/machine-learning databases/heartdisease/processed.cleveland.data

[16] A Wireless Early Prediction System of Cardiac Arrest through IoT Yosuf ElSaadany, AKM Jahangir A. Majumder, Donald R. UCCI Department of Electrical and Computer Engineering, Miami University, Oxford, OH, USA.