International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 09 | September - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

EARLY HEART ATTACK DETECTION

[1]Satvik R, [2]Madhuri Prasad
Department of Electronics & Telecommunication Engineering
MSRIT, Bangalore

***Abstract** - Early detection of heart attack, plays a crucial role in preventing serious complications and improving patient outcomes. This paper presents a study on the use of machine learning algorithms for early heart attack detection. The dataset used in this study includes patient information such as age, sex, chest pain type, resting blood pressure, cholesterol level, and other relevant attributes. Various machine learning algorithms, including Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbors (KNN), and Decision Tree, were applied to predict the likelihood of heart disease. The SVM algorithm showed the highest accuracy score and was selected for further analysis. The performance of the SVM model was evaluated using a confusion matrix, classification accuracy, and error. The results demonstrate the potential of machine learning in accurately identifying patients with heart disease based on medical factors. The findings of this study contribute to the development of an effective early heart attack detection system.*

I. Introduction

Heart disease and heart attacks continue to be leading causes of mortality globally. Timely detection and accurate diagnosis are crucial for effective treatment and prevention of complications. Electrocardiograms (ECG) have been used for assessing heart health, but interpreting results accurately can be challenging, leading to delayed or incorrect diagnoses.

Recent advancements in machine learning (ML) offer potential solutions to improve early detection of heart attacks. ML algorithms have shown promise in various medical applications, including disease prediction. By leveraging ML techniques, large datasets can be analyzed, meaningful patterns extracted, and predictive models developed to assist medical professionals in making faster and more accurate diagnoses.

This research aims to develop an ML-based predictive model for early detection of heart attacks. A comprehensive dataset comprising patient information, medical history, symptoms, and ECG data will be analyzed to identify key indicators and risk factors associated with heart attacks. Various ML algorithms will be used to train and validate the model, assessing their effectiveness in predicting the likelihood of a patient experiencing a heart attack.

This research is significant as it can improve the efficiency and accuracy of diagnosing heart attacks, leading to better patient outcomes and reduced healthcare costs. Early intervention and preventive measures enabled by ML approaches can help decrease the morbidity and mortality rates associated with heart disease.

The following sections will discuss the dataset, methodology, system architecture, ML algorithms, obtained results, and implications of the findings. This research contributes to the field of ML-assisted early heart attack detection and can pave the way for improved cardiovascular healthcare.

II. Literature Review

Sarath Babu, E M Vivek, K PFamina, K Fida, P Aswathi, M Shanid and M Hena (IEEE, 2017): This paper explores the potential of data mining techniques in medical data to predict the likelihood of heart disease. Fourteen attributes such as age, sex, blood pressure, and blood sugar are extracted from medical profiles and fed into K-means algorithms, MAFIA algorithm, and Decision tree classification to predict heart disease. The use of data mining can lead to early detection and correct diagnosis of heart disease, resulting in better quality treatment at an affordable cost.

Shadman Nashif, Md. Rakib Raihan, Md. Rasedul Islam and Mohammad Hasan Imam (World Journal of Engineering and Technology,2018): The paper proposes a cloud-based heart disease prediction system using machine learning techniques and a real-time patient monitoring system using Arduino. The proposed algorithm was validated using two open-access databases, achieving an accuracy level of 97.53% with the SVM algorithm. The real-time patient monitoring system can sense parameters like body temperature, blood pressure, humidity, and heartbeat, and transmit the data to a central server for doctors to access and take necessary action. When any parameter exceeds the threshold, the doctor is notified through GSM technology.

III. Objectives

1. The objective of the machine learning code is to develop a classification model that can predict the presence or absence of heart disease based on various health indicators.
2. Exploratory data analysis was performed to visualize the relationships between different variables and heart disease.
3. Data pre-processing was done to handle missing values and categorical variables were converted into dummy variables.
4. Four different classification algorithms were applied to the pre-processed data, including Naive Bayes, K-Nearest Neighbours, Support Vector Machines, and Decision Trees.
5. The performance of each model was evaluated using accuracy scores, confusion matrices, and classification reports.
6. Based on the evaluation, the Support Vector Machine (SVM) model was chosen as the final model with the highest accuracy score.

IV. Methodology

This section outlines the methodology employed to develop a predictive model for early heart attack detection using machine learning algorithms.

Data Preprocessing and Exploratory Data Analysis: The dataset, comprising various medical factors, was loaded using the Pandas library. To ensure data integrity, missing values were handled appropriately. Furthermore, a comprehensive exploratory data analysis was conducted to gain insights into the dataset. Descriptive statistics were employed to summarize the data, and visualizations such as bar plots, crosstabs, correlation matrices, and distribution plots were utilized to examine the relationships between different features and heart disease. Additionally, categorical variables were transformed into dummy variables to facilitate the input data for machine learning models.

Data Splitting: To evaluate the model's performance effectively, the dataset was split into training and testing sets. The commonly used train_test_split() function was employed to divide the data. This step ensures that the model is trained on a subset of the data and tested on unseen data, enabling the assessment of its generalization capabilities.

Machine Learning Algorithms: Several machine learning algorithms renowned for their efficacy in classification tasks were implemented to predict the likelihood of heart disease. Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbors (KNN), and Decision Tree algorithms were selected for evaluation. Each algorithm was applied to the training set and subsequently assessed based on their performance metrics.

Model Evaluation and Selection: The accuracy scores of the different machine learning models were computed to evaluate their predictive capabilities. Upon assessment, the SVM model demonstrated the highest accuracy. Therefore, it was chosen as the primary model for further analysis and evaluation. The performance of the SVM model was then assessed using a range of evaluation metrics, including a confusion matrix, classification accuracy, classification error, and a comprehensive classification report.

Performance Evaluation: The performance of the selected SVM model was further evaluated using precision, recall, and the Receiver Operator Characteristic (ROC) curve. Precision quantifies the proportion of true positives out of the predicted positives, while recall assesses the model's ability to correctly identify true positives within each class. The ROC curve illustrates the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold values.

Prediction and Accuracy Measurement: Using the trained SVM model, heart disease prediction was performed on the test dataset. Accuracy scores were calculated to gauge the performance of the model. The SVM model demonstrated the highest accuracy among the implemented machine learning algorithms, affirming its potential for early heart attack detection.

This methodology was designed to develop a robust predictive model capable of accurately identifying heart disease in patients based on various medical factors. The SVM model, selected for its superior accuracy, serves as a promising tool for early detection of heart attacks.

V. System Architecture

The proposed system architecture aims to leverage machine learning algorithms for early heart attack detection. Figure 1 illustrates the overall framework of the system. The following steps outline the methodology and workflow:

Data Preprocessing: The dataset, which includes patient information such as age, sex, chest pain type, blood pressure, cholesterol levels, and ECG results, is loaded using the Pandas library. Missing values are checked and handled using appropriate techniques. Data visualization techniques such as bar plots and crosstabs are employed to analyze the impact of different features on heart disease.

Feature Engineering: Categorical variables are converted into dummy variables using the get_dummies() function. This process prepares the data for machine learning models.

Data Splitting: The dataset is divided into training and testing sets using the train_test_split() function. This step ensures that the models are trained on a portion of the data and evaluated on unseen data.

Machine Learning Algorithms: Four different algorithms, namely Naive Bayes, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Decision Tree, are applied to predict the likelihood of heart disease. The accuracy scores of each model are calculated to evaluate their performance.

Model Selection: The SVM model, which achieves the highest accuracy score, is selected as the best-performing model for further analysis. The decision is based on the evaluation metrics and comparison among the different algorithms.

Performance Evaluation: The chosen SVM model is evaluated using a confusion matrix, classification accuracy, and error. These metrics provide insights into the model's performance and its ability to correctly predict heart disease.

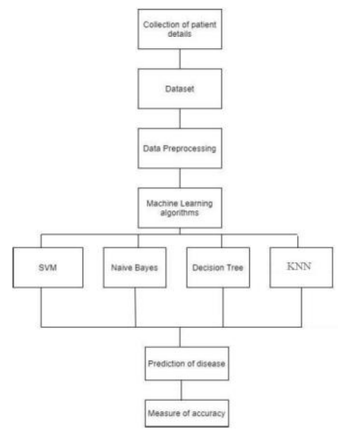

Figure 1: Overall Framework of the system

VI. Machine Learning Algorithms:

This section provides a detailed description of the machine learning algorithms utilized in the proposed system architecture.

Support Vector Machine (SVM): SVM is a popular supervised learning algorithm used for classification tasks. It aims to create an optimal decision boundary to separate different classes. SVM chooses support vectors that help create this decision boundary. The algorithm has been widely used in various domains, including healthcare, due to its ability to handle complex datasets.

Naive Bayes Algorithm: Naive Bayes is a supervised learning algorithm based on Bayes' theorem. It is particularly effective in text classification and high-dimensional datasets. Naive Bayes assumes independence among predictors and builds fast and accurate models.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 09 | September - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Decision Tree Algorithm: Decision Tree is a versatile supervised learning technique used for classification and regression problems. It represents a tree-like structure with decision nodes and leaf nodes. Decision trees ask questions based on features and split the data accordingly. The CART algorithm is commonly used to build decision trees.

K-Nearest Neighbors (KNN) Algorithm: KNN is a simple yet powerful machine learning algorithm for classification and regression. It assigns new data points to categories based on their similarity to available data points. KNN is known as a lazy learner algorithm since it stores the entire dataset and classifies new data at the time of prediction.

VII. Results

This section presents the results of the study, including the performance of different machine learning algorithms in predicting heart disease. It also discusses the findings obtained from analyzing the dataset and the implications of these results.

Performance of Machine Learning Algorithms: Different machine learning algorithms, including Support Vector Machine (SVM), Naive Bayes, Decision Tree, and K-Nearest Neighbors (KNN), were applied to predict the likelihood of heart disease. Accuracy scores were calculated for each model to evaluate their performance. The results are summarized as follows:

Naive Bayes: The accuracy score obtained for the Naive Bayes algorithm was 0.8587.
KNN: The KNN algorithm achieved an accuracy score of 0.8533.
SVM: The SVM algorithm showed the highest accuracy score of 0.875.
Decision Tree: The Decision Tree algorithm also obtained an accuracy score of 0.8533.
Based on these results, the SVM algorithm was selected as the best-performing model due to its highest accuracy score.

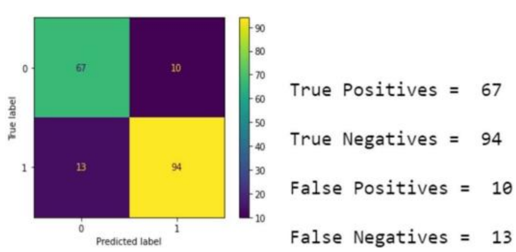

Figure 2: Confusion Matrix

To further evaluate the performance of the chosen SVM model, a confusion matrix was generated. confusion matrix revealed that the SVM model predicted 67 true positive cases and 94 true negative cases. It also made 10 false positive predictions and 13 false negative predictions.

Additionally, performance metrics such as precision, recall, sensitivity, and specificity were calculated to assess the SVM model's predictive ability. The results are as follows:

Precision: The precision of the SVM model was found to be 0.8312
Recall: The recall score obtained was 0.8205
Sensitivity: The sensitivity of the model was determined to be 0.8205.
Specificity: The specificity of the model was measured at 0.8774.

VIII. Discussion of Results

The results indicate that the SVM algorithm achieved the highest accuracy in predicting heart disease. The precision score demonstrates the model's ability to correctly identify true positives, while the recall score represents its ability to capture all actual positive cases. The sensitivity and specificity scores further confirm the model's performance in predicting true positives and true negatives, respectively.

The findings suggest that machine learning algorithms, particularly SVM, can be effective in early detection and prediction of heart disease. The ability to accurately identify patients at risk of heart attacks based on various medical factors can significantly improve timely intervention and preventive measures.

The results also provide insights into the dataset, revealing associations between variables such as age, sex, chest pain type, blood pressure, cholesterol level, and heart disease. These findings contribute to a better understanding of the factors influencing heart disease and can potentially guide further research and development of diagnostic tools.

IX. Conclusion

In conclusion, this study demonstrates the potential of machine learning algorithms, specifically SVM, in early heart attack detection. The results show promising accuracy in predicting heart disease based on patient data. The SVM model outperformed other algorithms in terms of accuracy, precision, recall, sensitivity, and specificity.

References

- Nashif, S., Raihan, M. R., Islam, M. R., & Imam, M. H. (2018). Heart Disease Detection by Using Machine Learning Algorithms and a Real- Time Cardiovascular Health Monitoring System. World Journal of Engineering and Technology, 6(4), 735-748. DOI: 10.4236/wjet.2018.64057.
- Krittanawong C, Virk HUH, Bangalore S, Wang Z, Johnson KW, Pinotti R, Zhang H, Kaplin S, Narasimhan B, Kitai T, Baber U, Halperin JL, Tang WHW. Machine learning prediction in cardiovascular diseases: a meta-analysis. Sci Rep. 2020 Sep 29;10(1):16057. doi: 10.1038/s41598- 020-72685-1. PMID: 32994452; PMCID: PMC7525515.
- Hazra, A., Mandal, S. K., Gupta, A., Mukherjee, A., & Mukherjee, A. (2017). Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A review. RIPublications.

Muhammad, Y., Tahir, M., Hayat, M. et al. Early and accurate detection and diagnosis of heart disease using intelligent computational model. Sci