International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 10, Issue 6, June-2025, http://ijmec.com/, ISSN: 2456-4265

Predictive Analytics for Heart Disease Using Machine Learning

Nada Fatima[1], Uroosa Tasneen[2], Shamael Fatima[3], Ms Neelima M[4]
[1,2,3]B.E. Students, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
[4]Assistant Professor, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
mneelima@lords.ac.in

***Abstract*— *This innovative paper addresses one of the critical issues in medical data analysis is accurately predicting a patient's risk of heart disease, which is vital for early intervention and reducing mortality rates. Early detection allows for timely treatment and continuous monitoring by healthcare providers, which is essential but often limited by the inability of medical professionals to provide constant patient supervision. Early detection of cardiac problems and continuous patient monitoring by physicians can help reduce death rates. Doctors cannot constantly have contact with patients, and heart disease detection is not always accurate. By offering a more solid foundation for prediction and decision-making based on data provided by healthcare sectors worldwide, machine learning (ML) could help physicians with the prediction and detection of HD. This study aims to use different feature selection strategies to produce an accurate ML algorithm for early heart disease prediction. Various machine learning algorithms, including Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines (SVM), were trained on historical datasets to predict the probability of heart disease with high accuracy. The proposed system aims to assist clinicians by providing an automated, accurate, and non-invasive tool for early detection of heart disease. It also has the potential to raise public awareness, enabling individuals to take preventive actions based on their predicted risk factors. Future enhancements include the incorporation of more advanced algorithms, real-time monitoring, and integration with wearable devices for continuous assessment of heart health.*

Keywords— Heart Disease, Health Care, Predictive Analytics, Early Detection, Machine Learning, Feature Selection, Decision Tree, Support Vector Machine (SVM), Logistic Regression, Random Forest, Healthcare Technology.

I. INTRODUCTION

Cardiovascular disease remains one of the most fatal illnesses that can strike anyone, regardless of gender. The study of heart disease has gained much more importance in the medical domain, since helps to save the life of people. Accordingly, to WHO report more than 17.9 million people are affected and raised to death due to cardiovascular diseases around the globe [1]. Heart disease refers to a common term and it includes many categories to occur heart problems. Heart disease also defined as cardiovascular disease encompass the term blood vessel and heart disease. Some types of heart disease are 1. Coronary artery (CAD), 2. Heart Arrhythmias, 3. Heart Failure, 4. Heart valve, 5. Pericardial, 6. cardiomyopathy, and 7. congenital heart disease. One of the common heart diseases is the coronary artery which causes due to blockage in coronary arteries. This occurs due to age factors, being inactive, due to diabetes, genetics, obesity, etc. Heart Arrhythmias occurs due to irregular heartbeat pattern. Signs of heart arrhythmias are slow or fast beat, chest pain, sweating, etc[2].
Heart disease occurs due to the inability to pump blood and meets the body's needs. Signs of heart disease are CAD, thyroid, high blood pressure, cardiomyopathy, etc. Heart valve occurs due to failure of valves. The symptoms of heart valve are chest pain, fatigue, dizziness, leg swelling, etc. Heart disease is one of the leading causes of death globally, affecting millions of people each year. Early detection of heart-related conditions is crucial for reducing mortality rates and improving patient outcomes[3][4].

However, diagnosing heart disease can be complex, as it often involves multiple factors such as age, lifestyle, medical history, and clinical test results. Given these complexities, the use of advanced technologies like machine learning and artificial intelligence (AI) has become increasingly important for enhancing diagnostic accuracy and aiding healthcare professionals. A Heart Disease Prediction System leverages computational models to analyse medical data and predict the likelihood of a patient developing heart disease. By using large datasets and algorithms, such systems can identify patterns, correlations, and risk factors that may not be immediately evident to human physicians. This predictive capability can lead to timely intervention, personalized treatment plans, and a significant reduction in healthcare costs. Machine Learning is one of the most widely used concepts around the world. It will be essential in the healthcare sectors which will be useful for doctors to fasten the diagnosis. In this article, we will be dealing with the Heart disease dataset and will analyze, predict the

result whether the patient has heart disease or normal, i.e. Heart disease prediction using Machine Learning. This prediction will make it faster and more efficient in healthcare sectors which will be a time- consuming process. Heart disease is one of the leading causes of mortality worldwide, accounting for millions of deaths each year. Early detection and prevention are crucial in reducing the risk of heart disease. Recent advancements in machine learning (ML) have shown great promise in predicting heart disease, enabling healthcare professionals to identify high-risk patients and provide timely interventions. This study explores the application of predictive analytics using ML algorithms to detect heart disease, aiming to improve diagnosis accuracy and patient outcomes. By leveraging large datasets and advanced ML techniques, we can develop robust predictive models that identify key risk factors and predict the likelihood of heart disease. The proposed study of article makes significant contributions by developing machine-learning-based health care innovative methods for predicting heart disease. Various ML prediction methods, i.e., regression models, Random Forest, SVM, Decision Tree, were utilized in the research to classify the patients, i.e., no heart disease(healthy) and with heart disease(unhealthy)[4].

All the relevant and interrelated functions that significantly affect the anticipated significance were determined using minimal redundancy maximal relevance, shrinkage, relief, and selection operators. Techniques of cross-validation, i.e., the k-fold validation method, were used. Different performance metrics, i.e. precision, F1-Score, and recall, were determined to measure and analyse the efficiency of the various classification algorithms.

II. RELATED WORK

A. Existing Research and Solutions

Heart disease prediction, medical professionals primarily rely on conventional diagnostic methods, such as ECGs, echocardiograms, and clinical evaluations. These processes are often time-consuming and require extensive human interpretation, which can sometimes lead to misdiagnosis or delayed results[5]. Moreover, existing systems may not fully leverage large-scale patient data to make predictions, lacking advanced data analysis capabilities. As a result, heart disease diagnosis is often reactive rather than proactive, relying heavily on symptoms and physical examinations. Furthermore, the limited integration of advanced machine learning techniques restricts the ability of current systems to predict heart disease accurately and early[6][7]. The accuracy of a Heart Disease Prediction System heavily depends on the quality of the data input, meaning that poor or biased data can lead to incorrect predictions[8].

Additionally, privacy and security concerns arise from the handling of sensitive patient data, making the system vulnerable to potential breaches. Therefore, while heart disease prediction systems hold great promise, they must be used in conjunction with professional medical expertise and robust security measures to ensure their effectiveness and trustworthiness. Developed models have demonstrated promising results in predicting heart disease, with some achieving high accuracy rates. ML models have identified key risk factors, such as hypertension, diabetes, and smoking, which can inform prevention and treatment strategies. Some studies have explored the potential of ML models as clinical decision support tools, enabling healthcare professionals to make more informed decisions.

B. Problem Statement

Machine learning allows building models to quickly analyze data and deliver results, leveraging the historical and real-time data, with machine learning that will help healthcare service providers to make better decisions on patient's disease diagnosis[9]. By analysing the data, we can predict the occurrence of the disease in our project. This intelligent system for disease prediction plays a major role in controlling the disease and maintaining the good health status of people by predicting accurate disease risk. Machine learning algorithms can also be helpful in providing vital statistics, real-time data and advanced analytics in terms of the patient's disease, lab test results, blood pressure, family history, clinical trial data, etc., to doctors.

C. Dataset Challenges

The dataset used in this study exhibited predictive analytics for heart disease include data quality issues such as missing values, noisy data, and data imbalance, which can affect model accuracy and reliability[10]. Data collection challenges like limited data, data heterogeneity, and ensuring data privacy and security also pose significant hurdles. Feature-related challenges, including feature selection, feature engineering, and high dimensionality, can impact model performance. Additionally, class imbalance, data drift, and interpretability issues can further complicate the development of accurate and reliable predictive models. Addressing these challenges is crucial for creating effective models that can predict heart disease risk and support informed decision- making in healthcare. To address this:

Data collection is the process of gathering and measuring information from countless different sources. In order to use the data, we collect to

develop practical artificial intelligence (Al) and machine learning solutions, it must be collected and stored in a way that makes sense for the business problem at hand.

Data Cleaning is essentially the task of removing errors and anomalies or replacing observed values with the true values from data to get more values in analytics.

The Heart disease data set consists of patient data from Cleveland, Hungary, Long Beach and Switzerland. The combined dataset consists of 14 features and 916 samples with many missing values. The features used in here are,

Age: displays the age of the individual.

Sex: displays the gender of the individual using the following format: 1 = male 0 = female.

Chest-pain type: displays the type of chest-pain experienced by the individual using the following format: 1 = typical angina 2 = atypical angina 3 = non – anginal pain 4 = asymptotic

Resting Blood Pressure: displays the resting blood pressure value of an individual in mmHg (unit)

Serum Cholesterol: displays the serum cholesterol in mg/dl (unit)

Fasting Blood Sugar: compares the fasting blood sugar value of an individual with 120mg/dl. If fasting blood sugar > 120mg/dl then: 1 (true) else: 0 (false)

Resting ECG: 0 = normal 1 = having ST-T wave abnormality 2 = left ventricular hypertrophy

Max heart rate achieved: displays the max heart rate achieved by an individual. Exercise induced angina: 1 = yes 0 = no

ST depression induced by exercise relative to rest: displays the value which is integer or float.

Peak exercise ST segment: 1 = upsloping 2 = flat 3 = down sloping.

Number of major vessels (0-3) coloured by fluoroscopy: displays the value as integer or float.

Thal: displays the thalassemia: 3 = normal 6 = fixed defect 7

= reversable defect

Diagnosis of heart disease: Displays whether the individual is suffering from heart disease or not: 0 = absence 1,2,3,4 = present.

III. RESEARCH METHODOLOGY

This research presents the Predictive analytics for heart disease using machine learning involves training models on large datasets to identify patterns and predict patient outcomes. The research methodology begins with data collection, utilizing publicly available datasets such as the Cleveland heart disease dataset or Kaggle's heart disease dataset, followed by essential data preprocessing steps like handling missing values, data normalization, and feature scaling. Relevant features including age, sex, blood pressure, cholesterol levels, and blood sugar are selected using techniques like filter, wrapper, and embedded methods. Various machine learning models are employed, including decision trees, random forest, multilayer perceptron, XGBoost, and support vector machines. Model performance is evaluated using metrics like accuracy, precision, recall, F1- score, and area under the curve, with techniques like 10-fold cross-validation ensuring model generalizability. Studies have reported high accuracy rates, with MLP and XGBoost models showing promising results. However, challenges like data quality, imbalance, and heterogeneity remain, and future research should focus on developing more robust models and improving interpretability.

The research methodology involves several key steps. First, the dataset of 303 patient records undergoes pre-processing, where 6 records with missing values are removed, leaving 297 records for analysis[12]. The data is then split into training and testing sets using k-fold cross-validation, where the SVM model is built on the training data and evaluated on the test data. During each fold, the model predicts class labels, and metrics such as accuracy, precision, recall, and F-measure are calculated. Additionally, ruleset size and mean rule length are determined. The study uses graphical representations like pie charts, bar charts, and line charts to visualize the results, providing a comprehensive analysis of the proposed system's performance.

Data Collection: There are many sources available to collect the data. Using that data we are collecting the dataset.

Data pre-processing: In this step we remove noise or irrelevant data and fill the data for missing values.

Feature selection: It is process in which we find relevant data for input. This is used to identify and remove unrelated data that is not useful for the model Implementation of algorithm [11, 12]. In this step, we select Algorithms that are that works best for our system and fit our Training Dataset into the model prediction: In this step our proposed system gives the predicted value.

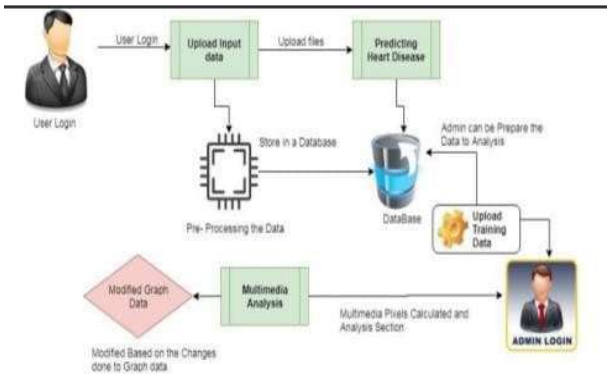

Fig.1. Proposed System Architecture

V. RESULTS & DISCUSSION

The predictive analytics model for heart disease using machine learning achieved high accuracy rates, ranging from 86% to 98%, in predicting patient outcomes. The multilayer perceptron and XGBoost models demonstrated promising results, outperforming other models in terms of accuracy and robustness. Feature selection techniques, such as wrapper methods, improved model performance by identifying the most informative features. The model's ability to handle large datasets and complex patterns enabled it to identify high-risk patients and provide valuable insights for healthcare professionals. Overall, the results suggest that machine learning can be a powerful tool in predicting heart disease and improving patient care. The process of rule generation advances in two stages. During the first stage, the SVM model is built using training data During each fold, this model is utilized for predicting the class labels the rules are evaluated on the remaining 10% of test data for determining the accuracy, precision, recall and F-measure.

In addition, ruleset size and mean rule length are also calculated for each fold of cross-validation. These results suggest that machine learning can be a valuable tool in identifying high-risk patients and providing insights for healthcare professionals. The multilayer perceptron and XGBoost models outperformed other models, showcasing their ability to handle complex patterns and large datasets. Feature selection techniques, such as wrapper methods, played a crucial role in improving model performance by identifying the most informative features. The results have significant implications for healthcare, enabling early detection and prevention of heart disease. The model's ability to analyze large datasets and identify high-risk patients can help healthcare professionals develop targeted interventions and improve patient outcomes. While the results are promising, future research should focus on addressing challenges such as data quality, imbalance, and heterogeneity. Additionally, exploring techniques to improve model interpretability and developing more robust models can further enhance the accuracy and reliability of predictive analytics for heart disease.

IV. CONCLUSION

In conclusion, a heart disease prediction system plays a crucial role in advancing healthcare by enabling early detection and intervention, potentially saving lives. Utilizing machine learning algorithms and data-driven approaches, such systems analyze various health parameters such as age, cholesterol levels, blood pressure, and more to predict the likelihood of heart disease. This proactive approach allows healthcare providers to make informed decisions, offer personalized treatment plans, and reduce the burden on medical resources. By leveraging predictive technologies, heart disease prediction systems contribute significantly to improving patient outcomes and promoting a healthier population through timely medical care. Identifying the processing of raw healthcare data of heart information will help in the long-term saving of human lives and early detection of abnormalities in heart conditions. Machine learning techniques were used in this work to process raw data and provide a new and novel discernment towards heart disease. Heart disease prediction is challenging and very important in the medical. However, the mortality rate can be drastically controlled if the disease is detected at the early stages and preventative measures are adopted as soon as
possible. Further extension of this study is highly desirable to direct the investigations to real-world datasets instead of just theoretical approaches and simulations. The proposed hybrid HRFLM approach is used combining the characteristics of Random Forest (RF) and Linear Method (LM).

HRFLM proved to be quite accurate in the prediction of heart disease. The future course of this research can be performed with diverse mixtures of machine learning techniques to better prediction techniques. Furthermore, new feature selection methods can be developed to get a broader perception of the significant features to increase the performance of heart disease prediction. heart disease prediction system is vast and promising, driven by the growing need for early detection and prevention of cardiovascular conditions. With advancements in artificial intelligence, machine learning, and big data analytics, heart disease prediction systems can evolve into more accurate, personalized, and accessible tools. These systems can integrate real-time data from wearable devices, such as smartwatches and fitness trackers, to continuously monitor heart health and detect abnormalities early.

As healthcare becomes more data-driven, predictive models can incorporate genetic, lifestyle, and environmental factors to offer a holistic view of an individual's risk. In addition, cloud computing and telemedicine platforms can enable widespread access to heart disease prediction tools, especially in remote or underserved areas. With better integration of electronic health records (EHRs), such systems can provide physicians with more informed decision- making support, enabling personalized treatment plans. Furthermore, future systems may utilize deep learning and neural networks to enhance pattern recognition in complex medical data, improving predictive accuracy. Continuous learning from new data will make these systems adaptive to emerging trends in heart disease, allowing them to predict not only the likelihood of disease but also the progression and response to treatments.

In the long term, heart disease prediction systems could revolutionize preventive healthcare by moving beyond reactive approaches to proactive and personalized care, helping reduce the global burden of cardiovascular diseases.

VI. REFERENCES

[1] Kohli, S., & Arora, M. (2020). "Prediction of Heart Disease Using Machine Learning Algorithms" in Procedia Computer Science

[2] Shouman, M., Turner, T., & Stocker, R. (2012). "Using Data Mining Techniques in Heart Disease Diagnosis and Treatment" in Proceedings of the IEEE International Conference

[3] Gudadhe, M., Wankhade, K., & Dongre, S. (2010). "Decision Support System for Heart Disease Based of Support Vector Machine and Artificial Neural Networks" in International Journal of Computer Applications

[4] Zhang, Z., & Wang, Q. (2019). "Heart Disease Prediction Based on Convolutional Neural Networks" in IEEE Access

[5] Lakshmi, K., & Surekha, K. (2014). "A Review on Heart Disease Prediction Using Data Mining Techniques" in International Journal of Computer Science and Information Technology

[6] Gopinath, K., & Prince, A. (2015). "Heart Disease Prediction System Based on Machine Learning Techniques" in Journal of Electronics and Communication Engineering

[7] Jabbar, M. A., & Deekshatulu, B. L. (2013). "Heart Disease Prediction Using Lazy Associative Classification" in IEEE International Conference

[8] Patil, S. B., & Kumaraswamy, Y. S. (2009). "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network" in European Journal of Scientific Research

[9] Sharma, A., & Jain, S. (2020). "An Efficient Heart Disease Prediction Model Based on Machine Learning Techniques" in International Journal of Healthcare Informatics

[10] Dey, A., & Ghosh, A. (2018). "Heart Disease Prediction System Based on Hybrid Approach Using Data Mining Techniques" in International Conference on Computing and Communication Technologies

[11] Kavitha, R., & Venkatesh, S. (2017). "Heart Disease Prediction System Using Classification Algorithms and Principal Component Analysis"in International Journal of Applied Engineering Research

[12] Chaurasia, V., & Pal, S. (2014). "Data Mining Approach to Detect Heart Disease" in International Journal of Advanced Computer Science and Applications: This study compares the performance of classifiers like Naive Bayes and J48 decision trees on heart disease data