Prediction of Heart Disease using Machine Learning Technique

Shachi Mall
Associate Professor
School of Computer
Science Engineering,
Galgotias University
Greater Noida, India
shachimall@gmail.com

Jitendra Nath Singh
Professor
School of
Computer Science
Galgotias University Gr.
Noida, U.P
singhjn2000@gmail.com

Anshika malik
School of Computer
Science Engineering,
Galgotias University
Greater Noida, India
anshikaru@gmail.com

Lakshay Mahur
Department of
Computer Science
Raj Kumar Goel
Institute of Technology
Ghaziabad, U.P
lakshfcs@rkgit.edu.in

Myasar Mundher
adnan Department of
Computer Science
Najaf, Iraq The Islamic
universityMaiser.month
er@iunajaf.edu.iq

Abstract— heart disease is a leading global cause of death, and predicting it is complex, requiring advanced expertise beyond doctors' ease. The medical environment remains "Information rich" and "Information poor". There is a lot of medical information available on the internet. However, there are no good analytical tools for discovering relationships and patterns in the data. Automated diagnostic systems offer greater efficiency and cost-effectiveness. This web application uses Cleveland Medical Research Foundation and Kaggle heart disease data to predict the likelihood of disease occurrence. The goal is to apply knowledge mining techniques to the dataset to extract hidden patterns important to heart disease and predict whether a patient has heart disease where heart disease is dedicated to something. Modern technology struggles to process and analyze the massive data volumes required for accurate heart disease prediction, highlighting the need for advanced tools to handle such complex healthcare challenges. Our goal is to find suitable machine learning methods with high performance and accuracy in predicting heart disease. Data mining enables the discovery of hidden patterns and relationships in large datasets by combining statistical analysis, machine learning, and advanced database techniques for insightful and efficient data interpretation.

Keywords— *Keywords*: Prediction, heart disease, machine learning, algorithm, analysis, database, data search.

1. INTRODUCTION

According to the World Health Organisation (WHO), the primary cause of death worldwide is cardiovascular diseases (CVDs), which include heart attacks brought on by ST-Elevation Myocardial Infarction (STEMI). According to the most recent data from the WHO, CVDs are thought to be responsible for 17.9 million deaths a year, or 32% of all deaths worldwide. STEMI makes up a sizable percentage of these because it is an acute condition that can have serious consequences if left untreated. We can use mining techniques to find information from datasets. Healthcare providers can use the findings to improve the quality of care. Doctors can also use the information discovered to reduce the number of adverse reactions in order to find cheaper treatments. Data mining techniques are essential in healthcare management because they allow for proactive decision-making and better care delivery by predicting a patient's future health outcomes based on their medical history. The main problem of healthcare institutions (hospitals, clinics) is to provide quality service at an affordable price. Good care means diagnosing patients correctly and administering the correct treatments. Poor medical decisions can be disastrous and therefore unnecessary. Hospitals also need to reduce diagnostic costs. They can reach these conclusions using appropriate computerized information and/or decision support [10,13,14,15,16]. The medical literature is quite extensive. It includes patient information, administrative information, and transaction information. Healthcare must analyze data of medical records of patients using machine learning algorithm. The records are in form of features. On the basis of the features the machine learning algorithm take decisions on the basis of the data of decision experience of physician's opinion and experience. This quality of services provided to patients in the medical decision reduce time and improve the patient's treatment experience. This recommendation aids medical decision-making by leveraging data modeling, analysis tools, and classification algorithms to enhance decision quality.

2. RELATED WORK

Smart heart disease assessment using information technology: The medical industry collects a wealth of medical information; Unfortunately, this information is not "scrubbed" to reveal confidential information for prudent decision making. Heart attacks, especially those brought on by ST-Elevation Myocardial Infarction (STEMI), are serious medical emergencies that need to be diagnosed and treated right away [1]. The hallmark of STEMI is a marked increase in the electrocardiogram's (ECG) ST segment, which indicates acute myocardial damage brought on by total coronary artery blockage. To enhance patient outcomes and lower mortality, STEMI must be detected early and accurately. Traditional

507

diagnostic techniques, however, might have drawbacks like their dependence on clinical judgement and interpretive variability [6,17,18]. The relationship between emotional distress and cardiovascular health has been widely explored in recent studies. Research indicates that psychological factors significantly influence physical health outcomes, particularly in individuals with chronic conditions like diabetes. Recent advancements in machine learning provide opportunities to identify and predict the impact of psychological distress on cardiovascular health. Predictive models incorporating emotional and clinical data can offer personalized insights, aiding in early interventions and improved management of diabetes and cardiovascular risks.

Use data mining technology to analyze data and generate risk scores for ischemic heart disease (IHD), categorizing the risk as low, intermediate, or high [19,20]

Technology Research Forecast: Cardiovascular disease remains a leading global cause of death, making its prediction a critical but challenging task that requires specialized skills and knowledge. It is often difficult for doctors to make accurate predictions. This article tackles the issue of heart disease prediction through data mining techniques. Results from performance factors such as SMO (Sequential Minimal Optimization) and Bayes Net show superior performance compared to K-Star, Multilayer Perceptron, and J48 techniques.

The effectiveness of these data mining methods is assessed using six standard datasets, with prediction accuracy, ROC curves, and AUC (Area Under the Curve) values as key evaluation metrics. Based on the performance of SMO and Bayesian network models, K-Star performs better than Multilayer Perceptron and J48 models. Cardiovascular disease was responsible for an estimated 8.2 million deaths in 2012, accounting for roughly 15% of global mortality [3].

This study introduces a machine learning algorithm designed to find the risk in heart disease [4]. UCI and Kaggle datasets are used to evaluate the planning process. The cardiovascular disease predictions made by the classification models in this study are compared with previous research [5]. Experimental results indicate that the proposed scheme provides higher accuracy in predicting risk compared to other approaches. Additionally, the impact of imputation on prediction performance was assessed, with the REMI method outperforming traditional methods [7].

3. OBJECTIVES

Ease of use: The main aim of the project is to create an easy-to-use platform because here the medical information of the patient must be provided and the feature extraction algorithm will diagnose heart disease and determine its type. Since the algorithm here completes this task, the training model rarely makes mistakes in predicting heart disease and its type, therefore accuracy increases correction, thus also saving time for both doctors and patients to find the heart disease.

No human intervention: age, cholesterol, etc. to diagnose heart disease. medical information must be provided and the algorithm will provide the results accordingly, the results are removed, hence there is a possibility of error. They can continue faster. It's a case of getting them quicker. This can save doctors and patients valuable time by making the prevention/prevention process of heart disease treatment faster so they can proceed with it.

Additional treatment and prevention measures: take steps to reduce risk of heart disease. Write, but attention needs to be paid to notes to reduce the risk of heart disease. Getting recommendations on prevention will help doctors and patients easily take further treatment steps. It is believed that successful training of machine learning algorithms requires thousands of training examples.

Therefore, the network and training strategy we use are based on the use of previous data to better leverage the recorded patterns. Because there is not much clinical data (more than thousands of models based on machine learning models) we use previous data to use better informed information. Preliminary data is essential for data mining techniques and involves converting raw data into an understandable format.

4. PROPOSED SYSTEM

Heart Disease is a web-based machine learning application that predicts heart disease risk based on user-input medical details. The system uses clustering and classification methods to calculate the likelihood of a heart attack.

By leveraging these algorithms, the application provides users with timely and accurate predictions, helping to reduce both the cost and time typically associated with disease prediction. The data used for training is sourced from the UCI dataset, which includes a variety of relevant health indicators.

The system will implement three key algorithms: Support Vector Machine (SVM)[13,14], Logistic Regression, and Naïve Bayes Algorithm. To ensure robust model performance, 75% of the data will be used for training the algorithms, while the remaining 25% will be reserved for testing and validating the model's accuracy.

These approaches aim to improve the precision of heart disease risk predictions, offering reliable results to users [5]. Steps will also be taken to optimize the algorithm and thus increase accuracy. This step involves cleaning the dataset and previous data.

These algorithms were determined to be accurate, and SVM was found to be the most accurate among our algorithms, with an efficiency of 64.4%. That's why it was chosen as the main application. The results are shown along with the accuracy of the prediction.

A. Online Website

The system will include a website where users will sign up to receive heart health information based on heart disease prediction.

Users must first fill out the registration form. After submitting their health information, patients will be able to view reports to find out their heart disease or percentage hmm. If the user's risk is higher than 60%, the user will be sent to another form where they must enter additional symptoms so that the system can predict two types of heart disease. Most groups are CAD (cardiovascular disease) and valvular disease.
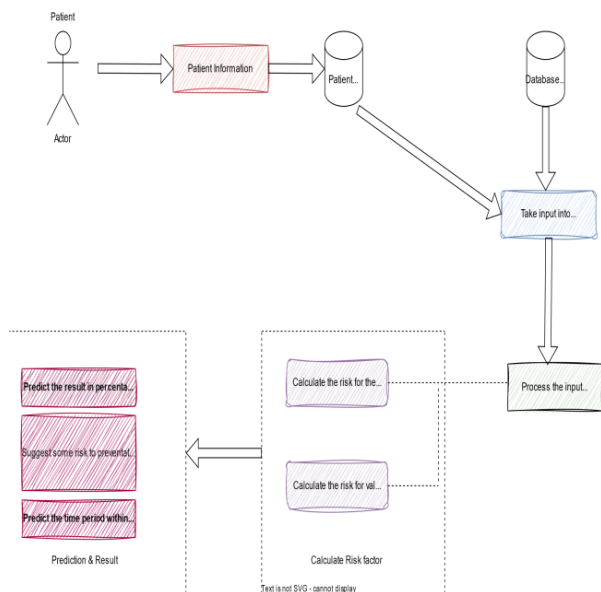

Fig.1. Block Diagram for Heart Disease Prediction

B. Database

The dataset is sourced from Kaggle, and MySQL will be utilized by the server to manage data. This information is stored in the system's database. The user's medical history, including details such as age, gender, resting blood pressure, cholesterol levels, fasting blood sugar, and senior status, will be included in the records.

C. Machine Learning Algorithm

Machine learning classification is a supervised learning method that involves predicting the category or class of an input based on labeled data heart disease prediction, the dataset may include features such as age, gender, blood pressure, cholesterol levels, and other medical factors [5].

5. METHODOLOGY

In order to get a very accurate forecast based on the data set collected, we find that the following activities need to be done. The activities we want to do are as follows. That's why we do the following for him first. Invalid fields will be skipped. We will replace this with the meaning of the column. In this way we will eliminate all costs of data. Features may not work properly in K-mean classifier. Therefore, the spacing of all features must be scaled to ensure that each feature is supported roughly proportional to the tip distance. The result is evaluated individually like, age has only two values; 0.1, cholesterol is like 100. So, we need to measure them to keep them close together.

Factoring: In this step, we give values particular meanings to avoid the algorithm misinterpreting them. In the age section, for example, we define the values 0 and 1 to guarantee that the algorithm makes the proper distinctions when processing data. This approach ensures accurate classification by preventing confusion. The Support Vector Machine (SVM), a potent learning method created for distribution control and inverse analysis, is the foundation of the analysis algorithm we employ. Two different groups of labelled training data are fed into the SVM model. After that, the SVM algorithm builds a model that assigns new data points to one of these two categories [9]. he SVM (Support Vector Machine) model maps data points in a high-dimensional space and transforms them into instances of distinct classes, aiming to spread them as widely as possible. In the context of heart disease prediction, the SVM algorithm learns to classify data based on features such as age, cholesterol levels, and other medical parameters. The model identifies an optimal hyperplane that separates these classes (e.g., "high risk" vs. "low risk") in the feature space, ensuring the maximum margin between them. New data points are then classified based on which side of the hyperplane they fall on. While the SVM method is effective in heart disease classification, its limitation lies in the difficulty of handling very large datasets with numerous features, as well as the need for careful parameter tuning. Additionally, SVM may struggle with noisy data or when classes are not well-separated in high-dimensional spaces [11,13],14]. Machine learning method is required that tries to find natural objects in the group and then combine the data with a new text for the design team. In this project, we use this algorithm to group patients according to their given risks. The accuracy of Naive Bayes was found to be 60%, Logistic Regression 61.45% and SVM 64.4%. For this reason, SVM was chosen as the best algorithm for web applications.

CONCLUSION

Prediction of heart disease is an important and important task in medicine [10]. If this disease is recognized in the early stages, preventive measures are taken and appropriate treatment is provided, the mortality rate will decrease. The aim of the plan is to accurately predict the disease. The decision-making process depends on age, BMI, cholesterol, etc. It turned out that using features can be effective in predicting the disease. The proposed approach offers more accurate cardiovascular disease predictions. Our project focuses on designing a system that processes and evaluates patients' real-time data, considering both past and present symptoms for various diseases [12]. So, by measuring the results the plan can predict further heart disease. Current signs and symptoms. The implementation requires less budget and can be easily implemented in Web-based systems.

R<small>EFERENCES</small>

[1] Sibhuraj Pappu, Hamza Girach, Mohd Zaid Gandhi, Adnan Hathiyari, Mohammed Juned "Heart Disease Prediction using Machine

[2] Learning", JETIR April 2022, Volume 9, Issue 4.

[3] Selina McKee "Cancer rates in women rising six times faster than men", Feb 2017.

[4] Soodeh Nikan, Femida Gwadry Sridhar, Michael A.Bauer "Machine Learning Application to Predict the Risk of Coronary Artery Atherosclrosis", IEEE, DEC 2016.

[5] Sanya Raghuwanshi "Understanding Data Preprocessing taking the Titanic Dataset", Sep 2020.

[6] Saini, K., Sonone, S. S., Sankhla, M. S., & Kumar, N. (Eds.). (2024). Artificial intelligence in forensic science: An emerging technology in criminal investigation systems. CRC Press.

[7] Saini, K., Bharadwaj, S., & Gupta, V. (2022, April). Face Mask Detection: A Deep Learning Concept. In 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM) (pp. 437-441). IEEE.

[8] Abhijeet Jagtap, Priya Malewadkar, Omkar Baswat, Harshali Rambade "Heart Disease Prediction using Machine Learning", Volume-2, Issue-2, February-2019.

[9] V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja "Heart disease prediction using machine learning tech: A survey" International Journal of Engineering & Technology, 7 (2.8), April 2018.

[10] Mall, S., & Singh, J. (2023, March). Heart diagnosis using deep neural network. In 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) (pp. 7-12). IEEE.

[11] Galla Siva Sai Bindhika, Munaga Meghana, Manchuri Sathvika Reddy "Heart Disease Prediction Using Machine Learning Techniques", (IRJET)April 2020.

[12] Rahul Vashistha, Aditya Randive, Pallavi Gade, Gaurav Pardeshi "International Journal of Advanced Research in Computer and Communication Engineering", Vol. 11, Issue 4, April 2022.

[13] Mall, S., & Singh, J. (2023, June). Comparative Study of Heart Failure Using the Approach of Machine Learning and Deep Neural Networks. In International Conference on Data Analytics & Management (pp. 627-641). Singapore: Springer Nature Singapore.

[14] Mall, S., Srivastava, A., Mazumdar, B. D., Mishra, M., Bangare, S. L., & Deepak, A. (2022). Implementation of machine learning techniques for disease diagnosis. *Materials Today: Proceedings*, *51*, 2198-2201.

[15] Kumar, S., Negi, A., J.N.Singh. (2019). Semantic Segmentation Using Deep Learning for Brain Tumor MRI via Fully Convolution Neural Networks. In: Satapathy, S., Joshi, A.(eds) Information and Communication Technology for Intelligent Systems. Smart Innovation, Systems and Technologies, vol 106. Springer, Singapore.

[16] Agrawal, V., Singh, J.N., Negi, A., Kumar, S. (2020). Comparison-Based Analysis of Travel Time Using Support Vector Regression. In: Yadav, S., Singh, D., Arora, P., Kumar, H. (eds) Proceedings of International Conference in Mechanical and Energy Technology. Smart Innovation, Systems and Technologies, vol 174. Springer, Singapore. https://doi.org/10.1007/978-981-15-2647-3_20.

[17] Kumari, R., Saini, K., & Anand, A. (2022). Predictive analytics to improve the quality of polymer component manufacturing. Measurement: Sensors, 24, 100428.

[18] Ahluwalia, J., Joshi, J., & Saini, K. (2022, December). Camera Nerve Center: A Modern CCTV Surveillance System. In 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 2235-2238). IEEE.

[19] Saini, K., Suresh, M., Sankhla, M. S., & Parihar, K. (2021, December). Analysis of Student Performance and Requirement Using Data Science. In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 66-70). IEEE.

[20] Singh, P., Nagill, J., & Saini, K. (2023, December). Using Supervised Learning for Breast Cancer Detection using AI&ML. In 2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 281-285). IEE