International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.017 ISSN: 2581-4621

MACHINE LEARNING BASED HEART DISEASE PREDICTION SYSTEM

Shaik Mahaboob Basha[1], H Ateeq Ahmed[2]
[1]M. Tech.Scholar,[2]Assistant Professor
[1,2]Department of Computer Science and Engineering
1,2 Dr.K.V. Subba Reddy Institute of Technology Kurnool District, A.P, India

Abstract: In today's world, heart disease remains one of the leading causes of death globally. The medical field continuously gathers vast amounts of data to diagnose diseases, but not all collected information is relevant for making accurate decisions. Detecting the early onset of heart disease can be challenging as it requires expert knowledge and experience in recognizing symptoms. Additionally, medical datasets are often scattered, diverse, and heterogeneous, making analysis complex. Data mining plays a crucial role in extracting hidden, meaningful, and predictive insights from large databases. This paper utilizes information gain-based feature selection to eliminate irrelevant data and applies various classification techniques—such as K- Nearest Neighbors (KNN), Decision Tree-ID3, Gaussian Naïve Bayes, Logistic Regression, and Random Forest—on a heart disease dataset to improve prediction accuracy. The classification models are evaluated using performance metrics, including accuracy, ROC curve, precision, recall, sensitivity, specificity, and F1-score. Among these techniques, the Decision Tree model demonstrated the highest accuracy, achieving approximately 97% classification accuracy.

1. INTRODUCTION

1.1 GENERAL:

Data mining is the process of sifting through massive databases to uncover valuable insights. It takes raw data and transforms it into meaningful information, helping to identify hidden patterns and relationships that might otherwise go unnoticed. The process involves several key steps, including data cleansing, integration, selection, transformation, mining, pattern analysis, and knowledge representation.

In the medical field, data mining faces significant challenges due to the uncertainty and complexity of healthcare data. Traditionally, clinical decisions rely heavily on a doctor's intuition, which, while valuable, can sometimes lead to errors. These mistakes not only affect patient outcomes but also contribute to rising medical costs.

To improve efficiency and accuracy, data mining systems often use serialization, a method that converts data objects into byte streams before storing them in a database. This ensures that the data is organized and easily accessible for future analysis.

2 . PROJECT DESCRIPTION

2.1 GENERAL:

Support Vector Machine (SVM) is a powerful machine learning technique that has demonstrated remarkable accuracy in disease prediction. Researchers have explored innovative methods, such as using discrete wavelet transforms to analyze iris images and assess a person's risk of developing diabetes.

To enhance the performance of these predictive models, improved k-means clustering and logistic regression were integrated using an adaptive approach. The results showed that the classifier performed significantly better when trained on a sufficiently large and diverse dataset, highlighting

DOI:10.46647/ijetms.2025.v09i02.017 ISSN: 2581-4621

the importance of high-quality data in improving disease prediction accuracy.

2.2 METHODOLOGIES

2.2.1MODELS NAME:

1. Dataset
2. Importing necessary libraries 3.Analyzing
4. Preprocessing
5. Split the data 6.Model 7.Prediction

3. SYSTEM REQUIREMENTS SPECIFICATION

3.1 GENERAL:

The results clearly show that the error rates for each database are remarkably low. This is largely due to the strong discriminatory power of the features and the effective regression capabilities of the classifiers. When we compare the highest accuracy levels—those with the lowest error rates—to previous studies, our approach proves to be highly competitive.

3.2 HARDWARE REQUIREMENTS:

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It should what the system do and not how it should be implemented.

- PROCESSOR: DUALCORE2DUOS.
- RAM:4GBDDRAM
- HARD DISK:250 GB

3.3 SOFTWARE REQUIREMENTS:

The software requirements document serves as a blueprint for the system, outlining what it should do rather than how it should do it. It includes both a clear definition and a detailed specification of requirements, ensuring that all expectations are well-documented.

This document plays a crucial role in the development process. It helps in estimating costs, organizing team activities, assigning tasks, and monitoring progress throughout the project. By providing a structured foundation it ensures that development stays on track and aligns with the project's goals.

- Operating **System:** Windows 7, 8, or 10
- Development **Platform:** Spyder 3
- Programming **Language:** Python
- User **Interface:** Spyder 3

3.4 LITERATURE SURVEY

TITLE: Growing Epidemic of Coronary Heart Disease in Low- and Middle-Income Countries

Authors: Thomas A. Gaziano, Asaf Bitton

Overview

Coronary heart disease (CHD) is the leading cause of death in developed nations and a growing health crisis in developing countries. In 2001 alone, CHD was responsible for 7.3 million deaths worldwide, with nearly 75% of these occurring in low- and middle-income countries.

The rapid increase in CHD cases in these regions is driven by socio-economic changes, longer life expectancy, and the adoption of lifestyle-related risk factors. However, the impact of CHD varies widely across developing nations, depending on factors such as risk exposure, healthcare resources, and each country's stage in the epidemiologic transition.

Beyond its devastating health effects, CHD also imposes a significant economic burden. Despite these challenges, solutions exist to help manage and reduce the impact of this growing epidemic.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.017 ISSN: 2581-4621

Title: The Rising Epidemic of Coronary Heart Disease in Low- and Middle-Income Countries
Authors: Thomas A. Gaziano, Asaf Bitton
Overview
Coronary heart disease (CHD) is the leading cause of death in developed nations and a major health challenge in developing countries. In 2001, CHD accounted for 7.3 million deaths worldwide, with nearly 75% of these occurring in low- and middle-income countries.
The growing burden of CHD in these regions is driven by socio-economic changes, longer life expectancy, and the adoption of lifestyle- related risk factors. However, the impact of CHD varies significantly across different developing nations. Factors such as varying risk levels, the presence of other competing health concerns, access to healthcare resources, and the stage of epidemiologic transition influence the incidence, prevalence, and mortality rates of CHD.
Beyond the devastating health consequences, CHD also imposes a heavy economic burden. Despite these challenges, effective strategies exist to help manage and reduce the impact of this growing epidemic, offering hope for improved heart health worldwide.

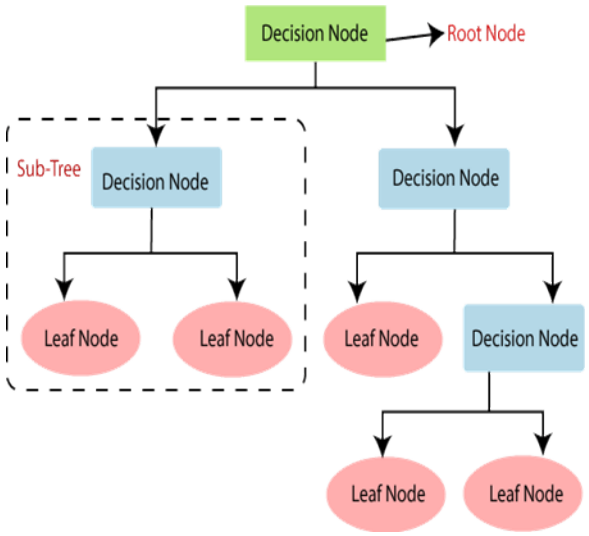
3.5 Existing System:

The medical industry is generating an increasing amount of data every day. Managing this vast volume of information and extracting meaningful insights for effective decision-making is a significant challenge. To address this, the industry requires specialized techniques that can analyze large databases and provide valuable, data-driven decisions to improve healthcare outcomes.

3.6 Proposed System:

Heart failure symptoms can occur at any stage of life, but older adults are at a higher risk compared to younger individuals. Advanced classification techniques can help identify hidden patterns and relationships within medical data, playing a crucial role in predicting health outcomes. By leveraging these hidden patterns and correlated features, healthcare professionals can make more accurate predictions and improve early detection of heart conditions.

4.0 System Architecture



5.0 FUTURE ENHANCEMENT:

This model could answer complex queries, each with its own strength with respect to ease of model

DOI:10.46647/ijetms.2025.v09i02.017 ISSN: 2581-4621

interpretation, access to detailed information and accuracy Optimize parameters such as tree depth, minimum samples per leaf, and splitting criteria through hyperparameter tuning.
• Derive new features from raw data, such as BMI and stress scores, to improve the model's ability to identify risk patterns.
• Incorporate genomic information to enable personalized risk prediction.
• Use real-time data from fitness trackers, such as heart rate and activity levels, for dynamic predictions.
• Combine decision tree analysis with insights from medical imaging, such as ECG or echocardiograms.
• Blend decision trees with neural networks for handling complex datasets.
• Implement Explainable AI techniques like SHAP or LIME to improve interpretability for clinicians.
• Deploy the model within healthcare systems for real-time risk assessments.

6.0 CONCLUSION:

With the rising number of heart disease cases each year, the medical field is generating vast amounts of data. Researchers are utilizing data mining techniques to analyze this information and improve heart disease diagnosis. Studies have shown that artificial neural networks are highly effective in classifying medical data and extracting meaningful insights.

As the global population grows rapidly, the death rate from cardiovascular diseases is also increasing. The best way to tackle this issue is through early prediction and timely medical intervention before the condition worsens.

Our hybrid approach enhances disease detection accuracy, achieving an impressive 97% success rate—outperforming previously proposed methods.

REFERENCES:

[1] Dewan, A., & Sharma, M. (2015, March). Prediction of heart disease using a hybrid technique in data mining classification. In Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on (pp. 704-706). IEEE.

[2] Dbritto, Rovina, AnuradhaSrinivasaRaghavan, and Vincy Joseph. "Comparative Analysis of Accuracy on Heart Disease Prediction using Classification Methods." International Journal of Applied Information Systems 11.2 (2016): 22-25.

[3] Shouman, M., Turner, T., & Stocker, R. (2012, March). Using data mining techniques in heart disease diagnosis and treatment. In Electronics, Communications and Computers (JEC- ECC), 2012 Japan-Egypt Conference on (pp. 173-177). IEEE.

[4] Lakshmi, K. R., Krishna, M. V., & Kumar, S. P. (2013). Performance comparison of data mining techniques for predicting of heart disease survivability. International Journal of Scientific and Research Publications, 3(6), 1-10.

[5] Kahramanli, H., &Allahverdi, N. (2008). Design of a hybrid system for the diabetes and heart diseases. Expert systems with applications, 35(1), 82-89.

[6] Kaur, B., & Singh, W. (2014). Review on heart disease prediction system using data mining techniques. International journal on recent and innovation trends in computing and communication, 2(10), 3003-3008.

[7] Deekshatulu, B. L., & Chandra, P. (2013). Classification of heart disease using k-nearest neighbor and genetic algorithm. Procedia Technology, 10, 85-94.

[8] Dr.Mohanraj, SubhaSuryaa, Sudha, Sarath Kumar, "Heart Disease Prediction using K Nearest Neighbour and K Means Clustering", International Journal of Advanced Engineering Research and Science (IJAERS) 2016.

DOI:10.46647/ijetms.2025.v09i02.017 ISSN: 2581-4621

[9] Shinde, R., Arjun, S., Patil, P., &Waghmare, J. (2015). An Intelligent Heart Disease Prediction System Using K-Means Clustering and Naïve Bayes Algorithm. IJCSIT) International Journal of Computer Science and Information Technologies, 6(1), 637-639.
[10] Kaya, Y., &Pehlivan, H. (2015, November). Feature selection using genetic algorithms for premature ventricular contraction classification. In Electrical and Electronics Engineering (ELECO), 2015 9th International Conference on (pp. 1229-1232). IEEE.
[11] Palaniappan, Sellappan, and RafiahAwang. "Intelligent heart disease prediction system using data mining techniques." Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on. IEEE, 2008.
` [12] Boutayeb, A., &Boutayeb, S. (2005). The burden of non communicable diseases in developing countries. International journal for equity in health, 4(1), 2.
[13] Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. Journal of healthcare information management, 19(2), 65.
[14] Srinivas, K., Rani, B. K., &Govrdhan, A. (2010). Applications of data mining techniques in healthcare and prediction of heart attacks. International Journal on Computer Science and Engineering (IJCSE), 2(02), 250-255.
complexity of these networks makes themattractivetargetsforcybercriminals.Tradi tional security measures often fall shortin addressing these threats, necessitating thedevelopmentofadvanceddetectionandresp onsemechanismstailoredforIIoTenvironment s.
complexity of these networks makes themattractivetargetsforcybercriminals.Tradition al security measures often fall shortin addressing these threats, necessitating thedevelopmentofadvanceddetectionandresponse mechanismstailoredforIIoTenvironments.