International Journal of Engineering and Techniques - Volume 12 Issue 4, July - August - 2026

Prediction of Heart Disease using Machine Learning

[1]Kshitija Chandrakant Gandle, [2] Dr. Hemant Wani, [3]Dr. Y. S. Angal, [4]Dr. T. K. Nagaraj

[1]*MTech Student at Dept. of Electronics & Telecommunication Engineering, JSPM's Bhivarabai Sawant Institute of Technology and Research, Pune, MH, India*

[2]*Assistant Professor, of Electronics & Telecommunication Engineering, JSPM's Bhivarabai Sawant Institute of Technology and Research, Pune, MH, India*

[3]*Associate Professor & Head, Dept. Of Electronics & Telecommunication Engineering, JSPM's Bhivarabai Sawant Institute of Technology and Research, Pune, MH, India*

[4] *Principal of JSPM's Bhivarabai Sawant Institute of Technology and Research, Pune, MH, India*

Corresponding author email: g.kshitija96@gmail.com and wanihemant1983@gmail.com

Abstract — On globally heart disease is one of the major causes of death and also early detection of heart disease plays a crucial role in reducing mortality rates. In the Traditional diagnostic methods depends largely on manual analysis, Limited statistical tools and clinical experience which is delayed and incorrect diagnosis. With the huge growth of healthcare data and machine learning techniques offers a effective solutions in early stage and accurate prediction of heart disease. In this project I proposed an advanced machine learning based system for predicting the "Heart Disease" using patient's medical attributes like age, cholesterol level, Heart rate, blood pressure and also other medical parameters. In This system employs multiple Machine Learning algorithms such as Random Forest, Support Vector machine (SVM), Gradient Boosting and Artificial Neural Network (ANN) to improve the accuracy of predicted data. Data pre-processing, feature selection and modern evaluation techniques are applied to enhanced System Performance. This system provides addition to disease prediction, The proposed system provides a personalized prediction, Recommendations and suggest suitable doctors and also patients condition. The results demonstrate that the proposed approach achieves higher accuracy and also reliability as compared to previous traditional methods. This system can serve as an effective decision-support tool for healthcare professionals and also can be further extended for real-time and large-scale healthcare applications.

Keywords —

Heart Disease Prediction, Machine Learning Algorithms, Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest, Healthcare Analytics, Disease Diagnosis, Data Preprocessing.

I. INTRODUCTION

Heart disease is one of the most serious and life-threatening health issues facing people around the world. The conditions that affect the structure and functioning of the heart, It includes the heart valve problem, coronary artery disease, heart rhythm disorders and heart failures. According to the stats of global health Organizations cardiovascular disease is one of the leading cause of death worldwide, its accounting the millions of deaths per year. For improve patients survival rate and reducing the mortality we have to focus on continuous monitoring and also early diagnosis. In previously heart disease diagnosis depends on multiple clinical examinations, laboratory tests and experience in medical professionals. These methods are effective but time consuming and prone to human error, especially when dealing with the large volumes of patient data. With the rapid growth of the digital healthcare systems, Hospitals and medical colleges and Institutes nowadays generate massive amount of health-related data. Extract the useful and meaningful data from this large Dataset using the conventional techniques is challenging. Machine Learning (ML) and Artificial Intelligence (AI) It is a powerful technique to analyze complex and large datasets. The techniques like a Machine Learning (ML) algorithms can identify the Hidden patterns, relationships between the data and trend's within the data those are not easily detectable by the Humans. By applied machine learning techniques to the healthcare data, it's possible to build intelligent systems that can be assist the doctors in predicting disease's in early stages. This project focuses on the prediction of heart disease using the machine learning algorithms to support medical decision making and enhancing the patient care.

II. RELATED WORK

Heart disease is one of the leading causes of death worldwide. Early diagnosis and risk prediction are crucial areas of research. Traditional diagnostic methods depend on clinical examinations, lab tests, and physician expertise. These methods can be time-consuming and may not always detect patients at an early stage. With the rapid growth of healthcare data and advances in artificial intelligence, machine learning has become an effective way to predict heart disease using patient medical records. Researchers have tested various classification algorithms to improve diagnostic accuracy and help healthcare professionals make timely decisions.

Several studies have used "Logistic Regression" for predicting heart disease because of its simplicity, interpretability, and effectiveness in binary classification problems. Logistic Regression looks at the relationship between multiple clinical factors like age, cholesterol level, blood pressure, fasting blood sugar, and chest pain type to assess the probability of heart disease. While this model performs well as a baseline and is easy to understand, its predictive ability may decline with complex nonlinear relationships among medical features.

Researchers have also tried "Decision Tree" and "Random Forest" algorithms for diagnosing heart disease. Decision Trees create clear and understandable decision rules, making them useful for medical applications where interpretability matters. However, Decision Trees can overfit when trained on complex datasets. Random Forest addresses this issue by combining several decision trees through ensemble learning, which leads to higher accuracy, improved stability, and better generalization. Many studies report Random Forest as one of the best-performing machine learning algorithms for heart disease prediction.

Another popular algorithm is the "Support Vector Machine (SVM)", which effectively handles high-dimensional medical datasets. SVM creates an optimal decision boundary to separate patients with and without heart disease. Its ability to classify complex data with precision has shown strong prediction performance in various heart disease studies. However, choosing suitable kernel functions and tuning model parameters are crucial for achieving the best accuracy.

The "K-Nearest Neighbors (KNN)" algorithm has also been investigated for heart disease prediction. KNN classifies new patient records by comparing them to similar cases in the training dataset. While KNN is easy to implement and often yields satisfactory results, its performance relies on selecting the right value of K and ensuring proper feature normalization. Researchers have found that KNN works well on smaller datasets but can become computationally expensive with larger healthcare databases.

Recently, "Artificial Neural Networks (ANNs)" and "Deep Learning" models have gained traction due to their ability to learn complex nonlinear relationships among medical factors. These techniques automatically discover hidden patterns in patient data and often achieve higher prediction accuracy than traditional machine learning algorithms when trained on large datasets. However, they require more computational resources and larger amounts of labeled medical data, which may not always be available in clinical settings.

Recent research has also looked into ensemble learning methods such as XGBoost, AdaBoost, Gradient Boosting, and Voting Classifiers. These methods combine predictions from multiple machine learning models to enhance overall performance, reduce overfitting, and boost robustness. Comparative studies consistently show that ensemble approaches achieve better accuracy, precision, and recall than individual classification algorithms.

Building on previous research, the proposed "Prediction of Heart Disease Using Machine Learning" system implements and compares multiple machine learning algorithms using patient health data. The model is trained and evaluated using standard metrics like accuracy, precision, recall, and F1-score to find the most suitable classifier for heart disease prediction. The goal of this system is to assist healthcare professionals by providing quick, accurate, and reliable predictions that support early diagnosis and timely medical intervention.

III. METHODOLOGY

In This Section a comprehensive methodological framework adopted for the design, development and evaluation of heart disease predictions. This Methodology follows the Systematic pathway that transfer the raw data into meaningful clinical predictions using machine learning algorithms. Heart disease data is complex because of heterogenous feature distribution, missing values, class imbalance and also inter feature correlations. In clinical systems demand transparency, where predictions must be explainable to medical professionals. For tackle this problem we proposed methodology integrate with advanced data preprocessing strategies, hybrid feature selection and ensemble learning model that can be capable to capture linear and non-linear relationships within the data. This Approach inspired from CRISP-DM (Cross-Industry Standard process for Data Mining) Framework which is widely accepted by machine learning and healthcare analytics. Using this methodology, we can tackle various problems using early detect the problem, reduced diagnostic delay and improved clinical decision support. The UCI Heart Disease dataset serves as primary experimental benchmark because of widespread acceptance in cardiovascular research. To further validate generalizability and robustness we employed auxiliary datasets and synthetic augmentation techniques. This methodology ensure that this developed system not merely a

predictive model but also decision support toll capable to providing actionable risk assessments.

IV. OBJECTIVE

The main objective of the project "Prediction of Heart Disease Using Machine Learning" is to develop an intelligent system that can predict the probability of heart disease using patient's medical data and machine learning algorithms with high accuracy.

The objectives in particular are:

- Gather and clean heart disease datasets, handling missing data, encoding categorical features, and normalizing numerical features.
- To identify the main risk factors of heart disease like age, blood pressure, cholesterol, type of chest pain, fasting blood sugar and max heart rate.
- To implement and compare various machine learning algorithms like Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN).
- To assess the performance of various models using metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC.
- The goal is to create an easy-to-use prediction system. Users can enter medical parameters and get immediate heart disease risk predictions.
- This will help healthcare professionals detect diseases early and support their decision-making.
- We aim to cut down diagnostic time and improve the heart disease screening process using artificial intelligence techniques.
- This will show how machine learning can improve healthcare results and preventive medicine

V. CONCLUSION AND FUTURE WORK

A. CONCLUSION

This research effectively developed an advanced machine learning-based analytical system for the early prediction of a heart disease. Integrating outdated clinical parameters with sophisticated computational models this study addresses the limitations of a manual diagnostic methods and also which are often prone to human error and delays. The implementation of a difficult preprocessing pipeline — incorporating SMOTE for class balancing and IQR for an outlier mitigation it ensured high data integrity. Furthermore, the hybrid feature selection strategy is optimized the model by identifying critical predictors such as age, chest pain type, and cholesterol levels. Among the evaluated algorithms, the ensemble approach and Artificial Neural Networks (ANN) demonstrated superior performance in capturing complex, non-linear relationships compared to baseline models, achieving the higher accuracy and reliability. Beyond the mere binary classification, the system's ability to provide a personalized recommendations and the risk social stratification transforms it into a feasible scientific decision support tool. This research bridges the gap between massive healthcare datasets and also actionable medical insights, particularly for the remote or underserved areas. Future enhancements will focus on the integrating real-time data from wearable sensors and implementing the Federated Learning to ensure patient privacy and ultimately facilitating for a large-scale transition toward proactive and personalized cardiovascular care.

B. FUTURE WORK-

Although the proposed system is demonstrates the promising results and several areas remain for further enhancement and Future work will focus on:

– For continuous monitoring we integrated with real-time ECG and wearable sensors.

– Implementation of the federated learning techniques to ensure the patient data its privacy and security.

– Deploying of the model within electronic health record (EHR) systems for the large-scale hospital use.

– Validating the multi-centre and diverse demographic datasets to improve the generalizability.

– The proposed framework can be serve as a scalable and reliable decision support tool for recent healthcare systems and also has strong potential for the real-world clinical deployment.

VI. DECLARATIONS

This academic project titled "Prediction of Heart Disease Using Machine Learning" is prepared to demonstrate the application of machine learning techniques in healthcare. It is designed and implemented using standard machine learning algorithms and publicly available datasets for educational and research purposes. The objective of this project is to assist in the early prediction of heart disease by analyzing various clinical parameters. Every effort has been made to ensure the implementation and evaluation of the machine learning models are correct. The system is able to predict based on the patterns of historical data and should be considered as an assistant for medical professionals. This project is not intended as a replacement for professional medical diagnosis,

consultation or treatment. Patients should seek qualified medical practitioners for the diagnosis and treatment of any health problems. The authors declare that the work presented in this project is original and that all sources (datasets, libraries and references) used in the development have been properly acknowledged.

VII. ACKNOWLEDGMENTS

I would like to express sincere gratitude to the Department of ENTC: VLSI and embedded system Engineering for providing to me necessary infrastructure and technical support to carry out this research work. I really thankful to my supervisor Prof.H.A.Wani and Head of Department Dr. Y. S. Angal under his direction we study project thoroughly and Thanks to Honourable Principal Dr.T.K.Nagaraj. whose always encourage the work and also make sure that necessary resources were available to us and also other faculty members for their valuable guidance, constructive feedback, and continuous encouragement throughout my study. I also acknowledge the contributions of the scientific community for providing publicly available datasets and the reference materials that supported the validating the proposed system.

VIII RESULT

The developed Heart Disease Prediction System can predict the chance of heart disease by analysing the patient's health parameters like age, gender, blood pressure, cholesterol level, maximum heart rate, fasting blood sugar, chest pain type and other medical attributes. We tested and trained several machine learning algorithms on a publicly available heart disease dataset. The Random Forest classifier was the best performing algorithm among all the algorithms tested including Logistic Regression, Decision Tree, K-Nearest Neighbours and Support Vector Machine in terms of prediction accuracy. The trained model showed reliable performance with high accuracy, precision, recall and F1-score making it suitable to assist healthcare professionals in preliminary diagnosis. The system provides instant prediction based on user's input which helps to identify the persons who are prone to heart disease. Though the model is not a substitute for professional medical diagnosis, it can be a useful tool for decision support to enable early detection and timely medical intervention.

Key Outcomes:

- Successfully processed and analysed patient health data.
- Developed several machine learning algorithms for comparison.
- Random Forest model was used to achieve high accuracy of prediction.
- Implemented a user-friendly interface for patient details entry and prediction results display.
- Demonstrated that machine learning techniques can be used to predict heart disease early

.IX. REFERENCES

1. A. Alghamdi, M. Alghamdi, and S. Alghamdi, " Heart Disease Prediction Using ML," Journal of Imaging, vol. 11, no. 10, p. 124, Oct. 2025. [Online]. Available: https://www.mdpi.com/2673-4591/107/1/124.

2. M. Benhar, A. O. Boudhir, and A. Haqiq, "A systematic review of machine learning in heart disease prediction," Frontiers in Artificial Intelligence, vol. 7, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12614364/

3. S. Shah, D. Patel, and R. Shah, " A comprehensive review of machine learning for heart disease prediction," Frontiers in Artificial Intelligence, vol. 8, p. 1583459, 2025. [Online]. Available: https://www.frontiersin.org/journals/artificialintelligence/ articles/ 10.3389/frai.2025.1583459/full

4. M. A. Alghamdi, M. Alghamdi, and S. Alghamdi, " A proposed technique for predicting heart disease using machine learning algorithms, " Scientific Reports, vol. 14, p. 74656, Oct. 2024. [Online].https://www.nature.com/articles/s41598-024-74656-2

5. R. Krishna, S. Kumar, and A. Reddy, " Accurate Heart Disease Prediction using Machine Learning Techniques on Clinical Data," International Journal of Computer Applications, vol. 187, no. 65, Dec. 2025. [Online]. Available: https://ijcaonline.org/archives/volume187/number65/accurateheart-diseaseprediction-using-machine-learning-techniques-on-clinicaldata/