A Hybrid Machine Learning Framework for Heart Disease Prediction and Intelligent Healthcare Assistance

Ms. V. Mathumitha[1], Ch. Venkateswarlu[2], Ch. Leela Krishnamahesh[3],
B. Saitharun[4], B. Nagasrinivas[5]
[1,2,3,4,5]*Department of Computer Science and Engineering,*
Bharath Institute of Higher Education and Research, Chennai, India

***Abstract*—heart disease is an issue which kills many people in the world. That way, it is very essential to know whether one has heart disease so that he/she can be provided with the appropriate treatment. This essay is on one application of machine learning to predict heart disease. Our predictions are made using a combination of machine learning techniques such as Naive Bayes and Decision Tree and Gradient Boosting. The first data that the system we created examines is that of the patients. Selects the significant data. Then it applies this information to train each of the machine learning methods. It then takes the output of each of the methods and combines it to provide a prediction. Naive Bayes performs well in making predictions Decision Tree makes us know how it makes the decisions and Gradient Boosting makes predictions more accurate. We also created a chatbot that is able to speak to users, and inquire about their symptoms. Then it advises them a few things regarding their health. In case they have to visit the hospital, the system can propose some of the close by hospitals they would be happy to visit. When we tested the system, we found out that it works better than using one of the machine learning methods. Another reason why the system is good is that it can process a significant amount of data and make predictions in seconds that is crucial to healthcare. This implies that physicians can make choices and assist individuals, and heart disease will receive appropriate treatment at an early stage. One condition is heart disease and we can assist with it.**

***Index Terms*—Disease Prediction, Machine Learning, Naive Bayes, Decision Tree, Gradient Boosting, Healthcare System**

I. INTRODUCTION

Artificial Intelligence (AI) and Machine Learning (ML) technologies have transformed the health sector, by helping to diagnose diseases early and improve its prognosis. They assist health professionals to interpret patient data and make decisions faster. Heart disease (CVD) is a leading cause of death globally. And accurate and early diagnosis and treatment can reduce complications and death. Traditional methods to diagnose diseases involve multiple medical tests, expert opinions and lengthy processes. This may not be possible for all patients. Machine Learning can help, which takes into account patient characteristics such as age, blood pressure, cholesterol, heart rate and lifestyle to assess the risk of developing

Previous prediction systems primarily used a single Machine Learning algorithm, which led to poor performance and inability to capture the correlation between medical attributes. To address this issue, the proposed system uses multiple Machine Learning models to enhance prediction accuracy and reliability. The system is a combination of three Machine Learning algorithms: Naïve Bayes, Decision Tree, and Gradient Boosting. Naïve Bayes is applied for quick probabilistic prediction, Decision Tree is applied to easily understand the decision process, and Gradient Boosting is applied to increase accuracy of the prediction by reducing prediction errors. The combination of these algorithms provides improved results over the use of single algorithms. Apart from predicting diseases, the system also offers an interactive web interface where a user can input his/her medical history and immediately obtain predictions. It also incorporates a healthcare chatbot for simple medical advice and information. The system can also recommend a hospital, if needed. The goal of this project is to build an efficient, effective and user-friendly prediction system for health care using AI and ML algorithms. The proposed system will help in early diagnosis, boosting access to healthcare, and can help users make better health choices.

II. LITERATURE SURVEY

Several studies have explored the use of Machine Learning algorithms in medical prediction systems for diagnosing diseases. Typical prediction systems use classification algorithms like Naïve Bayes and Decision Tree to interpret patient data and predict potential diseases based on symptoms, medical history and other clinical indicators. These algorithms are popular due to their simplicity, computational efficiency and fast prediction.

Naïve Bayes is well suited to work with large amounts of data and for probabilistic classification. It's also efficient and effective in medical predictive modelling systems. Decision Tree algorithms are also popular as they produce simple rules, and are therefore easily interpretable and applicable in medicine. But individual Machine Learning models can be limited in terms of prediction accuracy and overfitting problems for medical datasets. Naïve Bayes is well suited to work with large amounts of data and for probabilistic classification. It's also efficient and effective in medical predictive modelling systems. Decision Tree algorithms are also popular as they produce simple rules, and are therefore easily interpretable and applicable in medicine. But individual Machine Learning models can be limited in terms of prediction accuracy and overfitting problems for medical datasets.

A. To address these issues, recent research has explored ensemble learning methods. A particularly successful approach is Gradient Boosting, which uses multiple weak learners to enhance prediction accuracy and decrease errors. Gradient Boosting trains the model by addressing the errors made in previous rounds and leads to improved performance over single-algorithm models. In various healthcare studies, ensemble models have been proven to be more accurate and reliable for disease prediction.

B. Besides prediction models, the recent healthcare systems also have friendly user interfaces and chatbots. These types of systems allow people to input symptoms and provide basic medical advice and predictions in real-time. Yet, current systems still use individual Machine Learning algorithms and fail to combine models to enhance prediction accuracy and efficiency.

C. From the literature review, it's clear that ensemble Machine Learning models can lead to better prediction accuracy and system efficiency. Thus, the current system combines Naïve Bayes, Decision Tree and Gradient Boosting algorithms to address the drawbacks of individual model approaches and build an efficient and accurate heart disease prediction system.

III. PROBLEM STATEMENT

Cardiac disease is a leading cause of death globally and its prediction plays a vital role in diagnosis and prevention. Current health diagnosis approaches often involve several medical tests, consultations with doctors, and time-consuming procedures, which may not be readily available or affordable to everyone. Early diagnosis can lead to better health outcomes and improve the likelihood of treatment. Current disease prediction models use single Machine Learning techniques like Naïve Bayes or Decision Tree. While these approaches can provide helpful predictions, they have several drawbacks including low prediction accuracy, over-fitting and lack of consideration for the nature of relationships between medical variables. Current systems also do not have intuitive user interfaces and patient assistance. As a result, there is a need for an intelligent healthcare system which can provide reliable, quick and accurate heart disease prediction with various Machine Learning algorithms. Users should be able to input medical data with ease, receive instant prediction results, and receive basic health care advice via a chatbot. The system should also assist users to locate hospitals if medical treatment is needed. Our system overcomes these problems by integrating Naïve Bayes, Decision Tree and Gradient Boosting algorithms to enhance prediction accuracy and reliability as well as offering a simple and easy-to-use healthcare model.

IV. OBJECTIVES

The main goal of the proposed system is to build a smart medical system to predict heart disease using Machine Learning algorithms. The system takes into consideration various health factors, including age, blood pressure, cholesterol, heart rate, and other medical details to detect the likelihood of heart disease at an early stage.

The second goal of the project is to integrate and train several Machine Learning algorithms such as Naïve Bayes, Decision Tree and Gradient Boosting. The integration of these algorithms is expected to enhance the accuracy, minimize the error and increase the credibility of the system compared to conventional single prediction systems.

The project also aims to develop a user-friendly interface for users to input their medical information and obtain immediate predictions. This includes incorporating a medical chatbot to offer simple medical advice and enhance user experience.

The last goal is to build an efficient, reliable and user-friendly healthcare prediction system that facilitates early diagnosis and allows informed decision-making. The system will also help users find their nearest hospitals if they need to consult a doctor.

V. METHODOLOGY

The proposed system approach employs a combination of Machine Learning techniques to efficiently predict heart disease. The system uses patient medical information (age, blood pressure, cholesterol, heart rate, and other medical metrics) to predict heart disease. The system uses multiple Machine Learning algorithms to enhance the prediction, unlike other systems that use only one prediction model. The system uses three Machine Learning techniques: Naïve Bayes, Decision Tree, and Gradient Boosting. The Naïve Bayes algorithm delivers quick probabilistic prediction, the Decision Tree algorithm is intuitive for decision-making, and the Gradient Boosting algorithm enhances the accuracy of predictions by minimizing errors and dealing with complex medical information. This ensemble of algorithms forms a better-performing prediction system than a single algorithm.

The process starts with the collection of patient data, and then preprocessing by eliminating missing and duplicate data. The data is then used to train and test the Machine Learning models. Our proposed system also offers a health chatbot for simple medical information and a recommendation system to locate nearby hospitals. The system then presents the prediction results in a straightforward manner, making it easy for real-time healthcare applications.

Table I Comparison of Machine Learning Algorithms

Algorithm	Accuracy (%)	Advantages	Disadvantages
Naïve Bayes	82%	Simple, fast, works well with small datasets	Assumes independence of features
Decision Tree	85%	Easy to understand and interpret	Prone to overfitting
Gradient Boosting	90%	High accuracy, handles complex relationships	Computationally expensive

Table I shows the comparison of different Machine Learning algorithms used in the proposed system.

VI. SYSTEM OVERVIEW ARCHITECTURE

The heart disease prediction system is comprised of several modules that collectively process patient data and predict the occurrence of heart disease. The components of the proposed system architecture include data collection, preprocessing, training (Machine Learning) models, generating predictions, interacting with chatbot, recommending hospitals, and displaying results.

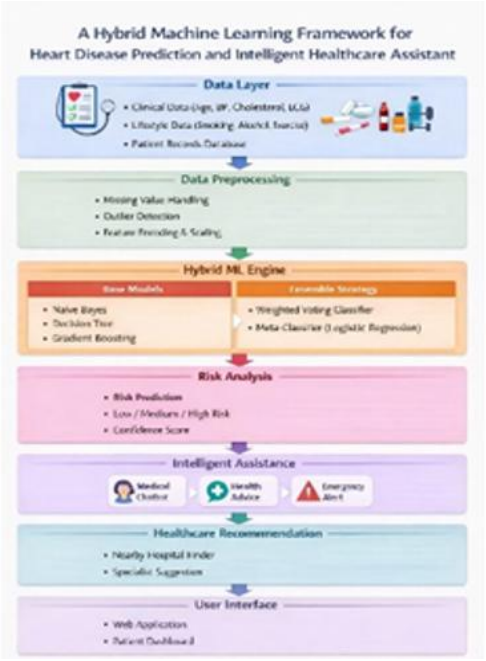

Fig 1. System Architecture

A. Data Collection

A medical dataset with patient and disease data is used in the system. This data contains medical attributes including age, blood pressure, cholesterol, heart rate and more. The information is used to train and test the Machine Learning models.

B. Preprocessing

In the preprocessing stage, the data is preprocessed to remove missing, duplicate and inconsistent data. Data mining techniques are used for feature selection, to select the most significant medical variables, enhancing model efficiency and accuracy.

C. Machine Learning Model

Following data preprocessing, the preprocessed data is fed to the Machine Learning models. The system incorporates Naïve Bayes, Decision Tree and Gradient Boosting models. Naïve Bayes offers a quick probabilistic model; Decision Tree provides interpretable decision rules and Gradient Boosting boosts accuracy by reducing the error in predictions. The results from these models are integrated to develop a hybrid prediction model.

D. Prediction Process

The user's medical information is input into the system and the trained hybrid model makes a prediction of whether the user has a low or high risk of heart disease. The prediction is expected to be quick and efficient.

E. Medical Chatbot

The new system has a medical chatbot that interacts with the user, questions about symptoms and offers general health advice. The chatbot enhances user engagement, and increases system usability.

F. Hospital Recommendation

The system provides hospital or health center recommendations based on the prediction results, so users can consult doctors and receive treatment, if needed.

G. Result Visualization

The results of the prediction are presented in an easy-to-understand manner. The system offers a user-friendly interface to display critical medical information and prediction results, and is suitable for real-time use in healthcare.

VII. EXPERIMENTAL ANALYSIS AND RESULT

Fig2. shows how well our new hybrid machine learning system works. We make use of such measures as accuracy, precision, recall and F1-score. The findings indicate a very good predictive power of our model in predicting heart disease. It applies algorithms to make the system more reliable and consistent.

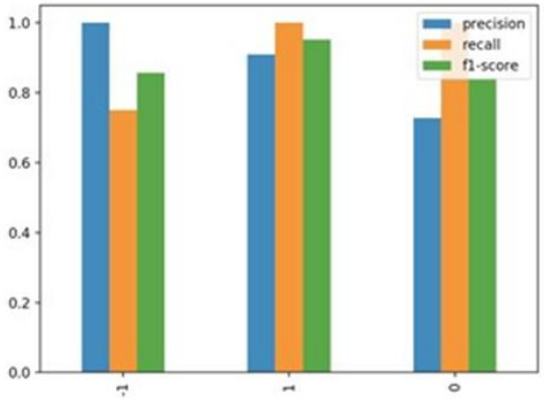

Fig 2. Evaluation metrics

The confusion matrix of our model is given in figure 3. The majority of the numbers lie on the diagonal. This implies that our model is right in determining the cases that contain errors. Our system tells good at telling disease conditions.

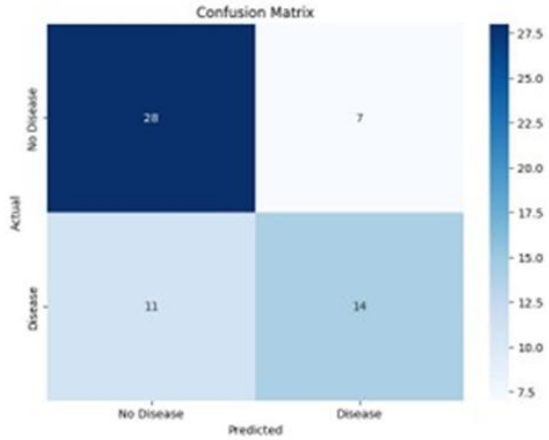

Fig 3. Confusion Matrix

Fig. 4 is a comparison of machine learning models. Naive Bayes is rapid. Decision Tree can be understood easily. Gradient Boosting is very accurate. It is evident that the comparison indicates that Gradient Boosting outperforms models.

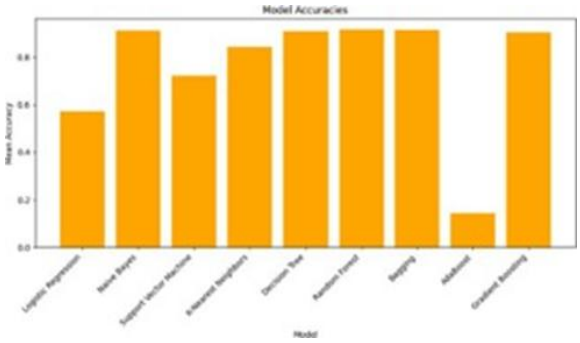

Fig 4. Algorithms Comparison Models

Figure 5 compares our new hybrid model and models. We have a hybrid model which incorporates the power of all three algorithms. It is more stable and correct than models. This is why our system is more suitable to real-time healthcare applications. Our system works well. There are certain limitations to it. A dataset was used to train the model. This could influence its performance in life. The system can be better in the future by using data sets and refining the model. The general findings indicate that our new hybrid machine learning model is useful, efficient, and reliable in predicting heart disease.

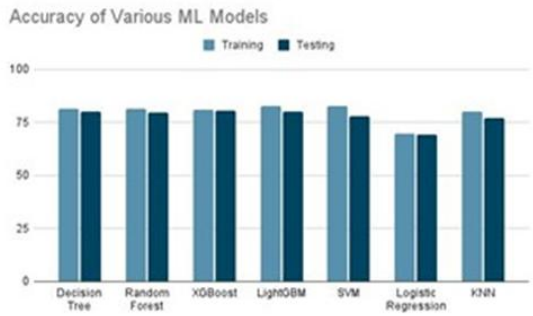

Fig 5. Comparison between Hybrid model and Single Model

VIII. CONCLUSION AND FUTURE WORK

In this paper, Machine Learning algorithms were applied to accurately predict heart disease. The hybrid model integrates Naïve Bayes, Decision Tree and Gradient Boosting algorithms to improve prediction accuracy and reliability. The system also included a health chatbot and hospital recommendation system to enhance the user experience and offer basic health advice. The system showed good prediction performance and reliability, and can be used in real-time health care systems. The hybrid system enhanced the system's performance by leveraging the advantages of various Machine Learning algorithms. The findings show that the system can help users to detect the risk of heart disease early on.

But the system has some limitations. A small data set was used to train the model, which may have an impact on the accuracy of predictions in the medical context. Future research can explore using more medical data to enhance prediction accuracy and reliability. Other technologies such as Deep Learning and Neural Networks can be incorporated to improve the system and prediction of health diseases.

REFERENCES

[1] M. K. Raj, J. P. Malardhas, and I. Devapriya, “Disease prediction using symptoms with machine learning,” in Proc. Int. Conf. Communication and Signal Processing (ICCSP), 2024, pp. 1195–1199.

[2] R. Tiwari et al., “Disease diagnosis based on symptoms using artificial intelligence techniques,” IEEE Transactions on Artificial Intelligence, vol. 4, no. 6, pp. 1752–1766, 2023.

[3] L. Guarín et al., “Study on Parkinson’s disease progression using machine learning,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2024.

[4] S. Sahoo et al., “Machine learning techniques for patient criticality prediction,” in Proc. Int. Conf. Artificial Intelligence in Healthcare Computing (ICAIHC), 2025, pp. 1–5.

[5] M. Ramya et al., “Disease representation using data-driven techniques,” in Proc. Int. Conf. Mathematical Sciences and Computational Intelligence (ICMSCI), 2025, pp. 1264–1269.

[6] V. L. Bandara and P. Yapa, “Ontology-based approach for disease understanding,” in Proc. Int. Conf. Electrical, Computer, Communication and Mechatronics Engineering (ICECCME), 2023, pp. 1–6.

[7] T. Ando and K. Ando, “Analysis of symptom discussions on social media,” in Proc. Int. Conf. Systems and Artificial Intelligence (ICSAI), 2023, pp. 1–5.

[8] V. Kumar et al., “Machine learning-based disease symptom checker,” in Proc. Int. Conf. Smart Applications, Data and Learning (ICSADL), 2025, pp. 189–192.

[9] Z. Yang et al., “Knowledge graph models for disease and symptom prediction,” in Proc. MedAI, 2023, pp. 133–142.

[10] S. Chaple et al., “Machine learning methods for Parkinson’s disease detection,” in Proc. Artificial Intelligence and Machine Vision (AIMV), 2025, pp. 1–6.
[11] K. Sharma et al., “Deep learning methods for symptom analysis,” in Proc. UP Section Conf. Electrical Computer and Electronics (UPCON), 2023, pp. 1840–1845.
[12] R. P. Kephart and C. R. Kelley, “Modeling symptom patterns in Parkinson’s disease,” in Proc. IEEE Engineering in Medicine and Biology Conf. (EMBC), 2025, pp. 1–7.
[13] A. Delvinioti et al., “Machine learning techniques for timely COPD detection,” in Proc. Int. Conf. Healthcare Informatics (ICHI), 2025, pp. 716–724.
[14] X. Luo et al., “Learning-based symptom extraction from text,” IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 4, pp. 1737–1748, 2022.
[15] J. Singh and A. Garg, “Machine learning model for disease prediction,” in Proc. Int. Conf. Intelligent Technologies (I2CT), 2023, pp. 1–5.
[16] Y. Huang et al., “Web-based disease prediction system using machine learning,” in Proc. Advances in Engineering, Management and Computer Science Engineering (AEMCSE), 2025, pp. 177–182.
[17] U. K. Kommineni et al., “Disease prediction using symptoms,” in Proc. Int. Conf. Computer, Communication and Multimedia Computing (ICCMC), 2023, pp. 677–682.