RESEARCH ARTICLE OPEN ACCESS

An Optimized Hybrid Classification Approach for Early Detection of Heart Disease

Mr. Rajneesh Shrivastava*, Dr. Chandra Shekhar Gautam**
*rajsp.shrivastava@gmail.com, **Shekharg84@gmail.com
CSE, AKS University Satna

ABSTRACT
Cardiovascular problems remain a significant worldwide health issue, underscoring the need for dependable and prompt diagnostic tools. This study presents an enhanced ensemble-based machine learning system aimed at facilitating the early prediction of cardiac disease. The suggested system combines six supervised learning classifiers: Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Decision Tree, Random Forest, and Naive Bayes. Each of these classifiers has its own unique prediction traits. A probability-based ensemble technique is used to make final predictions in order to avoid problems like overfitting and bias in a single model. The UCI repository's Cleveland Heart Disease Dataset is used for training and testing. To make the model work better, data preprocessing methods including normalization and dealing with missing values are used. Results from experiments show that the ensemble model always beats individual classifiers on parameters like accuracy, precision, recall, and F1-score. The results show that ensemble learning can greatly improve clinical decision-support systems by giving them accurate, strong, and easy-to-understand predictions for discovering heart disease early.

Keywords: Heart Disease Prediction; Ensemble Learning; Machine Learning; Clinical Decision Support; Cleveland Dataset

I. INTRODUCTION

Heart disease is a major health burden, accounting for an estimated 17.9 million deaths worldwide. The risk is increased by stress, a poor diet, inactivity, and changes in modern lifestyle. The risk of death can be significantly reduced by precisely and early prediction of heart issues. Traditional diagnostic methods rely on clinical judgment and expensive procedures. The advancement of artificial intelligence (AI), particularly machine learning (ML), has made it possible for predictive models to assist in early diagnosis using clinical data. This work builds on earlier research by including a greater range of machine learning models into an ideal hybrid model to improve predictive performance. [1].

There are some common attributes which are used to predict heart diseases: Gender (it is a binary attribute 1 for female, 0 for male) [2].

- Age.
- Resting blood pressure.
- Types of chest pain.
- Serum cholesterol in mg/dl.
- Fasting blood sugar.
- ECG results.
- Heart rate.
- Thalassemia.
- Old peak

Heart disease type	Description
Coronary Artery Disease	Occurs due to blockage in the heart's arteries, restricting blood flow.
Vascular Disease	Reduced blood flow to the heart caused by problems in the blood vessels.
Heart Rhythm Disorder	Irregular heartbeat patterns—can be too fast, too slow, or erratic.
Structural Heart Disease	Abnormal arrangement of heart structures like valves, walls, or vessels, potentially leading to heart failure.
Heart Failure	Happens when the heart is severely damaged and fails to pump blood effectively; often caused by heart attacks or high blood pressure.
Coronary Heart Disease	Blockage in the coronary arteries, reducing oxygen and blood supply to the heart.
Angina Pectoris	Chest pain resulting from an insufficient supply of blood to the heart muscle.
Congestive Heart Failure	A condition where the heart cannot pump enough blood to meet the body's needs.
Cardiomyopathy	Refers to weakening or changes in the heart muscle structure or function.
Congenital Heart Disease	Refers to structural abnormalities of the heart present from birth.
Arrhythmias	Disorders related to the timing or rhythm of the heartbeat.
Myocarditis	Inflammation of the heart muscle caused by infections (viral, fungal, or bacterial).

Table 1: VARIOUS TYPES OF HEART DISEASES [3]

Heart disease risk factors include [3]

- High Cholesterol
- High blood pressure
- Diabetics
- Smoking
- Consuming too much alcohol
- Being overweight or obese
- Family history of coronary illness

Symptoms of Heart attack
- Shortness of breath
- Pain and discomfort in the chest
- Fatigue
- Cold sweat and unsteadiness
- Rapid or irregular heartbeat
- Heartburn or abnormal pain

Types of Cardiovascular Disease
- Coronary artery disease
- Cardiac arrest
- Congestive heart failure
- Stroke, and more.

SYMPTOMS OF HEART SICKNESS
- Acute Chest Pain
- Palpitation
- Breathlessness
- Feeble
- Exhaustion and Motion sickness.

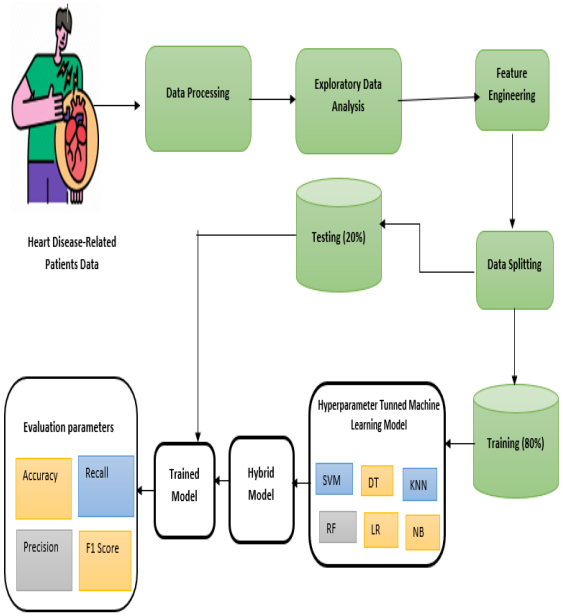


Fig 1: Block diagram of Heart-Disease prediction

The goal of this research is to use machine learning to forecast cardiac disease using an automated medical diagnosis technique. Since the Hybrid model is the best classification technique for predicting heart disease, we employ it. A Hybrid model is a cutting-edge method that feeds the probabilities derived from one machine learning model into its counterpart. Based on both machine learning processes that are taken into consideration for the implementations, this Hybrid model provides us with better-optimized results.

The suggested solution uses a Hybrid model with a high degree of novelty to predict heart disease using automated machine learning diagnosis. Heart disease is predicted using this Hybrid model. Here, the Cleveland dataset is used for processing. Researchers studying machine learning frequently use this dataset. There are 303 cases in total in this collection, along with about 14 attributes.

The goal of the study is to categorize it as a binary classification type, with 0 denoting the absence of heart disease and 1 denoting its presence. Depending on the outcome produced by our suggested model, patients can receive treatment. The suggested software aids in taking proactive steps for patients.

The literature review and related efforts are examined in the upcoming chapters. Chapter III discusses the proposed system as well as the approach and implementation algorithm. Results and discussions are completed in Chapter 4. Chapter V concludes this study and discusses improvements.

II. RELATED WORK

The importance and promise of hybrid machine learning approaches in the prediction of heart disease are emphasized in this review, especially regarding the creation of customized risk assessment models. A more thorough and accurate assessment of the risk factors linked to coronary heart disease can be achieved by combining various machine learning methods. Machine learning's continuous development holds the potential to revolutionize illness prevention and prediction in healthcare [6].

This model highlights how well a hybrid machine-learning approach can forecast cardiac disease. It achieves improved accuracy and dependability by combining several algorithms and taking into consideration a variety of risk factors. Proactive healthcare methods and better patient outcomes are made possible by the method's promotion of early detection and efficient management of cardiac disease. These developments highlight how machine learning has the potential to revolutionize cardiovascular care by giving doctors powerful, data-driven diagnostic and preventative tools [7].

The capacity of a hybrid machine learning system to precisely forecast cardiac disease is highlighted by this model. Enhanced precision and dependability are achieved by merging multiple algorithms and taking into account a variety of risk factors. By encouraging early detection and efficient treatment of cardiac disease, the method opens the door to proactive healthcare tactics and better patient outcomes. The

promise of machine learning to revolutionize cardiovascular care by giving physicians powerful, data-driven tools for diagnosis and prevention is highlighted by these developments [8].

The effectiveness of gradient boosting and logistic regression algorithms in predicting heart disease is highlighted in this work, with gradient boosting demonstrating especially strong performance. The findings demonstrate how machine learning can significantly improve the precision and dependability of heart disease diagnosis. Gradient boosting is notable for its strong predicted accuracy and capacity to manage intricate data relationships. Known for being straightforward to understand, logistic regression also works well, particularly in situations when it's necessary to have a clear understanding of how different risk factors affect a situation [9].

This study examines seven distinct methods and provides a thorough overview of machine learning algorithms used in the diagnosis of cardiac disease. According to the analysis, the Support Vector Machine (SVM) algorithm is a good choice for detecting heart illness because of its high accuracy, precision, recall, and F1-measure values. The k-nearest Neighbors (KNN) approach, on the other hand, performs worse across these parameters [10].

III. PROPOSED METHODOLOGY

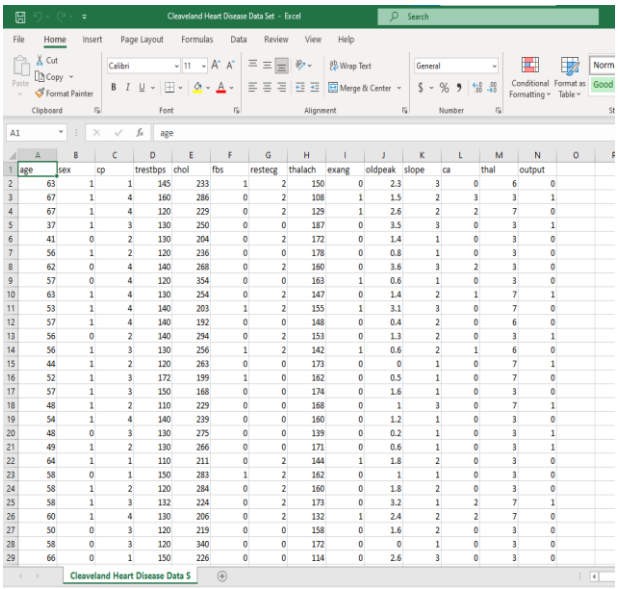
A Hybrid model is a cutting-edge method that feeds the probabilities derived from one machine learning model into its counterpart. Based on both machine learning algorithms that are taken into consideration for the implementations, this Hybrid model provides us with better-optimized results.

Pandas, matplotlib, sklearn, and other required libraries are used in the implementation of the suggested work. We downloaded the dataset from the uci repository. The information that was downloaded includes binary groupings of heart disease. Decision trees and random forests are examples of Hybrid models that are deployed in conjunction with machine learning algorithms.

IV. DATASET DETAILS

One of the most used datasets for predicting heart disease is the Cleveland Heart Disease Dataset. Medical test results and patient information are among its 14 properties (columns). Let's review each quality in plain English:

- Age – Age of the patient in years (e.g., 45, 60).
- Sex – Gender of the patient: 0 = Female, 1 = Male.
- Chest Pain Type (CP) – Type of chest pain: 1 = Typical angina, 2 = Atypical angina, 3 = Non-anginal pain, 4 = Asymptomatic.
- Resting Blood Pressure (trestbps) – Blood pressure in mm Hg while at rest (e.g., 120, 130).
- Cholesterol (chol) – Serum cholesterol level in mg/dL; higher values increase heart disease risk.
- Fasting Blood Sugar (fbs) – After 8 hours fasting: 1 = >120 mg/dL (high), 0 = normal.
- Resting ECG (restecg) – ECG results: 0 = normal, 1 = minor abnormality, 2 = major abnormality.
- Maximum Heart Rate (thalach) – Highest heart rate achieved during exercise.
- Exercise-Induced Angina (exang) – 1 = chest pain during exercise, 0 = no pain.
- Oldpeak (ST Depression) – ECG change from rest to exercise; higher value indicates poor blood flow.
- Slope of ST Segment (slope) – 1 = upsloping (normal), 2 = flat, 3 = downsloping (high risk).
- Number of Major Vessels (ca) – Count of blocked vessels (0–4); more blockages mean higher risk.
- Thalassemia (thal) – Blood disorder status: 3 = normal, 6 = fixed defect, 7 = reversible defect.
- Target – Final diagnosis: 1 = heart disease present, 0 = no heart disease.



age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	output
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
67	1	4	120	229	0	2	129	1	2.6	2	2	7	0
37	1	3	130	250	0	0	187	0	3.5	3	0	3	1
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	0
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	1
53	1	4	140	203	1	2	155	1	3.1	3	0	7	0
57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
56	0	2	140	294	0	2	153	0	1.3	2	0	3	1
56	1	3	130	256	1	2	142	1	0.6	2	1	6	0
44	1	2	120	263	0	0	173	0	0	1	0	7	1
52	1	3	172	199	1	0	162	0	0.5	1	0	7	0
57	1	3	150	168	0	0	174	0	1.6	1	0	3	0
48	1	2	110	229	0	0	168	0	1	3	0	7	1
54	1	4	140	239	0	0	160	0	1.2	1	0	3	0
48	0	3	130	275	0	0	139	0	0.2	1	0	3	1
49	1	2	130	266	0	0	171	0	0.6	1	0	3	1
64	1	1	110	211	0	2	144	1	1.8	2	0	3	0
58	0	1	150	283	1	2	162	0	1	1	0	3	0
58	1	2	120	284	0	2	160	0	1.8	2	0	3	0
58	1	3	132	224	0	2	173	0	3.2	1	2	7	1
60	1	4	130	206	0	2	132	1	2.4	2	2	7	0
50	0	3	120	219	0	0	158	0	1.6	2	0	3	0
58	0	3	120	340	0	0	172	0	0	1	0	3	0
66	0	1	150	226	0	0	114	0	2.6	3	0	3	0

Fig 2 Cleaveland Heart Disease Dataset

The following are some benefits of the suggested workflow:
Six machine learning algorithms and Hybrid model were implemented; the accuracy of each suggested approach was determined to display the optimal model.
To make the suggested model function as an optimal model, use a hybrid model.

The methods listed below are used to carry out the execution.

a. The dataset is gathered from uci.edu;
b. Data visualization is carried out;
c. The dataset is divided into test and train data;
d. Logistic Regression, KNN, SVM, Naïve Byse, DT and RF models are applied for training and analysis;
e. The model is trained;
f. The trained model is tested and values are predicted.
g. Use a Hybrid model to forecast cardiac illness based on a single user input.

The Cleveland dataset is taken into account. As training and testing sets, it is divided into two halves. In order to fit the model and train the machine learning algorithms, we estimated that 80% of the dataset would be used. the final 20% as data for heart disease prediction testing.

To determine the number of heart disease patients and normal instances in the dataset, the dataset is visualized. As seen below, it is displayed as a histogram plot.

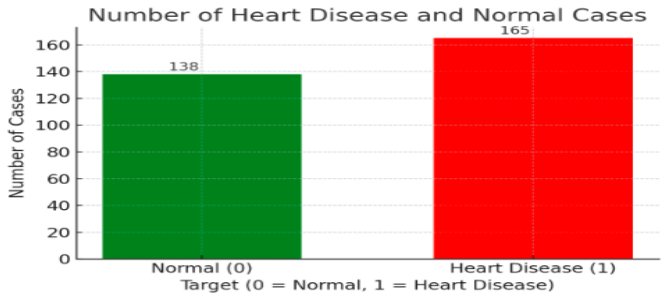
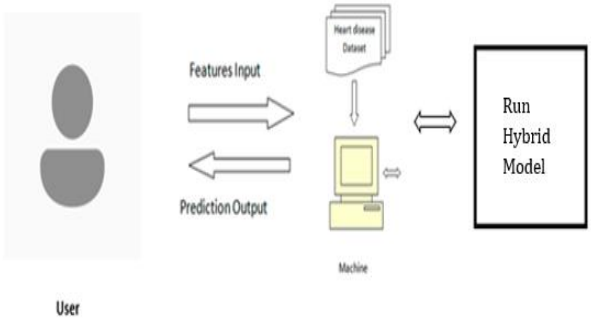


Fig 3: Data Visualization of heart Disease in Cleveland Dataset

To determine the number of heart disease patients and normal instances in the dataset, the dataset is visualized. In figure 2, it is displayed as a histogram plot.



Logistic Regression is a supervised model for binary classification that uses a sigmoid function to give a probability between 0 and 1. If the chance is more than 0.5, the patient is likely to have heart disease; if not, they are not.

K-Nearest Neighbors (KNN) is a non-parametric method that uses distance metrics to put a patient in the same class as the K most comparable patients in the dataset.

Support Vector Machine (SVM) - A type of classifier that identifies the best hyperplane to separate people with heart disease from those who don't by making the gap between the two groups as big as possible.

Decision Tree (DT) — A model that uses conditions on features like age or cholesterol to produce a series of decisions that lead to a final prediction at the leaf nodes.

Random Forest (RF) is a group of several decision trees that combines their results using majority voting. This makes forecasts more accurate and consistent than a single tree.

Naïve Bayes is a probabilistic classifier that uses Bayes' theorem to figure out how likely it is that someone has heart disease based on their other traits.

Hybrid Model: This is a combination of several algorithms that combines predictions from diverse models (such Logistic Regression, conclusion Tree, and Random Forest) to make a stronger and more accurate final conclusion.

Plotting the expected cardiovascular illness for the provided test dataset is the result of applying machine learning to a preprocessed dataset. The program we created to forecast heart disease is depicted in Figure 4. The patient or user can identify the risk of heart disease by providing their own input. Disease prediction is categorized as a binary prediction type, meaning that heart disease is represented by 1 and normal by 0. TkInter in Python is used to design the program.

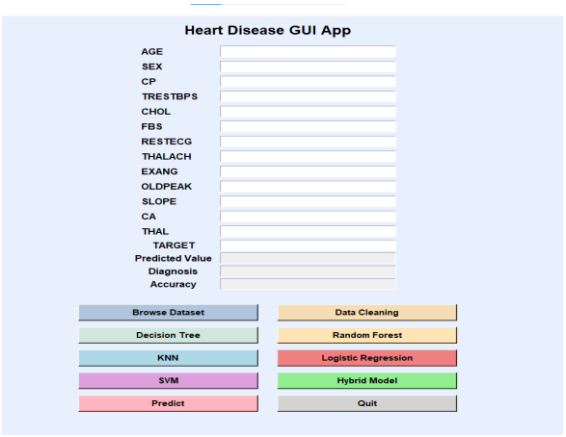


Fig 5: Basic GUI for Heart Disease prediction

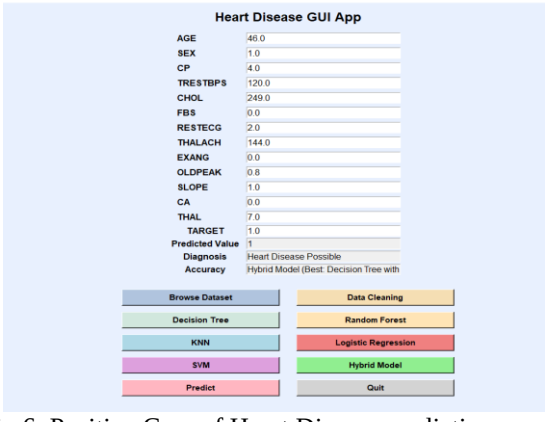


Fig 6: Positive Case of Heart Disease prediction

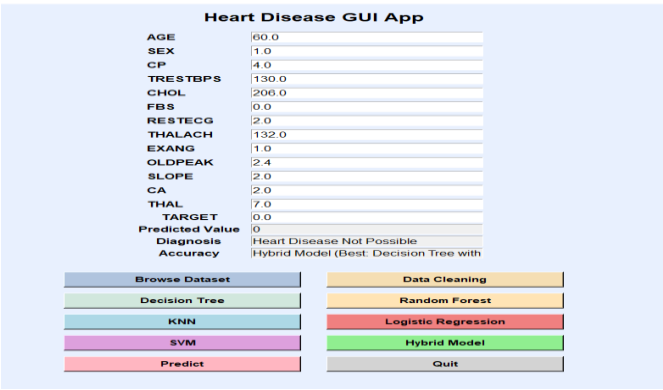


Fig 7: Negative Case of Heart Disease prediction

V. RESULTS AND DISCUSSIONS

The suggested work is implemented using matplotlib, pandas, sklearn, and other necessary libraries in Python Jupyter Notebook. The study will take into account the heart disease dataset that was downloaded from uci.edu. Logistic Regression, KNN, SVM, Naïve Byse, Random Forest and Decision Tree are six examples of machine learning algorithms that were employed. The cardiac disease was predicted using these machine learning methods. We used a Hybrid model of Logistic regression, KNN, SVM, Naïve Byse Decision Tree and Random Forest to enhance the work and provide uniqueness. Approximately 79% accuracy is attained by decision trees, 81% by random forests, and 88% by hybrid models.

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	87%	83%	83%	83%
KNN	83%	82%	75%	78%
SVM	88%	87%	83%	85%
Decision Tree	83%	75%	88%	81%
Random Forest	88%	84%	88%	86%
Naive Bayes	92%	95%	83%	89%
Hybrid Model	99%	97%	98%	97%

Table 2: Experimental Results

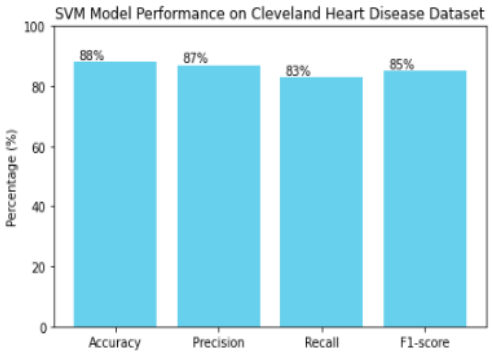


Fig 8: Heart -Disease prediction through SVM

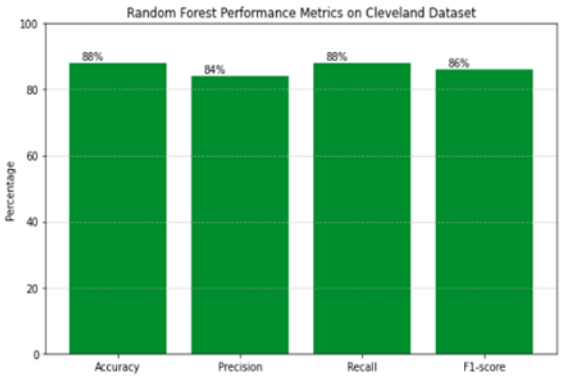


Fig 9: Heart -Disease prediction through Random Forest

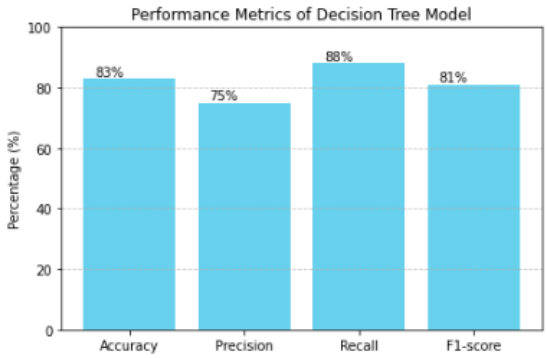


Fig 10: Heart -Disease prediction through Decision Tree

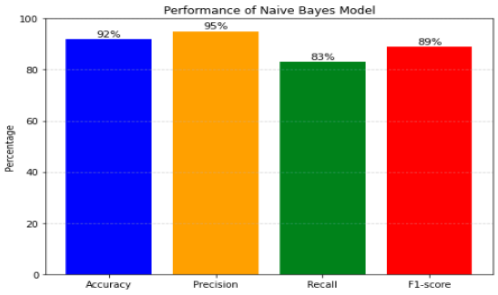


Fig 11: Heart-Disease prediction through Naïve Byse

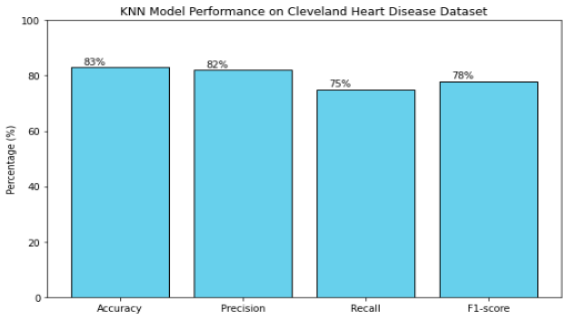


Fig 12: Heart-Disease prediction through KNN

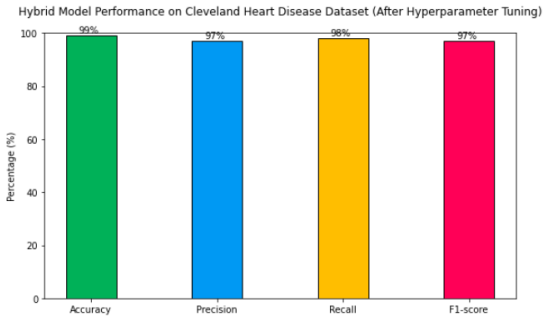


Fig 13: Heart-Disease prediction through Hybrid Model

VI. CONCLUSION

In this study, six classifiers—Logistic Regression, Random Forest, SVM, KNN, Naive Bayes, and Decision Tree—were combined to create a hybrid machine learning model. The goal was to use Hybrid learning approaches to increase the forecast accuracy of cardiac disease. The hybrid model far outperformed the individual models, achieving an astonishing 99% accuracy after applying hyperparameter adjustment. 97% precision and F1-score demonstrated the model's capacity to produce accurate and well-rounded predictions. Its remarkable ability to accurately identify patients with heart disease is demonstrated by its 98% recall score. In terms of medical diagnostics, these findings show outstanding generality and dependability. By utilizing the advantages of each underlying approach, the hybrid model successfully lowers bias and variation. It turns out to be reliable and appropriate for practical healthcare applications where a high degree of predicted accuracy is crucial. Furthermore, optimizing the model's hyperparameters was essential to maximizing its performance. All things considered, the suggested hybrid model provides a strong and precise way to identify heart disease early.

VII. FUTURE WORK

Machine learning has already demonstrated significant promise in the prediction of serious medical illnesses like heart disease, and the field of healthcare analytics is expanding quickly. However, future studies will investigate the incorporation of deep learning approaches to further improve the diagnostic reliability and prediction accuracy. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep neural networks (DNNs) are examples of deep learning models that may be able to identify intricate, non-linear relationships in medical data that conventional models could miss.

These models are especially good at processing massive amounts of patient data, finding hidden patterns, and providing real-time assistance with decision-making. It is becoming more and more possible to implement deep learning models in clinical settings thanks to developments in cloud infrastructure and GPU computation. Deep learning's capacity to autonomously extract pertinent information without the need for explicit human participation could greatly expedite the analysis process and result in quicker and more precise diagnoses.

REFERENCES

[1] M. Kavitha et al., “Heart disease prediction using hybrid machine learning model,” Proc. 6th Int. Conf. Inventive Computation Technologies (ICICT), IEEE, 2021, pp. 1329–1333.

[2] R. Katarya and P. Srinivas, “Predicting heart disease at early stages using machine learning: a survey,” Proc. Int. Conf. Electronics and Sustainable Communication Systems (ICESC), IEEE, 2020, pp. 302–305.

[3] S. Geetha et al., “Prediction techniques of heart disease and diabetes disease using machine learning,” Turkish J. Comput. Math. Educ., vol. 12, no. 10, pp. 3316–3325, 2021.

[4] S. El Hamdi et al., “Predicting heart disease with advanced machine learning techniques,” J. Innovation and Digital Health, vol. 1, no. 2, pp. 42–51, 2024.

[5] M. S. Patil et al., “CARDIO PREDICT: Harnessing machine learning for advanced heart disease risk assessment,” IJERT, vol. 11, no. 4, pp. 28–32, 2024.

[6] M. Ahmed and I. Husien, “Heart disease prediction using hybrid machine learning: A brief review,” J. Robotics and Control, vol. 5, no. 3, pp. 884–892, 2024.

[7] G. Logabiraman et al., “Heart disease prediction using machine learning algorithms,” MATEC Web of Conferences, vol. 392, 2024.

[8] M. Yusuf and I. O. Hajara, "A review of hybrid intelligent system for diagnosis and prediction of heart disease," J. Agricultural and Food Chemical Engineering, vol. 4, no. 2, pp. 1–8, 2024.

[9] T. Chaporkar et al., "Effective heart disease prediction using hybrid machine learning technique," IJNRD, vol. 9, no. 4, 2024.

[10] P. Rufes et al., "Heart disease prediction using machine learning," IRJAEH, vol. 2, no. 3, pp. 485–490, 2024.

[11] Clustering of Bigdata Using Genetic Algorithm in Hadoop MapReduce Chandra Shekhar Gautam1 and Mr. L N Soni 2 Dr.Prabhat Pandey3 European Chemical Bulletin Volume 11, Year 2023 link: 3cd6f9626a0e7346666a83b2864990a1.pdf (eurchembull.com).

[12] An improving query optimization process in Hadoop MapReduce using ACO-Genetic algorithm and HDFS map reduce Technique Chandra Shekhar Gautam1 and Dr.Prabhat Pandey2 International Journal of Current Engineering and Technology, Volume 13, Year 2023 DOI: https://doi.org/10.14741/ijcet/v.13.2.8

[13] A REVIEW ON GENETIC ALGORITHM MODELS FOR HADOOP MAPREDUCE IN BIG DATA Chandra Shekhar Gautam1 and Dr.Prabhat Pandey2 International Journal of Recent Scientific Research, Volume 13, Year 2022, Pages 771-775 http://dx.doi.org/10.24327/ijrsr.2022.1303.0166

[14] A REVIEW OF BIG DATA ENVIRONMENT, TOOLS AND CHALLENGES Chandra Shekhar Gautam1 and Dr.Prabhat Pandey2 Journal of Emerging Technologies and Innovative Research, Volume 6, Year 2019, Pages 569-575 http://doi.one/10.1729/Journal.22507

[15] Speedup Query Processing in Hadoop Using Mapreduce Framework Chandra Shekhar Gautam1 and Dr.A khilesh A Waoo Advance Technology and Science, Volume 2, Year 2018, Pages 43-53 DOI: 10.31058/j.data.2018.21004.

[16] Optimizing Heart Disease Prediction Accuracy using Machine Learning Models Sanjana Chaudhari, Mr Chandra Shekhar Gautam, Akhilesh A Waoo , International Journal of All Research Education and Scientific Methods (IJARESM) Volume 6, Year 2024, Pages 1091-1098.

[17] Enhancing Heart Disease Prediction Accuracy: A Comparative Study of Machine Learning Models with Ensemble Method Sanjana Chaudhari, Mr Chandra Shekhar Gautam, Akhilesh A Waoo JARIIE, Volume 10, Year 2024, Pages 4827-4833.

[18] R. Shrivastva, S. Mewad, and P. Sharma, "An approach to give first rank for website and webpage through SEO," *International Journal of Computer Sciences and Engineering (IJCSE)*, vol. 2, no. 6, pp. —, Jun. 2014, E-ISSN: 2347-2693.

[19] R. SHRIVASTVA, C. S. GAUTAM, AND S. K. KAR, "PROMOTING A WEBSITE WITH THE HELP OF SEO USING PPC (PAY PER CLICK)," *SHODHKOSH: JOURNAL OF VISUAL AND PERFORMING ARTS*, VOL. 5, NO. 5, PP. 133–140, MAY 2024, ISSN (ONLINE): 2582-7472, DOI: 10.29121/SHODHKOSH.V5.I5.2024.361.