



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: V Month of publication: May 2023

DOI: <https://doi.org/10.22214/ijraset.2023.51828>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Machine Learning for Personalized Risk Assessment and Early Detection of Heart Disease

Reshma Anilkumar¹, P. Vineetha Sankar²

Abstract: *Today in our busy life it very difficult for us to look after our health. Cardio vascular diseases are very common now days which cause loss of billions of lives world-wide. Our lifestyles have a major impact on our health causing various chronic diseases. Machine learning (ML) can revolutionize the field of cardiovascular disease prediction by providing more accurate, nuanced, and personalized risk assessment, leading to improved patient health. A properly trained machine learning model can easily detect or predict the heart disease. Within the context of using artificial intelligence for forecasting and assessing the occurrence of heart ailments, here are various factors that contribute to the heart disease to be considered, such as lifestyle choices and medical conditions to predict whether the person is at high risk or not. ML technology can also analyze large datasets of patient data to identify patterns and risk factors that cannot be easily distinguished, predicted, or detected using traditional methods. Estimators like k-NN classifier, Decision tree classifier, Gradient boosting classifier and Gaussian Naive bayes (NB) classifier, are provided with different characteristics or feature extracted from the dataset, demonstrated consistent and reliable performance in predicting heart disease. The utilization of these attributes can potentially aid the medical field for on time detection and diagnosis of the heart disease.*

Keywords: *Machine learning, Heart ailments, Classifiers, k-NN*

I. INTRODUCTION

Heart disease is a serious health problem that affects millions of people worldwide. Heart is an important organ in our body, it controls blood flow and provides smooth functioning of other organs. Any kind of disruption in its functioning can affect the body severely. Heart diseases are very fatal. Lifestyle factors such as immobile lifestyle, unhealthy diet, nicotine consumption, and alcohol dependency, as well as underlying diseases such as high blood pressure, diabetes, and obesity, are some of the main reasons for the development of heart disease.

Only the timely identification and accurate diagnosis of the heart disease can provide effective treatment and a complete cure of this condition.

The main challenges confronting the medical industry are the quality of service and lack of availability of technology. The quality of service referred to the aspects, such as reliability, responsiveness, accessibility, effectiveness, and efficiency. It becomes Perilous in the absence of trained personnel and technology to provide accurate diagnosis.

Machine learning is a much-awaited innovation in the medical industry, as it enables the quick analysis of large volumes of complex data, leading to improved efficiency, early accurate diagnosis, personalized treatment, and the discovery of new drugs.

The introduction and use of machine learning (ML) model has gained acceptance in medical industry as they offer promising capabilities for the early detection, diagnosis, and prediction of heart diseases.

In this research paper, we explore the use of various datamining methods such as data preprocessing, feature selection, classifications, and ML algorithms such as Gaussian Naive Bayes (NB), Gradient Boosting, k-Nearest Neighbor (KNN) and Decision Tree to detect heart disease. We also propose a hybrid approach combining gradient boosting and Gaussian naive Bayes to improve the accuracy of the detection system.

This hybrid model combines the feature of both algorithms, i.e., Gradient Boosting's ability to handle complex relationships between features and Naive Bayes' ability to handle missing values and noisy data. According to the computed result, proposed hybrid model overtakes the individual models with regards to predictive accuracy and can be used as an effective tool for early detection and diagnosis of heart disease.

A. Problem Statement

Heart disease is ranked among the primary factor of death worldwide. ML based prediction system holds great promise in predicting risks. Although many popular ML models are available to effectively predict heart disease, the accuracy of these models in prediction is comparatively low.

So, the problem of this study focuses to address this by building an accurate ml-based heart disease prediction system that can accurately predict heart disease risk in people. The accuracy and effectiveness of the system should be evaluated and compared with the existing model to determine practical usability.

B. Research Questions

- 1) How can we incorporate patient-specific factors, such as medical history, into machine learning models for heart disease prediction?
- 2) Which features are the most important for the study, and how can feature selection techniques be used to improve model performance?
- 3) What all new contributions can machine learning make in the field of heart disease diagnosis and prevention?
- 4) Which machine learning algorithm(s) is/are most effective for heart disease prediction?
- 5) Can we improve the accuracy of prediction by using a hybrid model that combines multiple machine learning algorithms?

II. LITERATURE REVIEW

Apurb Rajdhan et al,[1] proposed a study relies to develop a reliable system that could predict heart disease. In the near future, a web application utilizing Random Forest algorithm has been recommended for advanced developments, and an expanded dataset may potentially enhance the predictive accuracy of prediction which could help health professionals to predict heart disease more effectively and efficiently.

Pooja Anbuselvan. [2] Aimed to identify the most accurate algorithm for predicting the development of heart disease by analyzing various classification techniques. The study used pre-processed data and techniques like Logistic Regression, Naïve Bayes, Support Vector Machine, K-Nearest Neighbor, Decision Tree, Random Forest and XGBoost. The accuracy of each algorithm was compared, with Random Forest and XGBoost being the most efficient and K-Nearest Neighbor performing the worst. They also recommended incorporating other data mining techniques such as time series, clustering, and association rules to improve accuracy.

Megha Kamboj.[3] Built a model to predicting heart disease using supervised machine learning. The author has used six machine learning algorithms, including K-Nearest Neighbors (KNN), Random Forest, Support Vector Machine (SVM), Decision Tree, Naïve Bayes, and Logistic Regression. They have also processed the data appropriately before implementing these algorithms for prediction.

Rishabh Magar et al.,[4] Proposed a web-based machine learning application that predicts the risk of heart disease for a user based on their medical details. The application uses a UCI dataset and four algorithms, including Support Vector Machine, Decision Tree, Naïve Bayes, and Logistic Regression. The accuracy of each algorithm is displayed as a percentage, and the prediction result is binary (Yes/No) in nature. The application checks the format of user data and displays an error message if it is not in the required format.

Sai Bhavan Gubbala.[5] Built a model where Random Forest has the highest accuracy among all the classifiers at 85.22%. The models were trained and tested using Python, and the accuracy was measured. It is concluded that Random Forest is the best classifier, even though other classifiers like Logistic Regression, AdaBoost Classifier, Decision Tree, and Support Vector Machine also performed well with accuracy levels greater than 50%. The random forest approach has the potential to achieve even higher accuracy than other methods.

III. DATA COLLECTION AND PREPROCESSING TECHNIQUES

A. Data Source

An organized dataset named "heart_csv" has been selected, which contain information about the patients, their medical conditions and about the likelihood of having heart disease.

The dataset is small contain 303 entries and 14 columns. The objective of this study is to predict whether a patient has heart disease or not.

Using Dataset sourced from Kaggle repository. dataset consist of 14 columns representing various relevant medical attributes associated with heart disease.

These features include age, sex, types of Chest pain, resting BP, serum cholesterol, fasting BP, resting ECG results, heart rate achieved, exercise-induced angina, ST depression, slope of the peak exercise ST segment, number of major vessels, thallium heart scan results, and the target, a binary variable indicating presence or absence of heart disease in patients.

Sl.no	Feature code	Name of feature	Mean
1	age	Age	54.366
2	sex	Sex 0 = female. 1 = male	0.683
3	cp	Types of chest pain 1 = typical angina. 2 = atypical angina. 3 = non-anginal pain. 4 = asymptotic	0.966
4	trtbps	Resting blood pressure	131.623
5	chol	Cholesterol	246.264
	fbs	Fasting blood sugar	0.148
6	restecg	Resting electro cardio graph 0 = normal. 1 = having ST-T wave abnormality. 2 = left ventricular hypertrophy.	0.528
7	thalachh	Maximum heart rate achieved	149.646
8	exng	Exercise induced angina 0 = no. 1 = yes	0.326
9	oldpeak	ST depression induced by exercise relative to rest	1.039
10	slp	Slope 0-upsloping 1-flat 2-down slopping	1.399
11	caa	No of major vessel colored by fluoroscope (0-3)	0.729
12	thall	Defect type 1 = normal. 2 = fixed defect. 3 = reversible defect	2.313
13	output	Target or class column 0 = absence. 1 = present	0.544

Figure 1.Dataset

B. Preprocessing Techniques

The dataset does not contain any null values. Even there are no null values present there are many preprocessing steps to be performed to improve the quality and enhance the accuracy of the model.

Two preprocessing approaches were used, includes the feature selection and check for outliers and skewness in the model. Variable(feature) selection help to choose the most informative and most likely to contribute features from the dataset. Here the Pearson correlation method is utilized to determine the strength of relationship between features and to calculate the correlation coefficient. The feature with lowest correlation can be removed from data as it is not important for analysis i.e., it shows less correlation with the target variable. This helps to reduce the dimensionality of the dataset by removing redundant features or highly correlated features and improve performance and efficiency of the model and contribute to exploratory analysis to understand relationship between attributes, identifying the trends and patterns and proceed for further analysis.

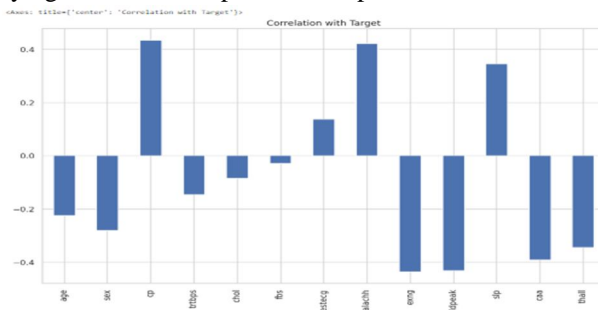


Figure 2.Correlation with target attribute

It is equally important to check the correlation between the attributes. It helps to find the relationships among attributes and also help to remove highest correlated attributes as they cause multicollinearity and leads to overfitting and poor performance of the model.

Outliers are the datapoint deviate significantly from rest of the data points and skewness refers to asymmetry in distribution of data. the performance of the model fully relies on the quality and relevance of the information that fully captured in the data. Therefore, it is necessary to plot and identify the outliers and skewness in data.

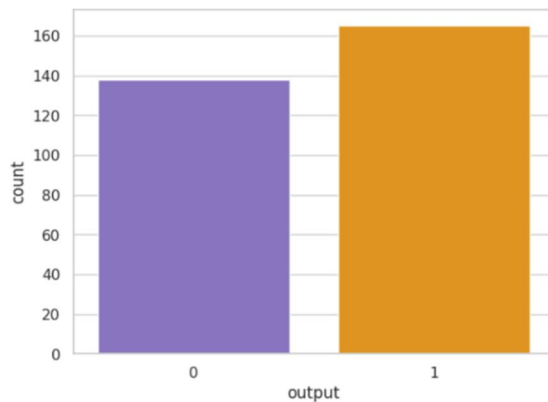


Figure 3. count plot of target attribute

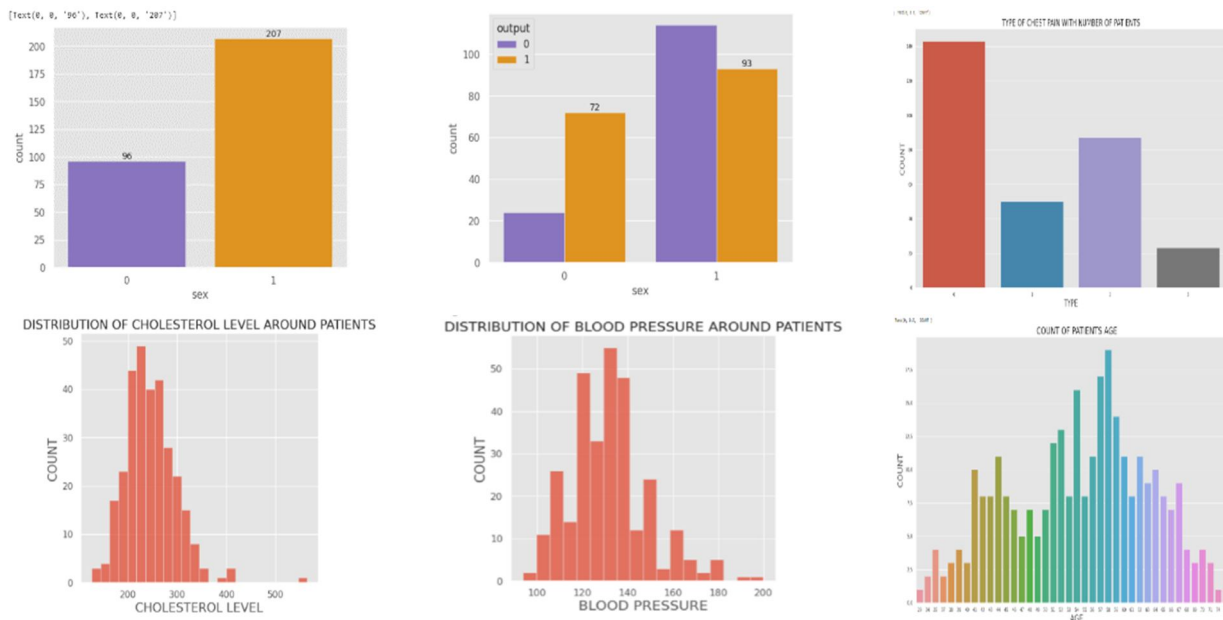


Figure 4. Count plot of various attributes

C. Feature Engineering

Feature engineering is an indispensable part in machine learning project, which refers to transforming the raw data into features for subsequent analysis of machine learning model. Here the feature engineering is performed on the basis of categorical and continuous variable and two methods used are “One-hot encoding” and “Scaling”. One-hot encoding is important for models that use numerical computations as many models cannot work directly on the categorical variables.

In this study the dataset column is separated as continuous and categorical variables based on the unique value in each attribute column. Each categorical variable is one-hot encoded. One-hot encoding creates binary features for each unique value of categorical variable, excluding the output variable since it is the target. And the continuous variables are scaled, which involves transforming the variables to a common scale, otherwise leading to biased results.

These techniques are essential for dataset preprocessing and preparation as they improve the accuracy and performance of the model. When used together they can improve the performance by having a comparable impact on the model by preventing some features from dominating others.

IV. METHODOLOGY

A. Machine Learning Algorithms

A machine learning algorithm typically describes the development, implementation, and evaluation of a specific algorithm or technique used to solve a particular problem in a given domain. Machine learning (ML) algorithms are a set of mathematical models and statistical techniques that allow computer to learn from a dataset without being explicitly programmed. They enable computers to recognize patterns, trends and relationships in large and complex data sets and make predictions based on those discovered patterns. Supervised learning algorithms learn from labeled data where each data point is tagged with the correct output or label. These algorithms are used to address both classification and regression task. The algorithms used in this study includes, decision trees, gradient boosting, and k-nearest neighbors (KNN), Gaussian naive bayes and hybrid of Gradient Boosting and Gaussian Naive bayes.

1) Gradient Boosting

Gradient boosting is a powerful algorithm based on supervised learning that can be used to solve both the classification and regression task. It is an ensemble method combine multiple weak models to create strong and efficient predictive model. Gradient boosting has the ability to handle both the continuous and categorical values.

The idea behind gradient boosting is to iteratively add weak models to base models and each of which focus on residues of previous models. This process is repeated until model achieves required accuracy.

The key advantage of using gradient boosting algorithm is its ability to handle complex data and missing values, making it ideal for Realtime data. It is less susceptible to overfitting i.e., having lower tendency to fit noisy data on comparing with other algorithms, which can result in better generalization performance.

2) Gaussian Naive Bayse Classifier

Gaussian Naive bayes (GNB) is a classification algorithm based on Bayes' theory of conditional probability. It is referred as "naive" because it makes assumption that the features in a dataset are independent of each other. The algorithm models the distribution of each feature in each class using a Gaussian distribution, assuming that each class is normally distributed. The model then uses Bayes' theorem to estimate the fraction of distribution of each class based on the characteristics of a new data point. GNB is a fast and efficient algorithm that performs well on small and medium-sized datasets with high dimensional feature spaces. It is mainly used for computer vision problem application.

3) K-Nearest Neighbour Classifier

Is a simple and effective algorithm for regression and classification task. It is a non-parametric algorithm, meaning that it does not make any assumptions about the underlying distribution of the data. Instead, it uses the proximity of nearby points to make predictions.it work by finding k nearest point by using Euclidean distance or Manhattan distance measure. Then value is predicted based on the not common class or average value of K nearest neighbor. K is an important parameter need to be considered when making predictions using this algorithm.

4) Decision Tree Classifier

Decision tree is a supervised machine learning algorithm.it can be used to solve both classification and regression challenges. This classifier model is tree like structure where the target variable is already known. Decision tree consist of different types of nodes with functionalities includes, root node, intermediate nodes, and leaf nodes. Leaf node labels specify which class does it belong. It works by splitting the dataset into small subset based on significant feature along each node. The goal is to develop a tree like structure that best predict the target variable by selecting the most appropriate feature that effectively divide the dataset.

5) Hybrid Model Of Gaussian Naive Bayes And Gradient Boosting

Hybrid model combine two or more classifiers to output the best result. In the Hybrid model of Gaussian naive bayes and Gradient boosting it combines the strength of both the models to improve the potential and performance of the model.

In this the GNB is used to preprocess and select the feature and which then fed as input to Gradient Boosting algorithm. GNB will generate ensemble of decision trees that are trained to classify the data. The prediction of each combined to generate the final output.

B. Feature Selection Method

Feature selection is a method of selecting significant variables from a large set of features to improve the performance and efficiency of a machine learning model and prevent overfitting. Various feature selection methods can be used depending on the complexity and specifications. Basic dataset and specific research question investigated. This study used two feature selection methods namely Correlation-based Feature Selection (CFS) and Recursive Feature Elimination (RFE). These methods can help identify the most important features that best fit the target variable, while avoiding overfitting and reducing the computational burden.

Correlation-based Feature Selection (CFS): It is a filter-based feature selection method that select features based on its relationship with the output variable. CFS evaluates the predictive power of a feature by measuring the correlation between the features and the features and target variable. The idea behind CFS is to choose features that are highly correlated with the target attribute and less correlated with each other, to avoid overfitting. In this paper the correlation is evaluated using a Heatmap.

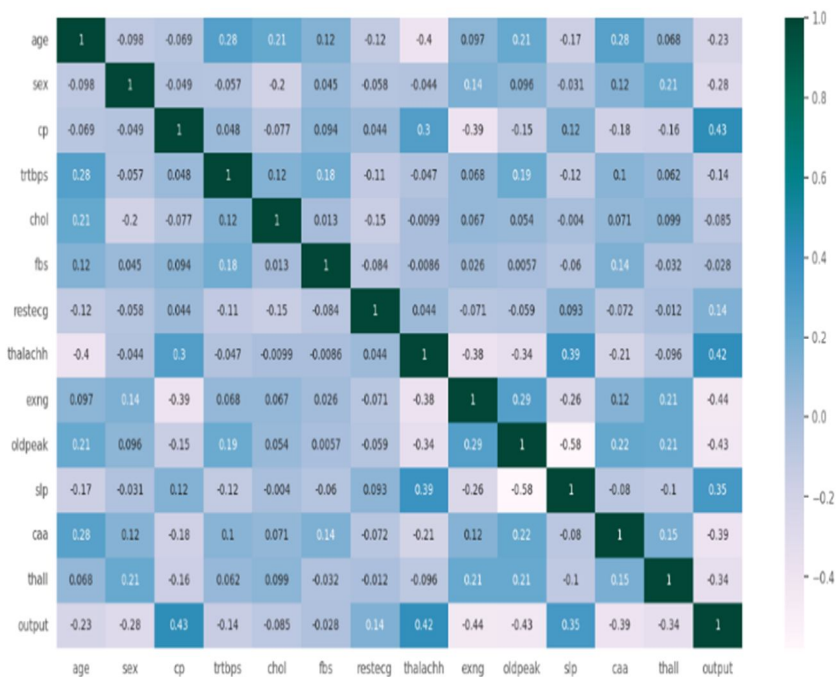


Figure 5. Correlation heatmap

Recursive Feature Elimination (RFE) is a wrapper-based feature selection method that selects features by iteratively removing the least significant features based on coefficients of an estimator. In wrapper-based feature selection, a subset of the data is selected, and a model is trained on this subset. The performance of the model is then evaluated, and the same is repeated for different subsets. At each iteration, a model is trained on the reduced set of features, and the feature with the lowest importance score is removed. It works by fitting an estimator to the data, ranking the features based on their importance, and iteratively eliminating the least important ones until the desired number of features is left.

C. Model Architecture

The proposed project creates an estimator to predict the heart disease at the earliest by using the mentioned algorithms. In the proposed model, Dataset containing medical history of the patients are collected. Which then undergo a series of preprocessing techniques such as feature selection, feature engineering and correlation analysis etc. The preprocessed data will be partitioned for training and testing, here in the project we have taken 20% for testing and 80% for training. The selected subset for training is fed to the classifier algorithms. Once the model is trained it can be tested using the test subset for validating the accuracy of the model.

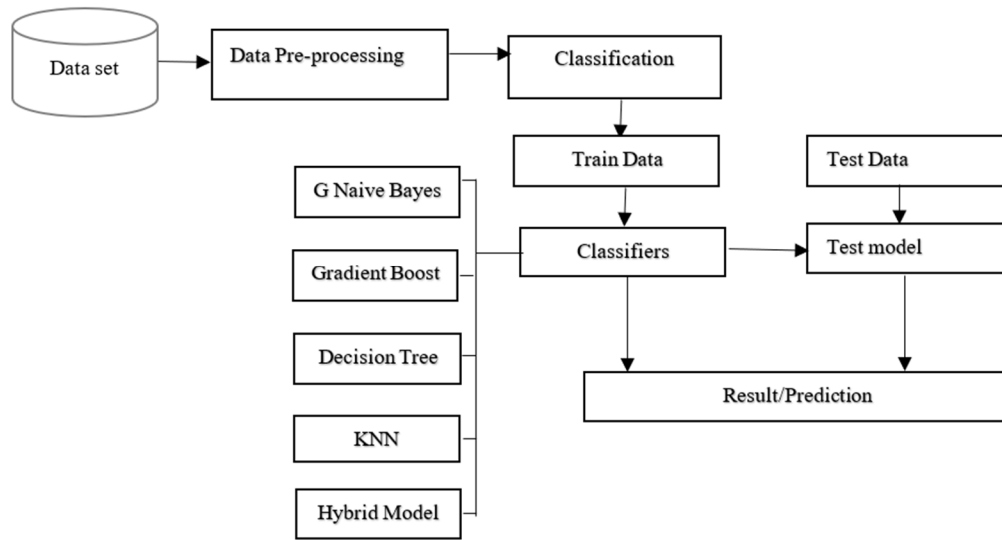


Figure 6. Proposed Model

D. Performance Matrix

Performance matrix are used in machine learning algorithm to measure accuracy and effectiveness of the model. They help to compare various machine learning models and evaluate the performance. The paper used performance matrix such as;

1) Accuracy

It is the common performance matrix used by ML Algorithms to measure the performance of the model. It refers to the ratio of correct prediction made from all the predictions. It accurately predicts the output if the data is balanced.

Defined by;

$$\text{Accuracy} = (\text{No of correct prediction}) / (\text{total no of predictions}).$$

In the paper We have used five classifiers to measure the accuracy such as decision trees, gradient boosting, and k-nearest neighbors (KNN), Gaussian naive byes and hybrid of Gradient Boosting and Gaussian Naive bayes.

Table 1. Accuracy of proposed classifiers

	Accuracy
Gaussian NB	0.846
Gradient Boosting	0.802
K-Neighbor Classifier	0.835
Decision Tree Classifier	.725
Hybrid model	.885

2) Confusion matrix

A confusion matrix is a tabular form often used to evaluate the performance of a classifiers in machine learning. It is a matrix with four entries, representing the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) produced by a classifier.

Table 2. Format of confusion matrix

Actual/predicted	positive	negative
positive	TP	FP
negative	FN	TN

TP represents the number of positive examples that were correctly classified as positive by the model. FP represents the number of negative examples that were incorrectly classified as positive. FN represents the number of positive examples that were incorrectly classified as negative. TN represents the number of negative examples that were correctly classified as negative.

In this study the confusion matrix of decision trees, gradient boosting, k-Nearest Neighbors (k-NN), Gaussian Naive bayes and hybrid of Gradient boosting and Gaussian Naive bayes are created.

Confusion Matrix for Heart Attack Detection using DecisionTree

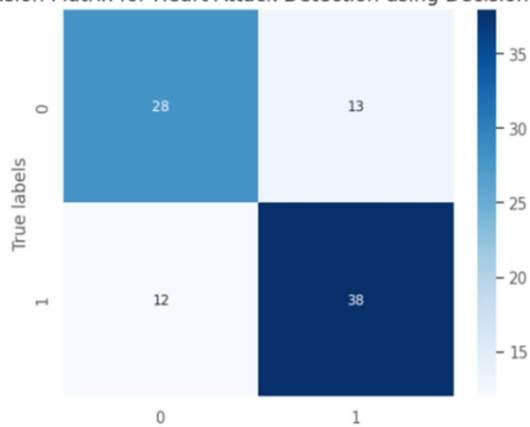


Figure 7. Performance analysis for Decision tree

Confusion Matrix for Heart Attack Detection using GradientBoosting

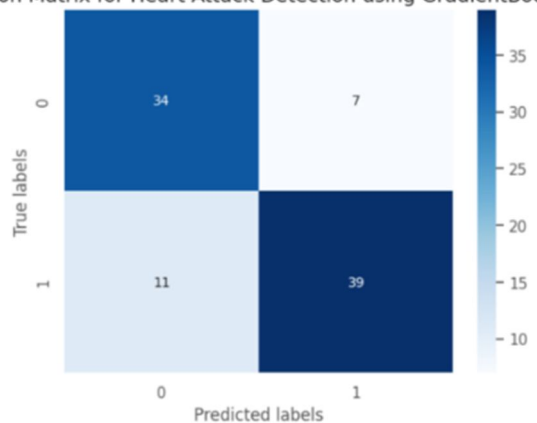


Figure 8. Performance analysis for Gradient boosting

Confusion Matrix for Heart Attack Detection using GaussianNB

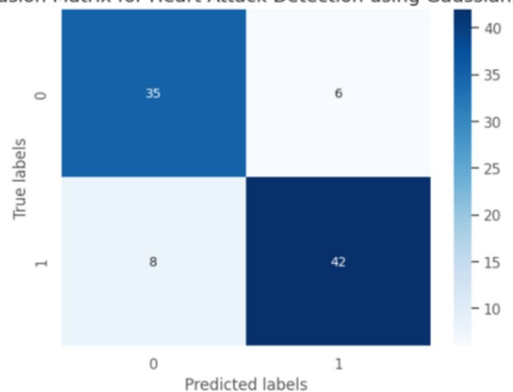


Figure 9. Performance analysis for Gaussian Naive bayes

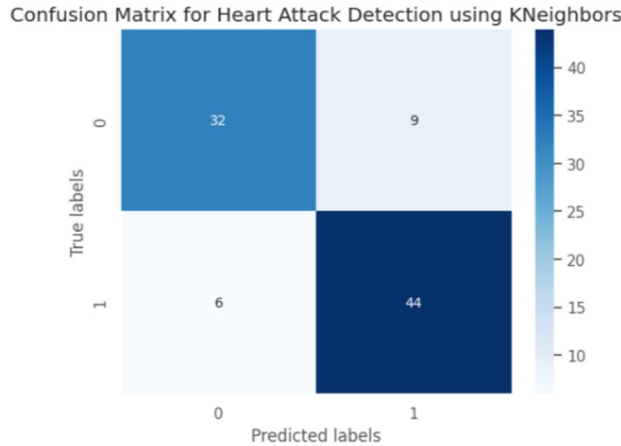


Figure 10. Performance analysis for k- Nearest Neighbors

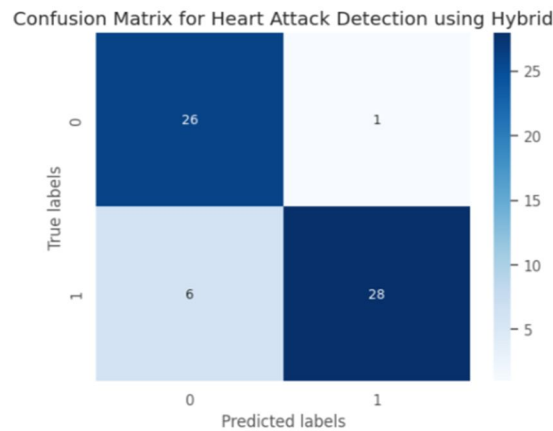


Figure 11. Performance analysis for Hybrid Model

3) Precision

In machine learning, precision is a performance measure of a classification model. It is calculated by dividing number of correctly predicted positive instances by the total number instances predicted positive by the model

Defined by,

$$\text{Precision} = \frac{TP}{(TP + FP)}$$

Table 3. Precision value of proposed classifiers

	Precision
Gaussian NB	0.8536
Gradient Boosting	0.8292
K-Neighbor Classifier	0.780
Decision Tree Classifier	0.6829
Hybrid model	0.965

4) Recall

In machine learning, recall is a performance measure of the completeness of a classification model. It is the number of TP predictions divided by the total number of actual positive examples in the dataset. In other words, recall measures the proportion of correctly predicted positive examples out of all positive examples in the dataset.

Defined by,

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Table 4. Recall value of proposed classifiers

	Recall
Gaussian NB	0.82
Gradient Boosting	0.75
K-Neighbor Classifier	0.84
Decision Tree Classifier	0.7
Hybrid model	0.82

V. RESULT

Based on the results obtained from the model, it appears that the Machine learning algorithms applied for predicting heart disease have varying levels of accuracy.

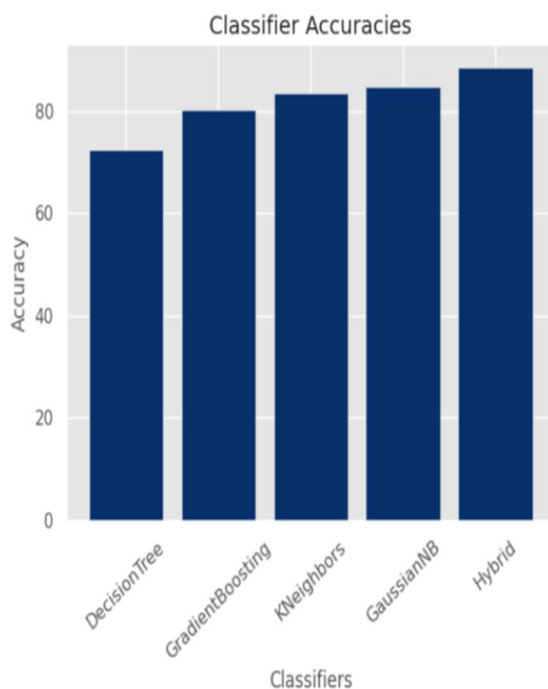


Figure 12. Bar chart visualization of proposed algorithm accuracies

The k-Nearest Neighbor (k-NN) algorithm achieved an accuracy of 83.51%, indicating that it can be able to correctly predict the presence of heart ailments. The Gaussian Naive Bayes algorithm performed slightly better with an accuracy of 84.61%. This indicates that the algorithm was able to accurately classify patients as having heart disease or not.

The Gradient Boosting algorithm achieved an accuracy of 80.21%, indicating that it was less accurate than the KNN and Gaussian algorithms. The Decision Tree algorithm had the lowest accuracy with only 72.5% accuracy.

Finally, with an aim to improve the accuracy a Hybrid algorithm by combining two of the best performed classifiers were created and achieved the highest accuracy of 88.5%. It combined both the strong features of Gaussian Naive Bayes and Gradient Boosting to achieve high accuracy.

Table 5.Comparative analysis of previous works

Year	Authors	Techniques Used & accuracy	Key findings of the proposed model
2018	Megha Kamboj [3]	K-Nearest Neighbor -87% Naive bayes - 80%. Decision tree -79%.	Gradient boosting - 80.21% Gaussian naive bayes - 84.61% Decision tree - 72.5% k-Nearest Neighbor - 83.51 Hybrid model - 88.5%
2020	Pooja Anbuselvan [2]	Naive bayes -77.5% Decision tree - 77.5% k- Nearest Neighbor -57.8%.	
2020	Apurb Rajdhan et al, [1]	Naive bayes-85.25% Decision tree-81.97%.	
2020	Rishabh Magar et al,[4]	Naive Bayes-80.43% Decision Tree -80.43%	
2022	Sai Bhavan Gubbala [5]	Decision tree -74.34%.	

Using only accuracy to compare the model is not a precise way. In this paper we have used a novel approach to implement a hybrid model which can outperform the individual models to provide a greater performance. Various feature selection measures and feature engineering is employed to identify the most informative features likely to contribute in the model and also used One-hot encoding to treat categorical value as continuous value to avoid the bias became a significant advantage in this proposed model.

Overall, these results indicate combining multiple algorithms to create hybrid model can outperform and lead to even greater accuracy. It also indicated that Gaussian Naive bayes classifier can perform efficiently to predict the heart disease among various machine learning methods.

VI. FUTURE SCOPE

There are many potential areas for further advancement and future improvements in this project. Includes;

- 1) *Algorithm selection*: There are many Machine Learning and Deep learning algorithms that could be applicable to this dataset such as Random Forest, SVM and Neural Networks. It would be worth comparing these algorithms with the proposed model.
- 2) *Hyperparameter tuning*: Performance of the model can be optimized by tuning hyper parameters of the classifiers. This includes testing different pairs of hyperparameters to best choose the optimal one.
- 3) *Real time prediction*: The current model was trained on existing data. But in future there could be requirements for real time predictions. Predicting a model that can predict the presence of heart disease in Realtime helps in early detection and diagnosis of the disease.
- 4) *Integrating External Data*: Integrating the data from devices like wearable devices, IoT devices and electronic medical records can improve the accuracy of the model.

Continued research and development in these areas could lead to more accurate and effective models for predicting the presence of heart disease.

VII. CONCLUSION

In this paper four classifiers were used to calculate the accuracy of the heart disease prediction system. Gaussian naive bayes outperformed all the three. On an attempt to increase the accuracy and performance two of the classifiers were combined to create a hybrid model achieved more accuracy. It shows that the model can perform better with a hybrid model by combining two strong classifiers. Promising results were achieved.

REFERENCES

[1] Rajdhan, A. et al. (2020) 'Heart Disease Prediction using Machine Learning', International Journal of Engineering Research & Technology (IJERT), 9(4). doi:01-05-2020.

[2] Anbuselvan, P. (2020a) 'Heart Disease Prediction using Machine Learning Techniques', International Journal of Engineering Research & Technology (IJERT), 9(11). doi:05-12-2020.



- [3] Kamboj, M. (2020) 'Heart Disease Prediction with Machine Learning Approaches', International Journal of Science and Research (IJSR), 9(7).
- [4] Magar, R., Memane, R. and Raut, S. (2020) 'HEART DISEASE PREDICTION USING MACHINE LEARNING', Journal of Emerging Technologies and Innovative Research, 7(6).
- [5] Gubbala, S.B. (2022) 'Heart Disease Prediction Using Machine Learning Techniques', International Research Journal of Engineering and Technology (IRJET), 9(10).
- [6] Iyer, S. et al. (2020) 'HEART DISEASE PREDICTION USING MACHINE LEARNING', International Research Journal of Modernization in Engineering Technology and Science, 2(7).
- [7] Julker Nayeem, Rana, S. and Rabiul Islam (2022) Prediction of Heart Disease Using Machine Learning Algorithms, View of prediction of heart disease using machine learning algorithms. Available at: <https://www.ej-ai.org/index.php/ejai/article/view/13/13> (Accessed: 03 May 2023).
- [8] Adil Hussain Seh, Pawan Chaurasia and Mustafa Shuaieb Sabri (2019) 'A Review on Heart Disease Prediction Using Machine Learning Techniques', International journal of Management, IT and Engineering, 9(4).
- [9] Rindhe, B. U., Ahire, N., Patil, R., Gagare, S., & Darade, M. (2021). Heart Disease Prediction Using Machine Learning. *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, 5(1). <https://doi.org/10.48175/IJARSCT-1131>
- [10] Shah, D., Patel, S. & Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. SN COMPUT. SCI. 1, 345 (2020). <https://doi.org/10.1007/s42979-020-00365-y>
- [11] Krittanawong, C., Virk, H.U.H., Bangalore, S. et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. Sci Rep 10, 16057 (2020). <https://doi.org/10.1038/s41598-020-72685-1>
- [12] C. Boukhatem, H. Y. Youssef and A. B. Nassif, "Heart Disease Prediction Using Machine Learning," 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 2022, pp. 1-6, doi: 10.1109/ASET53988.2022.9734880.
- [13] Mahmud, Tanjim & Barua, Anik & Begum, Manoara & Chakma, Eipshita & Das, Sudhakar & Sharmen, Nahed. (2023). An Improved Framework for Reliable Cardiovascular Disease Prediction Using Hybrid Ensemble Learning. 1-6. 10.1109/ECCE57851.2023.10101564.
- [14] Mane, V., Tobre, Y., Bonde, S., Patil, A., Sakhare, P. (2023). Heart Disease Prediction Using Machine Learning and Neural Networks. In: Shukla, P.K., Singh, K.P., Tripathi, A.K., Engelbrecht, A. (eds) Computer Vision and Robotics. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-19-7892-0_17
- [15] K. Joshi, G. A. Reddy, S. Kumar, H. Anandaram, A. Gupta and H. Gupta, "Analysis of Heart Disease Prediction using Various Machine Learning Techniques: A Review Study," 2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT), Dehradun, India, 2023, pp. 105-109, doi: 10.1109/DICCT56244.2023.10110139.
- [16] Jackins, V., Vimal, S., Kaliappan, M. et al. AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. J Supercomput 77, 5198–5219 (2021). <https://doi.org/10.1007/s11227-020-03481-x>



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)