HEART DISEASE PREDICTION AND RECOMMENDATION USING HYBRID MACHINE LEARNING MODEL

[1] Sethu Madhavan K, [2] Santhosh Kumar D, [3] Shailesh E, [4] Vidhya A

[1,2,3] SRMVEC Student, [4] Assistant Professor,

[1,2,3,4] Department of Computer Science and Engineering

[1,2,3,4] SRM Valliammai Engineering College, Chennai, Tamil Nadu, India

Abstract- Heart disease (HD) is a serious health problem that has affected many people all over the world. Shortness of breath, muscular weakness, and swollen feet are prominent signs of HD. Due to some factors, including accuracy and execution time, present heart disease diagnosis techniques are not very effective in early-time identification. Researchers are working to develop an effective method for the detection of heart disease. In this study, we suggested a supervised machine learning-based system that can quickly and accurately diagnose cardiac problems. The categorization techniques used in the system's development, including support vector machines, decision trees, and random forests, are all included. To increase accuracy, we combined all of the above-mentioned algorithms to create a hybrid algorithm. The classifier's performances are evaluated using the performance measurement metrics. On the features chosen via features selection algorithms, the classifier performances have been evaluated. We produce an enhanced performance level with an accuracy of 95.08% through the prediction model for heart disease with Hybrid Decision Tree, Random Forest, and SVM model. The flask framework website provides the final output. The suggested system can also recommend a food diet to patients who test positive.

Keywords – Heart Disease, Supervised machine learning algorithms, Feature Selection, Support Vector Machine, Random Forest, Decision Tree, Hybrid model, Flask Framework.

I. INTRODUCTION

The diagnosis and treatment of heart disease are extremely difficult when modern technology and medical experts are not available. A number of contributing risk factors, including diabetes, high blood pressure, excessive cholesterol, an irregular pulse rate, and many other factors make it challenging to diagnose heart disease. The severity of cardiac disease in humans has been determined using various data mining and neural network techniques. Many techniques, including the K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic Algorithm (GA), and Naive Bayes, are used to categorise the severity of the condition (NB). As cardiac illness has a complex character, it requires cautious management. Failure to do so could harm the heart or result in premature death. To identify different types of metabolic syndromes, data mining and the perspective of medical research are employed. Heart disease prediction and data analysis both greatly benefit from data mining with classification. Decision trees have also been used to accurately forecast events linked to heart disease in the past. There are numerous cardiovascular ailments, including arrhythmia, angina, heart failure, and cardiomyopathy. People of all ages are susceptible to heart disease, which is particularly prevalent in middle and advanced age. Men suffer from heart disease more frequently than women do. 30 percent of deaths in underdeveloped nations are thought to be attributable to heart disease, according to WHO estimates. Globally, heart disease is to blame for one-third of fatalities. In the US and other developed nations, heart disease is the leading cause of mortality. Worldwide, cardiovascular disease (CVD) claims the lives of almost 17 million individuals each year. Effective diagnosis and proper treatment can save the lives of many people. According to the European Society of Cardiology, 26 million approximately people with HD were diagnosed, and 3.6 million people per year. Most people in the United States

are suffering from heart disease. Diagnosis of HD is traditionally done by the analysis of the medical history of the patient, the physical examination report, and the analysis of concerned symptoms by a physician. But the results obtained from this diagnosis method are not accurate in identifying the patient of HD. Moreover, it is expensive and computationally difficult to analyze. Thus, to develop a non-invasive diagnosis system based on classifiers of machine learning to resolve these issues. Expert decision system based on machine learning classifiers and the application of artificial fuzzy logic effectively diagnosis the HD as a result, the ratio of death decrease.

II. LITERATURE REVIEW

The paper proposed by Senthil Kumar Mohan, Chandrasegar Thirumalai and Gautam Srivastava suggest a new approach to improve the precision of cardiovascular disease prediction by identifying key features using machine learning techniques. The prediction model is introduced together with several feature combinations and well-known classification methods. By using a Hybrid Random Forest with Linear Model as our heart disease prediction model, we achieve an improved performance level with an accuracy level of 88.7%. (HRFLM) [1]. The model proposed by Md. Touhidul Islam, Sanjida Reza Rafa, Md. Golam Kibria suggest a strategy for predicting the risk of heart disease that incorporates principal component analysis (PCA), k-means clustering, and a hybrid genetic algorithm. The proposed method incorporates characteristics including age, sex, cholesterol levels, blood pressure, and other medical parameters to predict the risk of heart disease. Accuracy, sensitivity, and specificity are just a few of the performance indicators that are used to assess the suggested approach. The findings demonstrate that the suggested approach outperforms other current machine learning models shows the risk of heart disease with high accuracy. The authors come to the conclusion that their method can be helpful in early heart disease prediction, which can help in heart disease prevention and management. Overall, the paper provides a valuable contribution to the field of machine learning for healthcare applications with an accuracy of 94.06% [2]. The model proposed by Dr. M. Kavitha, G. Gnaneswar, R. Dinesh, Y. Rohith Sai, R. Sai Sura presents a hybrid model for predicting the risk of heart disease. The goal of this model is to propose a unique machine learning approach. The Cleveland heart disease dataset was employed in the proposed study, and data mining techniques including regression and classification were applied. approaches for machine learning Decision Tree and Random Forest are used. The machine learning model's innovative technique is created. Three machine learning methods are employed in the implementation: Random Forest, Decision Tree, and Hybrid Model (Hybrid of random forest and decision tree). According to experimental findings, the hybrid model's heart disease prediction accuracy rate is 88.7% [3]. This paper suggested by Ke Yuan, Longwei Yang, Yabing Huang, Zheng Li presents a hybrid machine learning model which increase the precision of machine learning in predicting cardiac disease. The authors propose a new approach called hybrid gradient boosting decision tree with logistic regression (HGBDTLR) based on ensemble learning. The actual results show that the Cleveland heart disease dataset HGBDTLR algorithm's prediction accuracy can reach 91.8%. [4]. The following paper is suggested by Sanchayita Dhar, Krishna Roy, Tanusree Dey, Pritha Datta, Ankur Biswas aims to propose a useful method for foretelling cardiac problems using machine learning techniques. Hence, utilizing a Random Forest classifier and a straightforward k-means algorithm, we suggested a hybrid strategy for heart prediction. The J48 tree classifier and Naive Bayes classifier are two further machine learning techniques that are used to evaluate the dataset, and the results are contrasted. Results from the Random Forest classifier and the corresponding confusion matrix demonstrate the methodology's robustness [5].The model proposed by A. Pandiaraj, S.Lakshmana Prakash, P. Rajesh Kanna will manage, control, and foresee risk variables, data mining software analyses enormous amounts of real-time medical data. Using the use of a support vector machine (SVM) and a genetic algorithm (GA), this research also attempts to diagnose cardiac disease. The higher performance of the suggested technique in foretelling heart disease is shown by experimental data. [6]. In this paper we improved the accuracy of the predicting model and decrease the level of misdiagnosis. The proposed model will help the patients to understand whether they get heart disease or not. If the result shows that they are positive it will recommend a food diet and gives notification to the doctor associated with the model.

III. PROPOSED MODEL

To create the heart disease prediction model, we have used three widely used ML approaches. The details of the algorithms are as follows:

Decision Tree

Decision Tree method is used to create Classification models. The structure of a tree is the foundation of this categorization approach. This falls within the supervised learning category because the desired outcome is already known. The decision tree approach can be applied to both categorical and numerical data. The root node, branches, and leaf nodes make up a decision tree. The traversal path from the root to a leaf node is used to evaluate the data. A total of 283 tuples from our dataset were evaluated down the decision tree. They could have reached a favorable or unfavorable conclusion about the prognosis of heart disease. To verify for false positives or false negatives, these were compared to the actual parameters. This demonstrates the accuracy, specificity, and sensitivity of the model.One of the learning models used to solve the categorization problem is the decision tree. Using this method, we split the dataset into two or more sets. Internal nodes of a decision tree indicate a test of the features, a branch represents the result, and leaves reflect the decisions that are produced after further processing. The following is the decision tree algorithm:

i. Make the dataset's best feature the tree's root.

ii. The data set is divided into train and test sets. Subsets should be created so that each subset includes data with the aforementioned feature characteristic.

iii. The previous procedures are repeated on each subset until the tree has leaves. The anticipated class label record in the root of the decision tree. The values are contrasted with the root characteristics and the subsequent record attributes. In this comparison, the matching value of the following node departs.

Random Forest

To arrive at a single decision, Random Forest combines several decisions. Random sampling is carried out for training characteristics, followed by random sub-characteristics for sampling nodes. Separate the train from the test set in the dataset. Subsets should be created so that each subset includes a feature property similar to that. The previous procedures are repeated on each subset until the tree has leaves. Since the tree-building samples are generated via bootstrapping, the same feature may be taken into account more than once. There may be a numerical limit to the total number of node-splitting features. The fitting issue is simplified by this algorithm. A group of unpruned classification-based trees makes up Random Forest. Because of its insensitivity to dataset noise and extremely low risk of overfitting, it exhibits astounding performance in several real-life challenges. It operates more quickly than many other tree-based algorithms and typically increases accuracy for testing and validation data. Individual decision tree algorithms' predictions are combined to form random forests. When building a random tree, there are several options for adjusting the performance of a random forest.

Support Vector Machine

Support Vector Machine is a machine learning classification technique that is used to examine data and find patterns in regression analysis and classification. When data is described as a two-class problem, SVM is often considered. Finding the appropriate hyperplane that isolates every data point from one class to the other is how this technique characterizes data. The better the model is taken into account, the greater the separation or edge between the two classes. Support vectors refer to the data points that are located at the margin's edge. SVM is built on mathematical techniques for creating challenging real-world issues. Heart Disease Dataset we used for this project, includes multiple classes based on a variety of factors, and results. The hardest part of using SVM to build a model is choosing the kernel and strategy to avoid overfitting and underfitting problems. As a result, the final SVM model needs to be evaluated against real data. The hardest part of using SVM to build a model is choosing the kernel and strategy to avoid overfitting and underfitting problems.
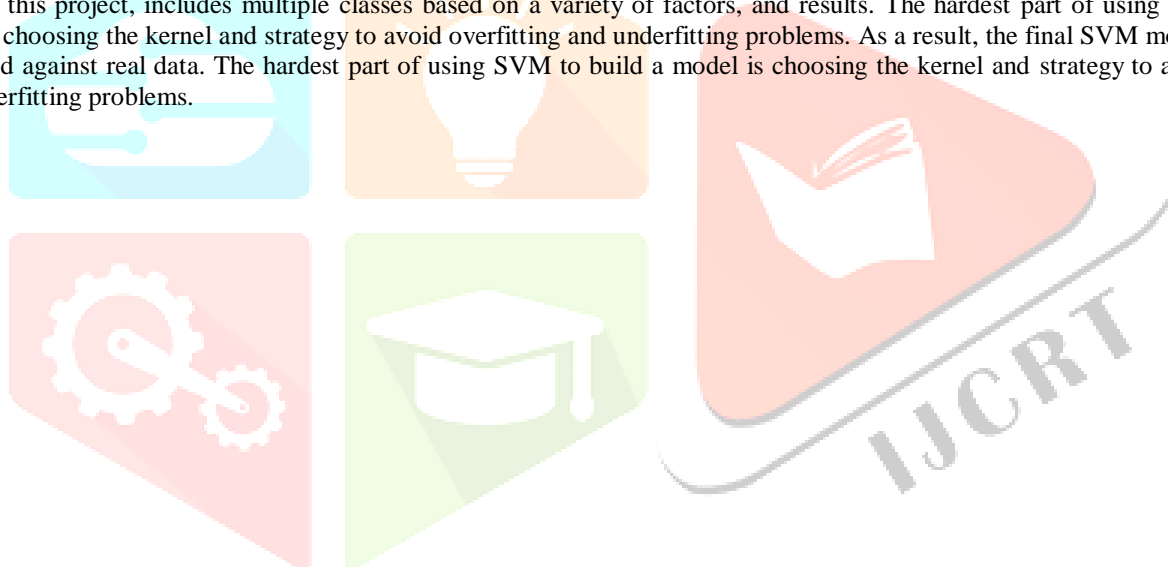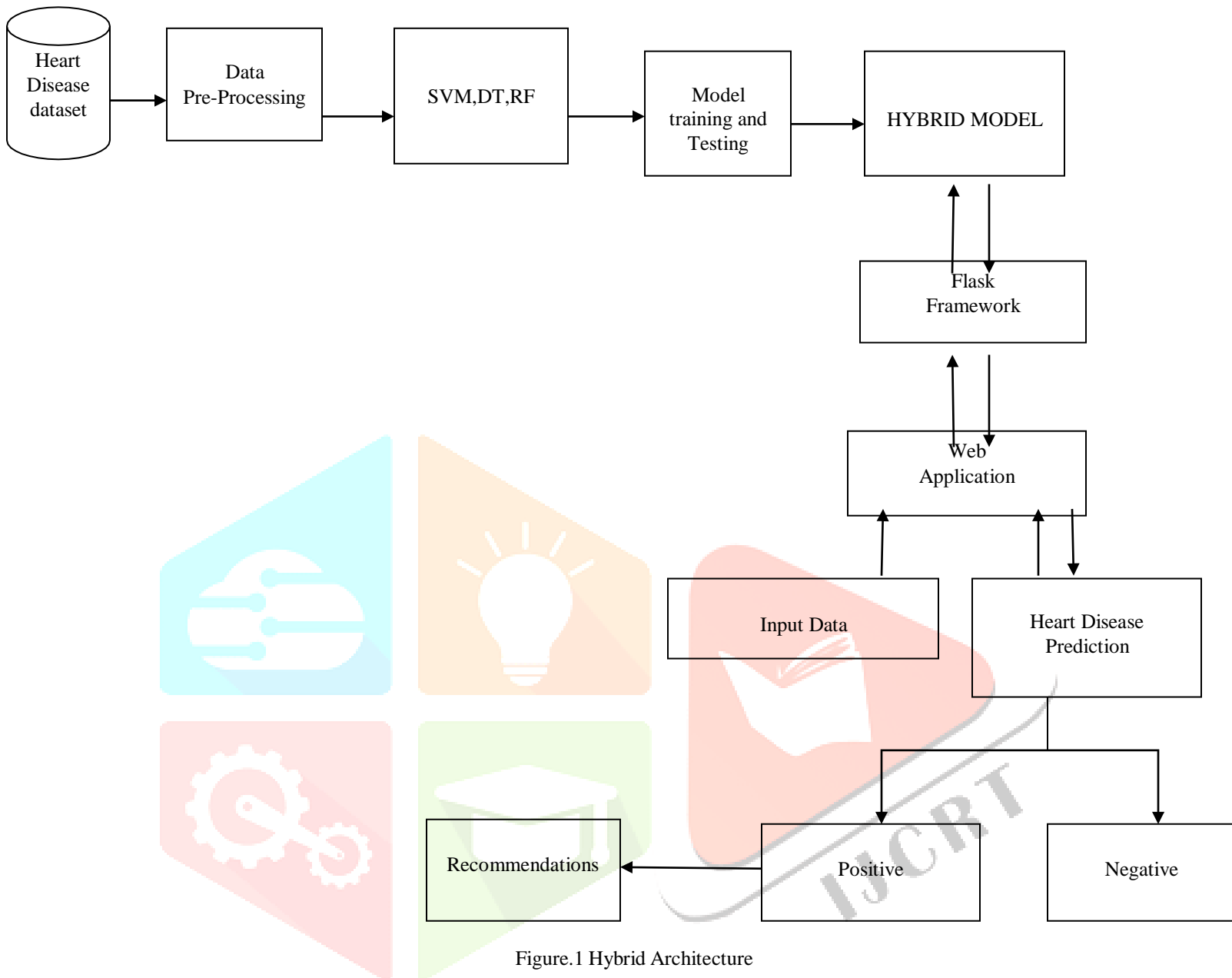
Hybrid Model

Figure.1 Hybrid Architecture

Figure 1 depicts the operation of the hybrid model in detail. The heart disease dataset, which was collected from the UCI Machine Learning repository, is used by this model. Data pre-processing happens following the data collection process. Once the pre-processing of the data is complete, we have a dataset that is entirely consistent, noise-free, and complete. The model training follows that. SVM, Decision Tree, and Random Forest are the three algorithms that make up the hybrid model. The dataset is split in half, with 80% of the data used for training and the remaining 20% for testing, in a ratio of 80:20. The Flask Framework acts as an interface for passing user input to the model, which then makes predictions and returns the results to the intermediary. The model performs the prediction and provides the output to the framework, which then displays the framework's output on the screen, using the Flask Framework as an intermediary. If the model predicts the output to be negative, there is no need to be concerned; however, if it predicts the output to be positive, it provides food diets, a doctor's guidance, and exercise recommendations that must be followed in order to recover.

IV. METHODOLOGY

Data Collection

Data collection is a critical step in machine learning-based heart disease prediction. The success of the prediction model depends heavily on the quality and quantity of data used for training and validation. To ensure the model is trained on data that is relevant to the problem at hand, the target population must be defined. This includes defining the age range, gender, and other relevant demographics of the population being studied. Relevant data sources for heart disease prediction may include electronic health records (EHRs), medical images, laboratory test results, genetic data, and lifestyle factors.

Table 1: Attributes

S.I No.	Attribute	Description	Measurement
1	Age	Age	In years
2	Sex	Patient's gender	Male, Female
3	cp	Chest pressure	Typical angina, Atypical angina, Non-angina, Asymptomatic
4	trstbps	Resting Blood Pressure in mm hg	mm Hg
5	chol	Serum Cholesterol in mg/dl	mg/dl
6	fbs	fasting blood sugar- 1 if >120 mg/dl, 0 if <120 mg/dl	mg/dl
7	restecg	Electrocardiographic Results	Normal, ST-T wave abnormal, Hypertrophy ,
8	thalach	Maximum Heart Rate observed	Numeric
9	exang	Exercise with angina has occurred	Binary
10	oldpeak	depression inducedthrough exercise	Numeric
11	slope	slope of the ST segment	Flat, Upsloping, Downsloping
12	thal	Heart Status	Normal, Fixed Defect, Reversible Defect
13	ca	Number of major vessels ranging from 0 - 3 color byfluoroscopy	Colored by fluoroscopy(0-3)
14	Target	Output	0=Absence,1=Presence

Data Pre-processing

Data preprocessing is an essential step in machine learning-based heart disease prediction. The first step in data preprocessing is to clean the data by removing any errors, missing values, or outliers. Missing values can be imputed using techniques such as mean imputation or regression imputation. Outliers can be identified and removed using statistical methods or visual inspection. After cleaning the data, the next step is to select the most relevant features for heart disease prediction. This can be done using statistical methods such as correlation analysis or feature importance ranking algorithms. Machine learning algorithms often perform better when the input features are on a similar scale. Therefore, feature scaling is often performed by rescaling the features to have zero mean and unit variance or by normalizing the features to a specified range.

Table 2: Sample Testing Data

52	1	0	129	202	0	1		168	0	1	2	2	3	0
50	1	0	137	203	1	0		155	1	3.1	0	0	3	0
77	1	0	149	164	0	1		125	1	2.6	0	0	3	0
49	1	1	125	259	0	0		144	0	0.8	2	0	3	0
58	1	0	128	246	0	0		116	1	3.2	1	2	2	0
73	0	0	116	139	0	1		125	0	1.6	1	0	2	1
44	0	1	130	341	1	0		136	1	3	1	0	3	0
34	1	1	114	210	0	1		192	0	0.7	2	0	2	1
36	1	0	120	249	0	0		144	0	0.8	2	0	3	0
56	1	1	132	286	0	0		116	1	3.2	1	2	2	0
66	0	0	122	149	0	1		125	0	1.6	1	0	2	1

Feature Extraction & Selection

Feature selection is an important step in building a heart disease model, as it helps to identify the most informative features that can accurately predict the presence of heart disease. The feature selection process involves selecting a subset of the most relevant features from the original dataset, based on their importance in predicting the target variable (heart disease). There are different methods for feature selection, including filter methods, wrapper methods, and embedded methods. In filter methods, features are selected based on statistical measures such as correlation, chi-square test, and mutual information. These methods are fast and efficient, but they do not take into account the interaction between features. Wrapper methods, on the other hand, choose features by assessing how well the model performs given various subsets of characteristics. Although they require expensive calculation, they can record feature interaction. The feature selection process is incorporated into the model-building process using embedded approaches. For instance, feature importance measures are used by decision tree algorithms like Random Forest and Gradient Boosting to choose the most pertinent features during training. The most significant risk factors for heart disease, such as age, gender, blood pressure, cholesterol levels, and smoking behaviors, can be found using feature selection in the context of heart disease. The performance and accuracy of the heart disease model can be improved by choosing the most useful features. Feature selection also helps to reduce the complexity of the model, which can improve its interpretability and reduce the risk of over fitting.

Model Training & Testing

Separating the data into two sets, one for training and the other for testing. The testing set is used to assess the model's performance while the training set is used to train the machine learning algorithm. Choosing an appropriate machine learning method for the problem of heart disease prediction. Logistic Regression, Decision Tree, Random Forest, Support Vector Machines (SVM), and Artificial Neural Networks are a few of the frequently utilized methods (ANN). creating a predictive model by applying the chosen machine learning algorithm to the training dataset. Modifying the machine learning algorithm's hyperparameters to enhance performance. In this step, several hyper parameter combinations are tested, and the model's performance on the validation set is assessed.

Evaluating the performance of the model on the testing dataset using evaluation metrics such as accuracy, precision, recall, and F1-score. Deploying the final model to make predictions on new data. During the testing phase, the model is used to predict heart disease based on the features in the testing dataset. The predicted outcomes are compared with the actual outcomes in the testing dataset to evaluate the performance of the model. The accuracy of the model is evaluated based on how well it predicts the actual outcomes. If the accuracy of the model is satisfactory, it can be deployed for practical use. If the accuracy is not satisfactory, the model may need to be retrained or hyperparameters may need to be adjusted.

Flask Framework

Flask is a well-known microweb framework. It is employed in the creation of web services, RESTful APIs, and other web applications. With its straightforward and flexible interface, Flask makes it easy to build web apps. It is intended to be compact and modular, enabling developers to pick the elements they require for their particular use case. Here are some of Flask's main characteristics: Routing: Flask offers a technique for converting URLs into Python functions. Templating: To render dynamic HTML pages, Flask employs the Jinja2 templating engine. Request management: Flask offers functionalities for simple request management. Data storage and retrieval are made possible via Flask's session management system, which it offers. Extensions: There are numerous extensions available for Flask. Request management: Flask offers functionalities for simple request management.

V. EXPERIMENTAL ENVIRONMENT

Datasets

Data on heart disease were gathered from the machine learning repository at UCI. There are four databases: the VA Long Beach, Switzerland, Hungary, and Cleveland. Because it is a popular resource for ML researchers and has extensive and full records, the Cleveland database was chosen for this study. There are 303 records in the collection. Although there are 76 attributes in the Cleveland dataset, only 14 of them are covered by the data set that is available in the repository. The Cleveland Clinic Foundation is the dataset's data source. The description and type of properties are shown in Table 1. 13 traits are used to predict heart disease, yet only one of those 13 attributes actually results in the presence of heart disease in a person. The Cleveland dataset has an element called num that displays the severity of patients' diagnoses of heart disease on a range of scales, from 0 to 4. In this instance, 0 denotes absence. People who have heart disease are represented by the values from 1 to 4, with the scale referring to the severity of the ailment (4 being the highest). The distribution of the num attribute over the 303 records found .

Graphical Representation & Performance Evaluation

Conveniently, the model analyses and presents data on the confusion matrix, true positive, precision, recall, falsenegative, etc. It is a portable, GUI-based software that is open source and equipped with a wealth of cutting-edge machine learning methods, including deep learning algorithms for image processing, etc.

Confusion Matrix

To assess how well a machine learning classifier is performing, a confusion matrix is employed. The actual and expected classes are represented, accordingly, in the matrix's rows and columns. A model's predicted labels are compared to the actual labels of the data using the confusion matrix. Confusion matrix has four cells:
True Positive (TP): Total number of correctly labeled positive predictive cases that are truly positive. **True Negative (TN):** Total number of correctly labeled negative predictive cases that are truly negative. **False Positive (FP):** Total number of truly negative cases that are labeled incorrectly as positive cases. **False Negative (FN):** Total number of truly positive cases that are labeled incorrectly as negative cases. **True Positive Rate:**

True Positive Rate is the proportion of observations that are correctly predicted to be positive.

$$TPR = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

False Positive Rate:

False Positive Rate is the proportion of observations that are incorrectly predicted to be positive.

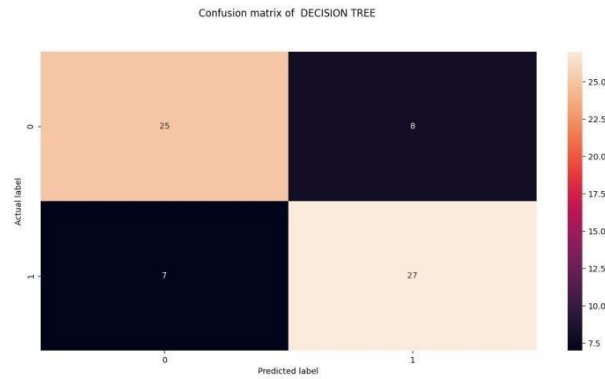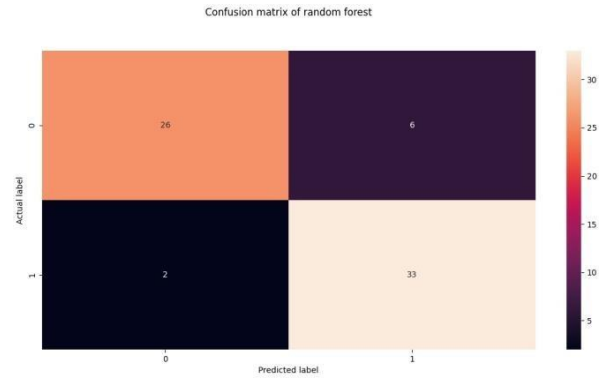$$FPR = \frac{False\ Positive}{True\ Positive + False\ Positive}$$
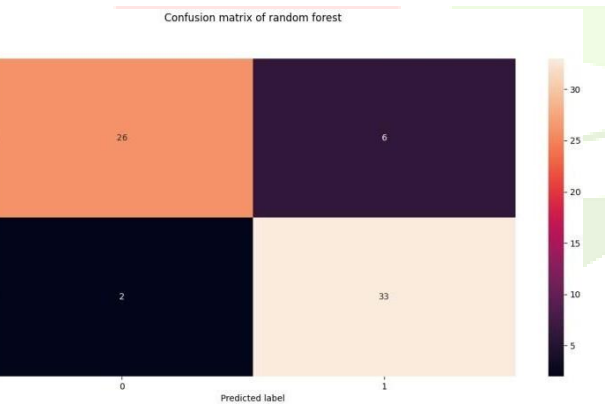

Figure 2: Confusion Matrix of DT
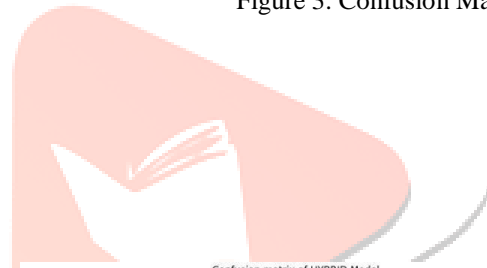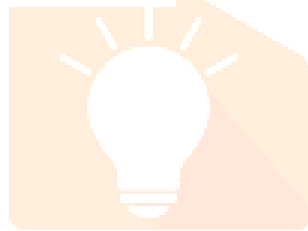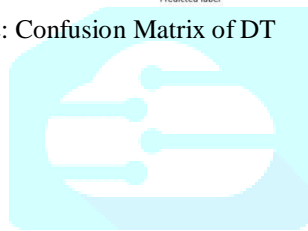

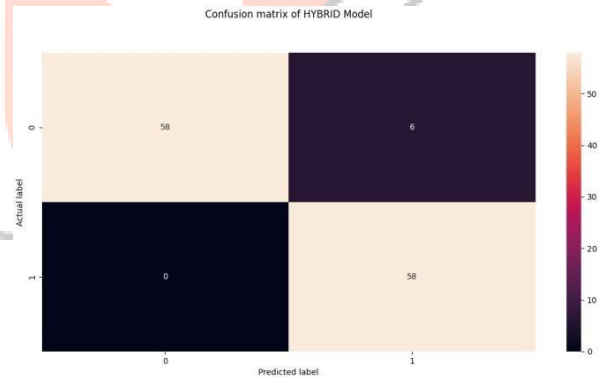Figure 3: Confusion Matrix of RF

Figure 4: Confusion Matrix of SVM

Figure 5: Confusion Matrix of HM

Precision: Shows the percentage of all anticipated positive samples that were successfully predicted.

$$Precision = \frac{TP}{TP + FP}$$

Recall: Shows the percentage of samples that were successfully categorised out of all samples.

Recall =   TP TP+FN
_____

Accuracy: Shows the percentage of all actual positive samples that were correctly predicted to be positive.

$$Accuracy = \frac{TP+FP}{TP + FP + TN + FN}$$

F1 score: F1 score is a metric used to evaluate the performance of a binary classification model. It is the harmonicmean of precision and recall.

$$F1 = \frac{2 * (precision * recall)}{(precision + recall)}$$

Table 3: Performance Measure Of Models

DECISION TREE	RANDOM FOREST	PORT VECTOR MACHINE	HYBRID MODEL
Accuracy : 77.61% **Precision** : 77.14% **Recall** : 79.41% **Confusion Matrix** :[[25 8] [7 27]]	**Accuracy** : 88.06% **Precision** : 94.29% **Recall** : 84.62% ion Matrix :[[26 6] [2 33]]	**Accuracy** : 91.04% **Precision** : 88.57% **Recall** : 93.94% sion Matrix :[[30 2] [4 31]]	**Accuracy:** 0.9508196721311475 **Precision :** 0.90625 **Recall :** 1.0 **F1 :** 0.9508196721311475 **Confusion Matrix :** [[58 6] [0 58]]

Density Plot

   A density plot is a graphic representation of a continuous variable's probability density function. It offers a smooth estimate of the density function and is used to show the distribution of a variable. A smoothed variant of a histogram, the density plot is particularly helpful when the underlying data distribution is not well represented by a normal distribution. The probability density function of the data is first estimated using a kernel density estimation method before the density plot is produced. Each data point in the kernel density estimation process receives a kernel function, which is then added together to form a smooth density function. The density plot's shape depends onthe bandwidth parameter and kernel function selection. The form of the distribution, the central tendency, and the distribution of the data are all displayed in the density plot that is produced. The density of the data at a given position is represented by the height of the curve there. The overall probability of the variable is represented by the area under the curve, which is equal to 1. To compare the distributions of various groupings, density graphs are helpful.
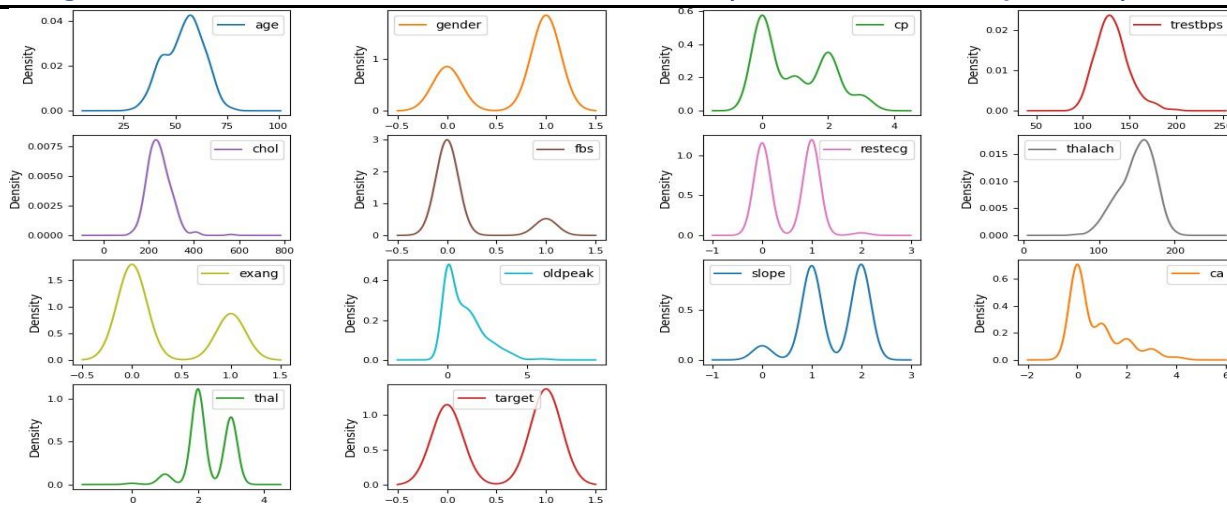
Figure 6: Density Plot

Correlation matrix

A correlation matrix or correlation graph may also be referred to as a correlation map. Each cell in the graph represents the correlation coefficient between two variables, and the entire dataset is a representation of the correlation between variables. A visual summary of the connections between variables in a dataset is provided by the correlation map. It is especially helpful when there are numerous variables to investigate in huge datasets. Patterns and links can be immediately found by looking at the map, which can help with additional analysis and modelling.
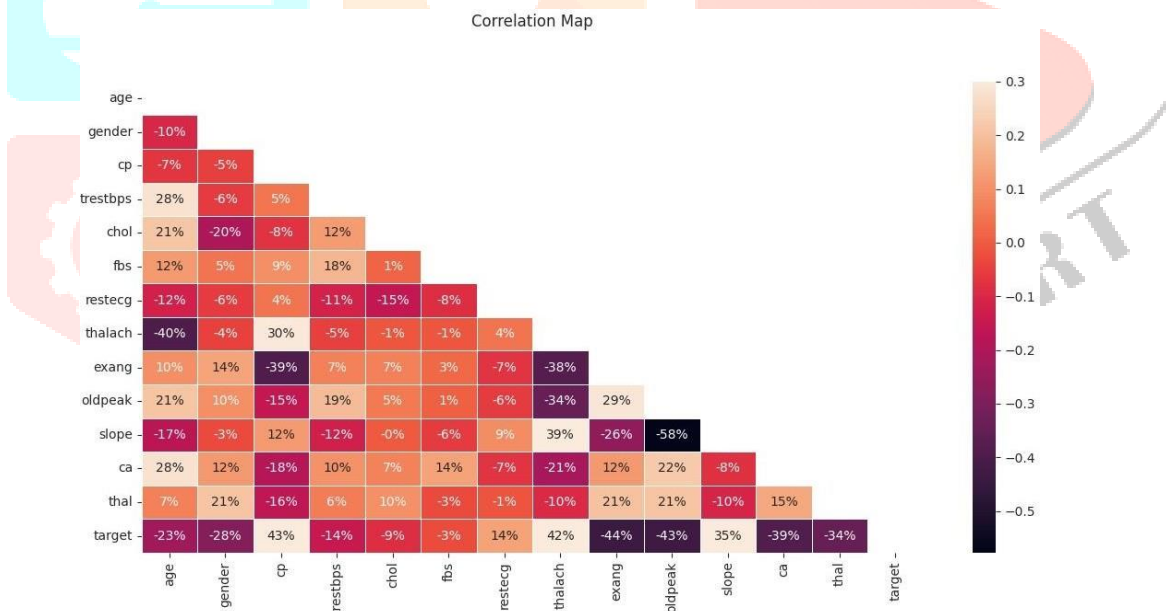

Figure 7: Correlation Matrix

VI. DISCUSSSION

Heart disease prediction using a hybrid model is an approach that combines multiple algorithms and techniques to improve the accuracy of predicting heart disease. We aim to improve the accuracy of the predicting model and decrease the level of misdiagnosis. The proposed model will help the patients to understand whether they get heart disease or not. If the result shows that they are positive it will recommend a food diet and notify the doctor who is associated with the system. To build a hybrid model various supervised machine learning algorithms are discussed and based on the literature survey the following three algorithms are finalized: Decision Tree, Random Forest, and Support Vector Machine. These mentioned algorithms showed better results and gave us a hybrid model which gives an accuracy of 95.08%.

VII. CONCLUSION

The goal of the project is to minimize misdiagnosis and to increase accuracy . This project focuses on a machine learning model for predicting heart disease. After data collection, we perform data pre-processing and analysis, model training, and evaluation. The dataset is visualized using a density map and a correlation map. The density map provides a visual representation of the distribution of each feature in the dataset, and the correlation map provides information about the correlation between the different features. In preprocessing we perform one-hot encoding on the categorical features implemented hybrid algorithm and flask framework prediction.

REFERENCES

[1] Senthil Kumar Mohan, Chandrasegar Thirumalai and Gautam Srivastava, "Effective Heart Disease Prediction using Hybrid Machine Learning Techniques", IEEE Access,vol.7, no.10, pp.30-50,2019.

[2] Md. Touhidul Islam, Sanjida Reza Rafa, Md. Golam Kibria, "Early Prediction of Heart Disease Using PCA and Hybrid Genetic Algorithm with k-Means", 23rd International Conference on Computer and Information Technology (ICCIT), pp:1615- 1666, 2020.

[3] Dr. M. Kavitha, G. Gnaneswar, R. Dinesh, Y. Rohith Sai, R. Sai Sura, "Heart Disease Prediction using Hybrid machine Learning Model", 6th International Conference on Inventive Computation Technologies (ICICT), pp:44-66, 2021,.

[4] Ke Yuan, Longwei Yang, Yabing Huang, Zheng Li," Heart Disease Prediction Algorithm Based on EnsembleLearning", 7th International Conference on Dependable Systems and Their Applications (DSA), pp: 30-55, 2020.

[5] Dhar, Krishna Roy, Tanusree Dey, Pritha Datta, Ankur Biswas, " A Hybrid Machine Learning Approach for Prediction of Heart Diseases ", 4th International Conference on Computing Communication and Automation (ICCCA), pp:53-75, 2018.

[6] A. Pandiaraj, S.Lakshmana Prakash, P. Rajesh Kanna, " Effective Heart Disease Prediction Using Hybridmachine Learning ", Third International Conference on Intelligent Communication Technologies andVirtual Mobile Networks (ICICV), pp. 93-113, 2021.

[7] Md. Istiaq Habib Khan, M. Rubaiyat Hossain Mondal, "Effectiveness of Data Driven Diagnosis of HeartDisease", 11th International Conference on Electrical and Computer Engineering (ICECE), 2020.

[8] Mahaveer, Puneet, Deepika, " Cardiovascular Disease Prediction Analysis using Classification Techniques",IEEE Delhi Section Conference (DELCON) , pp. 36-45, 2022.

[9] U B Mahadevaswamy, Keerthana R, Pooja B B, Sangatya V, Supritha S, "A Hybrid Model approach for Heart Disease Prediction", IEEE 2nd Mysore Sub Section International Conference (MysuruCon), pp. 976-980, 2022.

[10]Kanika Pahwa, Ravinder Kumar, "Prediction of heart disease using hybrid technique for selecting features", 4thIEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)., vol. 36, no. 4, pp. 1165-1188, 2017.

[11]Shruti Saravanan, Krithika Swaminathan, "Hybrid K-Means and Support Vector Machine to Predict HeartFailure", 2nd International Conference on Smart Electronics and Communication (ICOSEC) pp. 436-444, 2021.

[12]S. I. Ansarullah and P. Kumar, ''A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method,'' Int. J. Recent Technol. Eng., vol. 7, no. 6S, pp. 1009–1015, 2019.

[13]A. U. Haq, J. P. Li, J. Khan, M. H. Memon, S. Nazir, S. Ahmad, G. A. Khan, and A. Ali, ''Intelligent machine learning approach for effective recognition of diabetes in E-healthcare using clinical data,'' Sensors, vol. 20, no. 9, p. 2649, May 2020.

[14]A. U. Haq, J. Li, M. H. Memon, M. H. Memon, J. Khan, and S. M. Marium, ''Heart disease prediction system using model of machine learning and sequential backward selection algorithm for features selection,'' in Proc. IEEE 5th Int. Conf. Converg. Technol. (ICT), pp. 1–4, Mar. 2019.

[15]U. Haq, J. Li, M. H. Memon, J. Khan, and S. U. Din, ''A novel integrated diagnosis method for breast cancer detection,'' J. Intell. Fuzzy Syst., vol. 38, no. 2, pp. 2383–2398, 2020.