Coronary Artery Disease Prediction and Analysis using Machine Learning Techniques

Jasmine Jinitha A [1] , Dr. S. Mangayarkarasi [2]

Research Scholar [1], Assistant Professor [2]

Department of Computer Science[2]

Vels Institute of Science Technologies and Advanced Studies (VISTAS), Chennai

jasminejinitha@gmail.com[1], mangai.p.s@gmail.com[2]

Article Info
Page Number: 1207-1224
Publication Issue:
Vol. 71 No. 3 (2022)

Article History
Article Received: 12 January 2022
Revised: 25 Febuary 2022
Accepted: 20 April 2022
Publication: 09 June 2022

Abstract
Coronary Artery Disease (CAD) contains a huge variety of heart-associated illnesses are one of the leading causes of death globally in recent decades. Cardiovascular diseases account for 31 percent of all fatalities worldwide. The scientific affiliation generates a huge quantity of scientific information associated with cardiovascular disease, which need to be well tested to forecast cardiovascular disease. In current days, Machine learning (ML) has emerge as the number one method for the evolution of predictive models within the health-care industry, and it become determined to check how accurate their prediction scores are based on the data collected. The contemporary dataset from the UCI heart repository database is utilized in this proposal, which employs machine learning approaches. To look at the coronary illness, these strategies use 13 clinical parameters from the patient. As a result, supporting human beings in identifying whether or not or now no longer they are at threat for coronary heart illness is tremendously desirable. Gradient Boosting, Decision Tree, Random Forest, SVM, KNN, and Logistic Regression are some of the Supervised ML classifiers employed in this study to deploy a model for heart disease prediction. A 10-fold cross-validation testing option became used to assess the algorithms performance. Also researcher used tuning of the hyper parameter, number of nearest neighbors, namely k, the instance-based (KNN) classifier. Result indicates that compared to unique ML strategies, Gradient Boosting Classifier and Ada Boost Classifier algorithms gives 86.88%, Random Forest Classifier gives 88.15% and K Neighbors Classifier and SVM producing 90% accuracy in lots much less time for the prediction. This model (KNN) or SVM can be useful to the medical practitioners at their medical institution as Decision Making Support System.

Keywords— Machine Learning, Random Forest, KNN (K Nearest Neighbors), Logistic Regression, Decision Trees, Gradient Boosting, Ada Boost.

I INTRODUCTION

Coronary artery disorder is a sickness that impacts the heart particularly. In this study the necessary factors focused such as origins of coronary heart diseases, the complications faced through patients and medical doctors and the possible solutions like prediction in advance. Various dangerous activites like physical inactivity, unhealthy foods, nicotine use and

immoderate alcohol intake are the most common risk facors and cause for the coronary heart disorder and stroke. Patients may also revel in excessive blood pressure, excessive blood sugar, excessive blood lipids, and obese or weight problems due to behavioral risk factors [1]. The stroke may happen when the circulation of blood to the brain can be cut off by a blood clot in a cerebral artery. When a coronary artery will become clogged or narrowed to the factor in which blood float is stopped or significantly restricted, a coronary heart attack ensues. An artery that feeds blood to the coronary heart muscle is referred to as a coronary artery. When a blood clot blocks blood float to a coronary artery, it may motive blockage [2]. It is critical to consider matters just like the origins of coronary heart contamination, the various demanding situations that sufferers and doctors confront, and finally the available solution. One of its maximum key hazard elements for coronary artery disorder is excessive blood pressure (hypertension). Identifying people who are most at hazard for CVDs and making sure that they adequate remedy can assist to reduce early deaths. The intention of this idea is to apply Machine Learning techniques of strategies to forecast coronary artery disorder. ML strategies assist are expecting coronary contamination higher than different conventional fashions and enhance the capacity threat assessment of cardiovascular contamination for anticipatory care. It is important to note that such factors will be taken into account because the primary goal of this study is to develop a Coronary Artery Disease Prediction model that is powered by Boosting and Random Forest classification models and is specifically aimed at achieving better results in diagnosing heart disease effectively and efficiently.

II. MACHINE LEARNING

Machine Learning is one of the efficient technologies for constructing models the usage of training and testing. It is the subset of Artificial Intelligence (Al) that's one in all wide area of learning wherein machines emulating human abilities. On the other hand machine learning models are trained to analyze from facts and make choices as a result the aggregate of each generation is likewise referred to as Machine Intelligence. As the definition of machine learning, it learns from the natural things such as biological facts consisting of cholesterol, blood pressure, sex, age, etc. and on the idea of these, assessment is finished in terms of accuracy of algorithms such as in this paper have used 8 supervised classifier algorithms that are Logistic Regression, KNN, SVM, Naive Bayes, Decision Tree, Random Forest, Ada Boost and Gradient Boost. This proposed method calculate the accuracy of 8 distinct machine learning algorithms on a data set and decide which one is the best model. Machine learning algorithms takes training data and use it to create more precise models. While training a machine learning set of rules with facts, get a machine learning models. Machine learning algorithms take in training data and use it to create more precise models. When training the machine learning algorithm with data, you get a machine-learning model. When give a model an input after training set, get an output. The predictive algorithms will broaden many prediction models on this research, and the quality version can be chosen. Then feed facts into the predictive version, and also get a prediction primarily based on the facts that trained the version. Supervised learning, unsupervised learning, and Reinforcement learning strategies are the 3 types of machine learning algorithms.

Supervised learning strategies expect the output from enters that has already been tagged. Input capabilities with goal values make up the training facts. The set of rules is given enter facts and actual output to create a forecast version with the capacity to decide real output for future facts.

Unsupervised learning is an iterative procedure predicts the output from the facts with unknown labeled facts. The target value not specified that the patients either have heart disease or not. The learning learns to predict the output for the future data by similarities.

Reinforcement learning is a behavior-based learning approach. The algorithm collects input from the data analysis and directs the user to the best possible result. Because the system is not trained using the sample data set, reinforcement learning differs from other methods of supervised learning. Rather, the system learns through making mistakes.

Machine learning algorithms are being implemented to deal with the massive amount of data in clinical associations. Recent research guides have used modern-day neural network-related methodologies for high- precision evaluation of results. For each parameter of the dataset, the approach compared it to random forest with decision trees [4]. Although decision tree systems are a conventional approach, they produce a massive type of probabilistic results. Random Forest is a method for producing reliable results by combining several decision trees. In addition, awesome tool getting to know strategies collectively with Logistic Regression, K Nearest Neighbor, Support Vector Machine, XG Boost, Ada Boost, and Naive Bayes can be compared.

The execution of coronary artery disease predictions carried out with the below given methodologies and the following steps were carried out.

1. Dataset is collected from UCI Repository
2. Data Visualization is done by Seaborn and pair plot()
3. Splitting dataset into test and train set
4. Apply DT, RF.SVM. KNN, Ada Boost, Gradient Boost and Logistic Regression models
5. Train the model
6. Test the trained model and predict accuracy
7. Get single set of input from user and predict heart disease Using best model.

III METHODOLOGY

3.1 Description of the Dataset

The dataset used for this research work was the Public Health Dataset UCI repository, it includes seventy six attributes, including the target attribute. The target attribute refers to the patient's presence of cardiac disease. In this study experiments the use of chosen attribute subset of 14 of them.

Target is integer-valued zero represents “no disease” and 1 represents the affected person has “disease". The attributes that are used on this studies are defined in Table 1 and used python Jupiter with sci-kit learn packages to perform classification of coronary illness prediction. Seaborn function is used to visualise the data, used to create different ML models used on

this proposal. UCI dataset includes 303 records is loaded and performs out a few pre-processing performed on it. The dataset includes 303 records in which six rows include missing values and are supplanted with the mean estimation of all records with relating to features. The dataset is partitioned into 20% for testing it includes sixty one records and training set contains 242 records.

No	Dataset Description	Range of Values
1.	age(age of patient in years)	29 to 79
2.	sex (Female 0, Male 1)	0,1
3.	cp (Chestpain)	0,1,2,3
4.	trestbps (Resting Blood pressure)	94 to 200 (mm Hg)
5.	chol (Serum Cholestrol)	126 to 564 (mg/dl)
6.	fbs (fasting blood sugar)	larger than 120 mg/dl (1 true). Less than 100 mg/dL (5.6 mmol/L) is normal(0 false
7.	restecg (Resting ElectrocardiographicResults)	No=0,Yes=1
8.	thalach (Maximum heart rate achieved)	71 to 202
9.	exang (Exercise induced Angina)	No=0, Yes=1
10.	oldpeak (ST depression induced by exercise relative to rest)	0 to 6.2
11.	slope (Slope of the peak Exercise ST	0,1,2

	segment)	
12.	ca (Number of major vessels colored by fluoroscopy)	0 to 3
13.	thal (thalassemia) Normal=1, Fixed defect=2, Reversible Defect=3	1,2,3
14.	target (no disease =0 and disease =1)	No=0, Yes=1

Table 1: Description of data set

According to the figure 1, the age characteristic which variety begins off evolved from 29 and Gender is the essential aspect due to the fact greater male person are affected with coronary heart disorder while as compared with female, chest ache is the signal of coronary heart disorder, blood stress performs an essential aspect in any disease, cholesterol additionally performs an essential part, Fasting Blood sugar, Resting and different attributes additionally have correlation with every different which shown in Figure 3.

3.2. Preprocessing of the Dataset

There are not any null values within side the dataset. The skewness of the records, outlier detection, and records distribution had been all checked the usage of diverse plotting techniques. When filing records for type or prediction, all of those preprocessing techniques consisting of characteristic choice performs a vital role[5]. Before schooling the system gaining knowledge of models, a few specific variables need to be transformed to dummy variables, and scaling need to be performed with a popular scalar function. The get dummies approach changed into used to provide dummy columns for class variables, whilst the StandardScalar() approach changed into utilised to cut back the records. In the numerical price columns age, trestbps, chol, thalach, and oldpeak, a few capabilities have greater fluctuation.

3.2.1 Checking the Distribution of Data

The distribution of statistics in a dataset performs an critical function whilst the prediction or category of a trouble in machine learning algorithms. In this view that nearly 140 rows (patients) not have heart disease and more than 160 rows (patients) have heart disease out of 303 records which shows the dataset is balanced .Checking the data set is balanced or not by using plot() function which is shown in figure 1 where Target=l specifies Persons have heart disease and target=0 specifies Persons not have heart disease.

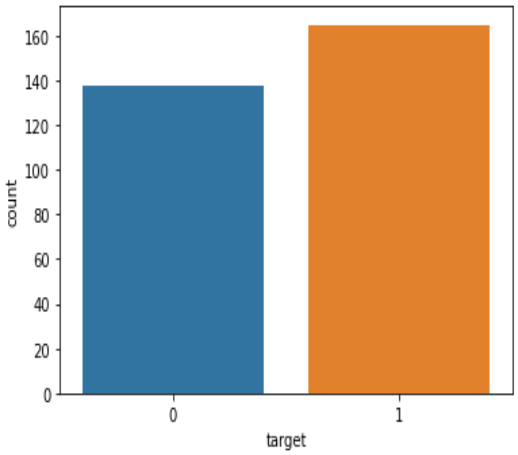


figure1: Distribution of target attributes

3.2.2. Distribution of the Data using histogram

Many distribution plots are plotted to check feature values and determine the facts distribution just so some interpretation of the facts can be observed. Different charts are displayed to provide an define of the facts. Figure 2 suggests the distributions of age,sex, chest pain, exang, oldpeak, slope and thal.

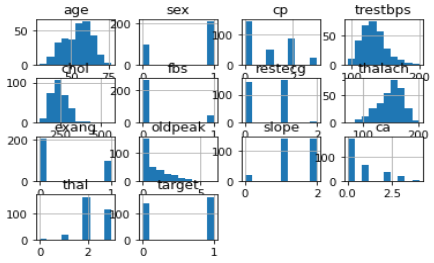


Figure 2: Histogram for distribution of attributes

Then carried out the correlation of the statistics attributes with every different and pick out simplest the great features. The correlation coefficient of two random variables X and Y can be calculated by

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

Here ρ is correlation coefficient and Cov is covariance and it can be calculated as

$$Cov(X,Y) = E((X - \mu x)(Y - \mu y))$$

3.2.3. Checking relationship between attributes.

The Corr ( ) function displays qualities that are both positively and negatively associated with the desired output. The target is moderately positive correlated with chest pain, ECG at rest, maximum heart rate, and ST segment during peak exercise, according to the correlation matrix. The target is moderately negative correlated with gender, age, blood pressure, cholesterol, blood sugar, ca, and the state of the heart, according to the correlation matrix.

Heatmaps are used to show the dataset's highly correlated and less associated properties using colours on a scale of one to ten. Heatmaps make data analysis easier by aggregating user behaviour and showing how each feature relates to the target attribute.

Figure 3: correlation matrix using heatmap

A. Random Forest

Random Forest is a selection version wherein every characteristic is represented via way of means of a selection tree. Each selection tree assigns a cost of one to a affected person who has a coronary artery sickness and a cost of zero to a affected person who does now no longer have a coronary artery sickness. This is a bagging approach that classifies samples via way of means of growing many selection bushes till the excellent end result is found. In a random forest, it can alter the inputs by adjusting factors such as the criterion, tree depth, maximum and minimum leaf, and so on. This algorithm was used to classify the data and it is not prone to over fitting and reliable approach that can fill in for missing values on its own The model is is more accurate when the number of trees are larger and the problem of over fitting is avoided. For a given dataset D, X = x1, x2, x3,...,xn, with goal values Y = y1, y2, y3,...,yn, bagging is done from I = 1 to N. The average of x predictions to discover the unseen samples.

i.e. $y = 1/N\sum_{i=1}^{n} fb(x)$ the error is derived by calculating the standard deviation of the trees

$$S.D. = \frac{\sum_{z=1}^{n}(fb(x) - f')^2 \quad 1/2}{N-1}$$

After transformation, Random Forest applied with stratified K-fold and Cross-Validation prevents data from over fitting and under fitting in which it split the percentage of data in train and test for the model and number of estimates specifies the parameter each tree produced.

Algorithm:

Step 1 Start with the selection of random samples from Processed Cleveland dataset from UCI.

Step 2: For each sample, create a decision tree.

Step 3: every anticipated result will be subjected.

Step 4: As the final prediction result, choose the one with the most points.

B. K-Nearest Neighbor (KNN)

The K-NN algorithm group all existing data and categories additional data points based on their similarity.The nearest neighbor set of rules is used for grouping and classification. It is usually utilized in predictive analytics. When new facts arrives, the ANN set of rules [8] identifies the closest present facts point. The `KNeighborsClassifier` with the characteristic n_neighbors = 10 changed into imported from the` sklearn library.

Algorithm.

Step 1: Pick a neighbor's number K.

Step 2: Determine the Euclidean distance between each of the K neighbours.

Step 3: Using the estimated Euclidean distance, find the K nearest neighbours.

Step 4: Count the number of data points in each category among these k neighbours.

Step 5: Assign the new data points to the category that has the highest number of neighbours.

Step 6: now have a completed model.

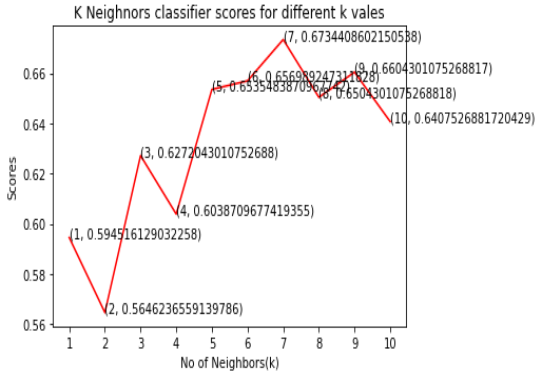


Figure 4: KNN applied with 10 nearest neighbours

C. ADA Boost

AdaBoost (Adaptive Boosting) is a method for finding a hypotheses with a low empirical risk while using a weak learner. The AdaBoost algorithm is fed a set of instances S = (x1, y1),...,(xm, ym), where yi = f(xi) for some labeling function f. The procedure of boosting is carried out in a series of cycles. The booster defines a distribution D(t) over the cases in S at round t. The booster then gives the weak learner the D(t) distribution and sample S.

Algorithm:

Step 1: First, assign the identical weight to every file withinside the dataset.

Step 2: Classify the pattern via way of means of stump. The selection tree is constructed at a intensity with

one node and leaves, additionally called stumps. Fit the version to the pattern and expect the elegance of the unique data.

Step 3: Calculate Total Error.Total Error = Weights of misclassified records

Step 4: Calculate Performance of the Stump.

Step 5: Update Weights and final predictions.

D. Gradient Boosting

To address classification and regression problems in Machine Learning, it is employ gradient boosting. It is a strategy for sequential ensemble learning in which the model's performance improves over time. The model is built in a stage-by-stage manner with this procedure. Gradient boosting is a method for constructing prediction models that is extremely reliable. It works with a variety of risk functions to improve the model's forecast accuracy. It allows the optimization of an absolutely variation error function, which infers the model. Likewise add each weak learner; a new model emerges that allows estimating the response variable more precisely. The following components are required for the gradient boosting technique to work.

Loss function optimization: it required to modify the loss function in order to minimize prediction errors. Gradient boosting does now no longer deliver the wrong end result a better weight, however it does attempt to decrease the loss characteristic via way of means of averaging the outputs from vulnerable freshmen. It wants vulnerable freshmen to make predictions with gradient boosting. Adding selection timber to gradient boosting tries to reduce the loss. By decreasing a parameters, additionally decreasing the mistake rate.As it bring about this situation, assemble the version in order that including a tree does now no longer adjust the present tree. Finally, the weights are up to date so one can lessen the computed error. Gradient boosting is much like selection frameworks in that it calculates the significance of variables related to coronary artery disease. It is used for its version overall performance and execution speed. The trees are constructed in this sort of manner that every one corrects the errors of the preceding one. Also each tree learns from its predecessors and corrects residual faults. As a result, the tree that follows with inside the grouping will advantage from a refreshed model of the residuals**.**

Algorithm:

Step 1: Calculate the common of the goal attribute.

Step 2: Calculate the residuals, residual = real cost expected cost.

Step 3: Next, construct a tree with the intention of predicting the residuals.

Step 4: Predict the goal label the usage of all the timber in the ensemble.

Step 5: Compute the brand new residuals.

Repeat steps 3 to 5 till the quantity of iterations suits the quantity particular through the hyper parameters (that is, the quantity of estimators).

Step 6: After training, use all the trees within side the ensemble to make very last predictions approximately the values of the goal variables.

E. Decision Tree

The decision tree method is a simple and efficient supervised learning method that continuously decomposes data points based on the specific parameters or problem an algorithm is trying to solve.

The most important thing to remember when designing a decision tree is to choose the best attribute for the root node. The extra buttons are selected from the full list of features.

Attribute Selection Measurement Method (ASM) is used to select the best attribute.

Information Gain and the Gini Index are two strategies for ASM. Entropy(S)- [(Weighted Avg) *Entropy(each feature) can be used to calculate Information Gain, where Entropy denotes the dataset's randomness. It's a statistic for determining impurity.

Gini index can be calculated using the below formula 1- $\sum_j P_j^2$Where pj stands for the probability. The entropy can be calculated by using: DT_Entropy = $-\sum_{z=1}^{n}$ pijlog$_2$p$_{ij}$

Algorithm:

Step 1; Begin the tree with the object of root node, which incorporates the whole information set. Step 2: Select the quality characteristic the usage of Attribute Selection Measurement (ASM). Step 3: Divide the basis into subsets containing viable values for the quality attributes. Step 4: Create a choice tree node containing the quality characteristic. Step five: Construct a brand new choice tree recursively the usage of subsets of the dataset created in step 3. Continue step five till the nodes can now not be ordered, wherein case the ultimate node is known as a leaf node.

F.Naive Bayes

A common data science classifier is the Naive Bayes algorithm. The idea behind this is to bring Bayesian theory to the surface. Bayes law is the sole basis of the naive Bayes algorithm. If there is one feature of the NAIVE BAYES model, it is irrelevant with or without the other features. This improvement enables functional predictor independence.

$$P(y|X) = \frac{P(X|y) * P(y)}{P(X)}$$

In classification, the goal of the predictive model is to identify the class that generated a particular instance. Consider such an instance x∈RN, a vector consisting of N features, x=[x1,x2,…,xN].

It is require to assign it to one of the M Classes C1,C2,…,CM depending on the values of the N features .

The training process of naive Bayes infers two quantitites

P(Cm) for all classes Cm∈[C1,C2,…,CM]

P(xn|Cm) for all features xn∈[x1,x2,…,xN], for all classes Cm∈[C1,C2,…,CM]

It has seen earlier that the calculation for P(Cm) is a straightforward ratio.

P(Cm)=Number of training instances belonging to Cm,

Total number of training examples

Figure 5: Architectural diagram.

G.Logistic Regression

Logistic regression is a statistical method for predicting outcomes. An analytical modeling technique known as logistic regression. Used to view data sets that have one or more independent factors that affect the results. Logistic regression taken from random null states. Then it fit the training model. The hypothesis class involved in logistic regression is the construction of the sigmoid function fsig: R [0, 1] for the class of linear functions. Specifically, the sigmoid function used for logistic regression is a logistic function and is defined as fsig(z) = 1/1 + exp(-z) H.**Support Vector Machine**

The purpose of the SVM algorithm creates a better decision restriction that can be split into a class allowing to easily classify new data points from the correct category. The best limitations for the solution are called SVM Hyperplan. Select the Extreme Dot / Vector that creates HyperPlane. This extreme case is called support vector, and the algorithm is called to enterprise support system. There is a specific SVM type that can be used for certain educational issues, such as support for SVC vector termination (SVR), such as support for the SVR (SVR), which is available for specific training issues. SVM is used to classify information with maximum benefits and find hyper plan. Training dataset= $\{x_i,y\}$

$X_{i->}$Represents all the attributes

y ->Represents the target

$f(x) = w^T x+ b$ is the equation of hyperplane which SVM locates by using the samples.

W is the weight factor and b is bias. This hyperplane is also called the optimal solution.

Step 1: Load the important libraries.

Step 2: Import dataset and extract the X variables and Y separately.

Step 3: Divide the dataset into train and test. .

Step 4: Initializing the SVM classifier model.

Step 5: Fitting the SVM classifier model.

Step 6: Coming up with predictions.

IV EXPERIMENTAL RESULTS

The distribution of coronary illness concerning age shows that age 41 to 64 having the most elevated level of cardio illness and the conveyance of males and females with cardio infection shows that the male having the most elevated level of coronary illness than female patients. The accuracy is determined after the learning, the test data samples are fed to the model to anticipate the classifications by comparing with the target value.

Accuracy= (TP+TN)/(TP+TN+FP+FN)

Precision=TP /TP+FP

Recall=TP/TP+FN

F1 Score= 2(Recall*Precision)/ (Recall + Precision) Precision specifies that the anticipation of the number of patients have coronary illness that actually belongs to the positive class. The major measure of accuracy is the precision with which a measured value compares to a known value. To improve the exhibition measurement esteems, it require to reduce the number of false negatives and false positives.

Figure 6. Confusion matrix description

The dataset contains 303 data instances is split into 80% for training which contains 242 instances and 20% for testing which contains 61 instances or 75% for training and 25% for testing.

RandomForestClassifier

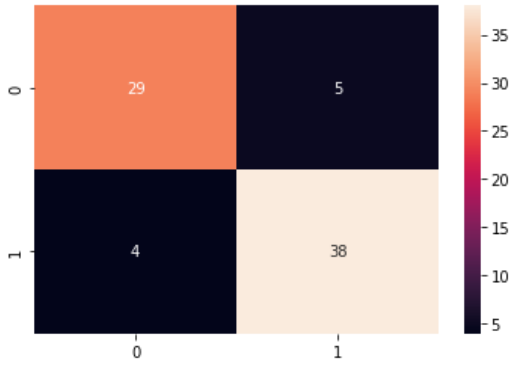


Figure 7: Confusion matrix for RF

Confusion matrix figure 7 shows that out of 76 test data 29 are found correctly the patient without heart disease. And 38 are found correctly with heart disease.

KNeighborsClassifier

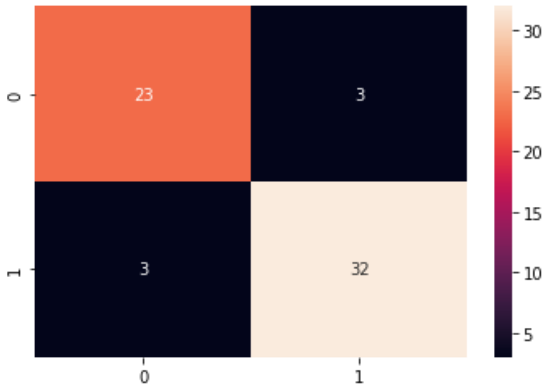


Figure 8: Confusion matrix for KNN

Confusion matrix figure8 shows that out of 61 test data 23 are found correctly the patient without heart disease and 32are found correctly with heart disease.

Logistic Regression

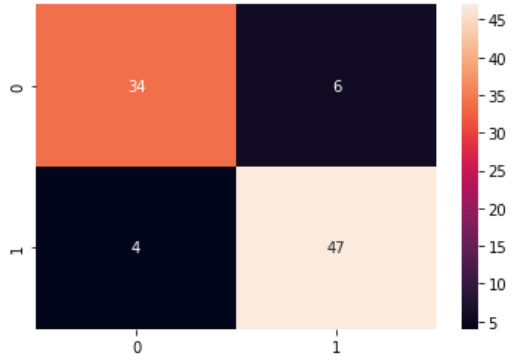


Figure 9: Confusion matrix for Logistic Regression

Confusion matrix figure 9 shows that out of 91 test data 34 are found correctly the patient without heart disease and 47 are found correctly with heart disease.

DecisionTree Classifier

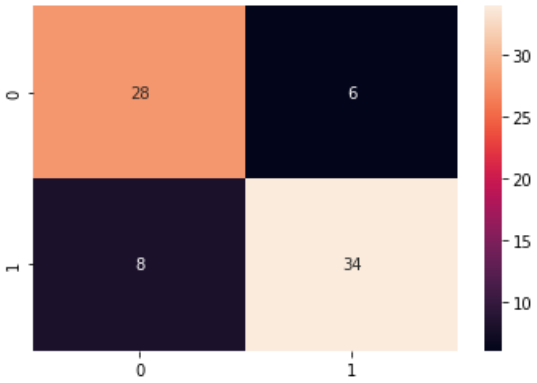


Figure 10 Confusion matrix for DecisionTree Classifier

Confusion matrix figure 10 shows that out of 76 test data 28 are found correctly the patient without heart disease and 34 are found correctly with heart disease.

Support Vector Machine

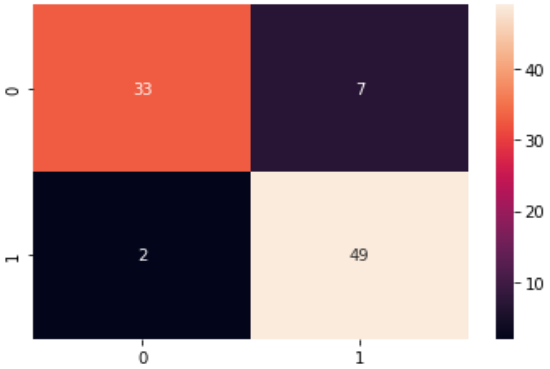


Figure 11: Confusion matrix for SVM Classifier

Confusion matrix figure 11 shows that out of 91 test data 33 are found correctly the patient without heart disease and 49 are found correctly with heart disease.

Naive Bayes Classifier

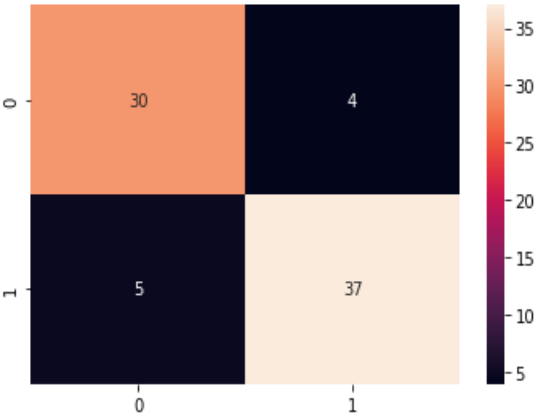
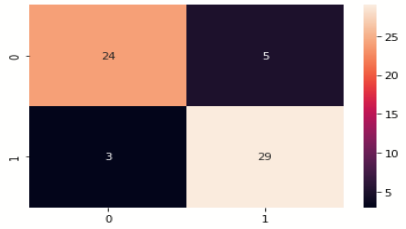


Figure 12: Confusion matrix for Naive Bayes Classifier

Confusion matrix figure 12 shows that out of 76 test data 30 are found correctly the patient without heart disease and 37 are found correctly with heart disease.

GradientBoostingClassifier



n_estimators=200, learning_rate=1.0

Figure 13: Confusion matrix for Gradient Boost

Confusion matrix figure 13 shows that out of 61 test data 24 are found correctly the patient without heart disease and 29 are found correctly with heart disease.

AdaBoostClassifier

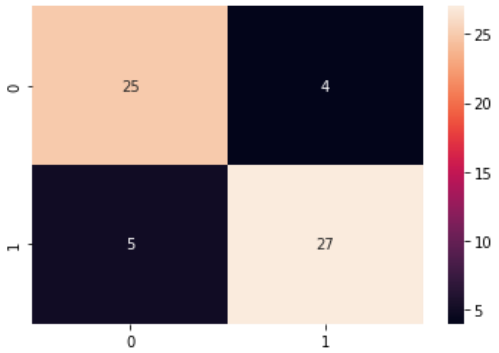


Figure 14: Confusion matrix for Ada Boost

Confusion matrix figure 14 shows from 61 test data 25 are found correctly the patient without heart disease and 27are found correctly with heart disease.

The predictive system can be generated with best suited model(SVM) and evaluated using the following python code:

```
input_data=(63,1,3,145,233,1,0,150,0,2.3,0,0,1)

input_data_as_numpy_array = np.asarray(input_data)

input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

prediction = svm.predict(input_data_reshaped)

print (prediction)

if (prediction[0]==0):

  print('Patient has no heart Disease')
```

else:

print(‘ Patient has heart Disease')

It produce the output as

[1] Patient has heart Disease

Table 2: Performances of different classifier

Algorithm	**Accuracy**
Logistic Regression	89.01%
KNeighborsClassifier	90.16%
DecisionTreeClassifier	81.58%
GradientBoostingClassifier	86.88%
AdaBoostClassifier	86.84%
Random forest	88.16%
Naive Bayes	88.16%
SVM	90.11%

V. CONCLUSION AND FUTURE WORK

The proposed method analyzes efficiency indicators for generating various models for training machines using a variety of ML procedures. Other algorithms are used to categorize support regression logistics, K nearest neighbor (KNN), Support Vector Machine (SVM), Random Forest, Gradient Boost and ADA Boost. The UCI data set contains 303 records with missing values, data pre-processing has been done and displacing the average estimate of all items. Data sets are divided into 20% for testing, and 61 elements and learning kits consist of 242 elements. The proposed model of the heart disease prediction SVM & KNN with 90% accuracy will help people, especially medical professionals expand various scenarios. This can help patients prevention to prevention of patients to make better understanding of human health. Easy to understand the characteristics associated with health risk associated with heart. On the other hand, the patient can prevent diseases recognized as consulting and testing doctors and doctors in advance. As a result, this model helps to form and develop people`s safety. In future studies, hybrid machine algorithms can be used to improve the accuracy and prediction of coronary heart disease models. Deep learning algorithms also play an important role in medical applications. As a result, predicting heart disease using critical learning procedures will yield the best results. This study can also perform problem classification in multiple classical problems to identify diseases.

REFERENCES

1. WHO. Available online: https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1
2. Health line. Available online: https://www.healthline.com/health/stroke-vs-heart-attack#treatment
3. Pasha, S.J.; Mohamed, E.S. Novel Feature Reduction (NFR) Model with Machine Learning and Data Mining Algorithms for Effective Disease Risk Prediction. IEEE Access 2020, 8, 184087–184108
4. Banumathi P. Miraclin Joyce Pamila J.C. , Coronary Illness Prediction and Analysis of Various Machine Learning Techniques 2020, IJSRCSE, Vol.8, Issue.3, pp.26-33
5. A. K. Garate-Escamila, A. Hajjam El Hassani, and E. Andres, "Classifification models for heart disease prediction using feature selection and PCA," Informatics in Medicine Unlocked,
   vol. 19, Article ID 100330, 2020.[6]
6. Rohit Bharti,Aditya Khamparia, Mohammad Shabaz, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning", Hindawi Computational Intelligence and Neuroscience, Volume 2021, Article ID 8387680, https://doi.org/10.1155/2021/8387680
7. Harvard Medical School 2020, Hungarian-Cleveland datasets were used for predicting heart disease using diffffferent machine learning classififiers and PCA was used for dimensionality reduction and feature selection.
8. Pabitra Kumar Bhunia "Heart Disease Prediction using Machine Learning"International Journal of Engineering Research & Technology (IJERT)
   ISSN: 2278-0181, Special Issue - 2021
9. Muhammad Saqib Nawaz , Bilal Shoaib , Muhammad Adeel Ashraf, " Intelligent Cardiovascular Disease Prediction Empowered with Gradient Descent Optimization" https://doi.org/10.1016/j.heliyon.2021.e06948
10. Dr. M. Kavitha , G. Gnaneswar, R. Dinesh , Y. Rohith Sai , R. Sai Suraj ," Heart Disease Prediction using Hybrid machine Learning Model ", IEEE Xplore Part Number: CFP21F70-ART; ISBN: 978-1-7281-8501-9
11. Archana Singh, Rakesh Kumar " Heart Disease Prediction Using Machine Learning Algorithms",2020 International Conference on Electrical and Electronics Engineering (ICE3-2020)
12. Anna Karen Garate-Escamilla, Amir Hajjam EL. Hassani, Emmanuel Andres, Classification models for heart disease prediction using feature selection andPCA, Informatics in Medicine Unlocked, Volume 19, 2020, 100330.
13. Harshit Jindal1, Sarthak Agrawal, "Heart disease prediction using machine learning algorithms" Harshit Jindal et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. **1022** 012072
14. Kaushalya Dissanayake, and Md Gapar Md Johar Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms, Applied Computational Intelligence and Soft Computing Volume 2021, Article ID 5581806 https://doi.org/10.1155/2021/5581806

15. Muhammad Azeem Sarwar, Nasir Kamal, Prediction of Diabetes Using Machine Learning Algorithms in Healthcare, Proceedings of the 24th International Conference on Automation & Computing, Newcastle University, Newcastle upon Tyne, UK, 6-7 September 2018.
16. S. J. Sushma , Tsehay Admassu Assegie, D. C. Vinutha , S. Padmashree, An improved feature selection approach for chronic heart disease detection Bulletin of Electrical Engineering and Informatics Vol. 10, No. 6, December 2021, pp. 3501~3506
ISSN: 2302-9285, DOI: 10.11591/eei.v10i6.3001
17. Vijeta Sharma, Shrinkhala Yadav "Heart Disease Prediction using Machine Learning Techniques"2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)
18. Ke Yuan,Longwei Yang "Heart Disease Prediction Algorithm Based on Ensemble Learning " 2020 7th International Conference on Dependable Systems and Their Applications (DSA) | 978-0-7381-2422-3/20/$31.00 ©2020 IEEE | DOI: 10.1109/DSA51864.2020.00052
19. Xiao-Yan Gao,Abdelmegeid Amin Ali "Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method", Hindawi Complexity Volume 2021, Article ID 6663455, 10 pages
https://doi.org/10.1155/2021/6663455