Journal of Machine Learning and Deep Learning (JMLDL)
Volume 1, Issue 1, Date: 31-December-2024
ISSN: 2978-4336, DOI: 10.64820/AEPJMLDL.11.10.19.122024

Prediction of Heart Diseases Using Cooperative AI Approach

Prabhudutta Ray [1,*], Dr. Raj Rawal [2], Dr. Brijesh Jajal [3], Dr. Ahesan Z. Rizvi [4]
[1, 3, 4] Computer Science & Engineering, Institute Of Advanced Research, Gandhinagar, India
[2] Gujarat Pulmonary and Critical Care Medicine, India
Email: [1,*] phdrayprabhu@gmail.com, [3] head.deps@iar.ac.in, [4] ahsan.rizvi@iar.ac.in
*Corresponding Author:

***Abstract*—Heart disease is a noticeable indicator of the rise of mortality rate. Deficiencies in pumping adequate blood throughout the body are the leading cause of heart disease. Major causes of mortality rate vary due to COPD disease in the world today. Prediction of cardiovascular disease is a major challenge to save lives at an early stage. It will be necessary and important to extract information from the raw medical data to predict the increasing mortality rate caused by heart diseases. Using critical data analysis by applying statistical techniques and providing correct features as input to the machine learning algorithm can predict the chances of heart attack and save many lives. Various machine learning methods are used to find out overall risk. In this paper, various machine learning techniques are applied and their comparative analysis is observed to correctly identify the CVD diseases.**

Keywords—Artificial Intelligence (AI), Cardiovascular Diseases (CVD), Machine Learning (ML), Co-operative AI.

I. INTRODUCTION

in the last few decades, along with the advancement of science and technology, medical science made a remarkable achievement in the prognosis or diagnosis of patients. This more or less accuracy in the field of medical science has become possible with the use of modern technology like ML and AI. ML and AI in the treatment of some critical diseases, especially diseases related to the heart. The application of AI has made the coordination between humans and machine easy and accurate. Moreover over application of AI especially in the healthcare industry along with the implementation of various tools like ML, NLP, and computer vision has made remarkable progress in the field of business and employment. AI is a smart technology which is able to process few amounts of data simultaneously. In other words, it is the technology of making the machine act like humans with the help of various applications [**1**]. AI is very prompt in making decisions, recognizing various patterns and at the same time is also capable of making judgements like a human being. It helps in providing better and more accurate search based on recommendations.

Co-operative AI is bliss in the medical field as it aims to improve the ability of AI systems along with engendering smooth cooperation between machines, humans and institutions. Further telemedicine is a new field in AI systems and has made a remarkable improvement in providing better treatment to patients especially those suffering from cardiac problems. A tele-ICU is a two-sided interaction between off-site clinicians and web-sided staff for providing consultation based on patient care. Here a centralized care team manages a large number of invisible ICUs situated in different parts of the world, to exchange health-related issues through electronic media in real time. As we all know the human heart is a vital organ of the human body. The heart circulates blood all through our body. It supplies oxygen and nutrients to the cells and absorbs carbon dioxide and other metabolic waste of our body. Our heart is also responsible for maintaining blood pressure. The degeneration of our heart occurs due to old age. At this time the coronary arteries of the heart either get damaged or blocked due to the deposition of fats and bad cholesterol in the body, illness, old age etc. When the flow of blood gets obstructed in the body, several disorders like – heart attack, excessive high BP, cerebral attack even leading to brain death may occur. Reports suggest that millions of people die globally due to CVD. The most important factors of stroke and heart diseases are generally due to lack of physical exercise, consumption of alcohol and tobacco, obesity, unhealthy diets etc. all these unhealthy habits lead to irregular blood pressure. Moreover, a hereditary history of CVD can lead to this disease for the next generation. According to the Centers for Disease Control and Prevention (CDC), CVD is the fourth leading cause of death in the USA.COPD is a collection of diseases which includes chronic bronchitis and emphysema making breathing difficult. Exposure to smoking, asthma, dust, and harmful chemicals can also lead to COPD. Spirometry helps to detect the severity of COPD. It is a type of lung function test which examines how much air a patient can exert. Individuals suffering from high BP are at a high risk of CVD. Moreover, high BP, high sugar, and lipids along with obesity increase the risk of CVD for this type of patient- as they enhance the chances of heart attack to a few more degrees. Abnormality in the flow of blood in various parts of the body like veins, arteries, and capillaries is also responsible for heart-related CVD diseases. Above 75% of the deaths are caused due to cardiac failure which is very common in developing and underdeveloped countries. The most common symptoms of CVD are chest pain leading to stroke, and difficulties in breathing leading to heart attack. To detect the severity of heart diseases and proper diagnosis of CVD various techniques like data mining, ML, and Neural networks along with new techniques recently introduced known as

Received: 15-08-2024
Revised: 14-09-2024
Published: 31-12-2024

cooperative AI have been applied to find out the extremity of heart diseases among people more accurately. Till now no proper solution has been detected to find out the proper cause of CVD all sorts of precautions must be taken to avoid it. Even science and technology are continuously striving to detect the causes of CVD along with proper prevention to reduce the rate of mortality. Complex measurement in medical applications along with ML and implementation of AI helps in providing advanced data analysis for healthcare providers to navigate the complexities of heart-related problems along with better accuracy and perfection[**2**].

II. RELATED WORK

Different researchers apply several types of techniques to solve coronary problems related to predictions and develop classification models. Table - 1 below describes related work, feature selection method, classifier used, and inference taken from the paper in tabular form.

Related work	**Method**	**Classifier used**	**Inference**
Predict severe cardiovascular event	Prediction and relation with respiratory problems [**3**]	CVD, ML, ML	ML give more prediction
Comorbid heart disease	chronic heart failure or ischemic heart disease [**4**]	ML, SVM, LR, DT	Strong clinical focus
Hybrid HRFLM-based techniques are used.	Discussed and designed new hybrid HRFLM techniques.[**5**]	RF and linear model	Accuracy achieved 88.7%
wearable system is based on heart disease prediction.	Proposed the wearable system-based CVD prediction[**6**]	NN, DNN,	Preprocessing feature selection and classification are performed
Hybrid ML technique used for heart disease prediction	Hybrid ML technique that combines classification and regression.[**7**]	RF, DT	Accuracy achieved 88.7% .
Prediction of CVD disorder	Describe a web-based solution for CVD disorder[**8**]	Electrocardiogram ; LSTM and CNN	MDCNN achieves an accuracy of 98.2.
DNN and linear SVC-based prediction techniques.	DNN and linear SVC-based integrated feature selection with gradient varnishing or explosion techniques.[**9**]	DNN, SVC	Achieved 98.56 % accuracy.
Early detection of CVD.	Suggested steps for detection of CVD problems associated with the lifestyle.[**10**]	Feature selection, prediction model, classification algorithms.	Prediction accuracy is 99.1%.
Heart disease disorder	Model-based on a recursion-enhanced random forest model for heart disease disorder.[**11**]	linear model with a random forest	Archived high accuracy and low classification error
NN-based approach for CVD prediction	Proposed ANN-based techniques for CVD disease prediction[**12**]	DNN, NN	Accuracy 98.4% achieved.
Feature Ranking Approach for Heart Diseases	Describe feature ranking techniques for CVD prediction using the Kaggel data set of 303 patients using AI approaches [**13**]	KNN, CNN, RF, DNN	Good accuracy scores for CNN for prediction.
CVD-based disease using specific features	Predicting heart disease using BMI is one of the important characteristics. [**14**]	SMOTE, KNN	Results show 99.1%, 98.0% and 95.5 % accuracy gain.

ML-based techniques for CVD prediction	compare the results of several data mining techniques.[**15**]	DT, NB, MLP, KNN, etc.	shows SVM using the boosting technique outperforms.
Evolutionary Algorithm-based CNN for Predicting CVD	Proposed CNN-JSO techniques for predicting cardiac diseases[**16**]	CNN-JSO	Predicting cardiac diseases with 94.12 % accuracy.
Range of peripheral factors for heart disease prediction	Estimation of Prediction for Heart Failure Chances Using Various Machine Learning Algorithms.[**17**]	NB, NN,RF, KNN	Extracts key information to increase the accuracy
CVD prediction for coronary blockages using mathematical rough set theory.	Developed CVD by RST and an RNN-based approach. [**18**]	RST-RNN	RST-RNN has shown highly accurate results for predicting heart diseases.
Heart disease prediction using a modeling framework	Developed predictive steps for big data clinical settings. [**19**]	LR, RF, SVM, NB	Good results in predictive modelling of CVD..
Class Imbalance in Heart Disease Prediction	Different learning rates were evaluated for the performance of MLP by using the sampling methods. [20**]**	RF, SVM, LR	Performance of MLP is evaluated
AI in CVD	Role of AI in CVD [**21**]	ML, NN, DT,	The potential of AI in medicine
AI in cardiology	AI's role in the future and present [**22**]	SVM, NN, RF	AI in future

III. METHODOLOGY

Classification is the process of categorizing a given set of data into classes. Classification can be performed for both structured and unstructured data. Predicting the class of the provided data points is the first step in the procedure. Common names for the classes include target, label, output, categories, etc. Different statistical and mathematical procedures such as linear programming, decision trees, and neural networks involved in classification. That notwithstanding, CVD detection can be recognized through classification procedures because it has two categories that is, one has CVD or not.

A. Support Vector Machines (SVM)

SVM is a technique for the ramification of both linear and nonlinear data. It applies a non-linear mapping method so that it can transform the training data into a higher dimension. A hyperplane is a kind of line that separates the input variable space in SVM. The hyperplane can separate the points in the input variable space containing their class which is either 0 or 1. In two dimensions, one can visualize this as a line and it is assumed that each input point can be completely separated by this line. The distance between the hyperplane and adjacent data coordinates is called the margin. The line which has the largest margin can distinguish between the two classes is known as the optimal hyperplane. These points are called support vectors, as they define or support the hyperplane. In practice, there is an optimization algorithm, which is used to calculate the values for the parameters that maximize the margin. This classifier separates the data point using the hyperplane with the largest amount of marginal differences. Support vectors check the data points which are close to the hyperplane. Depending on the several kernels the hyperplane can be decided. Kernels are different types like linear, polynomial, radial, and sigmoid [**32]**. SVM uses less memory because it uses a subset of training points in the decision phase. SVM is a technique for the ramification of both linear and nonlinear data. It applies a non-linear mapping method so that it can transform the training data into a higher dimension. A hyperplane is a kind of line that separates the input variable space in SVM. The hyperplane can separate the points in the input variable space containing their class which is either 0 or 1. In two dimensions, one can visualize this as a line and it is assumed that each input point can be completely separated by this line. The distance between the hyperplane and adjacent data coordinates is called the margin. The line which has the largest margin can distinguish between the two classes is known as the optimal hyperplane. These points are called support vectors, as they define or support the hyperplane. In practice, there is an optimization algorithm, which is used to calculate the values for the parameters that maximize the margin. Among the different classification techniques, the SVM is well known for its discriminative power for classification. The SVM is

widely considered in recent times due to its efficiency in most different pattern classification techniques. Kim et al. [**33**] proved that the SVM displays exceptional performance in the classification for prognostic prediction. We gave a brief mathematical description based on [**34**] of the SVM model. By assuming the binary classification of our response variable, CVD with the convention of linear divisibility for training samples, we have

$$S = \{(x_1, y_1), (x_2, y_2), \dots\dots\dots\dots (x_n, y_n)\} \text{ -------- } (\mathbf{1})$$

Where$x_i \in (|R|)$, such that the design matrix X belongs to the d-dimensional response space, and the response variable, CVD, is represented byy_i, which has a binary class in the vector Y with $y_i \in (0, 1)$in the study. The appropriate discriminating equation is given by

$f(x) = sign\{(z, x) + \beta\}]$ ---------- (**2**)

Similarly, Z represents the vector that determines the coordination of the hyperplane (discriminating plane), and so Z, X, and β are offsets. [**35**]. There are infinite numbers of possible hyperplanes that are efficiently classified by the training data which can be applied to the validation dataset. The optimal classifier identifies the similar optimal generalized hyperplanes that are nearer or even away from each cluster of objects [**36**]. The input set of coordinates is considered optimally separated by the hyperplane if there is accuracy in the separation with a maximum distance existing between the nearest components and the support vectors leading to the identification of a specific hyperplane [**37, 38**].

B. Random Forest (RF)

It is one of the ML algorithms that is called Bagging or Bootstrapping aggregation. To estimate a value from a data sample such as the mean, the bootstrap is a very powerful statistical approach. Here, lots of samples of data are taken, and the mean is calculated, after that, all of the mean values are averaged to give a better prediction of the real mean value. In bagging, the sampling method is used, but instead of estimating the mean of every data sample, decision trees are generally used. Here, numerous samples of the training data are considered and models are generated for every data sample. While a prediction for any data is needed, each model gives a prediction, and these predictions are then averaged to get a better estimation of the real output value.

C. Simple Logistic Regression

This method can be used for binary classification where values are distinguished with two classes. Logistic regression is similar to linear regression where the goal is to calculate the values of the coefficients within every input variable. Unlike linear regression, here the prediction of the output is constructed using a non-linear function which is called a logistic function. The logistic function transforms any value within the range of 0 to 1. The prediction made by logistic regression is used as the probability of a data instance concerning either class 0 or class 1. This can be necessary for problems where more rationale for a prediction is needed. Logistic regression works better when attributes are unrelated to the output variable and attributes correlated to one another are removed. Logistic regression is a classification technique based on a supervised learning approach. It uses probabilistic algorithmic calculation to predict outcomes. For estimating the probability it uses a logistic equation for measuring the relationship between the dependent variable and one or more independent variables [**39**]. It uses the sigmoid function for classification like

$$P = \frac{1}{1 + e^{-x}} \text{ ---------- } (\mathbf{3})$$

In this case, the logistic coefficients for each instance are given as$x_1, x_2, x_3, x_4, \dots\dots\dots\dots\dots x_n$ will be$b_0, b_1, b_2, b_3, \dots\dots\dots\dots\dots\dots b_n$ during the training phase [**40**]. Here stochastic gradient descent is used to calculate and update the coefficient values like

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + \cdots\dots\dots b_nx_n \text{ --- } (\mathbf{4})$$

Again

$$b = b + 1 * (y - p) * (1 - p) * p * x \text{ ------- } (\mathbf{5})$$

Here the output value is represented by y for each training phase, and all of the coefficients are initially set to 0. For b_0 biased input, 1 is the learning rate and x is always 1. By changing the coefficient values at the training level until it predicts the correct performance. Logistic regression depends on the proper representation of the data. To make the model useful important features are selected by using backward elimination and recursive elimination techniques are applied [**41**].

D. Decision Tree

this is the most important predictive modelling and classification method that is widely used in practical approaches. This algorithm can help to detect different ways by splitting the data sets based on numerous situations. In the classification tree, the responsive variable is considered a discrete set of values for tree methods. DT is a useful contemporary approach to solving decision-making challenges by building models that can be useful for prediction through systematic analysis. Branches represent the results, and leaves reflect the decisions that are produced after further computation. The steps for decision tre are as follows

a) divides the data set into two sub-data, that is, training and testing data sets.

b) in the initial stage, the entire training data are considered as the root.

c) continuous values are discretized before the model building, whereas categorical values are preferable for feature values.

d) establish subsets such that each subset includes data with the aforementioned feature attributes.

e) finally steps-(a) to steps (d) are repeated for each subset until we get the tree leaves.

In the DT, the prediction for a record class label begins at the root. the values are compared with the root features in the succeeding record characteristics. In this contrast the equivalent value of the next node to get analyzed.

IV. RESULTS AND DISCUSSION

In our current data set has the following attributes are taken into consideration which are as follows

```
Index(['age', 'sex', 'pheight ', 'pweight ', 'smoker', 'alcoholic ', 'mnocig',' cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalachh', 'exng', 'oldpeak', 'slope', 'ca', 'thal', 'output', 'oxyzen'], dtype='object').
```

The number of patients' details is shown in the graph in **Figure 1**

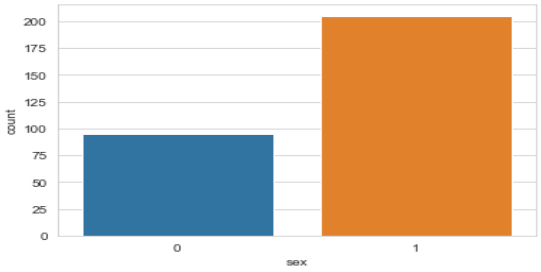

Figure 1: counterplot for patients with gender

Next, the heatmap representation of all the attributes is represented in **Figure 2** as follows

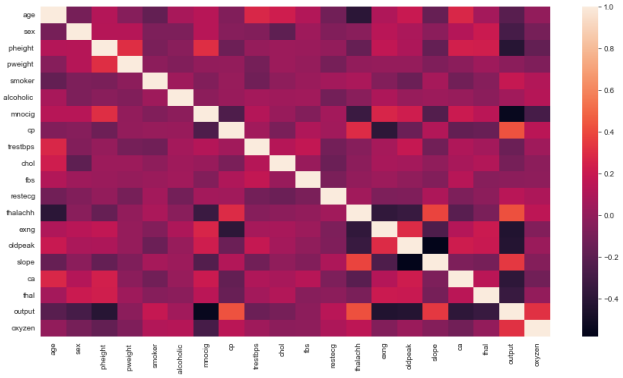

Figure 2: Representation of heat map from all the attributes

Next from the data set, it is required to identify the number of patients with different age groups which are as follows in Figure 3

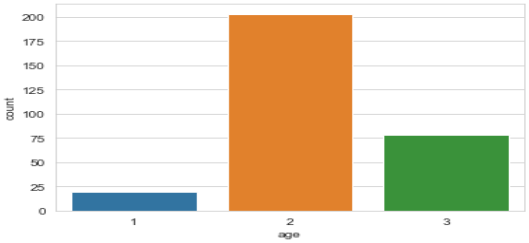

Figure 3: count of patients with different age groups

The next results show the attribute "exng" with the gender which is as follows in Figure 4 and the attribute chest pain with category in Figure 5

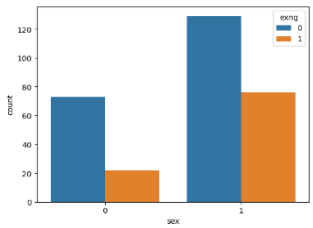

Figure 4: gender with exng

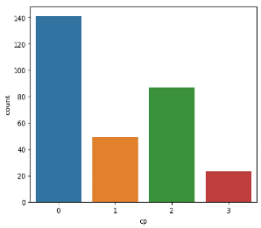

Figure 5: Chest Pain

The next results show the mortality rate of the data set in Figure **6**

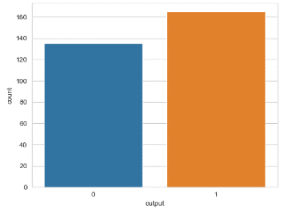

Figure 6: Mortality details

Nest GridsearchCv algorithm was applied with estimator is Logistic regression with the 5 fold cross-validation and the result

GridSearchCV
estimator: LogisticRegression

LogisticRegression
LogisticRegression()

Next, it will show the confusion matrix, accuracy score, and classification report [[30 2] [7 51]] The **Accuracy score is 90% and the corresponding performance details are below Table 1**

	precision	recall	F1-score	support
0	0.81	0.94	0.87	32

1	0.96	0.88	0.92	58
accuracy			0.90	90
macro avg	0.89	0.91	0.89	90
weighted avg	0.91	0.90	0.90	90

Figure 7 shows the performance graph for the above Table 1

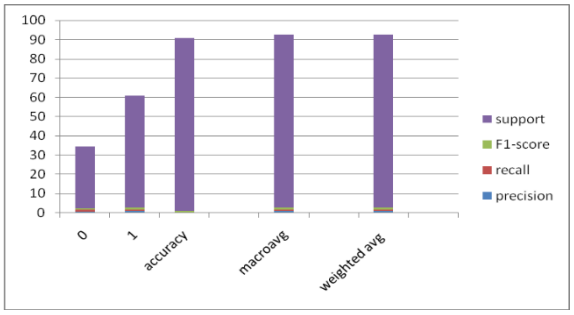

Figure 7: Performance Graph

Next use of mean square error is used to find the measure of deviation between the actual and predicted value by the use of the

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_{\cdot i})^2$$

formula and the calculated result for the **Mean squared error is mse = 0.1.**

5. Model creation and training, test, and spliting :

```
xtrain,xtest,ytrain,ytest=
train_test_split(x,y,test_size=0.2,random_state=42)
sc = StandardScaler()
xtrain = sc.fit_transform(xtrain)
xtest  = sc.fit_transform(xtest)
```

a) Logistic Regression

```
lr = LogisticRegression()
lr.fit(xtrain,ytrain)
lr.score(xtest,ytest)
The logistic regression score is 0.95
```

Predict some value

```
yp = lr.predict(xtest)
c = confusion_matrix(ytest,yp)
array([[26,  2],
       [ 1, 31]], dtype=int64)
```

Confusion Matrix

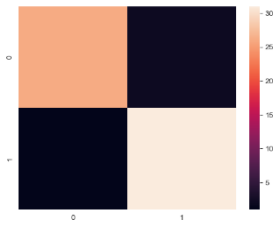

Figure 8: Confusion Matrix for LR

b) Decision Tree Classifier:

```
dt1 = DecisionTreeClassifier()
dt1.fit(xtrain,ytrain)
dt1.score(xtest,ytest)
Decision Tree classifier score is 0.85
```

The confusion matrix representation is

```
array([[25,  3],
       [ 6, 26]], dtype=int64)
```

Confusion Matrix

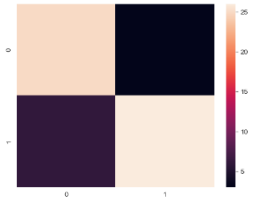

Figure 9 : Confusion Matrix for DT

c) svc algorithme

```
sv = SVC()
sv.fit(xtrain,ytrain)
sv.score(xtest,ytest)
SVC classifier score is
0.9666666666666667
ys=sv.predict(xtest)
sv1= confusion_matrix(ytest,ys)
array([[27,  1],
       [ 1, 31]], dtype=int64)
```

Confusion Matrix

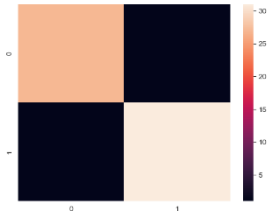

Figure 10: Confusion Matrix for SVC

d) Random forest classier

```
rfc = RandomForestClassifier(n_estimators = 200)
rfc.fit(xtrain,ytrain)
rfc.score(xtest,ytest)
Random Forest Classifier score is 0.9
yrfc=rfc.predict(xtest)
rfc1 = confusion_matrix(ytest,yrfc)
array([[26,  2],
       [ 4, 28]], dtype=int64)
```

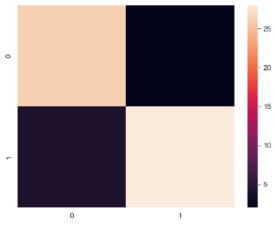

Figure 11 : Confusion Matrix for RF

6. Cross Validation Score

```
from sklearn.model_selection import cross_val_score
rfm = cross_val_score(rfc,X=xtrain, y=ytrain,cv=10)
array([0.95833333, 0.95833333, 0.9583333
3, 0.83333333,0.79166667,0.91666667,0.95
833333,0.95833333,0.83333333,0.95833333]
)
The mean value to find the accuracy of t
he rfc is 0.9125000000000002
```

The net accuracy of the random forest is 90%

e) k-nearest neighbor

```
from sklearn.neighbors import KNeighborsClassifier
kn = KNeighborsClassifier(n_neighbors=5)
kn.fit(xtrain,ytrain)
kn.score(xtest,ytest)
K-Nearest Neighbour classifier score is
0.9166666666666666
```

7. Prediction with the input value for the algorithm :

```
a = [[35,1,153,68,1,1,11,2,130,250,0,1,187,0,3.5,0,0,2,98.6]]
```

A) Prediction for logistic regression

```
lr.predict(a)
array([0], dtype=int64)
print(lr.predict(a))
if (lr.predict(a)==0):
   print('The person does not have a heart disease or healthy
heart')
else :
   print('The person has the heart Diseases')
[0]
The person does not have a heart disease
or healthy heart
```

B) Prediction for Decesion Tree

```
dt1.predict(a)
array([0], dtype=int64)
if (dt1.predict(a)==0):
   print('The person does not have a heart disease or healthy
heart')
else :
   print('The person has the heart Diseases')
[0] The person does not have a heart dis
ease or healthy heart
```

C) Prediction for SVM

```
sv.predict(a)
array([0], dtype=int64)
print(sv.predict(a))
if (sv.predict(a)==0):
   print('The person does not have a heart disease or healthy
heart')
else :
   print('The person has the heart Diseases')
[0] The person does not have a heart dis
ease or healthy heart
```

d) Prediction for random forest

```
rfc.predict(a)
array([0], dtype=int64)
print(rfc.predict(a))
if (rfc.predict(a)==0):
   print('The person does not have a heart disease or healthy
heart')
else :
   print('The person has the heart Diseases')

[0] The person does not have a heart dis
ease or healthy heart
```

V. COMPARISONS WITH ML ALGORITHMS

In this study, several ML models are taken into consideration with 10-fold cross-validation to choose the best models. The performance of each algorithm with the available dataset was selected and models with high accuracy scores were considered as best feature models, to predict the survival rate of patients with COPD. All the medical reports of the patient were carefully studied and it was found that patients with many blockages in the heart have the least chance of survival.

An ML algorithm is a set of rules used to conduct tasks to discover new data insights and patterns or predict output value from a given set of input variables. It acts as an exact list of instructions that conduct specified actions step by step in hardware- or software-based routines. They are in vast use in all IT areas. These models are trained accordingly based on the available set of data and features. The Logistic Regression adopts linear regression to estimate the probability of a class. This technique uses a sigmoid function where the outcome is limited to two sets of results- yes/ no, true/false, 0 or 1.

SVC is a type of supervised learning algorithm used in machine learning for solving classification and regression tasks. SVC is only concerned about the point nearest to the decision boundary. It is also helpful to find the best margin with separate classes and reduces the risk of error in the data. Unlike logistic regression, SVC works on the geometrical properties of data, while LR is restricted to statistical approaches only. In the below **table, -2** shows the performance of each algorithm and it was found that SVC provided the best results 96% which is expected for its performance and nature.

Table -2: Comparison among ML algorithms

SL No.	Algorithm	Accuracy score
1	Logistic regression	**95%**
2	Decision Tree	**85%**
3	SVC	**96%**
4	Random forest	**90%**
5	k-nearest neighbor	**91%**

From the above table-2, it is found that the support vector machine classifier gives a good accuracy core as compared to other algorithms. Further based on the same input vector a = [[35,1,153,68,1,1,11,2,130,250,0,1,187,0,3.5,0,0,2,98.6]] all the algorithms correctly classify the output.

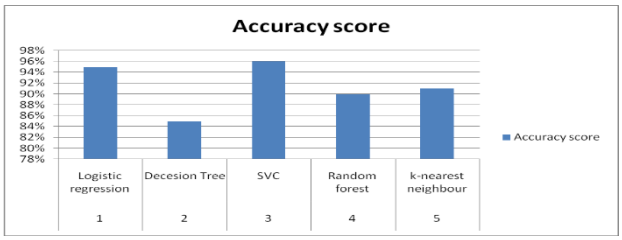

Figure 12: Comparison graph for ML algorithms with accuracy

Contributions: As stated earlier proper prediction of heart diseases is a very difficult task. It is found that a person affected with diabetes, obesity, high BP and without medication are at a high risk of heart disease. But without some medical examination, it is difficult to predict whether he suffering from co-morbidity or is really a cardiac patient. several models like naïve Bayes algorithm along with random forest with K-FOLD cross-validation for accurate prediction. According to industry standards, it was found that 70 % of the scores are acceptable but in our work, we have achieved 96% output which might be helpful for health care industries-specifically associated with the cardiac sector. In conclusion, it was finally achieved that ML-based algorithms have shown significant results in the health care sector, which will in turn lead to better success for more or less accurate prediction and treatment of diseases in the years to come.

VI. CONCLUSION

In this paper five machine learning methods were proposed in which comparative analysis was done and promising results were achieved. The collusion that was found is that the support vector machine learning algorithm performed better with 96% accuracy. The methods which are used for comparison are confusion matrix, precision, specificity, sensitivity, and F1 score. The 21 features that were used in the data set give good variation and better classification results. Risk assessment is important for early diagnosis and correct treatment of any kind of disease. CVD diseases show several disorders in arteries affecting the blood circulatory system in the heart.

The computational time was also reduced which is also helpful when developing the model. It was also found that statistical analysis is also important when a data set is analyzed and it should be a proper distribution technique.

VII. LIMITATIONS

In this work feature selection techniques are used to identify the important parameters but increasing more number of heart patients with other characteristic are also important for prediction. Here more different ways of normalizing the data and the corresponding results regarding comparison also need to be deeply examined.

VIII. FUTURE SCOPE

The dataset size can be increased and then deep learning with various other optimization techniques can be used in the future scope to achieve more promising results. Machine learning and various other optimization techniques can also be used so that the evaluation results can again be improved.

REFERENCES

[1] . Jamol Uzokov, Anis Alyavi , Bakhromkhon Alyavi , Akbar Abdullaev, " How artificial intelligence can assist with ischaemic heart disease ", Published: 10 February 2024 European Heart Journal, ehae030, https://doi.org/10.1093/eurheartj/ehae030

[2] . Saikumar, K., Rajesh, V. A machine intelligence technique for predicting cardiovascular disease (CVD) using Radiology Dataset. Int J Syst Assur Eng Manag 15, 135–151 (2024). https://doi.org/10.1007/s13198-022-01681-7

[3] . Anders LøkkeORCID Icon, Ole HilbergORCID Icon, Peter LangeORCID Icon, Rikke Ibsen, Gunilla TelgORCID Icon, Georgios Stratelis, "Exacerbations Predict Severe Cardiovascular Events in Patients with COPD and Stable Cardiovascular Disease–A Nationwide, Population-Based Cohort Study " Pages 419-429 | Received 16 Nov 2022, Accepted 27 Mar 2023, Published online: 31 Mar 2023.

[4] . Maaike Giezeman, Josefin Sundh , Åsa Athlin, Karin Lisspers, Björn Ställberg, Christer Janson , " Comorbid Heart Disease in Patients with COPD is Associated with Increased Hospitalization and Mortality – A 15-Year Follow-Up", Pages 11-21 | Received 23 Jun 2022, Accepted 05 Oct 2022, Published online: 08 Jan 2023.

[5] . Institute of Health Metrics and Evaluation. GBD Compare 2010. http://vizhub.healthdata.org/gbd-compare/. Accessed April 30, 2014

[6] . Junejo, A.R., Shen, Y., Laghari, A.A., Zhang, X. and Luo, H., 2019. Molecular diagnostic and using deep learning techniques for predicting functional recovery of patients treated with cardiovascular disease. IEEE Access, 7, pp.120315-120325.

[7] . Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y.R. and Suraj, R.S., 2021, January. Heart Disease Prediction using Hybrid Machine Learning Model. In 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 1329-1333). IEEE.

[8] . Khan, M.A., 2020. An IoT framework for heart disease prediction based on MDCNN classifier. IEEE Access, 8, pp.34717-34727.

[9] . M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," in *IEEE Access*, vol. 5, pp. 8869-8879, 2017, doi: 10.1109/ACCESS.2017.2694446.

[10] . Mohan, S., Thirumalai, C. and Srivastava, G., 2019. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, pp.81542-81554.

[11] . Nadakinamani, R.G., Reyana, A., Kautish, S., Vibith, A.S., Gupta, Y., Abdelwahab, S.F. and Mohamed, A.W., 2022. Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques. *Computational Intelligence and Neuroscience*, *2022*.

[12] . Nikam, A., Bhandari, S., Mhaske, A. and Mantri, S., 2020, December. Cardiovascular Disease Prediction Using Machine Learning Models. In 2020 IEEE Pune Section International Conference (PuneCon) (pp. 22-27). IEEE.

[13] . Praneetha, M., Jesudoss, A. and Mayan, A., 2021, May. Cardiovascular Disorder Prediction using Machine Learning. In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1665-1670). IEEE.

[14] . Rahim, A., Rasheed, Y., Azam, F., Anwar, M.W., Rahim, M.A. and Muzaffar, A.W., 2021. An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases. IEEE Access, 9, pp.106575-106588.

[15] . S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia and J. Gutierrez, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease," 2017 IEEE Symposium on Computers and Communications (ISCC), 2017, pp. 204-207, doi: 10.1109/ISCC.2017.8024530.

[16] . Samir, A.A., Rashwan, A.R., Sallam, K.M., Chakrabortty, R.K., Ryan, M.J. and Abohany, A.A., 2021. Evolutionary algorithm-based convolutional neural network for predicting heart diseases. Computers & Industrial Engineering, 161, p.107651.

[17] . T. P. Pushpavathi, S. Kumari and N. K. Kubra, "Heart Failure Prediction by Feature Ranking Analysis in Machine Learning," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 915-923, doi: 10.1109/ICICT50816.2021.9358733.

[18] . Talasila, V., Madhubabu, K., Madhubabu, K., Mahadasyam, M., Atchala, N. and Kande, L., 2020. The prediction of diseases using rough set theory with recurrent neural network in big data analytics. International Journal of Intelligent Engineering and Systems, 13(5), pp.10-18.

[19] . Nitten S. Rajliwall; Girija Chetty; Rachel Davey, "Chronic disease risk monitoring based on an innovative predictive modeling framework", IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1-8,2017.

[20] . Pinar Yildirim, "Chronic Kidney Disease Prediction on Imbalanced Databy Multilayer Perceptron", IEEE 41st Annual Computer Software and Applications Conference, pp. 193-198, 2017.

[21] . Yang YAN1,*, Jia-Wen ZHANG2,*, Guang-Yao ZANG3,*, Jun PU1, "The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine?", Journal of Geriatric Cardiology (2019) 16: 585-591 doi:10.11909/j.issn.1671-5411.2019.08.010

[22] . Loucia Karatzia, Nay Aung, Dunja Aksentijevic, "Artificial intelligence in cardiology: Hope for the future and power for the present", Front. Cardiovasc. Med., 13 October 2022 Cardiovascular Imaging ,Volume9-2022 https://doi.org/10.3389/fcvm.2022.945726

[23] . Maryam Soltanpour Gharibdousti, Kamran Azimi "Prediction of Chronic Kidney Disease Using Data Mining Techniques", Proceedings of the Industrial and Systems Engineering Conference, 2017.

[24] . M. Sharma, Tan, RS. Acharya, U.R. A new method to identify coronary artery disease with ECG signals and time-Frequency concentrated antisymmetric biorthogonal wavelet filter bank, Pattern Recognition Letters, Vol. 125, pp. 235-240(2019).

[25] . Pahulpreet Singh Kohli and Shriya Arora. (2018). "Application of Machine Learning in Diseases Prediction", 4th International Conference on Computing Communication And Automation(ICCCA).

[26] . A. Mishra, Abhishek Rai and Akhilesh Yadav, (2014). Medical ImageProcessing: A ChallengingAnalysis" International Journal of Bio-Science and Bio-Technology Vol.6, No.2.

[27] . Sneha A Mane, S R Chougule, (2016). NeuralNetworkk of Kidney Stone Detection" International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Volume 5 Issue 4.

[28] . Adam, Tijjani, U. Hashim, and U. S. Sani, (2012). Designing an Artificial Neural Network model for the prediction of kidney problems symptoms through patient's mental behavior for pre-clinical medical diagnosis. Biomedical Engineering (ICoBE), 2012 International Conference on. IEEE.

[29] . Aditi Gavhane, GouthamiKokkula, IshaPanday, Prof. Kailash Devadkar. (2018) "Prediction of Heart Disease using Machine Learning", Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology(ICECA), 2018.

[30] . Amandeep Kaur and Jyoti Arora. (2019)."Heart Diseases Prediction using Data Mining Techniques: A survey" International Journal of Advanced Research in Computer Science, IJARCS.

[31] . M. Nikhil Kumar, K. V. S. Koushik, K. Deepak. (2019). "Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools" International Journal of Scientific Research in Computer Science, Engineering and Information Technology, IJSRCSEIT.

[32] . Himanshu Sharma and M. A. Rizvi. (2017). "Prediction of Heart Disease using Machine Learning Algorithms: A Survey" International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 8, IJRITCC August 2017.

[33] . A. Sahaya Arthy, G. Murugeshwari, (April 2018): A survey on heart disease prediction using data mining techniques.

[34] . Monika Gandhi, Shailendra Narayanan Singh, (2015): Predictions in heart diseases using techniques of data mining.

[35] . Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava, (2019): Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, Digital Object Identifier 10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7.

[36] . M. Tanveer, T. Rajani, R. Rastogi, Y. H. Shao, and M. A. Ganaie, “Comprehensive review on twin support vector machines,” *Annals of Operations Research*, pp. 1–46, 2022.

[37] . B. M. Kim, B. Y. Kang, H. G. Kim, and S. H. Baek, “Prognosis prediction for Class III malocclusion treatment by feature wrapping method,” *&e Angle Orthodontist*, vol. 79, no. 4, pp. 683–691, 2009.

[38] . N. Cristianini and J. Shawe-Taylor, *An Introduction to SupportvVector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge, UK, 2000.

[39] . I. Steinwart and A. Christmann, *Support Vector Machines*, Springer Science & Business Media, Berlin, Germany, 2008.

[40] . M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, “Support vector machines,” *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.

[41] . J. Fan, J. Zheng, L. Wu, and F. Zhang, “Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, arti7cial and deep neural networks models,” *Agricultural Water Management*, vol. 245, Article ID 106547, 2021.

[42] . A. Kurani, P. Doshi, A. Vakharia, and M. Shah, “A comprehensive comparative study of arti7cial neural network (ANN) and support vector machines (SVM) on stock forecasting,” *Annals of Data Science*, pp. 1–26, 2021.

[43] . Ankur Gupta, Rahul Kumar, Harkirat Singh Arora, Balasubramanian Raman, MIFH: A Machine Intelligence Framework for Heart Disease Diagnosis, Publisher: IEEE, Page(s): 14659 – 14674, IEEE Access (Volume: 8), INSPEC Accession Number: 19313528, Electronic ISSN: 2169-3536, DOI: 10.1109/ACCESS.2019.2962755.

[44] . Rahul Katarya, Sunit Kumar Meena, Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis, Health and Technology volume 11, pages 87–97 (2021), Published: 19 November 2020.

[45] . Dr. R. Lakshminarayanan, Dr.L.Thanga Mariappan, N.Yuvaraj, Analysis On Cardiovascular Disease Classification Using Machine Learning Framework, Vol. 63 No. 6 (2020).