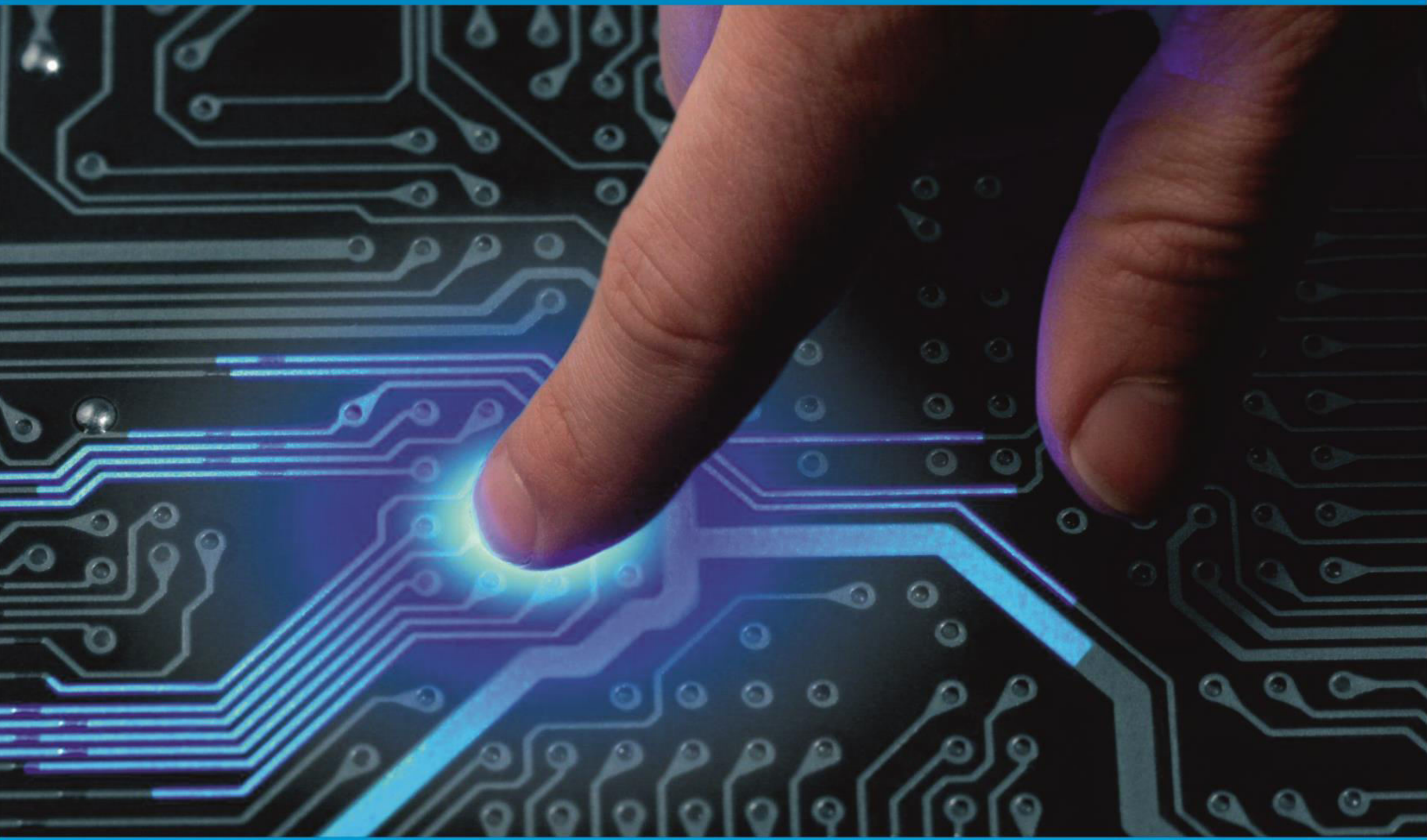
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 3, March 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

9940 572 462 6381 907 438 ijircce@gmail.com www.ijircce.com

| DOI: 10.15680/IJIRCCE.2021. 0903091 |

Prediction of Heart Disease Using Ensemble Feature Selection of Machine Learning Algorithms

P.Pugalendhi , Mr.V.Sornagopal, Mrs. K.Periyarselvam , Dr.P.Sivakumar

PG Student, Department of ECE, GRT Institute of Engineering and Technology, Tiruttani, India

Associate Professor, Department of ECE, GRT Institute of Engineering and Technology, Tiruttani, India

Associate Professor, Department of ECE, GRT Institute of Engineering and Technology, Tiruttani, India

Professor & Head, Department of ECE, GRT Institute of Engineering and Technology, Tiruttani, India

ABSTRACT: Machine Learning is used across many spheres around the world. The healthcareindustry is no exception. Machine Learning can play an essential role in predictingpresence/absence of Heart diseases and more. Heart disease is one of the complexdiseases and globally many people suffered from this disease. On time andefficient identification of heart disease plays a key role in healthcare, particularlyin the field of cardiology. In this an efficient and accurate system to diagnosisheart disease and the system are based on machine learning techniques. Theproposed framework based on novel Fast Conditional Mutual Information featureselection algorithm to solve feature selection problem. The features selectionalgorithms are used for features selection to increase the classification accuracyand reduce the execution time of classification system. This research paper aims toenvision the probability of developing heart disease in the patients. The resultsportray that the highest accuracy score is achieved with Gradient..

KEYWORDS: Machine Learning,Prediction,LT, Naïve bias, Decision Tree.

I. INTRODUCTION

According to the World Health Organization, every year 12 million deaths occur worldwide due to Heart Disease. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications. This project aims to predict future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithms.

Nowadays, data mining is becoming popular in healthcare domain as there is need of efficient analytical methodology for detecting unknown and valuable information in healthcare domain. Cardiovascular disease (also called heart disease) is a class of diseases that involve the heart or blood vessels (arteries, capillaries, and veins). Cardiovascular disease is the leading cause of deaths worldwide, since 1970s the cardiovascular mortality rates have declined in many high-income countries. At the same time, cardiovascular deaths and disease have increased at a fast rate in low and middle income group countries.

Although cardiovascular disease usually affects older adults, the antecedents of cardiovascular disease, notably atherosclerosis; begins at early stages of life making primary prevention efforts necessary from childhood. Therefore, increased emphasis on preventing atherosclerosis by modifying risk factors, evidence suggests a number of risk factors for heart disease such as age, gender, high blood pressure, high serum cholesterol levels, smoking, excessive alcohol consumption, sugar consumption, family history, obesity, lack of physical activity, psychosocial factors, diabetes mellitus, air pollution and using tobacco. The World Health Statistics 2012 report enlightens the fact that one in three adults worldwide has raised blood pressure – a condition that causes around half of the deaths from stroke and heart disease. Heart disease is the major cause of casualties in the different countries including India. Heart disease kills one person in every 34 seconds in the United States. Diagnosis is complicated and important task that needs to be executed accurately and efficiently.

II. RELATED WORK

Authors A.U. Haq, J. Li, M.H. Memon, J. Khan and S.M.Marium gives the development of the system we used Sequential backward selection feature algorithm to select important number of features for best classification accuracy. The supervised learning classifier K-NN was used in the system for classification with training and testing split method in order to train and test the classifier performance. The proposed system methodology have five steps Preprocessing, Feature selection, Training/testing split, classifier and classifier performance measuring metrics.

Author U. Haq, J. P. Li, M. H. Memon, S. Nazir and R. Sun proposes a hybrid ensemble model for heart disease detection and prediction which focuses on predicting labels of each SPECT image based on feature vector of the images and the labels that base classifiers assign to each image. To facilitate understanding of the proposed framework, in this section we describe the details of layout of the proposed model. A schematic illustration of proposed hybrid ensemble model can be seen. It consists of three modules, including partitioning module, inner classifiers module and fuser module. The initial dataset is first given to partitioning module to produce train and test subsets and prepare them for the next module. In inner classifiers module different classification algorithms are applied on the train and test datasets to produce input data for fuser module in which results of base classifiers next to initial feature vector of samples are considered simultaneously for building and adjusting components of the final classifier. In the rest of this section, a brief description of each component is given.

Author X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang, and Q. Wang tells the three advantages to combining ReliefF and RS (RFRS) approach as an integrated feature selection system for heart disease diagnosis.The RFRS method can remove superfluous and redundant features more effectively. The ReliefF algorithm can select relevant features for disease diagnosis; however, redundant features may still exist in the selected relevant features. In such cases, the RS reduction algorithm can remove remaining redundant features to offset this limitation of the ReliefF algorithm.

III. PROPOSED ALGORITHM

The proposed methodology a machine learning based diagnosis method for the identification of HD in this research work. Machine learning predictive models include LR, K-NN, SVM, DT, and NB are used for the identification of HD. Also proposed Fast Conditional Mutual Information (FCMIM) features selection algorithm for features selection. Apart from this, different performance assessment metrics have been used for classifiers performances evaluation. The proposed method has been tested on Cleveland HD dataset. Furthermore, the performance of the proposed technique has been compared with state of the art existing methods in the literature.

The contribution of the proposed research is to design a machine-learning-based medical intelligent decision support system for the diagnosis of heart disease. In the present study, various machines learning predictive models such as Logistic Regression, K-Nearest Neighbor, ANN, SVM, Decision Tree, Naive Bayes, and Random Forest have been used for classification of people with heart disease and healthy people. This feature selection algorithms, Relief, Minimal Redundancy-Maximal-Relevance (mRMR), Shrinkage and Selection Operator (LASSO), were also used to select the most important and highly correlated features that great influence on target predicted value. Cross-validation methods like k-fold were also used. In order to evaluate the performance of classifier, various performance evaluation metrics such as classification accuracy, classification error, specificity, sensitivity, Matthews' Correlation Coefficient (MCC), and Receiver Optimistic Curves (ROC) were used. Additionally, model execution time has also been computed. Moreover, data preprocessing techniques were applied to the heart disease dataset. 0e proposed system has been trained and tested on Cleveland heart disease dataset, 2016. UCI data-mining repository the dataset of Cleveland heart disease is available online.
All the computations were performed in Python.

Pseudo-Code of proposed heart disease diagnosis system

```
1.  Begin
2.  The pre-processing of heart disease dataset using preprocessing
    methods
3.  Features selection using standard state of the art and proposed FCMIM FS algorithms
4.  Train the classifiers using training dataset
5.  Validate using testing dataset
6.  Computes performance evaluation metrics
```

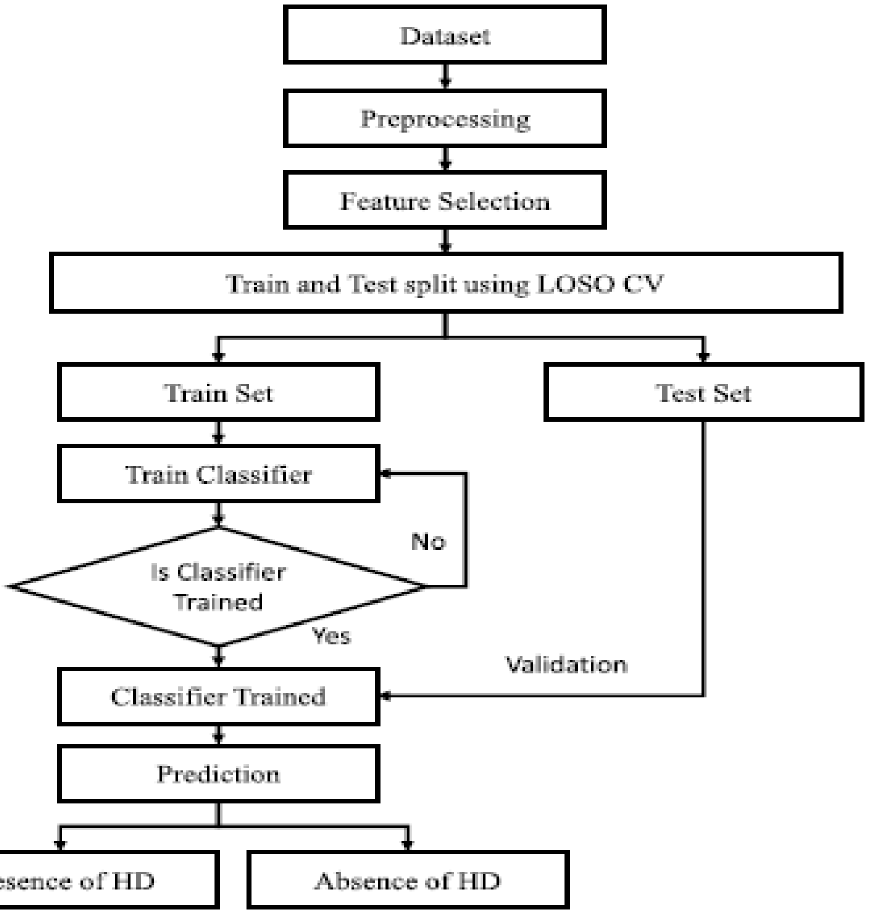
7. End

Logistic Regression

It is a statistical model. A logistic regression is a classification algorithm. For binary classification problem, in order to predict the value of predictive variable y when y ∈ [0, 1], 0 is negative class and 1 is positive class. It also uses multi classification to predict the value of y when y ∈ [0, 1, 2, 3]. In order to classify two classes 0 and 1, a hypothesis h(θ) = θ TX X will be designed and threshold classifier output is hθ(x) at 0.5. If the value of hypothesis hθ (x) ≥ 0.5, it will predict y =1 which mean that the person has heart disease and if value of hθ(x) < 0.5, then predict y = 0 which shows that the person is healthy. Hence, the prediction of logistic regression under the condition 0 ≤ hθ(x) ≤ 1 is done.

K- Nearest Neighbor

K-NN is a supervised learning classification algorithm. K-NN algorithm predicts the class label of a new input. K-NN utilizes the similarity of new input to its inputs samples in the training set. If the new input is same the samples in the training set. The K-NN classification performance is not good. Let (x, y) be the training observations and the learning function h: X ⟶Y, so that given an observation x, h(x) can determine y value.



IV. **PROPOSED BLOCK DIAGRAM**

Random Forest

Random forest is a tree based classification algorithm. As the name indicates, the algorithm creates a forest with a large number of trees. It is an ensemble algorithm which combines multiple algorithms. It creates a set of decision trees from a random sample of the training set. It repeats the process with multiple random samples and makes a final decision based on majority voting The Random forest algorithm is effective in handling missing values but it is prone to overfitting. Appropriate parameter tuning can be applied to avoid overfitting.

Support Vector Machine

SVM is a machine learning classification algorithm which has been mostly used for classification problems SVM used a maximum margin strategy that transformed into solving a complex quadratic programming problem. Due to the high performance of SVM in classification, various applications widely applied. In a binary classification problemwhich are normal to the hyper plane of the surface, b is offset value from the origin, and x is data set values. The SVM gets results of w and b. w can be solved by introducing Lagrangian multipliers in the linear case. 0e data points on borders are called support vectors.

Naive Bayes

The NB is a classification supervised learning algorithm. It is based on conditional probability theorem to determine the class of a new feature vector. 0e NB uses the training dataset to find out the conditional probability value of vectors for a given class. After computing the probability conditional value of each vector, the new vectors class is computed based on its conditionality probability. NB is used for text-concerned problem classification.

Artificial Neural Network

The artificial neural network is a supervised machine learning algorithm and is a mathematical model that integrates neurons that pass messages. 0e ANN has three components including inputs, outputs, and transfer functions. 0e input units take extraordinary values and weights, which are modified during the training process of the network. 0e output of the artificial neural network is calculated for the known class; the weight is recomputed using the error margin between the output of predicted and actual class. ANN is designed by the integration of neurons. 0is different combination of neurons from different structures is just like multilayer perception.

Decision Tree Classifier

A decision tree is a supervised machine learning algorithm. A decision tree shape is just a tree where every node is a leaf node or decision node. The techniques of the decision tree are simple and easily understandable for how to take the decision. A decision tree contained internal and external nodes linked with each other. The internal nodes are the decision-making part that makes a decision and the child node to visit the next nodes. The leaf node on the other hand has no child nodes and is associated with a label.

K-Nearest Neighbor

K-NN is a supervised learning classification algorithm. K-NN algorithm predicts the class label of a new input; K-NN utilizes the similarity of new input to its inputs samples in the training set. If the new input is same the samples in the training set. 0e K-NN classification performance is not good.

Fast Conditional Mutual Information feature selection algorithm

The feature selection problem, we proposed Fast Conditional Mutual Information (FCMIM) feature selection algorithm in this study. It is an efficient feature selection method which is designed from Conditional Mutual Information (CMI). The ``FCMIM" algorithm designing having the following procedures. Let us consider a dataset O(X; Y), where X instances and Y is output labels. The FCMIM high value shows that feature Xn is more relevant to output Y and is highly compatible with another selected feature Xj where j belongs to O.

Input: load the HD dataset, where O(X, Y) as a data matrix, X is instances and Y output labels. Maxnumerfeatures, selectedfeaturesubset, MI(Mutual Information), CMI(Conditional mutual Information), L(least used index), p(partial score)

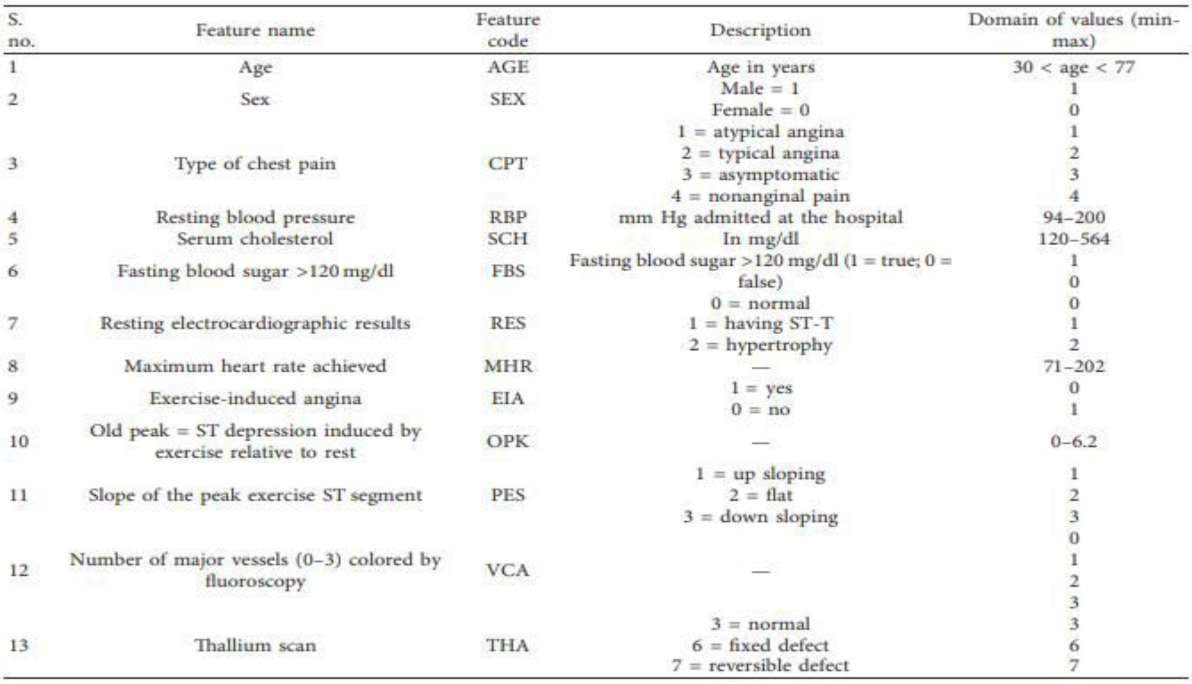
Output: selected featuresubsetO(xi; yi)

```
Pre-process the dataset
Initialize selected features D _
for features oi 2 O do
ComputeMi
set pi Mi
set Li 0
end for
for k 1 to K do Initialize scorei 0
for features oiinO do
while Pi >scorek And Li < k □ 1 do
set Li Li C 1
Calculate VUi between ok and oi
```

```
        Set pi   min(piCMik   )
    end while
    if pi >scorek then
        Set scorek D pi
        Selected featuressubset   Selected features
subsetUoi
    end if
    end for
end for
```

Dataset

"Cleveland heart disease dataset 2016" is used by various researchers and can be accessed from online data mining repository of the University of California, Irvine. 0is dataset was used in this research study for designing machine-learning-based system for heart disease diagnosis. 0e Cleveland heart disease dataset has a sample size of 303 patients, 76 features, and some missing values. During the analysis, 6 samples were removed due to missing values in feature columns and leftover samples size is 297 with 13 more appropriate independent input features, and target output label was extracted and used for diagnosing the heart disease. 0e target output label has two classes in order to represent a heart patient or a normal subject. 0us, the extracted dataset is of 297∗ 13 features matrix. 0e complete information and description of 297 instances of 13 features of the dataset is given in below table.



S. no.	Feature name	Feature code	Description	Domain of values (min-max)
1	Age	AGE	Age in years	30 < age < 77

2	Sex	SEX	Male = 1
			Female = 0
				1
				0

3	Type of chest pain	CPT	1 = atypical angina 2 = typical angina 3 = asymptomatic 4 = nonanginal pain	1 2 3 4
4	Resting blood pressure	RBP	mm Hg admitted at the hospital	94–200
5	Serum cholesterol	SCH	In mg/dl	120–564
6	Fasting blood sugar >120 mg/dl	FBS	Fasting blood sugar >120 mg/dl (1 = true; 0 = false)	1 0
7	Resting electrocardiographic results	RES	0 = normal 1 = having ST-T 2 = hypertrophy	0 1 2
8	Maximum heart rate achieved	MHR	—	71–202
9	Exercise-induced angina	EIA	1 = yes 0 = no	0 1
10	Old peak = ST depression induced by exercise relative to rest	OPK	—	0–6.2
11	Slope of the peak exercise ST segment	PES	1 = up sloping 2 = flat 3 = down sloping	1 2 3
12	Number of major vessels (0–3) colored by fluoroscopy	VCA	—	0 1 2 3
13	Thallium scan	THA	3 = normal 6 = fixed defect 7 = reversible defect	3 6 7

V. SIMULATION RESULTS

The dataset is publicly available on the Kaggle Website. The attributes include: age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting, sugar blood, resting electrocardiographic results, maximum heart rate, exercise induced angina, ST depression induced by exercise, slope of the peak exercise, number of major vessels, and target ranging from 1 and 2, where 1 is absence of heart disease. The data set is in csv (Comma Separated Value) format which is further prepared to data frame as supported by pandas library in python. The education data is irrelevant to the heart disease of an individual, so it is dropped. Further with this dataset pre-processing and experiments are then carried out.

DATA PREPARATION

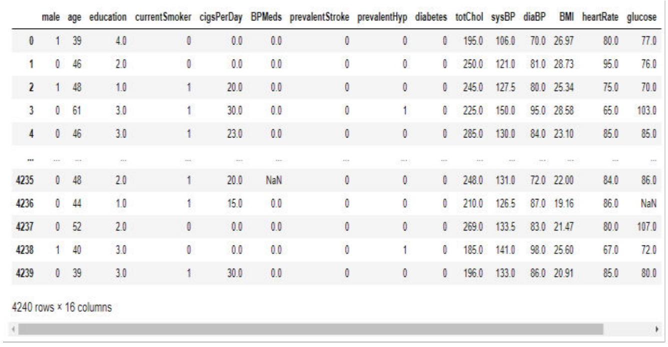
Cleveland Heart Disease dataset is considered for testing purpose in this study. During the designing of this data set there were 303 instances and 75 attributes, however all published experiments refer to using a subset of 14 of them. In this work, we performed pre-processing on the data set, and 6 samples have been eliminated due to missing values. The

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903091 |

remaining samples of 297 and 13 features dataset is left and with 1 output label. The output label has two classes to describe the absence of HD and the presence of HD. Hence features matrix 297*13 of extracted features is formed.



	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP	BMI	heartRate	glucose
0	1	39	4.0	0	0.0	0.0	0	0	0	195.0	106.0	70.0	26.97	80.0	77.0
1	0	46	2.0	0	0.0	0.0	0	0	0	250.0	121.0	81.0	28.73	95.0	76.0
2	1	48	1.0	1	20.0	0.0	0	0	0	245.0	127.5	80.0	25.34	75.0	70.0
3	0	61	3.0	1	30.0	0.0	0	1	0	225.0	150.0	95.0	28.58	65.0	103.0
4	0	46	3.0	1	23.0	0.0	0	0	0	285.0	130.0	84.0	23.10	85.0	85.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4235	0	48	2.0	1	20.0	NaN	0	0	0	248.0	131.0	72.0	22.00	84.0	86.0
4236	0	44	1.0	1	15.0	0.0	0	0	0	210.0	126.5	87.0	19.16	86.0	NaN
4237	0	52	2.0	0	0.0	0.0	0	0	0	269.0	133.5	83.0	21.47	80.0	107.0
4238	1	40	3.0	0	0.0	0.0	0	1	0	185.0	141.0	98.0	25.60	67.0	72.0
4239	0	39	3.0	1	30.0	0.0	0	0	0	196.0	133.0	86.0	20.91	85.0	80.0

4240 rows × 16 columns

Figure DATASET WITH SAMPLE ATTRIBUTES

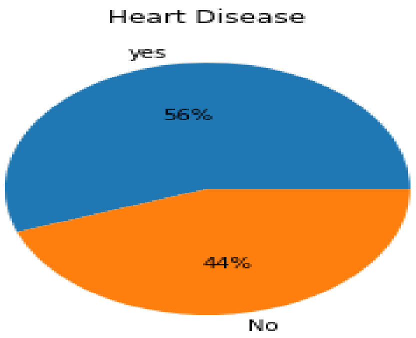
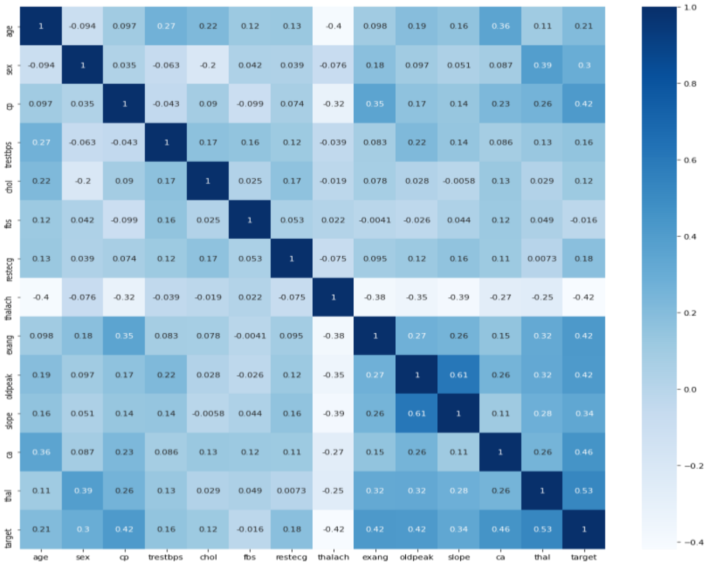


Figure HEART DISEASE

EXPLORATORY ANALYSIS

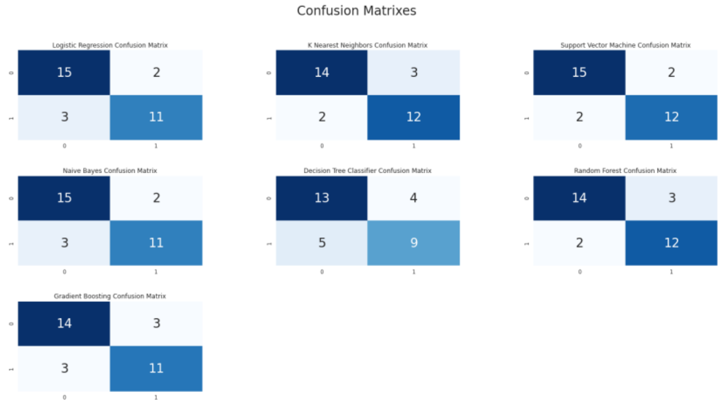
Correlation Matrix visualization Before Feature Selection. It shows that there is no single feature that has a very high correlation with our target value. Also, some of the features have a negative correlation with the target value and some have positive. The data was also visualized through plots and bar graphs.



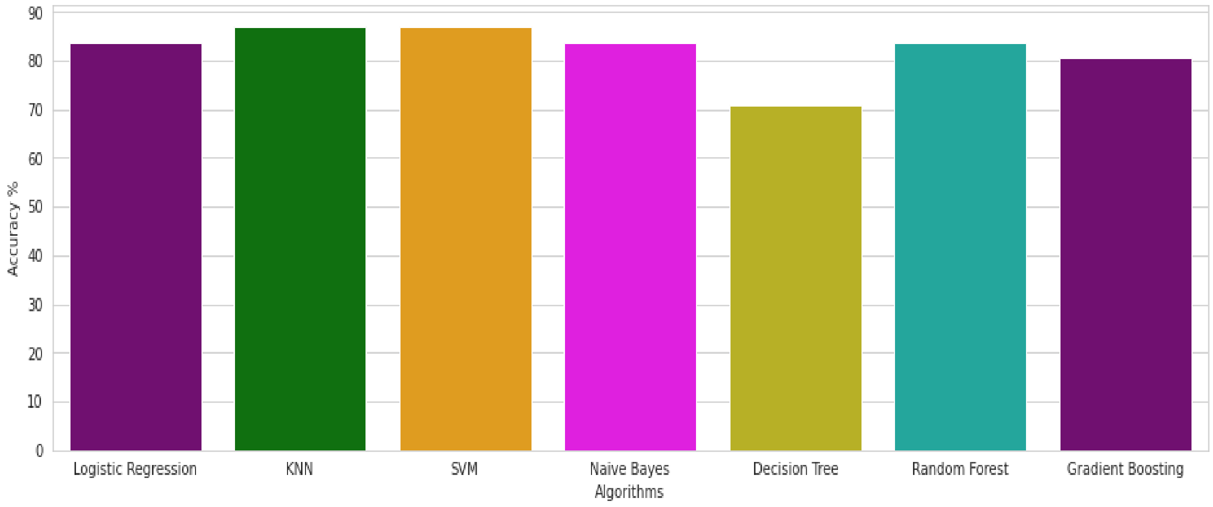
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
age	1	-0.094	0.097	0.27	0.22	0.12	0.13	-0.4	0.098	0.19	0.16	0.36	0.11	0.21
sex	-0.094	1	0.035	-0.063	-0.2	0.042	0.039	-0.076	0.18	0.097	0.051	0.087	0.39	0.3
cp	0.097	0.035	1	-0.043	0.09	-0.099	0.074	-0.32	0.35	0.17	0.14	0.23	0.26	0.42
trestbps	0.27	-0.063	-0.043	1	0.17	0.16	0.12	-0.039	0.083	0.22	0.14	0.086	0.13	0.16
chol	0.22	-0.2	0.09	0.17	1	0.025	0.17	-0.019	0.078	0.028	-0.0058	0.13	0.029	0.12
fbs	0.12	0.042	-0.099	0.16	0.025	1	0.053	0.022	-0.0041	-0.026	0.044	0.12	0.049	-0.016
restecg	0.13	0.039	0.074	0.12	0.17	0.053	1	-0.075	0.095	0.12	0.16	0.11	0.0073	0.18
thalach	-0.4	-0.076	-0.32	-0.039	-0.019	0.022	-0.075	1	-0.38	-0.35	-0.39	-0.27	-0.25	-0.42
exang	0.098	0.18	0.35	0.083	0.078	-0.0041	0.095	-0.38	1	0.27	0.26	0.15	0.32	0.42
oldpeak	0.19	0.097	0.17	0.22	0.028	-0.026	0.12	-0.35	0.27	1	0.61	0.26	0.32	0.42
slope	0.16	0.051	0.14	0.14	-0.0058	0.044	0.16	-0.39	0.26	0.61	1	0.11	0.28	0.34
ca	0.36	0.087	0.23	0.086	0.13	0.12	0.11	-0.27	0.15	0.26	0.11	1	0.26	0.46
thal	0.11	0.39	0.26	0.13	0.029	0.049	0.0073	-0.25	0.32	0.32	0.28	0.26	1	0.53
target	0.21	0.3	0.42	0.16	0.12	-0.016	0.18	-0.42	0.42	0.42	0.34	0.46	0.53	1

Figure CORRELATION MATRIX VISUALIZATION

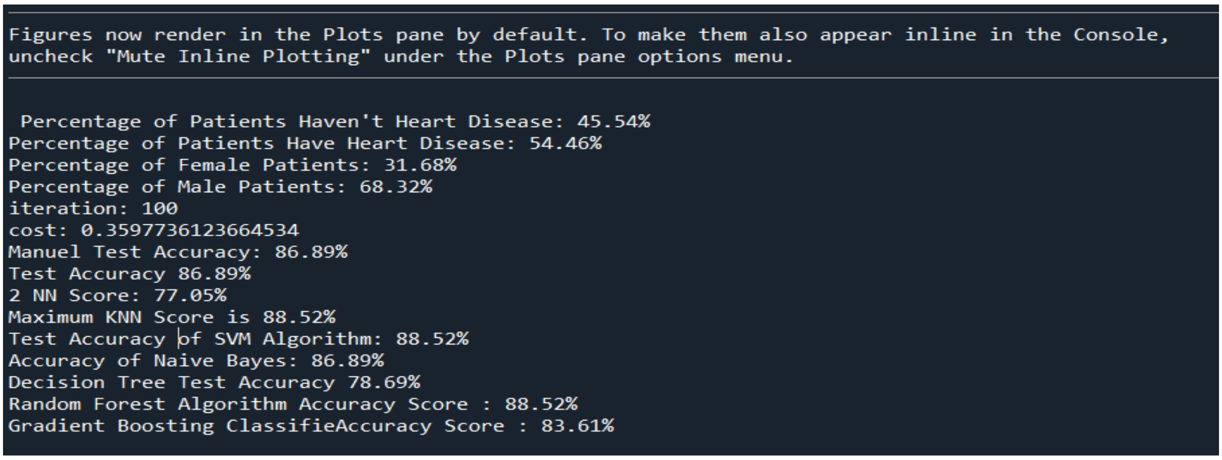
| DOI: 10.15680/IJIRCCE.2021. 0903091 |



CONFUSION MATRIX



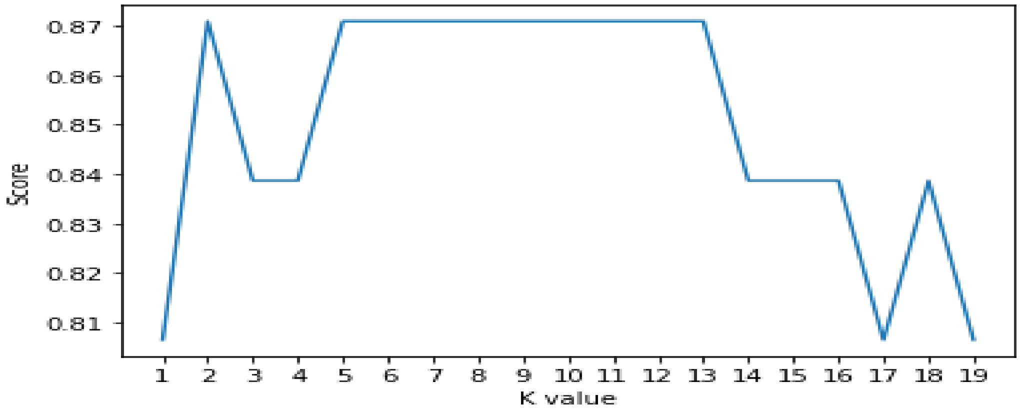
ACCURACY DETAILS



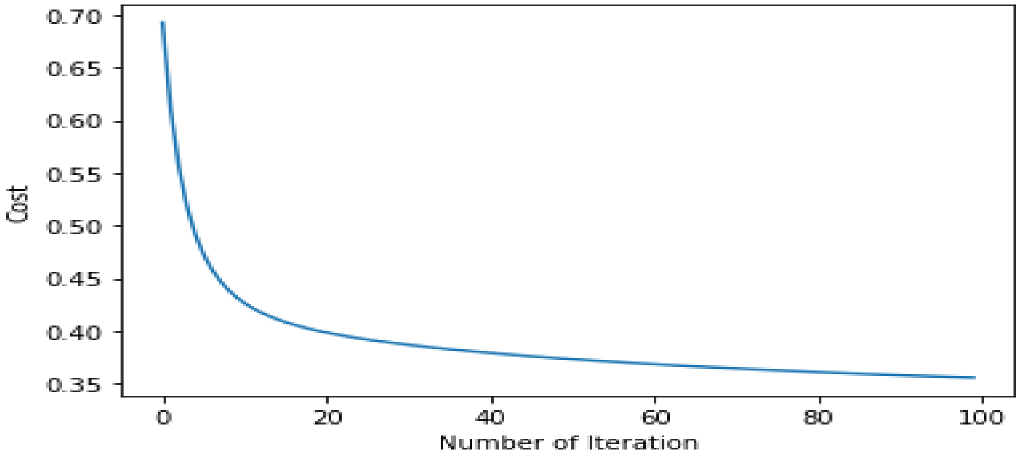
```
Figures now render in the Plots pane by default. To make them also appear inline in the Console,
uncheck "Mute Inline Plotting" under the Plots pane options menu.

 Percentage of Patients Haven't Heart Disease: 45.54%
Percentage of Patients Have Heart Disease: 54.46%
Percentage of Female Patients: 31.68%
Percentage of Male Patients: 68.32%
iteration: 100
cost: 0.3597736123664534
Manuel Test Accuracy: 86.89%
Test Accuracy 86.89%
2 NN Score: 77.05%
Maximum KNN Score is 88.52%
Test Accuracy of SVM Algorithm: 88.52%
Accuracy of Naive Bayes: 86.89%
Decision Tree Test Accuracy 78.69%
Random Forest Algorithm Accuracy Score : 88.52%
Gradient Boosting ClassifieAccuracy Score : 83.61%
```

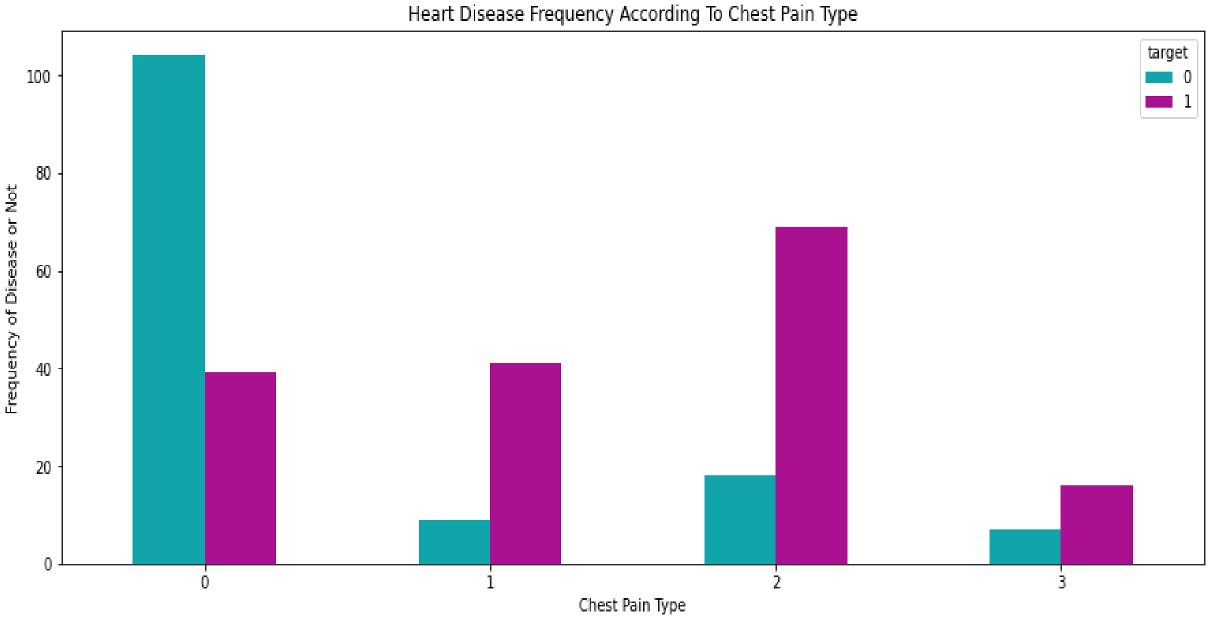
FINAL OUTPUT



KVALUE VS SCORE



NUMBER OF ITERATION VS COST



CHEST PAIN TYPE VS FREQUENCY OF DISEASE OR NOT

VI. CONCLUSION AND FUTURE WORK

In this study, an efficient machine learning based diagnosis system has been developed for the diagnosis of heart disease. Machine learning classifiers include LR, SVM, NB, and GB is used in the designing of the system. Four standard feature selection algorithms including proposed a novel feature selection algorithm FCMIM used to solve feature selection problem the early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes in high risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine. This project resolved the feature selection FCMIM behind the models and successfully predicts the h eart disease, with 85% accuracy. Further for its enhancement, we can train on models and predict the types of cardiovascular diseases providing recommendations to the users, and also use more enhanced models

REFERENCES

[1] A. L. Bui, T. B. Horwich, and G. C. Fonarow, ``Epidemiology and risk profile of heart failure,'' Nature Rev. Cardiol., vol. 8, no. 1, p. 30, 2011.

[2] M. Durairaj and N. Ramasamy, ``A comparison of the perceptive approaches forpreprocessing the data set for predicting fertility success rate,'' Int. J. Control Theory Appl., vol. 9, no. 27, pp. 255_260, 2016.

[3] L. A. Allen, L.W. Stevenson, K. L. Grady, N. E. Goldstein, D. D. Matlock, R. M. Arnold, N. R. Cook, G. M. Felker, G. S. Francis, P. J. Hauptman, E. P. Havranek, H. M. Krumholz, D. Mancini, B. Riegel, and J. A. Spertus, ``Decision making in advanceNUd heart failure: A scientific statement from the American heart association,'' Circulation, vol. 125, no. 15, pp. 1928_1952, 2012.

[4] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, ``Innovative artificial neural networks- based decision support system for heart diseases diagnosis,'' J. Intell. Learn. Syst. Appl., vol. 5, no. 3, 2013, Art. no. 35396.

[5] Q. K. Al-Shayea, ``Artificial neural networks in medical diagnosis,'' Int. J. Comput. Sci.Issues, vol. 8, no. 2, pp. 150_154, 2011.

[6] J. Lopez-Sendon, ``The heart failure epidemic,'' Medicographia, vol. 33, no. 4, pp. 363_369, 2011.

[7] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, ``Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quanti_cationof average Parkinson's disease symptom severity,'' J. Roy. Soc. Interface, vol. 8, no. 59, pp. 842_855, 2011.

[8] 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. 10091015, 2019.

[9] S. Nazir, S. Shahzad, S. Mahfooz, and M. Nazir, ``Fuzzy logic based decision support system for component security evaluation,'' Int. Arab J. Inf. Technol., vol. 15, no. 2, pp. 224231, 2018.

[10] R. Detrano, A. Janosi,W. Steinbrunn, M. Psterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, ``International application of a new probability algorithm for the diagnosis of coronary artery disease,'' Amer. J. Cardiol., vol. 64, no. 5, pp. 304310, Aug. 1989.

[11] J. H. Gennari, P. Langley, and D. Fisher, ``Models of incremental concept formation,'' Artif. Intell., vol. 40, nos. 13, pp. 1161, Sep. 1989.

[12] Y. Li, T. Li, and H. Liu, ``Recent advances in feature selection and its applications,'' Knowl. Inf. Syst., vol. 53, no. 3, pp. 551577, Dec. 2017.

[13] J. Li and H. Liu, ``Challenges of feature selection for big data analytics,'' IEEE Intell. Syst., vol. 32, no. 2, pp. 915, Mar. 2017.

[14] L. Zhu, J. Shen, L. Xie, and Z. Cheng, ``Unsupervised topic hypergraph hashing for effcient mobile image retrieval,'' IEEE Trans. Cybern., vol. 47, no. 11, pp. 39413954, Nov. 2017.

[15] S. Raschka, ``Model evaluation, model selection, and algorithm selection in machinelearning,'' 2018, arXiv:1811.12808. [Online]. Available: http://arxiv.org/abs/1811.12808

[16] S. Palaniappan and R. Awang, ``Intelligent heart disease prediction system using data mining techniques,'' in Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl., Mar. 2008, pp. 108115.

[17] E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, ``Heart diseases diagnosis using neural networks arbitration,'' Int. J. Intell. Syst. Appl., vol. 7, no. 12, p. 72, 2015.

[18] R. Das, I. Turkoglu, and A. Sengur, ``Effective diagnosis of heart disease through neural networks ensembles,'' Expert Syst. Appl., vol. 36, no. 4, pp. 76757680, May 2009.

[19] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, ``An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction,'' Expert Syst. Appl., vol. 68, pp. 163172, Feb. 2017.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING