International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST2183218

Heart Disease Prediction using Machine Learning Techniques

Ramesh Ponnala*[1], K. Sai Sowjanya[2]

*[1]Assistant Professor, Department of MCA, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad, Telangana State, India

[2]MCA Scholar, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad, Telangana State, India

Article Info
Volume 8, Issue 4
Page Number : 42-47

Publication Issue
July-August-2021

Article History
Accepted : 01 July 2021
Published : 05 July 2021

ABSTRACT

Prediction of Cardiovascular ailment is an important task inside the vicinity of clinical facts evaluation. Machine learning knowledge of has been proven to be effective in helping in making selections and predicting from the huge amount of facts produced by using the healthcare enterprise. on this paper, we advocate a unique technique that pursuits via finding good sized functions by means of applying ML strategies ensuing in improving the accuracy inside the prediction of heart ailment. The severity of the heart disease is classified primarily based on diverse methods like KNN, choice timber and so on. The prediction version is added with special combos of capabilities and several known classification techniques. We produce a stronger performance level with an accuracy level of a 100% through the prediction version for heart ailment with the Hybrid Random forest area with a linear model (HRFLM).

Keywords : Cardiovascular Disease (CVD), Heart disease prediction, Machine learning, Hybrid ML Techniques, Classification, Prediction

I. INTRODUCTION

Health care field has vast amount of data, for processing those data various techniques are used. Heart disease is leading cause of death worldwide. We have also seen ML techniques getting utilized in recent developments in several areas of the online of Things. Various studies give only a glimpse into predicting heart condition with ML techniques. It is difficult to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol and many other factors. Various techniques in Data mining and Machine learning have been employed to find out the severity of heart disease among humans.

The perspective of life science and data processing are used for discovering various kinds of metabolic syndromes. The nature of heart condition is complex and hence, the disease must be handled carefully. Not doing so may affect the center or cause premature death. Prediction of disorder could also be a critical challenge within the world of clinical data analysis.

Machine learning has been shown to be effective in assisting in making decisions and predicting from the large quantity of data produced by the healthcare Industry. In this paper, we propose a novel method that aims by finding significant features by applying ML techniques resulting in improving the accuracy in the prediction of heart disease. The severity of the heart disease is classified based on various methods like KNN, Decision trees etc. The prediction model is introduced with different combinations of features and variety of other known classification techniques. The data of heart condition patients collected from the UCI laboratory is employed to get patterns with NN, DT, Support Vector machines SVM, and Naïve Bayes. The results are compared for performance and accuracy with these algorithms. We produce an enhanced performance level with an accuracy level of 100% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM). The organization of this paper is as follows: in Section 2 Overview of Method and Dataset, in Section 3 Result and Discussion, in Section 4 Evaluation of Results and in section 5 Conclusion.

II. OVERVIEW OF METHOD AND DATASET

In the Existing system, prediction of heart disease was on different Machine learning algorithms like Decision trees, Random forest and many more. Hence the accuracy was predicted by using Decision trees and Random forest in the present system.

Drawback: The accuracy which is predicted by using Decision trees has very less scores when compared to other algorithms.

In this machine we need to examine the accuracy which is predicted by the use of decision trees and other machine mastering algorithms. right here, coronary heart ailment was collected from UCI Ml repository. There are 4 databases (i.e. Cleveland, Hungary, Switzerland, and VA long seashore). The Cleveland database became selected for this research because it is a commonly used database for ML researcher with comprehensive and complete statistics[1]. The category regulations generated based on the guideline after information pre-processing is achieved. After pre-processing, the statistics' 3 first-class ML strategies are selected and the outcomes are generated. The various datasets with DT, RF, LM are implemented to find out the excellent classification method. The results display that RF and LM are the first-rate. The LM approach for the dataset is the first-rate (one hundred%) compared to DT and RF techniques. We integrate the RF technique with LM and advocate HRFLM method to improve the results.

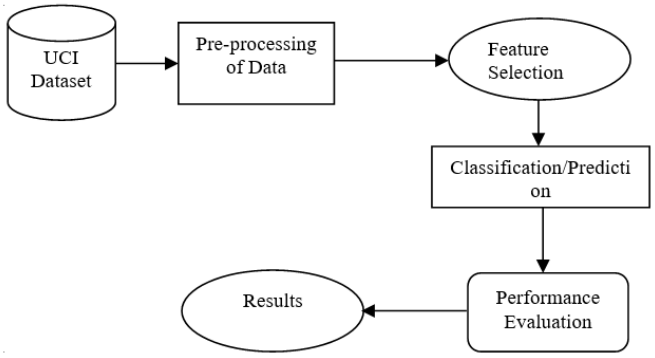

Figure 1 Architecture/Work Flow with UCI Dataset

In HRFLM, we use a computational technique with the three association rules of UCI Cleveland dataset. The available facts points to the deduction that women have less of a threat for heart sickness as compared to males. In coronary heart illnesses, correct diagnosis is number one. However, the traditional procedures are inadequate for correct prediction and prognosis.

HRFLM makes use of ANN with again propagation together with thirteen medical features because of the input. The obtained results are relatively analyzed against conventional strategies. The danger degrees turn out to be very high and a number of attributes are used for accuracy within the diagnosis of the disease the character and complexity of heart ailment require an mining specifically, apriori, predictive and

Tertius to discover the factors of heart disorder at the efficacious treatment plan. Statistics mining methods help in remedial situations within the scientific field. NN, SVM, and KNN. Amongst several employed strategies, the consequences from SVM prove to be beneficial in enhancing accuracy inside the prediction of disease. The nonlinear technique with a module for monitoring heart feature is added to detect the arrhythmias like bradycardia, tachycardia, atrial, atrial ventricular flutters, and many others. The overall performance efficiency of this method can be expected from the accuracy in the outcome effects based on ECG information. ANN training is used for the correct analysis of sickness and the prediction of viable abnormalities within the patient.

III. PROPOSED METHOD HRFLM

Random forests are a well-liked classification method supported an ensemble of one sort of decision tree. We propose a completely unique random forest algorithm, called a Hybrid random forest with linear model. Linear models describe endless response variable as a function of 1 or more predictor variables. They can help you understand and predict the behaviour of complex systems or analyse experimental, financial, and biological data.
We ensemble multiple sorts of decision trees into a random forest, and exploit diversity of the trees to reinforce the resulting model [1].

The results show that this method consistently outperforms when compared to these other methods.

Cleveland Dataset:

This Dataset contains 303 records. There are 76 attributes in total but here 13 attributes are considered for prediction of Heart disease, where only 1 attribute serves as the output.

The facts mining techniques are further used considering DT,

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num (the predicted attribute)
63	1	1	145	223	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	2
67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	3
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	2
53	1	4	140	203	1	2	155	1	3.1	3	0	7	1

Figure 2 UCI Cleveland Dataset attributes

A. Data Pre-processing, Feature Selection & Prediction:

Heart sickness statistics is pre-processed after collection of various statistics. The dataset includes a complete of 303 patient information, in which 6 statistics are with a few missing values. Those 6 records have been removed from the dataset and the staying 297 patient records are utilized in pre-processing.

The multi-elegance variable is used to test the presence or absence of heart disease. In the example of the affected person having coronary heart ailment, the value is about to at least 1, else the price is about to 0 indicating the absence of coronary heart disease in the affected person. The pre-processing of facts is done by way of converting scientific facts into prognosis values.

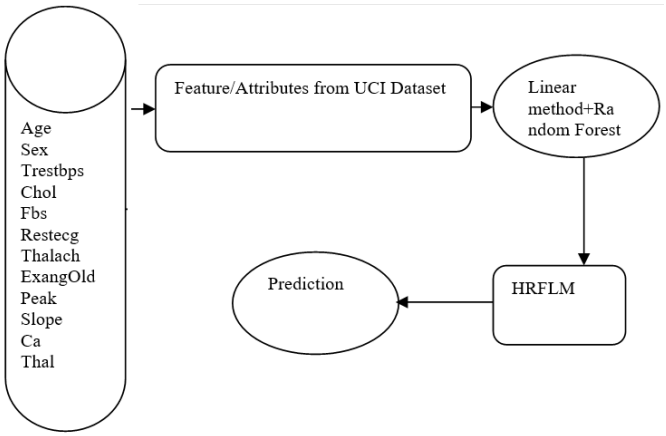

Figure 3 Workflow of predicting Heart Disease using HRFLM

B. Classification Modelling:

Clustering of datasets is done on the basis of the variables and criteria of selection Tree (DT), SVM capabilities. Then, the classifiers are carried out to every clustered dataset in order to estimate its overall performance. The first-class acting fashions are recognized from the above consequences primarily based on their low charge of error. The performance is in addition optimized through selecting the DT cluster with an excessive rate of blunders and extraction of its corresponding classifier functions. The overall performance of the classifier is evaluated for error optimization on this dataset.

Decision trees :

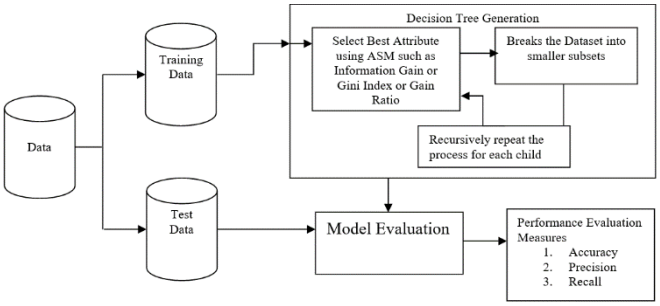

Figure 4 Decision trees Algorithm workflow

Algorithm 1 : Decision trees based partition

Tree-Learning(TR,Target,attr)

TR:Training examples
Target:target attribute
Attr: set of descriptive attributes
{
Create a Root node for the tree.
if TR have the same target attribute value ti,
Then return the single-node tree, ie. Root, with target attribute=ti
if Attr = empty (i.e., there is no descriptive attributes available),

Then return the single-node tree i.e., root with most common value of Target in TR
Otherwise {
Select attribute A from Attr that best classify TR based on an entropy-based measure
Set A the attribute for Root
For each legal value of A, vi, do
{
Add a branch below Root, corresponding to A=vi
Let TRvi be the subset of TR that have A=vi
If TRvi is empy,
Then add a leaf node below the branch with target value = most c common value of Target in TR0
Else below the brach, add the subtree learned by Tree-Learning(TRvi, Target,
Attr-{A})
}
}
Return Root }

Random Forest

Random Forest is a multifaceted machine learning system that performs regression and classification tasks. A group of real models combining to form a strong model is a type of ensemble learning process. In RF, in comparison to a single tree in the CART model, we rise many trees. Classification trees are decision trees which represent a significant advance in the exploration of information and the mining of data. The RF classifier makes high-level precision forecasts. Each tree has a gradation for classifying a new entity based on attributes, and we say the 'votes' tree. The forest selects the most common classification.

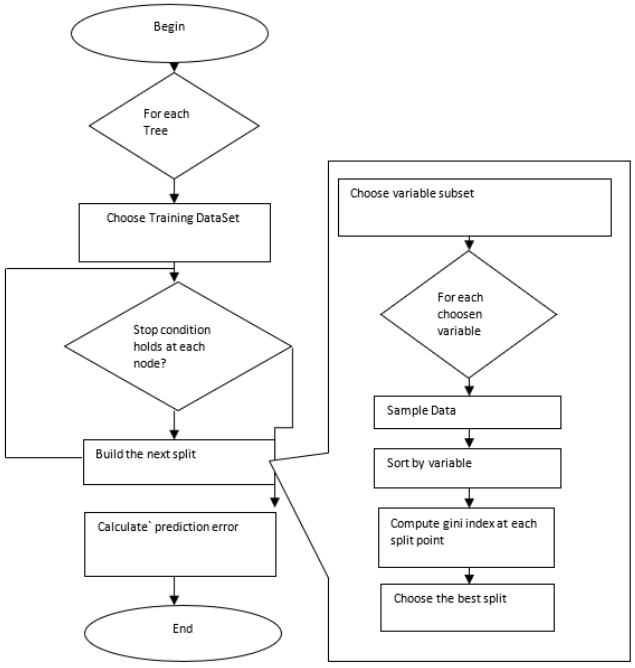

Figure 5 Random Forest Algorithm Workflow

Support Vector machine

SVM is a kind of supervised classification and regression problem Meiliana et al. [10]. It is also commonly used in problems of classification. This is used by the design of an N-dimensional hyper plane for correctly classifying unseen data. The key goal of SVM modeling is to find an optimum hyper plane dividing a cluster from a vector. Cases of a dependent variable form are often distinguished side by side from cases of a particular plane class. The word support vectors are used for all vectors near the hyper-airplane. SVM modelling finds a hyper plane directed to optimize the margins between the supporting vectors. SVM may use a kernel to partition the data in separate spaces and then use a hyper plane to support space to accommodate a non-linear separator between points Ramesh Ponnala et al.[14].

IV. EVALUATION OF RESULTS

```
[[33 11]
 [13 33]]
             precision    recall  f1-score   support

          0       0.71      0.80      0.76        46
          1       0.76      0.66      0.71        44

avg / total       0.74      0.73      0.73        90

0.7333333333333333
```

Figure 6 Decision Trees Accuracy Score

In the above result, Classification report is used to measure the quality of predictions from a classification algorithm called “Decision trees” and accuracy score is calculated.

```
Accuracy Score: 0.5
[[45  1]
 [44  0]]
             precision    recall  f1-score   support

          0       0.51      0.98      0.67        46
          1       0.00      0.00      0.00        44

avg / total       0.26      0.50      0.34        90
```

Figure7 Support Vector Machine Accuracy Score

In the above result, Classification report is used to measure the quality of predictions from a classification algorithm called “Support vector Machine” and accuracy score is calculated.

```
Accuracy Score: 0.7888888888888889
[[38  8]
 [11 33]]
             precision    recall  f1-score   support

          0       0.78      0.83      0.80        46
          1       0.80      0.75      0.78        44

avg / total       0.79      0.79      0.79        90
```

Figure 8 Hybrid Random Forest Linear Model Accuracy [HRFLM] Score.

In the above result, Classification report is used to measure the quality of predictions from a classification algorithm and a new model called “HRFLM” and accuracy score is calculated.

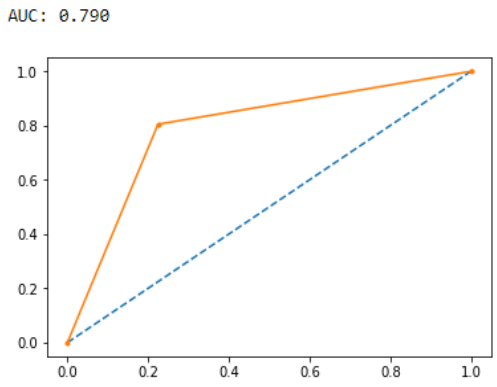

Figure 9 ROC Curve [HRFLM].

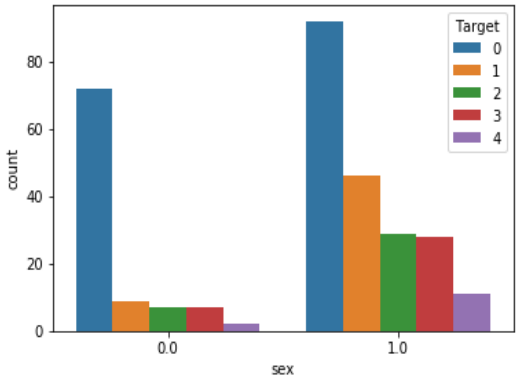

Figure 10 Count Plot

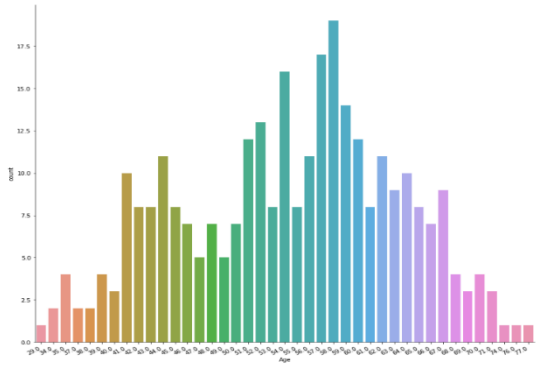

Figure 11 CAT plot

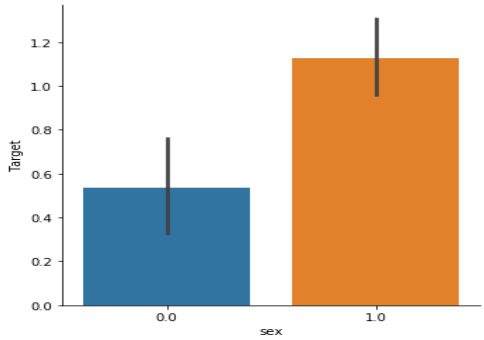

Figure12 CAT plot based on attributes

FINAL PREDICTION:

```
test = [[67.0,4.0,120.0,229.0,2.0,129.0,2.6,2,2]]
test = np.array(test)
result = RFC.predict(test)
result
if result[0]== 0:
    print('There is no Heart Disease')
elif result[0]==1:
    print('Disease with level 1')
else:
    print('Disease with level 2')
```

OUTPUT:

```
array([1], dtype=int64)
Disease with level 1
```

The above result predicts the heart disease based on certain feature values from dataset and finally result will be displayed in following 2 ways i.e. Disease with level-1 and Disease with level-2.

V. CONCLUSION

Predicting heart diseases is possible only by the consideration of attributes. This analysing method of the attributes can be achieved by the inclusion of data mining techniques. The overall objective of the work is to predict more exactly the occurrence of heart disease using Machine learning techniques. In this research work, the UCI data repository is used for performing the comparative analysis of these algorithms such as Random Forest, Decision trees linear model. From this research work, it has been experimentally proven that HRFLM (Hybrid Random Forest Linear Method) provides accurate results as compare to Decision trees. In this what we found is during small datasets in some other cases most of time decision trees direct us to a solution which is not accurate, but when we look at Hybrid Random Forest we are getting more accurate result.

VI. REFERENCES

[1]. Senthilkumar Mohan 1, Chandrasegar Thirumalai1, and Gautam Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", in IEEE Access, Special Section on Smart Caching, Communications, Computing and Cybersecurity for Information-Centric Internet of Things, Volume 7 June 2019, pp. 81542-81554

[2]. A. S. Abdullah and R. R. Rajalaxmi, "A data mining model for predicting the coronary heart disease using random forest classifier," in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22–25.

[3]. M. Durairaj and V. Revathi, ``Prediction of heart disease using back propagationMLPalgorithm,''Int. J. Sci. Technol. Res., vol. 4, no. 8, pp. 235_239,2015.

[4]. A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, ``Prediction of heart disease using machine learning,'' in Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA), Mar. 2018, pp. 1275_1278.

[5]. A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, "Using PSO algorithm for producing best rules in diagnosis of heart disease," in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306– 311. [3] N. Al-milli, "Backpropogation neural network for prediction of heart disease," J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.

[6]. C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, "Analysis of neural networks based heart disease prediction system," in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233–239.

[7]. P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules'. KingSaudUniv. - Comput.Inf.Sci., vol.24, no.1, pp.27–40, Jan.2012.doi: 10.1016/j.jksuci.2011.09.002.

[8]. L. Baccour, "Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets," Expert Syst. Appl., vol. 99, pp. 115–125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025.

[9]. C.-A. Cheng and H.-W. Chiu, "An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database," in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2017, pp. 2566–2569.

[10]. H. A. Esfahani and M. Ghazanfari, "Cardiovascular disease detection using a new ensemble classifier," in Proc. IEEE 4th Int. Conf. Knowl. Based Eng. Innov. (KBEI), Dec. 2017, pp. 1011–1014.

[11]. F. Dammak, L. Baccour, and A.M. Alimi, "The impact of criterion weights techniques in TOPSIS method of multi-criteria decision making in crisp and intuitionistic fuzzy domains," in Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE), vol. 9, Aug. 2015, pp. 1–8.

[12]. T. Mahboob, R. Irfan, and B. Ghaffar, "Evaluating ensemble prediction of coronary heart disease using receiver operating characteristics," in Proc. Internet Technol. Appl. (ITA), Sep. 2017, pp. 110–115.

[13]. Meiliana, S. Karim, H. L. H. S. Warnars, F. L. Gaol, E. Abdurachman, and B. Soewito, —Software metrics for fault prediction using machine learning approaches: A literature review with PROMISE repository dataset,‖ in 2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), 2017, pp. 19–23. doi:10.1109/CYBERNETICSCOM.2017. 8311708

[14]. Ramesh Ponnala, Dr. CRK Reddy, "Software Defect Prediction using Machine Learning

Algorithms :Current State of the Art", Solid State Technology, Volume 64 Issue 2 2021

Cite this article as :

Ramesh Ponnala, K. Sai Sowjanya, "Heart Disease Prediction using Machine Learning Techniques", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 8 Issue 4, pp. 42-47, July-August 2021. Available at doi : https://doi.org/10.32628/IJSRST2183218
Journal URL : https://ijsrst.com/IJSRST2183218