Effective Heart Disease Prediction using Hybrid Machine Learning Techniques

[1]Dr. J Rajaram, professor

[2]Dr M Bal Raju, assistant Professor

[3] Bandam Naresh, Assistant Professor

Departent:CSE

Nagole institute of engineering and technology,hydra bad.

Abstract:

One of the leading causes of death in the modern world is heart disease. Clinical data analysis faces a significant problem when predicting cardiovascular disease. It has been demonstrated that machine learning is good in helping with decision-making and prediction from the enormous amount of information generated by the healthcare sector. Only a few studies have looked into using machine learning to predict cardiac disease. In this study, we suggest a unique approach to improve the precision of cardiovascular disease prediction by identifying key features using machine learning techniques. The prediction model is introduced together with several feature combinations and well-known classification methods. By using a Hybrid Random Forest with Linear Model as our heart disease prediction model, we achieve an improved performance level with an accuracy level of 88.7%. (HRFLM).

1. Introduction

Massive amounts of data are analyzed using techniques like classification, clustering, and association to predict or explain the data. This process is known as mining. Using decision trees, it is estimated how accurately events connected to heart disease occur. Records from the medical field, such as Left Bundle Branch Block (LBBB), Right

The following terms are used to evaluate the patient's specific state in relation to heart catastrophe (BII): bundle branch block (RBBB), atrial fibrillation (AFIB), normal sinus rhythm (NSR), sinus bradycardia (SBR), atrial utter (AFL), premature ventricular contraction (PVC), and second-degree block. A radial basis function network (RBFN) dataset is used for classification, with 70% of the data being used for training and 30% being used for classification.

Using the well-known Cleveland dataset for experimental validation, which can be found in a UCI machine learning repository. The implementation of the Neural Networks ML technique results in performance that is more precise and reliable. The suggested technique for heart disease prediction contains 13 characteristics.

The results show a better level of efficiency compared to the currently used approaches. Neural network models are used, which incorporate both future likelihoods and predicable ideals from a number of precursor techniques. The grouping will initially distinguish between traffic produced by IoT devices and traffic produced by non-IoT devices. Each IoT device is assigned to a certain IoT device class in the second step.

2. Literature Review

The fields directly relevant to this study have a wealth of related work. In the field of medicine, ANN has been developed to produce predictions with the maximum degree of accuracy [6]. An ANN is utilized with back propagation multilayer perception (MLP).

A heart illness forecast. The resulting findings are contrasted with those of other models that have been used in the same domain, and it is discovered that they are better [10]. NN, DT, SVM, and Naive Bayes are utilized to find patterns in the data of heart disease patients collected from the UCI laboratory. The performance and accuracy of the output using various

methods are compared. The suggested hybrid method competes with the other known methods, yielding results for F-measure of 86.8% [7]. Convolutional Neural Networks (CNN) classification without segmentation is introduced. In the training phase, this approach takes into account the heart cycles with varying start locations from the electrocardiogram (ECG) signals. In the patient's testing phase, CNN can provide features with a range of positions [22], [41]. The medical sector generates a lot of data, but it hasn't always been handled well. The novel methods discussed here make heart disease more easily and effectively predictable while lowering costs.

The many research approaches taken into consideration in this work for the classification and prediction of heart disease utilizing machine learning (ML) and deep learning (DL) techniques are extremely accurate in determining the effectiveness of these methods.

3. Implementation of the system

The suggested research uses libraries like scikit-learn, pandas, matplotlib, and others and is built in Python 3.6.4. The data includes binary classifications of cardiac disease. The hybrid model is combined with algorithms like random forests and decision trees.

Age, Gender, Chest Pain, Blood Pressure Level, Cholesterol, Fasting Blood Sugar, Resting Electrocardiographic Results, Maximum Heart Rate Attained, Exercise Induced Angina, Old Peak, Slop, Number of Major Vessels, Thal, and Pred variables were all included in the dataset.

Project is to determine if a patient should be diagnosed with heart disease or not which is a binary outcome (0 – Heart disease doesn’t exists, 1-Heart disease exists).

The pre-processing of data is carried out by converting medical records into diagnosis values.

The proposed hybrid HRFLM approach is used combining the characteristics of Random Forest (RF) and Linear Method (LM). HRFLM proved to be quite accurate in the prediction of heart disease.

Classification issues are resolved using a Decision Tree (DT), a Supervised Learning Algorithm.

"Support is provided for both categorical and continuous input and output variables." The sample is divided into further consistent sets in accordance with the most significant splitter or discriminator in the input variables. This approach.

An evaluation of the attribute is represented by the internal node, the outcome is represented by the arm, and the conclusion is represented by the leaf in the decision tree.

When estimating a class grade for a record using decision trees, start at the tree's base. Next, the values of the record's attribute are contrasted with those of the root attribute.

Random Forest Model

1. Assume the training dataset contains n instances. Sub-samples are selected at random with substitution from these n instances. Individual trees are built using these random subsamples from the training dataset.
2. Given that there are k input variables, a number m is chosen such that m k. At each node, m variables are chosen at random from a pool of k variables. To split the node, the split that is the best of these m variables is chosen. While the forest grows, the value of m remains constant n Figure 4.
3. Each tree is allowed to grow to its full potential without being pruned. The new object's class is predicted based on the majority of votes earned from all of the decision trees combined.

HYBRID MODEL

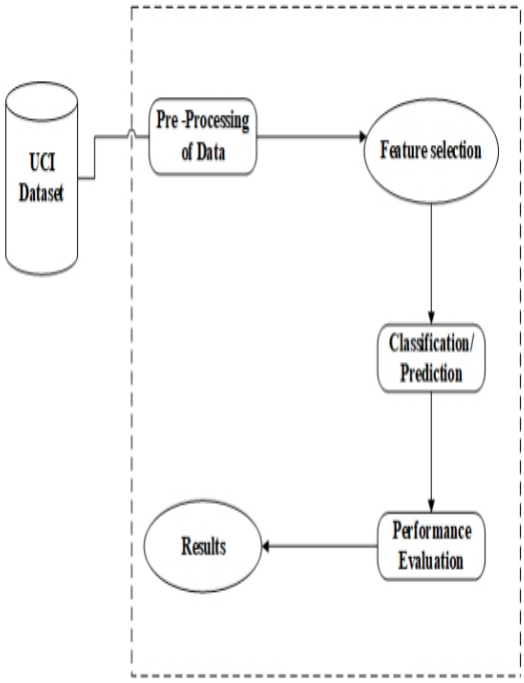
To construct a hybrid methodology, random forest probabilities are used in the combined model. The training data is combined with the random forest probabilities and cater into the decision tree model. In

a similar way, decision tree prospect is frequently described and cater to test data

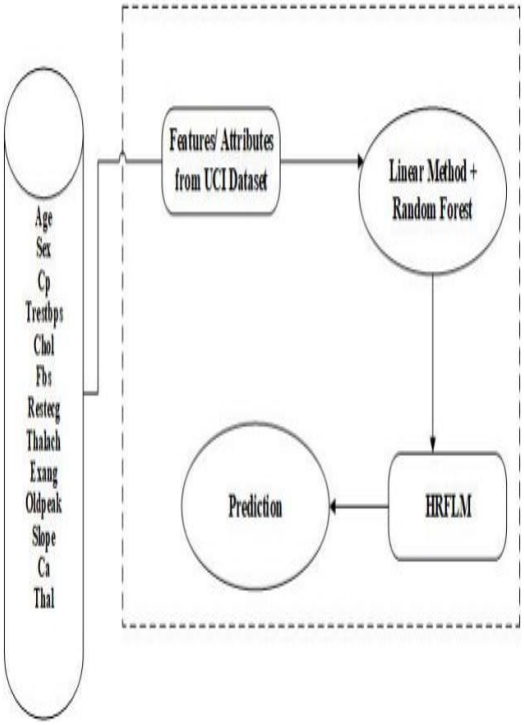
DATA PRE-PROCESSING OF DATA

After gathering multiple records, preprocessing of heart disease data occurs. There are 303 patient records in the dataset overall.

Where there are some missing data in 6 entries. The remaining 297 patient records are utilized for pre-processing after those 6 records were eliminated from the dataset. For the properties of the supplied dataset, multiclass variables and binary classification are introduced. The presence or absence of cardiac disease is determined using the multi-class variable. If the patient has heart disease, the value is set to 1, otherwise it is set to 0 to indicate that the patient is heart disease-free. Data pre-processing is carried out.



Experiment workflow with UCI dataset



Prediction of heart disease with HRFLM

Datasets

The UCI machine learning repository was used to gather information on heart illness. Four databases (including Cleveland, Switzerland, Hungary, and the VA Long Beach). Because it is a popular resource for ML researchers and has extensive and full information, the Cleveland database was chosen for this study. There are 303 entries in the collection. Although there are 76 attributes in the Cleveland dataset, only 14 of them are covered by the data set that is available in the repository. The Cleveland Clinic Foundation is the dataset's data source. The description and kind of properties are shown in Table 1. One characteristic acts as the output or the projected attribute to the existence of heart disease in a patient out of the 13 attributes that are included in the prediction of heart disease.

4. Results

The accuracy is calculated with the number of feature selection and the model generated results. HRFLM has no restriction in selecting of features to use. All the features selected in this model accomplish the best results

	Overall error rate			Best Model	Overall classification error rat		
Data Split	DT	RF	LM		DT	RF	LM
1	14.9	4	6.7	RF	14.9	14.9	16.2
2	34.9	12.2	22.6	RF	39.6	37.7	38.7
3	50	11.1	16.6	RF	50	27.8	50
4	62.5	20.9	29.2	RF	62.5	54.1	54.2
5	60	13.3	13.3	RF/ LM	60	53.4	53.3
6	54.6	12	18.1	RF	60.6	57.6	54.6
7	57.1	0	28.5	RF	57.1	28.5	42.8
8	36.4	18.2	9.1	LM	36.4	27.3	27.3

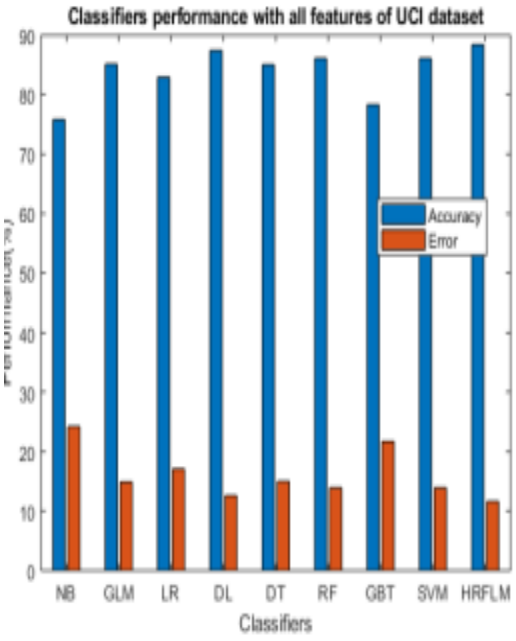
RESULT OF VARIOUS MODELS WITH PROPOSED MODEL

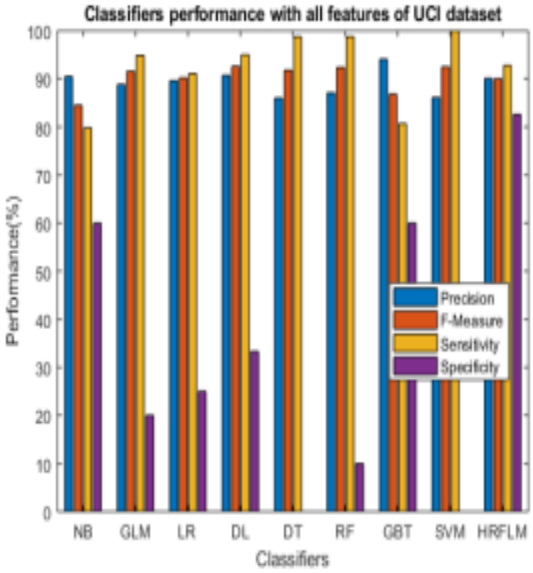
Sensitivity = (TP/TP+FN) =155/155+12 = 92.8

Specificity = (TN/TN+FP) =105/105+22 = 82.6

Precision = TP /TP+FP = 155/155+22==87.5

F-Measure= 2TP/ 2TP+FP+FN =310 /310+22+12 = 0.90





Performance comparison with various models

5. Conclusion

Identifying the heart's raw healthcare data processing Information is essential for the long-term preservation of human life. As well as the early identification of aberrant cardiac diseases. In this work, machine

learning techniques were utilized to process raw data and offer a fresh and innovative perspective on heart illness predicting heart disease is difficult, and highly significant in the medical industry. But the mortality rate rate might be significantly reduced if the illness is discovered Preventive actions are implemented as early as possible. Whenever it's feasible. Extension of this research is highly recommended. It is preferable to use real-world datasets for the research. Instead of only using simulations and theoretical techniques. The With the suggested hybrid HRFLM method, features of the Linear Method and Random Forest (RF) (LM). In the forecast of, HRFLM was highly successful. Heart illness this research's next directions can be carried out using a variety of machine learning approach combinations. To improve prediction methods. Additionally, new features election techniques may be created to obtain a more comprehensive understanding of the crucial features and boost the effectiveness of heart disease forecast.

REFERENCES

[1]. [1] Abdullah, A.S., 2012. A Data mining Model for predicting the Coronary Heart Disease using Random Forest Classifier. , (Icon3c), pp.22 −25.

[2]. [2] Alkeshuosh, A.H., Moghadam, M.Z. Mansoori, I. Al, 2017. Diagnosis of Heart Disease. , pp.306−311.

[3]. [3] Al-milli, N., 2013. Backpropogation Neural Network for Prediction of Heart Disease., 56(1), pp.131−135. [4] A. Devi, S. Rajamhoana, C, K. Umamaheswari, R. Kiruba, K. Karunya and R. Deepika, Analysis of Neural Networks Based Heart Disease Prediction System, 2018 11th International Conference on Human System Interaction (HSI), Gdansk, 2018, pp. 233-239.

[4]. [5] Anooj, P.K., 2012. Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. Journal of King Saud University - Computer and Information Sciences, 24(1), pp.27−40. Available at: http://dx.doi.org/10.1016/j.jksuci.2011.09.002.

[5]. [6] Baccour, L., 2018. Amende d fuse d TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets R. Expert Systems with Applications, 99, pp.115−125. Available at: https://doi.org/10.1016/j.eswa.2018.01.025.

[6]. [7] Cheng, C. Chiu, H., 2017. An Artificial Neural Network Model for the Evaluation of Carotid Artery Stenting Prognosis Using a National Wide Database. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp.2566−2569.

[7]. [8] H. A. Esfahani and M. Ghazanfari,"Cardiovascular disease detection using a new ensemble classifier," 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), Tehran, 2017, pp. 1011-1014.

[8]. [9] Dammak, F., Baccour, L. Alimi, A.M., The Impact of Criterion Weights Techniques in TOPSIS Method of Multi-Criteria Decision Making in Crisp and Intuitionistic Fuzzy Domains. 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), (9), pp.1−8.

[9]. [10] Das, R., Turkoglu, I. Sengur, A., 2009. Expert Systems with Applications Effective diagnosis of heart disease through neural networks ensembles. Expert Systems with Applications, 36(4), pp.7675−7680. Available at: http://dx.doi.org/10.1016/j.eswa.2008.09.013.

[10]. [11] Durairaj, M. Revathi, V., 2015. Prediction of Heart Disease Using Back Propagation MLP Algorithm. , 4(08), pp.235−239.

[11]. [12] Gandhi, M., 2015. Predictions in Heart Disease Using Techniques of Data Mining. 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp.520−525.

[12]. [13] Gavhane, A., 2018. Prediction of Heart Disease Using Machine Learning. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), (Iceca), pp.1275−1278. [14] Jpdlo, V. et al., 2018. Heart diseases prediction with Data Mining and Neural Network Techniques. , 6(7 2), pp.1−6.

[13]. [15] K, S.B.N., 2016. Prediction of Heart Disease at early stage using Data Mining and Big Data Analytics: A Survey. , pp.256−261.