ZKG INTERNATIONAL

ISSN: 2366-1313

Classification of Heart Beat disease with a Model based para-diagram of Machine Learning

Mr. Naresh Alvala, Yadlapati VaaniSri, Balina Vennela, Boddei Aravind Sai Srinivas, Gundapaneni Sai Shivani, Ravuri Kowsik

[1]Associate Professor, CSE Department, GITHAM University, Rudraram, Hyderabad 502329,India.

[23456]UG GITHAM University, Rudraram, Hyderabad 502329,India.

Abstract:

The most serious problem facing the world today is heart disease. The area of clinical data analysis is facing a major challenge in the prediction of cardiovascular disease. Machine learning has proven to be a useful tool in making predictions and decisions from large amounts of data generated by hospitals and healthcare sectors. Recent developments in various areas of the Internet of Things have seen ML techniques used. A variety of studies have shown that ML techniques can be used to predict heart disease. This paper proposes a narrative approach that uses machine learning to identify significant features and improve accuracy in the prediction for cardiovascular disease. The combination of several features and classification techniques is used to create the prediction model. The prediction model for heart disease using the ann and rnn along has an improved performance level of 95% and then followed by extra tree classifier and random forest with 92%.

Keywords: Cardiovascular Disease Prediction, Machine Learning Techniques, Extra Tree Classifier, Gradient Boosting Classifier, XGB Classifier, Ada Boost classifier, Random Forest Classifier, Logistic Regressions, Decision Tree Classifier, ANN, RNN

1 Introduction:

Heart disease prediction is a popular concept that has had a significant impact on society's health. The Random forest algorithm is used to determine the age and heart rate. This project demonstrates how the Heart rate and condition are estimated using inputs like blood pressure, and other information provided by users to the system. RFA is a better option than other algorithms. It provides a better experience and accurate results. This is useful in early diagnosis of disease. It is also used to calculate the heart rate based upon the patient's health.

Data mining is a powerful tool for extracting valuable information from large amounts of data. Data mining is used in nearly every area of life, including engineering, medicine, and business. It allows you to examine the data and extract crucial information that will help you make decisions. Machine learning algorithms can be used to reduce error and predict the truth. Machine learning algorithms are needed to help medical

ISSN: 2366-1313

professionals analyze medical data and make accurate and precise diagnosis decisions.

The specific issues that need to be addressed in order to examine cardiac disease mischance are those related to behaviors. Patients will also be subject to extensive exams, including blood pressure, glucose and vital signs, chest pain and electrocardiograms. However, the good news is that treatment can be achieved if the disease is detected early and prevented from becoming serious. Treatment for all cardiac patients depends on the clinical studies and patient history as well as the answers to the questions [4]. These techniques (history analysis and physical examination research, as well as medical professional evaluations) can often lead to incorrect diagnosis and delay in the testing. It is also more costly and takes longer to complete the evaluations [5].

It is difficult to determine the likelihood of developing cardiac disease by hand. A variety of data mining and machine learning techniques have been developed recently to address difficult problems [6, 7]. Advanced machine learning will allow us to recognize patterns and provide useful information. Machine learning has many uses, but it is most commonly used to predict the occurrence of heart disease. Many researchers are interested in machine learning to diagnose diseases. It reduces diagnostic time, is accurate and efficient. Machine learning can identify many diseases, however, heart disease diagnosis is the primary objective of this article. Heart disease is the leading cause for death today, and it is extremely helpful in saving lives [8].

Machine learning (ML), plays an important role in disease prediction [9]. Based on a simple learning algorithm, it can predict whether a patient is suffering from a specific disease type [7-10]. We use supervised learning techniques to predict the early stages of heart disease. To classify the individuals tested for heart disease and healthy, we use ensemble algorithms as well as several algorithms like k-nearest neighbour (KNN), support vector (SVM), decision trees (DT), Naive Bayes NB (NB), random forest (RF), and others. To select the most important features in the dataset, we use two methods for feature extraction: principal component analysis (PCA) and linear discriminant analysis.

The remainder of the paper is structured as follows: Section 2, which describes the literature review on the current research in this area. Section 3 explains the proposed architecture. Section 4 presents experimental results and a comparison of classification techniques. The paper's conclusion is described in Section 5.

II Related Work

Many literature articles have been published on heart disease diagnosis using data mining and machine-learning techniques [11]. Reddyet al. Reddyet al. Results show that RF achieved the best performance. Al-Mousa and Atallah [13] used stochastic gradient descent, KNN, logistic regression, RF, and voting ensemble learning for the prediction of cardiac diseases. The highest accuracy was achieved by the voting ensemble learning model, which has a 91% rate of accuracy. Pillaiet al. Pillaiet al. RNN achieved the best accuracy while K-mean achieved the lowest

ISSN: 2366-1313

accuracy. To predict heart disease, Kannan and Vasanthi [15] used four machine-learning algorithms: LR and RF, SVM and stochastic gradient booster (SGB). The best accuracy for model prediction was 86.5% for LR. Raza [16] used an ensemble learning model, multilayer perception, LR and NB to classify the heart diseases. Ensemble learning is more effective at predicting cardiac disease than other algorithms, according to the results. To predict heart disease, Oo and Win [17] used feature selection (CFS), sequential minimal optimization (SMO), and feature subset selection (CFS). CFS-SMO achieved the highest accuracy of 86.96%. Nalluri et al. To improve prediction of heart disease, [18] used XGBoost (or LR) as two methods. The results showed that LR had a better accuracy of 85.68% than XGBoost which was 84.46%. Bhatet al. Bhatet et al. (19) proposed a model that combines a multilayer perceptron (MLP), with a backpropagation algorithm for diagnosing heart disease. The model was found to have a lower error rate and an 80.99% accuracy. Abushariah et al. To predict the onset of cardiac disease, Abushariah et al. ANFIS achieved the lowest accuracy at 75.93%, while ANN has an optimal accuracy of 87.04%. Hasanet et al. [21] used MLP with backpropagation, SVM and MLP to classify heart diseases. MLP was found to have 98% accuracy. Chen et al. Chen et al. The best accuracy was achieved by ANN at 80%, according to the results. Sonawane & Patil [23] used vector quantum algorithm neural network to predict the occurrence of heart disease. Sapra et al. Sapra et al. Gradient boosted tree had 84% accuracy, which was higher than other algorithms. Haq et al. [25] used seven machine learning algorithms: LR, ANN, KNN, NB, SVM, DT, and RF with three feature selections: minimal-redundancy-maximal-relevance (mRMR), Relief, and Shrinkage and Selection Operator (LASSO) to predict heart disease. The highest accuracy was achieved by LR with Relief at 89%, compared to the other methods. In this system, the input details are obtained from the patient. Then from the user inputs, using ML techniques heart disease is analyzed. Now, the obtained results are compared with the results of existing models within the same domain and found to be improved. The data of heart disease patients collected from the UCI laboratory is used to discover patterns with NN, DT, Support Vector machines SVM, and Naive Bayes. The results are compared for performance and accuracy with these algorithms. The proposed hybrid method returns results of 87% for F-measure, competing with the other existing methods. Prediction of cardiovascular disease results is not accurate. Data mining techniques does not help to provide effective decision making. Cannot handle enormous datasets for patient records.

III Methodology:

After evaluating the results from the existing methodologies, we have used python and pandas operations to perform heart disease classification for the data obtained from the UCI repository. It provides an easy-to-use visual representation of the dataset, working environment and building the predictive analytics. ML process starts from a preprocessing data phase followed by feature selection based on data cleaning, classification of modelling performance evaluation. Random forest technique is used to improve the accuracy of the result. Increased accuracy for effective heart disease diagnosis. Handles roughest(enormous)

ISSN: 2366-1313

amount of data using random forest algorithm and feature selection. Reduce the time complexity of doctors. Cost effective for patients.

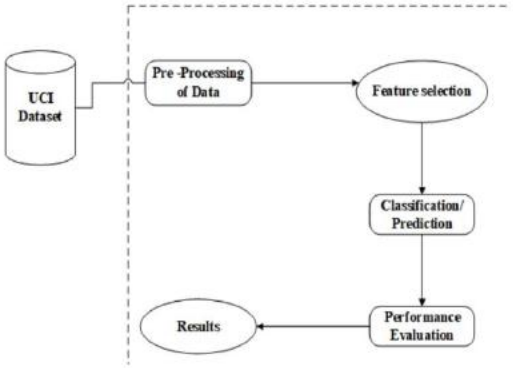

Figure 1 :Flow diagram

IV Proposed Method

Heart disease data is pre-processed by using various collection of records. The dataset contains a total of 303 patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 297 patient records are used in pre-processing. Among the 13 attributes of the data set, two attributes pertaining to age and sex are used to identify the personal information of the patient. The remaining attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of heart disease.

The clustering of datasets is done on the basis of the variables and criteria of Decision Tree (DT) features. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error.

AGE	Numeric [29 to 77;unique=41;mean=54.4;median=56]
SEX	Numeric [0 to 1;unique=2;mean=0.68;median=1]
CP	Numeric [1 to 4;unique=4;mean=3.16;median=3]
TESTBPS	Numeric [94 to 200;unique=50;mean=131.69;median=130]
CHOL	Numeric [126 to 564;unique=152;mean=246.69;median=241]
FBS	Numeric [0 to 1;unique=2;mean=0.15;median=0]
RESTECG	Numeric [0 to 2;unique=3;mean=0.99;median=1]
THALACH	Numeric [71 to 202;unique=91;mean=149.61;median=153]
EXANG	Numeric [0 to 1;unique=2;mean=0.33;median=0.00]
OLPEAK	Numeric [0 to 6.20;unique=40;mean=1.04;median=0.80]
SLOPE	Numeric [1 to 3;unique=3;mean=1.60;median=2]
CA	Categorical [5 levels]
THAL	Categorical [4 levels]
TARGET	Numeric [0.00 to 4.00;unique=5;mean=0.94;median=0.00]

For training samples of data D, the trees are constructed based on entropy inputs. These trees are simply constructed in a top down recursive divide and conquer (DAC) approach. Tree pruning is performed to remove the irrelevant samples on D.

Entropy = −Xm j=1 pij log2 pij

Algorithm for Decision Tree-Based Partition Require:

Input: D dataset – features with a target class

for ∀ features do

for Each sample

do Execute the Decision Tree algorithm

end for Identify the feature space f1, f2, . . . , fx of dataset UCI.

end for Obtain the total number of leaf nodes l1,l2,l3,. . . ,

ln with its constraints Split the dataset D into d1, d2, d3,. . . , dn

based on the leaf nodes constraints.

Output: Partition datasets d1, d2, d3,.

For given input features (xi, yi) with input vector xi of data D the linear form of solution f (x) = mx+b equation is solved by subsequent parameters:

m = P

This ensemble classifier builds several decision trees and incorporates them to get the best result.

ISSN: 2366-1313

For tree learning, it mainly applies bootstrap aggregating or bagging. For a given data, X = {x1, x2, x3, ..., xn} with responses Y = {x1, x2, x3, ..., xn} which repeats the bagging from b = 1 to B

Let the training samples having dataset Data = {yi, xi}; i = 1, 2, . . . , n where xi ∈ R n represent the i th vector and yi ∈ R n represent the target item. The linear SVM finds the optimal hyperplane of the form f (x) = w T x + b where w is a dimensional coefficient vector and b is a offset. This is done by solving the subsequent optimization problem:

Minw,b, ξi 1 2 w 2 + C Xn i=1 ξi s.t. yi , w T xi + b ≥ 1 − ξi, ξi ≥ 0, ∀i ∈ {1, 2, . . . ,}

	A	B	C	D	E	F	G	H	I	J
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	num
2	67	1	4	160	286	0	2	108	1	2
3	67	1	4	120	229	0	2	129	1	1
4	37	1	3	130	250	0	0	187	0	0
5	41	0	2	130	204	0	2	172	0	0
6	56	1	2	120	236	0	0	178	0	0
7	62	0	4	140	268	0	2	160	0	3
8	57	0	4	120	354	0	0	163	1	0
9	63	1	4	130	254	0	2	147	0	2
10	53	1	4	140	203	1	2	155	1	1
11	57	1	4	140	192	0	0	148	0	0
12	56	0	2	140	294	0	2	153	0	0
13	56	1	3	130	256	1	2	142	1	2
14	44	1	2	120	263	0	0	173	0	0
15	52	1	3	172	199	1	0	162	0	0
16	57	1	3	150	168	0	0	174	0	0
17	48	1	2	110	229	0	0	168	0	1
18	54	1	4	140	239	0	0	160	0	0
19	48	0	3	130	275	0	0	139	0	0
20	49	1	2	130	266	0	0	171	0	0
21	64	1	1	110	211	0	2	144	1	0
22	58	0	1	150	283	1	2	162	0	0
23	58	1	2	120	284	0	2	160	0	1

Figure 2 : Data set

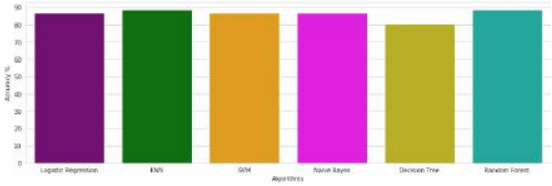

Figure 3 : Result comparison

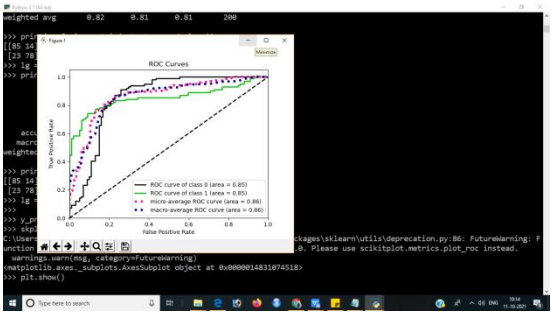

Figure 4 : ROC Curve under Regression model

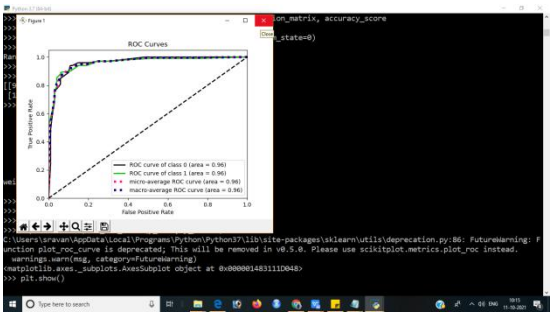

Figure 4 : ROC Curve under AdaBoost

Table Accuracy Comparison

Algorithms	Accuracy
Extra Tree Classifier	0.92
Gradient Boosting Classifier	0.825
XGB Classifier	0.915
Ada Boost Classifier	0.84
Random Forest Classifier	0.915
SVM	0.885
Decision Tree Classifier	0.80
Logistic Regression	0.815
ANN	0.93
RNN	0.95

ISSN: 2366-1313

Several standard performance metrics such as accuracy. precision and error in classification have been considered for the computation of performance efficiency of this model.

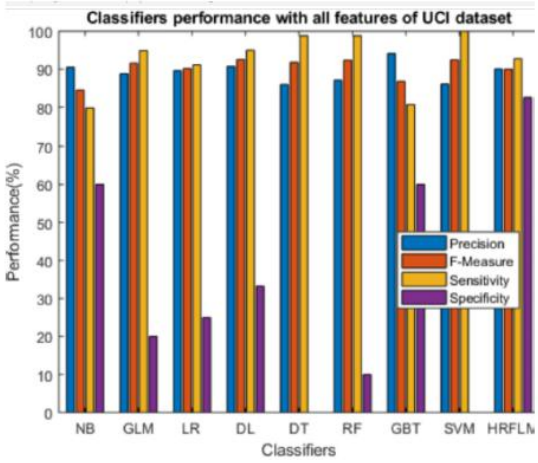

Figure 4 : Various models results

V Conclusion

In this paper, we proposed a method for heart disease prediction using machine learning techniques, these results showed a great accuracy standard for producing a better estimation result. By introducing new proposed Random forest classification, we find the problem of prediction rate without equipment and propose an approach to estimate the heart rate and condition. Sample results of heartrate are to be taken at different stages of the same subjects, we find the information from the above input via ML Techniques. Firstly, we introduced a support vector classifier based on datasets.

VI References

[1] WHO. The Top 10 Causes of Death. Accessed: Dec. 30, 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10causes-of-death

[2] C.Fryar,T.-C.Chen,andX.Li,‘‘Prevalenceofuncontrolledriskfactorsfor cardiovascular disease: United states, 1999-2010,’’ in NCHS Data Brief, vol. 103. Aug. 2012, pp. 1–8.

[3] MedicalProfessionals.CardiovascularDiseases.Accessed:Dec.29,2020. [Online]. Available: https://www.mayoclinic.org/medical-professionals/ cardiovascular-diseases

[4] 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,and E. P. Havranek, ‘‘Decision making in advanced heart failure: A scientific statement from the American heart association,’’ Circulation, vol. 125, p. E587, Apr. 2012.

[5] Q.K.Al-Shayea,‘‘Artificialneuralnetworksinmedicaldiagnosis,’’Int.J. Comput. Sci., vol. 8, no. 2, pp. 150–154, 2011.

[6] Centers for Disease Control and Prevention. Underlying Cause of Death 1999-2019. Accessed: Dec. 28, 2020. [Online]. Available: https: //wonder.cdc.gov/wonder/help/ucd.html

[7] S.S.Virani,A.Alonso,E.J.Benjamin,M.S.Bittencourt,C. W. Callaway, A. P. Carson, A. M. Chamberlain, A. R. Chang, S. Cheng, F. N. Delling, and L. Djousse, ‘‘Heart disease and stroke statistics—2020 update: A report from the American heart association,’’ Circulation, vol. 141, pp. E139–E596, Mar. 2020.

[8] E.J.Benjamin,P.Muntner,A.Alonso,M.S.Bittencourt,C.W.Callaway, A. P. Carson, A. M. Chamberlain, A. R. Chang, S. Cheng, S. R. Das, and F. N. Delling, ‘‘Heart disease and stroke statistics—2019 update: A report

ISSN: 2366-1313

fromtheAmericanheartassociation,''Circulation,vol. 139,pp.e56–e528, Mar. 2019.

[9] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, ''Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm,'' Comput. Methods Programs Biomed., vol. 141, pp. 19–26, Apr. 2017.

[10] R. Das, I. Turkoglu, and A. Sengur, ''Effective diagnosis of heart disease through neural networks ensembles,'' Expert Syst. Appl., vol. 36, no. 4, pp. 7675–7680, May 2009.

[11] B. Chapman, A. D. DeVore, R. J. Mentz, and M. Metra, ''Clinical profiles in acute heart failure: An urgent need for a new approach,'' ESC Heart Failure, vol. 6, no. 3, pp. 464–474, Jun. 2019.

[12] 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. 1009–1015, 2019.

[13] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, ''Can machine-learning improve cardiovascular risk prediction using routine clinical data?'' PLoS ONE, vol. 12, no. 4, 2017, Art. no. e0174944.

[14] M. M. A. Mary, ''Heart disease prediction using machine learning techniques: A survey,'' Int. J. Res. Appl. Sci. Eng. Technol., vol. 8, no. 10, pp. 441–447, Oct. 2020.

[15] T. Ahmad, A. Munir, S. H. Bhatti, M. Aftab, and M. A. Raza, ''Survival analysisofheartfailurepatients:Acasestudy,''PLoSONE,vol.12,no.7, Jul. 2017, Art. no. e0181001.