IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11112

Heart disease diagnosis using Hybrid machine learning algorithm

Devangam Sai Chaithanya[1], Kallupalli Lakshmi Narayana[2], Seshu Kotha[3], Kadiyala Surya Vardhan[4], Dr.Chandramma.R[5]

Department of computer science and Engineering , JAIN (Deemed-to-be university), Bangalore , India

Abstract: Heart is the a huge part in living creatures. Analysis and detection of heart related illnesses requires more accuracy, flawlessness and rightness on the grounds that a little slip-up can cause exhaustion issue or demise of the individual, there are various demise cases identified with heart and their number is expanding dramatically.. Predicting of Heart disease illness saves many lives recognizing Symptoms namely Raising in the heartbeat, Slow heartbeat ,Chest pain or discomfort ,Shortness of breath ,Light headache., Dizziness and so forth, is a basic challenge by the customary clinical information investigation. In this paper , we analysed the Machine Learning algorithms like K-KNN, NB,Decision Tree And Random Forest .and proposed a hybrid model which can predict the heart disease based on the basic symptoms like age, sex, pulse Rate etc. by comparing the accuracy we proven hybrid algorithm is the most accurate and reliable algorithm compared to all algorithms.

Keywords: K-Nearest Algorithm,Logistic Algorithm,Naïve Bayes ,Multi-Layer Perceptron,Machine Learning Algorithms.

I INTRODUCTION

Heart is one of the most sensitive and nessasary organs of the human body so the consideration of heart is more important. The vast majority of infections are identified with heartdiseases so the forecast about heart infections is important and for this reason relative review required in this field, today the greater part of patient are passed on in light of the fact that their illness are seen in the final stage on account of not appearing of accuracy of techniques so there is need to ponder the more compelling estimations for diseases.

Analysts attempt to run over a proficient procedure for the identification of coronary illness, as the current analysis strategies of coronary illness are very little compelling in right on time distinguishing proof because of a few reasons, like exactness what's more, execution time . The finding and treatment of heart sickness is amazingly troublesome when current innovation and clinical specialists are not accessible.As per the European Society of Cardiology, 26 million roughly individuals of HD 3.6 million were analyzed over every year . The vast majority of individuals in the United States are experiencing coronary illness . Determination of HD is generally done by the examination of the clinical history of the patient, actual assessment report and investigation of concerned indications by a physicianResearchers attempt to run over a proficient procedure for the identification of coronary illness, as the current conclusion strategies of coronary illness are very little viable in right on time.

II LITERATURE REVIEW

Mr.Santhana Krishnan.J,DR.Geeta.S [1] Proposed Prediction of Heart Disease Using Machine Learning Algorithms in which they proved that The Decision tree algorithm has predicted the coronary illness influenced individual with an accuracy level of 91% and Naïve Bayes classifier has expected coronary illness patient with a precision level of 87% they concluded that the decision tree is given more accuracy than Naïve Bayes classifier by using WEKA tool for analysis of UCI data set

Sanjay KumarSen,[2] author proposed "Predicting and Diagnosing of Heart Disease using Machine Learning Algorithms," in which he compared different ml algorithms applied common ml classifiers like Naïve Bayes, SVM, Decision Tree, and K-Nearest Neighbor in predicting and diagnosing of coronary illness. The tests were done by utilizing 14 info components of UCI Machine Learning Repository data set by means of the weka tool. The author presumed that Naive bayes algorithm performed better

DOI: 10.17148/IJARCCE.2022.11112

[4]proposed A Survey on Heart Attack PredictioNUsing

Machine Learning in which author compared two algorithms namely Random Forest ,K-Nearest Neighbours with the kaggle data set author found that the knn is showing more accuracy than random forest according to him Random Forest; got about 86.89% of accuracy with 200, 500 & 1000 estimators and K-Nearest Neighbours got 91.8% of accuracy with 8 neighbours

ArchanaSingh,RakeshKumar[3]developed Heart Disease Prediction Using k-nearest neighbor, decision tree, linear regression and support vector machine(SVM) algorithms by using UCI repository dataset in jupyter notebook with python programming author obtained greater accuracy with knn when compared to decision tree, linear regression and support vector machine(SVM) I.e 87%

P.Sai Chandrasekhar Reddy, [5]This author proposed Heart disease prediction using ANN's algorithms. With the rapid increase in the expenses of heart disease diagnosis, there is an availability to develop a new system which can predict heart disease. The Prediction of this disease is used to predict symptoms of the patient after checking on the basis of various parameters like pulse rate, BP(Blood pressure), cholesterol and so on. The accuracy of this algorithm has been proved in paper.

Binhu wanh [6],In this research paper , the authors proposed a technique of deep wide neural networks namely deep and wide neural networks to predict the Heart failures diseases. There are other deep neural networks like DWNN, which is a deep neural network model. They used these algorithms to get the accuracy of the heart disease prediction of more than the other algorithms .Their future preparation will include more information in EHR into their framework and improving the model of the deep wide neural networks learning techniques , these techniques results to further improve the accuracy performance of Prediction.

Anchana Khemphila , Veera Boonjing [7],By this paper ,we came to know that the authors used a Back-Propagation algorithm and a Feature selection algorithms using a classification approach called MLP(Multi-Layer Perceptron) to Diagnose Heart Diseases Their Paperwork is to trying to introduce the Multi-Layer Perceptron(MLP) with the algorithms like feature selection , Back-propagation algorithms. By taking the information in the different data training sets with Heart Failure Patients.The accuracy ranges between 13 and 8 features of data set is 1.2% with the data validation set of accuracy 0.83% .

Kavitha ,G.Gnaneswar, R.Dinesh, Y.Rohith Sai,R.Sai Suraj ,[8] proposed work, a novel machine learning approach is proposed to predict heart disease. These studies used a heart disease dataset, and data mining techniques such as linear regression and classification are used. Machine learning techniques Random Forest and Decision Tree are used. The techniques of the machine learning Algorithms are designed. With the implementation, three machine learning algorithms are used namely Random Forest,Decision Tree and Hybrid model (Hybrid of random forest and decision tree). Experimental results shows an accuracy level of 88.7% through the heart disease prediction model with the hybrid model.

CHUNYAN GUO , JIABING ZHANG[9] analysed Machine Learning Algorithms like random forest tree,Decision Tree,KNN and they found Accurate outputs with Recursion enhanced Random forest with an improved linear Model, with an Accuracy of 92% and 70.1% of Validation Stability ratio.

YUANYUAN PAN , MINGHUAN FU[10] Proposed Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform. By using EDCNN,artificial neural network,Deep Neural Networks and Recurrent Neural Network.Based on the analysis, EDCNN hyperparameters can achieve a precision of up to 99.1 % and efficiency of 95.4%.

AN XIAO , YI LI, AND YIMIN JIANG[11] Proposed "Heart Coronary Artery Segmentation and Disease Risk Warning Based on a Deep Learning Algorithm".This paper is based on an improved three-dimensional U-net convolutional neural network deep learning algorithm for heart coronary artery segmentation for disease risk prediction. U-net Network Architecture is Used in this Paper and The design of U-net is based on the fully convolutional neural network FCN.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11112

GIHUN JOO1 , YEONGJIN SONG1 , HYEONSEUNG IM 1 , AND JUN BEOM PARK 2 [12] Proposed a Machine Learning Algorithms like Random Forests, Logistic Regression and they found output with and they found output with Prediction Performance which is shown nearly 20% higher performance.

SYED ARSLAN ALI 1, BASIT RAZA 1 [13] Proposed a Ruzzo-Tompa and Stacked Genetic Algorithm for Heart Disease Prediction on the internet medical platform. By using this the have improved the precision of 93% and efficiency of 96%

III DIFFERENT CLASSIFIERS:

I..KNN Algorithm

The k-nearest Neighbors (KNN) algorithm is a simple,supervised machine learning algorithm that can be utilized to solve both classification and regression problems It's not difficult to implement and understand, however has a major drawback it becomes slow when size of dataset increases

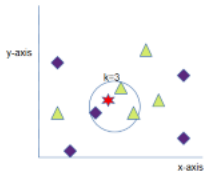

Fig-1

The k-nn algorithm stores all the data available and based on the similarity it classifies the new data points . This means when new data appears. Then it can be easily classified into a well-suited category by using K- NN algorithm

II.Decision Tree :

Decision tree is one of the learning models. Decision tree is a supervised algorithm. It can operate on both dates like categorical and numerical data. Based on certain conditions it gives a categorical solution such Yes/No, True or false, 1 or 0 We partition the dataset into at least two sets utilizing this method. In the Decision tree, inner hubs address a test on the attributes, the branch depicts the result, and leaves are the choices created after ensuing handling.

decision Tree classificsation is as follow

I. Set the dataset's best element as the base of the tree

ii. Dataset is parted into test and train sets. Subsets ought to be made so that every subset contains data with the component characteristic like that.

iii. On every subset, the means above are rehashed until we get leaves in the tree. The expectation for a record of a class mark in the Decision tree will begin from the root. The qualities are contrasted and the accompanying record credits with the root credits. The comparing worth of the following hub to go shows up in this correlation.

Information Gain = Class Entropy - Entropy Attribute

To find Class Entropy: **(Pi+Ni)=-P/P+Nlog2P/P+N-N/N+Nlog2N/P+N**

Here => P, Possibilities of Yes(1)

=> N, Possibilities of No(0)

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11112

To find Entropy Attributes:

Entropy attribute =∑**Pi+Ni/P+N**

IV. MLP(MULTI-LAYER PERCEPTRON)

The neural organizations have shown to be the most famous and developing part of Machine Learning in later considers. Presently in the proposed framework we utilized the neural network Algorithm Multi-Layer Perceptron (MLP) to prepare and test the dataset.

Multi-Layer perceptron Algorithm is a regulated neural organization algorithm in which there will be two layers like information and yield layers.They require at least one for stowed away layers between these two layers. Each node in the input layer is connected to output nodes through the hidden layers. The connections between any two nodes are assigned weights to it and the resultant input is calculated using the following formula

$$Yin=\sum wixi \ (i=0.\ldots n)$$

Where xi is the ith input and wi is the corresponding weight.

The activation function is applied on the weighted input to get the output. The role of hidden layer is to connect the input and output layer and to process the data internally.

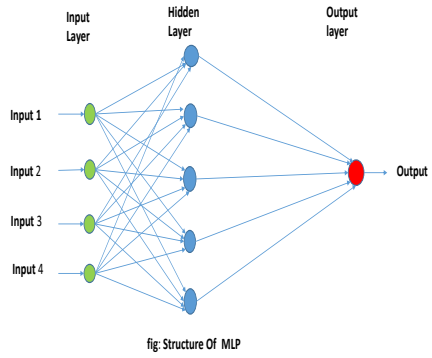

Fig-2

VI.RANDOM FOREST ALGORITHM

It is Supervised Learning Technique, It can be used for both Classification and Regression problems in ML. This algorithm is based on the concept of ensemble learning**,** it is a process of combining multiple classifiers which can solve a complex problem and also to improve the performance of the model.

It is a classifier which contains decision trees on various subsets of the dataset and it takes the average to improve the predictive accuracy of the dataset**.** Instead of depending on one decision tree, the random forest takes the prediction from each tree and based on the majority of predictions, it predicts the final output.

The Working procedure of Random Forest Tree

Step-1: In Random forest n number of random records are taken from the data set having k number of records..

Step-2: Decision trees are constructed Individually for each sample.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11112

Step-3: Output will be generated for every decision tree**.**

Step-4: Final output is considered based on Averaging for Classification and regression respectively..

Step-5: For the new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

VII NAÏVE BAYES CLASSIFICATION ALGORITHM:

It is a supervised learning algorithm, based on the Bayes theorem and used for solving classification problems It is majorly used in classification of text that includes a high-dimensional training dataset.The Bayes theorem is a rule or the mathematical concept that is used to get the probability is called Bayes theorem. Bayes theorem needs some independent presumption and it requires independent variables which is the principal assumption of Bayes theorem bayes theorem mathematical representation

P (A\B) = P(B/A)*P(A)/P(B)

P(A|B) is Posterior probability: Likelihood of hypothesis A on the event B.

P(B|A) is Likelihood probability: is the likelihood which is the probability of a predictor in given class.

P(A) is Prior Probability: Likelihood of hypothesis prior to noticing the proof.

P(B) is Marginal Probability: Probability of Evidence.

VIII. SUPPORT VECTOR MACHINE ALGORITHM

Support vector machines (SVMs) are powerful and flexible supervised machine learning algorithms which are used both for classification and regression. But generally, they are used in classification problems. In the 1960s, SVMs were first introduced but later they got refined in 1990. SVMs have their separate way of implementation as compared to other machine learning algorithms. SVM is very popular because of their ability to handle multiple continuous and categorical variables.

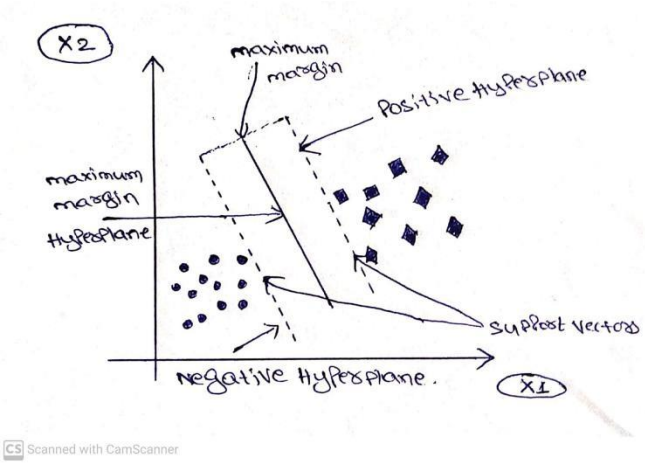

Fig-3

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11112

SVM chooses the extreme point vectors that help in creating the hyperplane. These kinds of extreme cases are known as support vectors, and. Consider the above diagram in which there are two different categories that are classified using a decision boundary or hyperplane

IV DATASET USED:

The dataset utilized mostly for predicting coronary illness is taken from UCI Machine learning repository. collection of databases that are utilized to execute ml algorithms. data set contains 14 attributes the detail explation about attributes are given in the table below

S. No.	Attribute	Desc.	Mean Value
1	age	in years	54.434
2	Sex	Male, Female	0.696
3	cp	Angina, abnang, notang, asympt	0.942
4	trstbps	Resting Blood Pressure in mm hg	131.612
5	chol	Serum Cholesterol in mg/dl	246
6	fbs	fasting blood sugar- 1 if >120 mg/dl, 0 if <120 mg/dl	0.149
7	restecg	Electrocardiographic Results	0.53
8	thalach	Maximum Heart Rate observed	149.114
9	exang	exercise with angina has occurred	0.337
10	oldpeak	ST depression induced through exercise	1.072
11	slope	slope of the ST segment	1.385
12	thal	Number of major vessels ranging from 0 - 3 color by fluoroscopy	0.754
13	ca	Heart status	2.34
14	Target	Output Class	

V METHODOLOGY:

We proposed a hybrid machine learning model which is the combination of random forest,SVM,MLP and Naive-bayes algorithm we attain the each model separately and with the voting classifier we get the final predicted out put of the model as shown in figure below:

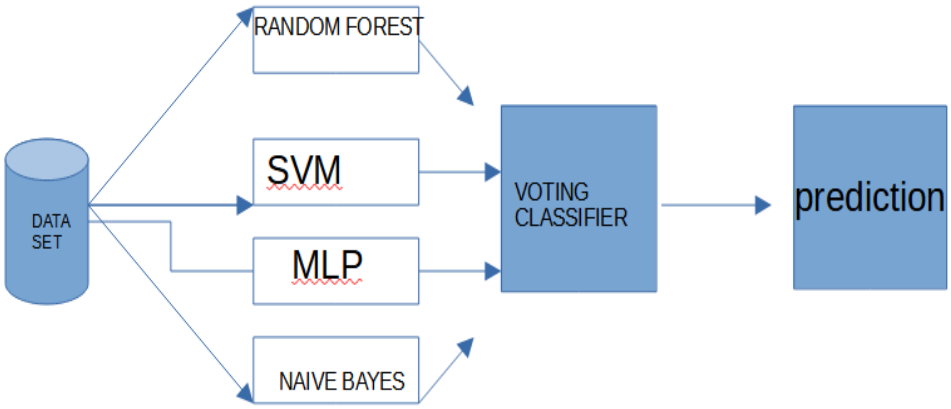

Fig-4

DISCUSSION:

By analysing All the Machine Learning Techniques namely K-Nearest Neighbours , Naïve Bayes, Support Vector Machine ,Decision Tree ,Random Forest ,MLP(Multi-Layer Perceptron) and our Hybrid Model , etc.. . in the jupiter note book by the predefined uci data set namely heart.csv data set which contains 303 elements and 14 features among them 242 are used for training the and the remaing 61 are used for testing and the accuracy is comapred By considering the

DOI: 10.17148/IJARCCE.2022.11112

accuracy we came to know that our hybrid ensemble model is giving us better accuracy than other machine learning models

CONCLUSION:

In this paper we analyzed the various ML-based prediction algorithm models to detect the Heart Diseases.. The Analysed algorithms are Random Forest, KNN, Decision Tree, ANN, MLP(MultiLayer Perceptron), Naive Bayes Classssifier. By analyzing these we have conclude that our hybrid algorithm is going to work better then the other machine learning models. So that we can use this hybrid model in the further research to detect the heart disease easily in the early stage more accurately than other models. This early stage detection may reduce death rate due to heart diseases by analysing the patients data . Hence with its precise accuracy in future it can replace the role of a doctor which is one of the project objective.

REFERENCES

[1]Mr.Santhana Krishnan.J ,Dr.Geetha.s "Prediction of Heart Disease Using Machine Learning Algorithms"in 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)

[2]Sanjay Kumar Sen, "Predicting and Diagnosing of Heart Disease using Machine Learning Algorithms," International Journal of Engineering and
Computer Science (IJECS), vol.6, pp.21623-21631, 2017.

[3]Archana Singh ,Rakesh Kumar " Heart Disease Prediction Using Machine Learning Algorithms " 2020 International Conference on Electrical and
Electronics Engineering (ICE3-2020)

[4]P.Santhi a , R.Ajayb ,D.Harshini c and S.S.Jamuna Sri d "A Survey on Heart Attack Prediction Using Machine Learning "Turkish Journal of
Computer and Mathematics Education : Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

[5]P.Sai Chandrasekhar Reddy "Multi Disease Prediction using Data Mining Techniques", International Journal of System and Software
Engineering, December 2016, pp.12-14

[6]BINHUA WANG 1,2,3, YONGYI BAI," A Multi-Task Neural Network Architecture for Heart disease Prediction in Heart Failure Patients"2019

Archana Khemphila and Veera Boonjing," Heart disease Classification using Neural Network and Feature Selection",21st International
Conference on Systems Engineering,2011.

[7]Kavitha ,G.Gnaneswar, R.Dinesh, Y.Rohith Sai,R.Sai Suraj," Heart Disease Prediction using Hybrid machine Learning Model", International
Conference on Inventive Computation Technologies,2021.

[8]CHUNYAN GUO , JIABING ZHANG , YANG LIU , YAYING XIE , ZHIQIANG HAN , AND JIANSHE YU "Recursion Enhanced Random Fores With an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform "Received March 2, 2020, accepted March 12, 2020, date of publication March 16, 2020, date of current version April 7, 2020

[9]YUANYUAN PAN , MINGHUAN FU, BIAO CHENG, XUEFEI TAO, AND JING GUO,"Enhanced Deep Learning Assisted ConvolutionalNeural Network for Heart Disease Prediction on the Internet of Medical Things Platform.Received July 24, 2020, accepted August 19, 2020, date of publication September 23, 2020, date of current version October 28, 2020

[10]AN XIAO , YI LI, AND YIMIN JIANG."Heart Coronary Artery Segmentation and Disease Risk Warning Based on a Deep Learning Algorithm".Received June 30, 2020, accepted July 12, 2020, date of publication July 21, 2020, date of current version August 11, 2020

[11]P. A. Wolf, R. D. Abbott, and W. B. Kannel, ''Atrial fibrillation as an independent risk factor for stroke: The framingham study,'' Stroke, vol. 22, no. 8, pp. 983–988, Aug. 1991.

[12]K. Polat, S. Şahan, and S. Güneş, ''Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour) based weighting preprocessing,'' Expert Syst. Appl., vol. 32, no. 2, pp. 625–631, Feb. 2007