Machine Learning Heart Disease Prediction Using KNN and RTC Algorithm

[1]Senthil G. A
Associate Professor,
Dept of Information Technology,
Agni College of Technology,
Chennai, India.
senthilga@gmail.com

[2]R. Prabha
Associate Professor,
Dept of Electronics and Communication Engineering,
Sri Sai Ram Institute of Technology,
Chennai, India.
r.praba05@gmail.com

[3]M Razmah
Assistant Professor,
Dept of Electronics and Communication Engineering,
Sri Sai Ram Institute of Technology,
Chennai, India.
razmah.ece@sairamit.edu.in

[4]T. Veeramakali
Associate Professor,
Dept of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology,
Chennai, India.
drveeramakali@gmail.com

[5]Sridevi S
Assistant Professor,
Department of Computer Science and Engineering,
Sri Ramachandra Engineering and Technology, Chennai, India.
sridevi@sret.edu.in

[6]Yashini R
Student, Dept of Electronics and Communication Engineering,
Sri Sai Ram Institute of Technology,
Chennai, India.
sit20ec082@sairamtap.edu.in

2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS) | 978-1-6654-6275-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICPECTS56089.2022.10047501

Abstract - Heart Disease or cardiovascular disease refers to the range of heart conditions like cardiac arrest, coronary artery disease. Heart disease can be very well hindered through certain lifestyle changes. There is a significant increase in the mortality rate recently due to the distinctive heart diseases. Machine learning uses mathematical models to work efficiently with the enormous amount of data. It plays a crucial role in medical science in the prediction of distinctive diseases. Cardiologists inspects the heart functionality using electrocardiography, computed tomography. These tests are quite expensive for a common man. Recent times, the life span of a human is guaranteed only with the support of medications. As prevention is better than the cure, machine learning helps to predict the vulnerability of a heart disease with few elemental symptoms and health factors. It is been fed by the basic data of the patients like age, sex. Machine learning helps to predict the vulnerability in advance which provides the cardiologists with great acumen for the adaption of the treatment. Machine learning algorithms have proven to produce reliable and accurate output with the help of the inputs. The algorithms used in the article include KNN and decision tree classifier which is compared to yield the desired and efficient output.

Keywords- *Machine Learning, Heart Disease Algorithms, KNN, Decision Tree Classifier Algorithm, Prediction Model.*

I. INTRODUCTION

For every 36 seconds, A human death is registered in the United States due to cardiovascular disease. Heart inflammations occur when parasites contact with the heart muscles. Healthy diet, regular exercises and stress management lowers the risk of cardiovascular diseases. Heart disease in India accounts for nearly 60% of global impact on cardiovascular disease. The primary factors of cardiovascular disease are obesity, cholesterol and high blood pressure. The elemental reason for the high mortality rate among the rural population is because of negligence and being left unnoticed. The medical centre collects all the basic information about the patient along with their medical history, but fails to use the data's effectively for the diagnosis. Machine learning algorithms utilizes all the useful information in order to make effective clinical decisions. Machine learning in healthcare easily diagnosis the disease which is considered as hard to detect and helps to provide the relevant ailment. The machine learning algorithm helps in reducing the manual medical errors with the help of the computer analysis.

Machine learning algorithm is an important aspect in finding the discrete patterns also by analyzing the given data set. The algorithm used in heart disease prediction are K-nearest neighbour (KNN) algorithm and decision tree algorithm. Since KNN algorithm is simplistic in using and as it depends on distance measures, it handles multiple class cases in the model, so as it efficiently works on the prediction model. Also, KNN considered the highly accurate algorithm. The next algorithm Decision tree classifier works by learning simple decision rules. It efficiently works and gets easily adapted to the dataset.

II. RELATED WORK

Since the heart disease prediction contains many open-source dataset there are some research articles available those are described below.

The dataset collected from alivkore a ECG visualizer which is connected with mobile phone, My heart a PDA for monitoring health with sensors ,Fitbit ,smartwatches and Health Gear an application to track medical test are collected for accurate prediction these data are cleaned and organized and finally the day is feed to a neural network with multilayer perceptron[1] which consist of input and output layer and a hidden layer which is used to connect the input and output layer and finally the activation function is applied on the weighted layer for prediction and this algorithm produced an accuracy of 91% which is quite low for real time applications and the use of a single algorithm for prediction is a drawback for this research work because if use of different algorithm may produce more accurate results.

978-1-6654-6275-4/22/$31.00 ©2022 IEEE

The heart disease prediction with the help of decision tree [2] algorithm was fit with the data which is collected from the database which provides screening clinical data and this data is pre-processed for effective mining with the help of KEEL tool to access the different data Mining algorithms with the decision tree algorithm the mode produced an accuracy of only 87%.

In this research paper the dataset used was small so the test and train contain is very less amount of data and it leads to less accurate result for real time classification and the author used a single algorithm for classification. The author has to use more data mining strategies to collect more instances of data and use of different algorithms to produce more accurate results.

The used hybrid machine learning model for predicting the heart disease[3] with the dataset that contains various useful features like sugar level, cholesterol for accurate classification of heart disease .here the dataset contains two classes that denotes whether the patient is affected by heart disease or not the data is trained with decision tree, random forest regression algorithm which produces an accuracy of 79% and 81% respectively by combined these algorithm to provide a hybrid algorithm and this model shows up a accuracy of 88% but using different hybrid algorithm the accuracy can be improved. However, this accuracy is skimpy for actual use [16].

Using k-nearest neighbour, support vector machine, linear regression and decision tree classifier the model for predicting the heart disease for the patient developed by Archana Singh et.al,[4] to find the presence of heart disease either its presence or absence using four machine learning algorithms and the accuracy of the algorithms are altogether compared between these models to levitate the prediction results for the model.

The dataset described for the model contains the symptoms like chest pain and the series of chemicals for the efficient calculation of the predicted result. Though the model used more than one algorithm in the research accuracy results obtained from the prediction are quite less. The highest accuracy of all the algorithm is obtained from the KNN algorithm which showed up the accuracy rate of 87% [17].

By using Cleveland dataset the author[5] used different feature selection methodologies like chi-squared measure which is used for feature selection this is a complicated process since medical data contains many features and elimination of important feature leads to issues in prediction and this chi-squared test is followed by genetic search which use optimal and near optimal features and select the most important features after this the data is trained with random forest regression which produced 83% accuracy which is low for crucial predictions[18].

The use of PSO algorithm[19-20] for rule production[6] is used to train the dataset collected from UCI machine learning repository database and the algorithm uses many local minima and locate the point of attraction with efficient gradient search for accurate prediction and this produced a accuracy of 96% and the decision tree algorithm produces 82% accuracy for best classes and the accuracy changes based on the classes .This different accuracy for different classes cause a desertion in real time prediction .

III. DATASET DESCRIPTION

The dataset consists of 1025 instances and 14 features the feature includes age, sex, cp-constrictive pericarditis which is a type of disease where the patient's heart tissues lose the elasticity, treetops which defines the amount of resting blood pressure, the cholesterol of the patient, fbs contains whether the patient is affected by fasting glucose level, the restECG shows the patient maximum heart rate achieved. The old peak measures the depression related to the rest which shows falls between zero to two and finally the target is the feature which shows the needed output.

IV. RESEARCH METHODOLOGY

Decision tree algorithm is categorized into supervised algorithm, used for the classification. It generates the output based upon the tree structure. The root node is found similar record for data for the optimal solution as shown in the fig 1.

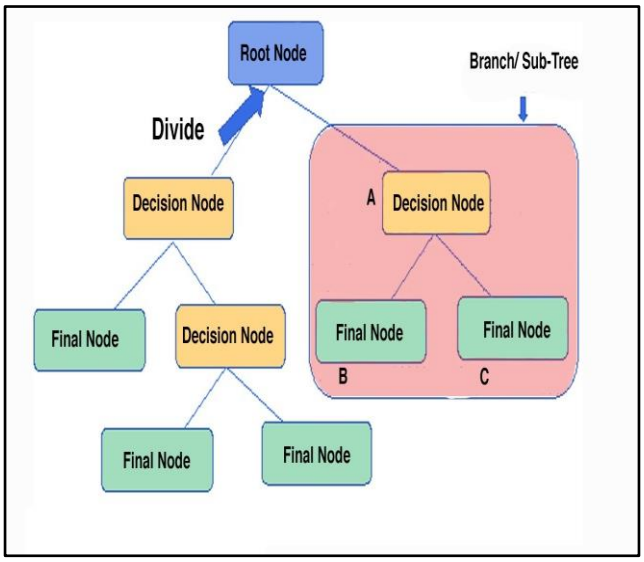

Fig. 1. Decision Tree Classifier

Decision tree classifier also falls under the category of supervised algorithm wherein it consists of two nodes namely decision nodes and leaf nodes.

The decision node represents the decision in a tree structured classifier whereas the leaf nodes represent the outcomes. The graph is obtained by comparing the score and max depth of the data set as shown in Fig.3.

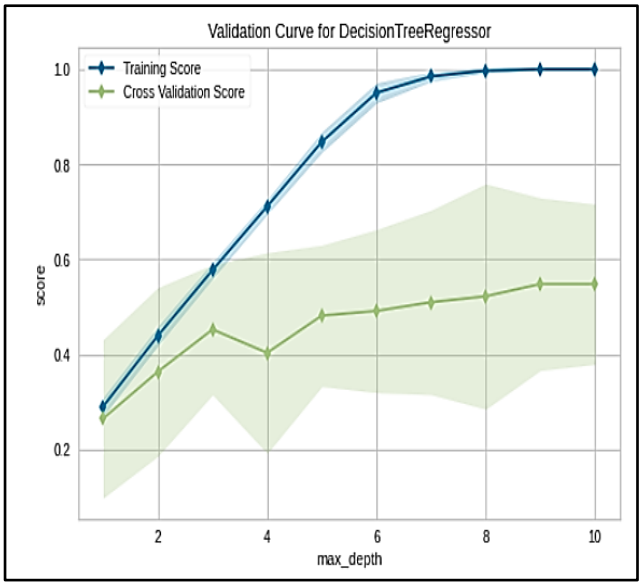

Fig. 2. Result obtained from DTC algorithm

KNN algorithm falls under the category of supervised algorithm which is utilized for both classified prediction and regression problems.

It categorizes the data and gives the new data point as shown in Fig 1. It is mainly used for classified prediction problems in the industry. KNN algorithm does not consist of a specialized training phrase and it doesn't assume conditions with the data.

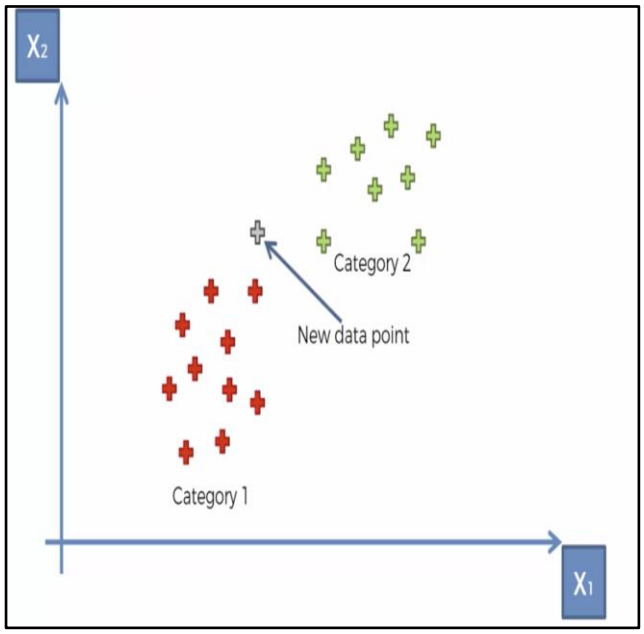

Fig. 3. KNN Algorithm

KNN algorithm falls under the category of supervised algorithm in which it makes a comparison between the new data with the existing data and groups the similar data together. KNN algorithm is simple to implement and considered as the best option for large training data. The resultant graph is shown in Fig.4.

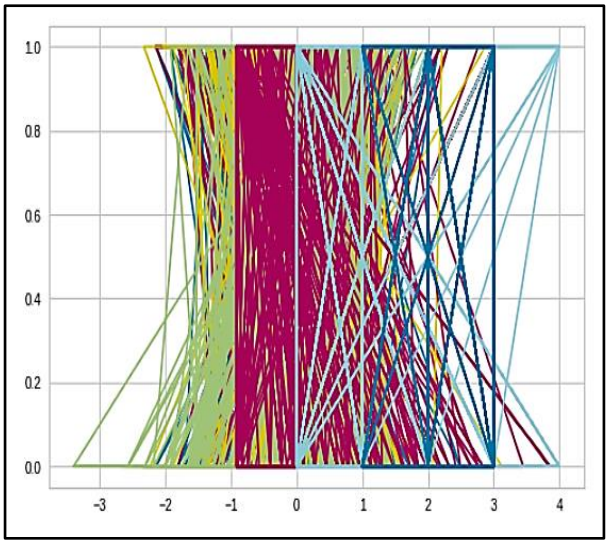

Fig. 4. Result Obtained from KNN Algorithm

MSE, MAE and RMSE values are basically used to predict the accuracy of a model. As squaring is one effective way of the removal of negative values, RMSE basically calculates the difference between expected and reality piece of data and determines the mean of the royalty with equation (1). The mean is square rooted.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|y(i) - \hat{y}(i)\|^2}{N}}, \tag{1}$$

Mean Squared Error (MSE) is generally the square value of all the errors. It is the average of the square of the difference between actual and predicted values as given by the formula in equation (2). Squaring is done in order to remove the negative values and blunders. It also produces more weight differences.

$$\sum_{i=1}^{n} \frac{\left(w^T x(i) - y(i)\right)^2}{n} \tag{2}$$

Mean Absolute Errors (MAE) is frequently used in regression models as an evaluation metric. It is the average of the absolute error, a non-negative value of the actual as well as predicted values with the equation (3).

$$mae = \frac{\sum_{i=1}^{n} abs\left(y_i - \lambda(x_i)\right)}{n} \tag{3}$$

V. RESULTS AND DISCUSSION

Machine learning algorithms play a vital role in health care. In future, we have decided to work on the prediction of a few chronic as well as fatal diseases like cancer, diabetics. Machine learning algorithms produce reliable and accurate optimal solutions. The MSE, MAE, RMSE are quite lower for the decision tree classifier when compared with the KNN. Thus, KNN are given the highest

priority in producing the optimal solution. The algorithm decision tree classifier with accuracy 0.85, whereas algorithm KNN with the accuracy 0.98. The accuracy of the decision tree classifier is quite impressive as it basically consists of two different nodes namely decision node and leaf node, as shown in the Table.1.

Fig. 5. Confusion Matrix

The decisions and leaf nodes represent the outcome. By comparing the accuracy of these algorithms, the Decision Tree Classifier is more efficient than KNN as shown in the Fig.6.

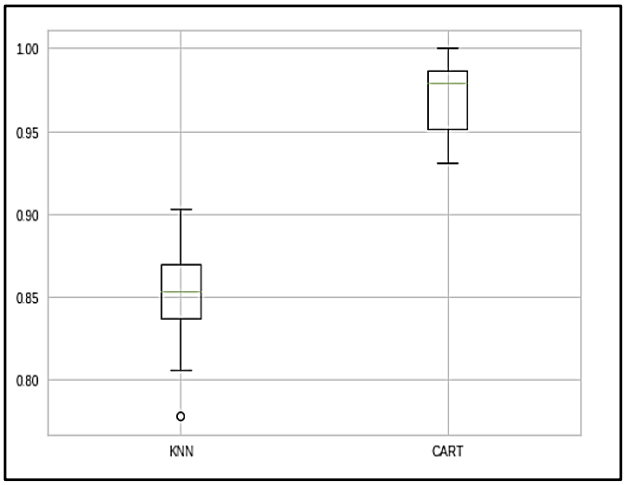

Fig.6. Comparison of KNN and DTC Algorithms

The accuracy of the model was predicted using MSE, MAE, RMSE methodology as mentioned in Table.2 and Table.3.

Table.1 - Accuracy of Two Algorithm

KNN	0.85714285714285714
DECISION TREE CLASSIFIER	0.9805194805194806

Table 2. Decision Tree Classification

MAE	0.0194805194
MSE	0.0194805194
RMSE	0.1395726315

Table 3. KNN Algorithm

MAE	0.14285714285
MSE	0.14285714285
RMSE	0.37796447300

VI. CONCLUSION

Cardiovascular diseases are the life-threatening disease in the real-world scenario. Using the proprietary sensors and machine learning algorithms, it outstands the human physicians in the prediction of cardiovascular diseases. Current lifestyle and lack of physical activities among the people adds more threat to the disease. There are a wide range of treatments in the modern science. But machine learning model is considered as the efficient and effective model because of its accurate and reliable output.

The heart prediction model is developed using the combination of two supervised algorithm namely K nearest neighbour and decision tree algorithm. The model produces optimal and feasible output with the given data input. As the heart disease is predicted in no time, cardiologists will provide the reliable treatment at the earlier stages and save the patient's life. Future Work, this heart disease prediction model offers a accuracy of 98% with decision tree algorithm and this model is adequate for real time use and in future the dataset can be modified with more features that are more relevant to the prediction and also collecting a huge dataset from various sources with the help of data mining techniques will produce more accurate prediction and also the use of Deep learning techniques and new algorithms for minimizing the errors.

REFERENCE

[1] Gavhane, Aditi, Gouthami Kokkula, Isha Pandya, and Kailas Devadkar. "Prediction of heart disease using machine learning." In *2018 second international conference on electronics, communication and aerospace technology (ICECA)*, pp. 1275-1278. IEEE, 2018. Efficient Heart Disease Prediction System using Decision Tree -Kanak Saxena,Kanak Saxena -2015

[2] Kavitha, M., G. Gnaneswar, R. Dinesh, Y. Rohith Sai, and R. Sai Suraj. "Heart disease prediction using hybrid machine learning model." In *2021 6th International*

Conference on Inventive Computation Technologies (ICICT), pp. 1329-1333. IEEE, 2021.

[3] Singh, A., & Kumar, R. (2020, February). Heart disease prediction using machine learning algorithms. In 2020 international conference on electrical and electronics engineering (ICE3) (pp. 452-457). IEEE. Intelligent heart disease prediction system using random forest and evolutionary approach -M.A.Jabbar 1;B.L.Deekshatulu2

[4] Li, N., Zhao, L., Chen, A.X., Meng, Q.W. and Zhang, G.F., 2009, July. A new heuristic of the decision tree induction. In 2009 International Conference on Machine Learning and Cybernetics (Vol. 3, pp. 1659-1664). IEEE.

[5] Alkeshuosh, Azhar Hussein, Mariam Zomorodi Moghadam, Inas Al Mansoori, and Moloud Abdar. "Using PSO algorithm for producing best rules in diagnosis of heart disease." In 2017 international conference on computer and applications (ICCA), pp. 306-311. IEEE, 2017.

[6] Tan, Yi, and Guo-Ji Zhang. "The application of machine learning algorithm in underwriting process." In 2005 international conference on machine learning and cybernetics, vol. 6, pp. 3523-3527. IEEE, 2005. Application of machine learning techniques to Web-based intelligent learning diagnosis systems- Chenn-Jung Huang;Ming-Chou Liu;San-Shine Chu;Chin-Lun Cheng-2019

[7] Katarya, Rahul, and Polipireddy Srinivas. "Predicting heart disease at early stages using machine learning: a survey." 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, 2020.

[8] Learning, Machine. "Heart disease diagnosis and prediction using machine learning and data mining techniques: a review." Advances in Computational Sciences and Technology 10.7 (2017): 2137-2159.

[9] Gramajo, M., L. Ballejos, and M. Ale. "Seizing Requirements Engineering Issues through Supervised Learning Techniques." IEEE Latin America Transactions 18, no. 07 (2020): 1164-1184.

[10] Shailaja, K., B. Seetharamulu, and M. A. Jabbar. "Machine learning in healthcare: A review." In 2018 Second international conference on electronics, communication and aerospace technology (ICECA), pp. 910-914. IEEE, 2018.

[11] Haq, A.U., Li, J.P., Memon, M.H., Nazir, S. and Sun, R., 2018. A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems, 2018.

[12] Higashi, Kotaro, Guanghao Sun, and Koichiro Ishibashi. "Precise heart rate measurement using non-contact Doppler radar assisted by machine-learning-based sleep posture estimation." 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019.

[13] Rahman, J. Riyazulla, Shridhar Sanshi, and N. Nasurudeen Ahamed. "Health Monitoring and Predicting System using Internet of Things & Machine Learning." In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 223-226. IEEE, 2021.

[14] Mohan, Narendra, Vinod Jain, and Gauranshi Agrawal. "Heart Disease Prediction Using Supervised Machine Learning Algorithms." 2021 5th International Conference on Information Systems and Computer Networks (ISCON). IEEE, 2021.

[15] Diwakar, Manoj, et al. "Latest trends on heart disease prediction using machine learning and image fusion." Materials Today: Proceedings 37 (2021): 3213-3218

[16] S. G. A, R. Prabha, M. Razmah, S. Sridevi, D. Roopa and R. M. Asha, "A Big Wave of Deep Learning in Medical Imaging - Analysis of Theory and Applications," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), 2022, pp. 1321-1327, doi: 10.1109/ICICCS53718.2022.9788412.

[17] R-Prabha M, Prabhu R, Suganthi SU, Sridevi S, Senthil GA, Babu DV. Design of Hybrid Deep Learning Approach for Covid-19 Infected Lung Image Segmentation. In Journal of Physics: Conference Series 2021 Oct 1 (Vol. 2040, No. 1, p. 012016). IOP Publishing. doi:10.1088/1742-6596/2040/1/012016.

[18] Prabha, R., G. A. Senthil, A. Lazha, D. VijendraBabu, and Ms D. Roopa. "A Novel Computational Rough Set Based Feature Extraction for Heart Disease Analysis." In I3CAC 2021: Proceedings of the First International Conference on Computing, Communication and Control System, I3CAC 2021, 7-8 June 2021, Bharath University, Chennai, India, p. 371. European Alliance for Innovation, 2021. DOI:10.4108/eai.7-6-2021.2308575.

[19] Senthil, G.A., Raaza, A. & Kumar, N. Internet of things multi hop energy efficient cluster-based routing using particle swarm optimization. Wireless Netw (2021). https://doi.org/10.1007/s11276-021-02801-0.

[20] Senthil, G.A., Raaza, A. & Kumar, N. Internet of Things Energy Efficient Cluster-Based Routing Using Hybrid Particle Swarm Optimization for Wireless Sensor Network. Wireless Pers Commun (2021). https://doi.org/10.1007/s11277-021-09015-9.