Cardio–Vascular Disease Prediction Using Machine Learning

Dr. Jyothi A P
Assistant Professor, Dept. of CSE,
Faculty of Engineering and Technology,
M S Ramaiah University of Applied Sciences,
Bangalore, Karnataka, India.
jyothiarcotprashant@gmail.com

Prakruthi L
Student, Dept. of CSE,
Faculty of Engineering and Technology,
M S Ramaiah University of Applied Sciences,
Bangalore, Karnataka, India.
l.prakruthi1999@gmail.com

Shilpa H H
Student, Dept. of CSE,
Faculty of Engineering and Technology,
M S Ramaiah University of Applied Sciences,
Bangalore, Karnataka, India.
Shilpahh204@gmail.com

ABSTRACT

The heart is a muscular organ that pumps blood through the arteries and veins called the cardiovascular system. Many people die because of heart disease. Symptoms like chest pain, insufficient blood pump, blockage in veins, irregular heartbeats, and many more. In this paper, we are using some machine learning algorithms like decision tree, K-neighbors classifier & logistic regression to predict cardiovascular disease. To get the best accuracy we are comparing the features of these algorithms.

Keywords – Machine Learning; Cardiovascular; decision tree; k neighbors classifier; logistic regression; accuracy.

I. INTRODUCTION

Heart disease is acknowledged as the most complicated and deadly diseases affecting people today. This illness causes the heart to operate improperly, resulting in clogged blood arteries, angina, heart attacks, and stroke. These days many people are dying due to heart attacks. So, by using some machine learning techniques we are predicting heart disease with high accuracy. Heart disease is predicted using symptoms like - age, sex, constrictive pericarditis, trestbps, cholesterol, fasting blood sugar, resting electrocardiographic results, thalach, exercise-induced angina, old peak, slope, ca, thalassemia, target (heart disease yes=1, no=0). There are now too many advanced tools, such as data mining, machine learning, deep learning, etc., to identify cardiac disease. Therefore, we shall give a basic overview of machine learning approaches in this work. Using machine learning resources, we train the datasets in this work.

II. LITERATURE REVIEW

In the medical center, many works have been completed related to disease prediction using machine learning techniques.

Limbitote M, Damkondwar K, Mahajan D, Patil P. [1] "A Survey on Prediction Techniques of Heart Disease using Machine Learning" in which the objective is to find the most accurate prediction techniques. The most used algorithms are Decision Tree, SVM, ANN, Naive Bayes, Random Forest, KNN. The main objective is to obtain the most accurate algorithm with less cost.

Alotaibi FS. "Implementation of Machine Learning Model" which aims to detect heart failure accuracy using multiple machine learning techniques. In this paper decision tree, logistic regression, random forest, naïve bayes, and SVM are the techniques used to detect heart failure accuracy.

Zriqat IA, Altamimi AM, Azzeh M. "A Comparative Study for Predicting Heart Diseases Using Data Mining Classification Methods", it aims to develop an effective intelligent medical decision system using data mining techniques. It uses 5 data mining techniques Naïve Bayes, Decision Tree, Discriminant, Random Forest, and Support Vector Machine. Here all 5 techniques are compared and the best is selected.

Rajdhan A, Agarwal A, Sai M, Ravi D. "Heart Disease Prediction using Machine Learning", it predicts the chances of heart disease and classifies the patient's risk level by implementing different data mining techniques such as Naive Bayes, Decision Tree, Logistic Regression and Random Forest. This paper compares all the techniques and one of the best techniques is selected.

Patel J, Upadhyay T, Patel S. "Heart Disease Prediction Using Machine learning and Data Mining Technique", to reduce the deaths caused by heart disease, some quick and effective algorithm is used. In this paper, the comparison of the Logistic model tree algorithm and Random Forest algorithm and effectiveness will be selected.

Amudhini VP, Santhini T, Kailash P, Nivetha D, Poonguzhali R. "Survey On Machine Learning Algorithms For Prediction of Cardiovascular Disease". In this paper, the survey of some related works is done by reviewing some projects.

TABLE 1: Research Analysis

Year	**Author Name**	**Purpose**	**Techniques**	**Accuracy**
2020	Limbitote M, Damkondwar K, Mahajan D, Patil P.	"A Survey on Prediction Techniques of Heart Disease using Machine Learning".	Decision Tree, SVM, ANN, Naive Bayes, Random Forest, KNN.	88%
2019	Alotaibi FS.	"Implementation of Machine Learning Model".	Decision Tree Logistic Regression Random Forest Naïve Bayes SVM	82.22% 82.56% 84.17% 84.24% 84.85%
2016	Zriqat IA, Altamimi AM, Azzeh M.	"A Comparative Study for Predicting Heart Diseases Using	Naïve Bayes, Decision Tree, Discriminant, Random Forest, and Support Vector Machine	99.0%

		Data Mining Classification Methods".		
2020	Rajdhan R, Agarwal A, Sai M, Ravi D.	"Heart Disease Prediction using Machine Learning".	Naive Bayes, Decision Tree, Logistic Regression, Random Forest	85.25% 81.97% 85.25% 90.16%
2015	Patel J, Upadhyay T, Patel S.	"Heart Disease Prediction Using Machine learning and Data Mining Technique".	Logistic model tree algorithm, Random Forest algorithm	55.77% 56.76%
2019	Amudhini VP, Santhini T, Kailash P, Nivetha D, Poonguzhali R.	"Survey On Machine Learning Algorithms for Prediction of Cardiovascular Disease".	SVM, Naïve bayes, The Neural network, Feed Forward, Back propagation NN, KNN, Decision Tree	50% -60% 52% 100% 57% 50-60% 80.6%

III. MACHINE LEARNING

It is one of the fastest-growing fields in computer science. Its application areas are abundant. A computer engineer writes a series of instructions that instruct a computer on how to transform input data into the desired output in traditional programming. The majority of instructions follow an IF-THEN structure: when certain conditions are met, the program performs a specific action. The Machine learning in contrast, side, is a computer-assisted process that allows machines to deal with problems with minimal or no human intervention and perform steps based on previous observations. It uses some statistical mathematical approaches to teach the model.

Each training sample contains an input as well as the desired output. When determining the labels for unseen data, a supervised machine learning algorithm analyses the sample data and makes an assumption - basically, an educated guess. The most typical and well-liked technique for machine learning is this one. Because these models need to be fed manually tagged sample data to learn from, it is "supervised." Data is labeled so that the machine will know what patterns to search for and connections to make. In unlabeled data, unsupervised learning reveals patterns and relationships. Models are given input data in this situation, but since the positive outcomes are unknown, they must draw conclusions based solely on corroborating evidence without any training or direction. The models must recognize patterns on their own because they weren't taught the "correct answer." For instance, the marketing department of an online retailer could use clustering to enhance customer segmentation. A machine learning model can locate communities of customers who exhibit comparable behaviors given a set of income and consumption data. Training data for semi-supervised learning is divided into two groups. A smaller set of data with labels and a larger set without labels. This method is becoming more and more popular, particularly for tasks requiring large datasets, like image classification. Semi-supervised learning is the best option for businesses that receive a lot of data because it doesn't require a lot of labeled data, is easier to set up, and is less expensive than supervised learning techniques. How a software agent should behave in a situation to maximize the reward is the focus of reinforcement learning. Refined machine learning models, in essence, look for the best course of action to take in a specific

circumstance. Trial and error are how they learn to do this. As there is no training data, machines learn from their mistakes and select actions that will produce the best result or the highest reward.

IV. MACHINE LEARNING IN CARDIOLOGY

Machine learning has a wide range of applications including in the medical field. Nowadays medical centers are working on the application of machine learning and artificial intelligence to avoid mankind's errors. Not only in Cardiology every branch of the medical field like Pediatrics, Orthopedics, Dermatology, Pathology, and so on are working based on Machine learning, Artificial intelligence techniques. In this paper, we are using Machine learning techniques to predict cardiovascular disease. Some of the algorithms which we are used decision tree, Logistic regression, and K neighbor classifier.

Decision Tree

Decision Trees are interpreted as supervised learning techniques that can be applied to regression. These models have a tree-like structure, with internal nodes standing in for a dataset's features, branches for previously learned decision rules, and leaf nodes for the results. The main problem faced during the implementation of the decision tree is how to choose an attribute as a root node, the frontier & leaf node. By using attribute selection measure, we can resolve this issue. There are two methods in attribute selection measure namely, 1. Gini index 2. Information gain.

Logistic Regression

Logistic Regression is the most popular supervised learning. Using a predetermined set of independent variables, it is used to predict the categorical dependent variable. In a classification rule, the output is predicted by logistic regression. With the exception of how they are applied, logistic regression and linear regression are very similar. While logistic regression is used to solve classification problems, linear regression is used to solve regression problems. Because it can classify new data using both continuous and discrete datasets, logistic regression is a significant machine learning algorithm.

K neighbor classifier

It is the simple algorithm in machine learning which follows supervised learning. K neighbor classifier makes the assumption that the new case and the existing cases are similar, and it places the new case in the category that is most like the existing categories. A new data point is classified using the K neighbor classifier algorithm based on similarity after all the existing data has been stored. This means that using the K neighbor classifier algorithm, new data can be quickly and accurately classified into a suitable category. Although the K neighbor classifier algorithm is most frequently used for classification problems.

After the analysis we got the accuracy from the decision tree is 94.34%, logistic regression is 88.68% and K neighbor classifier is 64.47%. So, we found the decision tree is the best technique to predict the heart disease.

V. CONCLUSION

We have outlined various machine learning algorithms for heart disease prediction. We developed a number of machine learning algorithms and then examined their features to determine which algorithm was the best. Every algorithm has produced a different result in a variety of circumstances. Here we get a decision tree with the best accuracy value and we observed the K neighbor classifier with less accuracy. Further analysis shows that the predictive model for heart disease only achieves marginal accuracy.

REFERENCES

[1] Limbitote M, Damkondwar K, Mahajan D, Patil P. (2020) A Survey on Prediction Techniques of Heart Disease using Machine Learning, The International Journal of Engineering Research & Technology, pp.9(6), 2278-0181.

[2] Amudhini, V.P. Santhini, T. kailash, P. N, D. and Poonguzhali, R. (2019) Survey on Machine Learning Algorithms For Prediction of Cardiovascular Disease, The International Journal of Research and Analytical Reviews, 6(1),2349-5138.

[3] Golande A, Kumar TP. (2019) Heart Disease Prediction Using Effective Machine Learning Techniques, The International Journal of Recent Technology and Engineering, 8, 2277-3878.

[4] Lakshmanarao A, Swathi Y, Sundareswar PS. (2019) Machine Learning Techniques for Heart Disease Prediction, The International Journal of Scientific & Technology Research,8(11).

[5] Krishnan SJ, Geetha S. (2019) Prediction of Heart Disease Using Machine Learning Algorithms, The International Conference on Innovations in Information and Communication Technology (ICIICT), doi:10.1109/ICIICT1.2019.8741465.

[6] Amin MS, Chiam YK, Varathan KD. (2019) Identification of significant features and data mining techniques in predicting heart disease, Telematics Inform., 36(1), pp. 82–93.

[7] Gavhane A, Kokkula G, Pandya I, Devadkar K. (2018) Prediction of Heart Disease Using Machine Learning, Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology. IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1- 5386-0965-1

[8] Weng SF, Reps J, Kai J. Jonathan M, Garibaldi, Qureshi N. (2017) Can machine-learning improve cardiovascular risk prediction using routine clinical data **https://doi.org/10.1371/journal.pone. 0174944.**

[9] Kishore A, Kumar A, Singh K, Punia M, Hambir Y. (2018) Heart Attack Prediction Using Deep Learning, The International Research Journal of Engineering and Technology (IRJET), 05(04).

[10] Oskouie SK, Prenner SB, Shah SJ, and Sauer, A.J., (2017) Differences in repolarization heterogeneity among heart failure with preserved ejection fraction phenotypic subgroups. Am J Cardiol;120(4):601–6.

[11] Gupta SD, Kumar A, Sharma, Performance analysis of various data mining classification techniques on healthcare data. The International journal of computer science & Information Technology (IJCSIT), 2011. 3(4).

[12] Kavita, et al., "APPLICATION OF DATA MINING IN HEALTH CARE", The international journal of advanced technology in engineering and science.

[13] Cleveland Clinic Foundation, "Heart Disease Data Set ", Available at: **http://archive.ics.uci.edu/ml/datasets/Heart+Disease**.

[14] Chitra R, Seenivasagam V. “REVIEW OF HEART DISEASE PREDICTION SYSTEM USING DATA MINING AND HYBRID INTELLIGENT TECHNIQUES”, The Journal on Soft Computing (ICTACT), 3(4), pp. 605–609.

[15] Dbritto R, Srinivasaraghavan A, Joseph V. “Comparative Analysis of Accuracy on Heart Disease Prediction using Classification Methods”, The International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868, The Foundation of Computer Science FCS, New York.