

Review Paper on Heart Disease Prediction using SVM

Dr. Radha Shirbhate¹, Dhiraj Chatur², Hanmant Karhale³, Neel Lihitkar⁴, Ranjit Kale⁵

Professor, Department of Computer Science¹

UG Students, Department of Computer Science^{2,3,4,5}

G. H. Raisoni Institute of Engineering and Technology, Pune, Maharashtra, India

Abstract: Increasing Heart diseases in urban areas is major concern. It is not possible for common man to frequently undergo test like ECG and so there is need of system which can predict heart diseases which is reliable and budget friendly. Data mining techniques can be used to identify whether a patient is normal or having heart disease. We can predict the vulnerability on the basic symptoms like age, sex, pulse rate, etc. Machine learning algorithms can be used to precisely predict heart diseases. This paper presents a survey of various Machine learning algorithms like Naive Bayes, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine.

Keywords: Heart Disease, Naive Bayes, Machine Learning, Support Vector Machine, Decision Tree

I. INTRODUCTION

Heart is one of the vital organ of human body, so the care of heart is essential. It pumps blood to every part of human body. If it fails to function properly then brain and other organs will stop working, within a few minutes person will die. Normal person heart beat is 72 per minute and normal blood pressure is less than 120/80mmHg. In humans, the heart is approximately the size of closed fist.

According to World Health Organization 31% deaths are caused due to heart diseases. Cardiovascular diseases(CVD) is a class that involves heart or blood vessels. CVD includes heart failure, stroke, abnormal heart rhythm, valvular diseases, venous thrombosis, carditis, etc. These diseases can be caused due to smoking, high blood pressure, pure diet, lack of exercise, obesity, high blood cholesterol, diabetes mellitus, etc.

Various machine learning algorithms under various conditions can be used to predict heart diseases. Algorithms such as Naive Bayes, Logistic Regression, Random Forest, Support Vector Machine. Naive Bayes is a classification technique based on Bayes' theorem with an assumption of independence among predictor. Logistic Regression computes the probability of even occurrence. Random Forest is a classifier that contains a number of decision tree on various subsets of the given data sets and takes the average to improve the predictive accuracy of the that data set. Support Vector Machine is one of the most popular Supervised Learning Algorithm which is used for classification as well as Regression problems.

II. RELATED WORK

Title	Author	Abstract	Algorithm
Heart Disease Prediction using Machine Learning Algorithm.	Archana Singh. Computer Science and Engineering Madan Mohan Malaviya University of Technology, Gorakhpur.	Heart plays significant role in living organisms. Diagnosis and prediction of heart related diseases requires more precision, perfection and correctness because a little mistake can cause fatigue problem or death of the person, there are numerous death cases related to heart and their counting is increasing exponentially day by day. To deal with the problem	Linear regression : It is the supervised learning technique. It is based on the relationship between independent variable and dependent variable variable "x" and "y" are independent and dependent variable and relation between them is shown by equation of line which is linear in nature that why this approach is called linear regression.
	Rakesh Kumar. Computer Science and Engineering		Decision tree : On the other hand

	Madan Mohan Malaviya University of Technology Gorakhpur,	there is essential need of prediction system for awareness about diseases. Machine learning is the branch of Artificial Intelligence(AI), it provides prestigious support in predicting any kind of event which take training from natural events. In this paper, we calculate accuracy of machine learning algorithms for predicting heart disease, for this algorithms are k- nearest neighbor, decision tree, linear regression and support vector machine(SVM) by using UCI repository dataset for training and testing. For implementation of Python Programming Anaconda (jupyter) notebook is best tool, which have many type of library, header file, that make the work more accurate and precise.	decision tree is the graphical representation of the data and it is also the kind of supervised machine learning algorithms. Support Vector Machine : It is one category of machine learning technique which work on the concept of hyperplane means it classify the data by creating hyperplane between them. K-nearest Neighbour : It work on the basis of distance between the location of data and on the basis of this distinct data are classified with each other. All the other group of data are called neighbor of each other and number of neighbor are decided by the user which play very crucial role in analysis of the dataset.
Prediction of Heart Disease Using Machine Learning Algorithms .	Mr, Santha Krishnn. PG Student Department of Computer Applications Anna University, BIT Campus	Health care field has a vast amount of data, for processing those data certain techniques are used. Data mining is one of the techniques often used. Heart diseases the Leading cause of death worldwide. This Health care field has a vast amount of data, for processing those data certain techniques are used. Data mining is one of the techniques often used. Heart diseases the Leading cause of death worldwide. This	Decision Tree Classification Algorithm: The decision tree is a supervised machine learning algorithm. It handles both the categorical data and numerical data. Based on certain conditions it gives a categorical solution such Yes/No, True or false, 1 or 0. For handling medical dataset the Decision tree Classification algorithm is widely used. The result of this model differing from the other models like the knn model, SVM model. Naïve Bayes Classification Algorithm: Naïve Bayes classifier is a supervised algorithm which classifies the dataset on the basis of Bayes theorem. The Bayes theorem is a rule or the mathematical concept that is used to get the probability is called Bayes theorem. Bayes theorem requires some independent assumption and it requires independent variables which is the fundamental assumption of Bayes theorem.
Prediction of Heart Disease	Himanshu Sharma, Department of Computer	According to recent survey by WHO organization 17.5 million people dead each year. It will increase to 75	Decision tree : Decision Tree is a graphical representation of specific decision situation that used for

using Machine Learning Algorithms : A Survey	Engineering and Applications, National Institute of Technical Teachers M A Rizvi, Department of Computer Engineering and Applications, National Institute of Technical Teachers' Training and Research.	million in the year 2030. Medical professionals working in the field of heart disease have their own limitation, they can predict chance of heart attack up to 67% accuracy, with the current epidemic scenario doctors need a support system for more accurate prediction of heart disease. Machine learning algorithm and deep learning opens new door opportunities for precise predication of heart attack. Paper provide slot information about state of art methods in Machine learning and deep learning. An analytical comparison has been provided to help new researches' working in this field.	predictive model, main component of decision tree involves root, nodes, and branching decision. There are few approaches for building tree such as ID3, CART, CYT, C5.0 and J48 as used the approaches to classify the dataset using J48, similarly have compared decision tree with classification output of other algorithm. Decision tree is used in those area of the medical science where numerous parameters involved in classification of data set. Support Vector Machine(SVM) : A SVM performs classification by finding the hyper plane that maximize the margin between two classes. The vectors that define the hyper plane are the support vectors. KNN : KNN is slow supervised learning algorithm, it take more time to get trained classification like other algorithm is divided into two step training from data and testing it on new instance . The K Nearest Neighbour working principles based on assignment of weight to the each data point which is called as neighbour. in K Nearest Neighbour distance is calculate for training dataset for each of the K Nearest data points now classification is done on basis of majority of votes there are three types of distances need to be measured in KNN
Effective Heart Disease Using Hybrid Machine Learning Techniques	Senthikar Mohan school of Information Technology and Engineering, VIT University, 632015, India Chandsagr Thirumai Department of Mathematics and Computer Science, Brandon University, Brandon, MB R7A 6A9, Canada	Heart disease is one of the most significant causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. We have also seen ML techniques being used in recent developments in different areas of the Internet of	Decision Trees For training samples of data D, the trees are constructed based on high entropy inputs. These trees are simple and fast constructed in a top down recursive divide and conquer (DAC) approach. Tree pruning is performed to remove the irrelevant samples on D. Support Vector Machine Let the training samples having dataset $Data \ y_i, x_i \ i \ 1, 2, \dots, n$ where $x_i \in R^n$ represent the i^{th} vector and y_i Represent the target item. The linear SVM finds the optimal hyperplane of

	Gautam Srivastava Research Center for Inter neural Computing, China Medical University, Taichung 40402, Taiwan	Things (IoT). Various studies give only a glimpse into predicting heart disease with ML techniques. In this paper, we propose a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques. We produce an enhanced performance level with an accuracy level of 88.7% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM).	the form $f(x) = w^T x + b$ where w is dimensional coefficient vector and b is a offset. This is done by solving the subsequent optimization problem Random Forest : This ensemble classifier builds several decision trees and incorporates them to get the best result. For tree learning, it mainly applies bootstrap aggregating or bagging. For a given data, $X = \{x_1, x_2, x_3, \dots, x_n\}$ with responses $Y = \{y_1, y_2, y_3, \dots, y_n\}$ which repeats the bagging from $b = 1$ to B. Navie Bayes: This learning model applies Bayes rules through independent features. Every instance of data D is allotted to the class of highest subsequent probability. The model is trained through the Gaussian function with prior probability $P(X_j = \text{priority}) \in (0 : 1)$
Heart Disease Prediction Using Machine learning and Data Mining Technique	Prof. Tejal Upadhyay Nirma University, Gujarat, India Principal Dr. Samir Patel Department of Computer Science and Engineering Jaymin Patel Grow More Faculty of Engineering, Ahmadabd,	Heart disease is the main reason for death in the world over the last decade. Almost one person dies of Heart disease about every minute in the United States alone. Researchers have been using several data mining techniques to help health care professionals in the diagnosis of heart disease. However using data mining technique can reduce the number of test that are required. In order to reduce number of deaths from heart diseases there have to be a quick and efficient detection technique. Decision Tree is one of the effective data mining methods used. This research compares different algorithms of Decision Tree classification seeking better performance in heart disease diagnosis using WEKA. The algorithms which are tested is J48 algorithm, Logistic model tree algorithm and Random Forest algorithm.	Classification Tree Algorithms Used J48 algorithm: J48 is an open source Java implementation of the C4.5 algorithm in the WEKA tool. This algorithm utilizes an avaricious method to make decision trees for classification and uses decreased-error pruning. Decision tree is built by examining information hubs, which are utilized to assess hugeness of existing highlights Logistic Model Tree Algorithm: Logistic Model Tree is the classifier for building logistic model trees, which consist of a decision tree structure with logistic regression function at the leaves. The algorithm can oversee parallel and multi-class target variables, numeric and nominal attributes and missing qualities. A combination of learners that rely on simple regression models if only little and/or noisy data is available and add a more complex tree structure if there is enough data to warrant such a structure. Random Forest Algorithm: Random forest is an ensemble classifier that

			consists of many decision trees. The output of the classes is represented by individual trees. It is derived from random decision a forest that was proposed by Tin Kam Ho of Bell Labs in 1995
--	--	--	---

III. SUPPORT VECTOR MACHINE

It works on the concept of Hyper plane means it classify the data by creating hyper plane between them. A simple SVM works by making a straight line between two classes. That means all the data points on one side of line will represent a category and the data points on the other side of the line will be put into different category. SVM Algorithm better because it chooses the best line to classify your data points. SVM are used in application like Handwriting recognition, Face Detection, Email Classification, Gene Classification and in Web Pages. There are advantages of SVM like :

- It is Effective on Datasets with multiple features like financial or medical data.
- It is Effective in cases where no of features is greater than no of data points.
- It uses Subset of training points in the decision function called Support Vectors which makes it memory efficient

Algorithm	Accuracy	Classification Error	Precision
Decision Tree	85	15.0	86
Naive Bayes	75.8	24.2	90.5
Logistic Regression	82.9	17.1	89.6
Random Forest	86.7	13.9	87.1
SVM	86.1	13.9	86.1

IV. CONCLUSION

Heart disease is crucial problem in human society. Prediction about heart diseases is also important concern so the accuracy of algorithm is one of the parameter for analysis . Accuracy depends on dataset that is used for training and testing purpose. There is huge scope for machine learning algorithm in predicting heart disease. When naive byes and Decision tree were applied on same dataset Decision tree accuracy is 91% were naive bayes has 87%. Decision tree algorithm is better for handling medical dataset . All algorithm mentioned have perform extremely well in some cases but poorly in some other cases. Decision tree with PCA performed well but Decision tree performed extremely poor in some other cases due to over fitting. Random forest in ensembles model they performed well because the solve overfitting problem by implying multiple algorithm. Models like naive bayes were computationally very fast and have also performed well. SVM performed extremely well for most of the cases. A lot of research is required to handle high dimensional data and over fitting problem.

REFERENCES

- [1]. Archana Singh, Rakesh kumar. "Heart Disease Prediction using Machine Learning Algorithm" 2020. International Conference on Electrical and ELectionics Engineering (ICE3-2020).
- [2]. Mr Santhana Krishnan, Dr Geetha. s. "Prediction of Heart disease using Machine Learning Algorithm ".
- [3]. Himanshu Sharma, M A Rizvi. "Prediction of Heart Disease using Machine Learning Algorithm. International Journal on Recent and Innovation trends in computing and communicate".
- [4]. Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava."Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques". Received May 13,2019,accepted june9,2019,date of publication June 19, 2019,date of current version July 3, 2019.
- [5]. Prof. Tejal Upadhyay, Dr. Samir Patel, Jaymin Patel."Heart Disease Predictio using Machine Learning and Mining technique."Volume 7 Sept 2015 - March 2016.

- [6]. D. Tain, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, "A dynamic and self-adaptive network selection method for multimode communication in heterogeneous vehicular telematics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3033-3049, 2015.
- [7]. M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, C. Youn, "Wearable 2.0: Enable Human-Cloud Integration in Next Generation Healthcare System," *IEEE Communications Magazine*, Vol. 55, No. 1, pp. 54-61, Jan. 2017.
- [8]. M. Chen, Y. Ma, J. Song, C. Lai, B. HU, "Smart Clothing: Connection Human With Clouds and Big Data for Sustainable Health Monitoring," *ACM/Springer Mobile Networks and Applications*, Vol. 21, No. 5, pp. 825-845, 2016.
- [9]. M. Chen, P. Zhou, G. Fortino, "Emotion Communication System," *IEEE Access*, DOI:10.1109/ACCESS.2016.2641480, 2016.
- [10]. J. Wang, M. Qui, and B. Guo, "Enabling real-time information service on telehealth system over cloud-based big data platform," *Journal of System Architecture*, vol. 72, pp. 69-79, 2017.
- [11]. Y. Zhang, M. Qui, C.-W. Tsai, M.M. Hassan, and A. Alamri, "Healthcps: Healthcare cyber-physical system assistance by cloud and big data," *IEEE Systems Journal*, 2015.
- [12]. K. Lin, J. Luo, L. Hu, M.S. Hossain, and A. Ghoneim, "Localization based on social big data analysis in the vehicular networks," *IEEE Transactions on Industrial Informatics*, 2016.
- [13]. K. Lin, M. Chen, J. Deng, M. M. Hassan, and G. Fortino, "Enhanced finger-printing and trajectory prediction for IoT localization in smart buildings," *IEEE Transactions on Automation Science and Engineering*, vol. 13, pp. 1294-1307, 2016.
- [14]. B. Quin, X. Wang, N. Cao, H. Li, and Y. G. Jaing, "A relatively similarity Based method for interactive patient risk prediction," *Data Mining and Knowledge Discovery*, vol. 29, no. 4, pp. 1070-1093, 2015.
- [15]. S. Bandyopadhyay, J. W. Wolfson, D.M. Vock, G. Vazquez-Benitez, G. Adomavicius, M. Elidrisi, P. E. Johnson, and P. J. O'Connor, "Data mining for censored time-to-event data: a Bayesian network model for predicting cardio-vascular risk from electronic health record data," *Data Mining and Knowledge Discovery*, vol. 29, no. 4, pp. 1033-1069, 2015.
- [16]. J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas and A. Vasilakos, "A manufacturing Big Data Solution for Active Preventive Maintenance," *IEEE Transactions on Industrial Informatics*, DOI: 10.1109/TII.2017.2670505, 2017.