Data Analytics and Artificial Intelligence
Vol: 2(5), 2022
REST Publisher; ISBN: 978-81-948459-4-2
Website: http://restpublisher.com/book-series/data-analytics-and-artificial-intelligence/

Heart Disease Prediction using Machine Learning

***Divyapraba. K, Sr. B. Arockia Valan Rani**
St. Joseph's College of Arts and Science for Women, Hosur, Tamil Nadu, India
*Corresponding author Email: diviyapraba442001@gmail.com

Abstract. Hearts are important to all living things. Heart related diseases requires more 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 increasingexponentially day by day. In recent years, numerous researchers have employed a variety of machine learning techniques to aid the medical community and specialists in the detection of heart-related disorders. This study surveys numerous models built on these methods and evaluates their functionality. Researchers are reported to be extremely interested in models based on supervised learning algorithms including Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Naive Bayes, Decision Trees (DT) and Random Forest (RF).

Keywords: Cardiovascular Diseases; Random Forest; Decision Tree, Support Vector Machine; K-Nearest Neighbour; Naïve Bayes

1. Introduction

One of the effective testing tool is machine learning, which is based on training and testing. Machine learning is a particular subset of Artificial Intelligence (AI), a large field of learning in which machines imitate human abilities.One of the main causes of death in today's modern society is heart disease. Smoking, drinking alcohol, eating a lot of fat, and living a sedentary lifestyle are all risk factors for heart disease. More than 10 million people worldwide pass away each year from heart disease, according to the World Health Organization. Only a healthy lifestyle and early prediction can stop heart-related diseases.The main objective of this study is to give clinicians a tool to identify cardiac disease at an early stage. As a result, patients will receive effective care and serious consequences will be avoided. In order to uncover hidden discrete patterns and analyse the provided data, ML plays a crucial role. Following data analysis, machine learning approaches aid in the early detection and prediction of cardiac disease.

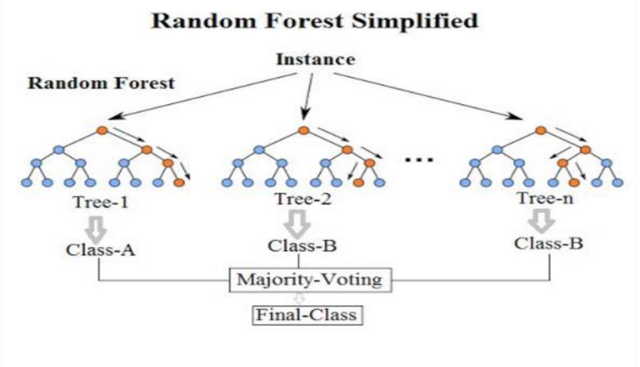
2. Related Works

Data mining has been used by some researchers to make predictions about cardiac disorders. In their research, Kaur et al. explain how the intriguing pattern and knowledge are gleaned from the sizable dataset. The results favour SVM when they compare the accuracy of several machine learning and data mining techniques to determine which is the best. SVM was found to be the best among the numerous machine learning and data mining algorithms that Kumar et al. had worked on; other algorithms included naive Bayes, KNN, and decision tree. These algorithms were trained by the UCI machine learning dataset, which has 303 samples and 14 input features. Using CAD technology, Gavhane et al. have developed a multi-layer perceptron model for the prediction of human cardiac illnesses and the accuracy of the method.If more people use prediction systems to diagnose their illnesses, then more people will be aware of the ailments, which will lower the death rate for cardiac patients. One or two illness prediction algorithms are the work of some researchers. In their project, Krishnan et al. demonstrated that the decision tree classification technique is more accurate than the naive bayes algorithm. Many researchers have worked on this, including Kohali et al., who used logistic regression to predict heart disease, support vector machines to predict diabetes, and Adaboot classifiers to predict breast cancer. They found that logistic regression had an accuracy of 87.1%, support vector machines had an accuracy of 85.71%, and Adaboot classifiers had an accuracy of up to 98.5 percent. A survey report on the prediction of cardiac ailments has demonstrated that hybridization performs well and provides better prediction accuracy than the older machine learning techniques.

3. Classification

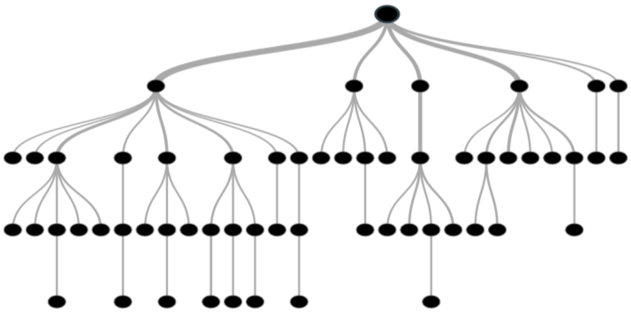
Eighty percent of the input dataset is used as training data, and the remaining twenty percent is used as test data. The dataset used to train a model is referred to as the training dataset. The performance of the trained model is evaluated using the testing dataset. Accuracy, precision, recall, and F-measure scores are just a few of the various metrics that are used to calculate and analyse the performance of each algorithm. The various algorithms that were investigated in this study are given below.

Random Forest: Regression and classification both employ Random Forest methods. The data is organised into a tree, and predictions are based on that tree. Even with a substantial number of record values missing, the Random Forest algorithm can still produce the same results when applied to huge datasets. The decision tree's generated samples can be preserved and used to different sets of data. In random forest, there are two stages: first, generate a random forest, and then, using a classifier produced in the first stage, make a prediction.



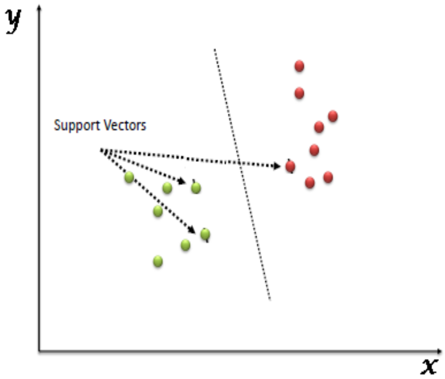
Decision Tree: A classification system known as a decision tree can be used with both category and numerical data. Tree-like structures are made using decision trees. A decision tree is a straightforward and popular tool for managing medical datasets. The data in a tree-shaped graph can be easily implemented and analysed. Three nodes serve as the basis for the decision tree model's analysis.

- Root node: main node, based on this all other nodes functions.
- Interior node: handles various attributes.
- Leaf node: represent the result of each test.



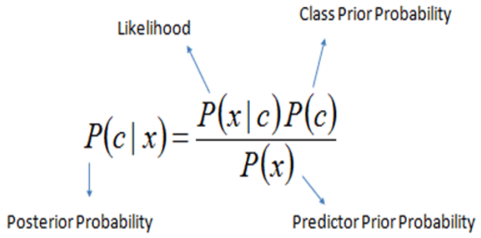
The outcomes are simpler to read and understand. As it analyses the dataset in the tree-like graph, this algorithm is more accurate than other algorithms. However, since only one attribute is checked at a time for decision-making, the data may be overclassified. The accuracy gained was extremely low, about 42.8954%, while Chauhan et aldecision .'s tree reached an accuracy of 71.43%.

Support Vector Machine: It falls within the domain of machine learning techniques that utilise the hyperplane concept, which classifies the data by establishing hyperplanes between it. (Yi, Xi) is a training sample dataset where I = 1, 2, 3, etc. and Xi = ith vector and Yi = target vector. The quantity of hyperplanes determines the kind of support vector; for instance, if a line is employed as a hyperplane, the technique is known as a linear support vector.



D.K-nearest Neighbour: The K-Nearest Neighbor rule, a nonparametric method for classifying patterns, was first described by Hodges et al. in 1951. One of the simplest but most powerful categorization methods is the K-Nearest Neighbor algorithm. It is typically used for classification jobs where there is little to no prior knowledge about the distribution of the data and makes no assumptions about the data. With this approach, the data point for which a target value is unavailable is located together with the k nearest data points in the training set, and the average value of those data points is then applied to that data point.When utilising the 10-cross validation procedure with a value of k equal to 9, KNN provides an accuracy of 83.16%. Ant Colony Optimization outperforms other methods in KNN with a 70.26% accuracy rate and a 0.526 error rate. A highly respectable efficiency of 87.5% was attained by Ridhi Saini et al.

E.Naïve Bayes: Naive Bayes is a simple but an effective classification technique which is based on the Bayes Theorem. It assumes independence among predictors, i.e., the attributes or features should be not correlated to one another or should not, in anyway, be related to each other. Even if there is reliance, each of these characteristics or properties still individually affects likelihood, which is why it is referred to as naive. This approach for supervised machine learning assumes that features are statistically independent of one another according to the Bayes' Theorem. When the input data is highly dimensional, the Naive Bayes Classifier is utilised. The naive Bayes approach has many applications in computer vision. Particularly, it has demonstrated that it is a classifier with effective outcomes.



$$P(c\,|\,x)=\frac{P(x\,|\,c)P(c)}{P(x)}$$

$$P(c\,|\,\mathrm{X}) = P(x_1\,|\,c)\times P(x_2\,|\,c)\times\cdots\times P(x_n\,|\,c)\times P(c)$$

.

4. Conclusion

According to the review above, machine learning algorithms have enormous potential for predicting cardiovascular disorders and heart-related ailments. Each of the aforementioned algorithms has done incredibly well in some situations while failing miserably in others. When combined with PCA, alternating decision trees have demonstrated exceptional performance, yet in other situations, decision trees have demonstrated extremely poor performance, which may be the result of overfitting. Because they employ numerous algorithms to address the issue of overfitting, Random Forest and Ensemble models have demonstrated excellent performance (multiple Decision Trees in case of Random Forest). Naive Bayes classifier-based models had excellent performance and were computationally quick. SVM excelled in the majority of the cases.Systems based on machine learning algorithms and techniques have been very accurate in predicting the heart related diseases but still there is a lot scope of research to be done on how to handle high dimensional data and overfitting. It is also possible to conduct extensive research on the ideal set of algorithms to employ for a specific kind of data.

References

[1]. Ramadoss and Shah B et al."A. Responding to the threat of chronic diseases in India". Lancet. 2005; 366:1744–1749. doi: 10.1016/S0140-6736(05)67343-6.

[2]. Global Atlas on Cardiovascular Disease Prevention and Control. Geneva, Switzerland: World Health Organization, 2011

[3]. Dhomse Kanchan B and Mahale Kishor M. et al. "Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis", 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication.

[4]. R.Kavitha and E.Kannan et al. "An Efficient Framework for Heart Disease Classification using Feature Extraction and Feature Selection Technique in Data Mining ", 2016

[5]. Shan Xu ,Tiangang Zhu, Zhen Zang, Daoxian Wang, Junfeng Hu and Xiaohui Duan et al. "Cardiovascular Risk Prediction Method Based on CFS Subset Evaluation and Random Forest Classification Framework", 2017 IEEE 2nd International Conference on Big Data Analysis.

[6]. Manpreet Singh, Levi Monteiro Martins, Patrick Joanis and Vijay K. Mago et al. " Building a Cardiovascular Disease Predictive Model using Structural Equation Model & Fuzzy Cognitive Map", 978-1-5090-0626-7/16/$31.00 c 2016 IEEE.

[7]. Kanika Pahwa and Ravinder Kumar et al. "Prediction of Heart Disease Using Hybrid Technique For Selecting Features", 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON).

[8]. Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez et al. " A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease", 22nd IEEE Symposium on Computers and Communication (ISCC 2017): Workshops - ICTS4eHealth 2017

[9]. Hanen Bouali and Jalel Akaichi et al. "Comparative study of Different classification techniques, heart Diseases use Case.", 2014 13th International Conference on Machine Learning and Applications

[10]. Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez et al. " A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease", 22nd IEEE Symposium on Computers and Communication (ISCC 2017): Workshops - ICTS4eHealth 2017

[11]. Houda Mezrigui, Foued Theljani and Kaouther Laabidi et al. "Decision Support System for Medical Diagnosis Using a Kernel-Based Approach", ICCAD'17, Hammamet - Tunisia, January 19-21, 2017.

[12]. Dr.(Mrs).D.Pugazhenthi, Quaid-E-Millath and Meenakshi et al. "Detection Of Ischemic Heart Diseases From Medical Images " 2016 International Conference on Micro-Electronics and Telecommunication Engineering.

[13]. J. Hodges et al. "Discriminatory analysis, nonparametric discrimination: Consistency properties," 1981.

[14]. S.Rajathi and Dr.G.Radhamani et al. "Prediction and Analysis of Rheumatic Heart Disease using kNN Classification with ACO ", 2016.

[15]. Puneet Bansal and Ridhi Saini et al. "Classification of heart diseases from ECG signals using wavelet transform and kNN classifier", International Conference on Computing, Communication and Automation (ICCCA2015).

[16]. Simge EKIZ and Pakize Erdogmus et al. "Comparitive Study of heart Disease Classification", 978-1-5386-0440-3/17/$31.00 ©2017 IEEE.

[17]. Renu Chauhan, Pinki Bajaj, Kavita Choudhary and Yogita Gigras et al. "Framework to Predict Health Diseases Using Attribute