International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

Heart Disease Prediction System

Sandhya Kumari[1], Dr. R. Viswanathan[2]

[1, 2]Computer Science and Engineerinig, Galgotias University, Greater Noida, India

Abstract: ***Cardiac disease is a major cause of death throughout the world. It is hard to know or predict by medical practitioners as it requires expertise and higher knowledge of prediction. The environment in healthcare sector is information rich but it have lacks of knowledge. A lot of data is available in healthcare systems over the internet but there is a lack of effective analysis tool to discover hidden patterns in data. An automated or self-made system will increase medical efficiency and decrease cost and time. This software goal to know or predict the occurrence of a disease based on the data which is collected from Kaggle. The Main aim is to extract the hidden patterns by applying Machine Learning techniques on the dataset and to predict or know the presence value on a scale. The prediction of heart disease requires a large size of data which is too massive and complex to process and Analysis by conventional technique. Our goal is to find out an suitable technique that is efficient and accurate for prediction of cardiac disease.***

Keywords: prediction, cardiac disease, machine learning, algorithms, analysis

1. Introduction

Data processing process involves mining/extracting of very significant, hidden and valuable information from large databases [1]. Usually the Healthcare sector involves abundant of knowledge regarding patients, various diagnosis of the diseases etc… [2]. Nowadays the hospitals are adopting the culture of hospital IMS (information management systems) so on handle their or Patients data systematically and effectively. [3]. large quantity of knowledge is produced by such systems that represented using charts, numbers, text and pictures. Though such quite data is hardly employed for creating any clinical Decisions[4]. This research focuses on heart condition diagnosis. Many techniques of knowledge processing are incorporated for diagnosing the disease thereby obtaining several probabilities [5].Concerning or regarding the middle disease prediction large systems are being recommended which are being deployed by the means of various techniques and algorithms. Gaining quality service at affordable price remains the prime and challenging concern for the health protection establishments. For offering quality services at per, there must be accurate or exact diagnosis of the patients in conjunction with effective dosage of medicines. Inferiority clinical diagnosis and treatment can yield in undesired and improper results. One solution for cut by Health protection establishments are often utilization of Computer generated data or use of DSS (decision support systems). Usually the Health protection sector involves abundant of knowledge regarding patients, various diagnosis of the diseases, resource management etc. This information or data must be further weakened by the Human services. Using computerized system, patients treatments records are often stored and using machine learning methods or algorithms one can acquire significant information or data and an odd queries concerning or regarding the hospital. Supervised and unsupervised learning are the two methods. Supervised learning involves usage of coaching for learning model parameters where else no training set is required in unsupervised learning. Classification and prediction are the essential approach of knowledge machine learning. The Classification models helps in classifying or predicting distinct, disorganized data values on the opposite hand prediction model anticipated values that are continuous. Following are the stages within the proposed approach: user registration and login supported Application, dataset collection, classification via Navies Bayesian, prediction and secure data transfer by the means of AES (Advanced Encryption Standard) and lastly output in PDF format. AES helps in transmitting user data to the database during a secured manner. From the safety point of view, patient's personalized data is replaced with some trial values. The study considers and employs medicine datasets performances for predicting cardiac condition in contrast to other Machine Learning techniques.

2. Literature Review

Monika Gandhi et.al, [1] used Naïve Bayes, Decision tree and neural network algorithms and analyses the medical dataset. There are many number of features involved. So, there is a need to scale back the amount of features. it will be done by feature selection. On doing this, they assert that point is decreased. They made use of decision tree and neural networks. J Thomas, R Theresa Princy [2] made use of K nearest neighbor algorithm, neural network, naïve Bayes and decision tree for heart health prediction. They made use of knowledge of machine learning algorithms to detect the guts disease risk rate. Sana Bharti, Shailendra Narayan Singh [3] made use of Particle Swarm Optimization, Artificial neural network, Genetic algorithm for prediction. Associative classification may be a new and efficient technique which integrates association rule machine learning and classification to a model for prediction and achieved good accuracy or average. Purushottam et.al, [4] proposed "An automated or self-made system in diagnosis would enhance medical aid and it also can decrease costs. During this study, we've developed a system which will efficiently or perfectly discover the principles to predict the danger level of patients supported the given parameter about their health. The principles are often prioritized supported the Clients requirement. The performance of the system is evaluated in terms of classification accuracy and therefore the results shows that the model has great potential in predicting the guts disease risk level more accurately or efficiently".

Sellappan Palaniyappan, Rafiah Awang [5] made use of decision tree Naïve Bayes, Decision tree, Artificial Neural Networks to create Intelligent heart health Prediction

Systems (IHHPS).To increase visualization and simple interpretation, it shows the results both in tabular and graphical forms. By providing effective treatments, it also helps to scale back treatment costs. Discovery or possibility of hidden patterns and relationships often has gone unexploited. Advanced Machine Learning techniques helped remedy this example. Himanshu Sharma, M A Rizvi [6] made use of Decision tree, support vector machine, deep learning, K nearest neighbour algorithms. Since the datasets contain or store noise, they tried to scale back the noise by algorithms and pre-processing the dataset and also tried to scale back the dimensionality of the dataset. They found that good accuracy or average is often achieved with neural networks. Animesh Hazra et.al, [7] discussed intimately the disorder and different symptoms of attack. The various sorts of classification and clustering algorithms and tools were used. V.Krishnaiah, G.Narsimha, N.Subhash Chandra [8] presented an analysis using Machine Learning. The analysis showed or Predict that using different techniques or algorithms and taking different number of attributes gives different accuracies or average for predicting heart diseases. J. Vijayashree and NCh. Sriman Narayana Iyengar [10] used machine learning. an huge amount of knowledge or skill developed on a day to day. As such, it can't be interpreted manually. Machine Learning is often effectively wont to predict diseases from these datasets. During this paper, different machine learning algorithms are analysis on heart health database. last, this paper analyses and compares how different-different classification algorithms work on a heart health database.

3. Proposed Methodology

In this paper, comparison of machine learning methods is completed for predicting the ten year risk of coronary heart condition of the patients from their medical data. The subsequent is that the flowchart for proposed methodology.

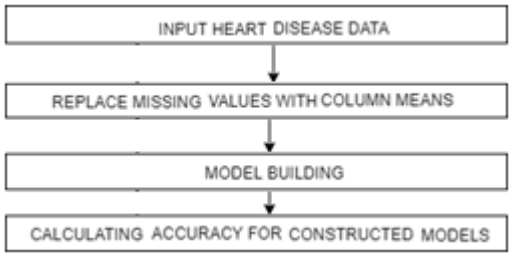

Figure: Proposed Work

The cardiac disease data set is taken as input. It's then pre-processed by replacing non-available values with column means. Two different methods were utilized during this paper. The methods used are depicted in figure 3..The output is that the accuracy matrices of the machine learning models. The model can then be utilized in prediction.

K-Nearest Neighbour (KNN) KNN is a non-parametric machine learning algorithm. The KNN algorithm is a supervised learning method. This means that all the data is labelled and the algorithm learns to predict the output from the input data. It performs well even if the training data is large and contains noisy values. The data is divided into training and test sets. The train set is used for model building and training. A k- value is decided which is often the square root of the number of observations. Now the test data is predicted on the model built. There are different distance measures. For continuous variables, Euclidean distance, Manhattan distance and Minkowski distance measures can be used. However, the commonly used measure is Euclidean distance. The formula for Euclidean distance is as follows:

$$d = \sqrt{\sum_{i=1}^{k} (xi - yi)2}$$

The ROC curve for k-nearest neighbour is depicted in figure 5.

Decision Tree- Decision trees is one of the ways to display an algorithm. it is a classic machine learning algorithm. In heart disease, there are several factors like cigarette, BP, Hypertension, age etc. The challenge of the selection tree lies within the choice of the idea node. This factor utilized in root node must clearly classify the data. We make use aged because the basis node. The ROC curve is given in figure 4.The decision tree is simple to interpret. They're non-parametric which they implicitly do feature selection.

4. Result Analysis and Comparison

The dataset used is Framingham taken from Kaggle [17].

There were 16 attributes were as follows: Male – gender 0 for female and 1 for male, Age – age of the patient, Education – values 1-5, education of the patient. Current smoker – 1 if current smoker and 0 otherwise, Cigarette per day – if current smoker then number of cigarette per day, BP Meds – vital sign, Prevalent BP – prevalent vital sign, Prevalent Hyp – prevalent hyper tension, Diabetes – 1 if diabetes 0 otherwise, Total cholesterol – cholesterol level, Sys BP – systolic vital sign, Dia BP – diastolic vital sign, BMI – body mass index, pulse – pulse or pulse of the patient, Glucose – glucose level, Ten Year CHD – has chronic heart condition or not. The machine learning models is evaluated using the AUC-ROC metric. This will be used to understand the model performance.

The ROC curve of the algorithms is as follows:

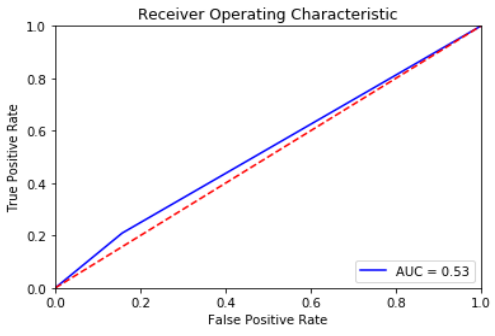

Figure 3: ROC Curve for Decision Tree

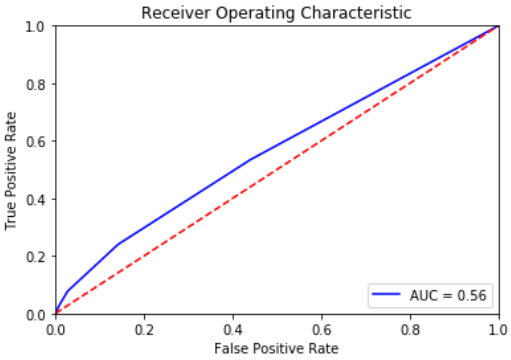

Figure 4: ROC Curve For KNN

The ROC curve is the Receiver Operating Characteristic curve. The AUC is the area under the ROC curve. If the AUC score is high, the model performance is high and vice versa. The figures 4, 5, 6 and 7 gives the ROC curve of the machine learning algorithms. The comparison of AUC score of the various algorithms is as follows:

Algorithm	*AUC score*
KNN	0.56
Decision Tree	0.53

The accuracy of the algorithms is calculated. The accuracy results are tabulated as follows:

Method	*Accuracy*
KNN	86.30%
Decision Tree	77.58%

5. Conclusion

This paper discusses the various machine learning algorithms such as decision tree and k- nearest neighbour which were applied to the data set. It utilizes the data such as blood pressure, cholesterol, diabetes and then tries to predict the possible in duration cardiac disease patient in next 10years. Family history of heart disease can also be a reason for developing or increasing a heart disease as mentioned earlier. So, this data of the patient can also be used for further increasing the accuracy of the model. This work will be helpful in recognising the possible patients who may suffer from cardiac disease in the next 10 years. This Prediction may help in taking preventive measures and hence try to avoid or decrease the chances of Cardiac disease for the patient. So when a patient is predicted as positive for Cardiac disease, then the medical data for the patient can be closely or nearly analysed by the doctors. An example would be - suppose the patient has diabetes which may be the cause for cardiac disease in future and then the patient can be given treatment to have diabetes in control which in turn may prevent the Cardiac disease.

The Cardiac disease prediction can be done using other machine learning algorithms. Logistic regression also perform well in case of binary classification problems such as cardiac disease prediction. Random forests can perform well than decision trees. Also, the ensemble methods and artificial neural networks can be applied to the data set. The results can be compared and improved also.

References

[1] Gandhi, Monika & Singh, Shailendra. (2015). Predictions in heart disease using techniques of Machine Learning. 2015 1st International Conference on Futuristic Trends in Computational Analysis and Knowledge Management, ABLAZE 2015. 520-525. 10.1109/ABLAZE.2015.7154917.

[2] Thomas,J.&Princy, R. (2016). Human heart disease prediction system using Machine Learning techniques. 1-5. 10.1109/ICCPCT.2016.7530265.

[3] S. Bharti and S. N. Singh, "Analytical study of heart disease prediction comparing with different algorithms," International Conference on Computing, Communication & Automation, Noida, 2015, pp. 78-82.

[4] Purushottam, & Saxena, Kanak & Sharma, Richa. (2015). Efficient heart disease prediction system using decision tree. International Conference on Computing, Communication and Automation, ICCCA 2015. 72-77. 10.1109/CCAA.2015.7148346.

[5] Mat Ghani, Mohd.& Awang, Raflah. (2008). Intelligent heart disease prediction system using Machine Learning.8.108 – 115.10.1109/AICCSA.2008.4493524

[6] SharmaHimanshu. “Prediction of Heart Disease using Machine Learning Algorithms: A Survey.” (2017).

[7] Hazra, Animesh& Mandal, Subrata & Gupta, Amit & Mukherjee, Arkomita& Mukherjee, Asmita. (2017). Heart Disease Diagnosis and Prediction Using Machine Learning : A Review. Advances in Computational Sciences and Technology. 10. 2137-2159.

[8] V Krishnaiah, G Narsimha and Subhash N Chandra. Article: Heart Disease Prediction System using Machine Learning Techniques and Intelligent Fuzzy Approach: A Review. *International Journal of Computer Applications* 136(2):43-51, February 2016. Published by Foundation of Computer Science (FCS), NY, USA

[9] Kaur, Ramandeep and Er. Prabhsharn Kaur. “A Review-Heart Disease Forecasting Pattern using Various Machine Learning.” (2016).

[10] Vijayashree, J. &Iyenger, N Ch Sriman Narayana. (2016). Heart Disease Prediction System Using Machine Learning: A Review. International Journal of Bio-Science and Bio-Technology. 8. 139-148. 10.14257/ijbsbt.2016.8.4.16.