

Heart disease prediction using machine learning algorithms

Abstract :- Day by day the cases of heart conditions are adding at a rapid-fire rate and it's veritably Important and concerning to prognosticate any similar conditions beforehand. This opinion is a delicate task i.e. it should be performed precisely and efficiently. The exploration paper substantially focuses on which case is more likely to have a heart complaint grounded on colorful medical attributes. We prepared a heart complaint. Vaticination system to prognosticate whether the case is likely to be diagnosed with a heart complaint or not using the medical history of the case. We used different algorithms of machine learning similar as logistic retrogression and KNN to prognosticate and classify the case with heart complaint. A relatively Helpful approach was used to regulate how the model can be used to ameliorate the delicacy of vaticination of Heart Attack in any existent. The strength of the proposed model was quiet satisfying and was suitable to prognosticate substantiation of having a heart complaint in a particular individual by using KNN and Logistic Retrogression which showed a good delicacy in comparison to the preliminarily used classifier similar as naive bayes etc. So a quiet significant quantum of pressure has been lift off by using the given model in chancing the probability of the classifier to rightly and directly identify the heart complaint. The Given heart complaint vaticination system enhances medical care and reduces the cost. This design gives us significant knowledge that can help us prognosticate the cases with heart complaint It's enforced on the. pynb format.

Introduction

"Machine Learning may be a way of Manipulating and extraction of implicit, previously unknown/known and potential useful information about data" [1]. Machine Learning may be a very vast and diverse field and its scope and implementation is increasing day by day. Machine learning Incorporates various classifiers of Supervised, Unsupervised and Ensemble Learning which are wont to predict and Find the Accuracy of the given dataset. we will use that knowledge in our project of HDPS because it will help tons of individuals. Cardiovascular diseases are quite common lately , they describe a variety of conditions that would affect your heart. World health organization estimates that 17.9 million global deaths from (Cardiovascular diseases) CVDs [2].

It is the first reason of deaths in adults. Our project can help predict the people that are likely to diagnose with a heart condition by help of their medical record [6]. It recognizes who all are having any symptoms of heart disease like pain or high vital sign and may help in diagnosing disease with less medical tests and effective treatments, in order that they will be cured accordingly. This project focuses on mainly three data processing techniques namely: (1) Logistic regression, (2) KNN and (3) Random Forest Classifier. The accuracy of our project is 87.5% that is best than previous system where just one data processing technique is employed . So, using more data processing techniques increased the

HDPS accuracy and efficiency. Logistic regression falls under the category of supervised learning. Only discrete values are utilized in logistic regression.

The objective of this project is to see whether the patient is probably going to be diagnosed with any cardiovascular heart diseases supported their medical attributes like gender, age, pain , fasting sugar

level, etc. A dataset is chosen from the UCI repository with patient's medical record and attributes. By using this dataset, we predict whether the patient can have a heart condition or not. To predict this, we use 14 medical attributes of a patient and classify him if the patient is probably going to possess a heart condition . These medical attributes are trained under

three algorithms: Logistic regression, KNN and Random Forest Classifier. Most efficient of those algorithms is KNN which provides us the accuracy of 88.52%. And, finally we classify patients that are in danger of getting a heart condition or not and also this method is completely cost efficient.

Related Work

A quiet Significant quantum of work related to the opinion of Cardiovascular Heart complaint using Machine Literacy algorithms has motivated this work. This paper contains a brief literature check. An effective Cardiovascular complaint vaticination has been made by using colorful algorithms some of them include Logistic Regression, KNN, Random Forest Classifier Etc. It can be seen in Results that each algorithm has its strength to register the defined objects (7). The model incorporating IHDPDS had the capability to calculate the decision boundary using the former and new model of machine literacy and deep literacy. It eased the important and the most introductory factors/ knowledge similar as family history connected with any heart complaint. But the delicacy that was Attained in similar IHDPDS model was far more lower than the new forthcoming model similar as detecting coronary heart complaint using artificial neural network and other algorithms of machine and deep literacy. The threat factors of coronary Heart complaint or atherosclerosis is linked by McPherson et al., (8) using the inbuilt

perpetration algorithm using uses some ways of Neural Network and were just directly suitable to prognosticate whether the test case is suffering from the given complaint or not.

Opinion and vaticination of Heart Disease and Blood Pressure along with other attributes using the aid of neural networks was introduced by R. Subramanian et al., (24). A deep Neural Network was Erected

incorporating the given attributes related to the complaint which were suitable to produce a affair which was carried out by the affair perceptron and nearly included 120 retired layers which is the introductory and utmost Applicable fashion of icing a accurate result of having heart complaint if we use the model for Test Dataset.

The supervised network has been advised for opinion of heart conditions (16). When the testing of the model was done by a croaker using an strange data, the model used and trained from the former learned data and prognosticated the result thereby calculating the delicacy of the given model.

Data Source

An Organized Dataset of individualities had been named Keeping in mind their history of heart problems and in agreement with other medical conditions (2). Heart complaint are the different conditions by which the heart is affected. According to World Health Organization (WHO), the topmost number of deaths in middle aged people are due to Cardiovascular conditions. We take a data source which is comprised of medical history of 304 different case of different age groups. This dataset gives us the important-required information i.e. the medical attributes similar as age, resting blood pressure, dieting sugar position etc. of the case that helps us in detecting the case that's diagnosed with any heart complaint or not. This dataset contains 13 medical attributes of 304 cases that helps us detecting if the case is at threat of getting a heart complaint or not and it helps us classify cases that are at threat of having a heart complaint and that who aren't at threat. This Heart Complaint dataset is taken from the UCI depository. According to this dataset, the pattern which leads to the discovery of patient prone to getting a heart complaint is uprooted. These records are resolve into two

corridor Training and Testing. This dataset contains 303 rows and 14 columns, where each row corresponds to a single record. All attributes are listed in ‘Table 1’.

Table 1. Various Attributes used are listed

S. No	Observation	Value
1.	Age	Age in Years
2.	Sex	Male/Female
3.	CP	Chest Pain
4.	TP	Resting Blood Pressure
5.	Chol	Serum Cholesterol
6.	FBS	Fasting Blood Sugar
7.	Resting	Resting Electrocardiograph
8.	Thalach	Maximum Heart Rate Achieved
9.	Exang	Exercise Induced Angina
10.	Oldpeak	ST Depression when Workload compared to the Amount of Rest Taken
11.	Slope	Slope of Peak Exercise ST segment
12.	Ca	Calcium the number of Major Vessels Coloured by Fluorography
13.	Thal	Thalium Stress Test Result
14.	Heart Disease	Heart Disease

Methodology

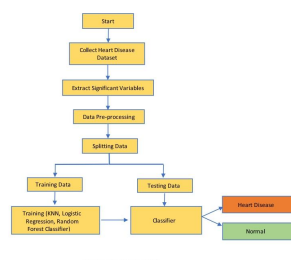
This paper shows the analysis of various machine learning algorithms, the algorithms that are used in this

paper are K nearest neighbors (KNN), Logistic Regression and Random Forest Classifiers which can be helpful for practitioners or medical analysts for accurately diagnose Heart Disease. This paperwork includes examining the journals, published paper and the data of cardiovascular disease of the recent times.

Methodology gives a framework for the proposed model [13]. The methodology is a process which includes steps that transform given data into recognized data patterns for the knowledge of the users. The proposed methodology (Figure 1.) includes steps, where first step is referred as the collection of the data than in

second stage it extracts significant values than the 3rd is the preprocessing stage where we explore the data. Data preprocessing deals with the missing values, cleaning of data and normalization depending on algorithms used [15]. After pre-processing of data, classifier is used to classify the pre-processed data the

classifier used in the proposed model are KNN, Logistic Regression, Random Forest Classifier. Finally, the proposed model is undertaken, where we evaluated our model on the basis of accuracy and performance using various performance metrics. Here in this model, an effective Heart Disease Prediction System (EHDPs) has been developed using different classifiers. This model uses 13 medical parameters such as chest pain, fasting sugar, blood pressure, cholesterol, age, sex etc. for prediction [17].



Results and Discussions

From these results we can see that although utmost of the experimenters are using different algorithms similar as SVC, Decision tree for the discovery of cases diagnosed with Heart complaint, KNN, Random Forest Classifier and Logistic regression yield a better result to out rule them (23). The algorithms that we used are more accurate, saves a lot of plutocrat i.e. it's cost effective and briskly than the algorithms that the former Experimenters used.

Also, the maximum delicacy attained by KNN and Logistic Regression are equal to 88.5 which is lesser or nearly equal to rigor attained from former inquiries. So, we epitomize that our delicacy is bettered due to the increased medical attributes that we used from the dataset we took. Our design also tells us that Logistic Retrogression and KNN outperforms Random Forest Classifier in the vaticination of the case diagnosed with a heart Complaint. This proves that KNN and Logistic Regression are better in opinion of a heart complaint. The following ' figure 2', ' figure 3', ' figure 4', ' figure 5' shows a plot of the number of cases that are been insulated and prognosticated by the classifier depending upon the age group, Resting Blood Pressure, Coitus, Casket Pain.

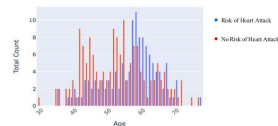


Figure 2. Shows the Risk of Heart Attack on the basis of their age.



Figure 3. Shows the Risk of Heart Attack on the basis of their Resting Blood Pressure.

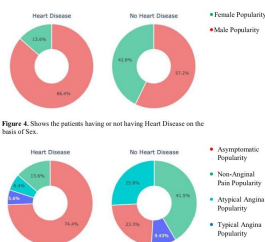


Figure 4. Shows the patients having or not having Heart Disease on the basis of Sex.

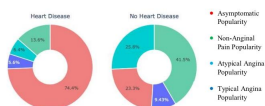
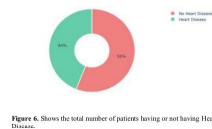


Figure 5. Shows the patients having or not having Heart Disease on the basis of type of Chest Pain.

Conclusion

A cardiovascular complaint discovery model has been developed using three ML bracket modelling ways. This design predicts people with cardiovascular complaint by rooting the patient medical history that leads to a fatal heart complaint from a dataset that includes cases' medical history similar as casket pain, sugar position, blood pressure, etc. This Heart Disease discovery system assists a case grounded on his/ her clinical information of them been diagnosed with a former heart complaint. The algorithms used in erecting the given model are Logistic retrogression, Random Forest Classifier and KNN (22). The delicacy of our model is 87.5. Use of further training data ensures the advanced chances of the model to directly prognosticate whether the given person has a heart complaint or not (9). By using these, computer backed ways we can Prognosticate the case presto and better and the cost can be reduced veritably much. There are a number of medical databases that we can work on as these Machine literacy ways are better and they can prognosticate better than a mortal being which helps the case as well as the croakers. Thus, in conclusion this design helps us prognosticate the cases who are diagnosed with heart conditions by drawing the dataset and applying logistic retrogression and KNN to get an delicacy of an normal of 87.5 on our model which is better than the Former models having an delicacy of 85. Also, it's concluded that delicacy of KNN is

loftiest between the three algorithms that we've used i.e.88.52.' Figure 6 shows 44 of people that are listed in the dataset are suffering from Heart Disease.



References

- [1] Soni J, Ansari U, Sharma D & Soni S (2011). Predictive data mining for medical diagnosis: an overview of heart disease prediction. *International Journal of Computer Applications*, 17(8), 43-8
- [2] Dangare C S & Apte S S (2012). Improved study of heart disease prediction system using data mining classification techniques. *International Journal of Computer Applications*, 47(10), 44-8.
- [3] Ordonez C (2006). Association rule discovery with the train and test approach for heart disease prediction. *IEEE Transactions on Information Technology in Biomedicine*, 10(2), 334-43.
- [4] Shinde R, Arjun S, Patil P & Waghmare J (2015). An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm. *International Journal of Computer Science and Information Technologies*, 6(1), 637-9.
- [5] Bashir S, Qamar U & Javed M Y (2014, November). An ensemble-based decision support framework for intelligent heart disease diagnosis. In *International Conference on Information Society (i-Society 2014)* (pp. 259-64). IEEE.
- [6] Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014). A coronary heart disease prediction model: the Korean Heart Study. *BMJ open*, 4(5), e005025.
- [7] Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Ärnlöv J & Ingelsson E (2013). Multilocus genetic risk scores for coronary heart disease prediction. *Arteriosclerosis, thrombosis, and vascular biology*, 33(9), 2267-72.
- [8] Jabbar M A, Deekshatulu B L & Chandra P (2013, March). Heart disease prediction using lazy associative classification. In *2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)* (pp. 40- 6). IEEE.
- [9] Dangare Chaitrali S and Sulabha S Apte. "Improved study of heart disease prediction system using data mining classification techniques." *International Journal of Computer Applications* 47.10 (2012): 44-8.
- [10] Soni Jyoti. "Predictive data mining for medical diagnosis: An overview of heart disease prediction." *International Journal of Computer Applications* 17.8 (2011): 43-8.
- [11] Chen A H, Huang S Y, Hong P S, Cheng C H & Lin E J (2011, September). HDPS: Heart disease prediction system. In *2011 Computing in Cardiology* (pp.557-60). IEEE.
- [12] Parthiban, Latha and R Subramanian. "Intelligent heart disease prediction system using CANFIS and genetic algorithm." *International Journal of Biological, Biomedical and Medical Sciences* 3.3 (2008).
- [13] Wolgast G, Ehrenborg C, Israelsson A, Helander J, Johansson E & Manefjord H (2016). Wireless body area network for heart attack detection [Education Corner]. *IEEE antennas and propagation magazine*, 58(5), 84-92.

- [14] Patel S & Chauhan Y (2014). Heart attack detection and medical attention using motion sensing device -kinect. International Journal of Scientific and Research Publications, 4(1), 1-4.
- [15] Zhang Y, Fogoros R, Thompson J, Kenknight B H, Pederson M J, Patangay A & Mazar S T(2011). U.S. Patent No. 8,014,863. Washington, DC: U.S. Patent and Trademark Office.
- [16] Raihan M, Mondal S, More A, Sagor M O F, Sikder G, Majumder M A & Ghosh K (2016,December). Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design. In 2016 19th International Conference on Computer and Information Technology (ICCIT) (pp. 299-303). IEEE.
- [17] Buechler K F & McPherson P H (1999). U.S. Patent No. 5,947,124. Washington, DC: U.S. Patent and Trademark Office.
- [18] Takci H (2018). Improvement of heart attack prediction by the feature selection methods. Turkish Journal of Electrical Engineering & Computer Sciences,26(1), 1-10.
- [19] Worthen W J, Evans S M, Winter S C & Balding D (2002). U.S. Patent No. 6,432, 124. Washington, DC: U.S. Patent and Trademark Office.
- [20] Acharya U R, Fujita H, Oh S L, Hagiwara Y, Tan J H & Adam M (2017). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Information Sciences, 415, 190-8.
- [21] Brown N, Young T, Gray D, Skene A M & Hampton J R (1997). Inpatient deaths from acute myocardial infarction, 1982-92: analysis of data in the Nottingham heart attack register. BMJ, 315(7101), 159-64.
- [22] Piller L B, Davis B R, Cutler J A, Cushman W C, Wright J T, Williamson J D & Haywood L J (2002). Validation of heart failure events in the Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) participants assigned to doxazosin and chlorthalidone. Current controlled trials in cardiovascular medicine, 3(1), 10.
- [23] Folsom A R, Prineas R J, Kaye S A & Soler J T (1989). Body fat distribution and self-reported prevalence of hypertension, heart attack, and other heart disease in older women. International journal of epidemiology, 18(2), 361-7.
- [24] Kiyasu J Y (1982). U.S. Patent No. 4,338,396. Washington, DC: U.S. Patent and Trademark Office