Proceedings of **"National Conference on Recent Trends in Science and Advances in Engineering"**
Organized by Fabtech Technical Campus, College of Engineering & Research, Sangola
International Journal of Innovations in Engineering Research and Technology [IJIERT]
ISSN: 2394-3696, Website: www.ijiert.org, June, 2022

TO STUDY DIFFERENT MACHINE LEARNING ALGORITHMS FOR PREDICTION OF HEART DISEASE

Dhanashri Mahesh Raut
Department of Electronics & Tlecommunication Engineering,Fabtech Technical Campus, College of Engineering and Research, Sangola,Dist.-Solapur,India
* dskraut@gmail.com

Trupti Madan Kulkarni
Department of of Electronics & Tlecommunication Engineering,
Sinhgad College of Engineering, Korti, Pandharpur, India
* kulkarni.trupti2011@gmail.com

ABSTRACT

Heart disease causes a significant mortality rate around the world, and it has become a health threat for many people. Early prediction of heart disease may save many lives; detecting cardiovascular diseases like heart attacks, coronary artery diseases etc., is a critical challenge for regular clinical data analysis. According to the World Health Organization report published in 2019, around 17.9 million people die every year worldwide due to heart disease [1]. There are various types of heart diseases such as coronary artery disease, congenital heart disease, arrhythmia, etc. The patient suffering from heart disease has various symptoms such as chest pain, dizzy sensations, and deep sweating. Smoking, high blood pressure, diabetes, obesity,etc. are the main reasons behind heart disease.

In India, more than 17 Lakh people die every year due to heart diseases and by 2030, the figure is expected to increase with 2.3 crore deaths. Invasive methods of predicting the disease are expensive and painful. Therefore, there is a need for a technique that can predict heart disease in a non-invasive manner at less cost.

Since the prediction of heart disease in people is very important, a method should be used in the right prediction of heart diseases that have the least errors in heart disease prediction.

Hence to overcome & reduce the chances of death occurs due to heart disease, machine learning comes in role to predict the chances of upcoming heart disease according to the health conditions as well as previous medical conditions of the patient.

Machine learning has rapidly developed within the last years and its extension into medical sciences offers the potential to revolutionize the way in which complex diagnostic and prognostic estimations at the level of the individual patient are performed. Machine Learning algorithms such as Random Forest, Support Vector Machine (SVM), Naive Bayes and Decision tree have been used for the development of model.

Keywords: Machine learning, heart disease, Random Forest, Support Vector Machine (SVM), Naive Bayes and Decision tree.

Introduction

Heart disease is one of the foremost vital causes of mortality within the world today. An estimate by the World Health Organization, that over 17.9 milliondeaths occur once a year worldwide due to upset the center may be a muscle which pumps blood from the lungs to the lower a part of the body round the body. If blood circulation within the body be inefficient, another organism like brain suffers from this problem. If the center stops working, death can occur in minutes. Life is completely obsessed on the work of the center, variety of things that increase the danger of cardiopathy include case history, smoking, poor diet, high force per unit area, high blood cholesterol, obesity, physical inactivity.

The most important dangerous factors of heart condition and attack are not diet food, no exercise, tobacco use and harmful use of drinks such as alcohol. This risk factors may show up in individuals as increasing pressure, increasing blood sugar, increasing blood lipids, and increasing weight and obesity. These "intermediate risks factors" will be measured in medical aid facilities and indicatean increased risk of attack, stroke, coronary failure and other complications.

In several cases, the prediction of cardiopathy is usually supported the results of the test of the patient. The prediction of cardiovascular disease could be a complex task that needs high skills and skill, early diagnosis and intensive medical aid of patients with this disease can largely prevents them from overtime.

Machine learning techniques will be accustomed design a choice network to detectcardiovascular disease. The Machine Learning algorithms are designed to perform an outsized number of tasks like prediction, classification, higher cognitive process etc.

Literature Review

There are many works studied through the researchers approximately cardiovascular ailment prediction and evaluation. A range of such works are addressed below.

In this paper[3] Researcher predicts if an person plagued by cardiopathy in phrases of chances the usage of statistics processing class strategies. Algorithm used for this approach is Decision Tree and KNN with the help of this technique coronary heart illnesses expected in current device.

Himanshu Sharma [4] centered on summarised country of artwork technology and to be had methods for prediction of assault ailment. They want used each gadget getting to know and deep getting to know ideas and essentially centered on comparative evaluation of coronary heart circumstance detection.

Kumar [5] Different gadget getting to know algorithms have been implemented for purchasing the effects then as compared with one another. Naïve base classifier is that the first-rate in overall performance.

Salman, Issam [6] especially centered on mortality prediction way to a assault the usage of one-of-a-kind gadget getting to know methods. to complete his paintings he used real-time statistics. For accomplishing higher accuracy he used differing types of classifier NB, Bayesian Network etc.

[7] The statistics of cardiopathy sufferers amassed from the UCI laboratory is hired to locate styles with DT and RF. The effects are as compared for overall performance and accuracy with those algorithms. The proposed hybrid technique returns higher accuracy.

[8] Mainly they`re centered on how early to are expecting coronary heart circumstance the usage of numerous statistics processing strategies and they used nearly four algorithms of gadget getting to know. The attributes they hired of their dataset like age, sex, strain level, and glucose. They carried out higher accuracy from the logistic regression.

[9] Researchers advanced ML primarily based totally prediction device for early prognosis of CVDs in India.The statistics of cardiovascular ailment sufferers amassed from clinic in South India.The take a look at changed into meted out with ML primarily based totally cardiopathy

Classification

- Linear Regression

Linear retrogression evaluation is used to prognosticate the fee of a variable grounded at the fee of any other variable. The variable you would really like to prognosticate is called the variable quantity. The variable you're the use of to prognosticate the opposite variable`s fee is nominated the variable. easy retrogression is often used for prophetic evaluation and modelling. as an illustration, it are often habituated quantify the relative affects old, gender, and diet(the predictor variables) on height(the outgrowth variable). The linearity supposition in direct retrogression manner the version is direct in parameters(i.e quantities of variables) & might also additionally

or won't be direct in variables. Linear retrogression is a gadget for prognosticating y from x. In our case, y is the based variable, and x is the unbiased variable. We need to prognosticate the fee of y for a given fee of x. Now, if the records had been impeccably direct, we should surely calculate the pitch intercept shape of the road in terms y = mxb.

- Logistic Regression

Logistic Retrogression is any other statistical evaluation machine espoused via way of means of Machine literacy. It`s used while our established variable is dichotomous or double. It simply method a variable that has best 2 labors. Logistic retrogression is used to prognosticate the class(or order) of individualities grounded on one or a couple of predictor variables(x). It's used to version a double outgrowth, that may be a variable, which can also additionally have best feasible values zero or 1, sure or no, diseased ornon-diseased. Logistic Retrogression can be a famous set of rules as it converts the values of the log of odds which can also additionally variety from- inf to inf to a variety among zero and 1. Since logistic capabilities affair the opportunity of condition of an condition, they may be carried out to numerous real- existence scripts hence those fashions are usually fashionable. hence those fashions are usually famous.

- Decision Tree

Decision Trees are a type of Supervised Machine literacy in which the records is continuously solve constant with a certain parameter. The tree can be described through manner of method of realities, videlicet desire bumps and leaves. A desire tree may be a veritably precise kind of opportunity tree that permits to offer a choice about some quite method. Decision Trees(DTs) are anon-parametric supervised literacy system used for bracket and retrogression. The difficulty is to form a model that predicts the without a doubt nicely really well worth of a purpose variable through manner of method of mastering clean desire rules inferred from the statistics skills. Decision wooden are used as an method in device literacy to form the set of policies. A desire tree set of policies is probably used to divide dataset skills through a price function. The desire tree is grown in advance than being optimised to get rid of branches that can use inapplicable skills, a method referred to as pruning

- SVM (Support Vector Machine) Algorithm

SVM is a supervised gadget studying set of rules which may be used for type or retrogression issues. It makes use of a gadget known as the kernel trick to transfigure statistics and additionally grounded on those metamorphoses it unearths an most effective boundary among the viable labors. SVC is a nonparametric clustering set of rules that doesn`t make any supposition at the wide variety or form of the clusters withinside the statistics. In our revel in it really works excellent for low- dimensional statistics, so if statistics is high- dimensional, a preprocessing step,e.g. the usage of pinnacle detail analysis, is normally needed.

- Naive Bayes Algorithm

A classifier assumes that the presence of a named element in a completely beauty is unconnected to the presence of the possibility element. Indeed if the ones talents are associated with one another, a classifier may want to probable do not forget all of those parcels singly at the same time as calculating the chance of a named outgrowth.

A model is straightforward to make and useful for large datasets. It`s clean and is concept to outperform in reality in large element today`s bracket patterns.

- KNN (K- Nearest Neighbors) Algorithm

This set of regulations may be enforced to every kind and retrogression problems. suppose, withinside the information Science assiduity, it's far lesser astronomically used to remedy kind problems. It`s an smooth set of rules that shops all reachable instances and classifies any new instances with the aid of using taking a adulthood vote of its ok neighbours. The case is likewise assigned to the elegance with which it`s the maximum in common. A distance feature performs this dimension. KNN are continuously with out issue understood via assessing it to reality. as an illustration, if you need information similarly than one person, it is smart to speak to his or her musketeers and associates.

Discussion

In today's world one in all the foremost common diseases are cardiovascular disease which its mortality and disability is high. Therefore, cardiovascular disease is one in all the biggest health

problems within the world. cardiovascular disease is one among the leading causes of fatality and within the absence of an accurate prediction. it's inferred from the present works that there's a requirement for novelty within the study, and a sturdy, optimized model is required for cardiovascular disease prediction.

Conclusion

In this paper, we've studied various Supervised ML classifiers namely, Random Forest, Decision Tree, SVM, and Logistic Regression are wont to deploy a model for prediction of cardiovascular disease.

This should be utilized in the correct prediction of cardiopathys that have the tiniest quantity errors in heart disease prediction. Machine learning techniques are often efficiently utilised for the prediction of presence or absence of cardiovascular disease. it had been also observed that researchers have tried singletechnique of machine learning.

There is a necessity of development of hybrid machine learning technique whichobtains the best possible accuracy throughout experimenting.

References

[1] World Health Organization, Cardiovascular Diseases, WHO, Geneva, Switzerland, 2020, https://www.who.int/healthtopics/cardiovascular-diseases.

[2] Archive.ics.uci.edu. UCI Machine Learning Repository: Heart Disease Data Set; 2019 [cited 17 December 2019]. Available from: https://archive.ics.uci.edu/ml/datasets/HeartþDisease.

[3] Ahmad Rufai , U. S. Idriss , Mahmood Umar "Using Artificial Neural Networks to Diagnose Heart Disease" International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 19, 2018

[4] Sharma, Himanshu, and M. A. Rizvi. "Prediction of heart disease using machine learning algorithms: A survey." International Journal on Recent and Innovation Trends in Computing and Communication 5.8 (2017): 99-104.

[5] S. Kumar, "Predicting and diagnosing of heart disease using machine learning algorithms", International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 6 Issue 6 June 2017.

[6] Salman, Issam. "Heart attack mortality prediction: an application of machine learning methods" ,Turkish Journal of Electrical Engineering Computer Sciences 27.6 (2019): 4378 4389.

[7] Dr. M. Kavitha, G. Gnaneswar, R. Dinesh, Y. Rohith Sai, R. Sai Suraj"Heart Disease Prediction using Hybrid machine Learning Model" International Conference on Inventive Computation Technologies [ICICT 2021] IEEE Xplore Part Number: CFP21F70-ART; ISBN: 978-1-7281-8501-9

[8] Ekta Maini,Bondu Venkateswarlu, Baljeet Maini, Dheeraj Marwaha "Machine learning based heart disease prediction system for Indian population: An exploratory study done in South India" Medical Journal Armed Forces India, https://doi.org/10.1016/j.mjafi.2020.10.013

[9] Sairabi H. Mujawar, P.R. Devale, "Prediction of heart disease using modified Kmeans and by using naïve bayes", International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 10,October 2015

[10] Singh, A., et .al (2020, February). Heart Disease Prediction Using Machine Learning Algorithms. In 2020 International Conference on Electrical and Electronics Engineering (ICE3) (pp. 452-457). IEEE.

[11] Shahab Tayeb, Matin Pirouz, Johann Sun, Kaylee Hall, Andrew Chang, Jessica Li,Connor Song, Apoorva Chauhan, Medical Ferra, Therasa Sugar, Justin Zhan, Shahram Latifi, "Towards Predicting Medical Condition Using K-Nearest Neighbor", in: 2017 International Conference on big Data. IEEE, 2017.

[12] Debabrata Swain, Santosh Kumar Pani, Debabala Swain "A Metaphoric Investigation on Prediction of Heart Disease using Machine Learning" 978-1-5386-5367,2018 IEEE

[13] Suraj Kumar Gupta, Aditya Shrivastava, S. P. Upadhyay, Pawan Kumar Chaurasia " A Machine Learning Approach for Heart Attack Prediction" International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958 (Online), Volume-10 Issue-6, 2021