International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 19th Feb 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 02

10.48047/IJIEMR/V12/ISSUE 02/76

Title **ANALYSIS OF CARDIOVASCULAR DISEASE**

Volume 12, ISSUE 02, Pages: 492-497

Paper Authors
Miss. Ramsha Syeda Hajira

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

ANALYSIS OF CARDIOVASCULAR DISEASE

Miss. Ramsha Syeda Hajira

UG Scholar, Department of Computer Science,

Stanley Women's College of Engineering and Technology, Hyderabad, Telangana, India.

E-mail: ramsha.saz@gmail.com.

Abstract

In the 21st century, cardiovascular disease has become a serious public health problem on a global scale. The rising cost of treating silent illnesses, such as cardiovascular disease, is putting a strain on national and corporate budgets for human services. Thus, there is a pressing need for early diagnosis and treatment of such conditions. Various mixtures of computations and algorithms are used to the data gleaned through hospital data analysis in order to forecast cardiovascular diseases at an early stage. One of these cutting-edge technologies that is finding widespread use is machine learning, which has several applications, including one in the healthcare industry that may help doctors forecast health problems. With this study, we assessed the performance of several machine learning algorithms for cardiovascular disease risk assessment. Standard machine learning methods including Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), support vector machine (SVM), and Decision Tree form the basis of the proposed experiment. There is usually some kind of connection between things in the environment, and discovering such connections may lead to useful judgements. Also, I'll try to use this data to build a model that can tell at a glance whether or not a patient has cardiovascular disease. In addition, Jupyter Lab is used for the data analysis, which is written in Python, to ensure that all of the Algorithms are correct.

1. INTRODUCTION

The manipulation and extraction of implicit, hitherto unknown/known, and potentially relevant information in data is known as machine learning [1]. Machine learning is a highly broad and complex topic, and its application and breadth are expanding daily. In order to forecast and determine the accuracy of the provided dataset, machine learning incorporates a variety of classifiers from supervised, unsupervised, and ensemble learning. Since content from this effort may be utilised in accordance with the rules of the CVD project, we may use that information. It will be very beneficial to many individuals. These days, a wide variety of disorders that potentially harm your heart are referred to as cardiovascular diseases. According to the World Health Organization, there are 17.9 million CVD-related deaths worldwide [2]. It is the main factor in adult fatalities. By using a person's medical history, our approach may identify those who are most likely to be diagnosed with a cardiac condition [6]. It may assist in identifying illness with less medical tests and effective therapies, so that patients can be treated appropriately. It can identify anybody who is experiencing any heart disease symptoms, such as chest discomfort or high blood pressure. Three data mining techniques—logistic regression, KNN is superior than prior KNN, and random forest classifier—are the major emphasis of this research. Our project's accuracy for the system using only one data mining approach is 87.5%. As a result, the HDPS accuracy and efficiency rose with the use of additional data mining tools. The field of supervised learning includes logistic regression. Logistic regression uses only discrete values. This project's goal is to determine, depending on the patient's medical characteristics—such as gender, age, chest discomfort, fasting blood sugar level, etc.—whether they are likely to be diagnosed with any cardiovascular heart illnesses. A dataset including the characteristics and medical background of the patient is chosen from the UCI repository. We make a prediction about the patient's potential for heart disease using this information. This is predicted using a patient's medical characteristics to determine if he is likely to get one of 14 heart diseases. Three algorithms—KNN, Random Forest Classifier, and Logistic Regression—were used to learn these medical characteristics. KNN is the most effective algorithm here, with accuracy of 88.52%. Finally, we categorise individuals according to whether they are at risk of developing a cardiac condition or not. This procedure is also very economical.

2. Literature review

Krittanawong et al. [1], e approach was developed utilising several datasets available in March 2019 to assess machine learning algorithms' overall prediction power of predicting cardiovascular illness. The prediction study employed the area under the curve metric to determine the capacity to forecast illnesses such coronary artery disease, cardiac arrhythmias, heart failure, and stroke. However, finding the best algorithm for cardiovascular illness remains difficult due to the variability of machine learning algorithms.

Duan et al. [2] examined the relationship between blood and urine heavy metal concentrations and death from cancer and cardiovascular disease. Datasets from the National Health and Nutrition Examination Survey were used for the investigation. To assess single and multiple metal exposure, Poisson's regression was performed. Ages of research participants varied from twenty-five to eighty-five. The research found a relationship with metal mixers in both blood and urine and cancer mortality after accounting for age, gender, education level, body mass index, serum cotinine, and medical comorbidities. The authors do, however, note that the need for greater research on cardiovascular illness served as the impetus for this study.

Lippi et al. [3] The government has been forced to enact different types of lockdown because to the statewide quarantine imposed by COVID-19, which has caused attention to the potential for cardiovascular illness. All residents stay at home as a consequence of these limitations, which leads to physical inactivity. Strict quarantine, on the other hand, has raised the risk of cardiovascular death even though the WHO has issued clear standards on the quantity of physical activity necessary to maintain good health. Negative health impacts are seen after quarantine. The authors' suggestion that physical activity should be maintained even during quarantine in order to prevent negative cardiovascular effects has inspired the design of the present research investigation.

Aryal et al. [4] faecal ribosomal RNA of 16S was examined from individuals with and without cardiovascular disease to propose a method employing machine learning algorithms to screen microbiome-based cardiovascular illness. The American Gut Project provided the samples used for analysis. In order to train decision trees, random forests, neural networks, elastic nets, and support vector machines, five main categories of machine learning techniques were used. Various forms of differentiated bacterial taxa were found. An improved characteristics curve of 0.70 was produced via random forest. Random forest and one other machine learning algorithm were used in the present investigation because of random forest's established ability to predict cardiovascular illness.

Joo et al. [5] the longitudinal cohort research on 3.6 million patients seeking admission to hospitals in England by the authors examined the reliability of machine learning algorithms for predicting the risks of cardiovascular disease. The 19 prediction models' calibration and discriminating abilities were assessed. The neural network prediction score, for instance, varied from 2.4 to 7.2 percent whereas the random forest tree prediction score ranged from 2.9 to 9.2 percent. It was proposed that while comparing different models, logistic models shouldn't be used to forecast long-term hazards and that the degrees of comparison across models should be periodically assessed.

3.Methodology

K closest neighbours (KNN), logistic regression, and random forest classifiers, which might be useful for practitioners or medical analysts to properly identify heart disease, are the machine learning algorithms that are analysed in this work. This paperwork include looking at journals, papers that have been published, and current cardiovascular disease statistics. The suggested model has a framework thanks to methodology. The methodology is a procedure that involves stages that convert provided data into acknowledged data patterns for the consumers' awareness. The suggested data-processing approach (Figure 1.) consists of three stages: the collecting of data, the extraction of important values, and the pre-processing stage when we look at the data. Depending on the methods employed, data pre-processing deals with missing values, data cleansing, and standardisation. Pre-processed data are then classified using a classifier. The classifiers employed in the proposed model are KNN, Logistic Regression, and Random Forest Classifier. The suggested model is then put into practise, and its accuracy and performance are assessed using a variety of performance indicators. An efficient cardiovascular disease prediction system is shown in this model.

ARCHITECTURE DIAGRAM

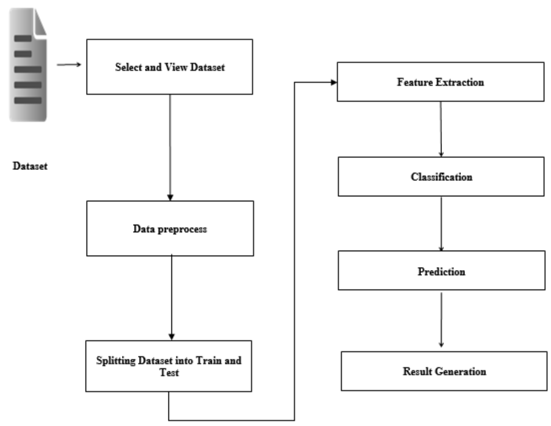


Fig1.Data flow diagram

Results and discussions

We can see from these results that even though the majority of researchers use different algorithms like SVC, Decision tree, and Logistic Regression to identify patients who have been diagnosed with Heart disease, KNN, Random Forest Classifier, and Logistic Regression produce a better more accurate result to outperform them . They also save a significant amount of money. The algorithms that we utilised are, for example, quicker and more cost-effective than the techniques that earlier researchers used. Furthermore, the combined maximum accuracy of KNN and Logistic Regression is 88.5%, which is higher than or nearly equivalent to the accuracy of earlier studies. So, to sum up that our accuracy is improved due to the increased medical attributes that we used from the dataset we took. Our project also tells us that Logistic Regression and KNN outperforms Random Forest Classifier in the prediction of the patient diagnosed with a heart Disease. This proves that KNN and Logistic Regression are better in diagnosis of a heart disease. The following 'figure 2', 'figure 3', 'figure 4', 'figure 5' shows a plot of the number of patients that are been segregated and predicted by the classifier depending upon the age group, Resting Blood Pressure, Sex, Chest Pain:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
0	28	1	2	130	132	0	2	185	0	0.0	?	?	?	0
1	29	1	2	120	243	0	0	160	0	0.0	?	?	?	0
2	29	1	2	140	?	0	0	170	0	0.0	?	?	?	0
3	30	0	1	170	237	0	1	170	0	0.0	?	?	6	0
4	31	0	2	100	219	0	1	150	0	0.0	?	?	?	0
5	32	0	2	105	198	0	0	165	0	0.0	?	?	?	0
6	32	1	2	110	225	0	0	184	0	0.0	?	?	?	0
7	32	1	2	125	254	0	0	155	0	0.0	?	?	?	0
8	33	1	3	120	298	0	0	185	0	0.0	?	?	?	0
9	34	0	2	130	161	0	0	190	0	0.0	?	?	?	0

Fig2.Loaded data set

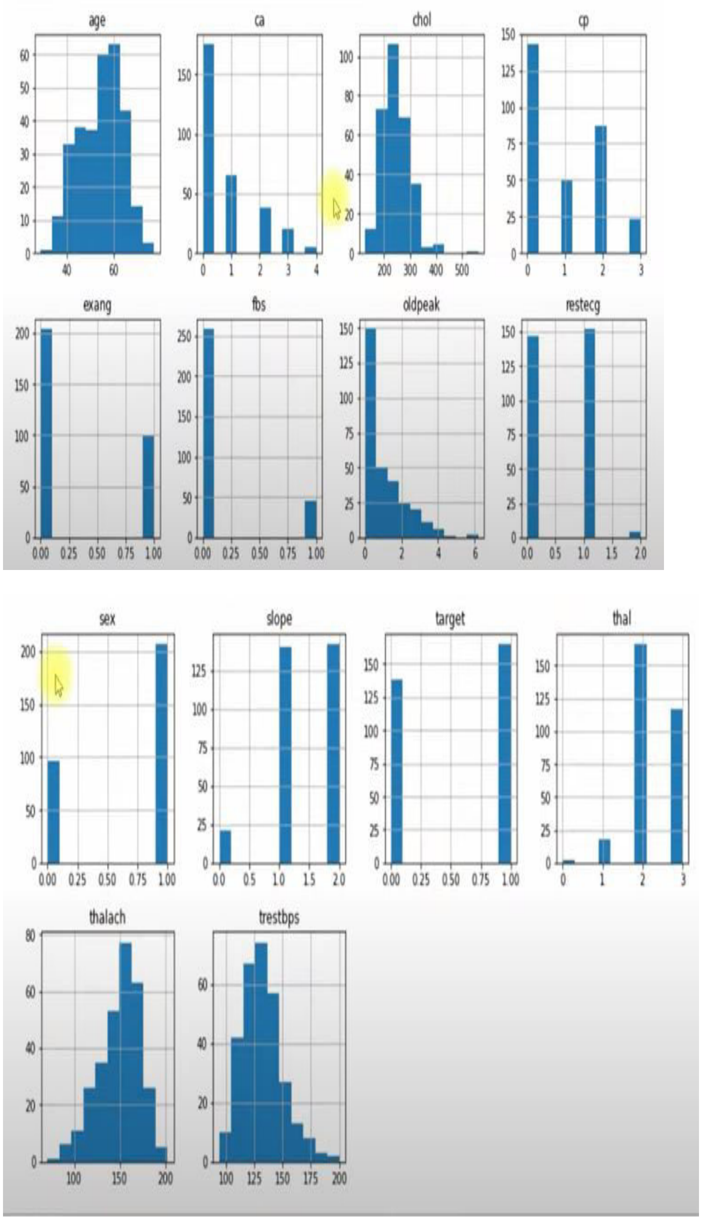
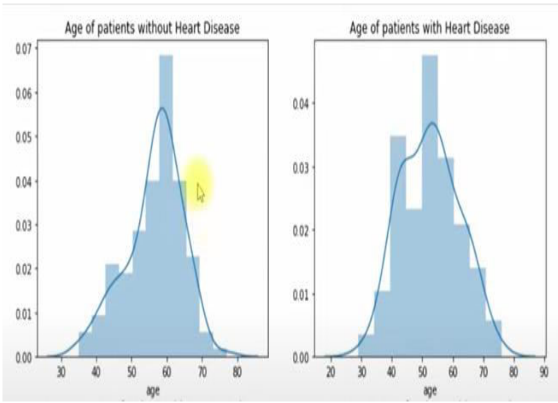


Fig3. Splitted data



International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

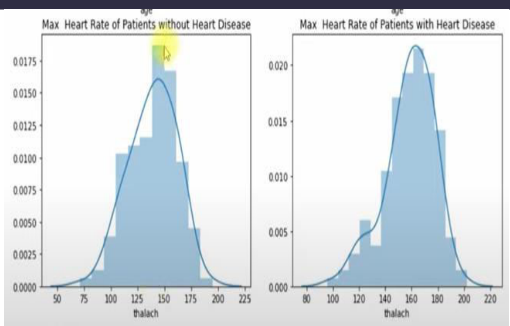


Fig4.Prediction

Conclusion

Three ML classification modelling approaches have been used to create a model for the identification of cardiovascular disease. By extracting the patient medical history that results in a fatal heart illness from a dataset that contains patients' medical history such as chest discomfort, sugar level, blood pressure, etc., this method predicts persons with cardiovascular disease. Based on clinical information about a patient's prior heart disease diagnosis, this heart disease detection system helps the patient. The proposed model was constructed using the methods of KNN, Random Forest Classifier, and Logistic Regression . Our model has an accuracy rate of 87.5%. The likelihood that the model will correctly identify whether a specific individual has heart disease or not increases with the use of additional training data [9]. These computer-assisted tools allow us to anticipate the patient quickly, more accurately, and at a far lower cost. We can work with a variety of medical datasets since machine learning methods are more advanced and can anticipate outcomes better than humans, benefiting both patients and medical professionals. As a result, our research aids in the prediction of patients who are diagnosed with heart problems by cleaning the dataset and using logistic regression and KNN to achieve an accuracy of an average of 87.5% on our model, which is better than the prior models' accuracy of 85%. Additionally, it is established that KNN's accuracy of 88.52% is the greatest of the three methods we utilised. Heart disease affects 44% of those that are identified in the dataset.

References

[1] G. Lippi, B. M. Henry, and F. Sanchis-Gomar, “Physical inactivity and cardiovascular disease at the time of coronavirus disease 2019 (COVID-19),” European journal of preventive cardiology, vol. 27, no. 9, pp. 906–908, 2020.

[2] S. Aryal, A. Alimadadi, I. Manandhar, B. Joe, and X. Cheng, “Machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease,” Hypertension, vol. 76, no. 5, pp. 1555–1562, 2020.

[3] D. Han, K. K. Kolli, S. J. Al’Aref et al., “Machine learning framework to

identify individuals at risk of rapid progression of coronary atherosclerosis: from the PARADIGM registry," Journal of American Heart Association, vol. 9, no. 5, Article ID e013958, 2020.

[4] G. Joo, Y. Song, H. Im, and J. Park, "Clinical implication of machine learning in predicting the occurrence of cardiovascular disease using big data (Nationwide Cohort Data in Korea)," IEEE Access, vol. 8, pp. 157643–157653, 2020.

[5] V.V. Ramalingam*, Ayantan Dandapath, M Karthik Raja; Heart disease prediction using machine learning techniques : a survey – March 2018 https://www.sciencepubco.com/index.php/ijet/article/vie w/10557

[6] A. S. Thanuja Nishadi University of Colombo, Faculty of Graduate Studies, Sri Lanka, Predicting Heart Diseases In Logistic Regression Of Machine Learning Algorithms By Python Jupyterlab- Volume 3 Issue 8, August 2019. http://www.ijarp.org/published-researchpapers/aug2019/Predicting-Heart-Diseases-In-LogisticRegression-Of-Machine-Learning-Algorithms-ByPython-Jupyterlab.pdf

[7] Reddy Prasad,Pidaparthi Anjali, S.Adil, N.Deepa. Heart Disease Prediction using Logistic Regression Algorithm using Machine Learning - ISSN: 2249 – 8958, Volume8, Issue-3S, February 2019. https://www.ijeat.org/wpcontent/uploads/papers/v8i3S/C11410283S19.pdf

[8] Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. (2019). Applications of Artificial Intelligence in Transport: An Overview. Sustainability, 11(1), 189. doi: 10.3390/su11010189 [9] Cunningham, Padraig & Delany, Sarah. (2007). kNearest neighbour classifiers. Mult Classif Syst.

[10] Sharma, Himani & Kumar, Sunil. (2016). A Survey on Decision Tree Algorithms of Classification in Data Mining. International Journal of Science and Research (IJSR). 5.