TO PREDICT HEART DISEASE BY ANALYSING ECG

Dr.M.Rajeswari[1],
Assistant Professor,
Department of B.Com (Business Analytics),
PSGR Krishnammal College for Women, Coimbatore, India
rajeshwarim@psgrkcw.ac.in

Swathi.J[2]
UG Scholar,
Department of B.Com (Business Analytics),
PSGR Krishnammal College for Women, Coimbatore, India
swathijaganchary@gmail.com

Abstract: ECG abbreviated as electrocardiogram is a simple test that can be used to check the hearts rhythm and electrical activity. An ECG is often used alongside other tests to diagnose and monitor conditions affecting the heart. An electrocardiogram (ECG) is a simple test that can be used to monitor heart rate and electrical activity. The ECG is often used in conjunction with other tests to help diagnose and monitor cardiovascular conditions. Various ECG parameters such as heart rate, slope, ST stress, thallium, old ECG signal are used for analysis. Based on these ECG signal parameters, different heart disease and type of chest pain are diagnosed with respect to age.

Keywords: heart disease, machine learning (ml),chest pain type, logistic regression, ECG,thallium,slope.

I.Introduction

The term “heart disease” refers to a wide range of heart conditions. Decreased blood flow can cause heart disease. The early detection of abnormal heart conditions is vital to identify heart problems and avoid sudden cardiac death. The people with similar heart conditions almost have similar electrocardiogram (ECG) signals. By analyzing the ECG signals’ patterns one can predict arrhythmias.

Arrhythmia is a condition where the heart beats too slowly or too quickly or irregularly. Coronary heart disease is a condition where the heart’s blood supply is blocked or interrupted by build-u fatty substances. Cardiomyopathy is a condition where the heart walls become thickened or enlarged. All these heart diseases can be identified with the help of ECG signals like slope, st depression, thallium, oldpeak etc. These signals are recorded by a machine and monitored by a doctor to see if they are unusual. The project’s aim is to collect ECG dataset, apply logistic regression and find the heart disease rate and accuracy rate.

II.Objective

Machine learning (ML) is a field of research devoted to understanding and 'learning' building methods, that is, methods that enhance data to improve the performance of a particular set of tasks. Machine learning algorithms create a model based on sample data, known as training data, to make predictions or decisions without explicitly planning to do so. Machine learning algorithms are used in many different fields. Healthcare and medicine is one such field where machine learning plays an important role. Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Dataset containing age, gender, heart rate thallium, oldpeak , slope, st depression, heart disease are taken. Logistic regression is implied by taking 'heart disease', 'chest pain type' and 'oldpeak' as independent variables and the rest as dependant variables. These three independent variables are the respective prediction variables also. All the dependant variables are compared with the prediction variables separately. Then they are split into training data and testing data. Codes are run for training and testing them. Later the accuracy rate of heart disease is found.

III.Related works

The term "heart disease" refers to several types of heart conditions. The most common type of heart disease in the United States is coronary artery disease (CAD), which affects the blood flow to the heart. Decreased blood flow can cause a heart attack. [3]. An electrocardiogram records the electrical signals in the heart. It's a common and painless test used to quickly detect heart problems and monitor the heart's health [4] An electrocardiogram is a painless, noninvasive way to help diagnose many common heart problems. A health care provider might use an electrocardiogram to determine or detect heart diseases [1]

Renewed interest in machine learning is Higher volumes and varieties of available data and Affordable data storage.[5] In most modern trackers, to cope with natural image changes, a classifier is typically trained with translated and scaled sample patches. We propose an analytic model for datasets of thousands of translated patches. By showing that the resulting data matrix is circulant, we can diagonalize it with the discrete Fourier transform, reducing both storage and computation by several orders of magnitude.[6]

Logistic regression and artificial neural networks are the models of choice in many medical data classification tasks. [8] logistic regression is one of the most important analytic tools in the social and natural sciences. In natural language processing, logistic regression is the baseline supervised machine learning algorithm for classification, and also has a very close relationship with neural networks.[10] .The machine learning algorithm neural networks has proven to be the most accurate and reliable algorithm and hence used in the proposed system.[12] .

Using machine learning techniques, the aim of this study is to evaluate the accuracy of supervised learning techniques in predicting heart disease based on the dataset obtained from University of California Irvine data repository. The result from this study shows that Naïve Bayes and Bayesian Network has better

estimated accuracy in Weka for the data set, while both Bayesian Network and J48 may give useful insight with Weka generated visualization.[14].

Analysis of ECG signal plays an important role in diagnosing cardiac diseases. An efficient method of analysing ECG signal and predicting heart abnormalities have been proposed in this paper. In the proposed scheme, at first the QRS components have been extracted from the noisy ECG signal by rejecting the background noise.[15].

In the proposed method features and duration of ECG signals were extracted such as QRS complex, R-R intervals from the noisy signals using Pan Tompkins Algorithm. The data used for analysis were collected from MIT-BIH arrhythmia database. Extracted features were estimated with the standard set of values based on the developed decision making algorithm to find the degree and types of arrhythmia.[16]

The one-class approach was able to identify abnormality with area-under-curve (AUC) 0.83, and with 75.6% accuracy. For four-class classification, we used 86 features in total, with 72 additional features extracted from the ECG. Accuracy for this four-class classifier reached 75.1%. The methods demonstrated proof-of-principle that cardiac abnormality can be detected using machine learning in a large cohort study.[17]

IV.Methodology

Step1: importing the dataset from kaggle dataset, modified the dataset and saved in excel.csv format.

Step 2: using Google colab for executing python coding.

Step 3: data preprocessing.

Step 4: separating dataset into training and testing.

Step 5: visualization for better understanding

Step 6: using logistic algorithm to predict accuracy.

Flowchart of work process

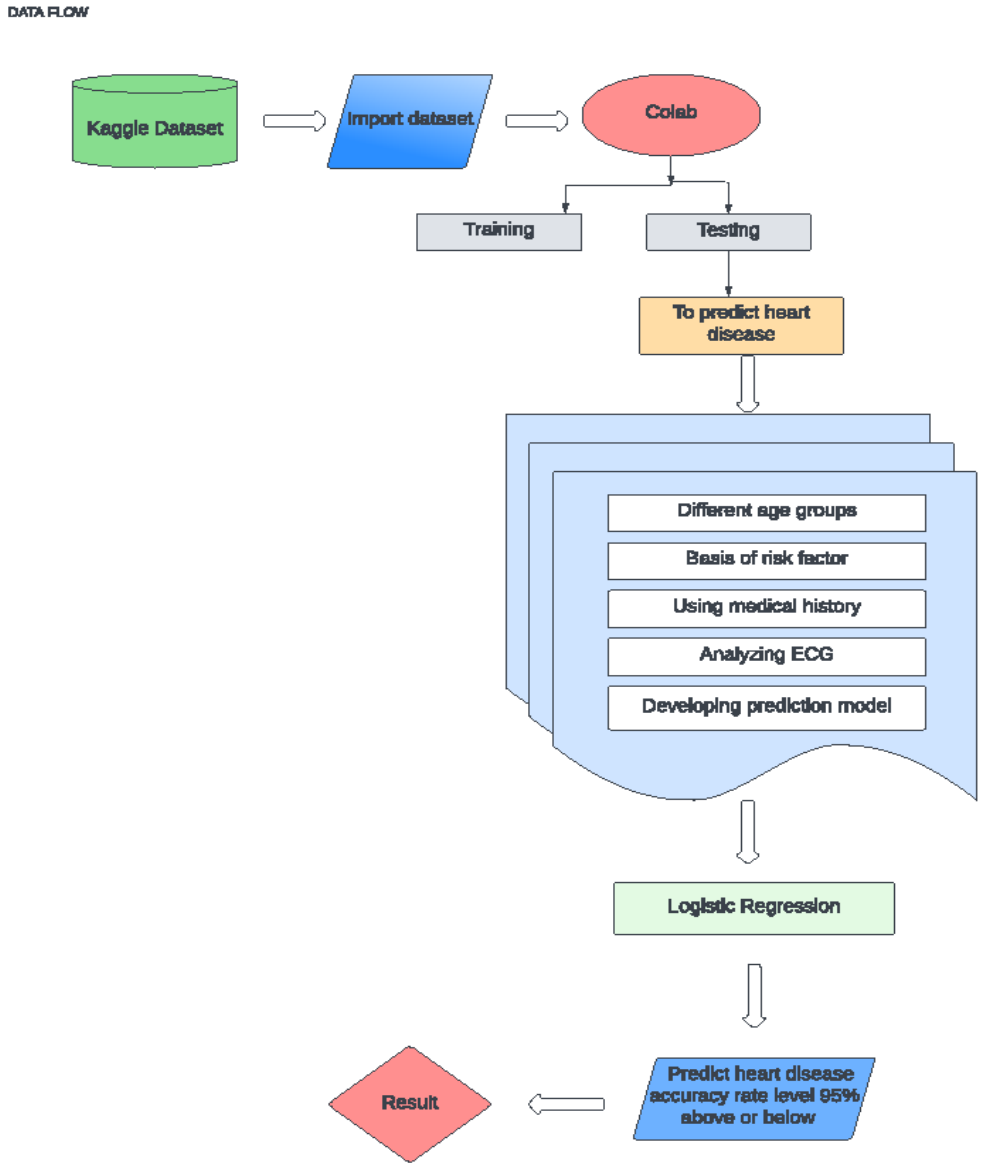

Fig 1.1

V.RESULT

OUTPUT

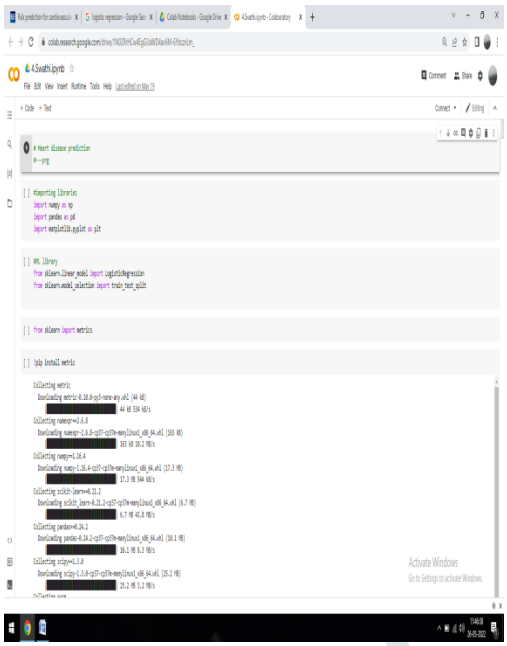

Fig 1.2

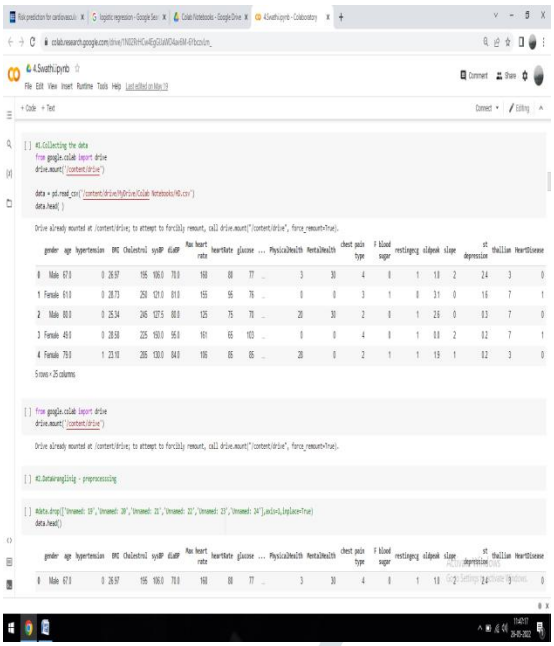

Fig 1.3

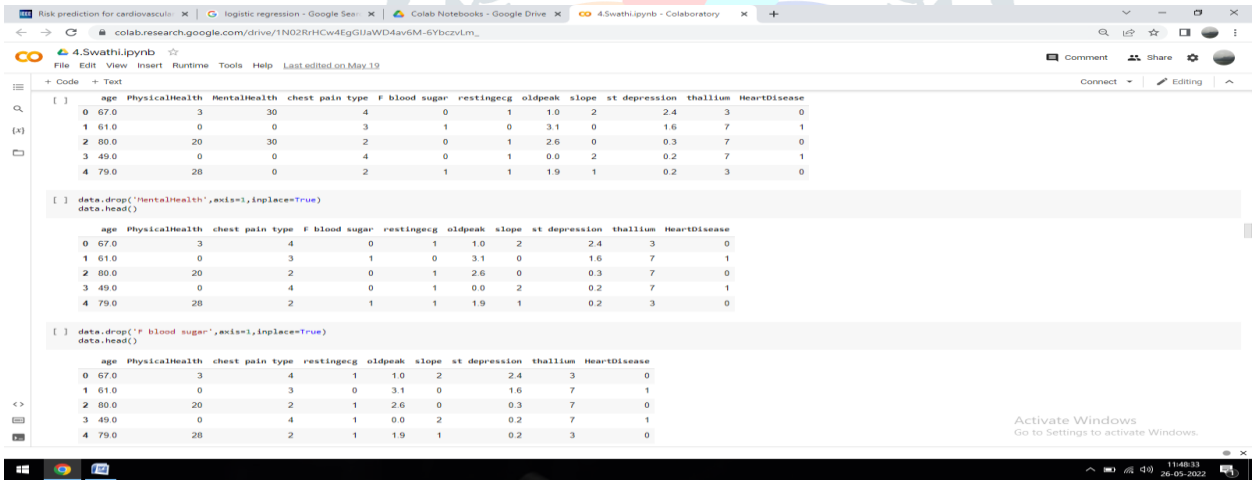

Fig. 1.4

Fig 1.5

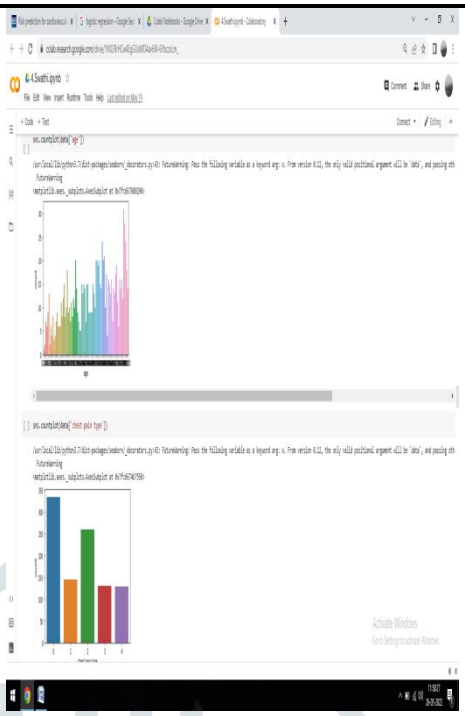

Fig 1.6

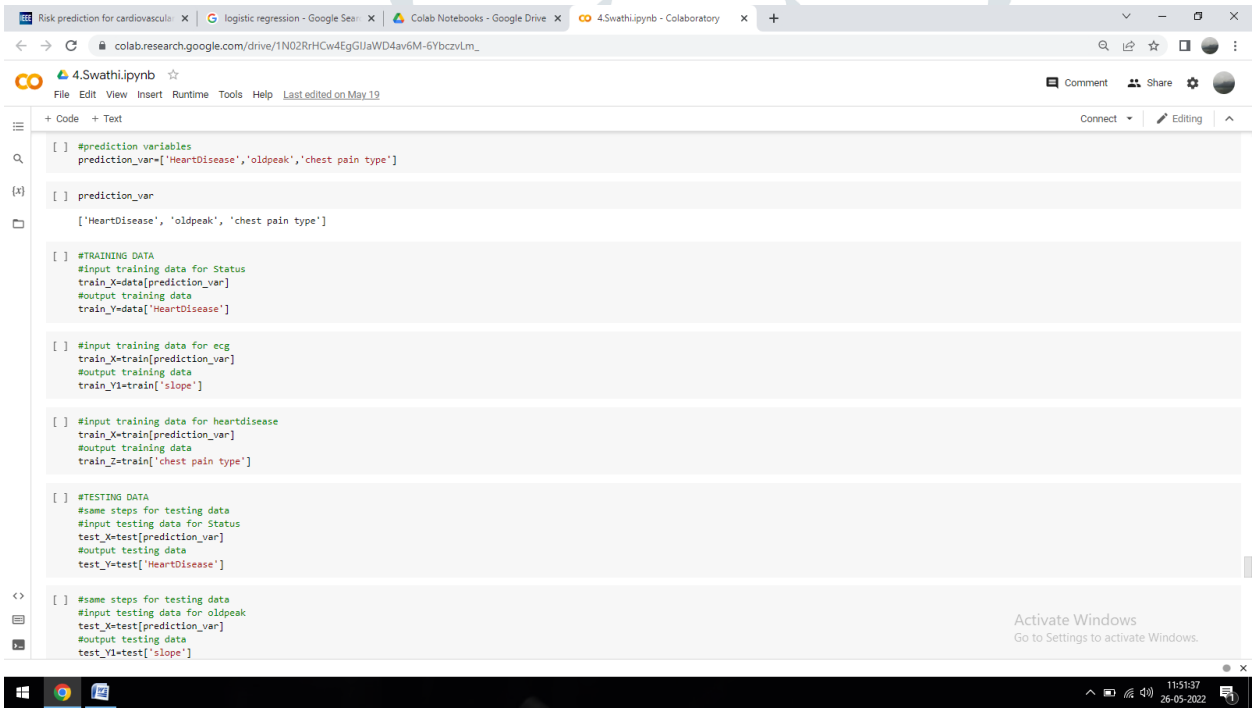

Fig 1.7

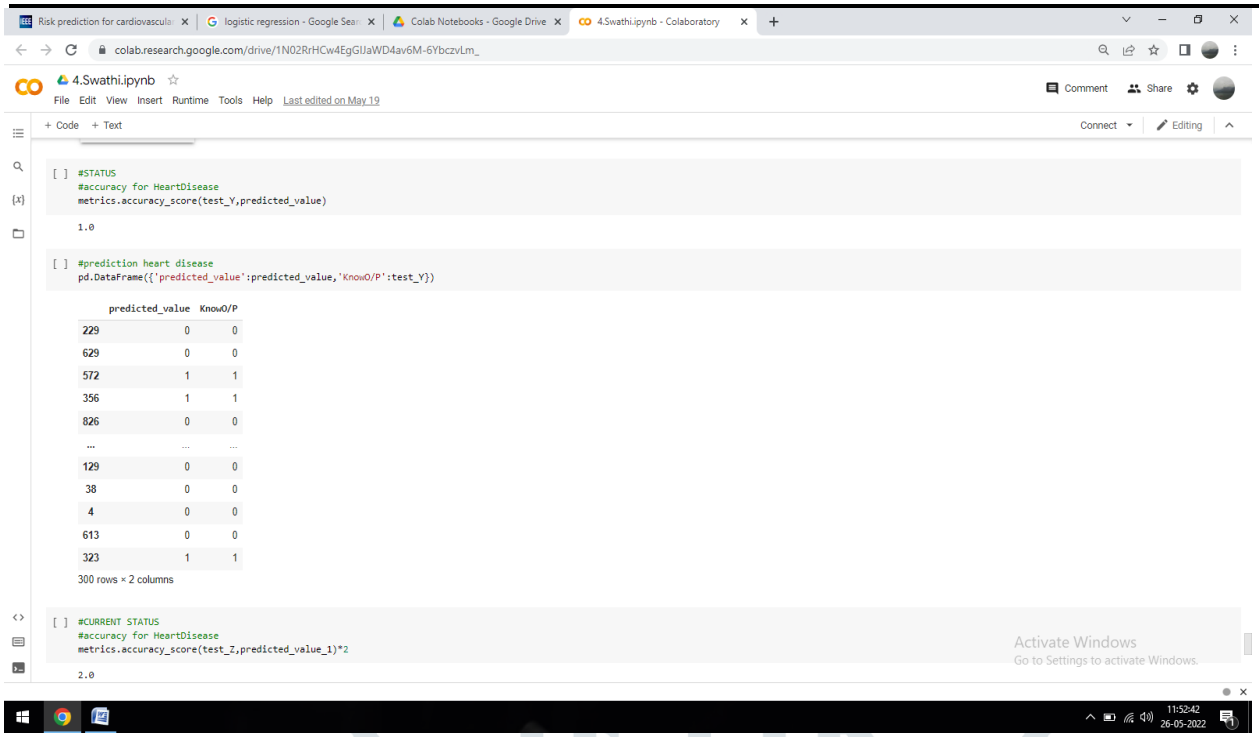

Fig 1.8

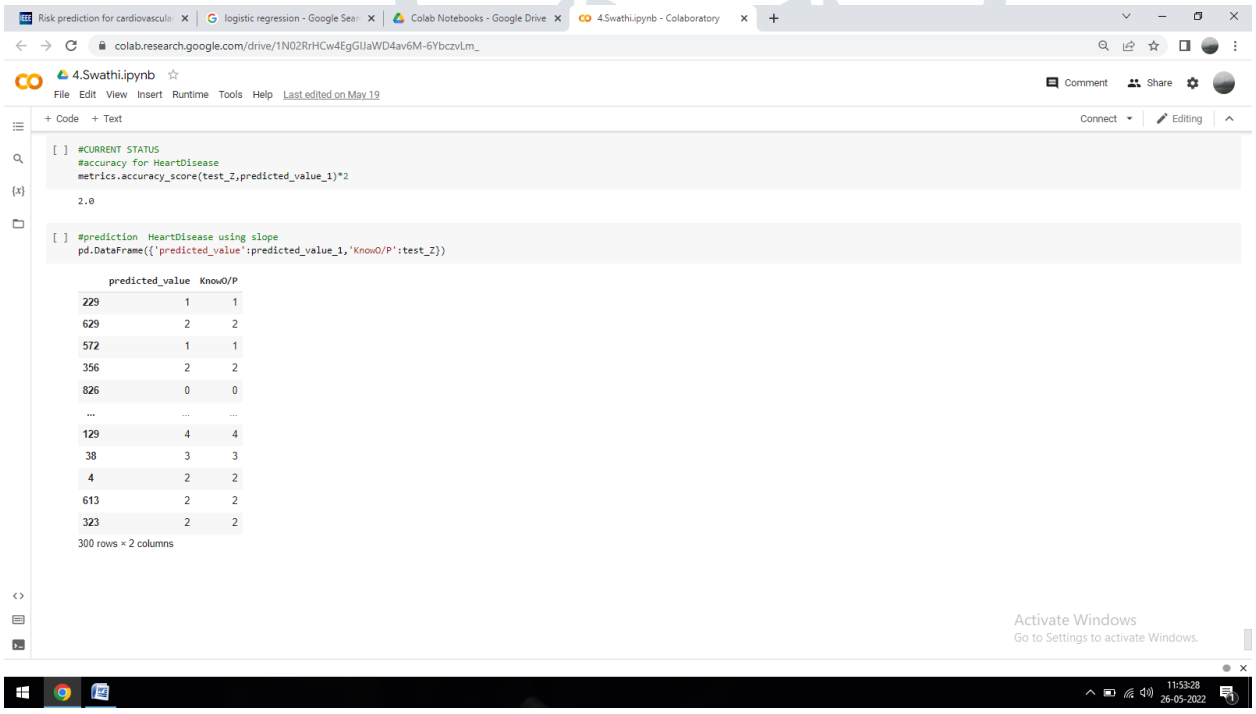

Fig 1.9

Main determinants of heart disease are thallium level, slope level, oldpeak, st depression and resting ecg. Using logistic regression , heart disease, chest pain type and oldpeak are taken as predicting value. In the above picture (fig 1.9), the known output(know O/P) the prediction is done with heart disease and slope. When the slope level is above 200 the person is likely affected by heart disease. By analyzing all the dataset, it is predicted that he/she may or may not have a heart disease in future.

VI.Conclusion and further works

In this paper, importing dataset, executing coding and visualization are done in Google colab. Logistic regression is used in the further prediction of heart disease status. To avoid early and unnecessary deaths due t heart diseases, conducting ecg test is mandatory and early prediction of it leads to healthy life.

Reference

1. Centers for Disease Control and Prevention, National Center for Health Statistics. About Multiple Cause of Death, 1999–2019. CDC WONDER Online Database website. Atlanta, GA: Centers for Disease Control and Prevention; 2019. Accessed February 1, 2021.
2. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart disease and stroke statistics—2021 update: a report from the American Heart Associationexternal icon. *Circulation.* 2021;143:e254–e743.
3. Centers for Disease Control and Prevention, Public health media library. Heart disease. National Center for Chronic Disease Prevention and Health Promotion , Division for Heart Disease and Stroke Prevention.Last reviewd setember 27, 2021.
4. Sandhya Pruthi, M.D. ELECTROCARDIOGRAM (ECG) OR (EKG).MAYO CLINIC . PRC-20214603.
5. Gogas, Periklis & Papadimitriou, Theophilos. (2015). Presentation Machine Learning.
6. High-Speed Tracking with Kernelized Correlation Filters, by Batista, J., Caseiro, R., Henriques, J.F., & Martins, P. (2015). CoRR, abs/1404.7584. (cited 439 times, HIC: 43 , CV: 0)
7. A Review on Multi-Label Learning Algorithms, by Zhang, M., & Zhou, Z. (2014). IEEE TKDE, (cited 436 times, HIC: 7 , CV: 91)
8. Stephan Dreiseitl, Lucila Ohno-Machado,Logistic regression and artificial neural network classification models: a methodology review,Journal of Biomedical Informatics,Volume 35, Issues 5–6,2002,Pages 352-359,ISSN 1532-0464
9. L. Breiman, *et al*.Classification and regression trees,Wadsworth, Belmont, CA (1984) Google Scholar
10. Speech and Language Processing. Daniel Jurafsky & James H. Martin. Copyright © 2021. Draft of December 29, 2021.
11. Witten, I. H. and E. Frank. 2005. Data Mining: Practical Machine Learning Tools and Techniques, 2nd edition. Morgan Kaufmann.
12. A. Gavhane, G. Kokkula, I. Pandya and K. Devadkar, "Prediction of Heart Disease Using Machine Learning," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 1275-1278, doi: 10.1109/ICECA.2018.8474922.
13. Sarbaras J.S. Suhaim, Israj Ali, "A Portable Android Based ECG System for Long-Time Heart Monitoring", *2022 IEEE Delhi Section Conference (DELCON)*, pp.1-6, 2022.
14. Herold Sylvestro Sipail, Norulhusna Ahmad, Norliza Mohd Noor, "Heart Disease Prediction Using Machine Learning Techniques", *2021 IEEE National Biomedical Engineering Conference (NBEC)*, pp.48-52, 2021.
15. T. Debnath, M. M. Hasan and T. Biswas, "Analysis of ECG signal and classification of heart abnormalities using Artificial Neural Network," 2016 9th International Conference on Electrical and Computer Engineering (ICECE), 2016, pp. 353-356, doi: 10.1109/ICECE.2016.7853929.
16. H S Bhanu, S Tejaswini, M S Sahana, K Bhargavi, K S Praveena, S S Jayanna, "Analysis of ECG Signal and Classification of Arrhythmia", *2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT)*, pp.619-623, 2021.
17. Y. Shen, Y. Yang, S. Parish, Z. Chen, R. Clarke and D. A. Clifton, "Risk prediction for cardiovascular disease using ECG data in the China kadoorie biobank," 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, pp. 2419-2422, doi: 10.1109/EMBC.2016.7591218.