Myocarditis prediction in Covid-19 patient using machine learning algorithm

Sreevidya B.
Department of Computer Science & Engineering
Amrita School of Engineering, Bengaluru
Amrita Vishwa Vidyapeetham, India
b_sreevidya@blr.amrita.edu

Rajesh M.
Department of Computer Science & Engineering
Amrita School of Engineering, Bengaluru
Amrita Vishwa Vidyapeetham, India
m_rajesh@blr.amrita.edu

Ruchi Chaurasiya
Department of Computer Science & Engineering
Amrita School of Engineering, Bengaluru
Amrita Vishwa Vidyapeetham, India
Bl.en.u4cse22278@bl.students.amrita.edu

Shakshi Yadav
Department of Computer Science & Engineering
Amrita School of Engineering, Bengaluru
Amrita Vishwa Vidyapeetham, India
bl.en.u4cse22286@bl.students.amrita.edu

Swagata Malik
Department of Computer Science & Engineering
Amrita School of Engineering, Bengaluru
Amrita Vishwa Vidyapeetham, India
Bl.en.u4cse22252@bl.students.amrita.edu

Abstract— Myocarditis, one of the great consequence of Covid-19 due to which many people got affected and loss their life .Here a few data are available in the literature about the incidence and clinical significance in patients affected by SARS-CoV-2. Myocarditis occurs when the heart muscle becomes inflamed and inflammation occurs when your body's immune system responds to infections . It can be diagnosed using cardiac magnetic resonance image (MRI), a non-invasive imaging technique with the possibility of operator bias. This paper proposes different machine learning algorithms to detect myocarditis . We evaluate our proposed approach on the "Characterization of Myocardial Injury in Patients With COVID-19" myocarditis dataset based on standard criteria and demonstrate that the proposed method gives superior myocarditis diagnosis performanc Further studies, specifically designed on this issue, are warranted.

Index Terms—Myocarditis diagnosis, MRI, Machine Learning algorithm

I. INTRODUCTION

Covid-19, a virus that started from a single country but later become the world concern which affected people physically,mentally,socially and economically around the world.Due to this virus people around the world start to suffer from many types of another diseases.such as heart,lungs, Respiratory,etc. Cardiac complications can be considered one of the major and serious problems caused by coronavirus infection. Arrhythmias,heart attack,stroke ,myocarditis, etc are one the mostly faced heart disease during pandemic. People had already researched on most of these diseases and their relation with covid-19. But a very few research have been done on Myocarditis. Myocarditis is a well-known cardiac complication of viral infections.That is one of the reason to work on this research paper.

Myocarditis mainly affect the T-cells of human immune system. Most of the previous researches and prediction model were considered based on ElectroCardiographic report meanwhile we have taken blood report into consideration to kickstart our research and to design our prediction model that will make the model more comfortable for the user's to take the first step towards their health and pre-idea about myocarditis hassle free.

However, it is unclear about the cause of myocarditis due to Covid-19 but study revealed that about 28% people among Covid-19 patients has chances of getting affected by myocardial infection. [1] Protein, named ACE2 which is found present in epithelial cells is also present in cardiomyocytes. During epidemics,a case study using endomyocardial biopsy revelaed that coronavirus also resides in myocardium. Among study of 20 Covid-19 patients heart sample, it was observed that,SARS-CoV2 virus has been found in myocardium of 7 patients which conclude that there is life-threatening chance of diagnosising myocarditis for people suffering from coronavirus.

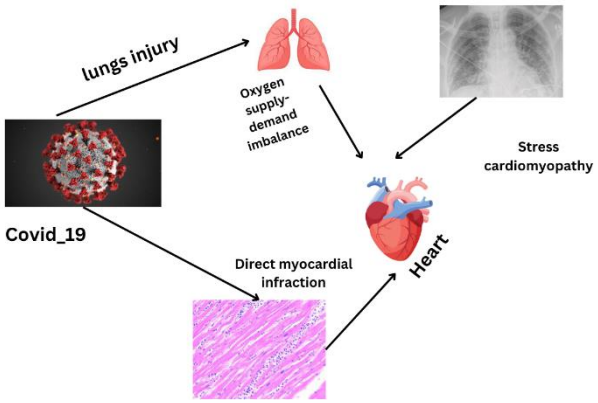


Figure:-

The figure depicts how Covid-19 infection is leading to myocardial infraction.

II. Related Works

According to the World Health Organization, an estimated 17.9 million people expired from CVDs in 2019, which represents 32% out of all global death[1]. In December 2019, coronavirus was first described in Wuhan, China, in case complaining of flulike symptoms. The virus was isolated and linked as a new strain of covid-19 now named SARS-CoV-2 (severe acute respiratory syndrome coronavirus[2]. It is one of the most serious worldwide concerns, which has created many problems in health, economics, and other aspects of human life around the world. It has shown to have a lot of manifestations during the infection, among which some of them are more critical. Cardiac issues can be considered one of the major and serious problems caused by coronavirus infection.

This[3] paper explained how myocardial infraction is becoming a life threatening disease that has 20% mortality rate among young people. Since, MRIs can detect myocarditis but its complex. Some models have been developed using learning algorithms and data mining techniques to detect myocarditis making it more reliable and accurate day by day.

MI, usually has severe impact which can lead to death due to coronary artery blockage[4] that is why it is very important to diagnose it in early stage and proper test. Sometimes, people die even before reaching to hospital. Sometimes, there is no complications of occurance of myocarditis which is one of the hindrance in prediction of MI. In[5,6], a large datasets that focused on ECG-ViEW II has been considered that assembled data from 19-year old boy. Different machine learning algorithms has been used to develop the model that includes logistic regression, Support vector machine, Decision tree, etc. and to distinguish which model is reliable.

However, paper[7,8] depicts that in the world of advance technology in medical science, even some doctors are also unable to predict the accurate outcome of myocarditis making it a complex problems for the treatment. Moreover, machine learning models has been developed for prediction but is also not free from variability and bias. They are also not 100% accurate to detect myocarditis. Myocardial infraction can be sometimes considered as silent[9], and undetectable which can lead to unaware death as it may or may not have severe complications and symptoms.

In addition, in most cases patients with acute myocarditis who tested positive for covid-19 were significally older and had lower levels of troponin in compared to those who tested negative.Study demonstrated that a patients with cardiovascular comorbidities were more at risk of developing COVID-19 myocarditis[9,10]. So far, a vast random forest algorithm have been used , whose aim is to diminish the error between the target and predicted outputs with appropriate blood report.

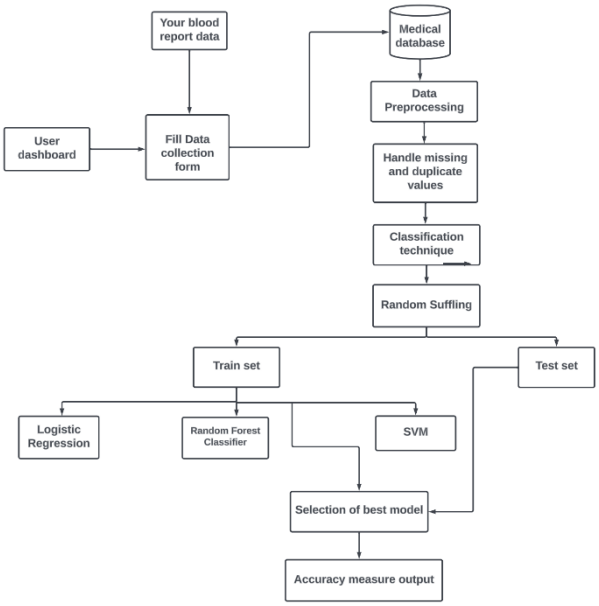
The estimated prevalence of coronary heart disease in Indonesia is around 883.447 or 0.5 percent [11]. These days deep learning model is becoming more effective than learning algorithms especially in case of finding death rates. However, machine learning has grown upto far extend that in this modern world it has become one of the best algorithms for prediction model[12].

However, it may sometimes gives false positive accuracy in case it is implemented in wrong or improper datasets[13] that is why it is very important to have a proper and genuine datasets in order to design the best prediction model and also it is very important to clean your datasets to get better and true positive accuracy. Cleaning datasets involves[14] data preprocessing and data manipulation. In a survey, it was found that about 65.2% programmer's user python language among roughly 2000 participants in a competition as it has free libraries and packages that fascinates user to code and design their model. Python has fast implementation and already existing which makes it easy for even non-computer scientists.

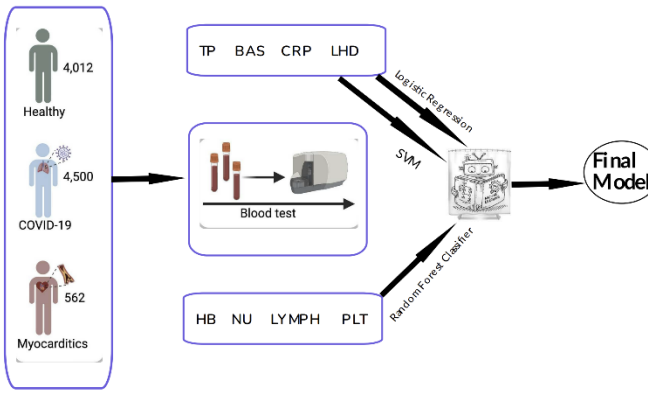
Although, machine learning has become very efficient in practical use but it is highly sensitive in terms of tuning hyperparameters[15,16]. For example, in random forest classifier parameters are categorize by number of trees and their depth, the number of features per tree, and the minimum samples per leaf node.

III. Design

In this proposed model, datasets based on Patient's blood report has been used for detecting myocarditis



The model involves data collection, data preprocessing along with handling missing values as well as duplicate values. It collects user's blood report data (specified as parameters explained in datasets section) as an input through user interface. The designed model, is then implemented to the fetched user's data for predicting the accuracy of user's chances of suffering from myocarditis and same is displayed to user through user interface. The fig 2. Shows indetailed explanation of model.



IV. Implementation

This section describes the practical realization of the proposed model. To develop this model, study of python, it's libraries, MySQL and machine learning has been done. Datasets collected from patients during pandemic has been used to predict myocardial infarction. To realize the practical implementation, we have applied three machine learning algorithm. They are

- Logistic Regression
- Random Forest Classifier
- Support Vector Machine

Firstly, all the essential libraries(numpy, pandas, scikit-learn, matplotlib, seaborn) were imported that is required for training the hyperparameters, and for data manipulation. In addition, PyMySQL python library was imported to connect our model to our database to store the user details, and also to read the datasets into pandas DataFrame.

To make our model more accurate and to prevent it form giving false positive accuracy we, then dropped all the rows containing null values(missing values) also we dropped duplicate values.

Secondly, data was splitted into train(70%) and test(30%) set using sklearn.model_selection library in python and then hyperparameters were trained using the above mentioned machine learning algorithms to predict the presence/absence of myocarditis as well as to check the accuracy of prediction model that best suits the datasets taken.

A. Logistic Regression

Data attributes taken from datasets were given for classification using logistic regression algorithm by dividing the hyperparameters into categorical and real number. Finally, the accuracy of the model has been achieved[6]. In LR, only two dependent possible binary values are considered i.e. 1/0. Here, in our model we have depicted 1 as person having myocarditis and 0 as person not having myocarditis.

Below given is the Pseudo code to design myocarditis prediction model using logistic regression

Start

- For n ← 1 to i
- For each training data sample d_i.
- Set the target valve for the regression to

$$Z_i \leftarrow \frac{y_j - p(1|d_J)}{[p(1|d_j) \ . \ (1 - p(1|d_j))]}$$

- Initialize the weight of sample d to $p(1|d_j).(1-p).(1|d_j)$
- Finalize a f(j) to the data with class value (z_j) & weights (w_j)
- Assign (class label:1) if $p(1|d_j)>0.5$,otherwise (class label:2)

End

B. Random Forest Classifier

Random Forest Classifer is considered as one of the fast, and accurate algorithm in machine learning as it combines outputs of multiple decision trees to reach at a single result. It furnishes with other learning algorithm to give better accuracy result.

Below given is the Pseudo code to design myocarditis prediction model using logistic regression

Input i- maximum numbers of samples
j-Total no. of features
k-size of sample
n=next node

Start

- Create i samples from your database
- Size of sample of random size k where k<j
- Split m features as selected in step 2 and detects the best split point for n node
- Until one leaf node is attained and the tree remains completed,split the tree iteratively
- Using trees classification predicated data is collected from the trained data (i)
- Final RF model is ready
- Retrun RF

End

C. *Support Vector Machine(SVM)*

From the datasets given, SVC takes different data points to form a hyperplane and correlate that data to classify the test sets which is depicted as given equation:

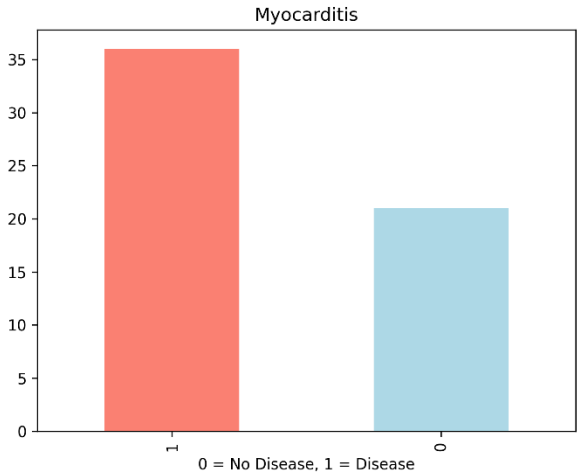
$$\beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta iXi = 0$$

Here β0, β1, β2 . . . βi = the hypothetical values cooperated with each variable X. X1. . . . . Xi = data points in sample space of i dimension.

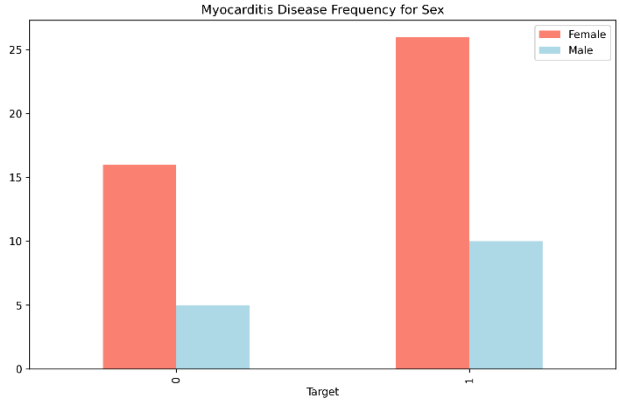
We checked the accuracy obtained by all three models and after analyzing the accuracy obtained from all the algorithms model we, then made our final prediction model using Random Forest Classifier algorithm being one of the best prediction model among these with having accuracy of 93.2%. Moreover, it reduces the risk of overfitting and the required training time. Additionally, it offers a high level of accuracy.

V. Results

To analyze our prediction model, among 4500 datasets taken, approx 1/3 of them were diagnosed myocarditis. In comparison, Random Forest Classifier algorithm gave 93.2% accuracy whereas Logistic regression and Support Vector Machine gave 91% and 90% accuracy respectively.



The graph shows the relationship between people suffering from Covid-19 as well as myocarditis and people suffering from Covid-19 but luckily saved from myocarditis.



This graph depicts the correlation between gender and the risk of suffering.

VI. Conclusion

By using machine learning algorithm, myocarditis prediction model has been designed. It has also depicted how Covid-19 has affected people's cardiac health leading to unaware death. The complications in parameters for predicting myocarditis has been reduced.

References

[1] Mohammed Ali1 , Haaris A. Shiwani2 , Mohammed Y. Elfaki3 , Moaz Hamid4 , Rebabonye Pharithi5 , Rene Kamgang5 , Christian BinounA Egom6 , Jean Louis Essame Oyono7 and Emmanuel Eroume-A Egom7, "COVID-19 and myocarditis: a review of literature", Ali et al. The Egyptian Heart Journal (2022) 74:23

[2] Zhu N, Zhang D, and Wang W, " A novel corona virus from patients with pneumonia in China", 2019. N Engl J Med 2020;382:727–733

[3] Saba Danaei, Arsam Bostani, Seyed Vahid Moravvej, Fardin Mohammadi, Roohallah Alizadehsani, Afshin Shoeibi, Hamid Alinejad-Rokny, and Saeid Nahavandi, "Myocarditis Diagnosis: A Method using Mutual Learning-Based ABC and Reinforcement Learning", IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics ,November 21-22

[4] Akashdeep Singh Chaudhary, Er. Ravinder Saini, Er. Nikita Kapoor, " Implementation of Decision Tree with Pruning for Predicting the Outcomes of Myocardial Infarction", 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)

[5] Lujain Ibrahim, Munib Mesinovic, Kai-Wen Yang, and Mohamad A. Eid, " Explainable Prediction of Acute Myocardial Infarction Using Machine Learning and Shapley Values", t New York University Abu Dhabi on November 24, 2020
[6] Sharmin Akter, Mahdia Amina and Nafees Mansoor, "Early Diagnosis and Comparative Analysis of Different Machine Learning Algorithms for Myocardial Infarction Prediction", /92021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC)
[7] M.SureshKumar M.Sella Vallal Sibi, R.Pugazh Bharathi, and P.Shanmugapriya, "Myocardial Prediction and Identification using Convolution Neural Networks", 022 International Conference on Communication, Computing and Internet of Things (IC3IoT)
[8] Shui-Hua Wang, Gerry McCann, and Ivan Tyukin, "Myocardial Infarction Detection and Quantification Based on a Convolution Neural Network with Online Error Correction Capabilities"
[9] Laganà N, Cei M, and Evangelista, "Suspected myocarditis in patients with COVID-19: a multicenter case series" (2021) Medicine 100(8):e24552–e24552. https://doi.org/10.1097/MD.0000000000024552
[10] Omidi F, Hajikhani B, and Kazemi, "COVID-19 and cardiomyopathy: a systematic review", (2021) https://doi.org/ 10.3389/fcvm.2021.695206
[11] Aditya Saxena, Manish Kumar, Pawan Tyagi, Kuldeep Sikarwar, and Aman Pathak, "Machine Learning based selection of Myocardial Complications to Predict Heart Attack",2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)
[12] Annisa Darmawahyuni, Siti Nurmaini, and Sukemi, "Deep Learning with Long Short-Term Memory for Enhancement Myocardial Infarction Classification",2019 6th International Conference on Instrumentation, Control, and Automation (ICA) Bandung, Indonesia. 31 July – 2 August 2019
[13] Peng Li†, Xi Rao‡, Jennifer Blase†, Yue Zhang†, Xu Chu†, and Ce Zhang‡, "CleanML: A Study for Evaluating the Impact of Data Cleaning on ML Classification Tasks",2021 IEEE 37th International Conference on Data Engineering (ICDE)
[14] Stančin, and A. Jović , "An overview and comparison of free Python libraries for data mining and big data analysis",MIPRO 2019, May 20-24, 2019, Opatija Croatia
[15] Sandeep Singh Sandha, Mohit Aggarwal†, Igor Fedorov†, and Mani Srivastava, "",niversity of California, Los Angeles †Arm Research 978-1-5090-6631-5/20/$31.00 ©2020 IEEE
[16] Guna Sekhar Reddy Thummala, Radhika Baskar, and RimlonShibi S, "Prediction of Heart Disease using Random Forest in Comparison with Logistic Regression to Measure Accuracy", 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)