

Predicting Covid-19 Patients Recovery Chance Referring the Precondition Using Multi-Layer Perceptron (MLP) and SMLT

M. Dhivya¹, S. Suruthi², S. Srija³, A.V. Sravani⁴

¹Assistant Professor, Dept. of Computer Science & Engineering, Panimalar Inst. of Technology, Chennai, India

^{2,3,4}Student, Department of Computer Science and Engineering, Panimalar Inst. of Technology, Chennai, India

Abstract - The Coronavirus disease 2019 (COVID-19) pandemic, which originated in Wuhan China, has had disastrous effects on the global community and has overburdened advanced healthcare systems throughout the world, WHO is continuously monitoring and responding to this pandemic. The current rapid and exponential rise in the number of patients has necessitated efficient and quick prediction of the possible outcome of an infected patient for appropriate treatment using ML techniques. The aim is to predict machine learning based techniques for covid-19 recovery chances possible or not in best accuracy. The analysis of dataset is done by supervised machine learning techniques(SMLT) like logistic regression, KNN, decision tree, SVM, confusion matrix and Multi-Layer Perceptron is also used to capture several information's like, variable identification, uni-variate analysis, bi-variate and multi-variate analysis, missing value treatments and analyze the data validation, data cleaning/preparing and data visualization will be done on the entire given dataset, and the recovery chances of a Covid-19 patient is predicted.

Key words:SVM, KNN, logistic regression, decision tree, confusion matrix, Multi-Layer Perceptron.

1. INTRODUCTION

Machine learning is a branch of artificial intelligence, in which the machine can perform any kind of task and challenge without human help. Deep learning is the subset of machine learning which contains some advanced algorithms for data processing. These algorithms can be used to process both structured and unstructured data (those are obtained from social media, images, videos, etc.), these data are called as big data.

The covid-19 pandemic is an ongoing global pandemic caused by severe acute respiratory syndrome coronavirus 2. As it is a most dangerous and deadly

disease, it is very important to find the recovery rate or recovery chances of a covid-19 affected patient. By using different machine learning and deep learning algorithms, we have created a platform to find the recovery chances of a covid-19 patient.

2. LITERATURE SURVEY

M. Qjidaa , Y. Mechbal , A. Ben-fares, H. Amakdouf and M. Maaroufi proposed an intelligent clinical decision support system, which has the capability to diagnose the covid-19 at its earliest stage with the help of chest x-ray. This idea came from the fact that, the other modes of detection of covid-19 are expensive and the people in rural area cannot afford to such expensive methods. For this method, three classes of samples are collected, they are the samples from covid-19 detected patients, pneumonia patients and normal people. This method produces the best accuracy rate of 99%.

Alaa A. R. Alsaedy and Edwin K. P. Chong produces a new strategy which uses existing cellular network functionality to find the area with high human crowd which is more prone to spread of covid-19. This strategy will focus on the cellular radiation, which can be captured by the user equipment carried by every person. This cellular radiation will find the region and categorizes it as the at-risk region and low-risk region.

Shuvankar Roy, Mahua Nandy Pal and Sonali Bhattacharyya developed a website, which provide each and every information about covid-19. It also gives accurate prediction details of confirmed and deceased cases of covid-19.

KhandakerFoysal Haque, Fatin Farhan Haque and Lisa Gandy proposed a CNN based model to detect covid-19 from patient’s chest x-ray. The sample x-ray collected are equally divided into two classes namely, covid-19 sample and normal sample. These samples are divided for training and testing phase. For detection of any disease time and accuracy are the two most important factors, with the help of CNN, those two factors are perfectly satisfied in this model.

Xi Ouyang, JiayuHuo, Liming Xia, Fei Shan proposed a diagnose method that distinguish covid-19 and pneumonia. Since, both pneumonia and covid-19 have some effects and symptoms in common, it is difficult to diagnose the correct one. Hence, a 3-D CNN model is developed to diagnose covid-19 and pneumonia with online attention refinement and dual sampling strategy.

3. EXISTING SYSTEM

The existing model contains a mathematical and numerical analysis to answer certain basic questions about the effect of Covid-19. This model includes a time-dependent SIR model which tracks the transmission rate and recovery rate at certain time. This model is further extended to analyze the spread of this infectious disease by considering two types of infected persons, (i.e.) detectable infected persons and undetectable infected persons. And also, the effect of social distancing in reducing the effective reproduction number is also explained. This model consists only the mathematical analysis and no machine learning or deep learning algorithms are used.

4. PROPOSED SYSTEM

The proposed method is building a machine learning model to classify the patient recovery chances. To make it more user friendly, GUI is used to create a user interface. The dataset is first preprocessed and the columns are analyzed to see the dependent and independent variable and then different machine learning algorithms would be applied to extract patterns and to obtain results with maximum accuracy.

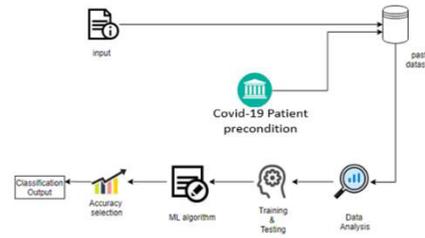


Fig-1: Architectural Diagram

5. RELATED WORKS

5.1 Logistic Regression

Logistic regression is similar to linear regression, except logistic regression predicts whether something is true or false instead of predicting something continuous like size. Also, instead of fitting a line to the data logistic regression fits an s-shaped logistic function. There are various types of logistic regression, they are: Binary logistic regression, Multinomial logistic regression, Ordinal logistic regression. The basic idea of this algorithm is to find and establish the relationship between the features and probability of particular outcome.

5.2 Decision Tree

The structure of the decision tree is similar to the structure of flow chart, which consist of internal node, branch and leaf node. The decision tree consists of three types of node, the first one is the decision nodes, which are generally represented by squares, the second node is the chance node, which is represented by circles, then finally the end nodes, which are represented by triangles. These decision trees are basically used in handling non-linear data sets and are widely used in all engineering areas and businesses.

5.3 K-Nearest Neighbor (KNN)

The most important supervised machine learning algorithm is the K-Nearest Neighbor algorithm, which is widely used to solve both classification and regression problems. Though this algorithm is used for both classification and regression problems, they are

prominently used for classification predictive problems in most of the industries. The KNN algorithm does not have separate training and testing phases, instead they utilize the whole dataset during the training phase, hence they are called as lazy learning algorithm.

5.4 Random Forest

Random forest is the classification predictive algorithm which consists of a group of decision trees that act as a single unit. Since this algorithm uses a group of decision trees, the result produced by this will be more accurate and stable.

5.5 Support Vector Machines

The support vector machine yields a high predictable rate and it categorizes the data set by setting an optimal hyper plane between the data. Though this support vector machine is used widely in classification challenges, it can also be used for regression and also in outlier detection.

6. FUTURE ENHANCEMENT

The future work will be automating this process by showing the prediction result in web application or desktop application, as it is more user friendly and can be easily used by all category of users. This work can also be optimized by implementing in Artificial Intelligence environment, as automation, smart decision making, enhanced customer experience, research and data analysis, solving complex problems, minimizing errors are simple and more efficient with artificial intelligence environment.

7. CONCLUSION

A machine learning model is built to classifying the covid-19 patients recovery chances, the dataset is preprocessed and analyzed, then the pattern is extracted using various machine learning and deep learning algorithms. By consistent comparing of the performance of different algorithms, the best one is finalized to produce maximum accuracy.

ACKNOWLEDGEMENT

We thank the management of Panimalar Institute of Technology and the faculties in computer science and engineering department for their support and guidance.

REFERENCES

- [1] H. Nishiura, T. Kobayashi, T. Miyama, A. Suzuki, S.-m. Jung, K. Hayashi, R. Kinoshita, Y. Yang, B. Yuan, A. R. Akhmetzhanov et al., "Estimation of the asymptomatic ratio of novel coronavirus infections (covid-19)," *International journal of infectious diseases*, vol. 94, p. 154, 2020.
- [2] T. Zhou, Q. Liu, Z. Yang, J. Liao, K. Yang, W. Bai, X. Lu, and W. Zhang, "Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-ncov," *Journal of Evidence-Based Medicine*, 2020.
- [3] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong et al., "Early transmission dynamics in Wuhan, china, of novel coronavirus–infected pneumonia," *New England Journal of Medicine*, 2020.
- [4] N. van Doremalen et al., "Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1", *New England J. Med.*, vol. 382, no. 16, pp. 1564-1567, Apr. 2020.
- [5] R. Sujatha, J. M. Chatterjee and A. E. Hassanien, "Correction to: A machine learning forecasting model for COVID-19 pandemic in India", *Stochastic Environmental Research and Risk Assessment*, vol. 34, pp. 959-972, 2020.

[6] W.-J. Guan et al., "Clinical characteristics of 2019 Novel Coronavirus infection in China", *New England J. Med.*, vol. 382, pp. 1708-1720, 2020.