

A Survey on Machine Learning based State Wise Forecast of Covid-19 Cases using Seir and Time Series Model

Shika A¹, Dimple Kanvar², Chandan³, Yogita Devoor⁴, Prof. Shruti.NG⁵

^{1,2,3,4} CSE Department, Sri Krishna Institute of Technology, B'lore-560090, India.

⁵ Faculty CSE Department, Sri Krishna Institute of Technology, B'lore-560090, India.

Abstract: Machine learning (ML)-based forecasting techniques have demonstrated their use in predicting perioperative outcomes and improving decision-making about future actions. Many application fields that required the detection and prioritization of adverse aspects for a threat have long used machine learning models. To deal with forecasting challenges, a variety of prediction approaches are widely utilized. This research illustrates the ability of machine learning models to predict the number of patients who would be afflicted by COVID-19, which is now regarded as a possible threat to humanity. Each model makes three types of predictions: the number of newly infected cases, the number of fatalities, and the number of recoveries. It's critical to examine the disease's spread in each state separately, as the circumstances are often quite different. To examine data on the number of people who have been infected in each Indian state and forecast the number of illnesses in that state. We anticipate that such state-by-state forecasts will assist the federal government in better allocating its limited healthcare resources. The extent of the pandemic, the recovery rate, and the death rate can all be predicted using data on COVID19 disease transmission.

Corona virus cases, deep learning, machine learning, SEIR, and Time Series Model are some of the terms used.

I. INTRODUCTION

India is a large country with a land area of 3,287,240 square kilometers and a population of over 1.3 billion people. The majority of Indian states are relatively substantial in terms of both area and population. When analyzing coronavirus infection data, assuming that India as a whole is on the same page may not provide us the best picture. This is due to the fact that each state's first infection, new infection rate, progression through time, and preventive measures adopted by state governments and the general population varies. We must address this issue. each state on its own. It will make it possible for government to make the most use of the few resources available.

Due to limited resources, the techniques to dealing with the two states must be distinct. Looking at when each state was originally infected is one technique to distinguish the state-by-state trends. The current approach employs a variety of methodologies to forecast future covid19 cases, including the number of newly infected patients, deaths, and recoveries. The logistic approach, the exponential method, the susceptible infectious recovered (SIR) method, linear regression (LR), least absolute shrinkage and selection operator (LASSO), support vector machine (SVM), and others are some of the strategies employed by existing systems.

They employ a Logistic Growth Curve model for shortterm forecasting, SIR models for forecasting the maximum number of active cases and peak time, and a Time Interrupted Regression model to assess the impact of lockdown and other interventions. The short-term outlook for India and high-incidence states is properly predicted by the logistic growth curve model. The SIR model's forecast might be utilized to plan and prepare health systems. The research also concludes that there is insufficient data to establish that lockdown has a favorable influence on the number of new cases.

The following aims are the major emphasis of the article:

- Predictive modelling of the covid19 lifetime pattern in different states.
- To anticipate the number of newly infected cases, the number of deaths, and the number of recoveries in the future.
- To assist the federal government in allocating resources based on forecasts.

The extraordinary 2019 new coronavirus epidemic, dubbed COVID-19 by the World Health Organization (WHO), has put a number of countries throughout the world in jeopardy. The impact of the COVID-19 epidemic, which was previously limited to Chinese nationals, has now become a source of severe concern for almost every country on the planet. A majority of these nations have gone into partial or total lockdown due to a lack of resources to deal with the COVID-19 epidemic and the concern of overwhelmed healthcare systems. We also take a wide look at the health of the world economy following the outbreak to calibrate the COVID-19's terrible impact.

II. BACKGROUND STUDY (LITERATURE)

This research presents an internet-based approach for detecting the Covid-19 virus. To raise awareness of the virus's propagation, it employs technologies such as web collecting data, model implementation, and a userfriendly online platform interface.

Future projection has a significant part in reducing the strain on the healthcare system. It will assist medical personnel in making the necessary plans for resource allocation. In situations like the worldwide pandemic caused by the covid-19

virus, when the scenario is constantly changing due to government constraints, dynamic projection is critical.

The suggested approach is divided into three stages:

1. Web scraping, also known as web harvesting, is a technique for extracting vast volumes of data from the internet and storing it in an organized fashion.
2. Model implementation: for covid-19 data prediction
3. User-friendly web platform interface: to allow individuals to share data.

Without any human intervention, the system refreshes the data from the source and teaches itself to forecast the outcome for the following 5 days. The dataset comes from the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) and is updated on a regular basis, thanks to the ESRI Living Atlas Team and the Johns Hopkins University Applied Physics Lab's help (JHU APL).

In this work, an intelligent clinical decision support system for corona virus identification in chest x-rays is proposed. The data is divided into three categories using deep learning algorithms: COVID-19, Pneumonia, and Normal. The work is based on artificial intelligence research in health services, which proposes employing chest x-ray machines to detect the covid-19 virus early in distant places. It can help reduce the growth of covid-19 cases in rural areas by serving as a clinical diagnostic prior to the covid-19 test.

For improved prediction, the dataset was separated between test and training sets using the stratify option. A total of 566 photos were used, with 70% of the data being used for training and 30% being used to test the model. TensorFlow and Keras(wrapper library) in Python were utilized to create the CNN model. The tests were conducted on a Lenovo ThinkPad p51 with an Intel® Core™ i7-7820HQ processor running at 2.90 GHz, an NVIDIA Quadro M2200 8GB GPU, and 8GB of RAM. The Adam optimization approach was utilized for hyperparameter optimization for training CNN models with cross-entropy as the loss function. After the data has been preprocessed, it is enhanced with rotation, horizontal flip, channel shift, and rescale.

The accuracy of machine learning algorithms in predicting the spread of the Covid-19 virus is demonstrated in this research. It suggests using four conventional forecasting models:

1. Linear regression (LR)
2. Least absolute shrinkage and selection operator (LASSO)
3. Support vector machine (SVM) and
4. Exponential smoothing (ES)

They forecast the number of new cases, fatalities, and recovered cases during the following 10 days. According to the findings, ES yielded the most promising outcomes, followed by LR and LASSO, however SVM fell short of expectations. The data set provided by John Hopkins University and GitHub was used to train this model.

The data set has been preprocessed and divided into training and test data. The data is utilized to train the model 85 percent of the time, and the remaining 15% is used to test it. Using the ml algorithms, the model evaluates the provided dataset and predictions future scenarios.

III. METHODOLOGY

The following mathematical model is used to determine the increasing and falling of covid-19 cases state wise in this system. Overall, for many there were major illness rate where the incubation period was lengthy enough during which infected individuals are not that infectious. This is said to be susceptible exposed infectious recovered (SEIR). At times, the person is in compartment E. (for exposed) The SEIR model is a significantly assessed model for characterizing a variety of illnesses. The population is classified based on the following groups per the model:

- Susceptible(S). These group of people are not infected yet.
- Exposed(E). The individuals who fall under this group are infected but not yet infectious
- Infectious(I). This group of inmates could infect the other people.
- Recovered(R). These group of people have recovered.

Five differential equations describe how with respect to time the number of people changes in these groups. We ignore the rigorous version rather we consider only simplified version because we ignore the birth and the death rate from other causes.

$$\begin{aligned}\frac{dS}{dt} &= -\beta S \frac{I}{N} \\ \frac{dE}{dt} &= \beta S \frac{I}{N} - \epsilon E \\ \frac{dI}{dt} &= \epsilon E - (\gamma + \alpha) I \\ \frac{dR}{dt} &= \gamma I \\ \frac{dD}{dt} &= \alpha I\end{aligned}$$

The SEIR model parameters are:

- Alpha(α) is a disease-induced mean fatality rate.
- Beta(β) is the probability of disease transmission per contact times the number of contacts per unit time.
- Epsilon(ϵ) is the rate of progression from infectious period of exposure. It is the inverse of the incubation period.
- Gamma(γ) is the recovery rate. It's the reciprocal of the period of infection.

In order to use the SEIR model for forecast the population we must master their parameters as well as the following values to be accounted in the beginning:

Exponential smoothing (ES) is a time series forecasting method for univariate information. It can extend to support data with a seasonal component or systematic trend. We will be implementing it using Time Series Forecasting Model which is used to predict the futuristic trends based on past behavior.

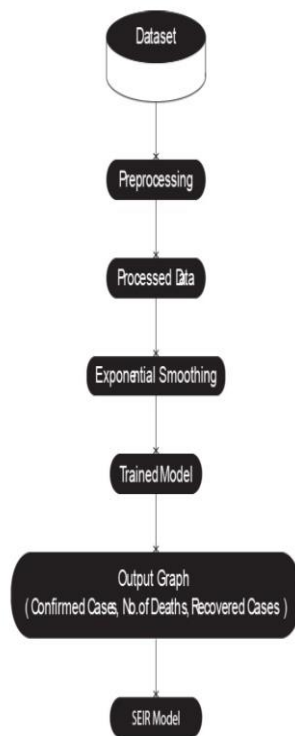


Fig 1. Proposed Work Flow

This study has some key findings which are listed below:

- ES performs best when there very limited entries in the time-series dataset.
- Different Machine Learning algorithms seem to act and perform better in changing class predictions.
- Most of the Machine Learning algorithms predict the future with ample amount of data, as the size of the dataset elevates the model performances also improves which is directly proportional.
- Machine Learning based forecasting can be very reliable and necessary tool for decision-makers to contain pandemics such like COVID-19.

IV. CONCLUSION

This paper objectifies to attain and develop a machine learning based time-series model for forecasting the futuristic covid19 cases such as newly infected cases, the number of death and the number of recoveries. In other words, this project aims at presenting a time series model for identifying and analyzing the corona virus cases using the pictorial graphs for consideration. The main motive of this model is to carry out and upfront the predictive modelling of the covid19 lifecycle pattern in state wise manner and also to assist the central government to allocate the resources.

V. ACKNOWLEDGEMENT

We would like to thank our guide Prof. Shruti NG for her valuable suggestion, expert advice and moral support in the process of preparing this paper.

REFERENCES

- [1] 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, p. 959-972, 2020.
- [2] S. Lalmuanawma, J. Hussain, and L. Chhakchuak, "Applications of machine learning and artificial intelligence for Covid 19 (SARS-CoV-2) pandemic: A review," *Chaos, Solitons & Fractals*, vol. 139, p. 110059, 2020.
- [3] A. Alimadadi, S. Aryal, I. Manandhar, P. B. Munroe, B. Joe, and X. Cheng, "Artificial intelligence and machine learning to fight COVID-19," *Physiological Genomics*, vol. 52, p. 200-202, 2020.
- [4] N. S. Pun, S. K. Sonbhadra, and S. Agarwal, "COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms," 2020.
- [5] F. A. B. Hamzah, C. H. Lau, H. Nazri, D. V. Ligot, G. Lee, C. L. Tan, M. K. B. M. Shaib, U. H. B. Zaidon, A. B. Abdullah, M. H. Chung, C. H. Ong, and P. Y. Chew, "CoronaTracker: Worldwide COVID-19 Outbreak Data Analysis and Prediction," *Bull World Health Organ*, 2020.
- [6] M. Mandal, S. Jana, S. K. Nandi, A. Khatua, S. Adak, and T. Kar, "A model based study on the dynamics of COVID-19: Prediction and control," *Chaos, Solitons & Fractals*, vol. 136, p. 109889, 2020.
- [7] V. Soukhovolsky, A. Kovalev, A. Pitt, and B. Kessel, "A new modelling of the COVID 19 pandemic," *Chaos, Solitons & Fractals*, p. 110039, 2020.
- [8] World Health Organization (WHO), "Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)," WHO, 2020.
- [9] World Health Organization (WHO), "Coronavirus disease 2019 (COVID19) Situation Report- 13," World Health Organization, 2020.
- [10] H. Dai, E. B. Khalil, Y. Zhang, B. Dilkina, and L. Song, "Learning combinatorial optimization algorithms over graphs," in *Proc. Adv. Neural Inf. Proc. Syst.*, Dec. 2017, pp. 6348-6358.
- [11] V. Gemmetto, A. Barrat, and C. Cattuto, "Mitigation of infectious disease at school: Targeted class closure vs school closure," *BMC Infectious Diseases*, vol. 14, no. 1, p. 695, Dec. 2014.
- [12] B. Wang, Y. Sun, T. Q. Duong, L. D. Nguyen, and N. Zhao, "Security enhanced content sharing in social IoT: A directed hypergraph-based learning scheme," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4412-4425, Apr. 2020.
- [13] D. K. Chu, E. A. Akl, S. Duda, K. Solo, S. Yaacoub, and H. J. Schünemann, "Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: A systematic review and metaanalysis," *Lancet*, Jun. 2020, doi: 10.1016/S01406736(20)31142-9.
- [14] L. F. R. Ribeiro, P. H. P. Saverese, and D. R. Figueiredo, "Struc2vec: Learning node representations from structural identity," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2017, pp. 385-394.
- [15] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, "Random graph models of social networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 1, pp. 2566-2572, Feb. 2002.
- [16] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck, "What's in a crowd? Analysis of face-to-face behavioral networks," *J. Theor. Biol.*, vol. 271, no. 1, pp. 166-180, Feb. 2011.
- [17] K. V. Aadithya, B. Ravindran, T. P. Michalak, and N. R. Jennings, "Efficient computation of the Shapley value for centrality in networks," in *Internet and Network Economics*, A. Saberi, Ed. Berlin, Germany: Springer, 2010, pp. 1-13.
- [18] J. B. Dowd, L. Andriano, D. M. Brazel, V. Rotondi, P. Block, X. Ding, Y. Liu, and M. C. Mills, "Demographic science aids in understanding the spread and fatality rates of COVID-19," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 18, pp. 9696-9698, Apr. 2020, doi: 10.1073/pnas.2004911117.
- [19] W. Song, C. Shi, Z. Xiao, Z. Duan, Y. Xu, M. Zhang, and J. Tang, "AutoInt: Automatic feature interaction learning via self-attentive neural networks," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 1161-1170, doi: 10.1145/3357384.3357925.
- [20] J. Gao. (2014). *Machine Learning Applications for Data Center Optimization*. Google. Mountain View, CA, USA. [Online]. Available: <https://static.googleusercontent.com/media/research.google.com/en/pubs/archive/42542.pdf>