

Prediction of Omicron cases in India using LSTM: An advanced approach of artificial intelligence

Nikku Yadav

Clinical Research Under Department of Community Medicine

Swami Rama Himalayan University

Dehradun

Uttarakhand

India

Dilip Kumar Jang Bahadur Saini *

Department of Computer Science and Engineering

Himalayan School of Science & Technology

Swami Rama Himalayan University

Uttarakhand

India

Akanksha Uniyal

Departments of Biostatistics

Swami Rama Himalayan University

Dehradun

Uttarakhand

India

Nidhi Yadav

Clinical Trail Center

Swami Rama Himalayan University

Dehradun

Uttarakhand

India

Maya S. Bembde

IT Department

JSPM's Rajarshi Shahu College of Engineering

Pune

Maharashtra

India

* E-mail: dilipsaini@gmail.com

Dharmesh Dhabliya

Department of Information Technology

Vishwakarma Institute of Information Technology

Pune

Maharashtra

India

Abstract

Several mutants of SARS COV-2 have been reported so far, each one having different severity and infectivity rate. Omicron has enhanced resistance to host immunity and displays higher transmission. 15 of the 37 known mutations in the spike protein's receptor-binding domain (RBD), the primary target of neutralising antibodies, are found there. An enhanced RNN, or sequential network, called a long short-term memory network, permits informatics data to endure. It is capable of resolving the RNN's vanishing gradient issue. RNN, also referred to as a recurrent neural network, is utilized for persistent memory. "In medication, the right analysis and the perfect opportunity are the keys for effective treatment. For the current review, the gathered Past information was utilized to train the model, and afterward this trained model was utilized to test new information and afterward for utilized for the advancement of the prediction model". The trained ML model's presentation or approval was assessed utilizing some piece of accessible past datasets and this method known as validation process. According to findings of the current study, Indian citizens can aid their nation in the fight against Omicron by adhering to the directions and instructions provided by the Indian government. We ought to cooperate to overcome the new lethal variation of Coronavirus.

Subject Classification: 97R40, 97R50.

Keywords: Artificial intelligence, Long short term memory (LSTM), Omicron, Covid 19 variant.

1. Introduction

The globe is currently experiencing a new pandemic brought on by SARS-CoV-2, a zoonotic virus, as a result of the appearance of the SARS-CoV-2 strain of the human coronavirus, and its variants has led to a phase of stagnancy in terms of growth worldwide by affecting more than 34.9Cr people and causing the death of more than 55.9 Lakh people [1]. December 2019 marked its beginning when people with pneumonia-like symptoms were observed in Wuhan, Hubei province of China. SARS-CoV-2 have wrapped infections with positive RNA having a place with the Coronaviridae family Nidovirales which are separated into four general (α , β , γ , and δ) [2]. Several mutants of SARS COV-2 have been reported so far, each one having different severity and infectivity rate.

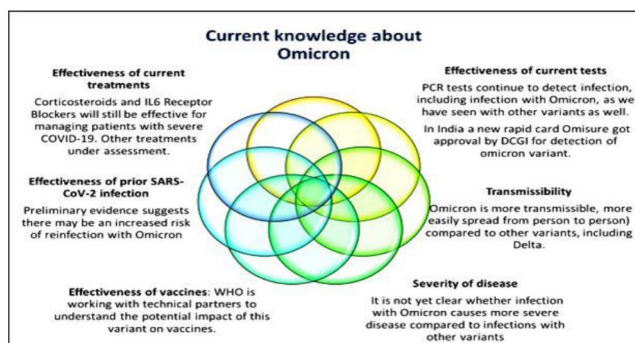


Figure 1

Summarizing the current information on Omicron Researchers in the current study questioned whether the RBD of the Omicron

On November 24, 2021, a new variation B.1.1.529 was detected in South Africa [3]. This variation has an enormous number of changes, some of which are concerning (as displayed in Figure 1). Furthermore, the preliminary proof proposes a high risk of disease with this new variation when contrasted with different variations like delta variation. The most broadly involved analytic technique for SARS Cov-2is the real time PCR, yet on account of Omicron the one of the three genes is unable to detected(called S gene or S target failure) as aftereffect of transformations in the S gene (<https://www.gisaid.org/>). Also, S gene based sequencing and distinguishing proof of change can be utilized for the discovery of this variation of concern Omicron has greater host immune resistance and higher transmission rates. 15 of the 37 known mutations in its receptor-binding domain (RBD), the primary target of neutralizing antibodies, are found in the spike (S) protein.

For the infection to start and enter the host cell, the RBD on the spike protein is essential. It is critical to look into the effects of these mutations in the Omicron form because the majority of medications target the spike protein.

Version binds strongly to the ACE2 and whether an alternate receptor may make it easier to infect healthy cells [4]. It is crucial to comprehend the processes by which these mutations arise in order to develop vaccinations and antibody therapies that are successful (Figure 1). One of the application of machine learning was Image Fusion, is a technique of selecting all the necessary details of several images and fusing them into one single image which is way more accurate, detailed and enhanced in comparison

to source images [5]. In the discipline of image restoration, noise in an image is crucial. The photos are distorted by various types of noise. One of the most difficult problems and least discussed ones is how to reduce noise in satellite photos. The term “Speckle noise” refers to the type of granular pattern seen in radar coherent pictures [6]. Long Short-Term Memory Network is an advanced RNN, a successive organization, that permits data to persist. It is equipped for taking care of the disappearing angle problem faced by RNN. A recurrent brain network otherwise called RNN is utilized for persistent memory. Likewise, RNNs work, they recollect the past data and use it for handling the current input [7]. The limitation of RNN is, they cannot recollect long term conditions due to vanishing gradient. LSTMs is expressly intended to keep away from long-term dependency problems. This prediction/forecast model will permit undertaking explicit judgments in light of transmission development, for example, growing the lockdown stage, playing out the disinfection plan, and offering everyday help and supplies. A study conducted by Wadha *et al* 2021 for prediction the time period of extension of lockdown due to increase in rate of COVID-19 [8].

1.1 *Significance of machine learning for covid-19*

“Machine learning is utilized in different fields, including medication and also to prognosticate illness and conjecture outcomes. In medication, the right conclusion and the perfect opportunity are the keys to fruitful treatment. On the off chance that the treatment has a high blunder rate, it might cause several deaths. Hence, analysts have begun involving man-made reasoning applications for medical treatment. The undertaking is muddled on the grounds that the scientists need to pick the right device: it involves life or death [7]. For this errand, ML accomplished an achievement in the field of medical care. ML methods are utilized to decipher and break down huge datasets and foresee their output. These ML apparatuses were utilized to recognize the side effects of illness and order tests into treatment gatherings. ML assists medical clinics with keeping up with managerial cycles and treats irresistible infection”.

ML procedures were recently used to oversee malignant growth, pneumonia, diabetes, Parkinson disease [9], neuromuscular issues, joint pain, and a lot more sicknesses; they give over 90% precise outcomes in expectation and guaging.

The deadly sickness “COVID-19” has claimed the lives of numerous people all over the world. This virus has no known cure. ML procedures

[10,11,12] have been utilized to anticipate whether patients are contaminated by the infection in view of side effects characterized by WHO and CDC [13]. ML is likewise used to analyze the illness in light of x-beam pictures. For example, chest pictures of patients can be utilized to distinguish whether a patient is infected with COVID-19 [14,15]. In addition, social separating can be checked by ML; with the assistance of this methodology, we can protect ourselves from COVID-19.

1.2 *“Different techniques for prediction and forecasting”*

Different ML methods are utilized to foresee and figure future occasions. Some ML strategies utilized for expectation are neural networks (multilayer perception), “K-nearest neighbor, decision trees (random forest etc), support vector machine, naive Bayes, and linear regression”.

1.3 *ML Application to Omicron data sets*

The data for Omicron cases from India were downloaded from Press Information Bureau (<https://www.pib.gov.in/allRel.aspx>) Govt website starting from 2 Dec 2021 In Figure 2 (The first case was reported in Karnataka, India) to 30 January 2022. The data was exported in excel then deep machine learning tools were applied. India is seeing a steep rise in COVID-19 Omicron cases Only 3 new cases were registered on 2 December 2021.

But in just 60 days, on 30 January 2022, a whopping 16300 new cases were registered. The curve of the third wave is very steep and that’s why it is a big concern right now. So, we thought about using the artificial recurrent neural network (RNN) architecture Long Short-Term Memory (LSTM) to predict how the COVID-19 graph will look in near future (next 90 days).

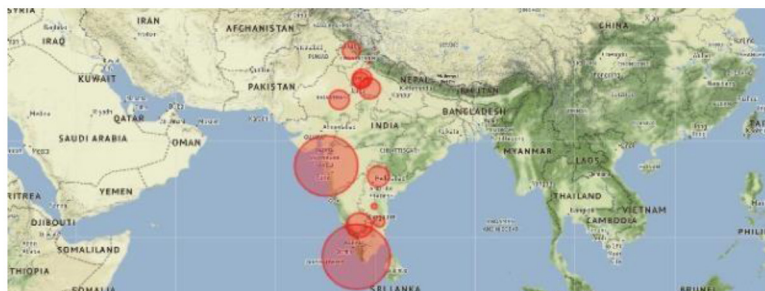
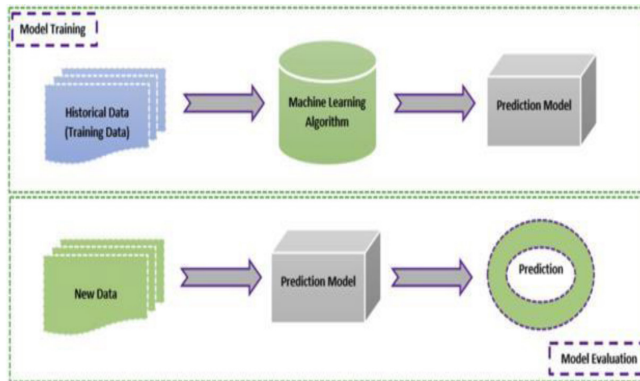


Figure 2

Omicron geographical region spread and hotspot zones

**Figure 3****Machine Learning Process**

The course of Machine Learning (ML) is basically gaining for a fact or perceptions from past work, like models, or guidance, to search for designs in information and with the assistance of models, [16] and Figure 2 shows the course of “Machine Learning (ML)”.

The model was trained using the previously gathered historical data, and it was then tested against fresh data before being used to create a prediction model. Using a fraction of the previously accessible data sets, the trained ML model’s effectiveness or validation was assessed (which is not present during training). The ML model’s performance was assessed during the validation process using metrics like accuracy. In terms of the ratio of characteristics that were properly predicted to all features that might be predicted, accuracy measured how well the ML model performed on unknown data [17] shown in Figure 3.

2. Methodology

Step 1: Import Libraries:

First of all imported the necessary libraries required for achieving the project objective.

Step 2: Data Pre-Processing:

“From the dataset all_totals.JSON, and created a data frame that contains new cases registered every day. The data frame named df_new contains new cases registered on each day (some dates are missing), starting from 2020-01-30”.

Step 3: ‘Date’ field the index.

We used the MinMaxScaler to bring our data in the 0–1 range. Then we reshaped it to make it just one column and n number of rows (n = number of elements in the array). After that, we converted the array to a list.

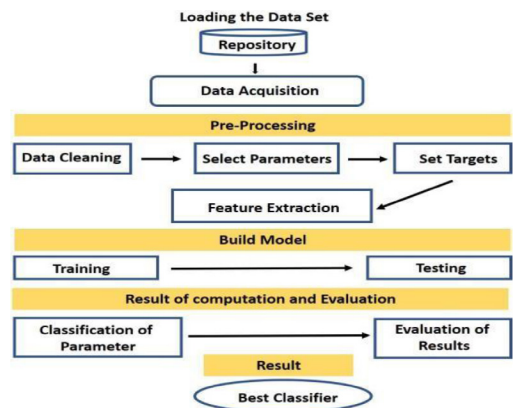


Figure 4
Machine learning process

Step 4: Training data: For every 90 points in one list, the 91st point was the label of them. Then we created a list of lists with all the lists containing 90 points. We created another list containing the labels. Later, all those lists are converted into NumPy arrays and NumPy arrays or arrays. Then plotted the predictions given by LSTM along with the real values of newly registered COVID-19 Omicron cases each day to validate the accuracy of the prediction model shown in Figure 4.

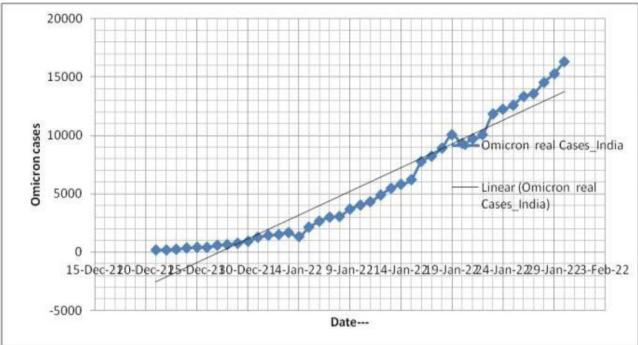


Figure 5 (i)
Omicron real cases in India

The projected values are highly correct, and the model is very promising, according to the graph (figure 5 i and ii). Therefore, it might be used to forecast the future.



Figure 5 (ii)
Omicron real cases versus predicted cases for Dec 2021

3. Future Predictions

Now we were trying to predict this curve in the near future for the January month prediction. We tried to predict how many cases new cases of COVID-19 Omicron were registered in India for Jan 2022. The prediction model was developed using a seed. A seed is the data of 45 days in the array. We used this seed to predict the data for the next day. When it was predicted then we updated our seed and made it contain the predicted data that was used in the next prediction.

The seed first contained the last 45 days of data (values of how many new covid cases are registered each day), ending on January 10, 2022. Using it, we predicted the value for 11 January 2022. Then we used the

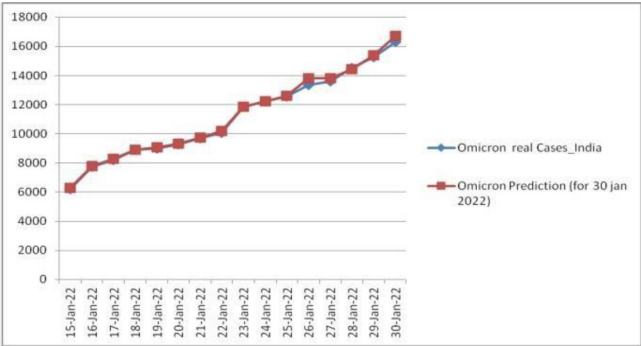


Figure 6
Omicron real cases versus predicted cases for Jan 2022

predicted data of 11 January 2022 as a data point in the seed and predicted the value data for 12 January 2022 and so on. This model was trained to predict the number of cases for the next 15 days shown in Figure 6.

4. Conclusion

To battle with the Omicron (we) residents of India can help our nation by adhering to the entryways and rules which are given by the Government of India. We ought to cooperate to overcome the new dangerous variation of Coronavirus. Due to our administration's stand now it appears to be that the third wave of COVID-19 won't show up in India, we as a whole maintain that it should work out, however we should be ready for it, we can't take the new variation of Coronavirus daintily. It is great to hear that even a month after the distinguishing proof of the new COVID- 19 variation India has just revealed 200 cases, however nobody understands what will occur from here on out on the off chance that the new variation of Coronavirus breakouts, the third flood of COVID-19 will definitely come in India.

References

- [1] Coronavirus disease (COVID-19): Variants of SARS-COV-2. 4 December 2021. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- [2] Pal, M., Berhanu, G., Desalegn, C., & Kandi, V. Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2): An Update. *Cureus*, 12(3), e7423-e7423 (2020). doi:10.7759/cureus.7423
- [3] Epidemiological Update Coronavirus disease (COVID-19). 2 December 2021
- [4] Holmes, E. C., Goldstein, S. A., Rasmussen, A. L., Robertson, D. L., Crits-Christoph, A., Wertheim, J. O., . . . Rambaut, A. The origins of SARS-CoV-2: A critical review. *Cell*, 184(19), 4848-4856 (2021). doi:<https://doi.org/10.1016/j.cell.2021.08.017>.
- [5] Tyagi, T., Gupta, P., & Singh, P. A hybrid multi-focus image fusion technique using SWT and PCA. Paper presented at the Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering, 491-497 (2020). doi:10.1109/Confluence47617.2020.9057960.

- [6] Singh, P., & Shree, R. Speckle noise: Modelling and implementation. *International Journal of Control Theory and Applications*, 9(17), 8717-8727 (2016).
- [7] Lupala, C. S., Ye, Y., Chen, H., Su, X. D., & Liu, H. Mutations on RBD of SARS- CoV-2 Omicron variant result in stronger binding to human ACE2 receptor. *Biochem Biophys Res Commun*, 590, 34-41 (2022). doi:10.1016/j.bbrc.2021.12.079
- [8] Wadhwa, P., Aishwarya, Tripathi, A., Singh, P., Diwakar, M., & Kumar, N. Predicting the time period of extension of lockdown due to increase in rate of COVID-19 cases in india using machine learning. *Materials Today: Proceedings*, 37(Part 2), 2617-2622 (2020). doi:10.1016/j.matpr.2020.08.509.
- [9] Pal, M., Berhanu, G., Desalegn, C., & Kandi, V. Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2): An Update. *Cureus*, 12(3), e7423-e7423 (2020). doi:10.7759/cureus.7423.
- [10] SARS-CoV-2 B.1.1.529 (Omicron) Variant - United States, December 1-8, 2021. *MMWR Morb Mortal Wkly Rep*, 70(50), 1731-1734 (2021). doi:10.15585/mmwr.mm7050e1
- [11] Miodrag Zivkovic , Nebojsa Bacanin , K Venkatachalam et. al. COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach. *Sustain Cities Soc*, pages-1-6 (2021). doi: 10.1016/j.scs.2020.102669.
- [12] D. Gupta, A. Julka, S. Jain, T. Aggarwal, A. Khanna, N. Arunkumar, V.H.C. De Albuquerque. Optimized cuttlefish algorithm for diagnosis of Parkinson's disease, *Cognit. Syst. Res.* 5236e48 (2018).
- [13] S. Tuli, S. Tuli, R. Tuli, S.S. Gill, Predicting the Growth and Trend of COVID-19 Pandemic Using Machine Learning and Cloud Computing, *Internet of Things* (2020).
- [14] N.S. Pun, S.K. Sonbhadra, S. Agarwal, COVID-19 Epidemic Analysis Using Machine Learning and Deep Learning Algorithms, *MedRxiv* (2020).
- [15] L. Jia, K. Li, Y. Jiang, X. Guo, Prediction and Analysis of Coronavirus Disease 2019. *arXiv Preprint*, p. 05447 (2020), arXiv:2003.
- [16] I. Turaiki, M. Alshahrani, T. Almutairi, Building predictive models for MERS-CoV infections using data mining techniques, *J. Infect. Public Health* 09 (2016) 744e748.
- [17] U. Kose, G.E. Guraksin, O. Deperlioglu, Diabetes Determination via Vortex Optimization AlgorithmBased Support Vector Machines: Medical Technologies National Conference, pp. 1e4 (2015).