Comparative Predictive Analysis of Mortality Rate after Covid-19 Vaccination Using Various Machine Learning Approaches

Nikita Agarwal
Center of Intelligent Systems and Robotics - ITER
Siksha 'O' Anusandhan (Deemed to be University)
Bhubaneswar – 751030, India
nikitaagarwal@soa.ac.in

Ritam Dutta
Center of Intelligent Systems and Robotics - ITER
Siksha 'O' Anusandhan (Deemed to be University)
Bhubaneswar – 751030, India
ritamdutta1986@gmail.com

2022 International Conference on Computer Communication and Informatics (ICCCI) | 978-1-6654-8035-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCCI54379.2022.9740854

***Abstract*— This work has mainly targeted in performing comparative real time predictive analysis of mortality rate after having COVID-19 vaccination using different machine learning approaches. In this paper various deep learning models viz. RNN, LSTM and CNN have been utilized to make future prediction on mortality rate on the basis of administered vaccine doses. Firstly, the dataset of confirmed active cases, death cases and administered vaccine doses have been converted from time-series format to supervised learning format, and secondly different deep learning models have been trained and compared based on the transformed dataset. The prediction analysis is performed strictly based on the newest COVID-19 Delta Variant infected cases. The predictive analysis has resulted 15.53% of reduction in mortality rate and 24.67% of reduction in confirmed active cases with increase in vaccination rate.**

Keywords—Mortality rate; Predictive analysis; RNN; LSTM; CNN; Supervised learning; Deep learning model

I. INTRODUCTION

The novel coronavirus SARS-CoV-2 first reported in Wuhan fist market situated in China in December 2019. The disease has got much genetic similarities with Bat SARS coronavirus WIV1 [1], is a strain of severe acute respiratory syndrome related to coronavirus. This deadliest virus has spread across the globe exponentially through human-to-human contact and has become a pandemic. The virus has mutated and generated several more threatening variants than its earlier variants cause fatal disease. Total 23,65,99,025 active cases and 48,31,486 cumulative death cases has been reported globally by World Health Organization, out of which total 3,39,15,569 active cases and 4,50,127 cumulative death cases has been reported in India as of Oct. 08, 2021 [2]. The top five countries having the maximum number of deaths recorded due to COVID-19 as of Oct. 08, 2021 [2] has shown in table 1. Despite of multiple preventive measures taken by the government of maximum countries, the latest variant of this fatal disease is now affecting children and elderly people. Non-vaccination of below 18 aged people has been primarily the cause of this new spike. Even if the infected person is recovered, then also it impacts their physical health in a longer period in future. These effects have been termed as long-COVID-19 side-effects. Vaccines are invented to fight against this fatal disease [3]. It is strongly expected that with the increased rate of COVID-19 complete vaccination dosage will reduce and restrict the probability of further spread of the virus, and thus, the further mutation of the virus could be prevented.

In this paper, we have collected an explicit dataset representing the number of confirmed cases, the number of death cases observed since mid-January 2020, along with the number of vaccine doses administered started from mid-January 2021. Total number of COVID-19 vaccines administered in India is 61,23,60,533 as of October 08, 2021 [4]. The United Nations Health Agency has confirmed four Variants of Concern - Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1) and Delta (B.1.617.2) and two Variants of Interest - Lambda (C.37) and Mu (B.1.621) as of October 08, 2021. Different machine learning approaches have been incorporated to perform predictive analysis on vaccine usage, active confirmed cases and death cases strictly based on the newest COVID Variant i.e., Delta Variant. Though, in coming days the modified variants of COVID-19 virus may appear and largely or lessly affect human health. These results are then analyzed and compared between the confirmed and death case rate to the given vaccine dose rate.

In the following sections, recent literature review of COVID-19 virus effects has been provided, where various machine learning based algorithms have been discussed; followed by our proposed work and the experimental results. An analysis of the predictive results has been performed, and conclusions are made based on this analysis.

TABLE I. LIST OF TOP 5 COUNTRIES AFFECTED BY COVID-19

Country	Total Confirmed Cases	Total Deaths
United States	4,37,92,254	7,03,599
India	3,39,15,569	4,50,127
Brazil	2,15,16,967	5,99,359
United Kingdom	80,46,394	1,37,417
Russia	77,17,356	2,14,485

2022 International Conference on Computer Communication and Informatics (ICCCI) | 978-1-6654-8035-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCCI54379.2022.9740854

II. Literature Review

R. Ross [5] has explicitly estimated the SARS death rates using predictive analysis approaches, where C. I. Van der Made et. al. [6] have introduced the typical presence of genetic variants among young men with Severe COVID-19 in their literature. As by now the vaccination has been well under process in several virus affected nations, therefore the correlation between vaccination dose rates with respect to mortality rates is now an exciting field of research, which can predict the future roadmap for us to battle this deadliest virus effects. M. Tolba et. al. [7] have introduced CNN based deep learning approach to assess and characterize the Post-COVID-19 manifestations in December 2020. There have been several research works has currently been performed by researchers to predict the recovered rate and mortality rate with respect to many socio-economic factors [8-14]. Moreover, P. Ghosh and R. Dutta [15] have emphasized the statistical machine learning forecasting algorithm for discipline prediction and cost estimation of COVID-19 pandemic, which can help to prepare the future ground work to post-economic crisis at several backward countries across the globe.

III. Methodology Used

This section has emphasized on brief description of the dataset used in this research work, its pre-processing steps, and the deep learning models which have been utilized for prediction analysis.

A. Dataset description

The datasets have been collected from the World Health Organization [2, 4]. The number of confirmed and death cases in India have been collected from W.H.O [2] from 15 January 2020 to 30 September 2021. The dataset related to vaccine doses administered from 15 January 2021 to 30 September 2021 have been collected on daily basis from W.H.O [4]. Figures 1-3 show the graphical representation of number of confirmed cases, death cases and the vaccine doses administered, respectively.

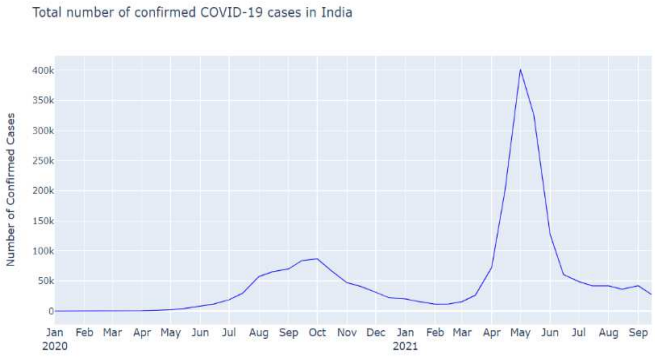

Figure 1. Total number of confirmed COVID-19 cases in India

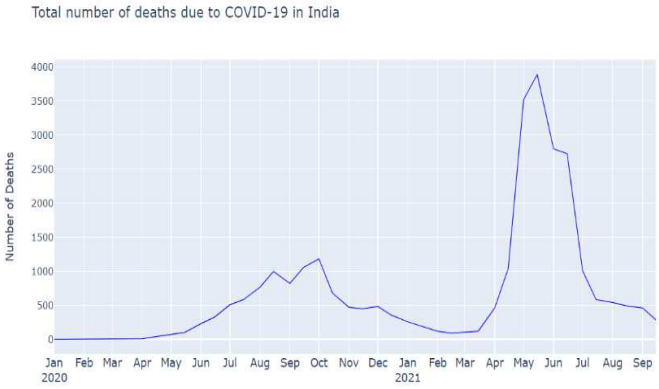

Figure 2. Mortality rate of COVID-19 Virus in India

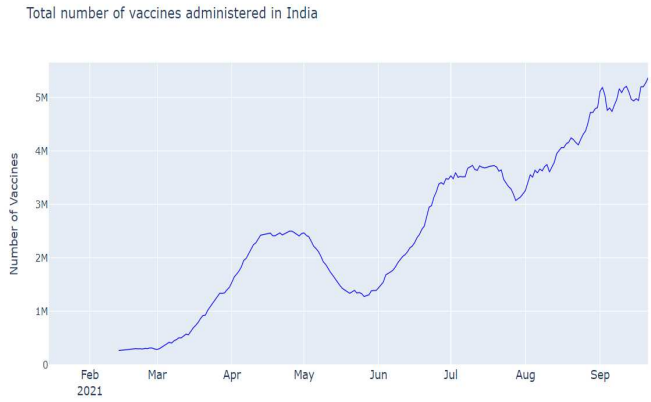

Figure 3. Vaccine doses administered in India

B. Data Pre-Processing

The dataset used in our work is in a time-series format. This data must be first converted to a supervised learning problem to use machine learning models. The dataset is arranged in such a way that the first five data is given as input, and the sixth data is considered as the output. The second to sixth values are then used as second sample and the seventh data as the output of this sample, and so on. This transformed data is then passed to train our machine learning models and make future predictions.

C. Models Used

This section describes the models used in our work. In this work, the deep learning models viz. RNN, LSTM, and CNN models were used to perform the predictive analysis on COVID-19 data and compare the results. RNN is the simplest recurrent deep learning model that can remember past information and make predictions. LSTM is the advanced version of the RNN that eliminates the problems of the RNN model. RNN model architecture is shown in figure 4.

978-1-6654-8035-2/22/$31.00 ©2022 IEEE

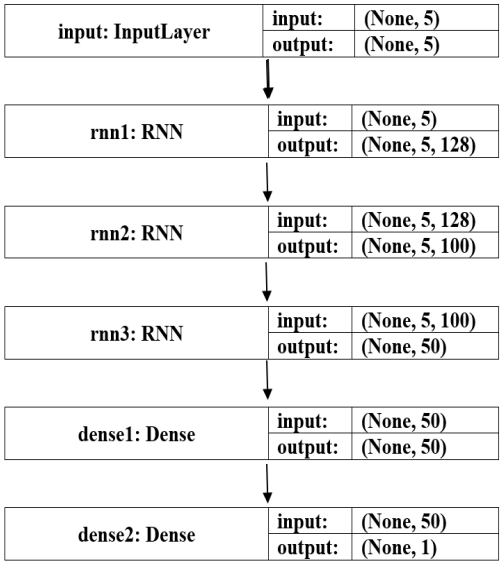

Figure 4. RNN Model used in our work

The LSTM and CNN models are primarily used deep learning models. CNN is a deep learning model that works well with image, text, numeric data. Both LSTM and CNN model architectures have been depicted in figure 5 and figure 6 below respectively.

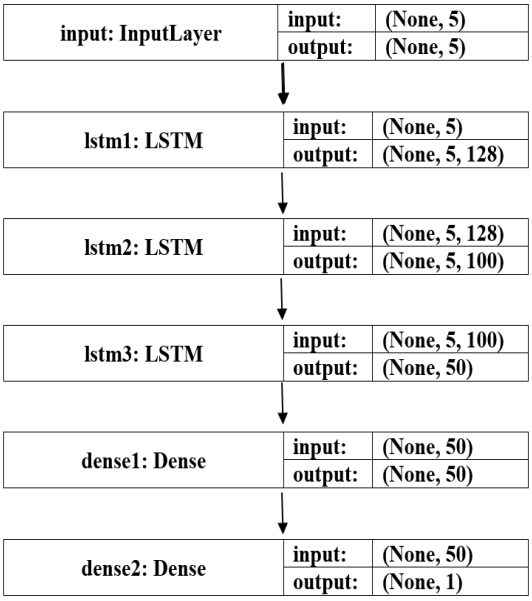

Figure 5. LSTM Model used in our work

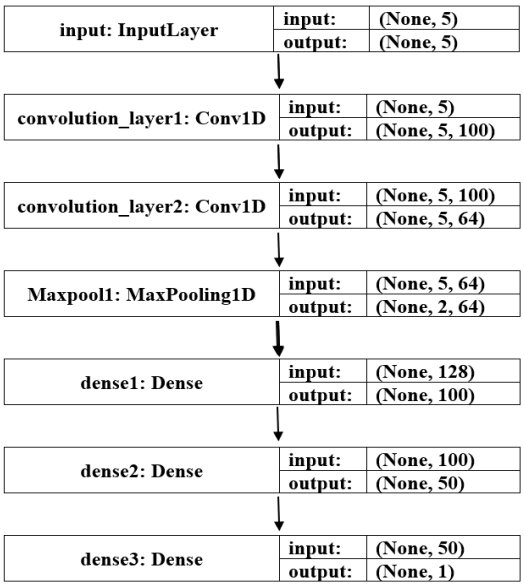

Figure 6. CNN Model used in our work

The predictive analysis results were analyzed to identify the rate at which people were getting affected by the virus, and the people died as a result of the infection. The rate has been calculated using the formula given below:

$$rate = c / p \quad (1)$$

where, c represents the number of confirmed cases/ death cases, p represents the population size of India. The population of India is taken from Govt. of India Census-2021 report as 1.39 billion as of Oct 2021 [16].

IV. Experimental Results

The experiment has been conducted to predict the number of confirmed COVID-19 cases, the death caused by the disease, and the number of vaccine doses to be administered in future. The timeline of future prediction has been considered till March 2022.

The machine learning models viz. RNN, LSTM, and CNN have been incorporated for the predictive analysis. The mean square error is used as the loss function and Root Mean Square Propagation (RMSprop) optimizer to optimize the models while training. Grid search has been used to optimize the parameters of the deep learning models. The parameters used in RNN, LSTM and CNN models have been mentioned in table 2 and 3. Figure 7-9 shows the future predictions made for confirmed COVID-19 cases, death cases and vaccine doses administered using these models respectively.

It has been observed that the CNN model has made good predictions considering all three parameters. The mortality rate and the rate at which people to be affected by the virus after making future predictions are 15.53% and 24.67%. It has been observed that the people getting affected by the COVID-19 virus and the mortality rate has been decreased as the number of vaccine doses increases.

978-1-6654-8035-2/22/$31.00 ©2022 IEEE

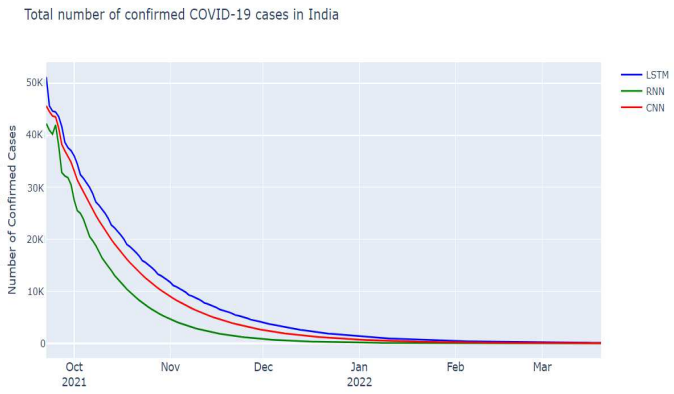

Figure 7. Future predictions of confirmed COVID-19 cases in India

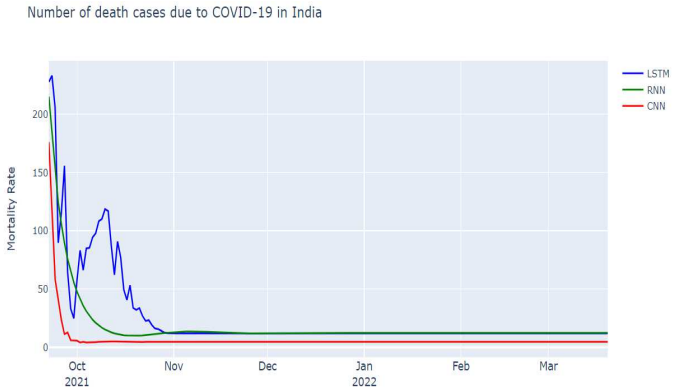

Figure 8. Future predictions of mortality rate due to COVID-19 in India

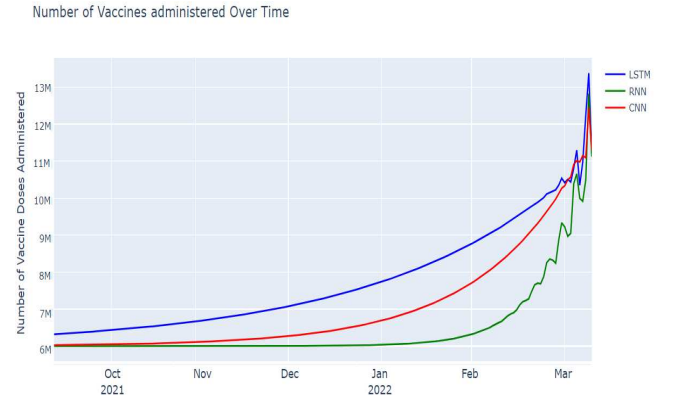

Figure 9. Future predictions of COVID-19 vaccines administered in India

TABLE II. PARAMETER USED IN CNN MODEL

Parameters	Values
Number of convolution layers	2
Number of max-pooling layers	1
Number of fully connected layers	3
Number of nodes in fully connected layers	100, 50, 1
Number of filters in each convolution layer	100, 64
Kernel/filter size in each convolution layer	3
Pool size in max-pooling layer	3
Strides	2
Activation function in convolution layers	Relu
Epochs	100

TABLE III. PARAMETER USED IN RNN AND LSTM MODELS

Parameters	Values
Number of layers	3
Number of nodes in each layer	128, 100, 50
Number of fully connected layers	2
Number of nodes in fully connected layers	50, 1
Activation function	Relu
Epochs	100

V. CONCLUSION AND FUTURE SCOPE

COVID-19 virus is one of the deadliest viruses that resulted massive sabotage and still irresistible. To reduce the rate of spread, many protocols have been followed, such as sanitizing hands, maintaining social distancing, and many more. Apart from this, there is a need for people to get vaccinated against the virus. This helps to boost the immunity, and the person can fight against the virus. The vaccine cannot eradicate the disease, but it can reduce the spread to a large extent.

This paper analyses the number of people affected by the virus and the mortality caused due to this fatal disease against the vaccine dose administered. The machine learning models such as RNN, LSTM and CNN have been used for the predictive analysis. It has been observed that the mortality rate and the people getting affected by the COVID-19 virus has been decreased by 15.53% and 24.67% as the number of vaccine doses increases.

ACKNOWLEDGMENT

The research work has been financially supported by the Center of Intelligent Systems and Robotics (CISR) – Institute of Technical Education and Research (ITER), affiliated to Siksha ‘O’ Anusandhan (Deemed to be University), Bhubaneswar, India.

Moreover, the authors deeply extend their gratitude towards World Health Organization (W.H.O) and all other (Govt. of India) official websites of COVID-19 for using their valuable dataset for their experimental research work.

REFERENCES

[1] H. Zhou, X. Chen, T. Hu, J. Li, H. Song, Y. Liu, P. Wang, D. Liu, J. Yang, E. C. Holmes, A. C. Hughes, Y. Bi, and W. Shi, “A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein”, Current Biology, Vol. 30, No. 11, pp. 2196-2203, 2020. Doi: 10.1016/j.cub.2020.05.023

[2] https://covid19.who.int/ [W.H.O Reports on COVID-19 as on Oct.8, 2021]

[3] A. K. Emel, K. Aynur, U. Huseyin, N. Karakaya, C. N. Zafer, H. Sadettin, “Comparison of an inactivated Covid19 vaccine-induced antibody response with concurrent natural Covid19 infection”, International Journal of Infectious Diseases, 2021. Doi: 10.1016/j.ijid.2021.09.060

[4] Vaccine data: https://ourworldindata.org/, (accessed on Sept.30 2021)

[5] R. Ross, “Estimates of SARS Death Rates Revised Upward, Center for Infectious Disease Research and Policy,” Pub: May 07, 2003. Available at: https://www.cidrap.umn.edu/news-perspective/2003/05/estimates-sars-death-rates-revised-upward. (Accessed 15 Sept 2021).

978-1-6654-8035-2/22/$31.00 ©2022 IEEE

[6] C. I. Van der Made, A. Simons and J. S. Hoeijmakers, "Presence of Genetic Variants Among Young Men with Severe COVID-19", JAMA. Vol. 324, No.7, pp. 663-673, 2020. doi:10.1001/jama.2020.13719

[7] M. Tolba, M. A. Omirah, A. Hussein and H. Saeed H, "Assessment and Characterization of Post-COVID-19 manifestations", International Journal of Clinical Practice, pp. 01-16, 2020. doi:10.1111/ijcp.13746

[8] A. V. Raveendran, R. Jayadevan and S. Sashidharan, "Long COVID: An overview", Diabetes & Metabolic Syndrome: Clinical Research & Reviews, Vol. 15, No. 3, pp. 869-875, 2021. doi:10.1016/j.dsx.2021.04.007

[9] G. M. Raza, Z. S. Butt, S. Latif and A. Wahid, "Sentiment Analysis on COVID Tweets: An Experimental Analysis on the Impact of Count Vectorizer and TF-IDF on Sentiment Predictions using Deep Learning Models", in International Conference on Digital Futures and Transformative Technologies (ICoDT2-2021), pp. 1-6, 2021. doi: 10.1109/ICoDT252288.2021.9441508.

[10] R. Dutta, S. Chowdhury and K. K. Singh, "IoT-based healthcare delivery services to promote transparency and patient satisfaction in a corporate hospital", Machine Learning and the Internet of Medical Things in Healthcare, pp. 135-151, 2021. doi: 10.1016/B978-0-12-821229-5.00001-X

[11] M. A. El-Affendi, K. Alrajhi and A. Hussain, "A Novel Deep Learning-Based Multilevel Parallel Attention Neural (MPAN) Model for Multidomain Arabic Sentiment Analysis", IEEE Access, Vol. 9, pp. 7508-7518, 2021. doi: 10.1109/ACCESS.2021.3049626.

[12] N. Panchal, R. Kamal, C. Cox and R. Garfield, "The implications of COVID-19 for mental health and substance use", Kaiser Family Foundation, 2021.

[13] B. Liang, R. Yin, J. Du, L. Gui, Y. He and R. Xu, "Embedding Refinement Framework for Targeted Aspect-based Sentiment Analysis", IEEE Transactions on Affective Computing, 2021. doi: 10.1109/TAFFC.2021.3071388.

[14] J. Pennington, R. Socher, C. D. Manning, "GloVe: Global Vectors for Word Representation", in Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP), pp. 01-12, 2014. doi: 10.3115/v1/D14-1162

[15] P. Ghosh and R. Dutta, "Statistical machine learning forecasting simulation for discipline prediction and cost estimation of COVID-19 pandemic", Data Science for COVID-19, pp. 147-173, 2021. doi: 10.1016/B978-0-12-824536-1.00019-8

[16] https://censusindia.gov.in/Census_Data_2001/Projected_Population/Projected_Population.pdf

978-1-6654-8035-2/22/$31.00 ©2022 IEEE