IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DATA ANALYSIS AND PREDICTION OF COVID-19 CASES IN INDIA

[1] Surabhi Shinde, [2]Sayali Shetye, [3]Taruna Rawat, [4] Prof. Prashant Nitnaware

[123]Student, [4]Professor

[1 23] Department of Information Technology, [4] Department of Computer Engineering

Pillai College of Engineering, University of Mumbai, New Panvel, India

Abstract: COVID-19 (Coronavirus) has affected the day to day life of everyone. This disease has affected the lives of thousands of people. COVID-19, described as -'CO' for corona, 'VI' for virus, and 'D' for diseases unfortunately, is rapidly increasing which has put the world in a different state. Broadcasting COVID information and analyzing the present data will help in predicting the future conditions as well as bring a control to this disease. Data Analysis plays an important role in analyzing the data. Data analysis can be done using various machine learning models. Machine learning forecasting models like Dynamic linear models, SIR model , LSTM and GRU model , ARIMA, SARIMA, GARCH, TBATS, Prophet, SIS model, etc. In this study, we used the epidemiological LSTM (Long-short-term-memory) model and GRU (Gated Recurrent unit) model for predicting the spread of COVID-19 cases. It is good to mention here that the forecasts done using these models will be as good as data quality which will be available and, hence, the progress and the prediction can also get affected due to the available quality of the data. The objective of this study is to use the LSTM and GRU model to forecast the analysis and the prediction of COVID -19 cases along with comparison of these two models .This website also provides visual information for better understanding of COVID-19 conditions in India.

Index Terms **- COVID-19, Data Analysis, Machine learning model, LSTM Model, GRU model.**

I. INTRODUCTION

On December 31, 2019, a new coronavirus known as COVID 19 was first reported in Wuhan, China. Coronavirus is a contagious virus. The first wave of COVID 19 in India was the native form of SARSCoV2 that affects the human respiratory system. Symptoms of COVID19 may or may not be seen in infected people, resulting in a faster infection rate. The second wave of COVID19 is a double mutant with higher infectivity. So far, the COVID-19 vaccine has not been given to the entire Indian population. Until then, only precautionary measures are safety measures. Despite ongoing efforts, the virus has spread to most of India's states and Union Territory, and the World Health Organization (WHO) has announced COVID 19 as a pandemic. Most countries in the world are working together to control this situation. Data scientists and data mining researchers can play an important role in this situation. You can integrate relevant data and technology to better understand viruses and their characteristics. This will help you make the right decisions and make concrete action plans.

II. LITERATURE SURVEY

In this chapter, we will review the relevant techniques in the literature. Describes the different techniques used in the work. Identify current literature on related domain issues. It identifies the techniques developed and demonstrates the various advantages and limitations of these methods, which are widely used in the literature. A summary of the literature at the end of this chapter. The Literature Review is an objective and critical summary of published research literature related to a research topic. Its purpose is to become familiar with current thinking and research on a particular topic and may guarantee future research in previously overlooked or unexplored territory.

2.1. Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India: This article uses a deep learning-based model to predict the number of new coronaviruses. (COVID19) Positive cases reported in 32 Indian states and Union Territory. Recurrent Neural Networks (RNN) -based long short-term memory (LSTM) variants such as Deep LSTM, Convolutional LSTM, Bi, etc.Directional LSTMs are applied to Indian datasets to predict the number of positive cases. The minimum error LSTM model has been selected to predict daily and weekly cases.

2.2.Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM: In this article, the COVID19 dataset is ARIMA, SVR with polynomials, and RBF kernel, LSTM, GRU and BiLSTM for future predictions of confirmed cases, mortality and recovery. For 10 countries around the world, comparison between different models using MAE, RMSE, and r2_score power measurements is shown. The ARIMA and SVR models do not follow trends in these features, resulting in higher prediction errors.

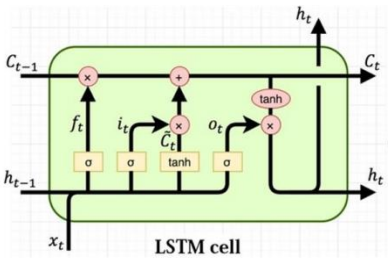
2.3.Time series predicting of COVID-19 based on deep learning: In this survey, the dataset is provided by European Disease Control Center Outbreak of COVID 19 in Malaysia, Morocco and Saudi Arabia. It also proposes a DL approach including recurrent neural network (RNN) and long short-term memory (LSTM) network for predicting the probable numbers of COVID-19 cases. It predicts the number of confirmed COVID-19 cases and deaths for a subsequent seven days.and the forecasts are presented using data available by December 3, 2020.

2.4.County-Level COVID-19 Case Predictions using Deep Learning: This project aims to train deep neural networks to accurately predict numbers. Percentage of 19 COVID cases in US states based on demographic and historical case data and national policy data. Trained feed forward 5-layer neural network and an LSTM network with data from 2,257 US counties. Here, goal is accurate time series prediction of cases and analysis of case characteristics Maximum effect.

III. PROPOSED WORK

Here we have used two models for predicting the COVID-19 cases. Comparison between the results of these two models have been shown. First let us understand what LSTM and GRU models are.

3.1 LSTM: A long short-term memory (LSTM) network used in the field of deep learning. This is a variety of recurrent neural networks (RNNs) capable of learning long-term dependencies, especially for sequence prediction problems.

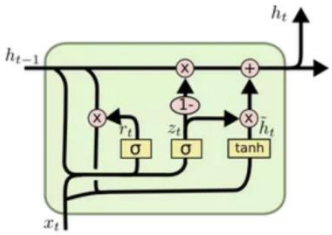


$$i_t = \sigma\left(x_t U^i + h_{t-1} W^i\right)$$
$$f_t = \sigma\left(x_t U^f + h_{t-1} W^f\right)$$
$$o_t = \sigma\left(x_t U^o + h_{t-1} W^o\right)$$
$$\tilde{C}_t = \tanh\left(x_t U^g + h_{t-1} W^g\right)$$
$$C_t = \sigma\left(f_t * C_{t-1} + i_t * \tilde{C}_t\right)$$
$$h_t = \tanh(C_t) * o_t$$

Fig 1 Mathematical equations for LSTM cell

$$i_t \rightarrow represents\ input\ gate.$$
$$f_t \rightarrow represents\ forget\ gate.$$
$$o_t \rightarrow represents\ output\ gate.$$
$$\sigma \rightarrow represents\ sigmoid\ function.$$
$$w_x \rightarrow weight\ for\ the\ respective\ gate(x)\ neurons.$$
$$h_{t-1} \rightarrow output\ of\ the\ previous\ lstm\ block(at\ timestamp\ t-1).$$
$$x_t \rightarrow input\ at\ current\ timestamp.$$
$$b_x \rightarrow biases\ for\ the\ respective\ gates(x).$$

3.2 GRU: A gated recurrent unit (GRU) is part of a particular recurrent neural network (RNN) model. It uses connections between a set of nodes to perform machine learning tasks related to memory and clustering. To solve the vanishing gradient problem, adjust the input weights of the neural network.



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$
$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Fig 2 Mathematical equations for GRU cell

IV. SYSTEM ARCHITECTURE

The system architecture is given in figure 3

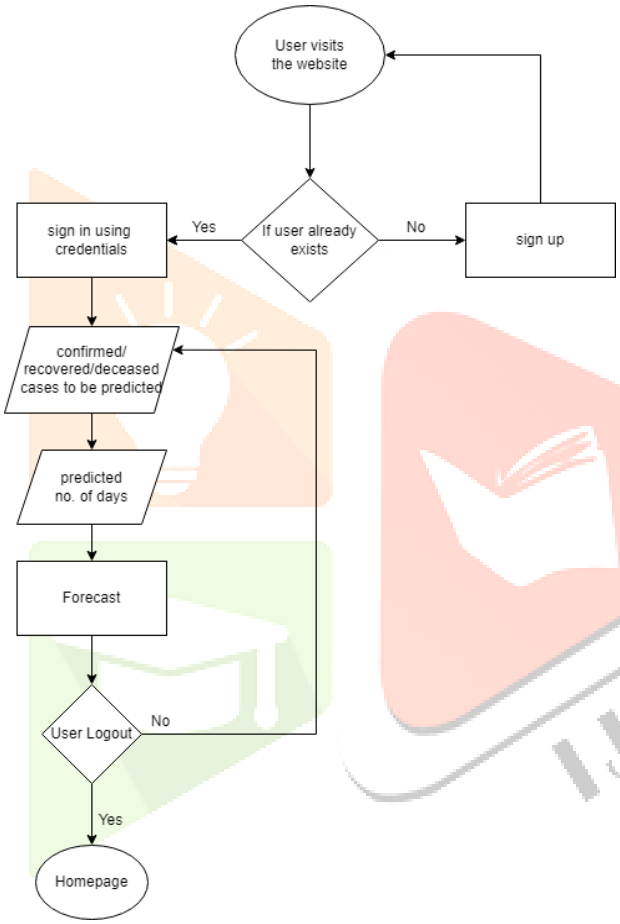


Fig 3 System Architecture

System Architecture for Prediction :

When a user tries to login with the registered credential,the database validates the user and signs in. If the user is new to the website, it asks for registration. After login, the forecasting page is displayed. Users can select out of confirmed,recovered and deceased cases. The user has to enter the number of days for which the prediction has to be made. For example if the user wants to forecast the confirmed cases for the next five days, the number of predicted cases for the next five days will come as an output.

V. RESULTS AND PERFORMANCE EVALUATION

5.1 Analysis

Analysis is done by comparing the loss and mean absolute error (MAE) between LSTM model and GRU model which is shown in the table below. Lesser the value of both the metrics, more accurate is the model.

5.2 Comparison of the Models

The next step of the study is to compare these competing models to evaluate which one of these models is more supported by data.

Classifier	Metrices					
	Daily confirmed Cases		Daily recovered Cases		Daily deceased cases	
	MAE	Loss	MAE	Loss	MAE	Loss
LSTM	0.0265	0.0021	0.0261	0.0020	0.0295	0.0031
GRU	0.0341	0.0037	0.0253	0.0018	0.0333	0.0038

Table 1. Comparison of LSTM and GRU model

5.3 Train Model

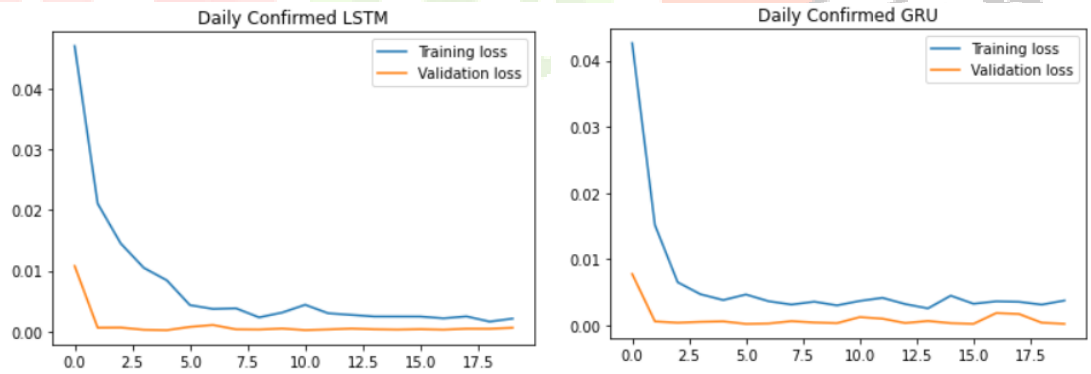


Fig 4. Training LSTM model and GRU model for confirmed cases

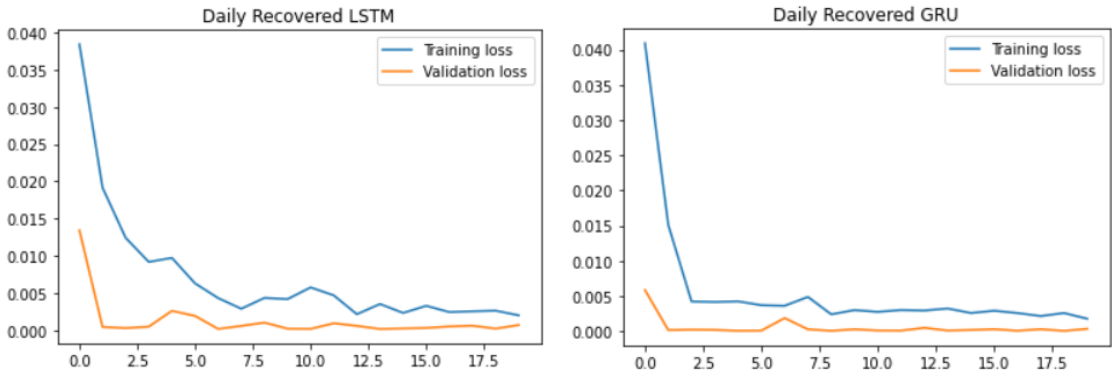


Fig 5. Training LSTM model and GRU model for recovered cases

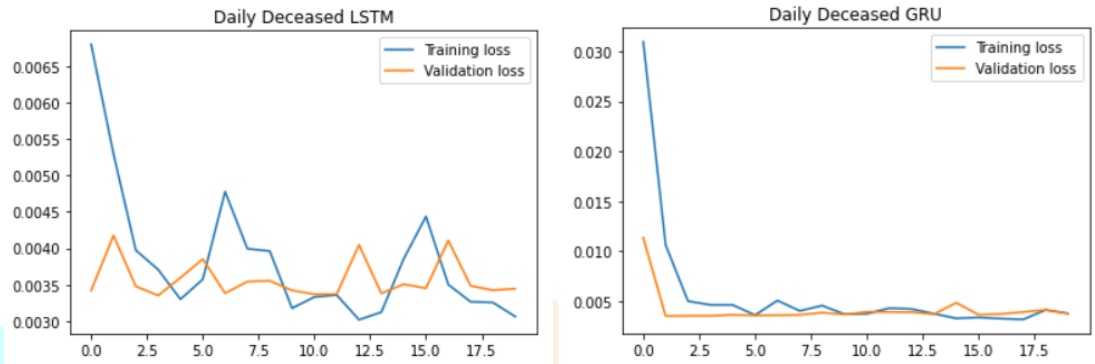


Fig 6. Training LSTM model and GRU model for deceased cases

5.4 Prediction using GRU Algorithm

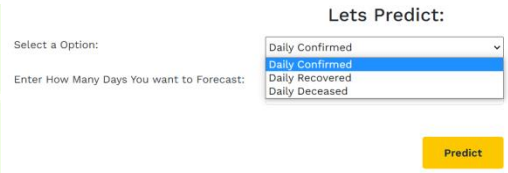


Fig 7. Checkbox for daily confirmed, recovered and deceased cases.

Lets Predict:

Select a Option: Daily Confirmed

Enter How Many Days You want to Forecast: 5

Predict

Predictions:

	Daily Confirmed Forecasted
0	14793
1	14732
2	14773
3	14868
4	14981

Table 2. Daily confirmed cases predicted for the next five days

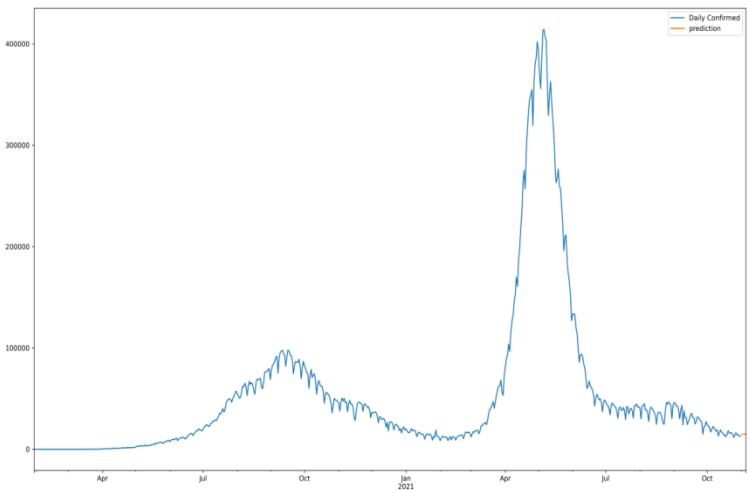
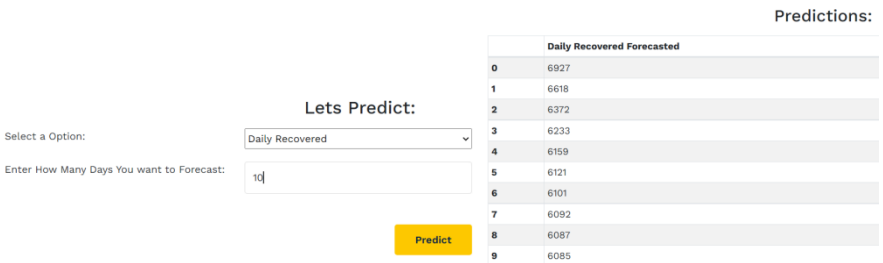


Fig 8. Forecasting daily cases for the next five days



	Daily Recovered Forecasted
0	6927
1	6618
2	6372
3	6233
4	6159
5	6121
6	6101
7	6092
8	6087
9	6085

Table 3. Daily recovered cases predicted for the ten next days

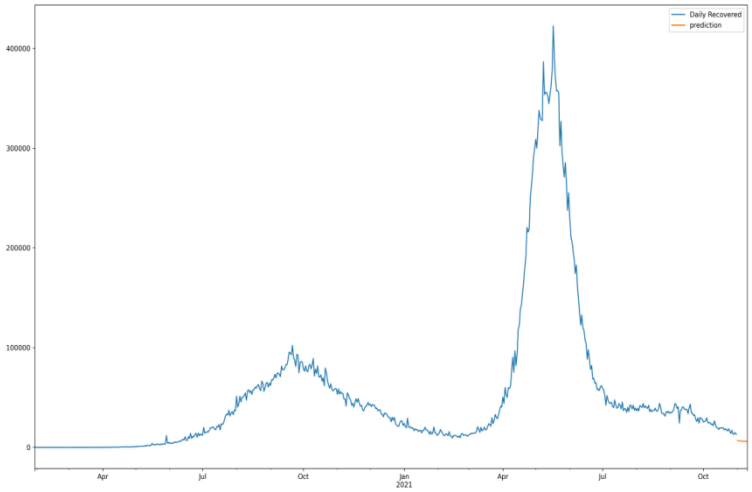


Fig 9. Forecasting daily cases for the next ten days

	Daily Deceased Forecasted
0	536
1	496
2	480
3	475
4	475
5	475
6	476
7	476
8	477
9	477
10	478
11	478
12	478
13	479
14	479

Table 4. Daily deceased cases predicted for the next fifteen days

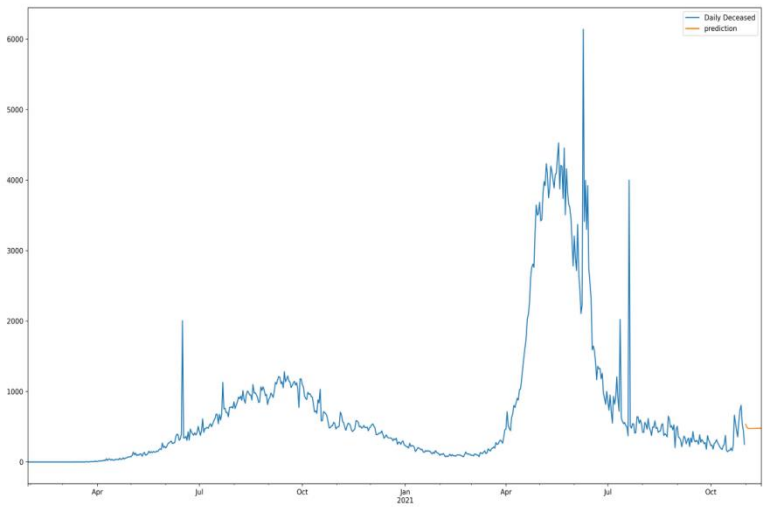


Fig 10. Forecasting daily cases for the next fifteen days

VI. CONCLUSION

From the study, we conclude that numbers of COVID- 19 Confirmed cases are increasing rapidly. The idea behind this work is to make predictions about the numbers of cases in the near future.By collecting knowledge about machine learning, predictive methods used in the past, and previous studies specific to COVID-19, we use machine learning predictive methods to register COVID-19 cases in respected regions. You can build a system to predict trends that will get registered in the respective local body. Analysts can use two ML algorithms, long short-term memory (LSTM) and gated regression unit (GRU), to make comparisons and use better algorithms to make predictions. The model helps to better understand the effect of COVID-19 in India. The strengthening of local governments indirectly contributes to the strengthening of the country and supports the preparation of the medical and administrative sectors.

VII. ACKNOWLEDGEMENT

We would like to thank our Guide Prof. Prashant Nitnaware, who gave us this golden opportunity to do this wonderful project on the topic “Data Analysis and prediction of COVID-19 cases in India”, which also helped us in doing a lot of Research and thereby increasing our knowledge.
We would like to express our gratitude to our H.O.D of Information Technology Dr. Satish Kumar Verma and our Principal Dr. Sandeep M. Joshi for encouraging and allowing us to present this work.

REFERENCES

[1] R.S. Walse, G.D. Kurundkar, P. U. Bhalchandra, "A Review: Design and Development of Novel Techniques for Clustering\ and Classification of Data," International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.1, pp.19-22, 2018

[2] Hemant Kumar Soni, "Machine Learning â€“ A New Paradigm of AI," International Journal of Scientific Research in Network Security and Communication, Vol.7, Issue.3, pp.31-32, 2019

[3]AarogyaSetu(2020) https://play.google.com/store/apps/details?id=nic.goi.aarogya setup &hl=en

[4]Androiddeveloper-Locations(2020)https://developer.android.com/reference/android/location/Location#distanceBetween(double,%2520double,%2520double,%2520double,%2520float[])

[5]https://matlabhelper.com/blog/matlab/COVID-19-sir-modelling-for-andhra-pradesh-india/

[6]Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET by Christophorus Beneditto Aditya Satrioa, *, William Darmawan , Bellatasya Unrica Nadia ,Novita Hanafiah.

[7] Colin Maclas Gonsalves’s ‘Prediction of COVID 19 cases using Machine learning’ Dissertation Report 2020-2021.