



# Predictive Modeling of the Spread of COVID-19: The Case of India

Sriram Sankaran<sup>1</sup>, Vamshi Sunku Mohan<sup>1(✉)</sup>, Mukund Seshadrinath<sup>2</sup>,  
Krushna Chandra Gouda<sup>3</sup>, Himesh Shivappa<sup>3</sup>, and Krishnashree Achuthan<sup>1</sup>

<sup>1</sup> Center for Cybersecurity Systems and Networks, Amrita Vishwa Vidyapeetham,  
Amritapuri, India

{srirams,vamshis}@am.amrita.edu, krishna@amrita.edu

<sup>2</sup> Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham,  
Amritapuri, India

mukund777@am.students.amrita.edu

<sup>3</sup> CSIR Fourth Paradigm Institute, Bangalore, India  
{kcgouda,himesh}@csir4pi.in

**Abstract.** COVID-19 has been the most notorious pandemic affecting the entire world resulting in numerous deaths thus crippling the world economy. While vaccines are in the process of being developed for protection, countries are implementing measures such as social distancing to prevent the spread of the virus. Also, there exists a need for developing mathematical models to predict the rate of spread of COVID-19 and quantify its impact on countries such as India. Towards this goal, we developed a realistic COVID-19 dataset consisting of state-wide distribution of number of cases in India from March-July 2020. Further, we conduct exploratory data analysis on the dataset to understand the states and their corresponding growth rates. This enables us to cluster states with exponential and non-exponential growth rates as well as assess the effectiveness of lockdown imposed to curb the spread of virus. Finally, we develop predictive models using Auto-Regressive Integrated Moving Average (ARIMA) and Long Short-Term Memory Networks (LSTM) on time-series data for top-10 affected states in India to predict the rate of spread and validate their accuracy. Finally, our models can be used to guide the development of mechanisms for optimal resource allocation of healthcare systems and response.

**Keywords:** COVID-19 · ARIMA · LSTM

## 1 Introduction

Promoting the health and well-being of citizens is of paramount importance for ensuring the safety and security of the country. The recent outbreak of the COVID-19 virus has caused a significant impact across the globe by affecting the physical and mental well-being of the people. Symptoms of the virus range

from simple cold and fever to a much more complex respiratory illness [1–4] thus requiring the use of ventilators for artificial respiration. While vaccines are being developed to make people of diverse health backgrounds resilient to the virus, a much more promising second line of defense is to promote the concept of social distancing and night curfew among people to limit the spread of the virus. Such measures can be taken only if the government has a estimate of the rise in the number of cases in the near future based on the recorded number of cases in the past.

The impact of COVID-19 varies for different countries thus requiring the use of techniques depending on the needs and requirements. Statistics show that there are about 26 million people affected by the virus due to the increase in the contact rates and about 863,000 deaths worldwide [5]. The United States occupies the topmost position in the world in terms of the number of infected people [6] followed by Brazil. While India remained at the bottom of the list initially, it soon climbed up due to the increase in the mobility of the citizens and the contact rates. Several countries impose a variety of techniques such as a mandatory 2-week institutional or home quarantine so that the symptoms of the virus wither away by the end of the quarantine.

While the impact of COVID-19 varies for different countries, there exists a need for exploratory data analysis to understand the characteristics of individual states as well as obtain a summary of the datasets. These characteristics are typically grouped for comparative analysis and represented in a pictorial manner through the use of scatterplots, bar plots, and histograms. Further, Exploratory data analysis often involves conducting statistical analysis to formulate a hypothesis that leads to conducting further experiments to validate the behavior. Thus, there exists a need for quantifying the impact of COVID-19 for countries such as India through the use of Exploratory data analysis.

While Exploratory data analysis shows promise, there exists a need for predictive modeling of a number of COVID-19 cases for each of the states in India to predict the rate of spread of the virus. Predictive modeling relies on exploratory data analysis to understand the data, analyze the trends, and obtain meaningful summarizations so that appropriate approaches can be applied. Approaches for predictive modeling range from simple linear regression to advanced techniques such as time series analysis and deep learning. Typically, predictive modeling classifies the data into training and testing sets and fits a model on the training set and evaluates the model on the testing set to determine the accuracy.

In this paper, we develop a predictive approach to modeling the impact of COVID-19 on India. First, we conduct exploratory data analysis to quantify the impact of COVID-19 on individual states in a comparative manner. In particular, we pictorially represent the top 10 states in India that are most affected by COVID-19 and analyze the growth rates. Further, we determine the effectiveness of the lockdown imposed at different time intervals for each of the states by observing the growth rates. Finally, we leverage techniques such as time series analysis and deep learning to predict the rate of spread of the virus using the number of cases for the individual states in India. These models can collectively

be used to guide the development of mechanisms for optimal resource allocation and response.

This paper is structured as follows. Section 2 investigates existing work in this area. Section 3 describes exploratory data analysis with particular emphasis on the number of cases, growth rates, and the impact of lockdown on growth rates. Section 4 discusses techniques based on time series analysis and deep learning to build predictive models for COVID-19. Section 5 concludes the paper followed by references.

## 2 Related Work

In this section, we review existing literature for COVID-19 that are very relevant to this paper. In particular, we divide our literature review into the pertinent areas of data analysis, mathematical modeling and predictive modeling.

### 2.1 Data Analysis

Statistics pertaining to COVID-19 for India have been curated on a constant basis in [5] in a crowd-sourced manner. In particular, [5] reports a number of cases on a daily as well as a cumulative basis along with the analysis of trends and patterns for each of the states in India. One of the extensively used apps for tracing COVID-19 hotspots is AarogyaSetu [7] launched by the Government of India. AarogyaSetu [7] is a location-aware app that relies on Bluetooth to identify hotspots in COVID-19 infected regions. While AarogyaSetu is primarily used in India, countries have mandated the usage of different apps for identifying COVID-19 hotspots through contact tracing. While AarogyaSetu shows promise, it has significant concerns in terms of privacy [8]. Ahmed et al. [8] has presented a survey of the list of contact tracing apps available in the market along with corresponding security and privacy concerns.

### 2.2 Mathematical Modeling

While data analysis is of significant importance, there exists a need for developing mathematical models and projections [9] to study the impact of virus spread. Few works [10–13] have developed models for analyzing the spread of COVID-19 in different countries. Chen et al. [10] have developed epidemiological models using Susceptible-Infected-Recovered (SIR) to study the impact of virus spread in China and analyzed the impact of lockdown under different conditions. Giamberardino et al. [11] formulated the problem of optimal resource allocation of healthcare resources during a pandemic as integer linear programming and developed heuristics. Further, the authors utilize the SIR model as the input for the optimal resource allocation problem. Li et al. [12] claim that popular social networking and streaming websites have been publishing fake information pertaining to COVID-19. Jin et al. [13] modeled the impact of news and rumors on Twitter using epidemiological models. However, their study does not consider the impact of fake information propagated during COVID-19.

### 2.3 Predictive Modeling

Few works [14–17] have developed techniques using machine learning to combat the spread of the virus. Rustam et al. [14] developed a suite of supervised machine learning techniques for forecasting the impact of COVID-19. Jamshidi et al. [15] leveraged deep learning techniques for the diagnosis and treatment of COVID-19. Hussain et al. [16] developed AI based techniques for COVID-19. Shoeibi et al. [17] presented a review of deep learning techniques for automated detection and forecasting of COVID-19.

In contrast to existing approaches, we conduct exploratory data analysis for studying the impact of COVID-19 for India. Further, we develop predictive models using Time Series Analysis and Deep learning to predict the rate of spread of COVID-19.

## 3 Exploratory Data Analysis

In this section, we conduct exploratory data analysis of COVID-19 dataset containing the number of states in India. We start with describing the dataset along with analyzing the number of cases and corresponding growth rates for the individual states in India. Finally, we examine the effectiveness of the lockdown on the growth states for the individual states.

### 3.1 Data Description

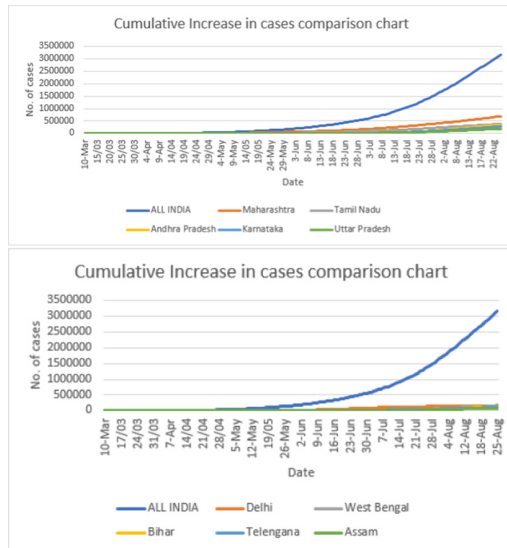
In this work the number of positive cases of COVID-19 reported is used for the period March-Aug 2020. The data is being collected from sources like Ministry of Health and Family Welfare (MoHFW), Government of India (<https://mohfw.gov.in>) [18] and Worldometers (<https://www.worldometers.info>) [19] at daily scale.

### 3.2 Increase in Cases

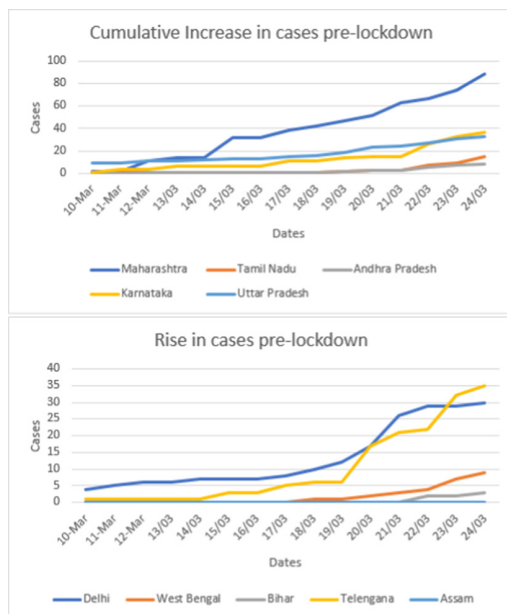
The day to day cumulative increase in the positive cases are presented in Fig. 1 for all India and 10 states worst affected i.e. Maharashtra, Tamilnadu, Andhra Pradesh, Karnataka, Uttar Pradesh, Delhi, West Bengal, Bihar, Telangana and Assam. The analysis clearly shows the exponential growth of the COVID-19 spread in India and the states.

Figure 2 represents the cumulative increase in number of cases for top 10 states which have been ranked based on maximum number of cases till date for the pre-lockdown phase. As we observe from the graphs, Maharashtra, Delhi, and Telangana has a sharp increase in the number of cases. This could be attributed to the high population density and lack of social distancing from its citizens.

Similarly Fig. 3 represents the number of newly infected patients recorded per day for the given time period. When we analyze the increase in cases per day, we can see that Maharashtra the greatest number of cases per day recorded in the

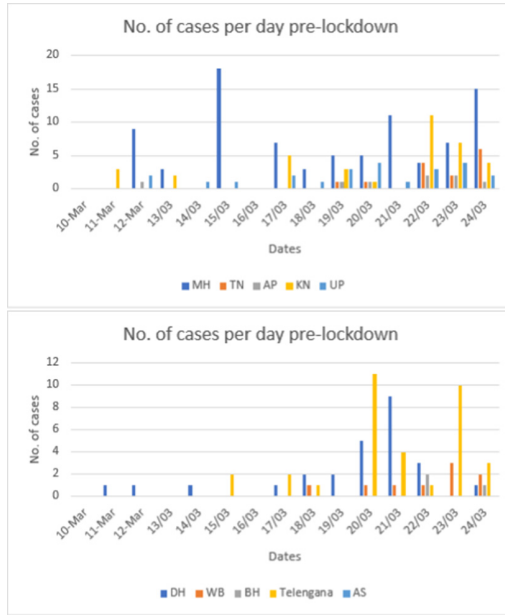


**Fig. 1.** The day to day increase of COVID-19 positive cases reported in India and 10 worst affected states



**Fig. 2.** The day to day COVID-19 positive cases reported during the pre-lockdown period in India and 10 worst affected states

country before lockdown was imposed by the government. Initially, Delhi and Telangana had a smaller number of cases per day then saw a drastic increase. Even Karnataka observed a steady rise in cases. Hence, lockdown had to be imposed to stem the rise in cases. Once the lockdown was lifted from 1st June, there has been a steady increase in cases. Maharashtra, Tamil Nadu, and Delhi continue to dominate the cases in India. West Bengal and Telangana have witnessed a major increase in the number of cases post lockdown (Fig. 4 and 5).



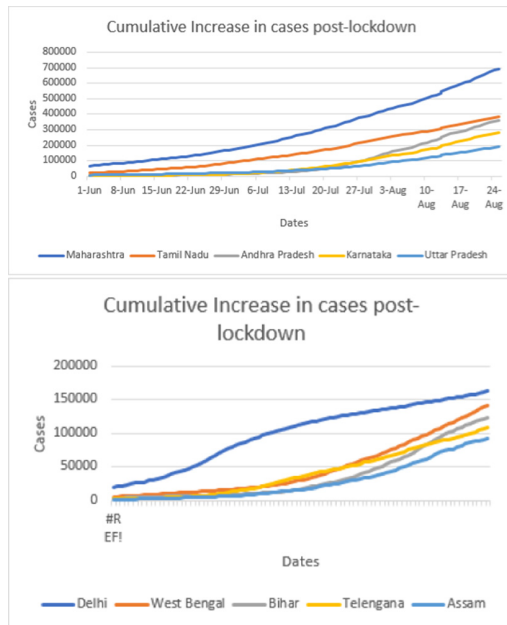
**Fig. 3.** Daily number of positive COVID-19 cases reported in India and 10 states during pre-lockdown period

### 3.3 Growth Rates

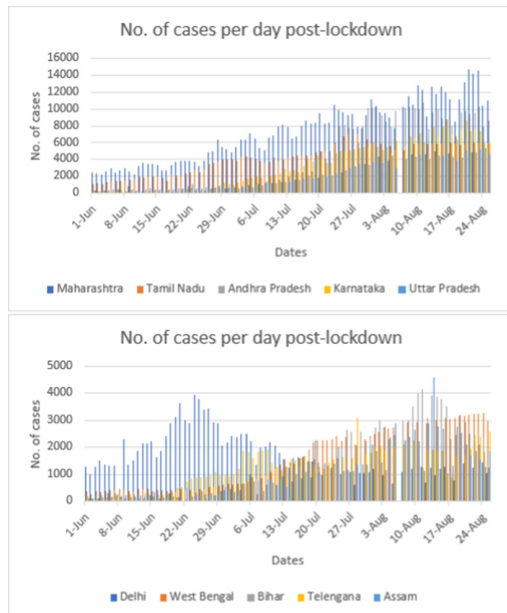
To get a measure of the severity of the spread, the growth rates of the COVID-19 cases are analyzed and presented in Fig. 6 and it is observed that Maharashtra and Telangana have a sharp increase in growth rates. These sharp increases represent a spike in cases. Similarly, this can be observed for West Bengal and Tamil Nadu towards the later stages and nearing towards the lockdown.

When we consider post lockdown (Fig. 7), we can see that Karnataka and Andhra Pradesh have prominent increases in growth rates, and the same carries for Assam and Telangana. Surprisingly, despite having the largest number of cases, Maharashtra, Delhi and Tamil Nadu maintain a consistent growth rate.

A deeper analysis from the above plots reveals that Maharashtra has a large spike in growth rate pre-lockdown whereas it maintains a rather consistent growth rate post lockdown. The reason is that, the state saw an increase



**Fig. 4.** The day to day increase of COVID-19 positive cases reported in India and 10 worst affected states during post-lock down period (i.e. June 01–24 August 2020)



**Fig. 5.** Daily number of positive COVID-19 cases reported in India and 10 states during post-lockdown period (i.e. June 01–24 August 2020)

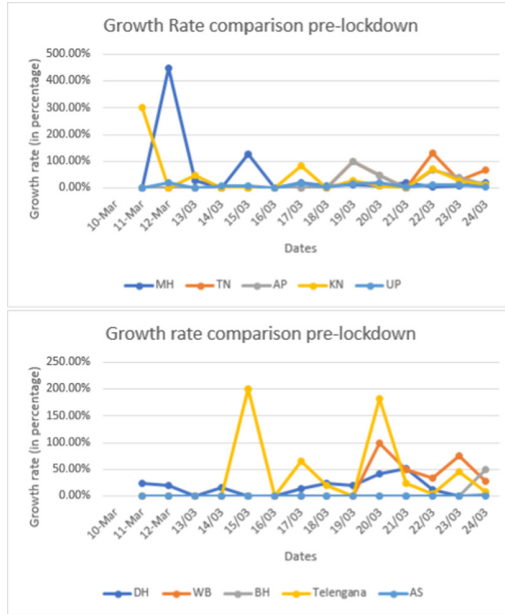


Fig. 6. Comparison of COVID-19 growth rate during pre-lockdown stage in India and states

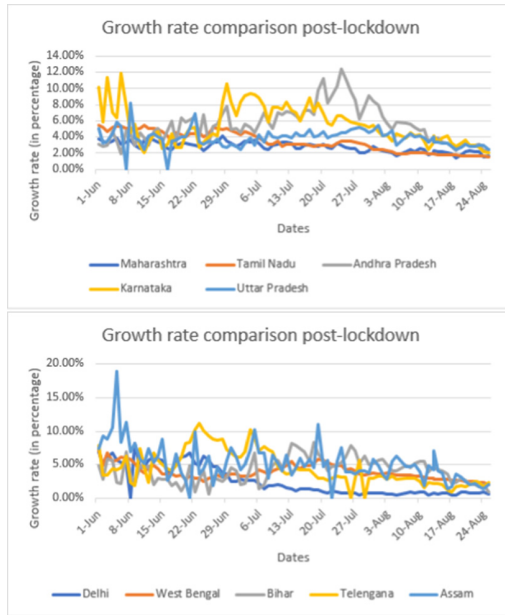


Fig. 7. Comparison of COVID-19 growth rate during post-lockdown stage in India and states

in newly infected people within a range of 0 to 20 in the pre-lockdown phase. Now for initial days of the pandemic, this was a huge increase as the number of cases was low. But when considering post lockdown phase, even though people are getting infected in the thousands, the number of cases is already high and hence the growth rate remains constant.

But this behavior is contrasting for the state of Assam. The state has a small growth rate during the pre-lockdown phase and hence the lesser number of cases. But when considering post lockdown phase, as the cases increase by the thousands, the graph shows major spikes in growth rate.

### 3.4 Impact of Lockdown

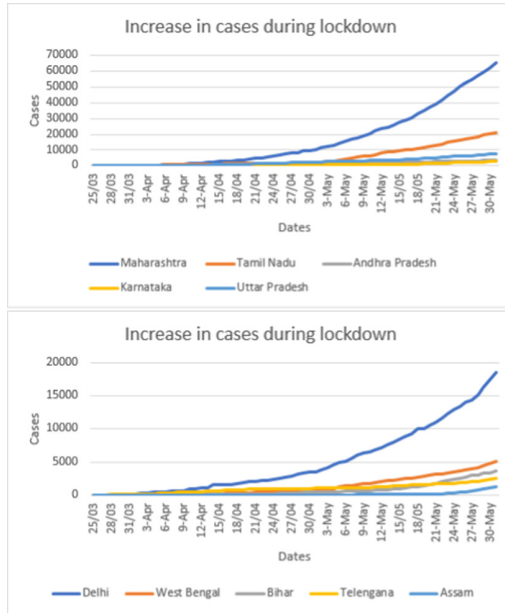
To assess the impact of lockdown in India the cases and growth rate during March-May 2020 is analysed and discussed. Figure 8 and 9 respectively represent the cumulative curve and the day to day reported cases in India and the states considered. As we can see from Fig. 8, Maharashtra and Delhi have a drastic increase in cases where as the remaining states have a rather consistent increase. This could be attributed to the amount of testing done as the increase in cases linearly dependent on the amount of testing done.

The above two graphs in Fig. 9 supplements to the above analysis during lockdown. Also, it was during this time that a lot of migrant workers were returning back to their hometowns and this could have contributed to the spread.

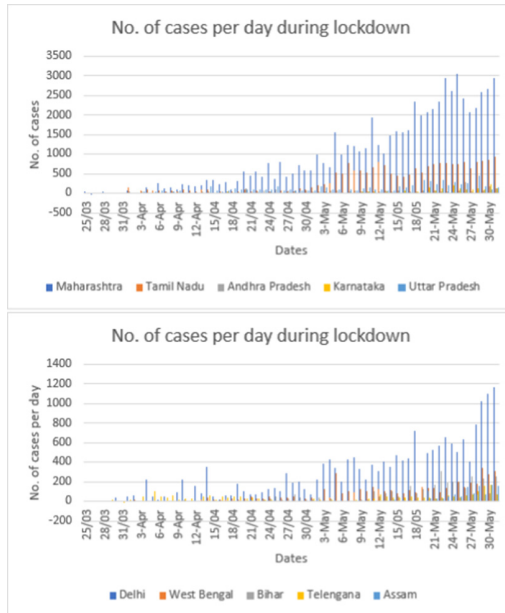
The growth rate during lockdown is presented in Fig. 10 and as per the analysis, Assam and Tamil Nadu saw a spike in cases where as Maharashtra and Delhi see a consistent growth rate despite having a large increase in cases. This could be attributed to the fact that Maharashtra and Delhi had a large number of cases going into lockdown whereas the remaining states had a rather smaller number. It can be attested that a smaller percentage increase in a large number is greater than a large percentage increase in a smaller number and hence the increase. For example, in the case of Maharashtra, on 12th March 2020, the cumulative number of cases rose from 2 to 11. That is a 450% increase in growth rate. Whereas, on 25th August 2020, the cumulative number of cases in Maharashtra rose from 6,82,383 to 6,93,398. 11,015 people were newly infected, but the growth rate for that day was only 1.61%.

## 4 Predictive Modelling

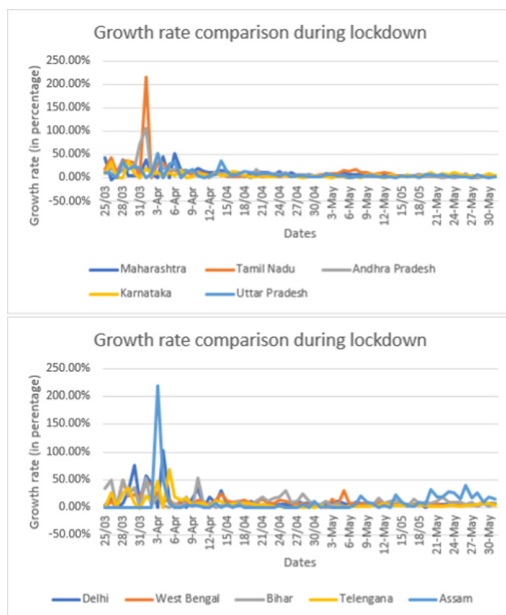
In this section, we develop and evaluate predictive models for forecasting the spread of virus for individual states in India using machine learning techniques. In particular, we have adopted time series analysis and deep learning approaches to predict the number of cases for a particular month based on the collective data of previous months.



**Fig. 8.** The day to day increase of COVID-19 positive cases reported in India and 10 worst affected states during lockdown phase



**Fig. 9.** Daily number of positive COVID-19 cases reported in India and 10 states during lock down phase



**Fig. 10.** Daily number of positive COVID-19 cases reported in India and 10 states during lock down phase

### 4.1 Time Series Analysis

We use ARIMA model to predict the rate of spread of COVID-19. There are two methods of prediction in ARIMA. The first method requires to differentiate the time series, apply auto-correlation function (ACF) and partial auto-correlation function (PACF) to determine d, q and p which represents the order of differencing, auto-regressive parameters and moving average parameter respectively. From the values obtained, ARIMA model is predicted to be ARIMA(p, d, q). One disadvantage in using this method is, the human error involved in deciding p and q from the graphs obtained by applying ACF and PACF. The predicted values differ significantly even with a small deviation in p and q. The second method uses an in-built function auto.arima where the model decides the perfect fit. Therefore, in this paper, auto.arima has been used. ARIMA works based on ‘Backward-shift operator (B)’, which back shifts the data by one time period. When applied to a time series Y and a period t, we get,

$$BY_t = BY_{t-1} \tag{1}$$

Multiplication by another B shifts the time period more by another time period,

$$B^2Y_t = B(BY_t) = BY_{t-1} = Y_{t-2} \quad (2)$$

For any time period n,

$$B^nY_t = Y_{t-n} \quad (3)$$

If y is the first difference of Y, then for time t,

$$y_t = Y_t - Y_{t-1} = Y_t - BY_t = B(1 - Y_t) \quad (4)$$

The differenced series  $y_t$  is obtained from the original time-series Y by multiplying by a factor of  $(1 - B)$ . Let z be defined as the first difference of y, i.e., z is the second difference of Y.

$$z_t = y_t - y_{t-1} = (1 - B)y_t = (1 - B)((1 - B)Y_t) = (1 - B)^2Y_t \quad (5)$$

Let us consider the equation for ARIMA (1, 1, 1) model.

$$y_t = Y_t - Y_{t-1} \quad (6)$$

$$y_t = \phi_1 y_{t-1} + e_t - \phi_1 e_{t-1} \quad (7)$$

where  $e_t$  is the random noise at time t. Using the backshift operator B,

$$y_t = (1 - B)Y_t \quad (8)$$

$$y_t = \phi_1 B y_t + e_t - \theta_1 B e_t \quad (9)$$

$$(1 - \phi_1 B)y_t = (1 - \theta_1 B)e_t \quad (10)$$

Writing the equation in terms of Y,

$$(1 - \phi_1 B)(1 - B)Y_t = (1 - \theta_1 B)e_t \quad (11)$$

## 4.2 Deep Learning

We use LSTM algorithm to predict results. The data was modelled using Simple, Stacked and Convolutional LSTM. Out of these models, Stacked LSTM predicted the values with the least error rate, hence, their results have been included in the paper.

The mathematical equations for various gates and cells in an LSTM model are defined below,

$$\begin{aligned} i^t &= \sigma(W^i x^t + U^i h^{t-1}) && \text{- Input Gate} \\ f^t &= \sigma(W^f x^t + U^f h^{t-1}) && \text{- Forgot Gate} \\ o^t &= \sigma(W^o x^t + U^o h^{t-1}) && \text{- Output/Exposure Gate} \\ \hat{c}^t &= \tanh(W^c x^t + U^c h^{t-1}) && \text{- New memory cell} \\ c^t &= f^t \hat{c}^{t-1} + i^t \hat{c}^t && \text{- Final memory cell} \\ h^t &= o^t \tanh(c^t) \end{aligned}$$

1. New memory generation - Input  $x^t$  and previous hidden state  $h^{t-1}$  generate a new memory  $c^t$ .
2. Input Gate - Uses input word and past hidden states to determine whether or not the input is worth preserving. It is used to gate the new memory. It produces  $i^t$  as an indicator for input gate.
3. Forget Gate - Similar to input gate except that, it does not make a determination of input word. It uses the input word and past hidden state to produce the forget gate  $f^t$ .
4. Final memory generation - This step takes the advice of forget gate and forgets the past memory  $c^{t-1}$ . Similarly, it takes advice from the input gate  $i^t$  and gates the new memory. The sum of these two results produces the final memory  $c^t$ .
5. Output/Exposure Gate - This gate separates the final memory from the hidden state. Hidden states are used in every gate of LSTM and makes an assessment of the parts of the memory  $c^t$  that is to be exposed or present in the hidden state  $h^t$ .  $o^t$  is used to gate the point-wise tanh of the memory.

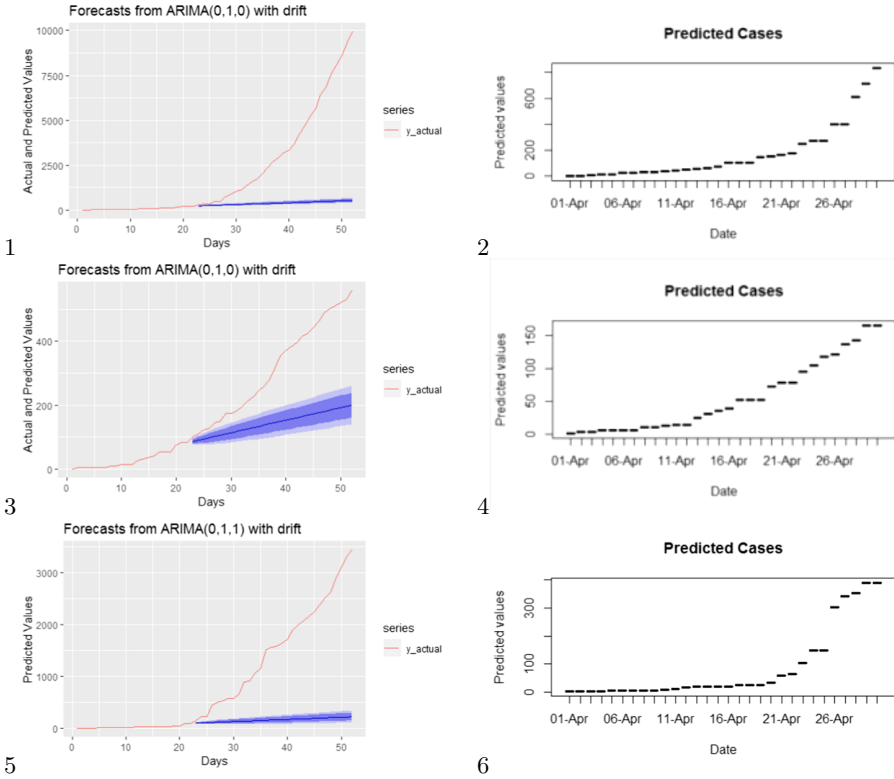
### 4.3 Evaluation of Predictive Models

The models have been evaluated and assessed for the top 10 COVID affected states, they are, Maharashtra, Tamil Nadu, Delhi, Andhra Pradesh, Karnataka, Uttar Pradesh, Gujarat, West Bengal, Telangana and Bihar. The number of cases for a particular month are predicted based on the collective data of all the previous months. In this paper, graphs of actual and predicted values and performance factors are presented for Maharashtra, Delhi and Karnataka for the months of April, August and November as shown in Figs. 11, 12, 13 and Table 1 respectively. Number of cases for the post-COVID period, i.e., the month of February, 2021 have been predicted using ARIMA, with the results recorded in Table 2. In the performance graphs, the red line indicates the actual number of cases. The blue portion represents the predicted values.

### 4.4 Model Validation

ARIMA and Stacked LSTM models have been used to predict the number of cases. It can be observed that the results obtained using Stacked LSTM have a wide deviation compared to the actual values, which are clearly visible in the loss graphs and performance factors. The results derived from ARIMA have high accuracy values, giving a reliable prediction.

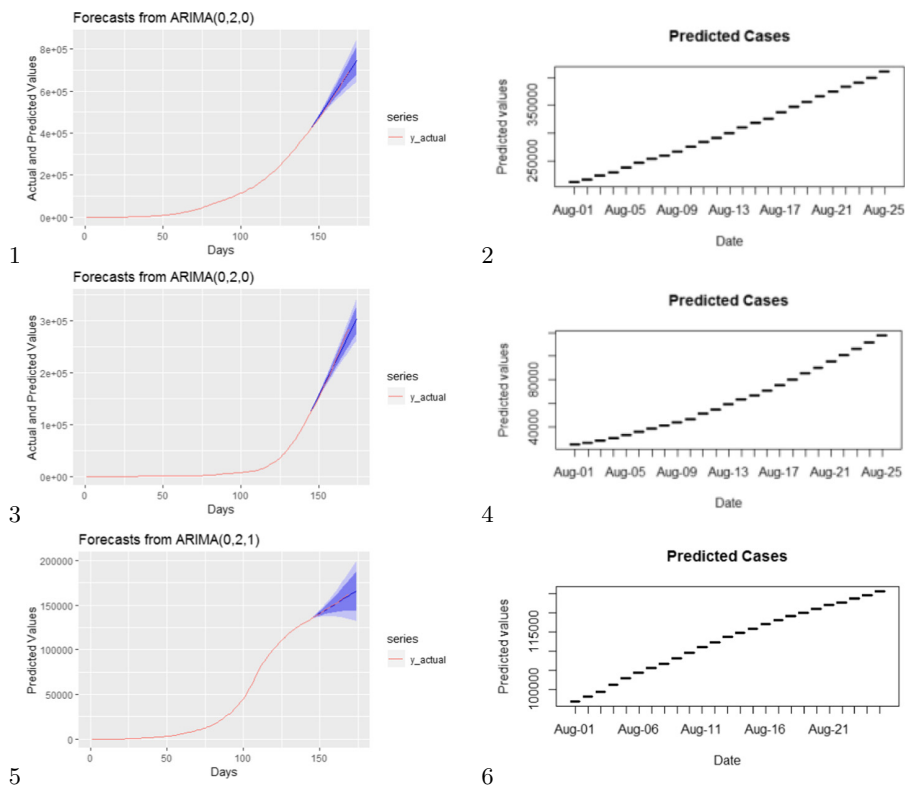
The results obtained by training the model for March and predicting for April shows that the actual cases deviate exponentially compared to the predicted cases, hence, not effective.



**Fig. 11.** Plots of actual and predicted values - Trained for March and predicted for April, (1) Maharashtra - ARIMA, (2) Maharashtra - LSTM, (3) Karnataka - ARIMA, (4) Karnataka - LSTM, (5) Delhi - ARIMA, (6) Delhi - LSTM (Color figure online)

While training for March and April and predicting for May, the results obtained are slightly effective for the states of Andhra Pradesh, Telangana, Gujarat and Uttar Pradesh. However, the predicted values deviate from the actual values for the remaining states.

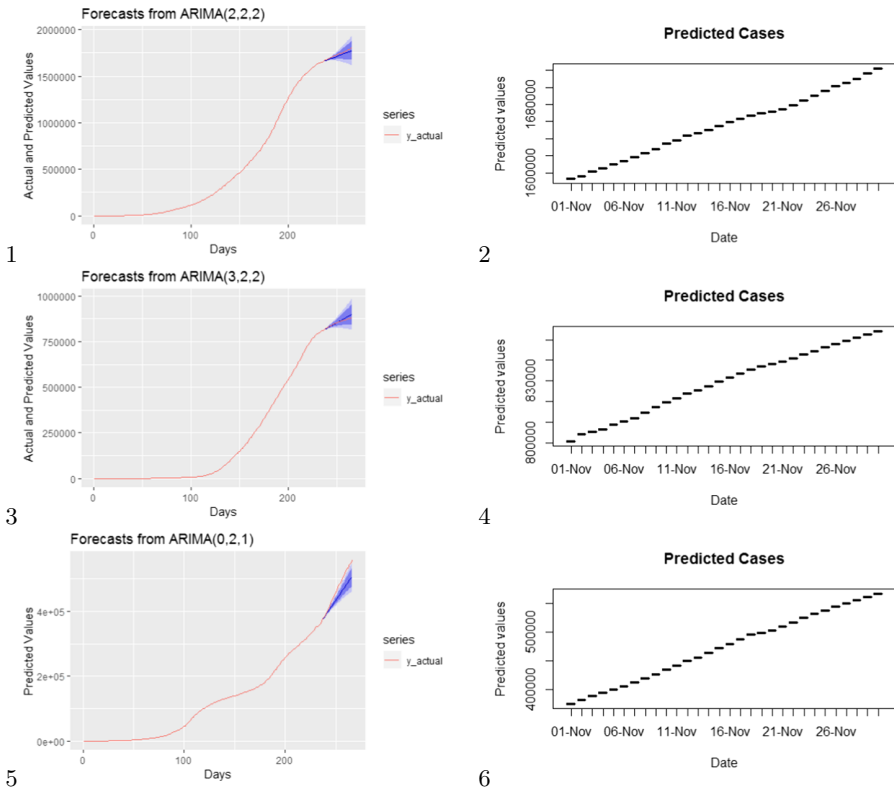
Training for March, April and May and predicting cases for June is applicable for Bihar, West Bengal, Gujarat, Maharashtra. The predicted values lie in the range of actual values. In the case of other states, the deviation between actual and predicted values has largely decreased and the predicted values are almost close to actual values.



**Fig. 12.** Plots of actual and predicted values - Trained for March, April, May, June and July and predicted for August, (1) Maharashtra - ARIMA, (2) Maharashtra - LSTM, (3) Karnataka - ARIMA, (4) Karnataka - LSTM, (5) Delhi - ARIMA, (6) Delhi - LSTM (Color figure online)

Training for March, April, May and June and predicting the cases in July is effective for Maharashtra, Tamil Nadu, Delhi, Karnataka, Andhra Pradesh, Gujarat and Telangana. Other states experience an exponential rise in the number of cases, due to which the model is not accurate.

Training for March, April, May, June and July and predicting for August works well for all the states. From the graphs, it is observable that the number of cases in Delhi is beginning to flatten.



**Fig. 13.** Plots of actual and predicted values - Trained for March, April, May, June, July, August, September and October and predicted for November, (1) Maharashtra - ARIMA, (2) Maharashtra - LSTM, (3) Karnataka - ARIMA, (4) Karnataka - LSTM, (5) Delhi - ARIMA, (6) Delhi - LSTM (Color figure online)

The performance values for the number of cases predicted from September onward shows that the model works well for all the states. It is observable that the curve starts to flatten. This forecast is clearly visible in the cases of Uttar Pradesh, Gujarat and Bihar based on the predictions for February 2021.

From the performance values obtained, it is observable that, with the increase in the cumulative training months data, the model predicts with lower error rate and higher accuracy.

**Table 1.** Predicted value errors when compared to the actual values

Train months	Test month	Method	Performance (%)			
			Parameter	Maharashtra	Karnataka	Delhi
March	April	ARIMA	ME	3193.98	169.21	1392.10
			MAPE	77.02	47.24	83.35
			Accuracy	10.48	45.89	10.34
		LSTM	ME	3418.79	256.73	-1466.82
			MAPE	96.88	86.25	96.44
			Accuracy	3.87	17.17	5.61
March, April, May, June, July	August	ARIMA	ME	-5557.64	1935.76	440.04
			MAPE	1.20	2.17	0.36
			Accuracy	90.59	87.86	97.30
		LSTM	ME	245279.00	137538.50	-30675.10
			MAPE	44.93	70.26	20.77
			Accuracy	64.56	72.09	74.57
March, April, May, June, July, August, September, October	November	ARIMA	ME	16287.55	-4292.50	26336.63
			MAPE	0.93	-0.49	5.29
			Accuracy	99.06	99.50	94.36
		LSTM	ME	108022.00	613790.40	324995.40
			MAPE	195.01	315.22	213.57
			Accuracy	86.99	83.93	85.19

**Table 2.** COVID predicted value for February 2021

Date	Maharashtra	Tamil Nadu	Delhi	Andhra Pradesh	Karnataka	Uttar Pradesh	Gujarat	West Bengal	Telangana	Bihar
01/02	2032149	847914	646027	892348	942941	33500	271404	556984	300015	266611
02/02	2035382	848878	646712	892673	943690	33507	272253	556421	300441	267088
03/02	2038614	849842	647397	892998	944450	33514	273101	555821	300867	267566
04/02	2041847	850805	648082	893323	945218	33522	273949	555187	301294	268043
05/02	2045080	851769	648767	893647	945987	33529	274798	554520	301720	268520
06/02	2048313	852733	649452	893972	946751	33537	275646	553819	302147	268998
07/02	2051546	853697	650137	894297	947504	33544	276494	553086	302573	269475
08/02	2054779	854660	650822	894622	948252	33551	277343	552321	303000	269952
09/02	2058012	855624	651506	894947	949003	33559	278191	551526	303426	270430
10/02	2061245	856588	652191	895272	949762	33566	279039	550699	303852	270907
11/02	2064478	857552	652876	895597	950529	33573	279888	549843	304279	271384
12/02	2067710	858515	653561	895922	951297	33581	280736	548958	304705	271861
13/02	2070943	859479	654246	896247	952060	33588	281584	548045	305132	272339
14/02	2074176	860443	654931	896572	952814	33595	282433	547103	305558	272816
15/02	2077409	861407	655616	896896	953563	33603	283281	546135	305984	273293
16/02	2080642	862370	656301	897221	954315	33610	284129	545139	306411	273771
17/02	2083875	863334	656986	897546	955073	33618	284978	544118	306837	274248
18/02	2087108	864298	657671	897871	955839	33625	285826	543071	307264	274725
19/02	2090341	865262	658355	898196	956607	33632	286674	541999	307690	275201
20/02	2093574	866225	659040	898521	957370	33640	287523	540902	308116	275680
21/02	2096806	867189	659725	898846	958125	33647	288371	539781	308543	276157
22/02	2100039	868153	660410	899171	958875	33654	289219	538637	308969	276635
23/02	2103272	869117	661095	899496	959626	33662	290068	537470	309396	277112
24/02	2106505	870080	661780	899821	960385	33669	290916	536281	309822	277589
25/02	2109738	871044	662465	900145	961150	33676	291764	535069	310249	278067
26/02	2112971	872008	663150	900470	961917	33684	292612	533836	310675	278543
27/02	2116204	872972	663835	900795	962680	33691	293461	532582	311101	279021
28/02	2119437	873936	664520	901120	963435	33698	294309	531307	311528	279498

## 5 Conclusion

In this paper, we analyzed the impact of COVID 19 for the individual states in India in a comparative manner. In particular, we developed a dataset containing the number of COVID-19 cases for India and conducted exploratory data analysis to analyze the growth rates and further examined the effectiveness of the lockdown on the growth rates for the states most affected by the virus. Finally, we developed predictive models to forecast the spread of virus for the states most affected by virus using time series analysis and deep learning and evaluated them in a comparative manner. Our models predict the cases with 97.30% accuracy, hence, can be used to guide the development of mechanisms for optimal resource allocation of healthcare systems and response.

## References

1. Gorbalenya, A., et al.: The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **5** (2020). <https://doi.org/10.1038/s41564-020-0695-z>
2. Is the coronavirus airborne? Experts can't agree. <https://www.nature.com/articles/d41586-020-00974-w>
3. Li, Q., et al.: Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New Engl. J. Med.* **382** (2020). <https://doi.org/10.1056/NEJMoa2001316>
4. Morawska, L., Cao, J.: Airborne transmission of SARS-CoV-2: the world should face the reality. *Environ. Int.* **139** (2020). <https://doi.org/10.1016/j.envint.2020.105730>
5. Coronavirus Outbreak in India. <http://www.covid19india.org/>
6. Killerby, M., et al.: Human coronavirus circulation in the United States 2014–2017. *J. Clin. Virol.* **101** (2018). <https://doi.org/10.1016/j.jcv.2018.01.019>
7. Aarogya Setu: MyGov (2020). <https://www.mygov.in/aarogya-setu-app/>
8. Ahmed, N., et al.: A survey of COVID-19 contact tracing apps. *IEEE Access* **8**, 134577–134601 (2020). <https://doi.org/10.1109/ACCESS.2020.3010226>
9. Kissler, S., Tedijanto, C., Goldstein, E., Grad, Y., Lipsitch, M.: Projecting the transmission dynamics of SARS-CoV-2 through the post-pandemic period (2020). <https://doi.org/10.1101/2020.03.04.20031112>
10. Chen, Y.-C., Lu, P.-E., Chang, C.-S., Liu, T.-H.: A time-dependent SIR model for COVID-19 with undetectable infected persons. *IEEE Trans. Netw. Sci. Eng.* **7**(4), 3279–3294 (2020)
11. Giamberardino, P., Iacoviello, D.: Optimal resource allocation to reduce an epidemic spread and its complication. *Information* **10**, 213 (2019). <https://doi.org/10.3390/info10060213>
12. Li, H., Bailey, A., Huynh, D., Chan, J.: YouTube as a source of information on COVID-19: a pandemic of misinformation? *BMJ Glob. Health* **5** (2020). <https://doi.org/10.1136/bmjgh-2020-002604>
13. Jin, F., Dougherty, E., Saraf, P., Cao, Y., Ramakrishnan, N.: Epidemiological modeling of news and rumors on Twitter. In: *Proceedings of the 7th Workshop on Social Network Mining and Analysis* (2013). <https://doi.org/10.1145/2501025.2501027>

14. Rustam, F., et al.: COVID-19 future forecasting using supervised machine learning models. *IEEE Access* **8**, 101489–101499 (2020). <https://doi.org/10.1109/ACCESS.2020.2997311>
15. Jamshidi, M., et al.: Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment. *IEEE Access* **8**, 109581–109595 (2020). <https://doi.org/10.1109/ACCESS.2020.3001973>
16. Hussain, A.A., Bouachir, O., Al-Turjman, F., Aloqaily, M.: AI techniques for COVID-19. *IEEE Access* **8**, 128776–128795 (2020). <https://doi.org/10.1109/ACCESS.2020.3007939>
17. Shoeibi, A., et al.: Automated detection and forecasting of COVID-19 using deep learning techniques: a review (2020)
18. Ministry of Health and Family Welfare, Government of India. <https://www.mohfw.gov.in/>
19. COVID-19 Coronavirus Pandemic, Worldometer. <https://www.worldometers.info/coronavirus/#countries>