Received: 14 August 2020 | Revised: 17 September 2020 | Accepted: 26 October 2020
DOI: 10.1002/pa.2537

ACADEMIC PAPER

WILEY

Analysis and prediction of COVID-19 trajectory: A machine learning approach

Ritanjali Majhi[1] | Rahul Thangeda[2] | Renu Prasad Sugasi[1] | Niraj Kumar[3]

[1]School of Management, National Institute of Technology Karnataka Surathkal, Mangalore, Karnataka

[2]National Institute of Technology Warangal, India

[3]Essex Business School, University of Essex, UK

Correspondence
Ritanjali Majhi, School of Management, National Institute of Technology Karnataka Surathkal, Mangalore 575025, Karnataka.
Email: ritanjalimajhi@gmail.com

The outbreak of Coronavirus 2019 (COVID-19) has impacted everyday lives globally. The number of positive cases is growing and India is now one of the most affected countries. This paper builds predictive models that can predict the number of positive cases with higher accuracy. Regression-based, Decision tree-based, and Random forest-based models have been built on the data from China and are validated on India's sample. The model is found to be effective and will be able to predict the positive number of cases in the future with minimal error. The developed machine learning model can work in real-time and can effectively predict the number of positive cases. Key measures and suggestions have been put forward considering the effect of lockdown.

1 | INTRODUCTION

In recent times, there has been no health hazard as the COVID-19 pandemic that deeply impacted human health worldwide. Every country is facing turbulent times in terms of ensuring the well-being of its citizens due to the widespread nature of the disease and the unavailability of drugs or vaccine for it. The safety measures considered by almost 162 countries across the globe are to avoid contacts and maintain social distancing. Over the last few months, the disease has impacted severely and there has been a continuous increase in the number of positive cases and deaths. According to WHO, globally 14,509 people have died with a total of 332,930 cases confirmed (WHO, 2020). India, the second-largest population in the world is also not an exception to the spread of the disease. It stands in the top five affected countries in the world. The lockdown measures considered in India in two phases could reduce the spread of the disease to a larger extent. However, as the lockdown was relaxed after two phases, there has been a spread of the disease. As of April 29, 2020 India crossed the 30,000 positive case mark and deaths of 1,000 (The Economic Times, 2020). Though the lockdown was effective to give the government time to prepare for the cure and control of the disease, India is still the third among the countries after China and Iran to see positive cases crossing 30,000 in Asia. One observation that can be made by considering the way the cases have grown during the lockdown is it took 12 days for China and 25 days for Iran to reach to a 30,000 mark from the 1000th case. In India, it has taken 31 days to reach the 30,000th mark from the 1,000th case and 48 days to reach the 1000th death which was 30 and 28 days respectively in China and Iran. The lockdown measures have shown significant results by not having a huge surge in the cases. In a recent report published by the United Nations, the global economy is expected to shrink by up to 1% due to COVID-19. The lockdown has restricted many economic activities and it can only become severe in the April to June quarter. In developing countries like India, COVID-19's effect is suggested to have impacted the economy significantly negatively (Dev & Sengupta, 2020). Hence unlocking the country with proper measures has already started. The resulting surge in infection and subsequent death cases has created worrisome policy dilemma for the government. In this context, forecasting accurately the future trajectory of the pandemic will give the government required tools to deal with it.

Against this backdrop, our paper addresses two important issues. Firstly, to develop a predictive model using the machine learning techniques to effectively predict the number of COVID_19 positive cases in the country. Second, is to propose a strategy in terms of applying lockdown to reduce the number of cases. This would be of immense help to the policymakers to plan effectively in terms of the allocations of medical equipment and infrastructure needed. Also, the paper proposes some important strategies to be followed going forward based on the observations made to sustain the economy and reduce the impact on the economy.

The next section of the paper is organized in the literature published in the area of predicting COVID-19 in different regions. Further sections would discuss the development of machine learning models and the analysis performed considering lockdown and its impact. Finally, the paper makes some of the key suggestions to the policymakers.

2 | LITERATURE REVIEW

There have been several researchers contributing to the areas of forecasting the pandemic. Remuzzi and Remuzzi (2020) have discussed the serious impact that Coronavirus2 (SARS-CoC-2) have caused to China and its further spread to Italy. They built a predictive model that will help understand the patient's raise. And this in turn would help the medical facilities to take decisions. Hall, Gani, Hughes, and Leach (2007) also studied the spread of the H5N1 influenza virus in birds and have forecasted using regression analysis the timing of the prevalence of the pandemic wave. A similar study by Roosa et al. (2020) have used validated phenomenological models to forecast the number of reported cases in Hubei Province. They have used data from the National Health Commission of China to provide forecasts for 5, 10, and 15 days. Their model showed that the containment strategies implemented in China are reducing the transmission and the pandemic has slowed down.

In the recent past, there is research considering China data to predict the number of cases. Hu, Ge, Jin, and Xiong (2020) have proposed an artificial intelligence-based method for real-time forecasting of COVID-19 to estimate the size, lengths, and ending time of COVID-19 across China. The data that was used was from the period of January 11 to February 27, 2020, published by the World Health Organisation (WHO). It was concluded in their work that the spread would reduce by April. Anastassopoulou, Russo, Tsakris, and Siettos (2020) stated that the publicly available data from the epidemiological data of Hubei, China, an estimation on case fatality and case recover recovery ratios were evolved. Their analysis reveals that there would be a drop in the number of cases by January 26, 2020. Fanelli and Piazza (2020) analyzed the temporal dynamics of the corona disease 2019 outbreak in China, Italy, and France between January to March 2020. The model assumes the peak in Italy to be around March 21, 2020, and assumes that an estimate of 2,500 ventilation units should be made ready to reach to the peak requirement.

Further et al. (2020) proposed an objective approach to forecast the continuation of COVID-19 using a live forecasting exercise. The authors in their work have also stressed that no forecast is certain as the future rarely repeats itself in the same way as the past. Especially in cases like COVID-19, it would be difficult to predict accurately. However, the implications of the research can be helpful for decision making. Bastos and Cajueiro (2020) developed a forecasting model using Brazilian data from February 25 to March 30, 2020. The model was built considering social distancing policies as a parameter. The results show that the social distancing policy can flatten the curve. However, if the policy is not lasting long, there is a possibility the curve may again peak up. Elmousalami and Hassanien (2020) used time series models and mathematical formulations, day level forecasting models on COVID 19 cases were forecasted. The authors have also stated that the number of cases will grow exponentially if public gatherings are not stopped and social distancing is not practiced. Dehning et al. (2020) have used Bayesian inference with epidemiological parameters to analyze the time dependence of the effective growth rate of new infections of COVID 19 in Germany. They propose that the model is ready and can be adapted to any country or region in the world.

Also, Benvenuto, Giovanetti, Vassallo, Angeletti, and Ciccozzi (2020) proposed a simple econometric model to predict COVID-19. The model was developed using the Auto-Regressive Integrated Moving Average (ARIMA) model on the Johns Hopkins epidemiological data to predict the trend of prevalence of COVID-19. Singh and Adhikari (2020) proposed a structured SIR model with social contact matrices obtained from surveys and Bayesian imputation to study the cases for COVID-19 in India. They have also studied the impact of social distancing and lockdown in India. They suggest sustained lockdown with periodic relaxation. Botha and Dednam (2020) developed a simple 3-dimensional iterative map model to forecast the Coronavirus disease. From their perspective lockdown measures are effective in postponing the large peak. However, if the relaxation is given there are high chances that the virus would exponentially grow. Peng, Yang, Zhang, Zhuge, and Hong (2020) examined the public data of the National Health Commission of China from January 20 to February 9,, 2020 to make predictions in five different regions. The authors have also made an inference that Beijing and Shanghai will soon stay safe and Wuhan will turn more severe till April.

Bayes and Valdivieso (2020) constructs a predictive Bayesian nonlinear model for the number of COVID-19 deaths in Peru. Liu et al. (2020) presented a timely and novel methodology that combines disease estimates from mechanistic models with digital traces with machine learning mythologies to forecast COVID-19 activity in the Chinese province. The model was able to give forecast 2 days ahead of the current time. Shim, Tariq, Choi, Lee, and Chowell (2020) studied the growth rate of the outbreak of COVID −19 in South Korea by identifying major clusters. They studied the transmission rate and estimated the reproduction number at 1.5 on average. Their estimates supported the implementation of social distancing measures in Korea. Perc, Gorišek Miksić, Slavinec, and Stožer (2020) proposed forecasts obtained with a simple iteration method that needs only the daily values of confirmed cases as input. The results show that the daily growth rates should be kept at least below 5% plateaus are to be seen anytime soon. Perone (2020) in his article forecasted the epidemic trend throughout April using Autoregressive integrated moving average (ARIMA). This model helps in understanding basic trends by suggesting the hypothetic epidemic's inflection point and final size.

Based on the aforesaid literature, it is observed that there are several methods to forecast COVID −19. However, this paper focuses on a simple machine learning model that forecasts the number of positive cases along with also focuses on discussing the impact it shows on the lockdown. The paper proposes to comparing the COVID-19 case with other recent outbreaks and then develop a forecasting model that can help predict the number of positive cases more effectively. There are several forecasting models built using effective statistical methods. This article focuses on using machine learning methods that can handle the nonlinearity better and produce more effective results. Also, the paper suggests some key guidelines appropriate to tackle the current crisis. The analysis will be helpful to policymakers and health authorities to allocate resources rightly in the next few days/weeks.

3 | THE GROWTH OF COVID-19

As a first step, it would be of use to look at the growth and impact of Corona throughout the world. Also, a comparison between the recent outbreaks like Severe Acute Respiratory Syndrome (SARS) and Ebola was taken into consideration to give a fair understanding of how rapid and contagious is the growth of COVID-19. The goal was to understand and answer a few questions like Which countries were affected, how many cases (confirmed/recovered/death) in all, understanding the trend of each outbreak? Going forward the section aims at looking into the details of each of the diseases and understand the patterns of the outbreaks. SARS cases were noticed in the year 2003 and the pattern of cases and deaths are shown in Figure 1.

Similarly, Ebola was another outbreak that happened during 2015, and following Figure 2 explains the patterns of deaths and confirmed cases. The recovered data were not available for analysis. Hence the plot below has only confirmed and death cases.

In the same way, COVID 19 cases were also studied for understanding the patterns between January to April 2020 as shown in Figure 3.

The observations from the above patterns indicate that the number of cases with COVID-19 is relatively more compared to SARS and Ebola. And though it is just close to 100 days of the prevalence of COVID-19, still there is a high impact of transmission of this disease both in terms of spread and deaths.

Geographically, which part of the world is high impacted was also a key question to assess. By April 1st, it could be noticed that the number of countries affected by COVID-19 was 181, whereas the number of countries affected by SARS was 36, and 10 were impacted by Ebola. These two observations help understand with evidence the spread of the diseases and its patterns of spread can be further studied as per regions too. The pictorial representation of the spread of COVID-19 can be obsessed with the following Figure 4.

The spread of the disease so devastatingly would need lots of planning to be made by the respective governments and policymakers. Especially in countries like India where the population is 1.4 Billion, it is need of the hour to understand the patterns of the spread to effectively tackle it by organizing the resources, like the hospital beds, ventilation services, and medical staff and accessories. Hence, in the next section, the attempt was made to build machine learning-based models that can predict the number of cases in advance more effectively.

4 | METHODS USED

Three machine learning based models have been chosen to predict the number of positive cases. The models were effective, less time consuming and have produced some interesting results which will be detailed in the next section. Nonlinear Regression, Decision Tree based regression, and random forest models have been used. This section aims at giving a high level understanding of the models used in the study.

Multiple regression models deal with data that are linearly related. In simple language, multiple regression model can be considered as a weighted average of its independent variables. These linear regression models also have the ability to be used to solve more complex, nonlinear, problems. The problem that is being addressed in this paper is also a nonlinear problem. Hence, Nonlinear regression model was the first model used in the study.

The function used for training is as follows:

$$F(x) = a1 + a2v1 + a3v2 + a4v3 + a5v4 + a6v5 + a7v6, \quad (1)$$

F(X) = The target/ dependent variable.

The next method used for analysis is Regression based Decision Tree. Decision Tree is one of the most commonly used, practical approaches for supervised learning. It can be used to solve both Regression and Classification tasks with the latter being put more into practical application. The decision tree model is very good at handling

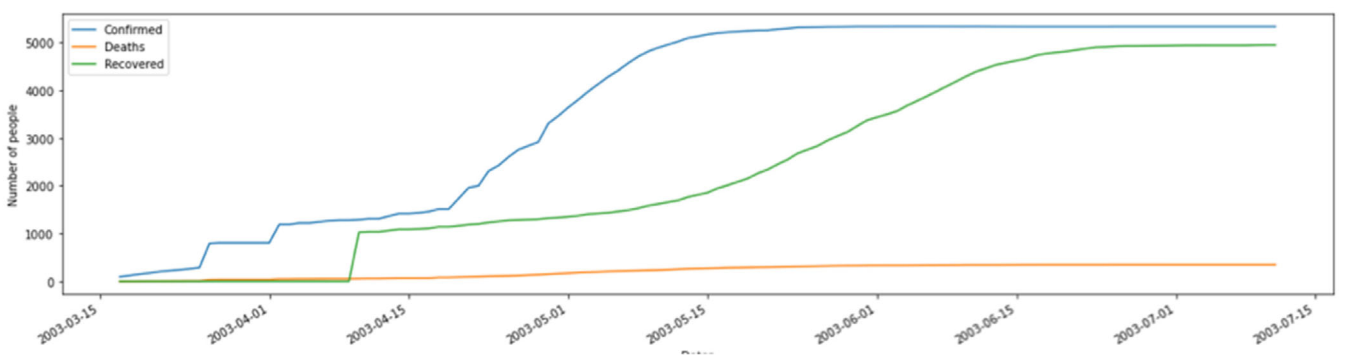

FIGURE 1 Patterns of cases for SARS

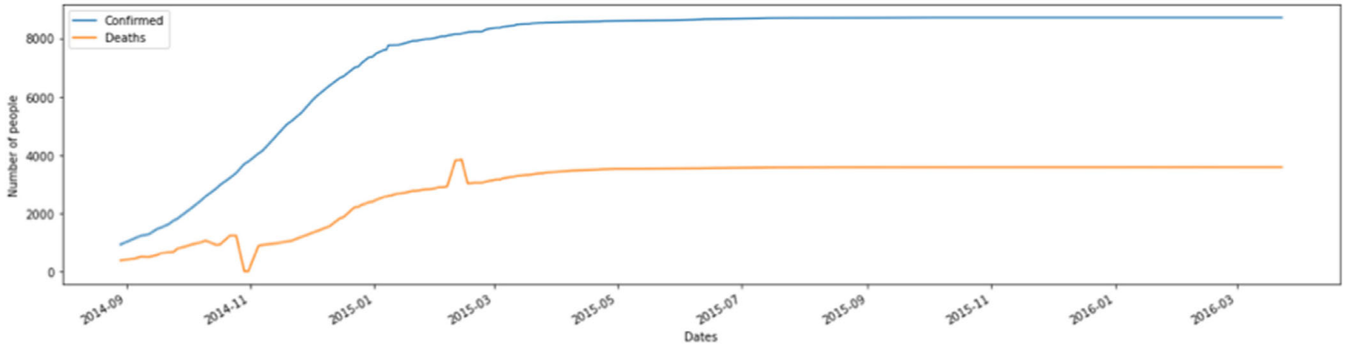

FIGURE 2 Patterns of cases for Ebola

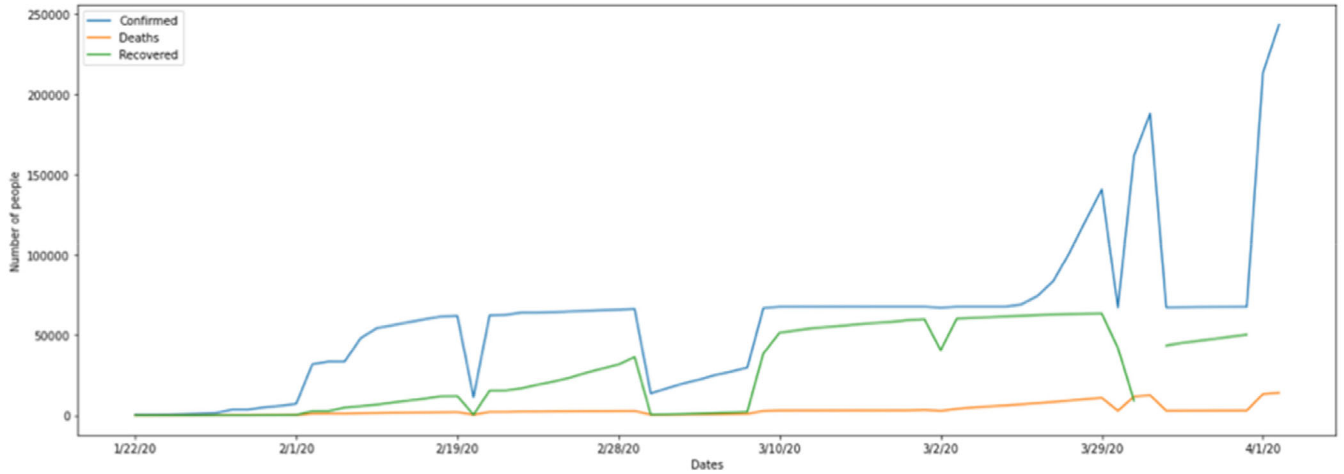

FIGURE 3 Patterns of cases for COVID 19

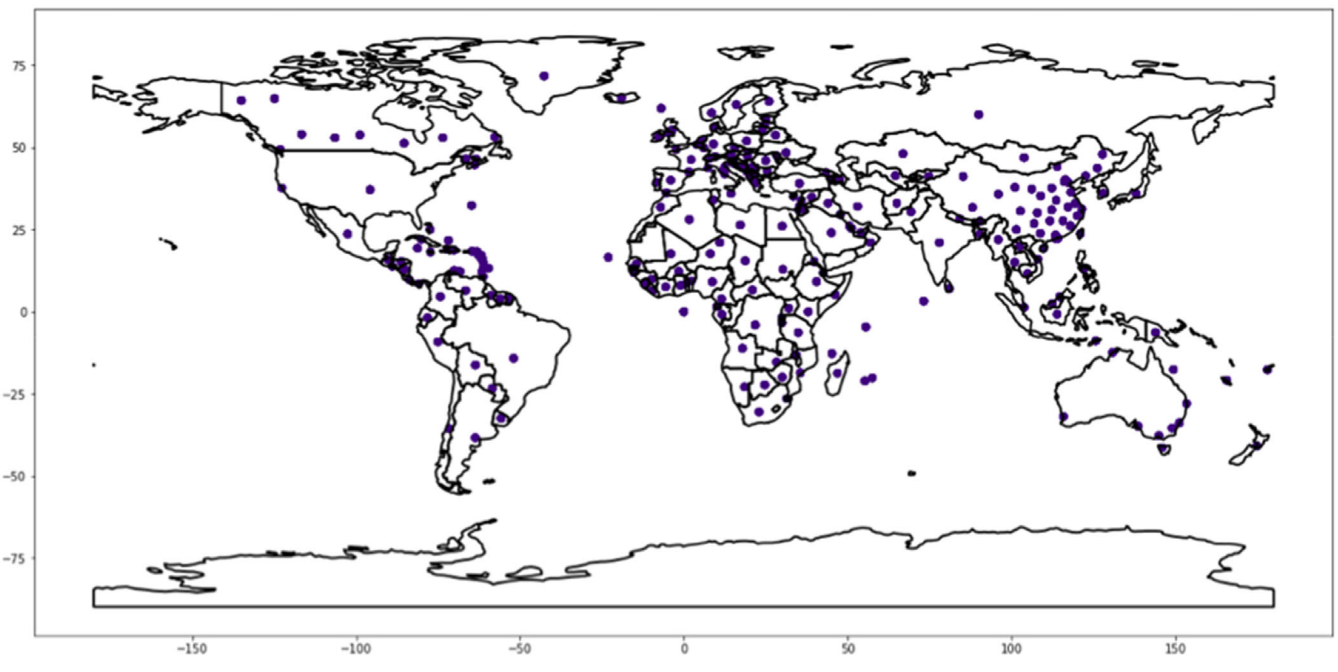

FIGURE 4 Regions where COVID-19 has impacted

tabular data with numerical features, or categorical features with fewer than hundreds of categories. Unlike linear models, decision trees are able to capture nonlinear interaction between the features and the target. Hence considering its ability to handle nonlinear data, this model also was used in the study. Another important and powerful model used in the study was Random Forest Model. Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction.

5 | MODEL DEVELOPMENT AND RESULTS

Many studies have used statistical and econometric (Bayes & Valdivieso, 2020; Benvenuto et al., 2020; Elmousalami & Hassanien, 2020). In this article, Machine learning based algorithms were used because of their predictability and handling nonlinearity. The second aspect of just considering previous data would always show growth in terms of forecast and hence it was decided to follow a pattern that already existed. Hence data from China were used to train the model. The reason China's data is considered because of the trend in the dataset. It had a decline of cases to imitate the pattern for India's condition and since during the time of analysis India had an increasing pattern for the active cases, it would not be apt to directly consider India's data for modeling. The second reason for considering China is because many of the features such as population, GDP, education levels are quite similar to India. Since the number of cases in India would not be a sufficient sample and to understand the flattening pattern, China's data were considered to build a model. So, China's data from January 15 was taken to date to imitate a similar pattern for India according to its features. Another important consideration to make for effective results of the model are the features that are to be considered to be fed into the model. In this case, the following features were developed to predict the possible increase in COVID-19 cases:

1. Deaths - Based on the number of deaths, the number of active cases vary as death might be premature and the active cases get reduced when a death occurs
2. Recovered - Since it takes 3–6 weeks for a patient recovers, it gives a better analysis to find the no of active cases
3. Confirmed - gives more dynamic info about the present scenario which is not reflected in deaths or recovered
4. Amount of testing - Since more testing gives more chances to find if patients exist, this is a very important feature
5. Lockdown - If a lockdown is present, the spread is reduced.
6. Lockdown features - The severity of the lockdown also plays an important role in the amount of spread of the disease.

The data were split into 80% training for developing the model and 20% testing to validate the effectiveness of the models developed. As discussed, the training was performed on the data from China. Three algorithms namely Nonlinear Regression, Decision Trees and Random Forests have been used to train the training dataset. The models were trained using programming language R. On the remaining 20% of the data, testing was performed with the three models. Table 1 explains the evaluation metrics obtained using the Mean Absolute Percentage Error (MAPE).

From Table 1, it is evident that the random forest model outperforms in comparison to the other two models. Hence, the Random forest was considered to predict the number of cases in India. The training of this model also was efficient and the convergent rate can be observed from Figure 5.

Figure 5 explains the convergence of the model. Also, the convergence of the model has happened considering less amount of time. This implies that the model can be used for quick results in real-time for different regions.

Another important observation behind building this model was to also explain the impact of lockdown on the number of cases. As discussed, the model was built using random forest and the following are the results obtained for India until May 3, 2020.

The Graph below shows the predictive analysis of the spread of the disease in India over the 40 days of Lockdown. It shows the effective rise in Active cases in India and its pattern over the lockdown period. The y-axis shows the number of active cases pattern over the lockdown and the x-axis shows the number of days elapsed from March 25th. The active cases reach a peak and tend to decrease as the lockdown period increases. The active cases decrease because of the patients' recovery or Death as seen in Kerala right now.

As shown in Figure 6, the graph was plotted considering the active cases. It is evident that the number of cases would flatten up or reduce because of the impact of lockdown. The number of positive cases would drastically reduce with the implementation of lockdown. It can be seen the peak will be somewhere around April 27th, and it will reduce over time and the new active cases per day will be almost zero. The rise in cases during the early stages is slow because of the

TABLE 1 Evaluation metrics of the three models used

Model	Nonlinear regression	Decision tree	Random forest
MAPE	0.24%	0.18%	0.02%

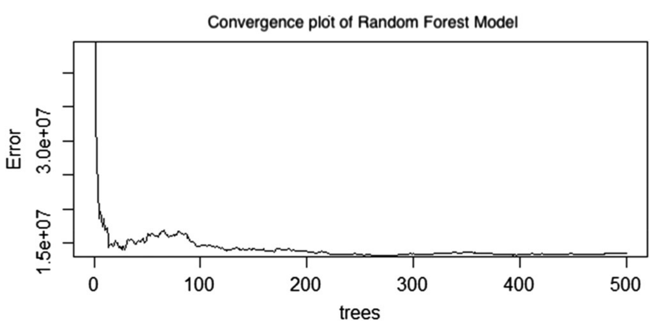

FIGURE 5 Convergence plot of random forest model

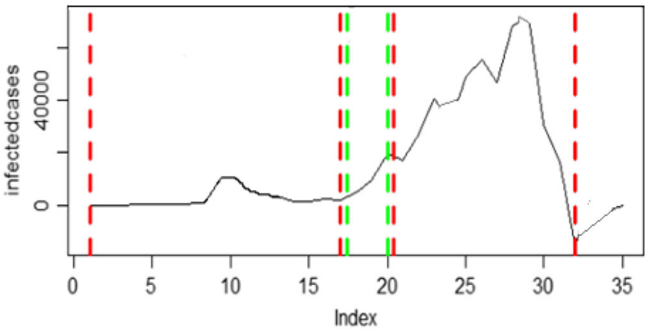

FIGURE 6 Random forest predictive analysis

lockdown and low transmission hosts which increases with time and hence the infection spreads faster which is shown in the graph. The active cases will have a maximum of 25 k cases before it starts coming down. In the next section, an attempt is made to discuss the steps to be implemented during these times.

6 | DISCUSSION AND RECOMMENDATIONS

From the above analysis, it can be interpreted that lockdown is one of the best ways to flatten the curve of positive cases and make sure the disease does not spread further. However, such lockdown for a longer time would impact the economy of any nation negatively as it is already witnessed by the negative growth of GDP in several countries including India. At this juncture, continuing with lockdown may not be the appropriate policy option. Until the vaccine for COVID-19 is fully commercialized, a balanced policy approach is imperative. Restriction on movements need to be lifted in a staggered manner to allow economic activities to function with a view to minimize the impact of lockdown on the economy.

Psychologically, people need to accept that corona is a disease that needs adequate preventive care to avoid the infection. Several initiatives have to be taken up at ground level to educate people about the severity of the disease and encourage citizens to follow measures such as physical distancing, using masks and gloves, sanitizing, and keeping themselves and their surroundings clean. This will reduce the reach of the disease in less affected places. From the economic perspective, whenever such pandemics have occurred in the past, there have been new business opportunities created. The government needs to focus on developing a fiscal stimulus package in which households who are deeply affected by COVID-19 get income support as in cash transfers to their bank accounts.

Going forward the government has to follow lockdown to containment zones where there are a large number of cases and strategize on developing certain practices for the flow of economic activities. At the same time as discussed above, this has to be restricted in places where the disease spread is high. At this juncture, the number of cases is on the surge and can only be reduced if necessary restriction is brought in high alert zones.

Hence, we propose a two-fold approach, where zones with high cases have to restrict openness and have strict measures and open up the economy in places where the cases are flattened. The opening up of the activities should be prioritized with key measures on the hold. Furthermore, there should be monitoring of the number of positive cases at micro level, preferably at Panchayat level, in close intervals of time. This will enable the government to take decisions on further course of actions as to whether to open up the economy further or to bring back the restrictions.

7 | CONCLUSION

The paper has made a comparison of different machine learning algorithms to predict the number of positive cases in India. One of the important contributions of the paper is to consider lockdown impact and apply it as a feature to the machine learning algorithm. Random forest Model performed better compared to the other two models. This model was used to develop the prediction of the number of cases for COVID-19 in India. It was observed that lockdown measures will flatten the curve. However, this paper focuses on suggesting strategies to the policymakers to lift the lockdown but periodically open up markets in few places where the number of cases is very low and monitor them constantly in addition to the citizens following precautions sincerely. In essence, the paper proposes a machine learning model that can predict the number of cases well in advance very effectively and also suggest some key inputs.

ORCID

Ritanjali Majhi https://orcid.org/0000-0001-6610-4732

REFERENCES

Anastassopoulou, C., Russo, L., Tsakris, A., & Siettos, C. (2020). Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS ONE*, *15*(3), e0230405.

Bastos, S. B., & Cajueiro, D. O. (2020). Modeling and forecasting the Covid-19 pandemic in Brazil. *arXiv preprint* arXiv:2003.14288.

Bayes, C., & Valdivieso, L. (2020). Modelling death rates due to COVID-19: A Bayesian approach. *arXiv preprint* arXiv:2004.02386.

Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*, *29*, 105340.

Botha, A. E., & Dednam, W. (2020). A simple iterative map forecast of the COVID-19 pandemic. *arXiv preprint* arXiv:2003.10532.

Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., & Priesemann, V. (2020). Inferring COVID-19 spreading rates and potential change points for case number forecasts. *arXiv preprint* arXiv:2004.01105.

Dev, M., & Sengupta, R. (2020. April). *Covid-19: Impact on the Indian economy* (WP-2020-013). Mumbai, India: Indira Gandhi Institute of Development Research.

Elmousalami, H. H., & Hassanien, A. E. (2020). Day level forecasting for Coronavirus Disease (COVID-19) spread: Analysis, modeling and recommendations. *arXiv preprint arXiv:2003.07778*.

Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals*, *134*, 109761.

Hall, I., Gani, R., Hughes, H., & Leach, S. (2007). Real-time epidemic forecasting for pandemic influenza. *Epidemiology and Infection*, *135*(3), 372–385. https://doi.org/10.1017/S0950268806007084

Hu, Z., Ge, Q., Jin, L., & Xiong, M. (2020). Artificial intelligence forecasting of covid-19 in China. *arXiv preprint arXiv:2002.07112*.

Liu, D., Clemente, L., Poirier, C., Ding, X., Chinazzi, M., Davis, J. T., … & Santillana, M. (2020). A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models. *arXiv preprint arXiv:2004.04019*.

Peng, L., Yang, W., Zhang, D., Zhuge, C., & Hong, L. (2020). Epidemic analysis of COVID-19 in China by dynamical modeling. *arXiv preprint arXiv:2002.06563*.

Perc, M., Gorišek Miksić, N., Slavinec, M., & Stožer, A. (2020). Forecasting COVID-19. *Frontiers in Physics*, *8*(127), 1–5.

Perone, G. (2020). *An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy* (No. 20/07). HEDG, c/o Department of Economics, University of York.

Petropoulos, F., Makridakis, S. (2020) Forecasting the novel coronavirus COVID-19. *PLoS ONE*, *15*(3), e0231236. https://doi.org/10.1371/journal.pone.0231236

Remuzzi, A., & Remuzzi, G. (2020). COVID-19, and Italy: What next? *The Lancet*, *395*(10231), 1225–1228.

Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J. M., … Chowell, G. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th. *Infectious Disease Modelling*, *5*, 256–263.

Shim, E., Tariq, A., Choi, W., Lee, Y., & Chowell, G. (2020). Transmission potential and severity of COVID-19 in South Korea. *International Journal of Infectious Diseases*, *93*, 339–344.

Singh, R., & Adhikari, R. (2020). Age-structured impact of social distancing on the COVID-19 epidemic in India. *arXiv preprint arXiv:2003.12055*.

The Economic Times. (2020). *India Covid count, April 29: Death toll crosses 1,000 mark, total cases over 31,000*. Retrieved from https://economictimes.indiatimes.com/news/politics-and-nation/india-covid-count-april-29-death-toll-crosses-1000-mark-total-cases-over-31000/articleshow/75441195.cms

WHO. (2020). *WHO Coronavirus Disease (COVID-19) Dashboard*. Retrieved from https://covid19.who.int/?gclid=Cj0KCQjwrIf3BRD1ARIsAMuugNuI-pJWSxU9q1fy7kJCoEgMh-_CXUU7sy1IMeHw_bSxN4L4PQlRaHYaAoArEALw_wcB

AUTHOR BIOGRAPHIES

Dr. Ritanjali Majhi is an Associate professor at School of Management, National Institute of Technology Karnataka, Surathkal. She has expertise in the field of Green Marketing, Big data Analysis, Consumer decision making, Time series Prediction, Artificial intelligence and soft and evolutionary application to Management, Marketing effectiveness metrics, E-Marketing, and Marketing Analytics. She has more than 15 years experience of research with publication in high impact factor journals. She has published 98 research papers in various referred International journals and conferences. Her total number of citations are 548 with h-index :10. She has reviewed many reputed journals.

Dr. Rahul Thangeda is Currently working as an Assistant Professor at NIT Warangal. He has also worked in Industries for many years and again the practical knowledge. He has published many papers in national and International Journals and conferences. He is a regular reviewers of many reputed journals.

Mr. Renu Prasad Sugasi did his B.Tech Mechanical from NItk, Surathkal. Currently working at Thorogood as Analytics and business consultant. He is interested in the field of data science and deep learning.

Niraj Kumar is Professor of Operations and Supply Chain Management at the Essex Business School (EBS). He is also the Associate Director of Research of the Management Science and Entrepreneurship (MSE) Group and the Director of Centre for Research on Entrepreneurship, Innovation Management and Internationalisation (REIMI). Previously, he held academic positions at the University of Liverpool Management School, Sheffield University Management School, Hull University Business School, University of Bath School of Management and University of Hong Kong. Professor Kumar has extensively published on various aspects of Operations and supply chain management in reputed international journals, books and conferences. Some of his research papers are highly cited and appeared in leading international journals such as International Journal of Operations and Production Management (IJOPM), Omega: the International Journal of Management Science, Transportation Research, International Journal of Production Economics, Expert Systems with Applications, International Journal of Production Research, Supply Chain Management and Technological Forecasting & Social Change. He currently serves on the editorial board of International Journal of Supply Chain and Operations Resilience (IJSCOR), and editorial review board of IJOPM and Sustainability. He also served on the academic and scientific committees of reputed international conferences such as European Operations Management Association (EurOMA) annual conference, International Symposium on Industrial Engineering & Operations Management, European Decision Sciences Institute annual conference, and Sustainable Operations, Logistics and Supply Chain Management conference. He is a member of the peer review group of Economic and Social Research Council (ESRC), Engineering and Physical Sciences Research Council (EPSRC), Global Challenges Research Fund (GCRF) and Newton Fund. Over the years, he has led and contributed to a range of research projects funded by European Commission (Horizon 2020), British Academy, Newton Fund, British Council, UGC and UKIERI. He has also worked on a variety of projects with several companies in Aerospace, Automobile, Retail, Construction, Food, IT and Waste management sectors in UK and abroad. In spare time, he actively participates in educational outreach and charitable activities. His research interests are primarily in Sustainable operations and supply chain management, Food supply chain, Supply chain innovation and Big data practices in supply chain. He has successfully supervised a number of PhD students in past.

How to cite this article: Majhi R, Thangeda R, Sugasi RP, Kumar N. Analysis and prediction of COVID-19 trajectory: A machine learning approach. *J Public Affairs*. 2020;e2537. https://doi.org/10.1002/pa.2537