Comparison of Traditional and Hybrid Time Series Models for Forecasting COVID-19 Cases

Samyak Prajapati[1*], Aman Swaraj[1], Ronak Lalwani[1], Akhil Narwal[1], Karan Verma[1], Ghanshyam Singh[2], Ashok Kumar[3]

181210046@nitdelhi.ac.in[1*], 182211001@nitdelhi.ac.in[1], 171210048@nitdelhi.ac.in[1], 171210005@nitdelhi.ac.in[1], karanverma@nitdelhi.ac.in[1], gsingh.ece@mnit.ac.in[2], kumarashoksaini@gmail.com[3],

National Institute of Technology, Delhi[1], Malaviya National Institute of Technology Jaipur[2], Government Mahila Engineering College, Ajmer[3]

*Corresponding Author

ABSTRACT

Background: Time series forecasting methods play critical role in estimating the spread of an epidemic. The coronavirus outbreak of December 2019 has already infected millions all over the world and continues to spread on. Just when the curve of the outbreak had started to flatten, many countries have again started to witness a rise in cases which is now being referred as the 2nd wave of the pandemic. A thorough analysis of time-series forecasting models is therefore required to equip state authorities and health officials with immediate strategies for future times.

Objective: This aims of the study are three-fold: (a) To model the overall trend of the spread; (b) To generate a short-term forecast of 10 days in countries with the highest incidence of confirmed cases (USA, India and Brazil); (c) To quantitatively determine the algorithm that is best suited for precise modelling of the linear and non-linear features of the time series.

Comparison: The comparison of forecasting models for the total cumulative cases of each country is carried out by comparing the reported data and the predicted value, and then ranking the algorithms (Prophet, Holt-Winters, LSTM, ARIMA, and ARIMA-NARNN) based on their RMSE, MAE and MAPE values.

Result: The hybrid combination of ARIMA and NARNN (Nonlinear Auto-Regression Neural Network) gave the best result among the selected models with a reduced RMSE, which proved to be almost 35.3% better than one of the most prevalent method of time-series prediction (ARIMA).

Conclusion: The results demonstrated the efficacy of the hybrid implementation of the ARIMA-NARNN model over other forecasting methods such as Prophet, Holt Winters, LSTM, and the ARIMA model in encapsulating the linear as well as non-linear patterns of the epidemical datasets.

Keywords: Hybrid Model, Forecasting, COVID-19, ARIMA, NARNN

1. INTRODUCTION

The novel coronavirus which first appeared in Wuhan, China in late 2019 has already infected over 104 million people and caused over 2.2 million deaths worldwide [1]. The ground-zero for the zoonotic spillover has been triangulated to the live-food markets of Wuhan [2], where the virus spread proximally due to direct exposure to animal shedding, bodily fluids, blood, and secretions [3]. In the absence of any tangible treatment, the pandemic has ruptured the concept of normal life while spreading with a rate of 1.8 (in India) [4].

To flatten the pandemic curve, several intervention policies have been implemented in many countries all over the world. However, these policies which include mobility and transportation restrictions, have provided temporary relief and the curve has started to rise again with possibilities of a second wave of the corona-virus (Fig.1). The situation has even more degraded in densely populated countries like India and Brazil which can't afford the luxury of lockdown due to socio-economic reasons. Therefore, rapid and predictable up-scaling of the healthcare framework is now most critical towards ensuring the availability of appropriate facilities during these demanding times.

Forecasting of epidemics and pandemics played a key role in curbing the spread of previous epidemics such as Ebola, Influenza etc. [5-11]. By providing insights into the severity of the infection and trend of the outbreak via simplified dashboards, such predictions not only help the general masses to acknowledge the severity of the pandemic but also prompts the state officials to take apt decisions in due time.

One widely used model in discerning the trends of an epidemic is the 'Susceptible–Exposed–Infectious–Resistant' (SEIR) model and researchers have actively employed the model for COVID-19 trend analysis as well [12-20]. While SEIR models are a proven tool in the analysis of these outbreaks, the algorithms and forecasting tools in the domain of Machine Learning and Artificial Intelligence have been considered equally important by researchers for forecasting [21-38].

Some of the commonly used techniques used in forecasting data are LSTMs [23-25], Exponential Smoothing [20, 30], Prophet [37, 38] and more. LSTMs are a form of (Recurrent Neural Networks) RNN with the ability to hold the previous data points for a short period of time which enables the concept of memory in forecasting the spread. Exponential Smoothing functions by making use of the previous lagged values in a weighted fashion; its ease of use and accuracy are some of the reasons justifying its popularity. Prophet is a fairly new technique developed by Facebook and built on Stan, this enables it to be extremely fast in parameter optimization and can easily handle irregular holidays and outliers in the data.

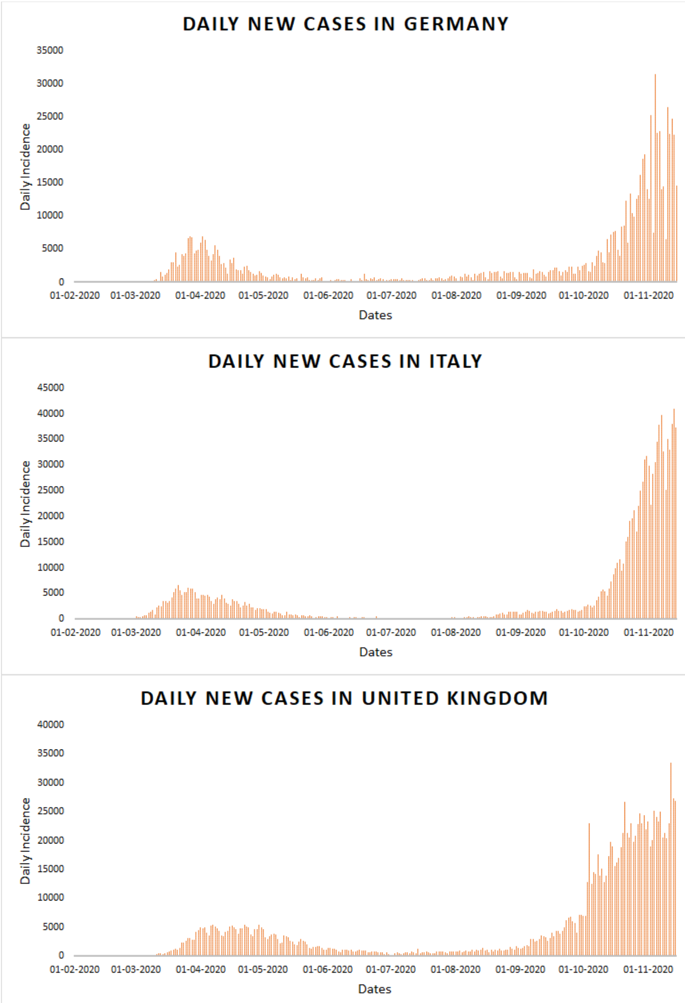

Figure 1: Second wave of COVID-19 rising in European countries respectively from top: Germany, Italy and UK (February 1 to November 14, 2020).

Upon analyzing the works, it is evident that ARIMA usually performs well in forecasting trends where linear patterns dominate the series attributing to its statistical properties and the well-known Box-Jenkins methodology [39]. However, ARIMA model assumes linear correlation structure among the time series values and therefore, it fails to capture the nonlinear patterns accurately. And thus, we propose a hybrid methodology of ARIMA and NARNN (Non-linear Autoregressive Neural Network), where we make use of NARNN for the non-linear patterns and use ARIMA for its forte, of forecasting of linear patterns in a time series.

With this motivation, we analyze and compare multiple time-series forecasting methods namely, Prophet, Holt-Winters, LSTM, ARIMA, and ARIMA-NARNN. Further, we make comparison between all the models and rank them according to their RMSE, MAE and MAPE values.

The rest of the paper is organized in the following sections: In Section 2, we depict the overall flow of the work along with an elaborate explanation of all the chosen models. In Section 3, we present the experiment analysis and results. Section 4 holds a discussion and Section 5 depicts the conclusion.

2. METHODOLOGY

This section elaborates on the data collection segment of our work, followed by a short description of the forecasting models that were used. The metrics used to assess the performance of the models are given at the end of this section. Figure 2 shows a summary of the workflow that was set up for each model. The time-series data was fed into multiple datasets and their results were compared based on their performance metrics and were used to rank them accordingly.

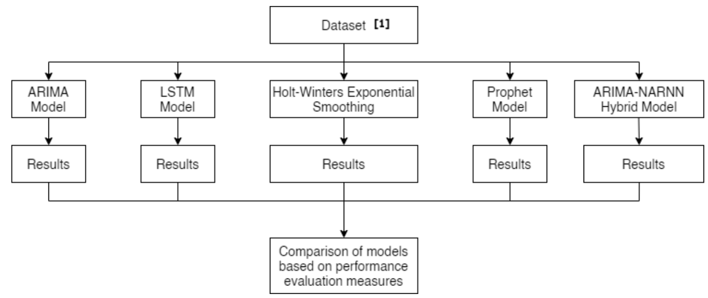

Figure 2: Flowchart indicating the various steps involved for forecasting

2.1 Dataset Description

The univariate time-series data of total cases of incidence was collected through the dataset published by John Hopkins University's Centre for System Science and Engineering [1]. For our study, we chose three countries that were severely affected by COVID-19, respectively the United States, India and Brazil. The models were analyzed on three different time intervals: (a) 6th May-15th May, (b) 21st July- 30th July and (c) 1st August- 10th August, and were then ranked accordingly based on Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). Each interval was split into training and test datasets, with the last 10 days of each interval being reserved as the test dataset, and the rest being used for training.

The motivation for providing short term accurate forecasts is to help the state authorities focus on a particular region of the country at a time. Since, all the three countries are densely populated, and on top of that, they are having a diverse demographic and dynamic mobility. Further, to meet the demands of accumulating enough COVID prevention resources (health centers, officials etc.) is also a challenge. Therefore, in such a scenario, providing short term forecasts suits the purpose best.

2.2 Models Analyzed

2.2.1 ARIMA

Auto-Regressive Integrated Moving Average (ARIMA) was proposed by Box and Jenkins in the 1970s [39] as a model which took varying trends, seasonal changes and random disturbances in account to predict the future values of the series. Due to these reasons, today, it is one of the most popular models that is used for forecasting time-series. It is denoted as ARIMA (p, d, q) where p and q are the orders of the AR and MA terms of the models respectively, and d represents the level of differencing used in the model. It can mathematically be represented as,

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_a Y_{t-a} + E_t - \theta_1 E_{t-1} - \theta_2 E_{t-2} - \cdots - \theta_C E_{t-C} \quad - \quad (1)$$

Where $\boldsymbol{Y_t}$ denotes the computed value of at the given time t, $\boldsymbol{\phi}$i and $\boldsymbol{\theta}$j are the coefficients of the AR and MA models respectively and $\boldsymbol{E_t}$ is the random error occurring at time $\boldsymbol{t}$.

2.2.2 LSTM

LSTMs are a form of a recurrent neural network (RNN) and as suggested by their name Long Short Term Memory, they allow for the model to retain information about the dataset that was previously computed. While most forms of RNNs can utilize the previous data in some form, LSTMs have the intrinsic ability to "store" the data for a short duration. This is achieved by the use of multiple "gates" and by modifying the cell state. Each gate is essentially a function which computes an output determining the way cell state has to be modified.

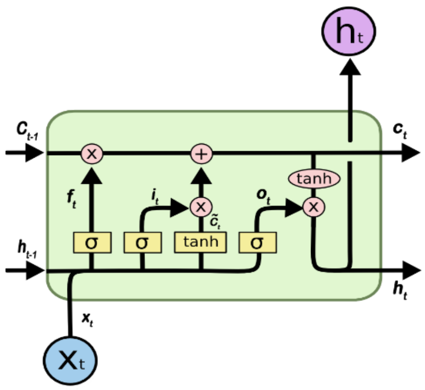

Figure 3: Pictorial Representation of an LSTM Node [40]

Each gate can easily be attributed to an activation function, where **x** is the feature vector, $\mathbf{h_{t-1}}$ is the output of cell **t-1**, $\mathbf{C_{t-1}}$ is cell state after cell **t-1**, $\mathbf{c_t}$ is the cell state after cell **t** and $\mathbf{h_t}$ is the output of cell **t**, thus the computations are,

$$f_t = \sigma\,(W_f \cdot [h_{t-1}, x_t] + b_f), \qquad - \quad (2)$$

$$i_t = \sigma\,(W_i \cdot [h_{t-1}, x_t] + b_i), \qquad - \quad (3)$$

$$\tilde{c}_t = tanh\,(W_c \cdot [h_{t-1}, x_t] + b_c), \qquad - \quad (4)$$

$$c_t = f * c_{t-1} + i_t * \tilde{c}_t, \qquad - \quad (5)$$

$$o_t = \sigma\,(W_o \cdot [h_{t-1}, x_t] + b_o), \qquad - \quad (6)$$

$$h_t = o_t * tanh\,(c_t) \qquad - \quad (7)$$

2.2.3 HOLT WINTERS

Exponential Smoothing is a univariate time-series modelling technique where close attention is paid to the precursing values and weights are assigned to them depending on the lag, these are then factored into the prediction of future values. There are mainly three versions, which selectively focus on the combinations of Level, Trend and Seasonality. Single Exponential Smoothing works by modelling along with the lags of the levels, whereas, Double Exponential Smoothing utilizes levels and trends, and Triple Exponential Smoothing and Holt-Winters Exponential Smoothing incorporates all three elements during its computation. It mainly has 3 parameters,

- α: Smoothing factor for the level,
- β: Smoothing factor for the trend,
- γ: Smoothing factor for the seasonality.

The mathematical equation for this is described as:

$$F_{i+k} = (L_i + k * B_i) * S_{i+k-m}, \qquad - \quad (8)$$

Where *m* is the period length of the seasonal variation, *k* is the number of steps ahead from any arbitrary step *i*, and,

$$B_i = \beta * [L_i - L_{i-1}] + (1-\beta) * B_{i-1}, \qquad - \quad (9)$$

$$L_i = \alpha * \frac{T_i}{S_{i-m}} + (1-\alpha) * [L_{i-1} + B_{i-1}], \qquad - \quad (10)$$

$$S_i = \gamma * \frac{T_i}{L_i} + (1-\gamma) * S_{i-m} \qquad - \quad (11)$$

2.2.4 PROPHET

Prophet is an open-source time-series forecasting library developed by Facebook which runs upon Stan. It is based on a decomposable additive model constituting three major components; trends, seasonality and holidays. The equation for the above can be interpreted as,

$$y(t) = g(t) + s(t) + h(t) + \varepsilon t, \qquad - \quad (12)$$

where, $g(t)$ represents the piecewise linear or the logistic growth curve for modelling the non-periodic changes in the time series, $s(t)$ is the periodical changes that occur with seasonality, $h(t)$ includes the effects of holidays (which can be provided by the user) along with schedules that may be irregular in nature and finally, εt is the error term which takes in consideration any irregular changes that may not be accommodated by the model.

2.3 Proposed Model

This illustrates the creation of a hybrid model between ARIMA (Auto-Regressive Integrated Moving Average) and a Non-linear Autoregressive Neural Network (NARNN) that can selectively work on a time series by isolating and working on individual areas of strengths. In general, a time series contains a linear auto correlated structure and a non-linear component as well, this can be represented as,

$$Z_t = LIN_t + NON_t, \qquad - \quad (13)$$

Where $\boldsymbol{Z_t}$ is the time-series having a linear component and a non-linear component, which are indicated respectively by $\boldsymbol{LIN_t}$ and $\boldsymbol{NON_t}$.

The first step in this hybridization is to create an ARIMA model with appropriate (p, d, q) parameters. The strength of ARIMA lies in the forecasting of linear dependencies, thus, fitting of the input features into this model will result in the generation of the linear component ($\boldsymbol{LIN_t}$). To isolate the non-linear dependencies, residuals of the ARIMA model must be generated as $\boldsymbol{RES_t}$.

$$RES_t = Z_t - \widehat{LIN}_t, \qquad - \quad (14)$$

Where $\boldsymbol{\widehat{LIN}_t}$ is the value forecasted by the ARIMA model at a time $\boldsymbol{t}$. By modelling the residuals using ANNs, the non-linear segments can be realized and thus, the residuals are fed into a NARNN model which comprises of n input nodes, modelling it into,

$$Rt = fx\,(R_{t-1}, R_{t-2}, \dots, R_{t-n}) + \epsilon_t, \qquad - \quad (15)$$

Where, $\boldsymbol{fx}$ constitutes as the non-linear function that is being evaluated by the NARNN model, and the error generated in doing so is represented by $\boldsymbol{\epsilon_t}$. The final equation then represented by the equation below, where $\boldsymbol{\hat{Z}_t}$ indicated the final forecast of the time-series at time $\boldsymbol{t}$ and $\boldsymbol{\widehat{NON}_t}$ is the residual forecast.

$$Zt = LINt + NONt, \qquad - \quad (16)$$

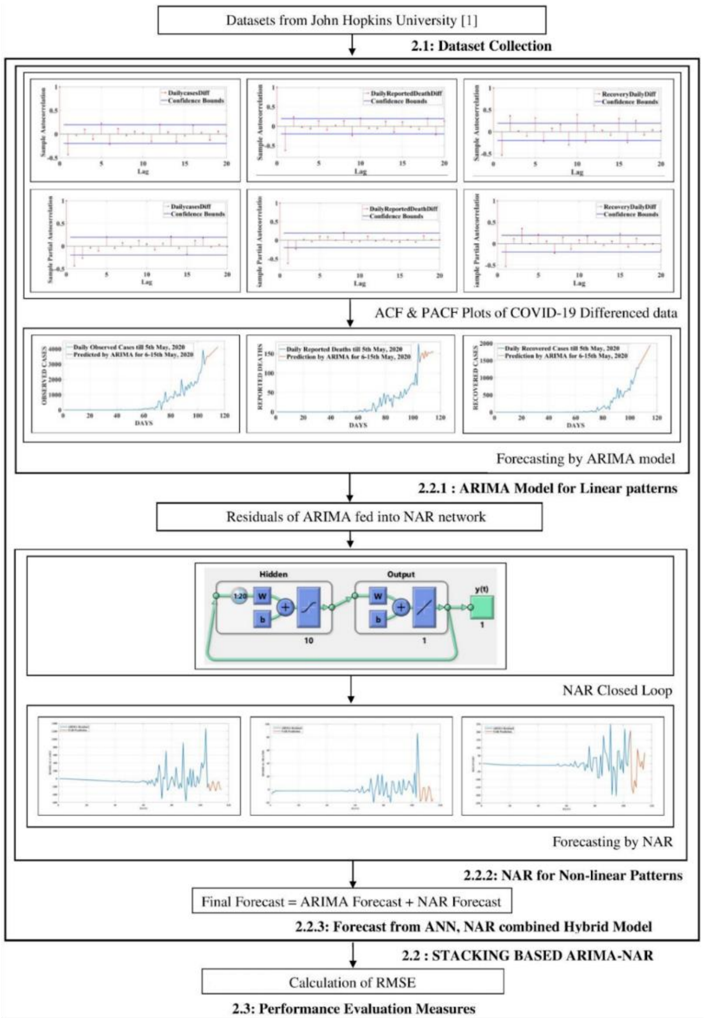

Figure 4: Pictorial representation of the process for the ARIMA-NARNN Hybrid model [41]

2.4 Performance Evaluation Measures

The true accuracy of a model is tested by comparing the prediction values with the true values. There lie several different performance parameters that can be used to generate an accuracy measure; however, this study aims at using a multiple performance metrics for the ease of ranking them, Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were chosen. These computations were implemented through scikit-learn's metrics module and the mathematical formulas for the same computations are shown as follows,

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Z_t - \hat{Z}_t\right)^2}, \quad - \quad (17)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|Z_t - \hat{Z}_t\right|, \quad - \quad (18)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Z_t - \hat{Z}_t}{Z_t}\right| \quad - \quad (19)$$

Here, $\boldsymbol{n}$ stands for the number of data points available, $\boldsymbol{Z_t}$ is representative of the observed value at time $\boldsymbol{t}$ and $\boldsymbol{\hat{Z}_t}$ denotes the estimated value at time $\boldsymbol{t}$. Lower values of RMSE, MAE and MAPE indicate the better fitting of the data to the model.

3. RESULTS

The dataset for the analysis was obtained through John Hopkins University's Centre for System Science and Engineering. A thorough analysis of these modelling techniques was done for the cumulative cases on three different intervals, (6th May- 15th May, 21st July- 30th July and 1st August- 10th August), prioritizing India and using the incidence count of the other countries to confirm the observations. This led us to achieve the following results as shown in Table 1(a) for India:

Table 1. (a) Prediction accuracy evaluation for cumulative cases of COVID-19 in India between the 6th and the 15th of May, 21st July and the 30th of July, and Aug 1st and the Aug 10th, 2020

Intervals	**Prophet** *(RMSE) (MAE) (MAPE) (Values in 90% Confidence Intervals)*	**Holt-Winters** *(RMSE) (MAE) (MAPE) (Values in 90% Confidence Intervals)*	**LSTM** *(RMSE) (MAE) (MAPE) (Values in 90% Confidence Intervals)*	**ARIMA** *(RMSE) (MAE) (MAPE) (Values in 90% Confidence Intervals)*	**Hybrid** *(RMSE) (MAE) (MAPE) (Values in 90% Confidence Intervals)*
6MAY-15MAY	4484.73 3206.9 4.330 *30%*	3556.82 279.2 0.436 *40%*	7755.93 5924.5 7.871 *20%*	502.30 459.6 0.718 *40%*	437.30 341.8 0.523 *40%*
21JUL-30JUL	35837.36 28300.4 1.946 *30%*	31493.75 24364.1 1.662 *50%*	132923.72 120040.6 8.474 *40%*	3961.78 3150.7 0.246 *40%*	3119.40 2566.7 0.197 *40%*
1AUG-10AUG	87260.64 64841.6 3.118 *60%*	10152.17 8256.5 0.420 *40%*	49587.15 43971.6 2.179 *40%*	3499.25 3041.5 0.156 *40%*	2825.65 2480.5 0.127 *40%*

From the above tabulated data in Table 1(a), albeit the number of elements in the 90% confidence intervals is on the lower end of the spectrum, it is clearly evident that ARIMA performs the best when compared with other popular time-series forecasting methods and in order to substantiate that our hybrid model is able to overcome the shortcomings of ARIMA, we compared the performance of ARIMA and our hybrid model on the same interval on USA and Brazil and the results are tabulated in Table 1(b) and 1(c) respectively.

Table 1. (b) Prediction accuracy evaluation for cumulative cases of COVID -19 in USA between the 6th and the 15th of May, 2020

Intervals	**ARIMA** *(RMSE) (MAE) (MAPE) (Values in 90% Confidence Intervals)*	**Hybrid** *(RMSE) (MAE) (MAPE) (Values in 90% Confidence Intervals)*
6MAY-15MAY	5676.96 4624.9 0.342 *40%*	3674.51 3009.9 0.223 *40%*

Table 1. (c) Prediction accuracy evaluation for cumulative cases of COVID -19 in Brazil between the 6th and the 15th of May, 2020

Intervals	**ARIMA** *(RMSE)* *(MAE)* *(MAPE)* *(Values in 90% Confidence Intervals)*	**Hybrid** *(RMSE)* *(MAE)* *(MAPE)* *(Values in 90% Confidence Intervals)*
6MAY-15MAY	10062.98 8729.90 5.175 *40%*	7803.56 6851.90 4.086 *40%*

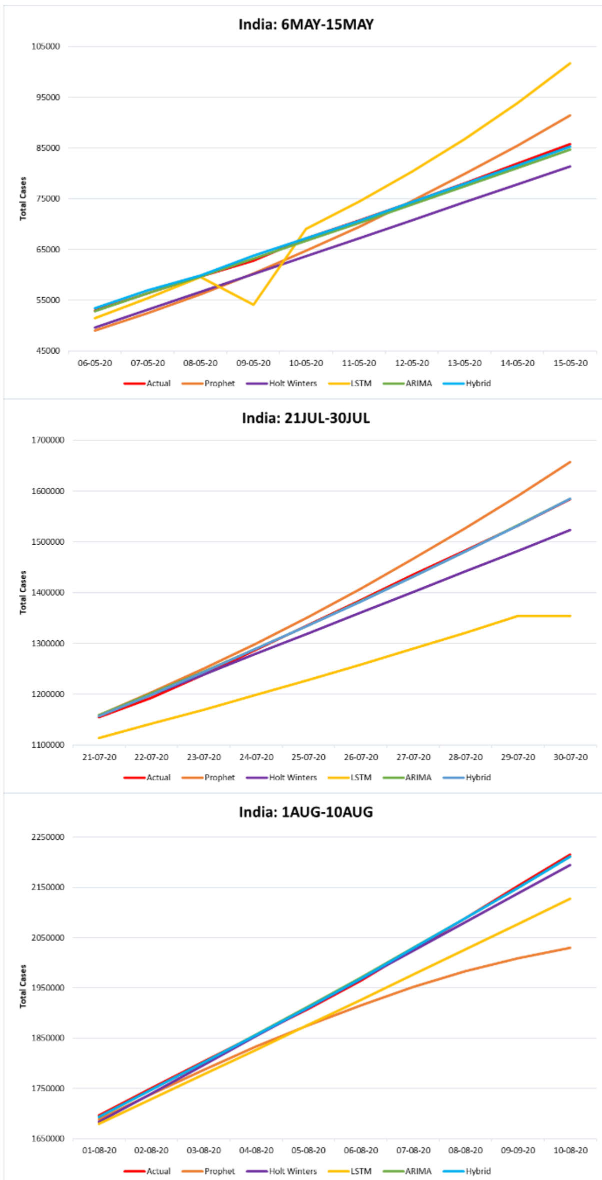

Figure 5(a), (b), (c): Graphical Illustration of the forecasts for India respectively on 6th May - 15th May, 21st July- 30th July and 1st August- 10th August from top

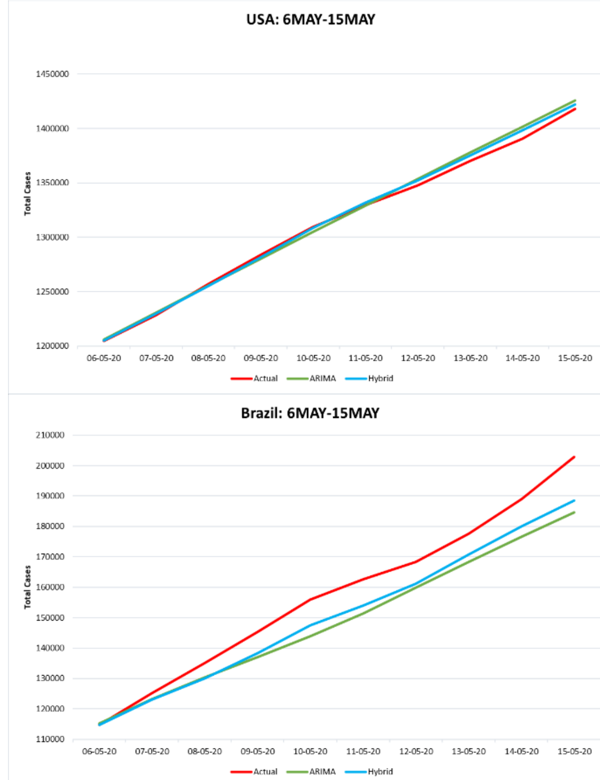

Figure 6(a), (b): Graphical Illustration of the forecasts for USA and Brazil respectively from top (between 6 - 15th May, 2020)

4. Discussion and Conclusion

Our study highlighted the key point of analyzing linear and nonlinear patterns in a time series forecasting model. From Table (1-a, b, c) we see clearly how RMSE value of the hybrid model is minimal when compared to the other stipulated models. This is attributed to the hybrid model having the intrinsic ability to detect and train itself on the non-linear features of the data as well.

The Indian time-series of total cases, there is a notable rise in the non-linear features along with linear features. While ARIMA is able to abstract out these non-linear features and perform well on just the linear features; the Hybrid model is able analyze these non-linear features as well and is able to substantially improve its performance, thus exhibiting the least RSME amongst the other models.

With most countries hitting their share of the surge of COVID cases, aptly named as the "second-wave", this surge is the result of multiple factors which range majorly from the softening of threat in the mindset of the common folk and the relaxations in the government-imposed policies due to economic slowdowns [42] in several sectors of the economy and its cascading effect on other sectors. In such times, where the governing bodies are struggling to stabilize economic recessions, true and precise forecasting of seasonal diseases and the spread of contagions is an ever-growing priority.

Among the chosen time-series modelling models, ignoring the outliers, ARIMA performed the best and its performance was then improved by our implementation of the ARIMA-NARNN Hybrid model. While it is important to model the long term spread of contagions, short-term forecasts play a vital role in the rapid deployment of resources and manpower.

The ever-increasing habitat loss of wildlife leads the animals in search of a new home, this search brings them closer to us; and a consequence of this is that it also exposes us to them. Keeping this in mind, it is plausible that we may get exposed to a lot more zoonotic pathogens in the coming future, and then the only way to circumvent another pandemic is to prepare ourselves in monitoring and curb the spread of infectious diseases.

Practical insights of how the spread of diseases may transpire would lead to the development of an understanding between the policymakers, and hence preferable allocation and management of crucial resources under tight time constraints.

5. Abbreviations

ARIMA: Auto-Regressive Integrated Moving Average
COVID: Corona Virus Disease
LSTM: Long Short Term Memory
NARNN: Non-linear Autoregressive Neural Network
RMSE: Root Mean Square Error
RNN: Recurrent Neural Networks
SEIR: Susceptible–Exposed–Infectious–Resistant

6. References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Inf Dis. 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1.
2. Novel Coronavirus Pneumonia Emergency Response Epidemiology Team (February 2020). "[The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China]". Zhonghua Liu Xing Bing Xue Za Zhi = Zhonghua Liuxingbingxue Zazhi (in Chinese). 41 (2): 145–151. doi:10.3760/cma.j.issn.0254-6450.2020.02.003. PMID 32064853. S2CID 211133882.
3. Kevin Berger. “The Man Who Saw the Pandemic Coming”. Nautilus. Issue 83.
4. Seema Patrikar, Deepti Poojary et al, “Projections for novel coronavirus (COVID-19) and evaluation of epidemic response strategies for India”, Medical Journal Armed Forces India, vol. 76, no. 3, 268-275, 2020
5. W. Jia, X. Li, K. Tan, and G. Xie, "Predicting the outbreak of the hand-foot- mouth diseases in china using recurrent neural network," in 2019 IEEE International Conference on Healthcare Informatics (ICHI). IEEE, 2019, pp. 1–4.
6. Shashvat, Kumar, RikmantraBasu, and Amol P. Bhondekar. "Application of time series methods for dengue cases in North India (Chandigarh)." Journal of Public Health (2019): 1-9.
7. Forna, P. Nouvellet, I. Dorigatti, and C. Donnelly, "Case fatality ratio estimates for the 2013–2016 west African Ebola epidemic: application of boosted regression trees for imputation," International Journal of Infectious Diseases, vol. 79, p. 128, 2019.
8. Shashvat, Kumar, et al. "Comparison of time series models predicting trends in typhoid cases in northern India." Southeast Asian Journal of Tropical Medicine and Public Health 50.2 (2019): 347-356.
9. S.-L. Jhuo, M.-T. Hsieh, T.-C. Weng, M.-J. Chen, C.-M. Yang, and C. H. Yeh, "Trend prediction of influenza and the associated pneumonia in Taiwan using machine learning," in 2019 International Symposium on Intelligent Signal Processing.
10. G. Machado, C. Vilalta, M. Recamonde-Mendoza, C. Corzo, M. Torremorell, A. Perez, and K. VanderWaal, "Identifying outbreaks of porcine epidemic diarrhoea virus through animal movements and spatial neighbourhoods," Scientific reports, vol. 9, no. 1, pp. 1–12, 2019.
11. Kalipe, G., Gautham, V., Behera, R. K. (2018, December). Predicting malarial outbreak using Machine Learning and Deep Learning approach: A review and analysis. In 2018 International Conference on Information Technology (ICIT) (pp. 33-38). IEEE.
12. Kuniya T. Prediction of the Epidemic Peak of Coronavirus Disease in Japan, 2020. J Clin Med. 2020; 9 (3): E789. Published 2020 March 13. doi: 10.3390 / jcm9030789.
13. Gupta, Rajan, et al. "SEIR and Regression Model based COVID-19 outbreak predictions in India." medRxiv (2020).
14. Yuan, George Xianzhi, et al. "The Framework for the Prediction of the Critical Turning Period for Outbreak of COVID-19 Spread in China based on the iSEIR Model." Available at SSRN 3568776 (2020).
15. Anastassopoulou, L. Russo, A. Tsakris, and C. Siettos, "Data-Based Analysis, Modelling and Forecasting of the novel Coronavirus (2019-nCoV) outbreak,"medRxiv, no. February, p. 2020.02.11.20022186, 2020, doi: 10.1101/2020.02.11.20022186.
16. Joseph T. Wu, Kathy Leung, Mary Bushman, Nishant Kishore, Rene Niehus, Pablo M. de Salazar, Benjamin J. Cowling, Marc Lipsitch& Gabriel M. Leung. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nature Medicine (2020), March 19 2020.https://doi.org/10.1038/s41591-020-0822-7.
17. J. T. Wu, Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020; 395: 689–97. https://doi.org/10.1016/ S0140-6736(20)30260-9.
18. Kiesha Prem, Yang Liu, Timothy W Russell, Adam J Kucharski, Rosalind M Eggo, Nicholas Davies. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health 2020. Published Online March 25, 2020.https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(20)30072-4/fulltext.
19. X. Liu, Geoffrey Hewings, Shouyang Wang, Minghui Qin, Xin Xiang, Shan Zheng, Xuefeng Li. Modelling the situation of COVID-19 and effects of different containment strategies in China with dynamic differential equations and parameters estimation. medRxiv preprint doi: https://doi.org/10.1101/2020.03.09.20033498, 2020.
20. Qianying Lin, Shi Zhao, Daozhou Gao, Yijun Lou, Shu Yang, Salihu S. Musa, Maggie H. Wang, Yongli Cai, Weiming Wang, Lin Yang, Daihai He. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. International Journal of Infectious Diseases (93) (2020), 211-216.
21. J.L. Murray. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator days and deaths by U.S. state in the next 4 months. MedRxiv. March 26 2020. doi:10.1101/2020.03.27.20043752.
22. Elmousalami, H. H., &Hassanien, A. E. (2020). Day Level Forecasting for Coronavirus Disease (COVID-19) Spread: Analysis, Modelling and Recommendations. ArXiv preprint arXiv:2003.07778.
23. Pal, Ratnabali, et al. "Neural network-based country wise risk prediction of COVID-19." arXiv preprint arXiv:2004.00959 (2020).
24. Bandyopadhyay, Samir Kumar, and Shawni Dutta. "Machine Learning Approach for Confirmation of COVID-19 Cases: Positive, Negative, Death and Release." medRxiv (2020).

25. Punn, Narinder Singh, Sanjay Kumar Sonbhadra, and Sonali Agarwal. "COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms." medRxiv (2020).
26. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. Data Brief. 2020; 29: 105340. Published 2020 Feb 26. doi: 10.1016 / j .dib.2020.105340.
27. Ding, Guorong, et al. "Brief Analysis of the ARIMA model on the COVID-19 in Italy." medRxiv (2020).
28. Perone, Gaetano. 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, 2020.
29. Dehesh, T., Mardani-Fard, H. A., &Dehesh, P. (2020). Forecasting of COVID-19 Confirmed Cases in Different Countries with ARIMA Models. MedRxiv.
30. Gupta, Rajan, and Saibal Kumar Pal. "Trend Analysis and Forecasting of COVID-19 outbreak in India." medRxiv (2020).
31. Tandon, Hiteshi, et al. "Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future." arXiv preprint arXiv:2004.07859 (2020).
32. Kumar, Pavan, et al. "Forecasting the dynamics of COVID-19 Pandemic in Top 15 countries in April 2020: ARIMA Model with Machine Learning Approach." medRxiv (2020).
33. Shi, Z., & Fang, Y. (2020). Temporal relationship between outbound traffic from Wuhan and the 2019 coronavirus disease (COVID-19) incidence in China. MedRxiv.
34. Al-Qaness, Mohammed AA, et al. "Optimization method for forecasting confirmed cases of COVID-19 in China." Journal of Clinical Medicine 9.3 (2020): 674.
35. Ceylan, Zeynep. "Estimation of COVID-19 prevalence in Italy, Spain, and France." Science of The Total Environment (2020): 138817.
36. Moftakhar, Leila, S. E. I. F. Mozhgan, and Marziyeh Sadat Safe. "Exponentially Increasing Trend of Infected Patients with COVID-19 in Iran: A Comparison of Neural Network and ARIMA Forecasting Models." Iranian Journal of Public Health 49 (2020): 92-100.
37. Tulshyan, V., Sharma, D., & Mittal, M. (2020). An Eye on the Future of COVID’19: Prediction of Likely Positive Cases and Fatality in India over A 30 Days Horizon using Prophet Model. Disaster Medicine and Public Health Preparedness, 1-20. doi:10.1017/dmp.2020.444
38. Xie, C., Wen, H., Yang, W. et al. Trend analysis and forecast of daily reported incidence of hand, foot and mouth disease in Hubei, China by Prophet model. Sci Rep 11, 1445 (2021). https://doi.org/10.1038/s41598-021-81100-2
39. Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. Time Series Analysis: Forecasting and Control. John Wiley & Sons.
40. Olah, C. (2015, September 27). Understanding LSTM Networks. https://colah.github.io/posts/2015-08-Understanding-LSTMs/.
41. Swaraj, Aman & Kaur, Arshpreet & Verma, Karan & Singh, Ghanshyam & Kumar, Ashok & Sales, Leandro. (2020). Implementation of Stacking Based ARIMA Model for Prediction of Covid-19 Cases in India. 10.21203/rs.3.rs-52063/v1.
42. “A second wave of coronavirus may force renewed lockdowns”. Pharmaceutical-technology.com. November 2nd 2020. https://pharmaceutical-technology.com/comment/coronavirus-second-wave-lockdowns