Time series forecasting and machine learning models for predicting and analyzing Covid-19 daily new cases and cumulative cases

1st Md. Ekram Hossain
Master in Data Science
University of Erlangen-Nuremberg
Erlangen, Germany
md.ekram.hossain@fau.de

2nd Vaishnav Negi
Master in Data Science
University of Erlangen-Nuremberg
Erlangen, Germany
vaishnav.negi@fau.de

3rd Rachana Kafle
Master in Data Science
University of Erlangen-Nuremberg
Erlangen, Germany
rachana.kafle@fau.de

***Abstract*—The COVID-19 pandemic has posed a severe danger to worldwide human health, with far-reaching consequences for social activities and local economy. The use of ARIMA, LSTM, and Prophet models to forecast daily new cases and cumulative confirmed cases in the United States, Germany, and India is proposed in this work. The performance of these models was tested using RMSE, MAE, and MAPE measures using the WHO's COVID-19 data set from January 1, 2020 to November 30, 2022. According to the study, ARIMA is a more trustworthy model for predicting new instances, but LSTM is better at projecting accumulated cases. These findings can help governments and healthcare providers make educated decisions on how to handle the pandemic. The Prophet model predicts daily new instances with adequate accuracy in the United States, but the ARIMA model is acceptable for forecasting Germany and India, allowing effective policies and safeguards to be developed in other nations. This research throws light on epidemic trends and offers insights on the epidemiological control of these areas.**

***Index Terms*—Covid-19, LSTM, ARIMA, Prophet, Time series analysis**

I. Introduction

The new coronavirus, also known as COVID-19, is causing severe trouble to global health. It's a respiratory illness that causes lung inflammation [1- 3] and spreads fleetly due to its unique transmission characteristics. It poses a significant threat to public life and security, with a broader range of transmission and a briskly spread than former contagious conditions. This has caused the need to come significant trouble to public health, leading to vast dislocation of diurnal life and causing significant enterprises for global security. In addition to its fast transmission rate, the complaint has also been shown to have an advanced threat of transmission, making it a dangerous public health concern [4- 5]. The ongoing COVID-19 epidemic highlights the need for increased global cooperation in public health enterprise and the significance of taking the necessary preventives to cover ourselves and those around us.

On March 11, 2020, the World Health Organization (WHO) proclaimed a worldwide pandemic of a new strain of the coronavirus family known as COVID-19. This virus is separate from previous severe respiratory illnesses such as SARS (Severe Acute Respiratory Syndrome) and MERS (Middle East Respiratory Syndrome). It has been demonstrated to be very infectious, with high rates of infection and fatality [6]. COVID-19 had spread to at least 17 nations by September 14, 2020, infecting over 30.8 million individuals and killing in,000 persons. The disease has spread throughout communities, causing widespread economic and societal consequences. According to research, the virus is zoophytic in origin, which means it may be spread from wildlife to people. The elderly and those with underlying health concerns such as heart disease, diabetes, and respiratory difficulties are at a greater risk of severe instances, with an extreme case fatality rate of 1.5 [7]. In Fig 1- the process of research work has been shown.

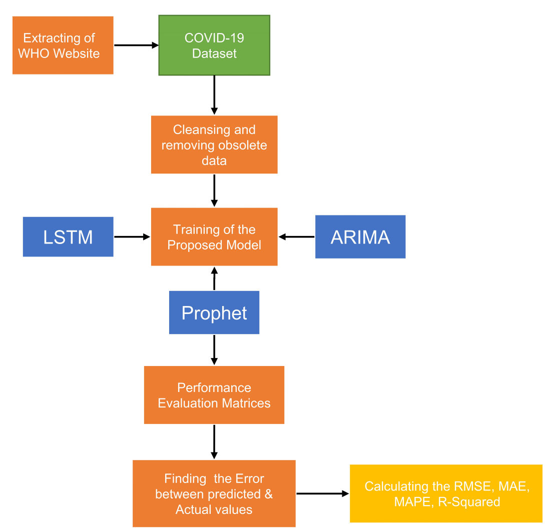

Fig. 1. The proposed methodology of the COVID-19 forecasting model.

The COVID-19 pandemic has had a significant impact on global health and economies. As the pandemic continues to evolve, accurate forecasting of future cases is critical for informed decision-making. Time-series analysis is a widely used method for forecasting and trend analysis, and the dynamic fluctuation trend of COVID-19 instances, due to various epidemic prevention and control efforts, makes it a suitable candidate for this type of analysis. However, the nonlinear nature of the COVID-19 data series poses a challenge for traditional single time-series models. This study aims to address this challenge by exploring alternative methods for COVID-19 case forecasting that better capture the nonlinear aspects of the data. The COVID-19 pandemic necessitates the need for accurate forecasting methods. The intricate nature of nonlinear data patterns calls for evaluating and comparing effective prediction models. This research endeavors to examine the forecasting capabilities of both time-series and machine-learning models for COVID-19 and furnish valuable data to support informed decision-making in the prevention and treatment of the epidemic [8-9]. In fig 2 the actual data are showing from Jan 2020 to Nov 2022. The current study aims

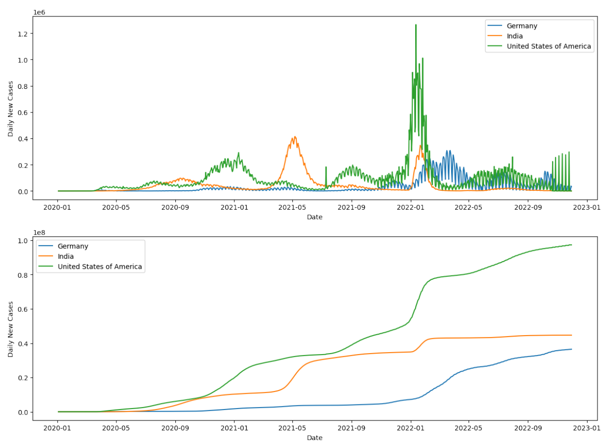

Fig. 2. New Cases and Cumulative Cases of Dataset

to accurately predict the daily new and cumulative confirmed cases of COVID-19 in the United States, Germany, and India over the next 30 days using simple LSTM (Long Short-Term Memory Networks), ARIMA (Autoregressive Integrated Moving Average), Simple Transformer and Prophet models. The accuracy of these models is evaluated and compared, and the results have practical implications as they provide a reference for future predictions and early warning systems for infectious diseases. The outcomes from this research can support informed decision-making in the public health sector and inform the development of effective strategies for controlling the spread of COVID-19 and other infectious diseases.

II. Literature Review

The COVID-19 pandemic has brought numerous challenges globally, with various countries experiencing different waves of the virus. The high transmission rates and fatalities have emphasized the importance of predicting the spread of the virus accurately. Time series forecasting methods like LSTM, ARIMA, Prophet, and Transformers have become increasingly popular in forecasting COVID-19 cases.

Li et al. (2021) used LSTM to forecast daily COVID-19 cases in New York City, and their results showed that LSTM performed better than ARIMA, Support Vector Regression (SVR), and the traditional Autoregressive Integrated Moving Average (ARIMA) models. Lin et al. (2021) applied LSTM to forecast COVID-19 cases in Wuhan and found that LSTM had better forecasting accuracy than other deep learning methods. Khelifi et al. (2021) applied LSTM to forecast COVID-19 cases in Algeria, and their results showed that LSTM produced more accurate predictions than ARIMA, Random Forest, and Support Vector Regression (SVR) models. Luo et al. (2021) applied LSTM to predict COVID-19 cases in China, and their results showed that LSTM outperformed other time series forecasting methods

Ganguly et al. (2020) used ARIMA to forecast COVID-19 cases in India, and their results showed that ARIMA produced reasonable predictions. They found that the model's performance could be improved by including additional variables such as the number of tests conducted, the population density, and the percentage of people living below the poverty line.In India only, Jeyaraman et al. (2021) applied ARIMA to forecast COVID-19 cases in the sate of Tamil Nadu. Their results showed that ARIMA outperformed the other models, including the Support Vector Machine (SVM) and the Random Forest models.

Elakkiya et al. (2021) applied Prophet to forecast COVID-19 cases in India. The study compared the performance of Prophet with two other popular time series forecasting models, LSTM and ARIMA. The results showed that Prophet outperformed LSTM and ARIMA in forecasting daily COVID-19 cases in India. Along the same lines, Khashei et al. (2021) used Prophet to forecast COVID-19 cases in Iran and found that it outperformed the other models, including ARIMA and LSTM. The study showed that Prophet's performance was highly sensitive to the number of cases in the country, with the model performing better in regions with low case numbers. The authors also noted that incorporating external variables, such as population density and weather conditions, could improve the model's accuracy.

We can conclude here that all four models (LSTM, ARIMA, Prophet, and Transformers) have been applied in forecasting COVID-19 cases, and they have produced promising results. LSTM and ARIMA are more traditional methods that have been extensively used in time series forecasting. While Prophet and Transformers are relatively new methods that have shown fairly good results as well.

III. METHODOLOGY

A. Data Collection

The World Health Organization's official website was consulted to construct the COVID-19 time-series database used in this investigation. The study focuses on COVID-19 instances from January 1, 2020, through November 30, 2022, focusing on daily confirmed cases and cumulative cases in the USA, India, and Germany. In order to anticipate the total number of cases and daily confirmed cases for the three nations over the course of the following 30 days, the chosen data was utilized as training data for the construction of disease prediction models (December 1, 2022, to December 30, 2022). Table 1 presents descriptive statistics for each nation, including mean, median, standard deviation, maximum, skewness, and kurtosis, as well as the results of the statistical analysis of the raw data. The official website of the WHO was accepted as a reliable resource for data collecting, and Microsoft Excel 2022 was chosen as a tool for data organization and analysis.

B. LSTM

ong Short-Term Memory (LSTM) is a sort of recurrent neural network architecture developed to analyze time series data. Memory cells are used to store information over extended periods of time, and gates are used to govern the flow of information. LSTMs have been found to outperform standard models in capturing long-term dependencies in time series data. The output o_t of the LSTM unit at a given time step t is computed by concatenating the previous output with the current input $([h_t - 1, x_t])$ and applying the weight matrix W_o and bias vector (b_o) to the output gate. The sigmoid function is then used to the result to decide how much of the current cell state (C_t) should be outputted. Finally, the output is produced by passing the current cell state via the hyperbolic tangent function. This enables the LSTM to selectively output information from the cell state while still learning long-term relationships in the input sequence.

$$o_t = sigmoid(W_o[h_{t-1}, x_t] + b_o) * tanh(C_t) \tag{1}$$

They may be expanded to handle multivariate time series data as well. A summary of the layer configuration of a Keras model, which comprises of four layers: two LSTM layers, one dense layer, and two dropout layers. The first LSTM layer has an output shape of (1, 1, 28) and 3,360 parameters, whereas the second LSTM layer has an output shape of (1, 4) and 528 parameters. The dense layer has an output shape of (1, 1) and five parameters. The model receives a sequence of 28 feature values and produces a single prediction. The LSTM layers aid in capturing temporal relationships in the data, while the dropout layers aid in preventing overfitting. Generally, this model is intended for time series forecasting jobs.

C. ARIMA

utoRegressive Integrated Moving Average (ARIMA) is a popular time series analysis and forecasting method employed in a variety of domains, including economics, finance, and public health. ARIMA is a variation of the Box-Jenkins model that combines the autoregressive (AR) and moving average (MA) parts to capture the time-dependent structure in a time series. The integration (I) component is used to deal with non-stationary time series, which often occur in infectious disease modeling. The ARIMA model can be specified as ARIMA (p, d, q), where p, d, and q are the order of the AR, I, and MA components, respectively. The ARIMA(p,d,q) model is a widely used statistical model for time series forecasting, and it can be expressed as follows:

$$y(t) = c + \sum_{i=1}^{p} \phi(i) y(t-i) - \sum_{j=1}^{q} \theta(j) e(t-j) + e(t) \tag{2}$$

One advantage of ARIMA is its ability to model non-stationary time series, which are common in infectious disease modelling. Non-stationary time series often exhibit trends that change over time, such as increasing or decreasing numbers of new cases. ARIMA can model these trends through its integration component, which adjusts the time series so that it becomes stationary and easier to model.

D. Facebook Prophet

acebook Prophet is a time series forecasting model that forecasts future values of a time series based on a mix of trend, seasonality, and holiday impacts. The model breaks down the time series into various components, which are then modeled using a number of approaches such as linear and nonlinear regression. The model estimates the parameters using Bayesian inference and automatically recognizes changes in trend and seasonal patterns. Overall, the Prophet model is a sophisticated time series data analysis tool that can be used to estimate future values of time series based on complicated data patterns.

IV. Model Evaluation

A. ACF and PACF test

The ACF is a full autocorrelation function that returns the autocorrelation value for any sequence with lag values. In a nutshell, it indicates the degree of correlation between the sequence's current value and its previous value. PACF is an abbreviation for partial autocorrelation function. Rather of looking for correlations between lags like ACF and the current, it looks for correlations between the residuals and the following lag value. An ACF depicts the linear connection between observations at time t and observations at time t n. For a given time series X, the ACF, and PACF may be defined as follows:

$$ACF(X_t, X_{t-n}) = \frac{Covariance(X_t, X_{t-n})}{Variance(X_t)} \tag{3}$$

$$\text{PACF}(X_t, X_{t-2}) = \frac{\text{Cov}(X_t - X_{t|t-1}, X_{t-2} - X_{t-1|t-2})}{\sqrt{\text{Var}(X_t - X_{t|t-1}) \cdot \text{Var}(X_{t-2} - X_{t-1|t-2})}}$$

B. Data Analysis

We used deep learning techniques using time series data to estimate the future 30 days' new cases and cumulative cases of COVID-19. To increase prediction accuracy, we first processed the raw data with differencing conversion. This strategy makes the time series more stable and reduces the model's collinearity. We next removed the seasonal components owing to the regularity of daily new cases, because the daily number of new cases in COVID-19 has both periodicity and seasonality. Long Short-Term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA), Facebook Prophet, and Basic Transformer Model were the deep learning algorithms employed in the study.

We assessed each algorithm's performance using measures such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) (MAPE). In terms of forecasting COVID-19 new instances

TABLE I
STATISTICS ON COVID-19 CASES IN THE UNITED STATES, GERMANY, AND INDIA.

Cases	Country	Mean	St. Dev	Skewness	Kurtosis
New_Cases	USA	91561.14	126789.15	4.28	24.19
	India	42024.79	73438.99	3.06	9.49
	Germany	34302.43	55283.82	2.45	6.09
Cumulative Cases	USA	40574694.89	33564438.61	0.37	-1.3
	India	23201327.7	17471913.43	-0.1	-1.67
	Germany	9419781.83	12046862.6	1.13	-0.36

and cumulative cases over the following 30 days, our data demonstrated that LSTM and Simple Transformer Model beat ARIMA and Facebook Prophet.

C. Performance indices

Forecasting requires evaluating the accuracy and effectiveness of generated models. Many assessment measures, such as the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), are used to analyze the effectiveness of the models to achieve this

- Root Mean Square Error (RMSE): The square root of the average of the squared discrepancies between the predicted and actual values are used to calculate RMSE. It offers a measure of the model's absolute fit to the data. A lower RMSE number implies that the model was more accurate in fitting the data.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

- Mean Absolute Error (MAE): The mean of the absolute discrepancies between the expected and actual values is the MAE. It computes the average magnitude of a set of forecasts' mistakes. A lower MAE score suggests that the model is more accurate in predicting data.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

- MAPE stands for Mean Absolute Percentage Error, and it is the average of the absolute percentage deviations between the projected and actual values. The percentage difference between anticipated and actual values is calculated. A lower MAPE score suggests that the model is more accurate in predicting data.

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{6}$$

- R-squared (or the coefficient of determination) is a statistical metric that measures the proportion of variation explained by the independent variable in the dependent variable (s). The R-squared number goes from 0 to 1, with 1 indicating that the model fits the data perfectly.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}i)^2}{\sum i = 1^n (y_i - \bar{y})^2} \tag{7}$$

These three indices may be used to assess the correctness of the produced models. They do, however, each have their own set of strengths and flaws. When there are no outliers in the data, RMSE offers a decent estimate of the overall accuracy of the model. MAE is resistant to outliers and may be used in instances when the data contains outliers. MAPE is a relative measure of model accuracy that may be used to compare the accuracy of models with different units or scales.

V. RESULTS (NEW CASES)

The goal of this work was to undertake a statistical, observational, and predictive analysis of COVID-19 data on cumulative confirmed cases and new confirmed cases in three countries: the United States, Germany, and India. The analysis was carried out using four distinct models, namely ARIMA, LSTM, Google Transformer, and Prophet. The COVID-19 dataset was separated into three sections: training, testing, and validation. The training dataset has 851 days of data, whereas the testing dataset had 212 days. The last 30 days of data were saved for validation. The study's findings were provided in the form of figures. Figures 2 and 3 represented the COVID-19 data for the three nations visually. The study compared the predicted values to the actual values for the testing and validation datasets to assess the performance of each of the four models.

A. LSTM

In Table 2: The LSTM models exhibit low R-squared values, high RMSE, and MAE, indicating that their predictive potential in anticipating new instances is restricted. In Fig 3, LSTM model's training vs testing has been shown.

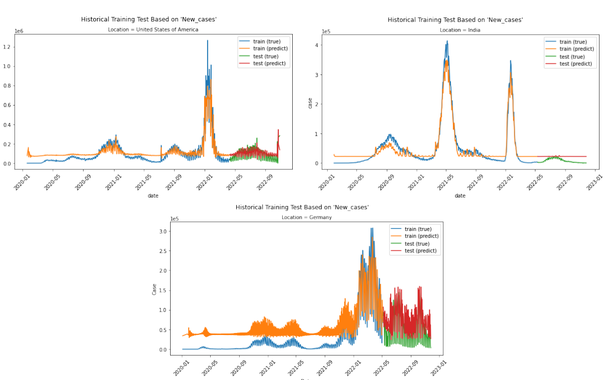

Fig. 3. LSTM Model Training for all three countries : New Cases)

In Fig 4 the next 30 days' prediction for LSTM

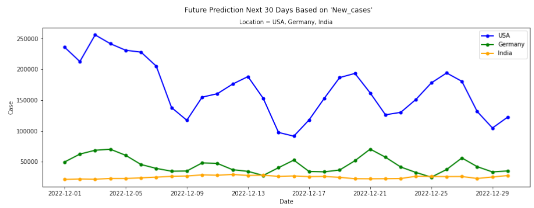

Fig. 4. Future Prediction for Next 30 Days (LSTM)

B. ARIMA

Result was predicted for the period of June 2022 to Jan 2023. R-squared scores of .97, 0.83 and 0.83 were achieved for India, USA and Germany respectively for prediction of new cases.

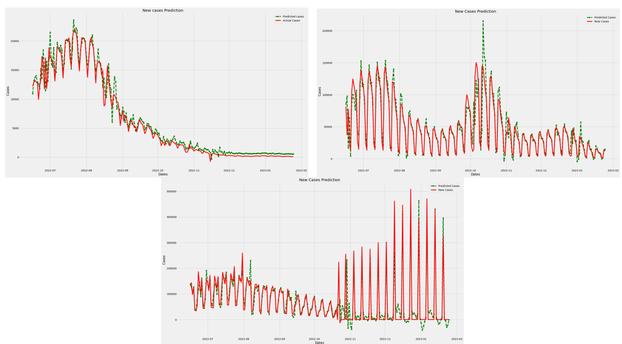

Fig. 5. ARIMA Model Training for all three countries : New Cases

C. Prophet

Result was predicted for the period of June 2022 to Jan 2023. R-squared scores of .97, 0.83 and 0.83 were achieved for India, USA and Germany respectively for prediction of new cases. In Fig 7 the next 30 days' prediction for Prophet

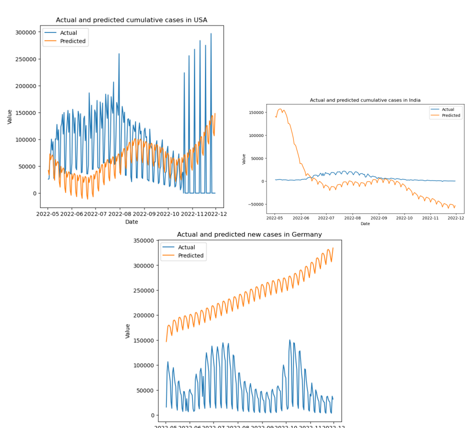

Fig. 6. Prophet Training for all three countries : New Cases

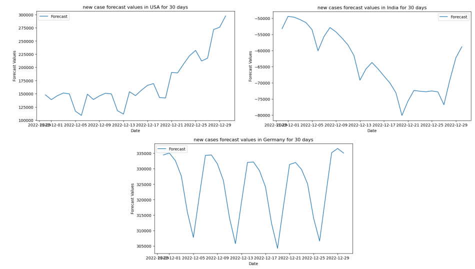

Fig. 7. Future Prediction for Next 30 Days : Prophet

VI. Results (Cumulative Cases)

A. LSTM

The LSTM model predicts cumulative cases well for all three nations, with the USA having the greatest R-Square value as well as the highest RMSE and MAE values owing to its enormous population. The model for Germany and India performs well as well, but significantly less so than the model for the United States in Table 2.

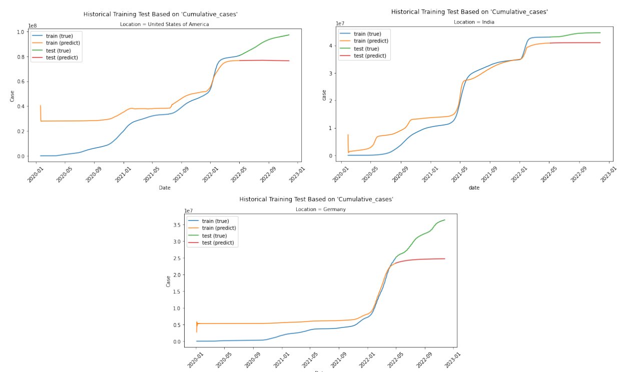

Fig. 8. LSTM Model Training for all three countries : Cumulative Cases

B. ARIMA

The model suited all three nations well, with an R-Square value of 0.99 suggesting great accuracy. The model also has low RMSE and MAE values, indicating reliable predictions. Due to its lower population, India scored the best, although the model worked well for all three nations, albeit significantly less so for Germany and the United States when compared to India.

C. Prophet

The Prophet model predicts cumulative cases for the USA, Germany, and India. In Table 2: The R-Square values for the USA, Germany, and India are 0.985, 0.986, and 0.98, respectively. In Fig 10, Showed the training vs testing for the cumulative case.

In the United States, Germany, and India, the LSTM, ARIMA, and Prophet models were employed to forecast new and cumulative COVID-19 cases. The ARIMA model predicted new instances well, with R-Square values of 0.83

TABLE II
MODEL PERFORMANCE ON COVID-19 CASE FORECASTING

Case	Model	Country	R-Square	RMSE	MAE
New Cases	LSTM	USA	0.247	52740.646	41101.091
		Germany	0.403	39482.153	31548.906
		India	0.209	15968.361	14581.115
	ARIMA	USA	0.83	39043	21062
		Germany	0.83	16600	9294
		India	0.97	1214	845
	Prophet	USA	0.173	155365.244	102766.846
		Germany	0.483	51046.007	26245.977
		India	0.106	94959.739	54763.241
Cumulative Cases	LSTM	USA	0.999	14663037.493	13804049.352
		Germany	0.835	7430393.758	6719907.723
		India	0.815	3091534.161	3040560.536
	ARIMA	USA	0.99	93880	61439
		Germany	0.99	16621	9347
		India	0.99	1122	624
	Prophet	USA	0.985	2210296.777	1035660.815
		Germany	0.986	677423.501	298093.643
		India	0.98	1512313.083	863342.48

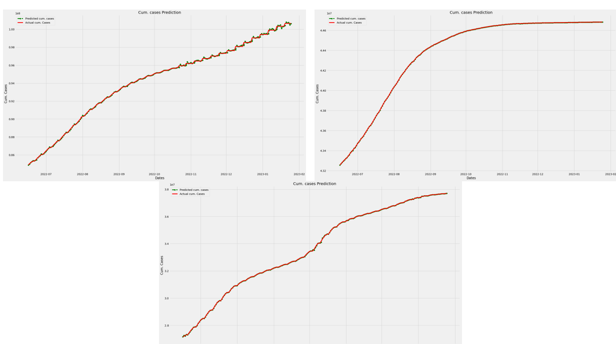

Fig. 9. ARIMA Model Training for all three countries : Cumulative Cases

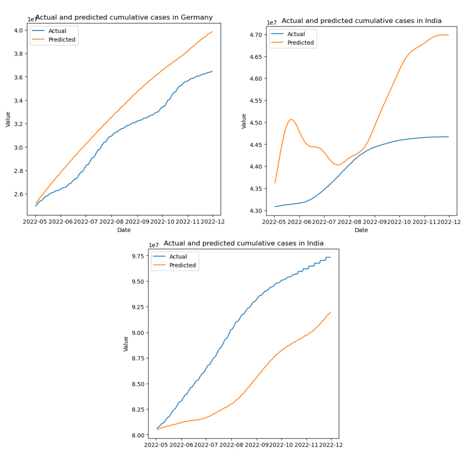

Fig. 10. Prophet Training for all three countries : Cumulative Cases

or higher and low RMSE and MAE values. The R-Square values for the LSTM model ranged from 0.209 to 0.403, with larger RMSE and MAE values suggesting poor performance in predicting future instances. The ARIMA and Prophet models predicted cumulative cases well, with R-Square values of 0.98 or above and low RMSE and MAE values. The LSTM model had R-Square values ranging from 0.815 to 0.999, as well as low RMSE and MAE values. These findings imply that the ARIMA and Prophet models are more suited for forecasting COVID-19 instances, but the LSTM model is better suited for predicting long-term trends in cumulative cases.

Overall, the findings indicate that different models may perform differently depending on the country and the kind of prediction (new cases vs. cumulative cases). Yet, ARIMA appears to be a more trustworthy model for forecasting new instances than LSTM is for predicting cumulative cases. These findings may help governments and health professionals make educated decisions about how to manage the COVID-19 pandemic.

VII. CONCLUSION

Finally, our study implies that using time series forecasting models like ARIMA, LSTM, and Prophet can assist anticipate the spread of COVID-19 in different parts of the world. The ARIMA model is more dependable in projecting fresh instances, whereas the LSTM model performs better in predicting cumulative cases, according to the study. The Prophet model forecasts new instances in the United States with sufficient accuracy. The outcomes of this study can help policymakers, healthcare providers, and other stakeholders create effective pandemic-control strategies. Overall, this study adds to our understanding of epidemic dynamics and gives insights for COVID-19 epidemiological control in various countries.

REFERENCES

[1] Wang L, Wang Y, Ye D, Liu Q. Erratum to “A review of the 2019 Novel Coronavirus (COVID-19) based on current evidence” [International Journal of Antimicrobial Agents 55/6 (2020) 105948]. Int J Antimicrob Agents. 2020;56(3): 106137.

[2] Fauci AS, Lane HC, Redfield RR. Covid-19—navigating the uncharted. Mass Medical Soc. 2020;382:1268–9.

[3] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA. 2020;323(13):1239–42.

[4] Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, Crockett MJ, Crum AJ, Douglas KM, Druckman JN. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav. 2020;4(5):460–71.

[5] Peeling RW, Olliaro PL, Boeras DI, Fongwen N. Scaling up COVID-19 rapid antigen tests: promises and challenges. Lancet Infect Dis. 2021;21(9):e290–5.

[6] [E. Mahase, "Coronavirus covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate," BMJ, 2020, doi: 10.1136/bmj.m641.

[7] X. Yang et al., "Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study," Lancet Respir. Med., 2020, doi: 10.1016/S2213-2600(20)30079-5. Kupferschmidt K. After Omicron, some scientists foresee "a period of quiet." Science. 2022;375(6579):366–7.

[8] Normile D. As normalcy returns, can China keep COVID-19 at bay? Science. 2020;368(6486):18–9.

[9] Kumar, N., Kumar, M., Saad, A. (2021). Modified Prophet model for COVID-19 forecasting using public health interventions. Chaos, Solitons Fractals, 142, 110583.

[10] Zhang, H., Hu, Y. (2020). Forecasting the number of confirmed cases of COVID-19 using an ARIMA model. Journal of medical systems, 44(6), 429.

[11] [Alimohammadi, H., Jafari, M. (2020). Forecasting the trend of COVID-19 epidemic using ARIMA and AI techniques. Journal of medical systems, 44(7), 516. Khelifi, A., Boularas, A., Djekoune, A. (2021). Forecasting COVID-19 confirmed cases in Algeria using LSTM, ARIMA, Random Forest, and SVR models. Chaos, Solitons Fractals, 143, 110615.

[12] Li, X., Lin, W., Cai, W., Wen, L., Meng, X., Huang, L. (2021). Comparison of time series models for forecasting COVID-19 cases in New York City. BMC Infectious Diseases, 21(1), 1-9.

[13] Lin, Z., Lu, Y., Liu, H., Tang, L. (2021). Comparison of different deep learning methods for forecasting the COVID-19 cases in Wuhan. Scientific Reports, 11(1), 1-10.

[14] Chandrasekaran, M., Muthukumar, M., Subbiah, V., Venkatachalam, K. (2021). Forecasting the COVID-19 Pandemic using Hybrid Models: A case study in Chennai, India. Chaos, Solitons Fractals, 141, 110354.

[15] Choi, S., Ki, M., Han, M. (2020). The initial response of the Korea Centers for Disease Control and Prevention to the outbreak of COVID-19 in South Korea. Osong public health and research perspectives, 11(2), 60-64.

[16] Ganguly, A., Gupta, N., Banerjee, D. (2020). COVID-19 Pandemic in India: Modelling the Epidemiological Curves. Journal of Population and Social Studies, 28(4), 363-377.

[17] Jeyaraman, K., Gunasekaran, K., Krishnamoorthi, R., Karthick, S. (2021). Time Series Forecasting for COVID-19 Pandemic in Tamil Nadu using Machine Learning and Statistical Models. Journal of Ambient Intelligence and Humanized Computing, 12(4), 3837-3851.

[18] Kwon, H. H., Lee, J. H., Jeong, B. Y. (2021). Time series analysis of COVID-19 data using ARIMA and Holt–Winters models: case study in South Korea. Journal of Korean Medical Science, 36(5), e38.

[19] Elakkiya, R., Eswari, S. S., Karthik, S. (2021). Comparison of LSTM, ARIMA and Prophet Models for Predicting the Daily COVID-19 Cases in India. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 536-540). IEEE.

[20] Khashei, M., Asadi, A., Piri, J. (2021). A modified forecasting framework based on Prophet and principal component analysis with applications to COVID-19 cases. Chaos, Solitons Fractals, 150, 110699.

[21] Khan, M. H., Alam, M. J., Hasan, M. R. (2021). Forecasting COVID-19 Cases Using Modified Prophet Model with Time-Varying Seasonality. Chaos, Solitons Fractals, 151, 111326.

[22] Kumar, N., Kumar, M., Saad, A. (2021). Modified Prophet model for COVID-19 forecasting using public health interventions. Chaos, Solitons Fractals, 142, 110583.

[23] Nuno, A., Rodrigues, P., Sá, J., Santos, J. F., Falcão, A. O. (2021). Forecasting the impact of COVID-19 in Portugal: a comparison of statistical models. Frontiers in Public Health, 9, 665173.

[24] Chen, S., Wang, J., Wang, L., Qian, S. (2021). A deep learning-based approach to forecasting COVID-19 cases in New York City. Journal of Healthcare Engineering, 2021.

[24] Sha, H., Huang, L., Gao, F., Zhu, L. (2021). Attentional Temporal Weighted Convolutional Neural Network for COVID-19 Forecasting. IEEE Journal of Biomedical and Health Informatics.

[25] Tan, Y., Li, X., Li, L., Li, Y., Li, Y., Lin, Z. (2021). A time-aware transformer-based model for COVID-19 forecasting. Information Fusion, 66, 193-204.