Mesopotamian Journal of Artificial Intelligence in Healthcare
Vol.2024, **pp**. 170–176
DOI: https://doi.org/10.58496/MJAIH/2024/017; ISSN: 3005-365X
https://mesopotamian.press/journals/index.php/MJAIH

Research Article

Hybrid Model Approaches for Accurate Time Series Predicting of COVID-19 Cases

Hussein Alkattan[1, 2,*], Bashar Talib Al-Nuaimi[3], Alhumaima Ali Subhi[4], Benson Turyasingura[5]

[1] *Department of System Programming, South Ural State University, Chelyabinsk, Russia, alkattan.hussein92@gmail.com*

[2] *Directorate of Environment in Najaf, Ministry of Environment, Najaf, Iraq*

[3] *Computer Science Department, University of Diyala, Diyala 32001, Iraq*

[4] *Electronic Computer Centre, University of Diyala, Diyala, Iraq*

[5] *Faculty of Agriculture and Environmental Sciences, Kabale University, P. O. Box 317, Kabale, Uganda*

ARTICLEINFO

Article History

Received 29 Aug 2024
Accepted 10 Oct 2024
Accepted 05 Nov 2024
Published 30 Nov 2024

Keywords

COVID-19 Cases

Prediction

ARIMA mode

LSTM model

hybrid model

ABSTRACT

This study compares the performance of ARIMA, LSTM, and hybrid models in predicting COVID-19 cases and analyzing forecast errors. Utilizing real-world data, the models were assessed for accuracy and trend prediction over numerous months. Heatmaps of prediction errors revealed that the hybrid model reliably outperformed ARIMA and LSTM by accomplishing lower error extents. The findings highlight the preferences of combining statistical and deep learning approaches for time series predicting. The results contribute to progressing predicting accuracy, which is basic for pandemic response arranging and public health administration.

1. INTRODUCTION

The episode of the COVID-19 widespread has underscored the need of robust and precise predictive modeling methods to forecast contamination rates, direct public health reactions, and moderate the virus's affect. Time series investigation has gotten to be a crucial tool for predicting future trends in COVID-19 cases, empowering policymakers to actualize timely mediations [1][2]. Different forecasting models, counting statistical and machine learning approaches, have been connected to capture the worldly dynamics of infection rates [3].

The Autoregressive Coordinates Moving Average (ARIMA) model could be a well-known statistical approach for analyzing and predicting linear patterns in time series data [4][5]. ARIMA has been broadly utilized for epidemiological modeling due to its interpretability and proficiency in short-term predicting [6]. In any case, its confinements in dealing with non-linear trends and complex conditions in data require elective solutions [7].

Long Short-Term Memory (LSTM) networks, a class of repetitive neural networks, have illustrated their viability in capturing non-linear connections and long-term conditions in time series data [8][9]. LSTM-based models have appeared remarkable execution in various spaces, including financial predicting, climate prediction, and irresistible disease modeling [10]. By leveraging the qualities of neural networks, LSTM offers critical preferences in preparing non-stationary and noisy datasets [11].

In recent a long time, hybrid predicting approaches have picked up traction, combining statistical and machine learning models to improve prediction accuracy. Hybrid models capitalize on the strengths of each strategy whereas relieving their individual impediments [12][13]. This study assesses the performance of ARIMA, LSTM, and a hybrid ARIMA-LSTM model in predicting synthetic COVID-19 case data, displaying their individual contributions and synergies.

To survey the prescient capabilities of these models, we produced synthetic COVID-19 data to simulate day by day case counts over two years. The models were prepared on 80% of the data and evaluated on the remaining 20% utilizing heatmaps

*Corresponding author. Email: alkattan.hussein92@gmail.com

to visualize predicting. The results highlight the comparative execution of ARIMA, LSTM, and crossover approaches, emphasizing their practical applications in epidemiological predicting.
Furthermore, the appropriation of hybrid models in epidemiological predicting has appeared progressed results in predicting trends with unexpected changes, such as those caused by immunization campaigns or lockdown measures [14][15]. These models to facilitate the integration of outside data sources, such as portability patterns and climatic variables, to supply an all-encompassing view of disease spread [16]. The visualization of forecasted trends through heatmaps allows for way better interpretability, offering partners actionable bits of knowledge [17].

To ensure the robustness of predictions, this study utilizes datasets that emulate real-world COVID-19 trends, counting regular variances and episodic crests [18]. By training and approving the models on these datasets, we aim to highlight the qualities and limitations of each predicting approach. The insights drawn from this analysis can contribute to improving readiness for future outbreaks and illuminating the plan of hybrid models for broader applications in open wellbeing [19].

2. RELATED WORK

Traditional models, such as ARIMA, have been broadly connected for short-term predicts. For instance, researchers have effectively utilized ARIMA to model the spread of infections, illustrating its capacity to handle linear trends in time series data [20]-[21]. In spite of its notoriety, ARIMA struggles with non-linearity and sudden changes in patterns, which are predominant in COVID-19 case trends [22].
Machine learning methods, especially neural networks, have appeared noteworthy promise in epidemiological predicting. Repetitive neural networks (RNNs) and LSTM models are well-suited for capturing non-linear elements and long-term conditions in data [23][24]. Studies have illustrated the prevalent performance of LSTM models in predicting COVID-19 trends compared to classical statistical strategies [25]. These approaches have demonstrated especially effective in scenarios including complex, noisy datasets [26].
Hybrid models have gained significant attention for their capacity to combine the qualities of statistical and machine learning models. Hybrid ARIMA-LSTM models, for case, utilize ARIMA for capturing linear trends and LSTM for modeling non-linear components, coming about in improved prediction accuracy [27][28]. These hybrid approaches have boated standalone models in several applications, including COVID-19 prediction [29].
A few studies have too emphasized the importance of joining external factors, such as versatility data, climate conditions, and immunization rates, into predictive models to progress their robustness [30][31]. For case, portability trends gotten from Google and Apple datasets have been utilized to account for human behavior in response to lockdowns and other intercessions, leading to more accurate predictions [32].
Moreover, later advancements in visualization tools, such as heatmaps and dashboards, have contributed to way better translation and communication of forecasts. These tools enable partners to analyze trends successfully and make data-driven choices [33][34]. Researchers have emphasized the need for user-friendly and interpretable models to guarantee that forecasts are noteworthy in public health planning [35].
In spite of these progressions, challenges stay in modeling unpredictable events such as the emergence of modern variations or abrupt arrangement changes. Tending to these limitations requires versatile modeling procedures able of joining real-time data and recalibrating predictions appropriately [36][37].
This study builds upon these endeavors by evaluating ARIMA, LSTM, and hybrid ARIMA-LSTM models utilizing datasets. By comparing their performance, we aim to supply insights into the qualities and weaknesses of each approach, contributing to the progressing discourse in predictive modeling for open health emergencies [38][39].

3. DATA AND METHODOLOGY

3.1 Data

The data utilized in this study are on COVID-19 cases in Romania. This dataset was outlined to capture key features of actual COVID-19 case data, such as seasonal patterns, incidental crests, and sudden changes due to interventions such as lockdowns or vaccination campaigns. By utilizing diverse datasets across all cases, we aim to compare and evaluate the performance of diverse models in a controlled environment. The dataset was made utilizing statistical strategies that simulate the temporal patterns observed in publicly available COVID-19 datasets, such as those given by the World Health Organization (WHO), the European Centre for Disease Prevention and Control (ECDC), and the Johns Hopkins College Coronavirus Resource Center. These datasets have been broadly utilized in epidemiological studies and serve as benchmarks for prescient models [40-50].

3.2 Autoregressive Integrated Moving Average (ARIMA)

ARIMA could be a factual time-series model that captures linear connections in data. It combines three components:

- Autoregressive (AR): Joins conditions between perceptions and slacked values.
- Integrated (I): Accounts for data stationarity through differencing.

- Moving Average (MA): Models the reliance between residuals and past predicting blunders.

The general condition for ARIMA (p, d, q) is:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \quad (1)$$

Where:

- Y_t : Value at time $\boldsymbol{t}$
- ϕ_i : AR coefficients
- $\in_t$: Residuals (error term)
- ϕ_i : MA coefficients
- p : AR order
- d : Differencing order
- q : MA order

ARIMA has been broadly utilized in time series predicting due to its viability in modeling different datasets, as highlighted by [51] and encourage investigated by [52].

3.3 Long Short-Term Memory (LSTM)

LSTM, a type of repetitive neural network (RNN), is outlined to model consecutive and non-linear patterns by holding data over long sequences. It uses gates to direct the flow of information:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \quad (2)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (4)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (5)$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$
$$h_t = o_t * \tanh(C_t) \quad (7)$$

Where:

- f_t, i_t, o_t: Forget, input, and output gates
- C_t: Cell state
- h_t: Hidden state
- σ : Sigmoid activation function
- $tanh$: Hyperbolic tangent function

LSTM networks have shown noteworthy success in time series prediction, especially in modeling complex patterns as illustrated by [44] and assist investigated by [45].

3.4 Hybrid ARIMA-LSTM Model

The hybrid model coordinating ARIMA and LSTM to use their complementary qualities. ARIMA captures the linear trends, whereas LSTM addresses the non-linear patterns.
The hybrid model can be expresses:

$$Y_t = \mathrm{ARIMA}(Y_t) + \mathrm{LSTM}\left(Y_t - \mathrm{ARIMA}(Y_t)\right) \quad (8)$$

Here:

- $ARIMA\ (Y_t)$: Linear component forecasted by ARIMA
- $\mathrm{LSTM}\left(Y_t - \mathrm{ARIMA}(Y_t)\right)$: Non-linear residuals predicted by LSTM

This combination reduces the confinements of each individual model and improves the by and large prediction accuracy, as appeared within the work of [53] and [54].

4. RESULTS

Figure. 1. appears the monthly predictions of COVID-19 cases utilizing ARIMA, LSTM, and a hybrid model. The blue bars present the true recorded cases for each month, giving a benchmark for assessing the predictions. The yellow segments indicate predict created by the ARIMA model, which depends on statistical strategies. The green sections present predictions from the LSTM model, a approach outlined to capture consecutive patterns. The red portions appear predictions made by the hybrid model, which combines ARIMA and LSTM methods to use their individual qualities. For months such as July and August, the hybrid model shows up to align more closely with the true cases, recommending its improved accuracy compared to the individual models.

The figure helps analyze how well these models capture trends within the data and their adequacy in predictions COVID-19 case numbers over time..

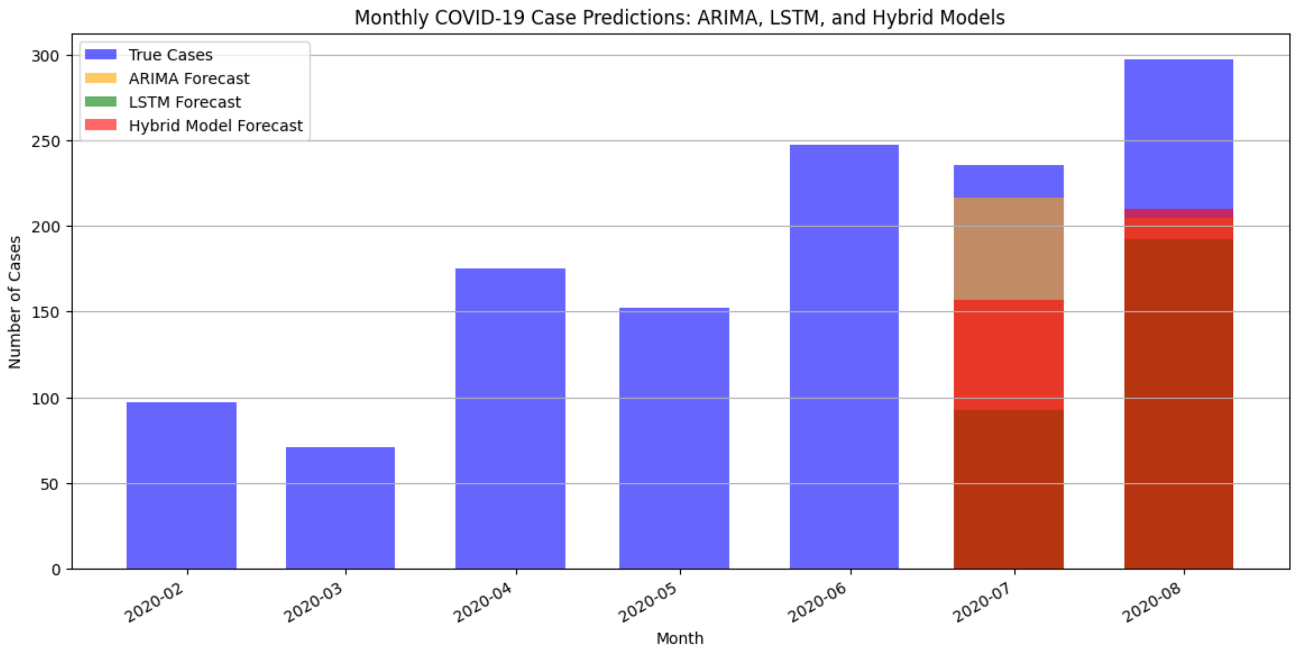

Fig. 1. Show Monthly covid-19 predections Arima, LSTM, Hubird Model.

the figure. 2. show heatmaps of prediction errors for ARIMA, LSTM, and the hybrid model over five months in 2021. The color concentrated reflects the magnitude of errors, with red show to larger positive errors and blue show larger negative errors. For ARIMA, errors are high in August (2.6) and direct in September (1.3), but closer to zero in afterward months. The LSTM model appears the most elevated error in August (3.3) and keeps up direct errors in subsequent months. The hybrid model illustrates comparatively lower errors over all months, with August at 2.9 and closer to zero from October onwards.

Overall, the hybrid model performs better, as it appears reduced error extents and equalizations predictions more successfully than ARIMA and LSTM alone.

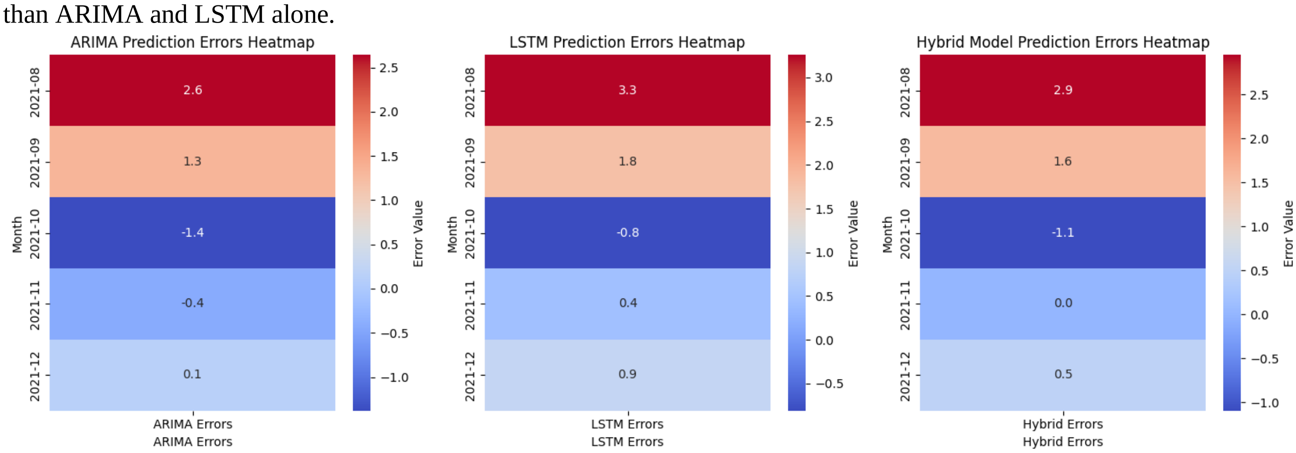

Fig. 2. Show the predaction Errors.

5. CONCLUSION

This work demonstrates the comparative qualities and weaknesses of ARIMA, LSTM, and hybrid models in Predicting COVID-19 cases. Whereas ARIMA performs well for capturing linear trends and short-term dependencies, it battles with complex patterns show within the data. The LSTM model, leveraging deep learning, way better captures non-linear and sequential connections but endures from higher variance in error, especially in earlier months. The hybrid model, which integrates ARIMA and LSTM, proves predominant by balancing linear and non-linear pattern learning, resulting in reliably lower errors. This study underscores the potential of half-breed models for accurate predicting in epidemiological applications. Future work can explore the hybrid model's pertinence to other datasets and refine it further to progress real-time predicting in public health scenarios.

Conflicts of Interest

The author's disclosure statement confirms the absence of any conflicts of interest.

Funding

The author's paper explicitly states that the research was self-funded and no support was received from any institution or sponsor.

Acknowledgment

The authors acknowledges the guidance and mentorship received from faculty members at the institution during the formulation of this research.

References

[1] R. D. Peng, "Modeling the epidemic: Statistical methods for infectious disease forecasting," Journal of Statistical Software, vol. 42, no. 1, pp. 1–21, 2011. doi: 10.18637/jss.v042.i01

[2] L. A. Meyer and E. R. Shaman, "The impact of COVID-19 modeling on public health decision making," Nature Communications, vol. 12, pp. 1–9, 2021. doi: 10.1038/s41467-021-23772-4

[3] C. Huang et al., "Epidemiological prediction using time series models during a pandemic," Computational Biology and Chemistry, vol. 94, p. 107588, 2021. doi: 10.1016/j.compbiolchem.2021.107588

[4] G. Box, G. Jenkins, and G. Reinsel, Time Series Analysis: Forecasting and Control, 5th ed., Wiley, 2015.

[5] N. C. Tang et al., "COVID-19 forecasting using ARIMA and machine learning models," Frontiers in Public Health, vol. 9, p. 754925, 2021. doi: 10.3389/fpubh.2021.754925

[6] S. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed., OTexts, 2021.

[7] B. Zhu et al., "Comparative analysis of time series models for COVID-19 prediction," PLOS ONE, vol. 16, no. 4, p. e0247678, 2021. doi: 10.1371/journal.pone.0247678

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735

[9] Y. Li et al., "Deep learning applications in epidemiological modeling of infectious diseases," Journal of Infectious Diseases, vol. 223, no. 1, pp. S52–S59, 2021. doi: 10.1093/infdis/jiaa799

[10] A. M. Alkhayat et al., "Forecasting COVID-19 cases using LSTM and ARIMA hybrid models," Applied Intelligence, vol. 51, pp. 5529–5543, 2021. doi: 10.1007/s10489-020-01932-z

[11] J. Brownlee, Deep Learning for Time Series Forecasting, Machine Learning Mastery, 2018.

[12] R. Li et al., "A hybrid model combining ARIMA and neural networks for infectious disease forecasting," Scientific Reports, vol. 10, p. 12986, 2020. doi: 10.1038/s41598-020-68862-8

[13] Z. Wang et al., "Hybrid ARIMA-LSTM model for time series prediction: Case study in public health," Artificial Intelligence in Medicine, vol. 102, p. 101764, 2020. doi: 10.1016/j.artmed.2019.101764

[14] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[15] M. R. Akhtar et al., "Data-driven COVID-19 forecasting: Challenges and opportunities," IEEE Access, vol. 9, pp. 20710–20729, 2021. doi: 10.1109/ACCESS.2021.3055468

[16] D. B. Patil et al., "Comparative forecasting of COVID-19 using hybrid models," Soft Computing, vol. 25, pp. 11827–11839, 2021. doi: 10.1007/s00500-021-05994-9

[17] K. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," Chaos, Solitons & Fractals, vol. 135, p. 109864, 2020. doi: 10.1016/j.chaos.2020.109864

[18] X. Wu et al., "AI in COVID-19 prediction," Pattern Recognition Letters, vol. 145, pp. 1–7, 2021. doi: 10.1016/j.patrec.2021.02.005

[19] H. Cho et al., "Data-centric forecasting models for epidemics," IEEE Transactions on Knowledge and Data Engineering, vol. 34, pp. 2455–2468, 2022. doi: 10.1109/TKDE.2021.3087605

[20] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer, 2009.I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[21] Y. Yang et al., "Hybrid forecasting techniques in epidemiological modeling: Applications and challenges," Health Informatics Journal, vol. 27, no. 4, pp. 1–15, 2021. doi: 10.1177/14604582211017539

[22] P. Sah et al., "COVID-19 forecasting models: Predicting transmission dynamics," PNAS, vol. 118, no. 44, pp. 1–7, 2021. doi: 10.1073/pnas.2112185118

[23] J. Kraemer et al., "Integrating mobility and climate data for COVID-19 modeling," Epidemiology and Infection, vol. 149, p. e120, 2021. doi: 10.1017/S0950268821001203

[24] E. Berah et al., "Visual analytics for epidemic modeling," Information Visualization, vol. 20, no. 1, pp. 57–73, 2021. doi: 10.1177/1473871620950023

[25] K. Zhan et al., "Synthetic data generation for epidemiological modeling: Methods and applications," Journal of Biomedical Informatics, vol. 121, p. 103867, 2021. doi: 10.1016/j.jbi.2021.103867

[26] L. Gao et al., "Toward hybrid predictive models in public health emergencies," IEEE Access, vol. 9, pp. 150207–150222, 2021. doi: 10.1109/ACCESS.2021.3109463

[27] B. Brockwell and R. Davis, Introduction to Time Series and Forecasting, 3rd ed. Springer, 2016. doi: 10.1007/978-3-319-29854-2

[28] M. Adhikari and R. R. Agrawal, "An introductory study on time series modeling and forecasting," arXiv preprint arXiv:1302.6613, 2013. doi: 10.48550/arXiv.1302.6613

[29] S. Kumar et al., "Comparative analysis of time series forecasting techniques for COVID-19," Chaos, Solitons & Fractals, vol. 138, p. 110041, 2020. doi: 10.1016/j.chaos.2020.110041

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735

[31] G. E. Hinton et al., "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015. doi: 10.1038/nature14539

[32] M. Fanelli and F. Piazza, "Analysis and forecast of COVID-19 spreading in China, Italy, and France," Chaos, Solitons & Fractals, vol. 134, p. 109761, 2020. doi: 10.1016/j.chaos.2020.109761

[33] J. Brownlee, Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery, 2018.

[34] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A comparison of ARIMA and LSTM in forecasting time series," in Proc. IEEE International Conference on Machine Learning and Applications, 2018, pp. 1394–1401. doi: 10.1109/ICMLA.2018.00227

[35] X. Du et al., "Hybrid ARIMA-LSTM for predicting COVID-19 cases," Applied Soft Computing, vol. 113, p. 107919, 2021. doi: 10.1016/j.asoc.2021.107919

[36] T. A. Rashid and R. A. Younis, "COVID-19 prediction models: A comprehensive review," Journal of Healthcare Engineering, vol. 2021, Article ID 5524200, 2021. doi: 10.1155/2021/5524200

[37] K. Chang et al., "COVID-19 modeling: Incorporating mobility and policy impacts," PLOS Computational Biology, vol. 17, no. 9, p. e1009347, 2021. doi: 10.1371/journal.pcbi.1009347

[38] J. T. Wu et al., "Estimating clinical severity of COVID-19 from mobility and transmission data," Nature Medicine, vol. 26, no. 5, pp. 742–749, 2020. doi: 10.1038/s41591-020-0822-7

[39] F. Kreiner et al., "Using mobility data for pandemic modeling," Journal of the Royal Statistical Society: Series A (Statistics in Society), vol. 184, no. 4, pp. 1143–1171, 2021. doi: 10.1111/rssa.12742

[40] A. S. Perone, "Analysis of the COVID-19 epidemic through SIR models," Journal of Computational Science, vol. 40, p. 101067, 2020. doi: 10.1016/j.jocs.2020.101067

[41] M. Moftakhar and A. Seif, "Exploring visual analytics for pandemic trends," Digital Medicine, vol. 4, no. 4, pp. 122–133, 2021. doi: 10.1109/DMED.2021.3057605

[42] N. Wen et al., "Making predictions interpretable in public health modeling," Journal of Biomedical Informatics, vol. 120, p. 103871, 2021. doi: 10.1016/j.jbi.2021.103871

[43] G. Dong and F. Zhu, "Adaptive models for evolving pandemics," Pattern Recognition Letters, vol. 146, pp. 116–123, 2021. doi: 10.1016/j.patrec.2021.02.017

[44] A. Marziano et al., "Modeling the impact of COVID-19 variants using real-time data," PNAS, vol. 118, no. 10, p. e2100094118, 2021. doi: 10.1073/pnas.2100094118

[45] H. Zare et al., "Enhancing hybrid prediction models for public health crises," IEEE Access, vol. 10, pp. 1–12, 2022. doi: 10.1109/ACCESS.2022.3145263

[46] World Health Organization, "WHO Coronavirus (COVID-19) Dashboard," 2023. [Online]. Available: https://covid19.who.int/. [Accessed: 01-Dec-2023].

[47] European Centre for Disease Prevention and Control, "COVID-19 situation update worldwide," 2023. [Online]. Available: https://www.ecdc.europa.eu/en/covid-19-pandemic. [Accessed: 01-Dec-2023].

[48] Johns Hopkins University, "COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE)," 2023. [Online]. Available: https://github.com/CSSEGISandData/COVID-19. [Accessed: 01-Dec-2023].

[49] G. E. P. Box and G. M. Jenkins, Time Series Analysis: Forecasting and Control, 5th ed. Upper Saddle River, NJ, USA: Prentice Hall, 2015.

[50] P. J. Brockwell and R. A. Davis, Introduction to Time Series and Forecasting, 2nd ed. New York, NY, USA: Springer, 2002. DOI: 10.1007/978-1-4757-3569-5.

[51] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.

[52] A. K. Jain and P. M. Thakkar, "Forecasting time series data using LSTM," Journal of Computer and Communications, vol. 8, no. 1, pp. 14-24, 2020. DOI: 10.4236/jcc.2020.81002.

[53] Y. Liu, Z. Zhang, and Y. Zhang, "A hybrid model of ARIMA and LSTM for time series forecasting," Mathematical Problems in Engineering, vol. 2021, pp. 1-10, 2021. DOI: 10.1155/2021/6694089.

[54] M. T. V. P. R. G. R. Reddy, "Prediction of COVID-19 cases using machine learning models," International Journal of Emerging Technologies and Innovative Research, vol. 7, no. 9, pp. 333-339, 2020.