INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Progression of COVID-19 Cases in Telangana State by using ARIMA, MLP, ELM and LSTM Prediction Models by Retrospective Confirmation

OPEN ACCESS

Received: 28-01-2024

Accepted: 23-02-2024

Published: 14-03-2024

Citation: Rajendar M, Reddy DM, Nagesh M, Nagaraju V (2024) Progression of COVID-19 Cases in Telangana State by using ARIMA, MLP, ELM and LSTM Prediction Models by Retrospective Confirmation. Indian Journal of Science and Technology 17(12): 1159-1166. https://doi.org/10.17485/IJST/v17i12.211

* **Corresponding author**.

rajender.sas@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

M Rajendar[1,2]*, D Mallikarjuna Reddy[3], M Nagesh[1,2], V Nagaraju[4]

1 Research Scholar, Department of Mathematics, School of Science GITAM: Deemed to be University, Hyderabad, Telangana, India
2 Assistant Professor, Department of Mathematics, CMR Technical Campus, Hyderabad, 501401, Telangana, India
3 Department of Mathematics, School of Science GITAM: Deemed to be University, Hyderabad, 502329, Telangana, India
4 Assistant Professor, Department of Mathematics, Nalla Malla Reddy Engineering College, Hyderabad, 500088, Telangana, India

Abstract

Objective: The importance of this research article is to evaluate efficient model for diagnosing pandemic COVID-19 positive cases in Telangana State, India. **Method:** Neural Network models (Extreme Learning Machine and Multi-Layer Perception), Deep Learning Neural Network model (Long Short Term Memory-LSTM) and traditional Auto Regressive Integrated Moving Average (ARIMA) models were applied and the data was converted from non-linear to linear (stationarity) for forecasting Covid-19 positive cases. The study of the data covered from 1$^{st.}$ Dec 2020 to 30th May 2021. 80% of train data was taken to fit the models and then 20% of the test data was used to predict the values. The deviation between original test data and predicted data led to an error. Among these error values, the model which had minimum errors was considered as the best of the four models. **Findings:** LSTM model was proved to be the most efficient model, as a result of the least Root mean square error (RMSE = 71.12) compared to ARIMA (258.20), ELM (553.67) and MLP (641.86) values. **Novelty:** These forecasting methods succour to predict the Covid-19 positive cases in the forthcoming days. This has been suggested for taking the better preventive steps to control the Covid-19 positive cases.

Keywords: COVID19; ARIMA; LSTM; MLP; ELM Forecasting

1 Introduction

According to World Health Organization, Corona virus disease known as Covid-19 it was declared a pandemic and it spread out whole World. COVID-19 disease was detected by Deep learning techniques while analyzing the images. The Neural Network (NN) approach predicted the mortality risk in patients with COVID-19 [1]. It started in

2019 and was named as SARS CoV-2 (Severe Acute Respiratory Syndrome Corona virus-2). Initially it started spreading in Hubei, one of the provinces of China which claimed numerous human lives. It discusses machine learning methods to help medical decision-making by predicting the mortality risk in patients with COVID-19 [2,3].

The accurate model was found by the Deep Neural Networks with X-Ray images to detect the COVID-19 positive case. The Gaussian Process Regression (GPR) was another method used to find the accurate forecasting COVID-19 positive cases spread. The GPR based model helped towards, the results. It was revealed that significant change was obtained and showed the predictions between 95% Confidence Interval [4]. The investigation of a novel hybrid model that borrowed a high structured Auto-regressive and LSTM models which succor for the prediction of COVID-19 positive cases [5].

The ARIMA forecast method helped to predict the Covid-19 positive cases; the results were clear and at the range of 95% Confidence interval. In order to make the further plan, the need of hospitals and pharmaceutical facilities to be improved [6]. The article describes ARIMA, Seasonal Auto Regressive Integrated Moving Average (SARIMA) and Prophet Models were examined to find the best fit models. The data was taken from the World Health Organization which published Covid-19 positive cases between the period (May 1st, 2020 and Nov. 30th, 2021). The comparison of the model performance was evaluated by using the following errors: RMSE, Moving Average Error (MAE) and Moving Average Percentage Error (MAPE) [7]. The impact of COVID-19 had a significant effect on economic and social activities throughout the World. Covid-19 affected the existing inadequate medical facilities. The traditional ARIMA model, Deep learning models, Neural Network models and LSTM models were utilized to predict the Covid-19 positive cases. The hybrid models were utilized for forecasting the data of COVID-19 positive cases. The accuracy of these models was assessed by the three above error values. This paper reveals that ARIMA (7, 1, 0) is the optimum model. The performance of the models was appraised by employing errors [7,8]. The lower values have been found from fitted multiple linear regression (MLR) Prophet and ARIMA models. The result has shown by comparing the models, that ARIMA was the best forecasting performance among the models [9,10]. Regression models and Machine learning were useful for the prediction of COVID-19 positive cases. Aim of research paper is to identify significant predictors for daily active cases [11–17].

In this study covered severity of COVID-19 positive cases and presented by improving the mathematical modeling with the Cloud computing and Machine learning. It was globally recognized as a threat; several studies have been conducted. Different kinds of models were protected. The mathematical models were subject to potential bias. The study states and suggested ARIMA model was the best approach to predict it [18–20]. The research paper examined an analysis of regular statistics of people who were affected by the COVID-19 pandemic disease. The linear and multiple linear regression models were built for the positive cases of COVID-19 about the State Odisha as well as India. The prediction of the best model performance was decided by Coefficient of determination (R^2) value [21]. Covid-19 disease was detected by computer tomography images. In this research they developed hybrid Neural Networks, Artificial (NN) and Deep Neural Network models were utilized to predict the COVID-19 positive cases. Deep Neural Networks were seen to be efficient models; this paper also provided predictive models. Evaluation for prediction of death status based on food categories were analyzed by linear regression [22–25].

2 Methodology

2.1 ARIMA model

Model identifications: This includes the AR and MA components ware identifying the most suitable lags values, and AR (p) model equation $\widetilde{Y}_t = \omega_1 Y_{t-1} + \omega_2 Y_{t-2} + \cdots + \omega_p Y_{t-p} + \varepsilon_t$

MA (q). The model equation is: $\widetilde{Y}_t = e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}$

To determine whether the variable essentials stationary, the 1st differentiation made the best model. By using Auto Correlation function (ACF) and the Partial Auto Correlation function (PACF). The estimations have been done for usually, it consists of the least squares estimation method. Diagnostic checking out: This has generally used to test for autocorrelation if those components fail, then the way to return the choice of the identity and it starts once more generally through the addition of more variables. Forecasting: For forecasting lagged variables are used in ARIMA model such as ADF test.

The ARIMA model is follows:

$$\widetilde{Y}_t = \omega_1 Y_{t-1} + \omega_2 Y_{t-2} + \cdots + \omega_p Y_{t-p} + \varepsilon_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \cdots + e_{t-q} + c$$

2.2 Multi-Layer-Perception (MLP)

The MLP model belongs to a type of ANN known as a feed forward neural network. A feed forward neural network is a form of Neural network that may approximate continuous and integrable functions. MLP's network architecture is composed of layers of neurons. All of the input nodes are in one layer in the MLP model, while the hidden layer has been divided into one or more hidden layers [26,27]. [Figure 1]

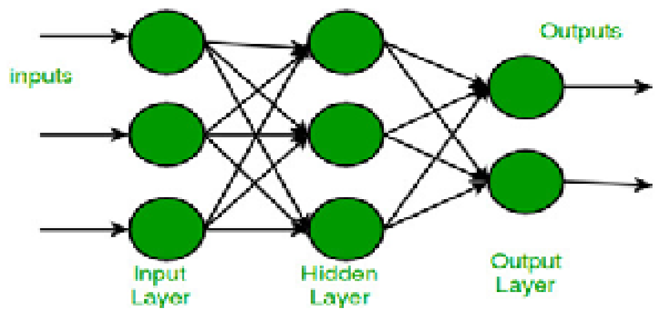

Fig 1. (Multi-Layer Perception model): Artificial neural network general architecture

2.3 Extreme Learning Machines (ELM):

ELM Model has a novel machine learning neural network algorithm used to model and forecast a time series were the Extreme-learning-machines (ELM) algorithm it was proposed by Huang. This algorithm well suited for Single hidden layer feed-forward Neural network (SLFN), which is identical to the feed-forward Neural networks. The most important feature of ELMs was the input weights and the hidden layer bias could be attributed randomly. Therefore, the architecture of the network resembles the resolution of a linear system. The unknown weights connect the hidden layer with the output layer. [Figure 2]

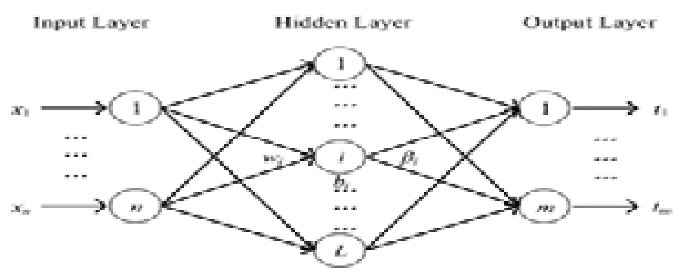

Fig 2. Artificial Neural Network (Extreme Learning Perception model) general architecture

2.4 Long Short-term Neural Network (LSTM)

LSTM is an advance form of Recurrent neural networks (RNN), this model solves gradient vanishing problem by using long short term neural networks. These models long term dependencies and calculate optimum time lag of Time series by giving memory unit ability to decide, forget or remember some of the information. This complex position in RNN allow LSTM network ability to recall past data, this makes creation of connections easy between past and current data points and permitting network discover patterns over time [28]. [Figure 3]

3 Results and discussion

The data collected from Telangana State, official website (https://covid19..telangana.gov.in) daily from 1st Dec.2020 to 31st May2021. The following statistical parameters computed, and it has been displayed in the [Table 1], [Figure 4 **(a)**].

The above graph describes trend of the number of positive cases versus the number of days from31st Dec 2020. The increasing rate of the no. of positive cases was constant and low, later it increased rapidly and then decreased.

Input files: [Table 2]. Total data divided into 80% for train data and 20% for test data were utilized. Plots have been drawn for the Covid-19 Positive Cases by applying Auto Correlation Function and Partial Auto Correlation Function. It has shown in the following figures [Figure 4 **(b), (c)**].

To test the Stationarity, ADF test is used to train data set. The following results were given:

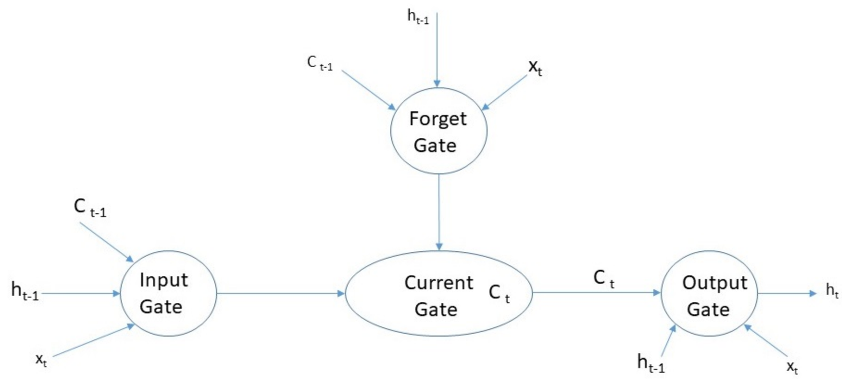

Fig 3. The General Architecture Long Short-term Neural Network

Table 1. Description of data for Covid-19 positive cases of Telangana state

	Date	Positive cases	Recovered	Deaths
Minimum	1/12/20	99	157	0
I^{-} Quarter	15/1/21	196	197	2
Median	1/3/21	463	397	3
Mean	1/3/21	1688	1539	10
3rd Quarter	15/4/21	3043	1198	9
Maximum	30/15/21	10122	9122	59

Table 2. Train and Test data for covid-19 positive cases

Set	Percentage	Observations
Training set	80%	145
validation	20%	36
total	100%	181

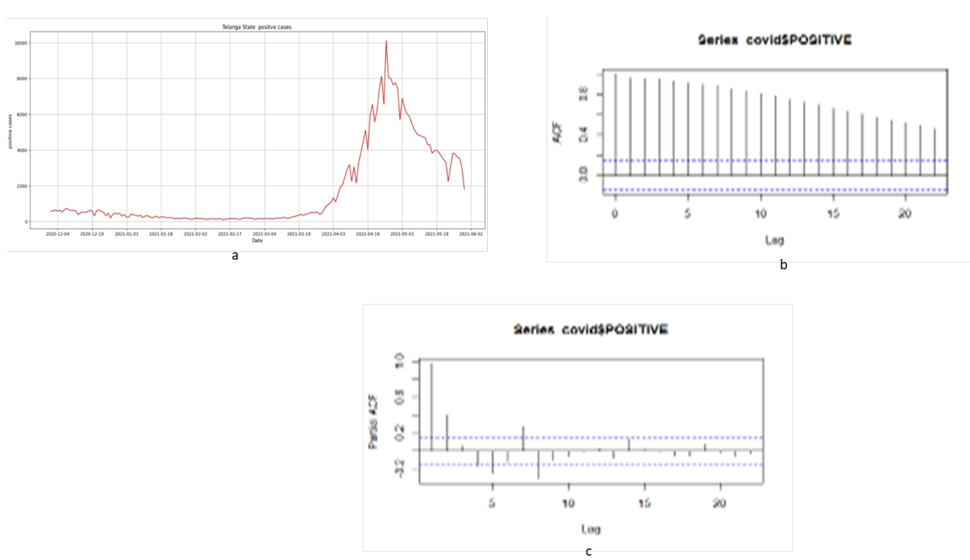

Fig 4. (a).

Hypothesis: Non stationary, after applying the Augmented Dickey Fuller test, it has observed that Augmented Dickey Fuller Test (ADF) Dickey Fuller = -1.877, Lag order value = 5, Value (p) = 0.627.

Alternative hypothesis: Stationarity, ADF test of the first difference has become Stationarity and the following results were given, ADF Value = -4.813, lag order = 5 and value (p) = 0.01 were accepted. Therefore, the data has converted to Stationarity.

By using R-Code, the auto ARIMA was fitted with the differences of Covid-19 Positive cases, Telangana State. Iterative procedure, the suitable model was found to be ARIMA (4, 1,2). [Table 3]

Table 3. ARIMA coefficients and SE Table

	AR1	AR2	AR3	AR4	MA1	MA2
Coefficients	-0.24865	-0.7024	-0.1060	0.1096	-0.2008	0.9742
SE	0.0770	0.0794	0.0806	0.0815	0.0232	0.0357

SE → Standard Error

The results were represented as the Estimated value of σ^2 = 169159: log likelihood value = -1337.91,
AIC = 2689.82, AICc = 2690.47, IC = 2733.20 and BIC = 2712.17.

3.1 MLP model fitting

There are five hidden nodes and 20 repetitions were used in the MLP fit. Differences-based series model: D1.

The Univarate lags: The median operator used to combine (1, 3, 4) has been predicted. The Root Mean Square Error =15156.85. [Figure 5**(a)**, (**b**)]

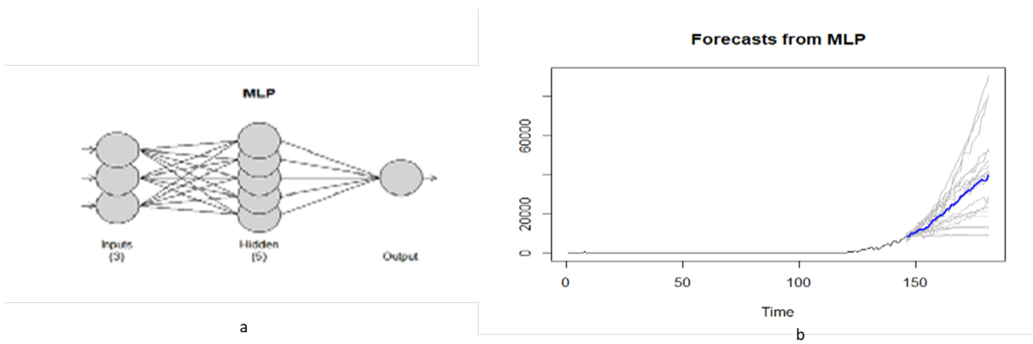

Fig 5. (a).Multi-Layer Perception Model for Covid-19 positive cases, (b). Multi-Layer Process forecasting values

3.2 ELM model fitting

There were 100 hidden nodes and 20 repetitions used in the ELM fit. Differences-based series model D1. The univariate lags: the median operator output weight estimation was used to combine (1, 3, 4) predicted. The Root Mean Square Error = 104281.16. [Figure 6].

3.3 LSTM model fitting

For the evaluation of model 80% of the data has taken trained and 20% was tested data. Data is normalized using MinMaxScaler in Python, the best hyper parameter tuning of LSTM, determined by manual search is "Batch size = 32, dropout = 0.2, 20 epochs and 50 units". In LSTM model total 10451 parameters trained when model fitted for the Telangana State, India. 10400 parameter were trained in hidden layers, 51 parameters were trained in output layer. Effectiveness of loss is measured using mean square error of ADAM optimizer. LSTM model for prediction of Covid-19 Positive cases is obtained in [Figure 7].

3.4 Computation of Errors

$$MAE = \frac{\sum_{i=1}^{n} |X_i - X_i^{\wedge}|}{N} \quad (1)$$

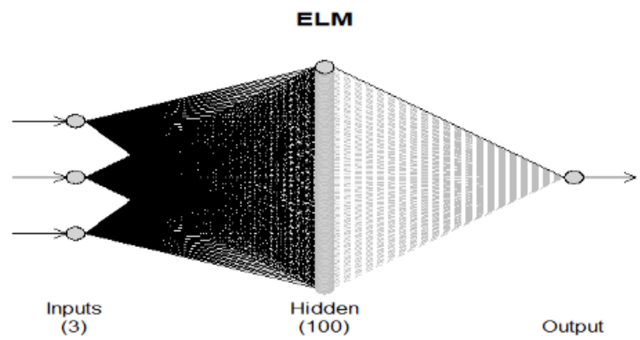

Fig 6. Extreme Learning Machines forecasting values

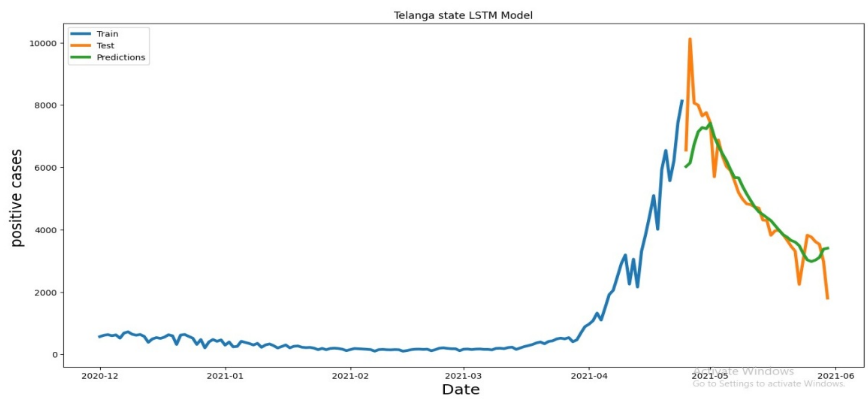

Fig 7. LSTM forecasting values

Where $X_i^{\wedge}$= Observed values; X_i = Expected values

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{T}\left(S_{it} - S_{it}^{\wedge}\right)^2} \tag{2}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{T}\frac{\left|S_{it} - S_{it}^{\wedge}\right|}{S_{it}} \times 100 \tag{3}$$

Where S_{it} indicate actual value. T and $S_{it}^{\wedge}$ are total number of days and modeled values.

Error comparison table for ARIMA, MLP, ELM and LSTM error values represented in [Table 4].

Table 4. Error comparison table for ARIMA, MLP, ELM and LSTM models

Model	RMSE	MAE	MAPE
ARIMA	258.2	125.9	17.28
MLP	641.86	277.10	24.87
ELP	553.67	259.47	26.73
LSTM	71.12	34.67	4.57

Result-I: By comparing Root mean square error of ELM and MLP; (553.67 < 641.86) Extreme learning machine is efficient model.

Result-II: By comparing Root mean square error of ARIMA, ELM and MLP; (258.2 < 553.67 < 641.86) ARIMA is efficient model.

Result-III: By comparing Root mean square error of LSTM, ELM and LSTM; (71.2< 553.67 < 641.86) Long short term memory is efficient model.

Result-IV: By comparing Root mean square error of LSTM, ARIMA, MLP and ELM; (71.12 < 258.2< 553.67 < 641.86) Long Short term memory is the best model. Conclude that LSTM has been found to be the best model.

4 Conclusion

Among the four, the best procedure has been revealed LSTM model, this method is the latest and advance Recurrent neural network. It is a Deep learning neural network model; the method is best because the complex position in RNN allows LSTM, its ability to recollect the past data. The above four methods observed RMSE value of LSTM has been shown the minimum error (71.12) which is in the result-IV represented. The application of time series traditional ARIMA model is also the good performance shown in Result-II as compared to the Artificial Neural Network (ANN) models. The Extreme Learning Machine (ELM) and Multi-Layer-Perception (MLP) were recently implemented for the Neural Network models. By comparing their error values of MAE, MAPE and RMSE, minimum error considered to be the best performance model. The architecture of the Neural Networks consists of an Input Layer, a hidden layer and output layer were taking number of neurons, during learning the time and cost was too high of MLP and ELM as compared to ARIMA models.

References

1) Chang TY, Huang CK, Weng CH, and JYC. Feature-based deep neural network approach for predicting mortality risk in patients with COVID-19. *Engineering Applications of Artificial Intelligence*. 2023;124:1–11. Available from: https://doi.org/10.1016/j.engappai.2023.106644.
2) Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health*. 2021;20:1–8. Available from: https://doi.org/10.1016/j.smhl.2020.100178.
3) Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim OR, Acharya UR. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*. 2020;121:1–11. Available from: https://doi.org/10.1016/j.compbiomed.2020.103792.
4) Alali Y, Harrou F, Sun Y. A proficient approach to forecast COVID-19 spread via optimized dynamic machine learning models. *Scientific Reports*. 2022;12(1):1–20. Available from: https://doi.org/10.1038/s41598-022-06218-3.
5) Zhang Y, Tang S, Yu G. An interpretable hybrid predictive model of COVID-19 cases using autoregressive model and LSTM. *Scientific Reports*. 2023;13(1):1–12. Available from: https://doi.org/10.1038/s41598-023-33685-z.
6) Darapaneni N, Reddy D, Paduri AR, Acharya P, Nithin HS. Forecasting of COVID-19 in India Using ARIMA Model,& quot; 2020 11th IEEE Annual Ubiquitous Computing. In: 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE. 2020. Available from: https://doi.org/10.1109/UEMCON51285.2020.9298045.
7) Wang Y, Yan Z, Wang D, Yang M, Li Z, Gong X, et al. Prediction and analysis of COVID-19 daily new cases and cumulative cases: times series forecasting and machine learning models. *BMC Infectious Diseases*. 2022;22(1):1–12. Available from: https://doi.org/10.1186/s12879-022-07472-6.
8) Jin YC, Cao Q, Wang KN, Zhou Y, Cao YP, Wang XY. Prediction of COVID-19 Data Using Improved ARIMA-LSTM Hybrid Forecast Models. *IEEE Access*. 2023;11:67956–67967. Available from: https://doi.org/10.1109/ACCESS.2023.3291999.
9) Zhao D, Zhang R, Zhang H, He S. Prediction of global omicron pandemic using ARIMA, MLR, and Prophet models. *Scientific Reports*. 2022;12(1):1–13. Available from: https://doi.org/10.1038/s41598-022-23154-4.
10) Hasan I, Dhawan P, Rizvi SAM, Dhir S. Data analytics and knowledge management approach for COVID-19 prediction and control. *International Journal of Information Technology*. 2023;15(2):937–954. Available from: https://doi.org/10.1007/s41870-022-00967-0.
11) Semwal J, Bahuguna A, Uniyal A, Vyas S. A Study to Analyse Covid-19 Outbreak Using Multiple Linear Regression: A Supervised Machine Learning Approach. *National Journal of Community Medicine*. 2023;14(02):82–89. Available from: https://doi.org/10.55489/njcm.140220232656.
12) Kumar RP, Rithesh A, Josh P, Raj B, John V, Prasad DS. Sleep Track: Automated Detection and Classification of Sleep Stages. In: 15th International Conference on Materials Processing and Characterization (ICMPC 2023);vol. 430. 2023;p. 1–13. Available from: https://doi.org/10.1051/e3sconf/202343001020.
13) Bhattamisra SK, Banerjee P, Gupta P, Mayuren J, Patra S, Candasamy M. Artificial Intelligence in Pharmaceutical and Healthcare Research. *Big Data and Cognitive Computing*. 2023;7(1):1–20. Available from: https://doi.org/10.3390/bdcc7010010.
14) Khoojine AS, Shadabfar M, Vahid R, Hosseini H, Kordestani. Network Autoregressive Model for the Prediction of COVID-19 Considering the Disease Interaction in Neighboring Countries. *Entropy*. 2021;23(10):1–18. Available from: https://doi.org/10.3390/e23101267.
15) Shetty RP, Pai PS. Forecasting of COVID 19 Cases in Karnataka State using Artificial Neural Network (ANN). *Journal of The Institution of Engineers (India): Series B*. 2021;102:1201–1211. Available from: https://doi.org/10.1007/s40031-021-00623-4.
16) Oshinubi K, Amakor A, Peter OJ, Rachdi M, Demongeot J. Approach to COVID-19 time series data using deep learning and spectral analysis methods. *AIMS Bioengineering*. 2022;9(1):1–21. Available from: https://doi.org/10.3934/bioeng.2022001.
17) Zheng N, Du S, Wang J, Zhang H, Cui W, Kang Z, et al. Predicting COVID-19 in China Using Hybrid AI Model. *IEEE Transactions on Cybernetics*. 2020;50(7):2891–2904. Available from: https://doi.org/10.1109/TCYB.2020.2990162.
18) Tuli S, Tuli S, Tuli R, Gill SS. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet of Things*. 2020;11:1–16. Available from: https://doi.org/10.1016/j.iot.2020.100222.
19) Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497–506. Available from: https://doi.org/10.1016/s0140-6736(20)30183-5.

20) Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*. 2020;29:1–4. Available from: https://doi.org/10.1016/j.dib.2020.105340.
21) Rath S, Tripathy A, Tripathy AR. Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. 2020;14(5):1467–1474. Available from: https://doi.org/10.1016/j.dsx.2020.07.045.
22) Ozyurt F, Tuncer T, Subasi A. An automated COVID-19 detection based on fused dynamic exemplar pyramid feature extraction and hybrid feature selection using deep learning. *Computers in Biology and Medicine*. 2021;132:1–10. Available from: https://doi.org/10.1016/j.compbiomed.2021.104356.
23) Alassafi MO, Jarrah M, Alotaibi R. Time series predicting of COVID-19 based on deep learning. *Neurocomputing*. 2022;468:335–344. Available from: https://doi.org/10.1016/j.neucom.2021.10.035.
24) Shams MY, Elzeki OM, Abouelmagd LM, Hassanien AE, Elfattah MA, Salem H. HANA: A Healthy Artificial Nutrition Analysis model during COVID-19 pandemic. *Computers in Biology and Medicine*. 2021;135:1–16. Available from: https://doi.org/10.1016/j.compbiomed.2021.104606.
25) Yadav M, Perumal M, Srinivas M. Analysis on novel coronavirus (COVID-19) using machine learning methods. *Chaos, Solitons & Fractals*. 2020;139:1–12. Available from: https://doi.org/10.1016/j.chaos.2020.110050.
26) Anne WR, Jeeva SC. ARIMA modelling of predicting COVID-19 infections. . Available from: https://www.medrxiv.org/content/10.1101/2020.04.18.20070631v1.full.pdf.
27) Khan FM, Gupta R. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *Journal of Safety Science and Resilience*. 2020;1(1):12–18. Available from: https://doi.org/10.1016/j.jnlssr.2020.06.007.
28) Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*. 2020;135:1–6. Available from: https://doi.org/10.1016/j.chaos.2020.109864.