



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Alexandria University  
**Alexandria Engineering Journal**

[www.elsevier.com/locate/aej](http://www.elsevier.com/locate/aej)  
[www.sciencedirect.com](http://www.sciencedirect.com)



# Soft computing techniques for forecasting of COVID-19 in Pakistan



**Muhammad Naeem<sup>a,b</sup>, Wali Khan Mashwani<sup>a,\*</sup>, Mohammad ABIAD<sup>c</sup>,  
 Habib Shah<sup>d</sup>, Zardad Khan<sup>b</sup>, Muhammad Aamir<sup>b,\*</sup>**

<sup>a</sup> *Institute of Numerical Sciences, Kohat University of Science & Technology, Kohat, KPK, 26000, Pakistan*

<sup>b</sup> *Department of Statistics, Abdul Wali Khan University Mardan, Pakistan*

<sup>c</sup> *Department of Mathematics and Statistics, American University of the Middle East, Egaila, Kuwait*

<sup>d</sup> *College of Computer Science, Department of Computer Science, King Khalid University, Abha, Kingdom of Saudi Arabia*

Received 19 February 2022; revised 16 July 2022; accepted 18 July 2022

Available online 22 July 2022

## KEYWORDS

Confirm Cases;  
 COVID-19;  
 Deaths Cases;  
 Forecasting;  
 Kalman Filter;  
 Neural Networks;  
 Pakistan;  
 Prediction;  
 Recoveries

**Abstract** Novel Pandemic COVID-19 led globally to severe health barriers and financial issues in different parts of the world. The forecast on COVID-19 infections is significant. Demeanor vital data will help in executing policies to reduce the number of cases efficiently. Filtering techniques are appropriate for dynamic model structures as it provide reasonable estimates over the recursive Bayesian updates. Kalman Filters, used for controlling epidemics, are valuable in knowing contagious infections. Artificial Neural Networks (ANN) have generally been used for classification and forecasting problems. ANN models show an essential role in several successful applications of neural networks and are commonly used in economic and business studies. Long short-term memory (LSTM) model is one of the most popular technique used in time series analysis. This paper aims to forecast COVID-19 on the basis of ANN, KF, LSTM and SVM methods. We applied ANN, KF, LSTM and SVM for the COVID-19 data in Pakistan to find the number of deaths, confirm cases, and cases of recovery. The three methods were used for prediction, and the results showed the performance of LSTM to be better than that of ANN and KF method. ANN, KF, LSTM and SVM endorsed the COVID-19 data in closely all three scenarios. LSTM, ANN and KF followed the fluctuations of the original data and made close COVID-19 predictions. The results of the three methods helped significantly in the decision-making direction for short term strategies and in the control of the COVID-19 outbreak.

© 2022 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Alexandria University This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The pandemic, COVID-19, emerged for the first time in China in December 2019. The COVID-19 spread in China very quickly and affected other countries including Pakistan. After

\* Corresponding authors.

E-mail addresses: [mashwanigr8@gmail.com](mailto:mashwanigr8@gmail.com) (W. Khan Mashwani), [aamirkhan@awkum.edu.pk](mailto:aamirkhan@awkum.edu.pk) (M. Aamir).

Peer review under responsibility of Faculty of Engineering, Alexandria University.

<https://doi.org/10.1016/j.aej.2022.07.029>

1110-0168 © 2022 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Alexandria University This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

some time, the mortality rate exceeded fifty thousand figure, and about more than a million individuals got infected across the world. To control the COVID-19, various steps were taken globally. The government of Pakistan initially managed to overcome the pandemic with the help of smart lockdown and by constructing exceptional emergency clinics, isolated offices, COVID-19 testing and laboratories [41].

The authors (M.-T. [22] proposed a model for Wuhan cases, which was adopted by the local government. It was helpful in limiting the disease from spreading in all cases. In the proposed model, the new cases were predicted to range from  $0.7 \times 10^5$  to  $14 \times 10^4$  by 29 Feb 2020. Similarly, the number of confirmed cases reached to 32,000 by Feb 29, 2020. According to the model, death cases got control with the passage of time. The model also helped us in the sense that with the help, we could avoid such outbreaks in the future. Their model proved successful in predicting the COVID-19 pattern of the original number of cases [25]. Daily increase in the figure of confirmed cases increased to 34,598 on Feb 8, 2020. The authors in [2], evaluated and forecasted the number of cases in the approaching ten days of COVID-19 in advance of the real cases in China, with a new forecasting model. The authors in (Anastassopoulou, Russo, Tsakris, & Siettos, 2020) have developed a model to estimate and determine the occurrence of COVID-19 infection for the Hubei region, China, from 11 Jan 2020 to 10 Feb 2020. They predicted the key factors involved in the transmission of the disease like the ratio of death cases, with 90% certainty stretch, for the Susceptible Infected Recovered Dead (SIRD) model.

[21] organized and combined the SARS-COV-2 data, and developed a model with four datasets, which originated from the city of Wuhan. In addition, the authors forecasted COVID-19 in December 2019 and February 2020. It was found that 50% of the COVID-19 disease spread from local people.

United States of America overtook South Korea concerning confirmed cases of COVID-19 by reaching a maximum number of 33,500 cases on 23 March 2020. Then the number of cases were more than 240,000, with 5800 deaths on 02 April 2020. Italy was the third country affected by COVID-19 for being unwilling to adopt precautionary measures to face the outbreak, i.e., through the complete lockdown. On 23 March 2020, the total number of deaths recorded in Italy were 5,400, and the total number of confirmed cases recorded were  $105 \times 10^3$ .

The applications of soft computing are used widely almost in all fields. The AI applications are helpful in various predictions and diagnoses of diseases. Besides fighting against the virus infection, these tools have the strength to support us in preventive measures. There are some advantages and disadvantages of traditional algorithms. The researchers developed various techniques to overcome this problem. The authors [40] proposed the self-adaptive extreme learning machine (SaELM) algorithm that selects the most appropriate number of neurons in the hidden layers. Moreover, there is no need to adjust the other parameters for choosing the number of neurons in the hidden layers in the method. If the neurons are improperly selected, then bad predictions are generated. To resolve these types of problems, the authors [43] developed the improved probabilistic neural networks with self-adaptive strategies, and it has multiple uses in various fields. The authors compared

the technique with known methods and proved the proposed method based on their performance. The authors in [19] developed and studied the soft computing techniques for the COVID-19 outbreak. The authors used techniques including machine learning (ML), deep learning (DL), support vector machine (SVM) and internet of things (IoT) to analyze COVID-19. They concluded that these soft computing techniques were helpful in health care policies. [4] ML and DL methods were used to classify the COVID-19 outbreak infected lung computed tomography segmentation. The key objective of study is two-folded; firstly, to discover the application of soft computing techniques, the outline of machine learning (ML), deep learning (DL), support vector machines (SVM), and internet of things (IoT), which have been interpreted about COVID-19. Secondly, to examine performance, the authors made a relative study that has been represented to express that the soft computing models outperform various successive analytical models on the COVID-19. Then to predict a COVID-19, A Cauchy distribution probabilistic approach is used. The authors in (Ahmadini et al., 2021) used Kalman filter algorithms for the top four affected countries of COVID-19 with the help of both methods (smoothed and filtered). The filter algorithms followed the actual trend of the data and forecasted new confirmed, death and recovered cases for the next three weeks. It suggests that the Kalman filter algorithm can make predictions for the days to come, and the government can use it to make counter arrangements for meeting future attacks of COVID-19. The authors [32] have constituted a forecast model on the aggregate adjustment of Kalman Filter concerning the periodic events of flu. They utilized aggregate changes of Kalman Filter for studying susceptible-infected-recovered-susceptible and flu-like disease by putting forward week-based forecasts for the occurrence of flu.

## 2. Literature review

[3] developed an improved strategy to estimate confirm cases of COVID-19 in China. On Feb 8, 2020, the confirmed cases reached gradually up to 34,598. The authors developed new forecast model which could measure and forecast the number of COVID-19 cases for the upcoming ten days in China. They utilized a slap swarm calculation to improve the versatile neuro-fluffy inference framework, and utilized better flower contamination calculations. The FPASSA-ANFIS model proved to be superior of all the models. This model used Mean Absolute Percentage Error (MAPE), Root Mean Squared Relative Error (RMSRE), and  $R^2$  for coefficient of assurance (Ekici & Aksoy, 2011).

Forecasting in medical situations is not often wide in range and the probability of occurrence of an event may be miscalculated. Forecasting about the future disease and epidemic is more challenging. In the past, with respect to measures taken to fight SARS, the proficiency and the use of preventive measures proved useless [27]. In March 2020, the COVID-19 was declared as a global epidemic. The death cases were more than 4,607 and  $125 \times 10^3$  number of confirm cases were noted in more than 118 countries. The predictive model proposed that the number of COVID-19 cases increased exponentially in countries that did not implement preventive measures [14].

The death rate recorded in China (1%) was smaller than that in Italy (3%), however, with the passage of time, the rate

of infected cases and mortality decreased. In the same way (L. [22]), also studied the progress in spread of the COVID-19.

COVID-19 spread repeatedly in China, Italy and Iran. Pakistan being in the neighborhood of China and Iran, was at a potential risk. Pakistan was not expecting similar rapid incidence of COVID-19 and also had no such timely arrangements to cope with it. When the first case was reported in Pakistan, smart lockdown was imposed and isolation centers were established in hospitals on emergency basis. Percentage of cases reported in people coming back from Iran to Pakistan was 78. Because Pakistan is a less developed country, the imposition of smart lockdown in Pakistan proved to be a successful measure to cope with a serious pandemic like COVID-19; however, COVID-19 still affected the daily wagers badly [45].

The authors in [45] developed an artificial neural network-based learning unit to enhance the prediction accuracy of the Kalman Filter algorithm and found that with this the performance of the model improved from 4.41% to 11.19% and the same like model used by (Ahmadini et al., 2021) for the top four affected countries. The authors in [44] applied Auto-Regressive Integrated Moving Average Model (ARIMA) using data from National Institute of Health (NIH), Pakistan to measure COVID-19 cases. Their model showed high exponential growth for the number of confirm, death and recovered cases and found that the estimates of COVID-19 cases would reach their peak in the month of May 2020. The health authorities and government ought to utilize new techniques to control the pandemic from spreading further until an appropriate treatment for it or immunization to it is discovered [44].

Machine learning and deep learning methods are widely used to predict and forecast in time series data [29]. The authors in [28] developed an application to assist the people activities. The authors used a Bidirectional residual LSTM-based method to recognize human activities. The authors also use d ML models to predict the COVID-19 cases among most affected countries. Furthermore, the authors showed that the COVID-19 cases will be declined in the future. The wrong reports have a big impact on every field especially on economy and public safety. For this purpose, [5] used a hybrid attention neural network (HANN) method to find the rumors related to COVID-19 on media. A similar study is carried out by (Almars, Atlam, et al., 2022) to find users view and sentiments on public media related to COVID-19 vaccine. Their results showed that majority of the people have a positive outlook towards COVID-19 vaccine.

LSTM model is one of the most popular method used in time series data [15]. The forecasting model, recurrent neural network (RNN), is an alternatives of long short-term memory (LSTM). The LSTM is both used for time series analysis (RNN) and Convolutional neural network (CNN) classification problems [26,30,34].

### 3. Data description and methods

#### 3.1. Data

In this study, we took data from GitHub website (<https://github.com/CSSEGISandData/COVID-19/tree/master/csse-covid-19-data/csse-covid-19-time-series#update-frequency>) which was accessed on 18 August 2021. The number of deaths

showed in Pakistan were 66 with 3974 cases of confirmed cases and 3122 cases of recovery. We predicted the main epidemiological factors, i.e. the number of confirmed, death and recovered cases on daily basis for the next three weeks. In this study, we divided the data in two phases. In the first phase, we divided the data into 70% training and 30% testing parts. Then in the second phase we divided the data into 80% training and 20% testing parts. Kalman Filter, a well-known method for efficient forecast accuracy, was found by Rudolf Emil Kalman in 1960. It is used in modern technology related to direction finding, mainframe visualization, econometrics, and time series analysis. ANN is also one of the powerful techniques, which is usually used for grouping and prediction problems. It is widely used in the economic and financial studies. In the same way, the LSTM and SVM are widely used for time series modelling. LSTM is one of the most powerful model in time series analysis. In this study, we used the KF, ANN, LSTM and SVM methods for the COVID-19 daily cases of confirmed, recovered and death in Pakistan.

#### 3.2. Kalman Filter algorithms

Consider an equitably common state-space model description [17]. Let.

$$\beta_t = \delta_t + T_t \beta_{t-1} + J_t \gamma_t \quad (1)$$

$$y_t = \omega_t + Z_t \beta_t + \epsilon_t \quad (2)$$

Where  $\gamma_t \sim N(0, L_t)$  and  $\epsilon_t \sim N(0, M_t)$ . The equation (1) of states defines the dynamics of  $\beta_t$ , the state vector, determined with the help of deterministic input  $\delta_t$  and stochastic inputs  $\gamma_t$ . The sample equation, the observed response  $y_t$ , by state direction over error term  $\epsilon_t$  and inputs  $\omega_t$ . The matrices  $J_t, T_t, Z_t, L_t$  and  $M_t$  might rest on the parameters vector  $\theta$  and remain time changing or determined beside time. The error terms  $\gamma_t$  and  $\epsilon_t$  remain imaginary successively and also equally uncorrelated, i.e.,  $E[\gamma_t \epsilon_s] = 0 \forall t, s$ . The past postulation and the gaussianity of  $\gamma_t$  and  $\epsilon_t$  remain distributed [8].

#### 3.3. The Kalman filter

The Kalman Filter equation through small notational deviations remain ordinary in some textbook: [8], [17], [12], [36], [9] and [37], by replication those equalities by means of repetitive reference, we get the result:

$$B_{t-1} = E[\beta_{t-1} | y_0, \dots, y_{t-1}] \quad (3)$$

$$\pi_{t-1} = E[(\beta_{t-1} - C_{t-1})(\beta_{t-1} - C_{t-1})^T] \quad (4)$$

It computes the state vector, besides at time t it is covariance matrix through data obtainable at time  $t-1$ ,  $C_{t|t-1}$  and  $\pi_{t-1}$  respectively that remains identified via the time update the equations:

$$C_{t|t-1} = T_t C_{t-1} + \delta_t \quad (5)$$

$$\pi_{t|t-1} = T_t \pi_{t-1} T_t^T + J_t L_t J_t^T \quad (6)$$

Suppose  $E_t = Z_t \pi_{t|t-1} Z_t^T + M_t$ . The condition is that, the new sample is accessible at time t, formerly  $C_{t|t-1}$  and  $\pi_{t|t-1}$  can be updated with the degree of update the equations:

$$C_t = C_{t|t-1} + \pi_{t|t-1} Z_t^T E_t^{-1} (y_t - Z_t C_{t|t-1} - d_t) \quad (7)$$

$$\pi_t = \pi_{t|t-1} - \pi_{t|t-1} Z_t E_t^{-1} Z_t^T \pi_{t|t-1} \quad (8)$$

In equations (5), (6), (7), and (8) composes the Kalman filter. Substituting equation (5) in (6) and equation (7) in (8). We got a single set equation involving  $C_{t-1}$  and  $\pi_{t-1}$  to  $C_t$  and  $\pi_t$  respectively. For replication, primary values of  $C_{t-1}$  and  $\pi_{t-1}$  or ( $C_{0|-1}$  and  $\pi_{0|-1}$ ) are being used. The equations of the filter i.e. (6) and (8) spread the state covariance matrix to define covariance filter. The resultant equations that spread the matrix  $\pi_t^{-1}$ , gives data filter. A better advantage is that it suggests a usual method to state whole unlikelihood about the prime value of component of the state: The affirming diagonal term can be set to zero in the data matrix. With the help of a covariance filter, the equivalent variance can be set in  $\pi_{0|-1}$  to a large number else verbose initialization can be used, a possibility is defined below.

Prediction results directly builds up the Kalman Filter into mainframe code. The abovementioned just observed that subsequent programs agonized after a mathematical improbability [10]. Accumulation of fluctuating argument inaccuracies in Eq (8) finally yield non-positive  $\pi_t$  matrices. Equation (8) more expanded as below:

$$\pi_t = (I - N_t Z_t) \pi_{t|t-1} (I - N_t Z_t)^T + N_t M_t N_t^T \quad (9)$$

ANN is a scientific tool that have predictors layer and response layers, and a hidden layer and has been usually used for grouping and prediction problems. A grouping of distinct hidden layers can choose a well fit MLP architecture network. The hidden layers play a key role in various successful applications of neural networks. ANN is commonly used in the economic and financial studies [18,31]. The number of hidden layers depends upon the nature of the problem. (Zhang, Patuwo, & Hu, 1998) used two hidden layers in finding better model prediction accuracy. Xu et al., 2020 used  $2 \times k + 1$ , where  $k$  is the number of predictors. For an ideal outcome of ANN, trial and error method is applied determining the number of hidden nodes. It is used in searching the architecture having the least MAPE among the models (Güler & Übeyli, 2005). In this study, there are four hidden layers and 8 neurons in the hidden layers is used for daily cases (death) using test and retest technique with 10,000 iteration while 2 hidden layers and 4 neurons in the hidden layers for recovered cases in similar pattern.

Long Short-Term Memory (LSTM) Neural Network is one of the most advanced models to forecast time series data. The prediction accuracy of LSTM model is better than the other methods, as it permits more constraints to be learned. Therefore, it is more powerful and widely used for forecasting and prediction. Recurrent neural network (RNN) is an important part of deep learning methods. RNN is used to find chronological data relationships in time series prediction [24,35]. It contains hidden layers that are spreader in chronological way that able to forecast the upcoming events with high accuracy rate than the other classical exponential smoothing methods [33,38,42,47]. There is no ending rule of thumb to choose on how many layers and nodes of the model. LSTM structure composed of four units namely input gate ( $i_g$ ), forget gate ( $f_g$ ), output gate ( $o_g$ ) and an activation function. We train and test the model on the COVID-19 data for Pakistan, with

sound parameter features which provides well match to actual data on trial and error based as shown in the Figs. 5, 6, and 7. There are four number of blocks, with 100 epochs size is used. The tanh activation function and 'adam' optimizer is used and other functions are kept as defaults.

Support vector machines are capable devices for the forecasting of financial time-series as it practice a risk utility covering of the observed error and a standardized term which results in fundamental risk minimization norm [20,39].

The Table 1, show the performance metrics for ANN and KF. This table used to evaluate the methods performance using test data.

#### 4. Results

We measured confirmed, death and recovered cases from the Pakistan COVID-19 dataset. The KF, ANN, LSTM and SVM models were applied for daily confirmed, death and recovered cases. The summary statistics is displayed in the Table 1. In this study, we assess the analytical results of soft computing models using five popular forecasting parameters, i.e. Mean error (ME), root mean square error (RMSE), mean absolute error (MAE), mean percentage error (MPE) and mean absolute percentage error (MAPE) are presented in Table 2, 3, 4, and 5.

In the Table 1, the minimum value of all the cases is zero while the maximum value of confirmed, death and recovered cases are 12073, 313 and 19,772 respectively. Similarly, the mean and median for the number of confirmed, death and recovered cases are 1929.56, 44.04, 1840.60 and 1592, 37 and 1282 respectively. While the first quartile value of daily confirm, death and recover cases are 665.5, 11 and 461.5 respectively. The third quartile value of number of daily confirm, death and recover cases are 2963.5, 66 and 2402 respectively. Moreover, the standard deviation of confirm, death and recover cases are 1674.98, 40.35 and 2132.76 respectively.

COVID-19 is initial as a viral disease and is a different set of warning to the world. We calculate the five values of ME, RMSE, MAE, MPE and MAPE in Table 2. In Table 2, the Mean Error (ME) value of Kalman Filter, ANN, LSTM and SVM for the daily confirm cases is 15.34, 15.09, 14.01 and 15.10 respectively. This indicates that ME value LSTM is lower than KF, ANN and SVM. Similarly, the value of RMSE and MAE value of Kalman Filter, ANN, LSTM and SVM are 744.98, 666.02, 500.76, 669.45, 541.09, 425.77, 399.03 and 469.32 respectively. Both these values revealed that LSTM is better than Kalman filter, ANN and SVM. The MPE and MAPE also indicate that the difference between the actual and forecasted value of LSTM is more consistent as compared to Kalman Filter, ANN and SVM. Comparing the five assessment parameters of Kalman Filter, ANN, LSTM and SVM models, it can be observed that LSTM performs better in terms of small error rate among the rest of soft computing models.

In Table 3, the results of death cases show that performance of LSTM in terms of RMSE, MAE, MPE and MAPE is better than the Kalman Filter, ANN and SVM. While, the ME value of ANN is better than KF, LSTM and SVM. The evaluated parameters of Kalman Filter, ANN, LSTM and SVM models, on average it can be showed that LSTM performs better in terms of high accuracy rate among the other soft computing models. Furthermore, LSTM is a better choice for prediction.

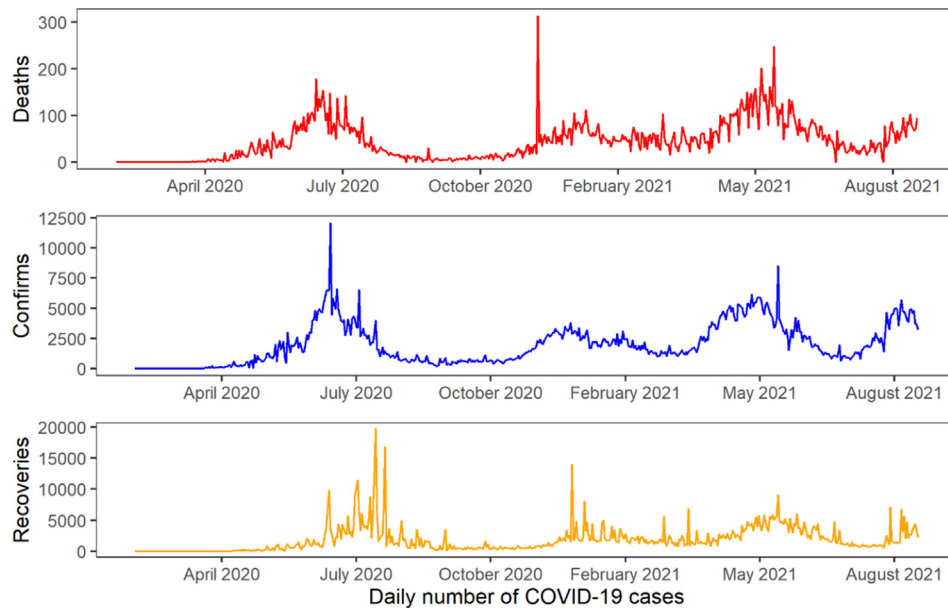


Fig. 1 Daily number of COVID-19 cases for Pakistan.

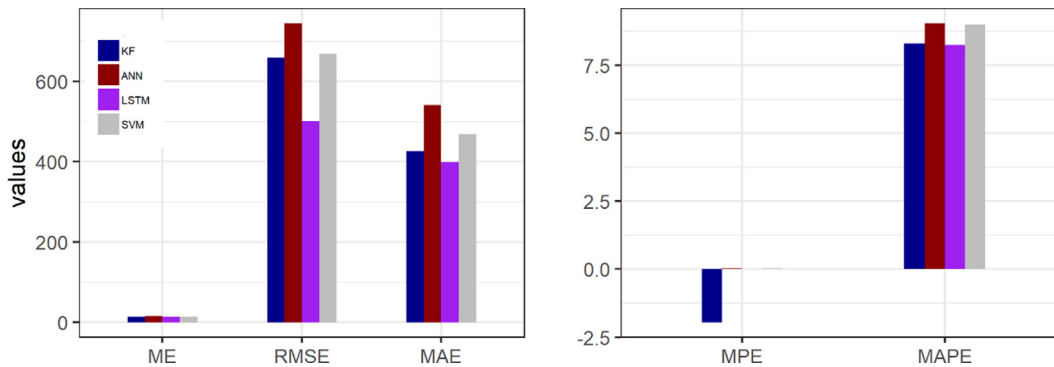


Fig. 2 Forecasting accuracy measures of KF, ANN, LSTM and SVM models for confirm cases.

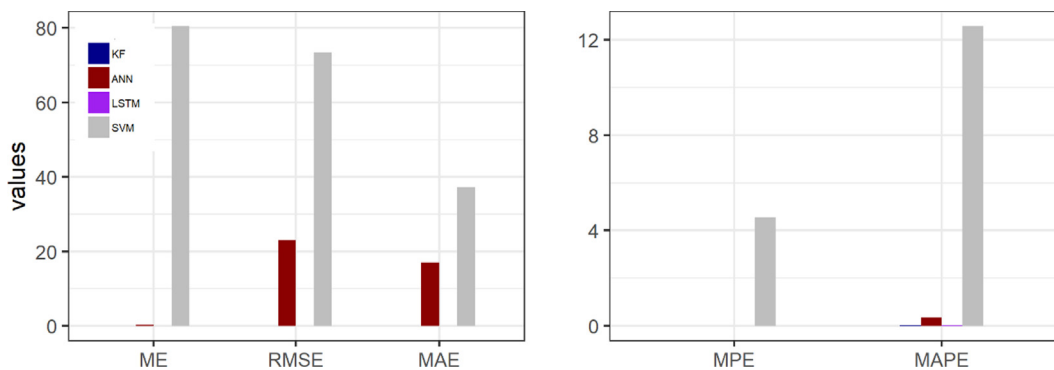
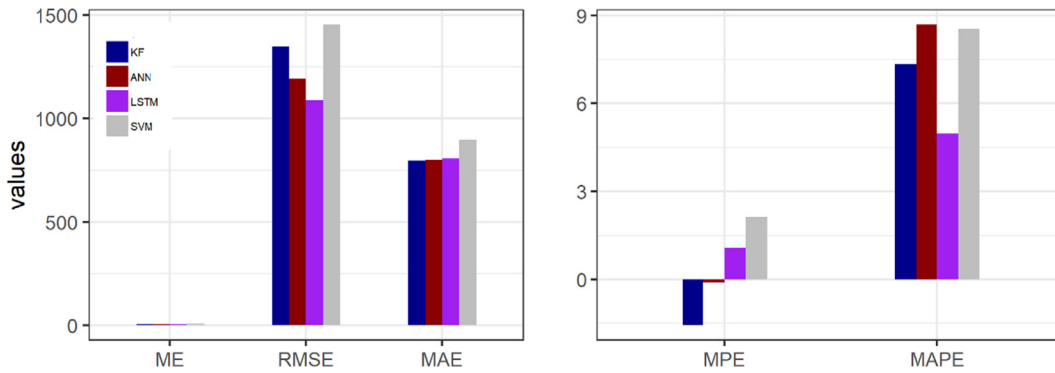


Fig. 3 Forecasting accuracy measures of KF, ANN, LSTM and SVM models for death cases.

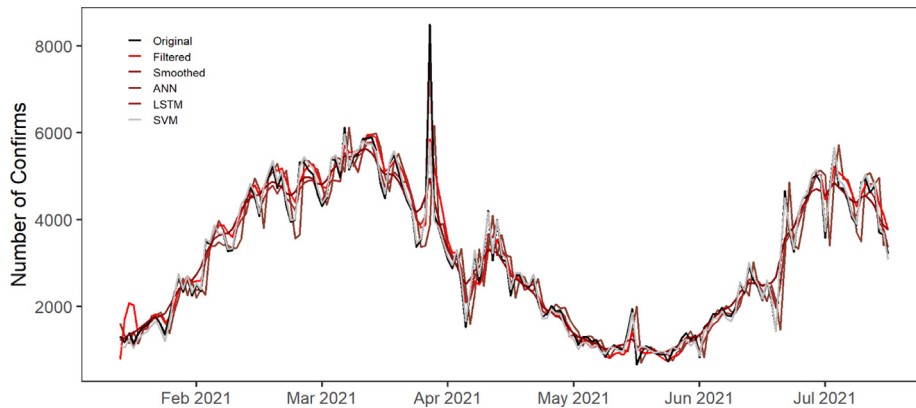
In Table 4, the results of recovered cases show the values of ME, RMSE, MPE and MAPE of LSTM are better than the Kalman Filter, ANN and SVM. While, the value of MAE of ANN is lower than the other methods. Therefore, LSTM is better choice for prediction as compared to the two methods. The evaluated five parameters of Kalman Filter, ANN, LSTM

and SVM models, it is indicates that LSTM is better in terms of low error rate among the KF, ANN and SVM. Furthermore, LSTM is a better choice for estimating the number of recovered cases of COVID-19 for Pakistan.

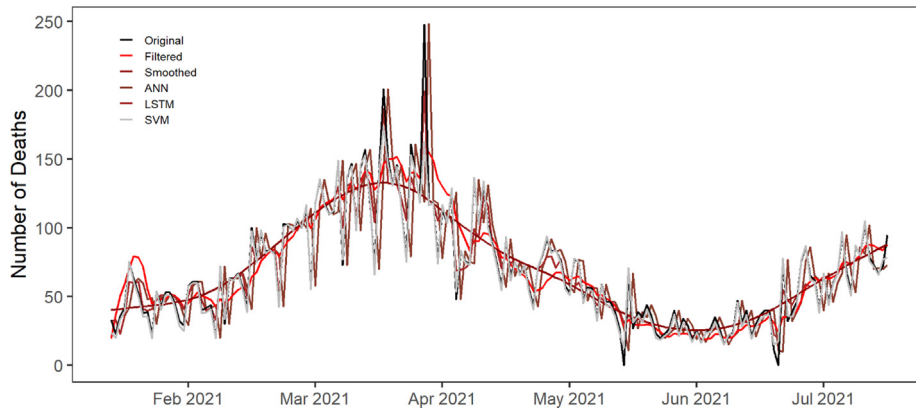
In Table 5, the results of confirmed cases used 80% training and 20% testing parts. The values of ME, RMSE, MPE and



**Fig. 4** Forecasting accuracy measures of KF, ANN, LSTM and SVM models for recovered cases.



**Fig. 5** Actual and predicted values of KF, ANN, LSTM and SVM models for confirmed cases of COVID-19 of the 70% training and 30% testing set.

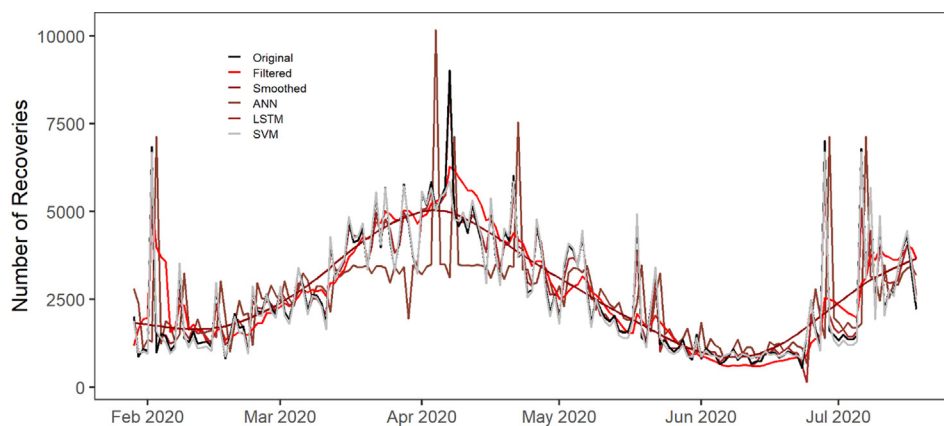


**Fig. 6** Actual and predicted values of KF, ANN, LSTM and SVM models for death cases of COVID-19 of the 70% training and 30% testing set.

MAPE of LSTM are better than the Kalman Filter, ANN and SVM. While, the MAE of ANN is better than Kalman filter, LSTM and SVM model. So, LSTM is a better choice for prediction. The parameters for the evaluation of Kalman Filter, ANN, LSTM and SVM models are given in the Table 5, the values indicates that LSTM is better in terms of low error rate among the KF, ANN and SVM. It shows that LSTM is a better choice for estimating the number of recovered cases of COVID-19 for Pakistan.

The Diebold-Mariano test proposed by [11] define the test of equality between the two forecasts under the null hypothesis  $H_0 : E(d) = 0$ . The results show that there is difference between all the methods in prediction. We can conclude that LSTM is better than the other methods as in the performance metrics Tables 3, 4, and 5. In the above Table 6, we included those and compared those models that are statistically significant.

The performance of ANN, KF, LSTM and SVM in bar chart is shown below.



**Fig. 7** Actual and predicted values of KF, ANN, LSTM and SVM models for recovered cases of COVID-19 of the 70% training and 30% testing set.

**Table 1** Forecasting performance metrics.

Criterion	Formula
Mean error	$ME = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)$
Standard deviation	$SD = \sqrt{\frac{\sum (Y_t - \bar{y})^2}{n-1}}$
Root mean square error	$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{Y}_t - Y_t)^2}$
Mean absolute error	$MAE = \frac{1}{n} \sum_{t=1}^n  \hat{Y}_t - Y_t $
Mean percentage error	$MPE = \frac{1}{n} \sum_{t=1}^n \left( \frac{\hat{Y}_t - Y_t}{Y_t} \right) * 100$
Mean absolute percentage error	$MAPE = \frac{1}{n} \sum_{t=1}^n \left  \frac{\hat{Y}_t - Y_t}{Y_t} \right  * 100$

**Table 2** Descriptive statistics of COVID-19 daily confirmed, death and recovered cases of Pakistan.

	Confirms	Deaths	Recoveries
Min.	0	0	0
1st Qu.	665.5	11	461.5
Median	1592	37	1282
Mean	1989.73	44.04	1840.60
3rd Qu.	2963.5	66	2402
Max.	12,073	313	19,772
SD	1674.98	40.35	2132.76

4.1. Discussion

The COVID-19 dataset, based from 22 January 2020 to 18 August 2021, used 70% training and 20% testing parts. In soft

computing, we use COVID-19 data to find models pattern and apply these learned models to forecast the different cases of COVID-19. In Fig. 8, the daily new confirm cases are forecasted for the upcoming three weeks. The filter and smooth model structures are able to produce epidemic bends that look like sound through the chronological time series. ANN fits the actual line throughout the series. While, LSTM fits an efficient way of the actual data. There are highest spikes in the month of July 2020 and the end of May 2021. Moreover, the ANN, KF, LSTM and SVM show that, the next three weeks forecast ahead showed that the number of confirmed cases are increased. (See Fig. 9.).

In Fig. 8, the number of daily confirmed cases is plotted. All the four models captured the real pattern of the data. The daily confirmed cases of COVID-19 for Pakistan is fluctuated with respect to time. There is high number of daily cases showed in the start of July 2020 and then again observed at the end of June 2021. In addition, the daily confirmed cases are forecasted for the three weeks using the four models. All the models shows that the new confirmed cases will be increased and then then decrease except SVM model at the end of last week.

In Fig. 10, the number of daily death cases is showed. The filtered and smoothed Kalman follow the original shape of the data. Similarly, the ANN and SVM follows the original data. LSTM tracked the original observations in all the cases. The highest number of cases were recorded at the end of November 2020, but after that, the number got reduced and then fluctuated again. The forecast of the four methods for three weeks indicates that daily death cases would increase in the upcoming days in Pakistan. (See Fig. 1. Fig. 2. Fig. 3. Fig. 4.).

In Fig. 11, number of recovered cases shown. Kalman Filter and ANN indicates fluctuations along time and captures the original shape of the data. In the same way, SVM and LSTM model also follow the real time line of the recovered

**Table 3** Daily confirmed cases using 70/30 percent training and testing data.

	ME	RMSE	MAE	MPE	MAPE
Kalman Filter	15.34	744.98	541.09	0.020	9.04
ANN	15.09	660.02	425.77	-1.96	8.30
LSTM	14.01	500.76	399.03	0.009	8.24
SVM	15.1	669.45	469.32	0.0234	9.009



**Table 4** Daily COVID-19 death cases using 70/30 percent training and testing data.

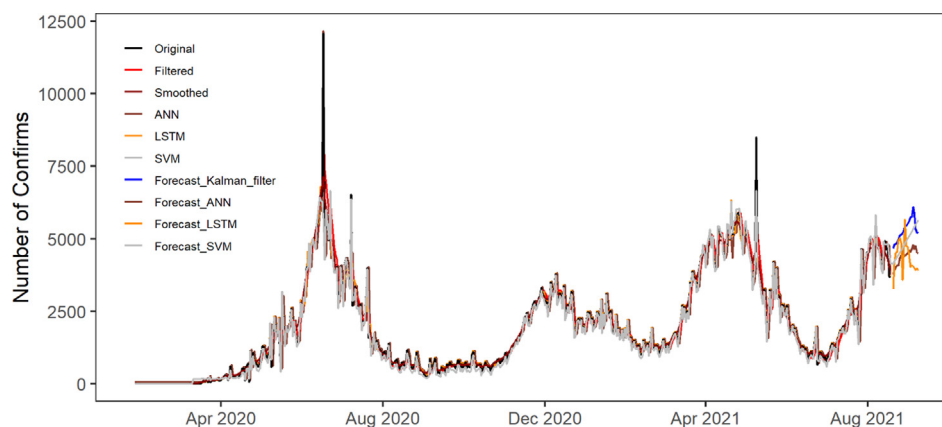
	ME	RMSE	MAE	MPE	MAPE
Kalman filter	0.263	23.05	17.03	0.010	0.356
ANN	-0.005	0.030	0.004	0.011	0.020
LSTM	0.023	0.027	0.002	0.004	0.018
SVM	80.44	73.55	37.223	4.543	12.559

**Table 5** Daily COVID-19 recovered cases using 70/30 percent training and testing data.

	ME	RMSE	MAE	MPE	MAPE
Kalman filter	6.87	1191.63	799.21	-0.110	8.69
ANN	6.19	1349.84	796.09	-1.57	7.35
LSTM	5.65	1089.45	808.09	1.09	4.98
SVM	7.54	1454.32	897.33	2.14	8.55

**Table 6** Daily COVID-19 confirmed cases using 80/20 percent training and testing data.

	ME	RMSE	MAE	MPE	MAPE
Kalman filter	-28.57	746.98	490.32	-0.018	19.05
ANN	-35.04	718.09	458.32	-4.77	16.02
LSTM	26.98	687.54	485.08	3.43	14.98
SVM	28	709	499	2.012	9.04453

**Fig. 8** Actual and three weeks ahead forecasted values of KF, ANN, LSTM and SVM models for confirmed cases of COVID-19 of the 70% training and 30% testing set.

cases of COVID-19. The number of daily recovered cases in Pakistan increases for the upcoming three weeks forecast for all the methods. (See Fig. 12.).

The predicted curve for the three weeks rises which is a good signal to decrease and control the COVID-19. The main results of the study are as under:

- The LSTM, Kalman Filter algorithm, ANN and SVM were capable to follow the original data pattern in almost all the situations in Pakistan.

- The LSTM, Kalman Filter, ANN and SVM forecast shows that the number of confirmed, death and recovered cases will increase in Pakistan in the next three weeks.
- LSTM provides better forecast almost in all cases as compared to Kalman Filter, ANN and SVM for the short term. Hence, policy makers may use LSTM for short term decision making.

Thus, the next three weeks summary of the confirmed, death, and recovered cases is presented in the following Table 7. (See Table 8.).

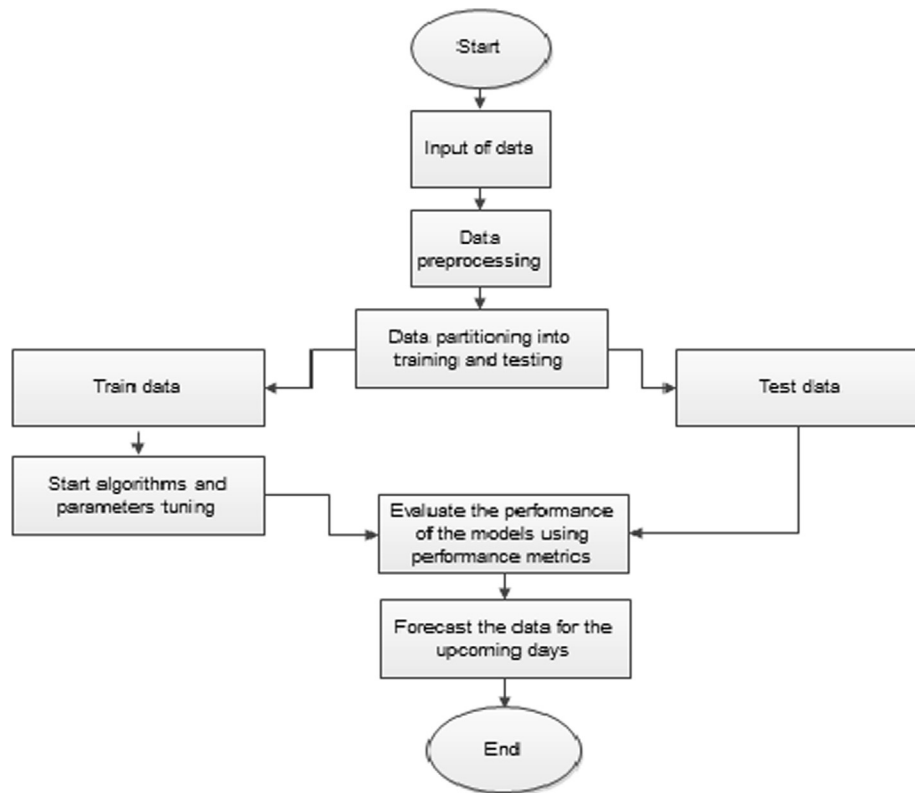


Fig. 9 Flowchart of algorithms for COVID-19 outbreak.

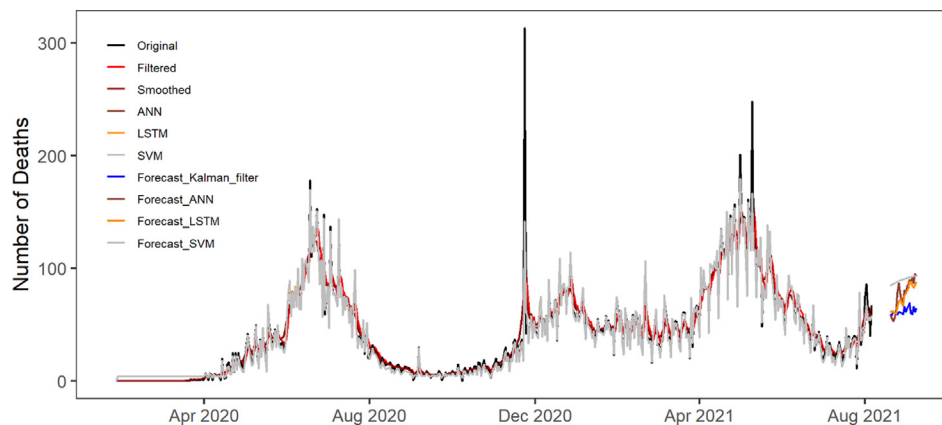


Fig. 10 Actual and three weeks ahead forecasted values of KF, ANN, LSTM and SVM models for death cases of COVID-19 of the 70% training and 30% testing set.

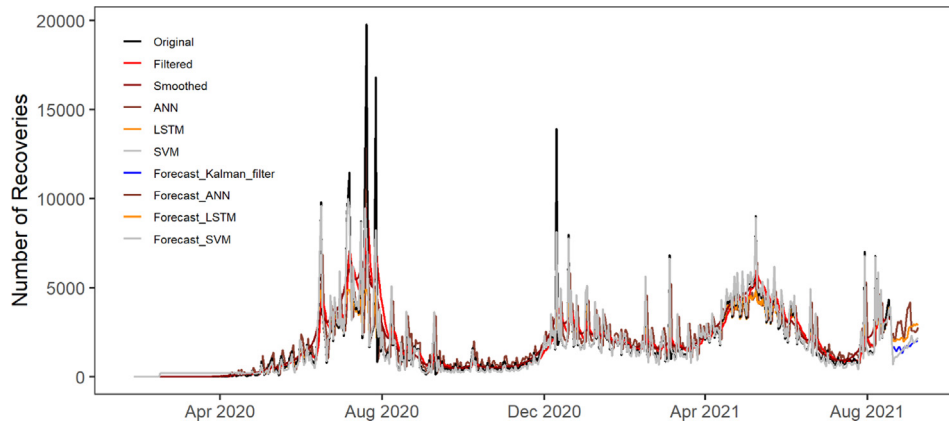
## 5. Conclusion

COVID-19 has shaken the world owing to its upsetting epidemic nature. Each and every country has to learn to live with this pandemic and make the precautionary measures for it as their part of life since this outbreak can occur for years to arise and we cannot close all the things for so long as it also suffers the economy of the country besides controlling the spreads of the disease.

In this study, we used four soft computing models to examine the upcoming trend of COVID-19 in Pakistan to interpret

the complex trend in time series by verifying with performance metrics, and finally, predicting the daily confirmed, death and recovered cases for the next three weeks.

LSTM, ANN, Kalman Filter and SVM methods for the time series fluctuations predictions give good results. LSTM, ANN, Kalman Filter and SVM have better fluctuation in the time series estimation. In this study, the LSTM, ANN, KF and SVM is implemented to determine the trend of the pandemic COVID-19. In addition, LSTM, ANN, KF and SVM methods were used to forecast the number of confirm, recovered and death cases in Pakistan. COVID-19 was well tracked



**Fig. 11** Actual and three weeks ahead forecasted values of KF, ANN, LSTM and SVM models for recovered cases of COVID-19 of the 70% training and 30% testing set.

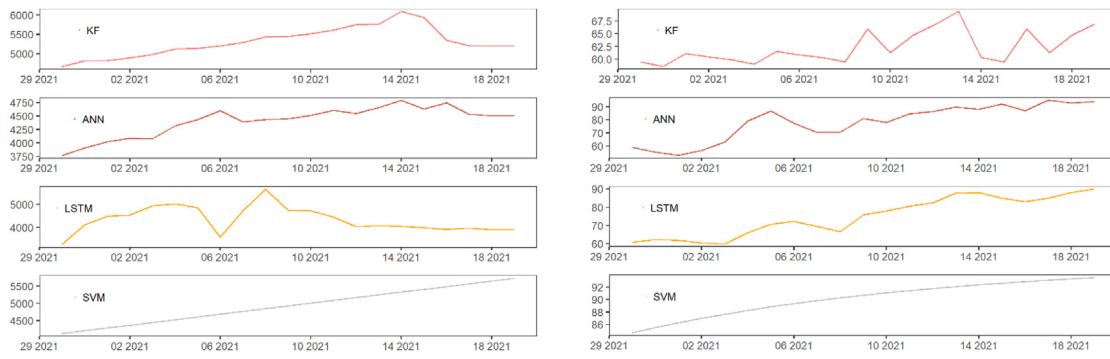
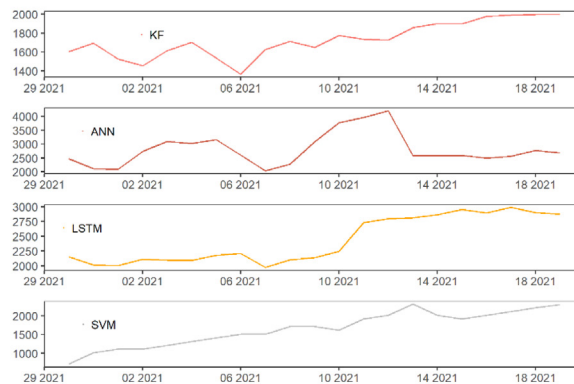


Figure 12. (a) Three weeks forecasted plots for confirmed cases of KF, ANN, LSTM and SVM (b) Three weeks forecasted plots for death cases of KF, ANN, LSTM and SVM



(c) Three weeks forecasted plots for recovered cases of KF, ANN, LSTM and SVM

**Fig. 12** (a) Three weeks forecasted plots for confirmed cases of KF, ANN, LSTM and SVM.

by LSTM, ANN, SVM and KF algorithms. The number of confirm, recovered, and death cases in Pakistan for the upcoming three weeks were predicted. LSTM, ANN, KF and SVM algorithms are helpful in the short-term forecasting for coronavirus. In addition, LSTM, ANN, KF and SVM methods can help decision-makers in formulating short term strategies about the current number of disease occurrence. In this study, all the methods showed that the COVID-19 cases are increased

in Pakistan based on 22 Jan 2020 to 18 August 2021 data. We observed that it was because of government relaxation in lockdowns, lack of awareness in the people, people didn't stop their jobs/duties because of poverty and many other factors. And now, all the COVID-19 cases are decreased, as because of most of the people are vaccinated.

In conclusion, the current study is the first comparative study using the soft computing techniques for Pakistan to pre-

**Table 7** Diabold Mariano (DM) test for the models.

Forecast Models	DM values	p-values
KF-ANN	-3.085	0.00237
KF-LSTM	6.155	5.204e-09
KF-SVM	4.768	0.00065
KF-LSTM	5.769	4.65e-05
ANN-LSTM	2.391	0.01788

**Table 8** Summary of the forecasts for the next three weeks.

Countries\Cases	Confirm	Recover	Deaths
Pakistan	Increase	Increase	Increase

dict the future time line of COVID-19. Our study will come to be useful for the country to take all preventive measurements before receiving subjugated by COVID-19.

## 6. Limitations

We did not consider the vaccinated people and the number of new dailies confirm death and recovered cases. Moreover, we did not measure the number of lockdowns in the cities and across the country during the pandemic.

## 7. Future work

Future work may focus on vaccinated persons and the daily confirm deaths or recovered cases. Comparisons can also be drawn between the types of viruses and the number of daily confirmed, death and recovered cases. Moreover, we can use some features selection techniques that might be useful in the early detection of COVID-19. We can also study the total economic destruction of the country during the COVID-19 in various economic sectors and create a suitable strategy to improve it, which can support the country to grow up economy rate. We design to estimate the possible COVID-19 cases for other countries besides possible viral transmission of COVID-19 can be proved.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups RGP.2/212/1443.

## References

- [2] M.A. Al-qaness, A.A. Ewees, H. Fan, M. Abd El Aziz, Optimization method for forecasting confirmed cases of COVID-19 in China, *J. Clin. Med.* 9 (3) (2020) 674.
- [3] M.A.A. Al-qaness, A.I. Saba, A.H. Elsheikh, M.A. Elaziz, R.A. Ibrahim, S. Lu, A.A. Hemedan, S. Shanmugan, A.A. Ewees, Efficient artificial intelligence forecasting models for COVID-19 outbreak in Russia and Brazil, *Process Saf. Environ. Prot.* 149 (2021) 399–409.
- [4] Ali, A., Mashwani, W. K., Naeem, S., Uddin, M. I., Kumam, W., Kumam, P., . . . Chesneau, C. (2021). COVID-19 infected lung computed tomography segmentation and supervised classification approach.
- [5] A.M. Almars, M. Almaliki, T.H. Noor, M.M. Alwateer, E. Atlam, HANN: hybrid attention neural network for detecting covid-19 related rumors, *IEEE Access* 10 (2022) 12334–12344.
- [8] B.D. Anderson, J.B. Moore, *Optimal filtering*, Courier Corporation, 2012.
- [9] A.P. Andrews, *Kalman Filtering: theory and practice*, Wiley, Using MATLAB, 2001.
- [10] R.S. Bucy, P.D. Joseph, *Filtering for stochastic processes with applications to guidance*, (Vol. 326):, American Mathematical Soc, 2005.
- [11] F. Diebold, R. Mariano, C.P. Accuracy, C.F. Baum, data before using {cmd: dmario}; see help {help tstset}, *J. Business and Economic Statistics* 13 (3) (1995) 253–263.
- [12] J. Durbin, S.J. Koopman, *Time series analysis by state space methods*, Oxford University Press, 2012.
- [14] Elmousalami, H. H., & Hassani, A. E. (2020). Day level forecasting for Coronavirus Disease (COVID-19) spread: analysis, modeling and recommendations. *arXiv preprint arXiv:2003.07778*.
- [15] F.A. Gers, D. Eck, J. Schmidhuber, Applying LSTM to time series predictable through time-window approaches *Neural Nets WTRN Vietri-01*, Springer, 2002, pp. 193–200.
- [17] A.C. Harvey (Ed.), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, 1990.
- [18] W. Huang, K.K. Lai, Y. Nakamori, S. Wang, L. Yu, Neural networks in finance and economics forecasting, *Int. J. Information Technol. Decision Making* 06 (01) (2007) 113–140.
- [19] Kasinathan, P., Montoya, O. D., Gil-González, W., Arul, R., Moovendan, M., Dhivya, S., . . . Angalaeswari, S. (2020). APPLICATION OF SOFT COMPUTING TECHNIQUES IN THE ANALYSIS OF COVID-19: A REVIEW. *European Journal of Molecular & Clinical Medicine*, 7(6), 2480-2503.
- [20] K.-J. Kim, Financial time series forecasting using support vector machines, *Neurocomputing* 55 (1-2) (2003) 307–319.
- [21] A.J. Kucharski, T.W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R.M. Eggo, F. Sun, M. Jit, J.D. Munday, N. Davies, A. Gimma, K. van Zandvoort, H. Gibbs, J. Hellewell, C.I. Jarvis, S. Clifford, B.J. Quilty, N.I. Bosse, S. Abbott, P. Klepac, S. Flasche, Early dynamics of transmission and control of COVID-19: a mathematical modelling study, *Lancet. Infect. Dis* 20 (5) (2020) 553–558.
- [22] L. Li, Z. Yang, Z. Dang, C. Meng, J. Huang, H. Meng, D. Wang, G. Chen, J. Zhang, H. Peng, Y. Shao, Propagation analysis and prediction of the COVID-19, *Infectious Dis. Modelling* 5 (2020) 282–292.
- [24] Y. Li, Z. Zhu, D. Kong, H. Han, Y. Zhao, EA-LSTM: evolutionary attention-based LSTM for time series prediction, *Knowl.-Based Syst.* 181 (2019) 104785, <https://doi.org/10.1016/j.knsys.2019.05.028>.
- [25] Q. Lin, S. Zhao, D. Gao, Y. Lou, S. Yang, S.S. Musa, L. Yang, A conceptual model for the outbreak of Coronavirus disease 2019 (COVID-19) in Wuhan, China with individual reaction and governmental action, *Int. J. Infectious Dis.* (2020).
- [26] J. Luo, Z. Zhang, Y. Fu, F. Rao, Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms, *Results Phys.* 27 (2021) 104462, <https://doi.org/10.1016/j.rinp.2021.104462>.

- [27] S. Makridakis, A. Wakefield, R. Kirkham, Predicting medical risks and appreciating uncertainty, *Foresight: Int. J. Appl.* 52 (2019) 28–35.
- [28] Z. Malki, E. Atlam, G. Dagnev, A.R. Alzighaibi, E. Ghada, I. Gad, Bidirectional residual LSTM-based human activity recognition, *Comput. Information Sci.* 13 (3) (2020) 40.
- [29] M. Naeem, J. Yu, M. Aamir, S.A. Khan, O. Adeleye, Z. Khan, Comparative analysis of machine learning approaches to analyze and predict the COVID-19 outbreak, *PeerJ. Comput. Sci.* 7 (2021) e746.
- [30] M.A. Ozdemir, M. Degirmenci, E. Izci, A. Akan, EEG-based emotion recognition with deep convolutional neural networks, *Biomed. Eng./Biomed. Technik* 66 (1) (2021) 43–57.
- [31] M. Qi, 18 Financial applications of artificial neural networks, *Handbook of Statistics* 14 (1996) 529–552.
- [32] J. Shaman, A. Karspeck, Forecasting seasonal outbreaks of influenza, *Proc. Natl. Acad. Sci.* 109 (50) (2012) 20425–20430.
- [33] S. Shastri, A. Sharma, V. Mansotra, A model for forecasting tourists arrival in J&K, India. *Int. J. Comput. Appl.* 129 (15) (2015) 32–36.
- [34] S. Shastri, K. Singh, S. Kumar, P. Kour, V. Mansotra, Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study, *Chaos, Solitons Fractals* 140 (2020) 110227, <https://doi.org/10.1016/j.chaos.2020.110227>.
- [35] I. Shiroky, A. Papangelo, N. Hoffmann, O. Gendelman, Nucleation and propagation of excitation fronts in self-excited systems, *Physica. D* 401 (2020) 132176.
- [36] R.H. Shumway, D.S. Stoffer, *Time series analysis and its applications: with R examples*, Springer, 2017.
- [37] D. Simon, *Optimal state estimation: kalman, H infinity, and nonlinear approaches*, John Wiley & Sons, 2006.
- [38] K. Singh, S. Shastri, A.S. Bhadwal, P. Kour, M. Kumari, A. Sharma, V. Mansotra, Implementation of exponential smoothing for forecasting time series data, *Int. J. Sci. Res. Comput. Sci. Appl. Manag. Stud.* 8 (1) (2019) 1–8.
- [39] U. Thissen, R. van Brakel, A.P. de Weijer, W.J. Melssen, L.M. C. Buydens, Using support vector machines for time series prediction, *Chemometrics and intelligent laboratory systems* 69 (1-2) (2003) 35–49.
- [40] G.-G. Wang, M. Lu, Y.-Q. Dong, X.-J. Zhao, Self-adaptive extreme learning machine, *Neural Comput. Appl.* 27 (2) (2016) 291–303.
- [41] A. Waris, U.K. Atta, M. Ali, A. Asmat, A. Baset, COVID-19 outbreak: current scenario of Pakistan, *New Microbes and New Infections* 35 (2020) 100681, <https://doi.org/10.1016/j.nmni.2020.100681>.
- [42] Yadav, A., Jha, C., & Sharan, A. (2020). Optimizing LSTM for time series prediction in Indian stock market. *Procedia Computer Science*, 167, 2091-2100.
- [43] Yi, J.-H., Wang, J., & Wang, G.-G. (2016). Improved probabilistic neural networks with self-adaptive strategies for transformer fault diagnosis problem. *Advances in Mechanical Engineering*, 8(1), 1687814015624832.
- [44] M. Yousaf, S. Zahir, M. Riaz, S.M. Hussain, K. Shah, Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan, *Chaos, Solitons Fractals* 138 (2020) 109926, <https://doi.org/10.1016/j.chaos.2020.109926>.
- [45] S.A.Y. Zai, A.S. Akhunzada, Challenges and responses of higher education institutions to align higher education with COVID-19 Realities: a thematic analysis, *Int. J. Distance Edu. E-Learn.* 6 (1) (2020) 109–129.
- [47] Zhao, Y., & Liu, T. (2007). Zhao. *Berry fruit, value-added products for health promotion*.

### Further reading

- [1] Ahmadini, A. A. H., Naeem, M., Aamir, M., Dewan, R., Alshaq, S. S. A., & Mashwani, W. K. (2021). Analysis and Forecast of the Number of Deaths, Recovered Cases, and Confirmed Cases From COVID-19 for the Top Four Affected Countries Using Kalman Filter. *Frontiers in Physics*, NA-NA.
- [6] A.M. Almars, E.-S. Atlam, T.H. Noor, G. Elmarhomy, R. Alagamy, I. Gad, Users opinion and emotion understanding in social media regarding COVID-19 vaccine, *Computing* 104 (6) (2022) 1481–1496.
- [7] C. Anastassopoulou, L. Russo, A. Tsakris, C. Siettos, S. Othumpangat, Data-based analysis, modelling and forecasting of the COVID-19 outbreak, *PLoS ONE* 15 (3) (2020) e0230405.
- [13] B. Bektas Ekici, U.T. Aksoy, Prediction of building energy needs in early stage of design by using ANFIS, *Expert Syst. Appl.* 38 (5) (2011) 5352–5358.
- [16] I. Guler, E.D. Ubeyli, An expert system for detection of electrocardiographic changes in patients with partial epilepsy using wavelet-based neural networks, *Expert Systems* 22 (2) (2005) 62–71.
- [23] M.-T. Li, G.-Q. Sun, J. Zhang, Y. Zhao, X. Pei, L. Li, Z. Jin, Analysis of COVID-19 transmission in Shanxi Province with discrete time imported cases, *Math. Biosci. Eng* 17 (4) (2020) 3710.
- [46] G. Zhang, B. Eddy Patuwo, M. Y. Hu, Forecasting with artificial neural networks: The state of the art, *Int. J. Forecast.* 14 (1) (1998) 35–62.