



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.



Deep learning-based forecasting model for COVID-19 outbreak in Saudi Arabia



Ammar H. Elsheikh^{a,*}, Amal I. Saba^b, Mohamed Abd Elaziz^c, Songfeng Lu^d, S. Shanmugan^e, T. Muthuramalingam^f, Ravinder Kumar^g, Ahmed O. Mosleh^h, F.A. Essaⁱ, Taher A. Shehabeldeenⁱ

^a Department of Production Engineering and Mechanical Design, Faculty of Engineering, Tanta University, Tanta, 31527, Egypt

^b Department of Histology, Faculty of Medicine, Tanta University, Tanta, 31527, Egypt

^c Department of Mathematics, Faculty of Science, Zagazig University, Zagazig, Egypt

^d School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, 430074, China

^e Research Centre for Solar Energy, Department of Physics, Koneru Lakshmaiah Education Foundation, Green Fields, Guntur District, Vaddeswaram, Andhra Pradesh, 522502, India

^f Department of Mechatronics Engineering, Kattankulathur Campus, SRM Institute of Science and Technology, Chennai, 603203, India

^g Department of Mechanical Engineering, Lovely Professional University, Phagwara, Jalandhar, 144411, Punjab, India

^h Shoubra Faculty of Engineering, Benha University, Shoubra St. 108, Shoubra, P.O. 11629, Cairo, Egypt

ⁱ Mechanical Engineering Department, Faculty of Engineering, Kafrelsheikh University, Kafrelsheikh, 33516, Egypt

ARTICLE INFO

Article history:

Received 9 June 2020

Received in revised form 22 October 2020

Accepted 23 October 2020

Available online 1 November 2020

Keywords:

COVID-19

Forecasting

Deep learning

Saudi Arabia

ABSTRACT

COVID-19 outbreak has become a global pandemic that affected more than 200 countries. Predicting the epidemiological behavior of this outbreak has a vital role to prevent its spreading. In this study, long short-term memory (LSTM) network as a robust deep learning model is proposed to forecast the number of total confirmed cases, total recovered cases, and total deaths in Saudi Arabia. The model was trained using the official reported data. The optimal values of the model's parameters that maximize the forecasting accuracy were determined. The forecasting accuracy of the model was assessed using seven statistical assessment criteria, namely, root mean square error (RMSE), coefficient of determination (R^2), mean absolute error (MAE), efficiency coefficient (EC), overall index (OI), coefficient of variation (COV), and coefficient of residual mass (CRM). A reasonable forecasting accuracy was obtained. The forecasting accuracy of the suggested model is compared with two other models. The first is a statistical based model called autoregressive integrated moving average (ARIMA). The second is an artificial intelligence based model called nonlinear autoregressive artificial neural networks (NARANN). Finally, the proposed LSTM model was applied to forecast the total number of confirmed cases as well as deaths in six different countries; Brazil, India, Saudi Arabia, South Africa, Spain, and USA. These countries have different epidemic trends as they apply different policies and have different age structure, weather, and culture. The social distancing and protection measures applied in different countries are assumed to be maintained during the forecasting period. The obtained results may help policymakers to control the disease and to put strategic plans to organize Hajj and the closure periods of the schools and universities.

© 2020 Institution of Chemical Engineers. Published by Elsevier B.V. All rights reserved.

1. Introduction

In December 2019, Wuhan, Hubei province, China has been reported as the center of the COVID-19 outbreak (Collivignarelli et al., 2020). Three months later, that outbreak was declared as a global pandemic by the world health organization (Shi et al., 2020).

More than 37.40 million confirmed COVID-19 cases and one million deaths worldwide have been officially reported in 10 October 2020. Covid-19 is a fatal and contagious disease that has spread to more than 200 countries. Therefore, it has been considered as the most critical universal crisis since the World War-II (Boccaletti et al., 2020).

COVID-19 infection affects respiratory system and has many symptoms such as fever, cough, flu, diarrhea, headache, myalgia, and dyspnea. It is a highly contagious disease with moderate fatality rate. It transmits among humans via touching contaminated bod-

* Corresponding author.

E-mail address: ammal_elsheikh@f-eng.tanta.edu.eg (A.H. Elsheikh).

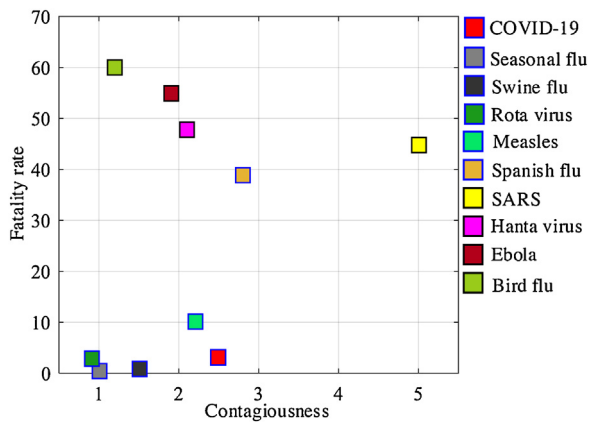


Fig. 1. Contagiousness and fatality rate of different communicable diseases (Thomas-Rüddel et al., 2020).

ies with viral particles or contacting infected patients (Wu et al., 2020). The viral particles may exist for long period on solid surfaces ranges from 12 h to two days. Moreover, the incubation period of this disease ranges from one day to fourteen days (Kang et al., 2020). The main problem related to the prevalence of this disease is that it may transmit from diseased carriers which have not any disease symptoms to other contacted persons (Cao et al., 2020). Infected patients with severe symptoms may be subjected to respiratory failure which requires artificial ventilation. Severe complications may occur in elderly people with other health problems such as hepatic disease, hypertension, diabetes, bronchial asthma, cancer, autoimmune, and cardiovascular diseases. The fatality rate of COVID-19 is about 2.89 % as reported by WHO on 10 October 2020 which is a moderate rate compared with other communicable diseases as shown in Fig. 1.

The prediction of the spreading of COVID-19 disease as well as investigating its epidemiological features are serious topics that should be seriously studied to control the outbreak prevalence, regulate production activities, and allocate medical resources (Cássaro and Pires, 2020; Materassi, 2019; Qi et al., 2020; Xu et al., 2020).

Statistical and mathematical modeling approaches have been employed to predict the spreading of the outbreak in different countries. COVID-19 epidemic curve was determined by applying the exponential growth method (Zhao et al., 2020). The confirmed cases were estimated to increase by about 2000 % within less than one month. The COVID-19 under-ascertainment rate was computed considering the evacuation flights data of Japanese inhabitants from Wuhan in the last week of January 2020 using mean serial interval method (Nishiura et al., 2020). The patient information based approach was employed to determine the death rate of COVID-19 in China (Wang et al., 2020). The proposed model was utilized to forecast case fatality, severe cases increasing rate, and the death rate of the infected patients.

Auto-regressive integrated moving average (ARIMA) is a statistical-based approach that used in time series forecasting problems. ARIMA has been proposed by many researchers to forecast the epidemiological behavior of different diseases such as influenza viruses (He and Tao, 2018), SARS (Earnest et al., 2005), and HIV (Yu et al., 2013). An ARIMA approach was employed to predict the confirmed and recovered cases of COVID-19 as well as the number of deaths in Pakistan (Yousaf et al., 2020). The authors warned against the severe prevalence of the outbreak due to the lack of medical facilities and social gathering restriction rules. (Abdulmajeed et al., 2020) used the same approach to forecast the COVID-19 confirmed cases in Nigeria and South Africa. The number of forecasted cases in South Africa was higher than Nigeria. This is due to two main reasons; the first is the existence of much more infected cases in South

Africa than Nigeria, the second is the more tests carried out in South Africa compared with Nigeria. Seasonal ARIMA forecasting package with R statistical approach was implemented by (Chintalapudi et al., 2020a) to forecast the total and recovered cases in Italy after two months of lockdown. It was recommended to extend the quarantine period and prohibit all transportations between cities. The effect of different ARIMA parameters on the prediction accuracy of COVID-19 cases in Spain, Italy, and France was investigated by (Ceylan, 2020).

In another study, ARIMA was compared with five other statistical approaches to predict the COVID-19 prevalence in Brazil (Ribeiro et al., 2020). These models are random decision forests approach, stacking-ensemble learning model, ridge regression, cubist regression, and support vector regression. The ranking of tested models, from the worst to the best regarding forecasting accuracy, is random decision forests approach, ridge regression, cubist regression, ARIMA, stacking-ensemble learning approach, and support vector regression.

The incorporation between ARIMA and α -Sutte indicator method to forecast the confirmed cases in Spain was proposed by (Ahmar and del Val, 2020). The proposed hybrid approach succeeded to predict the COVID-19 cases with a higher prediction accuracy compared with conventional ARIMA approach. The mean absolute percentage error for the predicted cases was 0.066 for conventional ARIMA approach and only 0.036 for the proposed hybrid approach. Another incorporation between wavelet-based forecasting and ARIMA was presented by (Chakraborty and Ghosh, 2020) to forecast of the daily COVID-19 cases in United Kingdom, France, Canada, South Korea, and India. The forecasted results revealed the importance of social distancing measures in controlling the prevalence rate of the disease. Moreover, the obtained results may help government policymakers in preparing allocation plans of health care resources. It was also recommended to extend the lockdown period in case of sharply increase in the number of confirmed cases and deaths. The incorporation between discrete wavelet decomposition and ARIMA to forecast deaths due to COVID-19 in France, Spain, Italy, United Kingdom, and USA was presented by (Singh et al., 2020). The results obtained by the proposed hybrid model reduced the forecasting errors by about 50 % compared with standalone ARIMA model. The performance of the suggested hybrid model is improved by about 80 % for Spain, Italy, and United Kingdom and by about 50 % for USA and France compared with that of ARIMA model.

The incorporation between reduced-space Gaussian process regression and chaotic Bayesian approach to predict the cases of COVID-19 and deaths in USA was presented by (Arias Velásquez and Mejía Lara, 2020). The prevalence of the outbreak was expected to dramatically increase and thus new restricted quarantine actions should be implemented.

An autoregressive time series model incorporated with two-piece scale mixture normal distributions was employed to forecast the COVID-19 cases in U.S.A, Spain, Italy, and Iran (Maleki et al., 2020). The proposed hybrid model had reasonable forecasting accuracy compared with other forecasting models based on statistical selection approaches such as Bayesian information approach, Akaike information approach, and Box–Pierce and Ljung–Box tests. The total cases of COVID-19 in Italy, United Kingdom and USA were forecasted using three different grey prediction models (Utkucan and Tezcan, 2020). These models are nonlinear and fractional nonlinear grey Bernoulli models in addition to conventional grey model. The fractional nonlinear model had the best prediction accuracy compared to the others based on different statistical measures such as root mean squared error, mean absolute percentage error, and coefficient of correlation. The increase in the prevalence outbreak was expected to continue until the end of May 2020. Ecological niche models were utilized to detect and forecast the

potential risk regions of the COVID-19 infection in three Chinese cities Beijing, Shenzhen, and Guangzhou (Ren et al., 2020). Transmission rates of COVID-19 outbreak in Marche, Italy was forecasted using R statistics approach (Chintalapudi et al., 2020b). Daily and total cases for the next month were also forecasted; which provide good insights into transmission dynamics of the virus to implement official regulations to reduce the spreading of the outbreak. In another study, susceptible exposed infectious recovered approach was employed to predict the infected COVID-19 cases in Delhi, India (Marimuthu et al., 2020). The proposed model is a compartmental model which used historical data of Wuhan, China to predict the prevalence of the outbreak in Delhi.

Artificial neural networks (ANN) are powerful information processing paradigms that mimic the human brain in processing data. ANNs have been employed to model different engineering problems (Babikir et al., 2019; Elsheikh et al., 2020b; Shehabeldeen et al., 2019). ANN has a number of advantages over other traditional modeling approaches such as handling enormous amounts of data, generalization capabilities, identifying complex relationships between dependent and independent variables, and detecting the inherent interactions between process variables (Elaziz et al., 2019).

A NARANN model was applied to forecast the total COVID-19 cases in Egypt (Saba and Elsheikh, 2020). The model has a better accuracy than that of conventional ARIMA model. Some policies were recommended by the authors to reduce the prevalence of the outbreak such as; increasing the overnight curfew period, applying more restricted lockdown regulations, establishing more quarantine hospitals, prohibiting all types of public transportation between Egyptian governorates, and extending closure period of mosques, schools, and universities.

LSTM neural network has been employed to predict the prevalence trends and to estimate the stopping time of the outbreak in Canada (Chimmula and Zhang, 2020). The transmission rate of the disease in Canada is compared with that of USA and Italy. The rate of infected cases has a linear tendency in Canada, while it has an exponential tendency in USA and Italy. It was expected that the number of daily cases will be declined due to the following of Canadians to the rules and regulations issued by the government. Genetic evolutionary programming has been applied to forecast the spread of COVID-19 outbreak in India (Salgotra et al., 2020). The proposed model is highly reliable in forecasting both total cases and death cases. It was recommended to strictly apply social distancing and lockdown so that the outbreak can be controlled.

A hybrid artificial intelligence-based model was developed by (Al-qaness et al., 2020a) to predict the COVID-19 prevalence in China. The proposed hybrid model is composed of adaptive neuro-fuzzy inference system incorporated with two different metaheuristic optimization approaches; salp swarm and flower pollination algorithms. The former algorithm is used to prevent flower pollination algorithm to be trapped into the local optima; and the enhanced flower pollination algorithm is used to predict the optimal parameters of the prediction model. The hybrid model had a better accuracy than the standalone model. Another attempt to improve the forecasting capabilities of the neuro-fuzzy system was done by (Al-qaness et al., 2020b) via incorporation with marine predators algorithm. The suggested model was employed to predict the COVID-19 cases in Italy, South Korea, Iran, and USA. A high coefficient of correlation was obtained for all forecasted results for all investigated countries; it was 96.48 %, 98.74 %, 98.59 %, and 96.96 % for South Korea, Iran, Italy, and USA, respectively.

In this study, long short-term memory (LSTM) neural network as a deep learning model is proposed to forecast the total number of confirmed cases, total recovered cases, and total deaths in Saudi Arabia. The forecasting accuracy was assessed using different statistical assessment criteria. The optimal hidden units' number

and initial learning rate value that maximize the forecasting accuracy are determined. The network with the optimal parameters was used to forecast the COVID-19 for three weeks ahead.

2. Study area

Saudi Arabia is the fifth-largest country in Asia with a total area of about 2.15 million km² and total population of 34.77 million people. It lies between longitudes 34° and 56 °E and latitudes 16° and 33 °N. Saudi Arabia is boarded by Iraq and Jordan to the north, Yemen to the south, Qatar, the United Arab Emirates, and Bahrain to the east, Oman to the southeast, and Kuwait to the northeast. It is also bounded by the Red Sea and the Arabian Gulf to the west and east, respectively. It has the largest economy in the Middle East. Saudi Arabia is the second largest oil producer (12 million bbl/day) and is the first largest exporting country (10.6 million bbl/day) in the world. It has one of the youngest populations in the world: 50 % of its population is under 25 years old. It hosts one of the Islam's pillars, which obliges all Muslims all over the world to pilgrimage to Mecca. The number of pilgrims in 2019 Hajj was officially reported as 2.48 million people from different countries all over the world. The total number of Umrah performers in 2019 was 18.31 million people. About 30 % of Umrah performers were over 50 years old. Overcrowding due to the increased pilgrim numbers in recent years causes numerous accidents.

Saudi Arabia has a national health care system in which health care services are provided through government agencies. Expenditure on healthcare in Saudi Arabia as a percentage of gross domestic product is 5.23 %, which is one of the highest expenditure among all Near East countries.

Saudi Arabia is considered as one of the most COVID-19 hotspots in Asia as it has 9408 confirmed cases per million people reported by the Saudi ministry of health in 17 September 2020 as shown in Fig. 2. It has 56,106 total confirmed cases. Riyadh, the capital of Saudi Arabia, was registered the highest number of COVID-19 cases followed by Jeddah, Mecca, and Medina. The cases in these four cities represent more than 70 % of all confirmed cases in Saudi Arabia as shown in Fig. 3(a). Very high recovery rates were recorded in Saudi Arabia; as the daily reported cases (1881 cases) is balanced with the daily reported recoveries (1864 cases) in 1 June 2020. The daily reported recoveries (1203) exceeded the daily confirmed cases (593) in 16 September 2020. Moreover, 93.61 % of the total reported cases were recovered in 16 September 2020. The recovered cases distribution in different Saudi Arabia cities is shown in Fig. 3(b). The most important observed phenomena in COVID-19 outbreak in Saudi Arabia is that; however it has one of the highest total cases per million people in the world, it also has one of the lowest fatality rate in the world as presented in Fig. 4. The total number of active cases in Saudi Arabia began to decline in 10 May 2020 due to the firm rules implemented by the governorates which could be summarized as follows:

- Closing Medina and Mecca against Umrah pilgrims.
- Suspending the issue of Umrah visas.
- Closing all mosques across the country.
- Closing all schools and universities.
- Suspending all international flights.
- Quarantining all repatriated citizens.
- Suspending all taxis, buses, trains, and domestic flights.
- Applying full curfew (24-h) in hotspot provinces.
- Banning all public gathering and social events.
- Suspending all sports competitions.
- Reducing cost-of-living allowance by about 250 \$/month.
- Increasing the value-added tax by about 10 %.

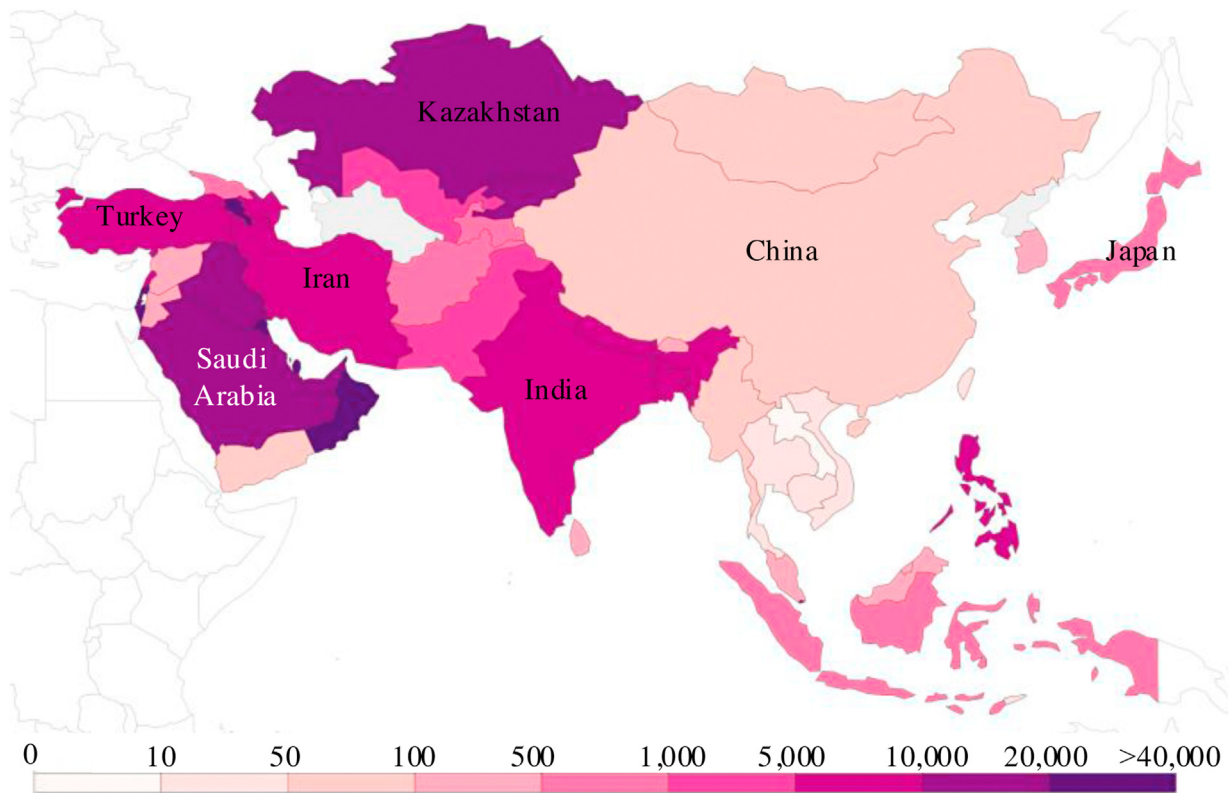


Fig. 2. Total confirmed COVID-19 cases per million people in Asia countries, 10 October 2020.

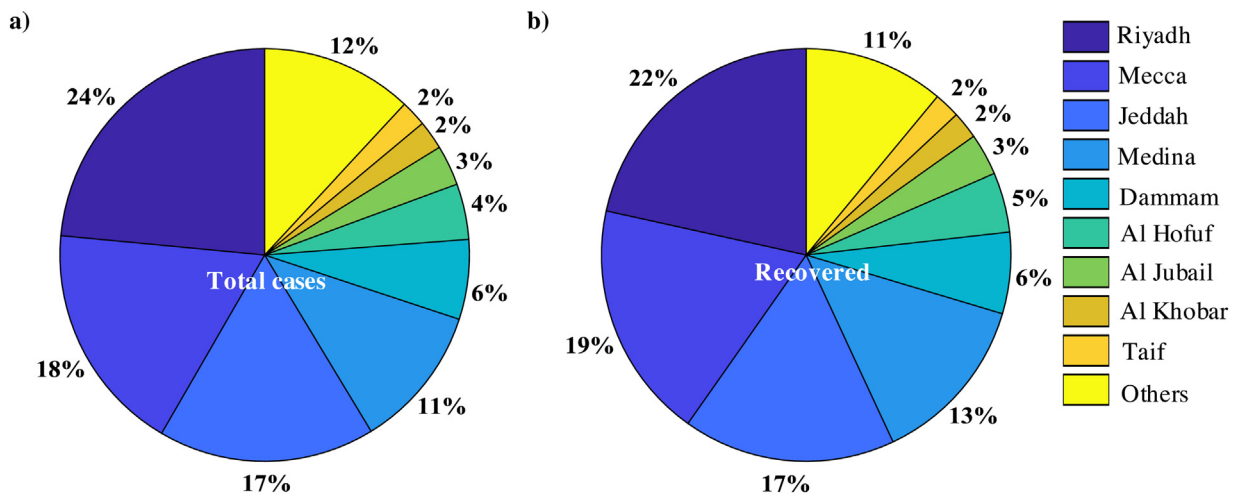


Fig. 3. Pie chart showing the COVID-19 distribution in Saudi Arabia cities (10 October 2020): a) total cases; b) recovered.

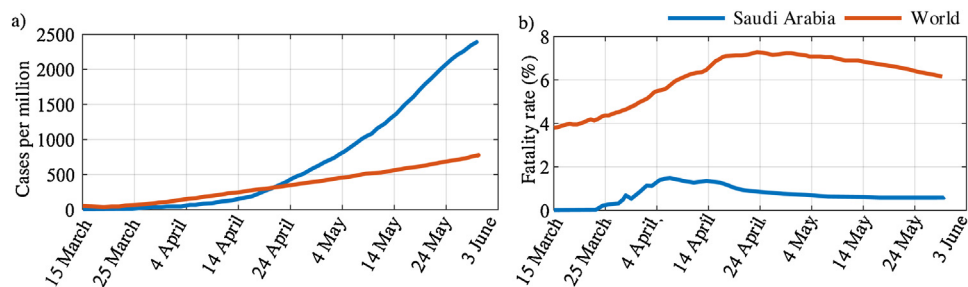


Fig. 4. a) Total COVID-19 cases per million people; b) Fatality rate due to COVID-19.

3. Data collection

The official reported COVID-19 confirmed cases, recovered cases, and deaths by the Saudi ministry of health (<https://covid19.moh.gov.sa/>) was utilized as a time series data to train the suggested models. Three periods have been considered in this study; from 2 March to 31 May, 2020, from 2 March to 15 September, 2020, and from 1 January to 10 October, 2020. In the first period, the total number of confirmed cases, recoveries, and deaths declared by the Saudi ministry of health was used as model inputs of LSTM, NARANN, and ARIMA models. Then the proposed models were trained and used to forecast the total number of confirmed cases, recoveries, and deaths for the next ten days. The optimal parameters of LSTM were selected using the forecasting results of this period. In the second period, the total number of confirmed cases, recoveries, and deaths reported by the Saudi ministry of health was used as model inputs of LSTM to forecast the prevalence of the epidemic for three weeks with a confidence level of 95 %. In the third period, the total number of confirmed cases as well as the total number of deaths in six different countries (Brazil, India, Saudi Arabia, South Africa, Spain, and USA) was used as model inputs of LSTM to forecast the prevalence of the epidemic for one month. The forecasting accuracy was evaluated using different statistical measures.

4. Long short term memory neural network model

Artificial neural networks (ANN) are computing techniques that mimic the behavior of the central nervous system of humans in data processing. Many ANN architectures have been proposed in literature as predictive tools to model different engineering problems (Abd Elaziz et al., 2020; Elsheikh et al., 2020a; Essa et al., 2020a; Shehabeldeen et al., 2020). A typical ANN model is composed of many interconnected neurons connected with each other with synaptic weights. Most of ANNs architectures are composed of three layers; input, hidden, and output layers. The training process of ANN model is accomplished using an optimization approach such as stochastic gradient descent approach and the synaptic weights are updated using backpropagation approach. The main advantage of ANNs over other statistical-based predictive approaches is its capability to detect the complex nonlinear behavior of a certain system without involving in solving complicated mathematical models.

The recurrent neural network (RNN) is an ANN type that has chain architecture with multiple identical modules which used as memories to store processed data from preceding processing stages (Sherstinsky, 2020). RNNs have a feedback loop which gives this type of ANN an important advantage over other conventional ANN to process a sequence of inputs. However, these feedback loops result in enormous updates of the model parameters which make the network unstable. This is the main drawback of RNN which impairs the network training capabilities.

LSTM is an evolved version of recurrent neural network (Le et al., 2019). It was developed to overcome the problems related to the conventional RNN by adding more module interactions. LSTMs have high capabilities to learn dependencies and to remember enormous quantities of information for long period (Elsheikh et al., 2021). A typical LSTM model has a chain structure form consists of multiple modules. These modules have different structures from those of conventional RNN as they have four uniquely connected interacting layers. The structure of a typical LSTM neural network is presented in Fig. 5; it consists of memory modules called cells. Two states (hidden state and cell state) are transferred from the preceding cell to the next one. The cell state is responsible for flowing of data in the forward direction. The data may be

subjected to some linear mathematical transformations. The data can be removed from or added to the cell state using activation gates. These activation gates applied sigmoid activation functions to data. Each gate has some individual weights and employs multiple matrix operations. These gates are also used to control the memorizing process and consequently enhance LSTMs capabilities to avoid the dependency problems.

LSTM network begins to identify the trivial information that may be excluded from the processing. This trivial information is neglected from the cell processing in that step. The identification of the importance of the data is employed via sigmoid activation function which makes a decision on the received output from the last LSTM module h_{t-1} at a time step $t - 1$ and the current input signal X_t at time step t . The sigmoid activation function may also called forget gate as it eliminates the trivial part of the preceding output. The output of this function is represented by f_t which has numerical values ranges between 0 and 1 and is defined as follows:

$$f_t = \sigma (W_f [h_{t-1}, X_t] + b_f) \tag{1}$$

where, σ denotes the sigmoid activation function, b_f is the gate applied bias, and W_f is the gate weight matrix.

The new input X_t to the next module will be subjected to two different activation functions; sigmoid and tanh functions. These activation functions will decide which part will be stored and which part will be declared from the input. The sigmoid activation function will decide whether the new input should be declared or updated (0 or 1). The tanh activation function will define a new weight to the input and assign an importance index to it, which ranges between -1 and 1. The output of the two activation functions are multiplied and added to the old memory C_{t-1} to produce a new updated cell state C_t . The mathematical calculations are executed as follows:

$$i_t = \sigma (W_i [h_{t-1}, X_t] + b_i) \tag{2}$$

$$N_t = \tanh (W_n [h_{t-1}, X_t] + b_n) \tag{3}$$

$$C_t = C_{t-1}f_t + N_t i_t \tag{4}$$

where, C_t and C_{t-1} are the cell states at time steps t and $t - 1$, respectively, σ and \tanh are the sigmoid and tanh activation functions, respectively, b_n is the gate applied bias, and W_n is the gate weight matrix.

Finally, the output h_t can be calculated as a function of the output of the cell state O_t and the modified version of C_t after passing through tanh activation function. The output of the cell state is calculated by passing preceding output via a sigmoid activation function. The mathematical calculations are executed as follows:

$$O_t = \sigma (W_o [h_{t-1}, X_t] + b_o) \tag{5}$$

$$h_t = O_t \tanh (C_t) \tag{6}$$

where, b_o is the gate applied bias, and W_o is the gate weight matrix.

5. Model assessment criteria

In order to evaluate the performance of the proposed LSTM model, seven statistical criteria were employed, namely, Root mean square error (RMSE), coefficient of determination (R^2), mean absolute error (MAE), efficiency coefficient (EC), overall index (OI), coefficient of variation (COV), and coefficient of residual mass (CRM) (Elsheikh et al., 2019; Essa et al., 2020b).

RMSE represents the error between the forecasted and reported data and can be calculated using:

$$RMSE = \sqrt{\frac{1}{n_s} \sum_{i=1}^{n_s} (d_i - y_i)^2} \tag{7}$$

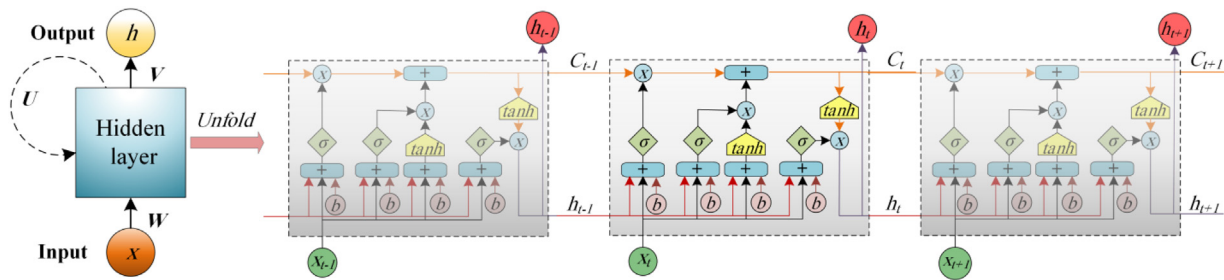


Fig. 5. The structure of a typical LSTM neural network.

R^2 represents the square correlation between forecasted and reported data and has values ranges between 0 and 1.0, and can be calculated using:

$$R^2 = \frac{(\sum_{i=1}^{n_s} (d_i - \bar{d})(y_i - \bar{y}))^2}{\sum_{i=1}^{n_s} (d_i - \bar{d})^2 \times \sum_{i=1}^{n_s} (y_i - \bar{y})^2} \quad (8)$$

MAE is used to evaluate the model accuracy and goodness fit of the datasets, and can be calculated using:

$$MAE = \frac{1}{n_s} \sum_{i=1}^{n_s} |d_i - y_i| \quad (9)$$

EC gives an overall sight to the model accuracy. The best accuracy can be obtained if the EC values equal the unity. EC can be calculated using:

$$EC = 1 - \frac{\sum_{i=1}^{n_s} (d_i - y_i)^2}{\sum_{i=1}^{n_s} (d_i - \bar{d})^2} \quad (10)$$

CRM is used to describe difference between forecasted and reported data, in which the value of zero indicates to the optimal accuracy. CRM can be calculated using:

$$CRM = \frac{\sum_{i=1}^{n_s} y_i - \sum_{i=1}^{n_s} d_i}{\sum_{i=1}^{n_s} d_i} \quad (11)$$

OI and COV are independent statistical criteria as they are computed as functions of RMSE. OI and COV can be calculated using:

$$OI = \frac{1}{2} \left(1 - \left(\frac{RMSE}{d_{max} - d_{min}} \right) + EC \right) \quad (12)$$

$$COV = \left(\left(\frac{RMSE}{y} \right) \times 100 \right) \quad (13)$$

Where, n_s , d , and y refers to the number of data, reported and the forecasted values, respectively. Moreover, the d_{min} , d_{max} , and \bar{d} denote the minimum, maximum and average values of the reported values, while \bar{y} denotes the average values of the forecasted values.

6. Results and discussion

LSTM network has been applied to predict the number of confirmed cases, recovered cases, and death on Saudi Arabia. The model was trained using the reported data for two different periods: 91 days (2 March- 31 May 2020) and 199 days (2 March- 15 September 2020). After the training process is accomplished, the trained network could be used to perform multi-day-ahead forecasting. Finally, the proposed LSTM model was applied to forecast the total number of confirmed cases as well as deaths in six different countries; Brazil, India, Saudi Arabia, South Africa, Spain, and USA. The model trained by the reported data from 1 January to 10 October, 2020.

To verify the forecasting accuracy of the model, the LSTM is trained using 90 % of the reported data and 10 % of the data was set aside for validation. The proposed model succeeded to predict the total cases for one week ahead with better accuracy compared with NARANN and ARIMA as shown in Fig. 6. The RMSE of the forecasted data using LSTM was less than 11 and 28 % of that of ARIMA and NARANN, respectively; which revealed the outperformance of the proposed method over other tested statistical and artificial intelligence based methods. The optimal parameters of the network such as the number of hidden units and the initial learning rate have been also determined via conducting a parametric study. First, a constant initial learning rate of 0.005 is used while the number of hidden units is changed and has values of 10, 50, 100, and 150. The forecasting accuracy of the model was assessed using different statistical criteria. The network with 100 hidden units had the highest accuracy among all other networks (with different hidden units). It had an R^2 , RMSE, MAE, COV, EC, and OI of 0.989, 473.454, 425.905, 0.577, 0.980, and 0.967, respectively, as tabulated in Table 1. It is noticed that the network with 100 hidden units has the highest R^2 which approaches the unity, the lowest RMSE, the lowest MAE, the lowest COV which approaches zero, the highest EC which approaches the unity, and the highest OI which approaches the unity. The values of these assessment criteria for the LSTM network with different hidden units are plotted in Fig. 7 (a). Moreover, the network with 0.005 initial learning had the highest accuracy among all other networks (with initial learning rate). It had an R^2 , RMSE, MAE, COV, EC, and OI of 0.989, 473.454, 425.905, 0.577, 0.980, and 0.967, respectively, as tabulated in Table 1. The computed values of all statistical assessment criteria for the network with the initial learning rate of 0.005 are the optimal compared with the others as shown in in Fig. 7(b). Therefore, to maximize the forecasting accuracy of the LSTM network, it is recommended to use 100 hidden units and an initial learning rate of 0.005.

Now, the accuracy of the proposed model to forecast COVID-19 cases has been verified by comparing the forecasted data with the reported data, also, the optimal network parameters that maximize the forecasting accuracy have been determined. Thus, the next step is to apply the proposed model with the determined optimal parameters to forecast the total confirmed cases, total recovered cases, and total deaths for ten days ahead. In this stage, the model is trained using 100 % of the reported data (2 March-31 May 2020). The forecasting results of total confirmed cases, total recovered cases, and total deaths are plotted in Fig. 7. It is estimated that the number of total confirmed cases will increase at a diminishing rate in the next ten days to reach 95,044 with only 536 daily cases at the last day. Moreover, the number of total recovered cases and total deaths will increase to reach 85,821 cases and 877 cases, respectively. Thus, at the end of the forecasting period, the total active cases are estimated to decline to 8346 as shown in Fig. 8. These forecasted results could help the policy makers and public health care providers in Saudi Arabia to make necessary arrangements in the future. Moreover, it may help to put suitable regulations and policies to organize the Hajj.

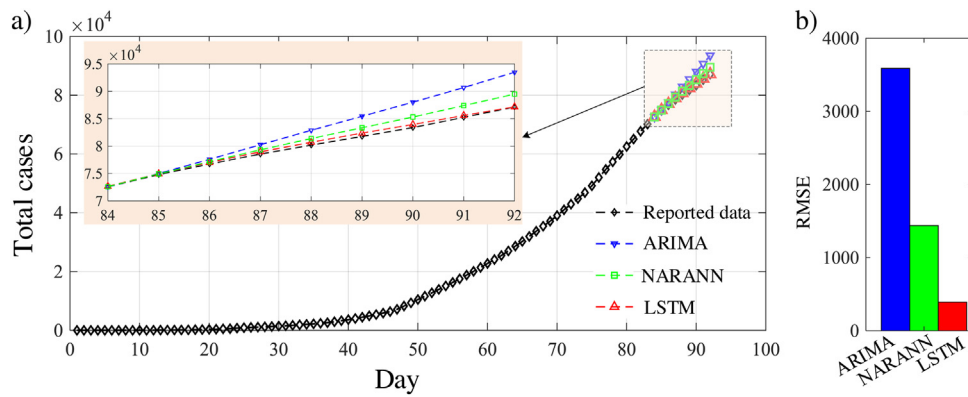


Fig. 6. a) Time series plot of the total confirmed COVID-19 cases data and forecasted data using ARIMA, NARANN, and LSTM; b) the RMSE between the forecasted and the reported data.

Table 1
Statistical evaluation of the developed model for different values of the model's parameters.

Factor	Value	R ²	RMSE	MAE	COV	EC	OI
Hidden units	10	0.979	2,373.551	2,016.164	2.972	0.518	0.645
	50	0.979	1,062.292	657.405	1.307	0.903	0.900
	100	0.989	473.453	425.903	0.577	0.980	0.967
	150	0.987	724.144	430.022	0.888	0.955	0.942
Initial learning rate	0.001	0.972	2,692.112	2,198.213	3.379	0.381	0.561
	0.002	0.994	1,413.930	1,329.624	1.699	0.829	0.846
	0.005	0.989	473.453	425.9055	0.577	0.980	0.967
	0.01	0.976	2,010.142	1,575.846	2.503	0.654	0.730

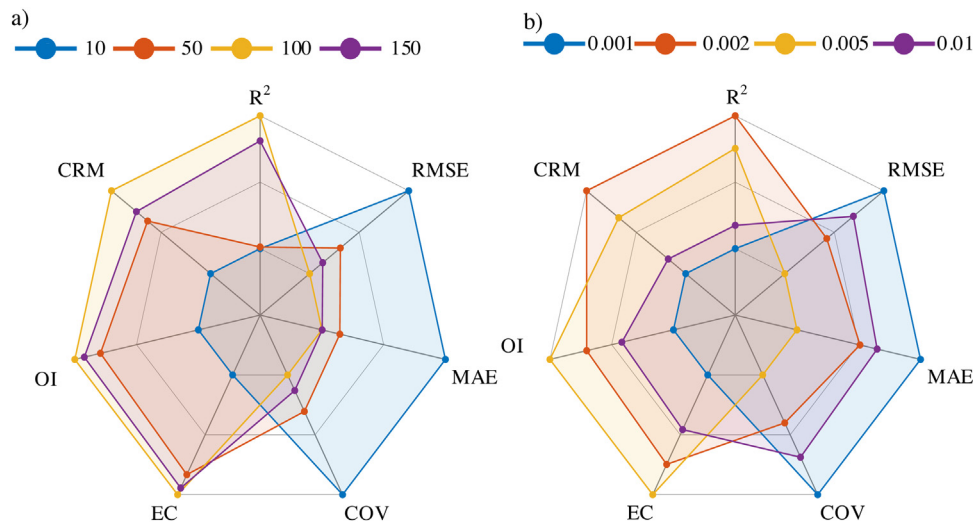


Fig. 7. Assessment criteria for the LSTM network with: a) different hidden units; b) different initial learning rate.

Fig. 9 shows the forecasted total confirmed cases, total recovered cases, and total deaths of COVID-19 in the next three weeks based on the declared cases in Saudi Arabia. Lower and upper forecasting limits are also plotted for the three cases using 95% confidence limit. The total number of the forecasted confirmed cases will reach 327,551 cases with lower limits of 337,101 cases and upper limit of 338,228 cases, respectively. The total number of the forecasted recoveries will reach 318,546 cases with lower limits of 316,292 cases and upper limit of 320,800 cases, respectively. The total number of the forecasted deaths will reach 4820 cases with lower limits of 4751 cases and upper limit of 4889 cases, respectively. Therefore, the total number of active cases is expected to decline to 4185 cases by the first week of October 2020. The statistical analysis of the forecasted results of the second forecasting period is summarized in Appendix A. The standard deviation and

mean of the forecasted results based on five runs of the model are tabulated.

The previous results showed the outperformance of the LSTM model over ARIMA and NARANN to forecast the prevalence of the outbreak in Saudi Arabia. Moreover, the optimal model parameters that maximize the model accuracy have been determined. To assess the robustness of the proposed LSTM model, it was applied to forecast the total number of confirmed cases as well as deaths in six different countries; Brazil, India, Saudi Arabia, South Africa, Spain, and USA. The selected countries have applied different protection measures to restrict the prevalence of the outbreak and consequently they have different epidemic trends. Moreover, they have different age structure, weather, and culture. To test the model, the reported data in 2020 (from 1 January to 10 October) has been used to train the model and about 10% (28 days) of the used data has

Table 2
Statistical evaluation of the LSTM model for different countries.

	Brazil		India		Saudi Arabia		South Africa		Spain		USA	
	Cases	Deaths	Cases	Deaths	Cases	Deaths	Cases	Deaths	Cases	Deaths	Cases	Deaths
R2	0.979	0.983	0.996	0.998	0.999	0.998	0.979	0.945	0.976	0.969	0.996	0.971
RMSE	18,609.734	388.201	34,625.082	735.183	160.608	100.039	5,346.588	143.234	8,257.561	112.784	9,205.816	709.711
MAE	15,520.000	319.656	29,856.400	667.806	145.737	89.611	5,232.888	109.895	6,247.400	93.846	8,092.200	603.603
COV	0.376	0.264	0.514	0.708	0.048	2.122	0.788	0.841	1.003	0.347	0.122	0.337
EC	0.811	0.838	0.883	0.705	0.992	0.136	-5.969	0.190	0.829	0.741	0.981	0.570
OI	0.823	0.844	0.880	0.756	0.982	0.424	-2.962	0.429	0.844	0.777	0.966	0.666
CRM	-0.003	-0.002	-0.004	-0.006	0.000	-0.019	-0.008	-0.006	0.001	0.003	0.001	-0.003

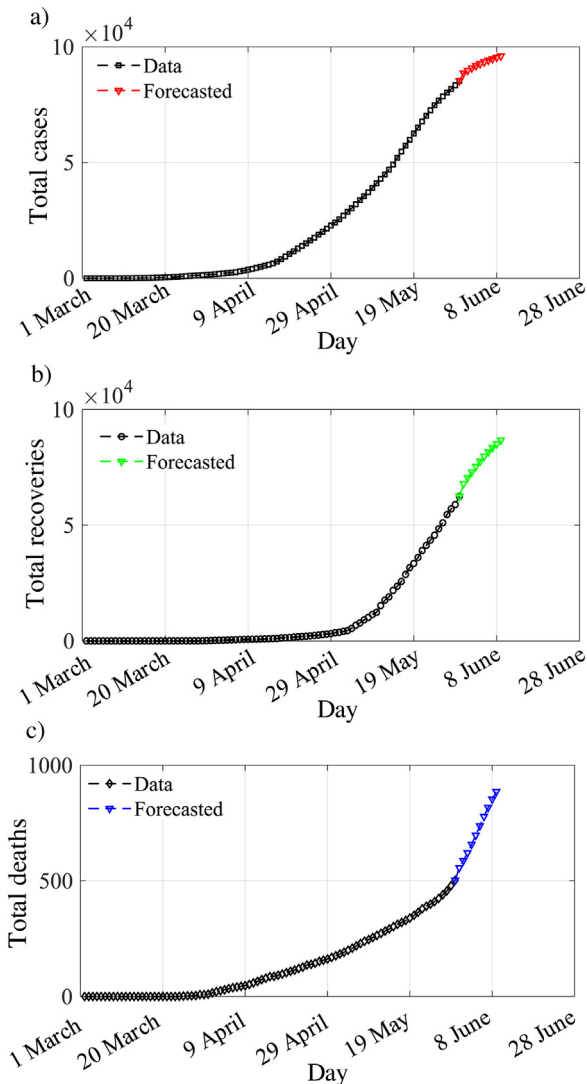


Fig. 8. Time series plot of the official reported data and forecasted data for the first time series; a) total confirmed cases; b) total recovered cases; c) total deaths.

been used to test the model accuracy. The model showed a good forecasting accuracy for all investigated countries based on different statistical measures as tabulated in Table 2. For all cases, R² approaches the unity which indicates the accuracy of the model. R² ranges from 0.976 to 0.998 for total cases and from 0.944 and 0.998 for total deaths. RMSE, MAE, COV, EC, OI, and CRM have reasonable values. The time series plots of the official reported data and forecasted data (accumulated cases and deaths) for six countries are shown in Fig. 10.

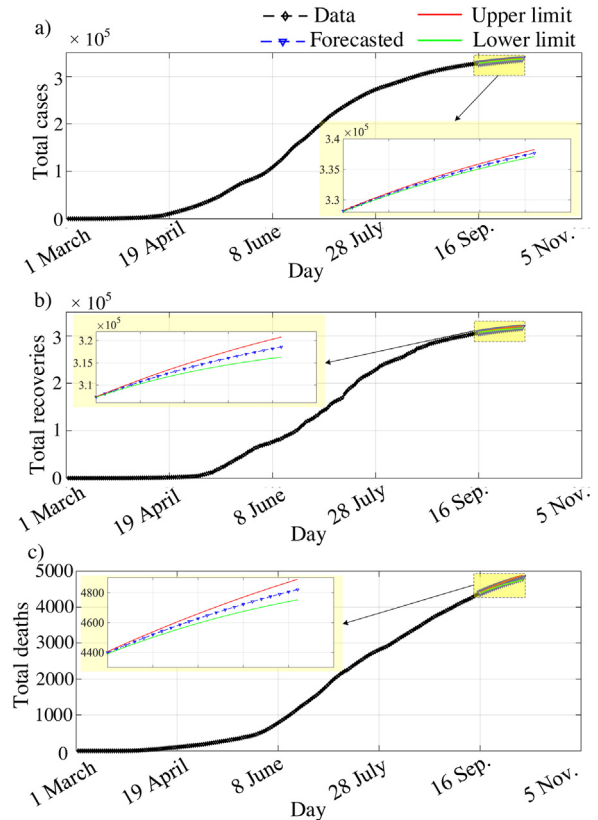


Fig. 9. Time series plot of the official reported data and forecasted data for the second time series with 95 % confidence level; a) total confirmed cases; b) total recovered cases; c) total deaths.

7. Conclusion

In this study, a deep learning model has been employed to forecast the number of total confirmed cases, recovered cases, and deaths of COVID-19 in Saudi Arabia. The official reported data was used to train the network. The optimal number of hidden units as well as the optimal value of the initial learning rate that maximize the forecasting accuracy was determined. It was recommended to use 100 hidden units and 0.005 as an initial learning rate. The network with the optimal parameter values was used to forecast the number of total cases, total recovered cases, and total deaths. The proposed LSTM model succeeded to forecast the total cases for one week ahead with better accuracy compared with NARANN and ARIMA. The RMSE of the forecasted data using LSTM was less than 11 and 28 % of that of ARIMA and NARANN, respectively; which revealed the outperformance of the proposed method over other tested statistical and artificial intelligence based methods. The obtained results of the study can help politics to put strategic plans not only to control the outbreak and to organize the 2020 Hajj

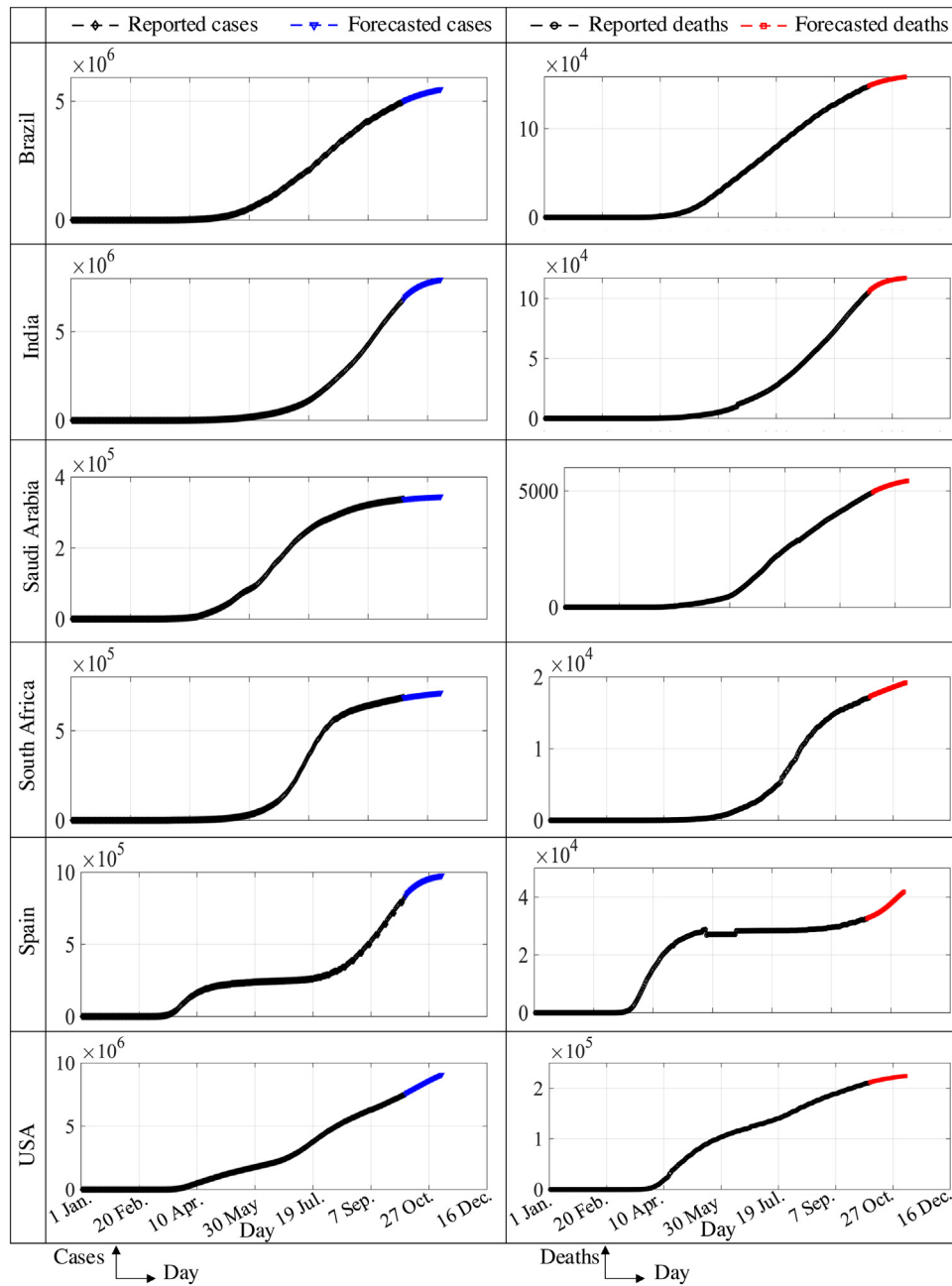


Fig. 10. Time series plot of the official reported data and forecasted data (accumulated cases and deaths) for six countries; Brazil, India, Saudi Arabia, South Africa, Spain, and USA.

pilgrimage but also for organizing the closure periods of the schools and universities. Moreover, the proposed LSTM model was applied to forecast the total number of confirmed cases as well as deaths in six different countries; Brazil, India, Saudi Arabia, South Africa, Spain, and USA. These countries have applied different policies to control the prevalence of the outbreak which results in different epidemic trends. They also have different age structure, weather, and culture. For all investigated cases, R^2 approaches the unity with a minimal value of 0.976 for total cases and 0.944 for total deaths which indicates the accuracy of the model. The main advantage of LSTM over conventional feedforward networks is the existence of the feedback connections which enables travelling of data signals in backward directions and consequently enhances the forecasting accuracy. Another advantage of the LSTM over other conventional

forecasting models is its capability to learn the nonlinearity from the training data. For the future work, it is recommended to investigate the effects of different protection measures such as applying contact restriction or lockdown on the outbreak prevalence via exploiting the high capabilities of LSTM to learn dependencies and to remember enormous quantities of information for long period. Moreover, the incorporation between LSTM and different advanced metaheuristic optimization approaches will be a good research topic to enhance the forecasting capabilities of the model and to include more variable context of the epidemic.

Appendix A

Table A1

Table A1
Statistical analysis of the forecasted results for the second forecasting period.

Date	Total confirmed cases		Total recoveries		Total deaths	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
16-Sep	328,192.824	46.727	307,304.432	38.205	4,398.669	3.590
17-Sep	328,778.844	56.014	308,035.332	53.911	4,423.523	4.717
18-Sep	329,349.7	65.672	308,745.234	91.660	4,448.236	5.813
19-Sep	329,905.745	75.563	309,430.322	135.042	4,472.367	6.962
20-Sep	330,446.9	85.695	310,092.431	180.954	4,495.965	8.156
21-Sep	330,973.646	96.026	310,732.337	228.711	4,519.034	9.397
22-Sep	331,486.246	106.549	311,352.133	278.079	4,541.583	10.684
23-Sep	331,984.734	117.258	311,951.314	328.957	4,563.616	12.017
24-Sep	332,469.7	128.152	312,531.653	381.279	4,585.139	13.398
25-Sep	332,941.343	139.237	313,091.954	434.933	4,606.155	14.824
26-Sep	333,399.853	150.510	313,634.542	489.878	4,626.667	16.297
27-Sep	333,845.656	161.976	314,159.523	545.975	4,646.680	17.814
28-Sep	334,279.535	173.643	314,667.323	603.175	4,666.198	19.376
29-Sep	334,700.156	185.510	315,158.523	661.329	4,685.225	20.980
30-Sep	335,109.443	197.584	315,633.633	720.348	4,703.765	22.625
01-Oct	335,507.567	209.847	316,093.233	780.158	4,721.824	24.310
02-Oct	335,893.424	222.311	316,537.243	840.630	4,739.408	26.031
03-Oct	336,268.734	234.966	316,966.821	901.688	4,756.522	27.782
04-Oct	336,633.243	247.796	317,382.214	963.242	4,773.174	29.575
05-Oct	336,987.234	260.821	317,783.528	1,025.172	4,789.369	31.394
06-Oct	337,331.345	274.014	318,171.523	1,087.390	4,805.115	33.240
07-Oct	337,664.945	287.357	318,546.623	1,149.845	4,820.421	35.110

Declaration of Competing Interest

The authors report no declarations of interest.

References

- Abd Elaziz, M., Shehabeldeen, T.A., Elsheikh, A.H., Zhou, J., Ewees, A.A., Al-qaness, M.A.A., 2020. Utilization of Random Vector Functional Link integrated with Marine Predators Algorithm for tensile behavior prediction of dissimilar friction stir welded aluminum alloy joints. *J. Mater. Res. Technol.* 9, 11370–11381.
- Abdulmajeed, K., Adeleke, M., Popoola, L., 2020. Online forecasting of COVID-19 cases in Nigeria using limited data. *Data Brief* 30, 105683.
- Ahmar, A.S., del Val, E.B., 2020. SutteARIMA: short-term forecasting method, a case: Covid-19 and stock market in Spain. *Sci. Total Environ.* 729, 138883.
- Al-qaness, M.A., Ewees, A.A., Fan, H., Abd El Aziz, M., 2020a. Optimization method for forecasting confirmed cases of COVID-19 in China. *J. Clin. Med.* 9, 674.
- Al-qaness, M.A., Ewees, A.A., Fan, H., Abugalig, L., Abd Elaziz, M., 2020b. Marine predators algorithm for forecasting confirmed cases of COVID-19 in Italy, USA, Iran and Korea. *Int. J. Environ. Res. Public Health* 17, 3520.
- Arias Velásquez, R.M., Mejía Lara, J.V., 2020. Forecast and evaluation of COVID-19 spreading in USA with reduced-space Gaussian process regression. *Chaos Solitons Fractals* 136, 109924.
- Babikir, H.A., Elaziz, M.A., Elsheikh, A.H., Showaib, E.A., Elhadary, M., Wu, D., Liu, Y., 2019. Noise prediction of axial piston pump based on different valve materials using a modified artificial neural network model. *Alexandria Eng. J.* 58, 1077–1087.
- Boccaletti, S., Ditto, W., Mindlin, G., Atangana, A., 2020. Modeling and forecasting of epidemic spreading: the case of Covid-19 and beyond. *Chaos Solitons Fractals* 135, 109794.
- Cao, D., Yin, H., Chen, J., Tang, F., Peng, M., Li, R., Xie, H., Wei, X., Zhao, Y., Sun, G., 2020. Clinical analysis of ten pregnant women with COVID-19 in Wuhan, China: a retrospective study. *Int. J. Infect. Dis.* 95, 294–300.
- Cássaro, F.A.M., Pires, L.F., 2020. Can we predict the occurrence of COVID-19 cases? Considerations using a simple model of growth. *Sci. Total Environ.* 728, 138834.
- Ceylan, Z., 2020. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci. Total Environ.* 729, 138817.
- Chakraborty, T., Ghosh, I., 2020. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: a data-driven analysis. *Chaos Solitons Fractals* 135, 109850.
- Chimmula, V.K.R., Zhang, L., 2020. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals* 135, 109864.
- Chintalapudi, N., Battineni, G., Amenta, F., 2020a. COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: a data driven model approach. *J. Microbiol. Immunol. Infect.*
- Chintalapudi, N., Battineni, G., Sagaro, G.G., Amenta, F., 2020b. COVID-19 outbreak reproduction number estimations and forecasting in Marche, Italy. *Int. J. Infect. Dis.* 96, 327–333.
- Collivignarelli, M.C., Collivignarelli, C., Carnevale Miino, M., Abbà, A., Pedrazzani, R., Bertanza, G., 2020. SARS-CoV-2 in sewer systems and connected facilities. *Process. Saf. Environ. Prot.* 143, 196–203.
- Earnest, A., Chen, M.I., Ng, D., Sin, L.Y., 2005. Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. *BMC Health Serv. Res.* 5, 36.
- Elaziz, M.A., Elsheikh, A.H., Sharshir, S.W., 2019. Improved prediction of oscillatory heat transfer coefficient for a thermoacoustic heat exchanger using modified adaptive neuro-fuzzy inference system. *Int. J. Refrig.* 102, 47–54.
- Elsheikh, A.H., Sharshir, S.W., Abd Elaziz, M., Kabeel, A.E., Guilan, W., Haiou, Z., 2019. Modeling of solar energy systems using artificial neural network: a comprehensive review. *Sol. Energy* 180, 622–639.
- Elsheikh, A.H., Sharshir, S.W., Ismail, A.S., Sathyamurthy, R., Abdelhamid, T., Edreis, E.M.A., Kabeel, A.E., Haiou, Z., 2020a. An artificial neural network based approach for prediction of the thermal conductivity of nanofluids. *SN Appl. Sci.* 2, 235.
- Elsheikh, A.H., Shehabeldeen, T.A., Zhou, J., Showaib, E., Abd Elaziz, M., 2020b. Prediction of laser cutting parameters for polymethylmethacrylate sheets using random vector functional link network integrated with equilibrium optimizer. *J. Intell. Manuf.*
- Elsheikh, A.H., Katekar, V.P., Muskens, O.L., Deshmukh, S.S., Elaziz, M.A., Dabour, S.M., 2021. Utilization of LSTM neural network for water production forecasting of a stepped solar still with a corrugated absorber plate. *Process. Saf. Environ. Prot.* 148, 273–282.
- Essa, F.A., Abd Elaziz, M., Elsheikh, A.H., 2020a. An enhanced productivity prediction model of active solar still using artificial neural network and Harris Hawks optimizer. *Appl. Therm. Eng.* 170, 115020.
- Essa, F.A., Abd Elaziz, M., Elsheikh, A.H., 2020b. Prediction of power consumption and water productivity of seawater greenhouse system using random vector functional link network integrated with artificial ecosystem-based optimization. *Process. Saf. Environ. Prot.* 144, 322–329.
- He, Z., Tao, H., 2018. Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: a nine-year retrospective study. *Int. J. Infect. Dis.* 74, 61–70.
- Kang, D., Choi, H., Kim, J.-H., Choi, J., 2020. Spatial epidemic dynamics of the COVID-19 outbreak in China. *Int. J. Infect. Dis.* 94, 96–102.
- Le, X.-H., Ho, H.V., Lee, G., Jung, S., 2019. Application of long short-term memory (LSTM) neural network for flood forecasting. *Water* 11, 1387.
- Maleki, M., Mahmoudi, M.R., Wraith, D., Pho, K.-H., 2020. Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel Med. Infect. Dis.*, 101742.
- Marimuthu, Y., Nagappa, B., Sharma, N., Basu, S., Chopra, K.K., 2020. COVID-19 and tuberculosis: a mathematical model based forecasting in Delhi, India. *Indian J. Tuberc.*
- Materassi, M., 2019. Some fractal thoughts about the COVID-19 infection outbreak. *Chaos Solitons Fract.* 4, 100032.
- Nishiura, H., Kobayashi, T., Yang, Y., Hayashi, K., Miyama, T., Kinoshita, R., Linton, N.M., Jung, S.-m., Yuan, B., Suzuki, A., 2020. The Rate of Underascertainment of Novel Coronavirus (2019-nCoV) Infection: Estimation Using Japanese Passengers Data on Evacuation Flights. *Multidisciplinary Digital Publishing Institute.*
- Qi, H., Xiao, S., Shi, R., Ward, M.P., Chen, Y., Tu, W., Su, Q., Wang, W., Wang, X., Zhang, Z., 2020. COVID-19 transmission in Mainland China is associated with temperature and humidity: a time-series analysis. *Sci. Total Environ.* 728, 138778.
- Ren, H., Zhao, L., Zhang, A., Song, L., Liao, Y., Lu, W., Cui, C., 2020. Early forecasting of the potential risk zones of COVID-19 in China's megacities. *Sci. Total Environ.* 729, 138995.
- Ribeiro, M.H.D.M., da Silva, R.G., Mariani, V.C., Coelho, L.D.S., 2020. Short-term forecasting COVID-19 cumulative confirmed cases: perspectives for Brazil. *Chaos Solitons Fractals* 135, 109853.

- Saba, A.I., Elsheikh, A.H., 2020. Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. *Process. Saf. Environ. Prot.* 141, 1–8.
- Salgotra, R., Gandomi, M., Gandomi, A.H., 2020. Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. *Chaos Solitons Fractals*, 109945.
- Shehabeldeen, T.A., Elaziz, M.A., Elsheikh, A.H., Zhou, J., 2019. Modeling of friction stir welding process using adaptive neuro-fuzzy inference system integrated with Harris hawks optimizer. *J. Mater. Res. Technol.* 8, 5882–5892.
- Shehabeldeen, T.A., Elaziz, M.A., Elsheikh, A.H., Hassan, O.F., Yin, Y., Ji, X., Shen, X., Zhou, J., 2020. A novel method for predicting tensile strength of friction stir welded AA6061 aluminium alloy joints based on hybrid random vector functional link and henry gas solubility optimization. *IEEE Access* 8, 79896–79907.
- Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D* 404, 132306.
- Shi, P., Dong, Y., Yan, H., Zhao, C., Li, X., Liu, W., He, M., Tang, S., Xi, S., 2020. Impact of temperature on the dynamics of the COVID-19 outbreak in China. *Sci. Total Environ.* 728, 138890.
- Singh, S., Parmar, K.S., Kumar, J., Makkhan, S.J.S., 2020. Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19. *Chaos Solitons Fractals* 135, 109866.
- Thomas-Rüddel, D., Winning, J., Dickmann, P., Ouart, D., Kortgen, A., Janssens, U., Bauer, M., 2020. Coronavirus disease 2019 (COVID-19): update for anesthesiologists and intensivists March 2020. *Der Anaesthesist*.
- Utkucan, Ş., Tezcan, Ş., 2020. Forecasting the cumulative number of confirmed cases of COVID-19 in Italy, UK and USA using fractional nonlinear grey Bernoulli model. *Chaos Solitons Fractals*, 109948.
- Wang, L., Li, J., Guo, S., Xie, N., Yao, L., Cao, Y., Day, S.W., Howard, S.C., Graff, J.C., Gu, T., Ji, J., Gu, W., Sun, D., 2020. Real-time estimation and prediction of mortality caused by COVID-19 with patient information based algorithm. *Sci. Total Environ.* 727, 138394.
- Wu, Y., Jing, W., Liu, J., Ma, Q., Yuan, J., Wang, Y., Du, M., Liu, M., 2020. Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. *Sci. Total Environ.* 729, 139051.
- Xu, H., Yan, C., Fu, Q., Xiao, K., Yu, Y., Han, D., Wang, W., Cheng, J., 2020. Possible environmental effects on the spread of COVID-19 in China. *Sci. Total Environ.* 731, 139211.
- Yousaf, M., Zahir, S., Riaz, M., Hussain, S.M., Shah, K., 2020. Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan. *Chaos Solitons Fractals* 138, 109926.
- Yu, H.-K., Kim, N.-Y., Kim, S.S., Chu, C., Kee, M.-K., 2013. Forecasting the number of human immunodeficiency virus infections in the Korean population using the autoregressive integrated moving average model. *Osong Public Health Res. Perspect.* 4, 358–362.
- Zhao, S., Musa, S.S., Lin, Q., Ran, J., Yang, G., Wang, W., Lou, Y., Yang, L., Gao, D., He, D., 2020. Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of January 2020: a data-driven modelling analysis of the early outbreak. *J. Clin. Med.* 9, 388.