Decision Analytics Journal 1 (2021) 100007

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

A predictive analytics model for COVID-19 pandemic using artificial neural networks

Yusuf Kuvvetli [a], Muhammet Deveci [b,*], Turan Paksoy [c], Harish Garg [d]

[a] Department of Industrial Engineering, Cukurova University, 01330 Adana, Turkey
[b] Department of Industrial Engineering, Turkish Naval Academy, National Defence University, 34940 Istanbul, Turkey
[c] Department of Aviation Management, Faculty of Aviation and Space Sciences, Necmettin Erbakan University, Konya, Turkey
[d] School of Mathematics, Thapar Institute of Engineering & Technology, Deemed University, Patiala, Punjab, India

ARTICLE INFO

Keywords:
Predictive analytics
COVID 19 pandemic
Artificial intelligence
Neural networks
Decision support
Healthcare management

ABSTRACT

The COVID-19 pandemic spread rapidly around the world and is currently one of the most leading causes of death and heath disaster in the world. Turkey, like most of the countries, has been negatively affected by COVID-19. The aim of this study is to design a predictive model based on artificial neural network (ANN) model to predict the future number of daily cases and deaths caused by COVID-19 in a generalized way to fit different countries' spreads. In this study, we used a dataset between 11 March 2020 and 23 January 2021 for different countries. This study provides an ANN model to assist the government to take preventive action for hospitals and medical facilities. The results show that there is an 86% overall accuracy in predicting the mortality rate and 87% in predicting the number of cases.

1. Introduction

The novel coronavirus (nCoV), which is a new respiratory and systemic disease [1], was first reported on 31 December 2019 in Wuhan, Hubei Province, China. It was temporarily named as "2019-nCoV". The novel coronavirus disease of 2019 was subsequently called "COVID-19". The World Health Organization (WHO) announced COVID-19 as a global pandemic on 11 March 2020. After COVID-19 cases were detected in China, it spread to other countries and continents including Europe, Australia and North America. According to the U.S.' Johns Hopkins University report published on 31 May 2020, the number of countries where COVID-19 cases have been detected has reached 187. Some of these countries such as United States, Brazil, Russia, United Kingdom, Spain, Italy, India, France, Germany and Turkey have been heavily affected. Since its first reporting at the end of December 2019 and as of 8 September 2021, over 222 million people have been infected, around 199 million people recovered, and about 4.591.915 people died due to COVID-19 pandemic.[1]

People around the world experienced several outbreaks such as Spanish flu, Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) in the last century. It is estimated that 500 million people worldwide were infected by the Spanish flu outbreak and 50–100 million people died from this infection between 1918 and 1920 [2]. The more recent outbreaks have been SARS in 2003 and MERS in 2012, and COVID-19 in 2019. All three diseases (COVID-19, SARS and MERS) are caused by various forms of Coronavirus and do not appear to be very different from each other in terms of their clinical features [3]. The major difference and ability of COVID-19 from other SARS and MERS diseases is that it is spreading rapidly by way of human contact (human-to-human transmission) and leave about 20% infected subjects as symptom-less carriers [4,5]. Also, distinguishing features of COVID-19 from other pandemics are as follows: *(i)* high infection rate, *(ii)* incubation period, *(iii)* patients to be contagious during the incubation period, and *(iv)* symptomatic infection [6]. In addition, clinical studies and experience from infected patients have identified that complications of COVID-19 are most likely to be effective in the elderly people and those who have weakened immune systems [5]. People with special health conditions such as cancer, hypertension, severe asthma, cardiovascular disease, lung conditions, heart disease, diabetes, neurological conditions, HIV/AIDS infection, pregnancy are more vulnerable to the serious effects of COVID-19 [7,8]. The most common symptoms are fever, dry cough, dyspnea, sore throat, loss of taste and tiredness, and all infected patients have at least one of these symptoms [3].

Each outbreak has its own characteristics, however previous outbreaks have caused major negative impacts on the population health, social–cultural activities and global economy. Artificial intelligence network can be an effective approach to monitor the outbreak and

* Corresponding author.
E-mail addresses: ykuvvetli@cu.edu.tr (Y. Kuvvetli), muhammetdeveci@gmail.com (M. Deveci), tpaksoy@erbakan.edu.tr (T. Paksoy), harish.garg@thapar.edu (H. Garg).

[1] https://coronavirus.jhu.edu/map.html

https://doi.org/10.1016/j.dajour.2021.100007
Received 17 August 2021; Received in revised form 2 October 2021; Accepted 2 October 2021
Available online 30 October 2021

predict the spread of the pandemic. In this study, we proposed an Artificial Neural Network (ANN) model to determine the number of new daily cases and predict the mortality rate caused by COVID-19.

It should reveal the nonlinear relations between attributes obtained with the advantage of an artificial learning method. In addition to classical artificial learning, many different model parameters are investigated to optimize the prediction model with the grid search approach. Besides, it is necessary to envision re-learning in the changing spread. In this respect, the proposed methodology results will be used in pandemic situations to assist decision making for policy makers.

The study aims to predict number of confirmed cases and fatality in the current pandemic for different countries. The deep learning and artificial neural networks can be useful approach to analyze the non-linear relations in the historical cases data as aforementioned. Therefore, three different artificial neural network and deep learning methods are applied for this prediction. The originality of this prediction is that the paper considers 10 days forecasts according to the current prospective. Therefore, policy makers can be used these predictions for monitoring the spread of the pandemic.

The rest of this paper is structured as follows. Section 2 provides a literature review about the prediction of COVID-19 using artificial intelligence techniques. The detailed information about COVID-19 in Turkey is presented in Section 3. Section 4 introduces the methods and dataset used in this study. The results are given in Section 5. Section 6 discusses the findings and conclude the study.

2. Literature review on COVID-19 and machine learning

Since the first day (31 December 2019) of the appearance of COVID-19, many studies on the spreading dynamics of COVID-19 have been published in the literature recently. Yang et al. [9] proposed an artificial intelligence (AI) model for the prediction of the epidemic trend of COVID-19 in China under public health interventions. Srinivasa Rao and Vazquez [10] introduced a machine learning algorithm to develop possible COVID-19 case identification faster using a mobile phone–based web survey. Vaishya et al. [11] applied AI for the development of drugs and vaccines, and the reduction of workload of healthcare workers. Randhawa et al. [12] improved a machine learning algorithm for the identification of intrinsic COVID-19 virus genomic signature.

Madurai Elavarasan and Pugazhendhi [13] conducted research on uncovering the hidden roles of technologies helping to control the pandemic. They illuminate a variety of applied technologies that are helping health systems, government and the public in various ways to cope with COVID-19. Chimmula and Zhang [14] improved a forecasting model of COVID-19 pandemic in Canada with the help of state-of-the-art Deep Learning models. In this study, the trends and possible downtime of the current COVID-19 outbreak were estimated. Cobb and Seale [15] examined trends across US states and COVID-19 growth rates in relation to the presence of shelter-in-place orders in that county using machine learning. Ardakani et al. [16] applied deep learning technique to manage COVID-19 in routine clinical practice using CT images. R et al. [17] aimed to analyze and visualize the impact of COVID-19 in the world by applying machine learning methods in sentiment analysis on the tweet dataset in order to understand the very positive and very negative views of the public in the world. Sujath et al. [18] proposed a machine learning prediction model for the COVID-19 pandemic in India. Here is a model including linear regression, Multilayer perceptron and Vector autoregression method that may be useful for predicting the spread of COVID-2019. Mollalo et al. [19] investigated the COVID-19 case rates in the United States with the multilayer perceptron (MLP) of the Artificial Neural Network. A database of 57 candidate explanatory variables was created to examine the model's performance in predicting cumulative COVID-19 case rates. Shen et al. [20] collected COVID-19-related posts on Weibo, a popular Twitter-like social media site in China, and analyzed them with a machine learning classifier. Kang et al. [21] examined a set of features extracted from CT images to diagnose COVID-19 using Structured Hidden Multi-View Representation Learning. Pinter et al. [22] implemented hybrid machine learning approach for COVID-19 pandemic prediction for Hungary. Vaid et al. [23] proposed artificial intelligence to investigate the strictness of physical distancing policies in North America and to examine the risk of a second wave of COVID-19 infection. Fayyoumi et al. [24] conducted a case study of machine learning and statistical modeling for the prediction of new COVID-19 patients in Jordan.

Aljaaf et al. [25] presented a fusion of data science and feedforward neural network-based modeling of COVID-19 pandemic prediction in Iraq. Alodat [26] proposed some models including Machine Learning (ML), Deep Learning (DL), Convolutional Neural Networks using Tensorflow (CNN-TF) to create robust models in real-time to support Telemedicine. Dash et al. [27] focused on improving an intelligent computing model on time-series data analysis to predict the pandemic of COVID-19. Elhag et al. [28] applied the ANNs networks and logistic regression model for optimization studying COVID-19. Data from the WHO website of 32 European countries from 11 January 2020 to 29 May 2020 were used for the analysis. Elleuch et al. [29] proposed an integrated ANN and Fuzzy Interval Mathematical model to help health managers with the Prioritization and Scheduling COVID-19 patient's problem. Hamadneh et al. [30] developed an ANN with prey predator algorithm for prediction of the COVID-19 cases of Brazil and Mexico. Kolozsvári et al. [31] explored the use of artificial intelligence in modeling COVID-19. Nabi et al. [32] studied four deep learning models including LSTM, GRU, CNN and MCNN to predict COVID-19 cases. Namasudra et al. [33] proposed a new Nonlinear Autoregressive (NAR) Neural Network Time Series (NAR-NNTS) model to predict COVID-19 cases. Rasjid et al. [34] studied the prediction of death and infected COVID-19 cases in Indonesia using time series correction and LSTM neural network. Shamman et al. [35] aimed to discuss the myriad aspects of today's modern technology used to combat COVID-19 emergencies at various scales, such as medical image processing, disease tracking, expected outcomes, computational science, and drugs. Shawaqfah and Almomani [36] improved an ANN architecture to predict the COVID pandemic impact in Qatar, Spain, and Italy. Tang et al. [37] focused the COVID-19 pandemic and analysis of economic impacts using self-correcting error-based prediction model. Toğa et al. [38] presented Autoregressive Integrated Moving Average (ARIMA) and ANN in Turkey for the COVID-19 prevalence forecasting. The proposed models are compared with respect to correlation coefficient and mean square error (MSE). Zisad et al. [39] presented an integrated ANN and SEIR (Susceptible, Exposed, Infected, Removed) model to predict COVID-19.

The existing literature generally focuses to predict cases or spread for specific countries with different prediction methods such as artificial neural networks, time series methods or other pandemic spread methods (SEIR). This study differs from the studies with considering ten days forecast for creating a beneficial input for the pandemic policy of countries. The proposed approach includes ANN, deep learning and LSTM methods to find the best predictions; therefore, this approach can potentially use in different countries with different spreads.

3. COVID-19 pandemic in Turkey and the world

COVID-19 has spread rapidly to other countries around the world since it emerged in China at the end of the year 2019. The key milestones in the spread of the coronavirus pandemic are illustrated in Fig. 1 as a timeline of WHO alerts.[2] As can be seen on the timeline, the first case was reported in China on December 2019. Turkey's monthly summary report is presented in Table 2.[3] When the dates 3th of February 2020 and 2th of February 2021 are compared, the number

[2] https://www.who.int/

[3] https://covid19.saglik.gov.tr/

Table 1
Summary of the available ANN and the prediction of COVID-19.

Author(s)	Year	Research focus	SA (Yes/No)	CA (Yes/No)	Method(s)	Application	
						Country	Type
Aljaaf et al.	2021	COVID-19 outbreak forecasting	No	Yes	ARIMA and ANN	Iraq	Real life
Alodat	2021	Support telemedicine	No	Yes	Deep learning model	Oman	Real life
Dash et al.	2021	Prediction of COVID-19 pandemics	No	No	Intelligent computing model	Six countries	Real life
Elhag et al.	2021	Optimization studying COVID-19	No	No	ANN and Logistic regression	European countries	Real life
Elleuch et al.	2021	Real-time prediction of COVID-19	No	Yes	ANN and Fuzzy interval	Not provided	Real life
Hamadneh et al.	2021	Prediction of the COVID-19	No	Yes	ANN and Prey predator algorithm	Brazil and Mexico	Real life
Kolozsvári et al.	2021	Predicting the SARS-CoV-2 diseases	No	Yes	Recurrent Neural Networks	European countries and USA	Real life
Nabi et al.	2021	Forecasting COVID-19 cases	No	Yes	Deep learning	Brazil, Russia and UK	Real life
Namasudra et al.	2021	Prediction of COVID-19 cases	No	Yes	Nonlinear neural network	–	–
Rasjid et al.	2021	Prediction of the COVID-19 Cases	No	Yes	LSTM-NN	Indonesia	Real life
Shamman et al.	2021	Tackling against COVID-19 pandemic	–	–	Machine learning	–	–
Shawaqfah and Almomani	2021	Forecasting COVID-19 cases	Yes	Yes	ANN model	European countries	Real life
Tang et al.	2021	Prediction for the COVID-19 pandemic	No	No	Deep deterministic policy gradient	U.S	Real life
Toğa et al.	2021	COVID-19 prevalence forecasting	No	Yes	ARIMA and ANN	Turkey	Real life
Zisad et al.	2021	Prediction of COVID-19 cases	No	Yes	SEIR model and Neural network	Bangladesh	Real life
Proposed study		Prediction of confirmed cases and deaths	Yes	Yes	ANN, LSTM and deep learning models	Ten countries	Real life

ANN: Artificial Neural Network; CA: Comparative Analysis; SA: Sensitivity Analysis.

Table 2
Summary table of COVID-19 in Turkey.

Date	Number of tests	Total number of cases	Total number of deaths	Rate of pneumonia in patients (%)	Number of severe patients	Total number of patients recovered	Daily number of patients	Daily number of tests	Daily number of deaths	Daily patients recovered
3 Feb 2021	30.061.437	2.501.079	26.354	4,9	1.523	2.387.384	632	148.192	117	8.314
2 Feb 2021	29.913.245	2.492.977	26.237	4,9	1.592	2.379.070	630	140.12	120	8.639
1 Feb 2021	29.773.125	2.485.182	26.117	4,9	1.615	2.370.431	636	141.703	124	8.016
31 Jan 2021	29.631.422	2.477.463	25.993	4,7	1.634	2.362.415	641	136.418	128	7.006
30 Jan 2021	29.495.004	2.470.901	25.865	4.7	1.692	2.355.409	658	148.785	129	7.1
29 Jan 2021	29.346.219	2.464.030	25.736	4,7	1.74	2.348.309	664	165.094	131	8.093
28 Jan 2021	29.181.125	2.457.118	25.605	4,7	1.751	2.340.216	670	173.21	129	8.902
27 Jan 2021	29.007.915	2.449.839	25.476	4,7	1.765	2.331.314	675	179.419	132	8.803
26 Jan 2021	28.828.496	2.442.350	25.344	4,7	1.791	2.322.511	681	180.303	134	8.108
25 Jan 2021	28.648.193	2.435.247	25.21	4,7	1.808	2.314.403	671	151.109	137	6.682
24 Jan 2021	28.497.084	2.429.605	25.073	5,1	1.905	2.307.721	684	148.425	140	5.86
23 Jan 2021	28.348.659	2.424.328	24.933	5,1	1.962	2.301.861	723	152.758	144	5.811
22 Jan 2021	28.195.901	2.418.472	24.789	5,1	2.003	2.296.050	734	163.342	149	6.018
21 Jan 2021	28.032.559	2.412.505	24.64	5,1	2.074	2.290.032	743	165.109	153	6.113
20 Jan 2021	27.867.450	2.406.216	24.487	5,1	2.102	2.283.919	752	168.894	159	5.932
19 Jan 2021	27.698.556	2.399.781	24.328	5,1	2.162	2.277.987	761	175.133	167	7.218
18 Jan 2021	27.523.423	2.392.963	24.161	5,1	2.183	2.270.769	749	151.342	164	7.905
17 Jan 2021	27.372.081	2.387.101	23.997	4,5	2.201	2.262.864	803	148.636	165	8.812
16 Jan 2021	27.223.445	2.380.665	23.832	4,5	2.265	2.254.052	902	156.792	168	8.005
15 Jan 2021	27.066.653	2.373.115	23.664	4,5	2.311	2.246.047	921	167.211	169	9.109
14 Jan 2021	26.899.442	2.364.801	23.495	4,5	2.512	2.236.938	958	169.847	170	9.011
13 Jan 2021	26.729.595	2.355.839	23.325	4,5	2.652	2.227.927	971	173.603	173	9.463
12 Jan 2021	26.555.992	2.346.285	23.152	4,5	2.701	2.218.464	983	179.208	171	10.013
11 Jan 2021	26.376.784	2.336.476	22.981	4,5	2.783	2.208.451	1.003	180.303	174	10.301
10 Jan 2021	26.196.481	2.326.256	22.807	4,3	2.811	2.198.150	1.017	162.786	176	8.103
9 Jan 2021	26.033.695	2.317.118	22.631	4,3	2.903	2.190.047	1.103	168.289	181	7.902
8 Jan 2021	25.865.406	2.307.581	22.45	4,3	3.094	2.182.145	1.291	184.193	186	9.894
7 Jan 2021	25.681.213	2.296.102	22.264	4,3	3.201	2.172.251	1.37	183.003	194	8.211
6 Jan 2021	25.498.210	2.283.931	22.07	4,3	3.303	2.164.040	1.458	182.645	191	8.702
5 Jan 2021	25.315.565	2.270.101	21.879	4,3	3.41	2.155.338	1.477	183.413	194	8.908
4 Jan 2021	25.132.152	2.255.607	21.685	4,3	3.522	2.146.430	1.508	181.323	197	9.896
3 Jan 2021	24.950.829	2.241.912	21.488	3,8	3.612	2.136.534	1.515	138.941	193	10.102
2 Jan 2021	24.811.888	2.232.035	21.295	3,8	3.764	2.126.432	1.713	149.218	202	11.672
1 Jan 2021	24.662.670	2.220.855	21.093	3,8	3.891	2.114.760	1.908	158.103	212	14.11

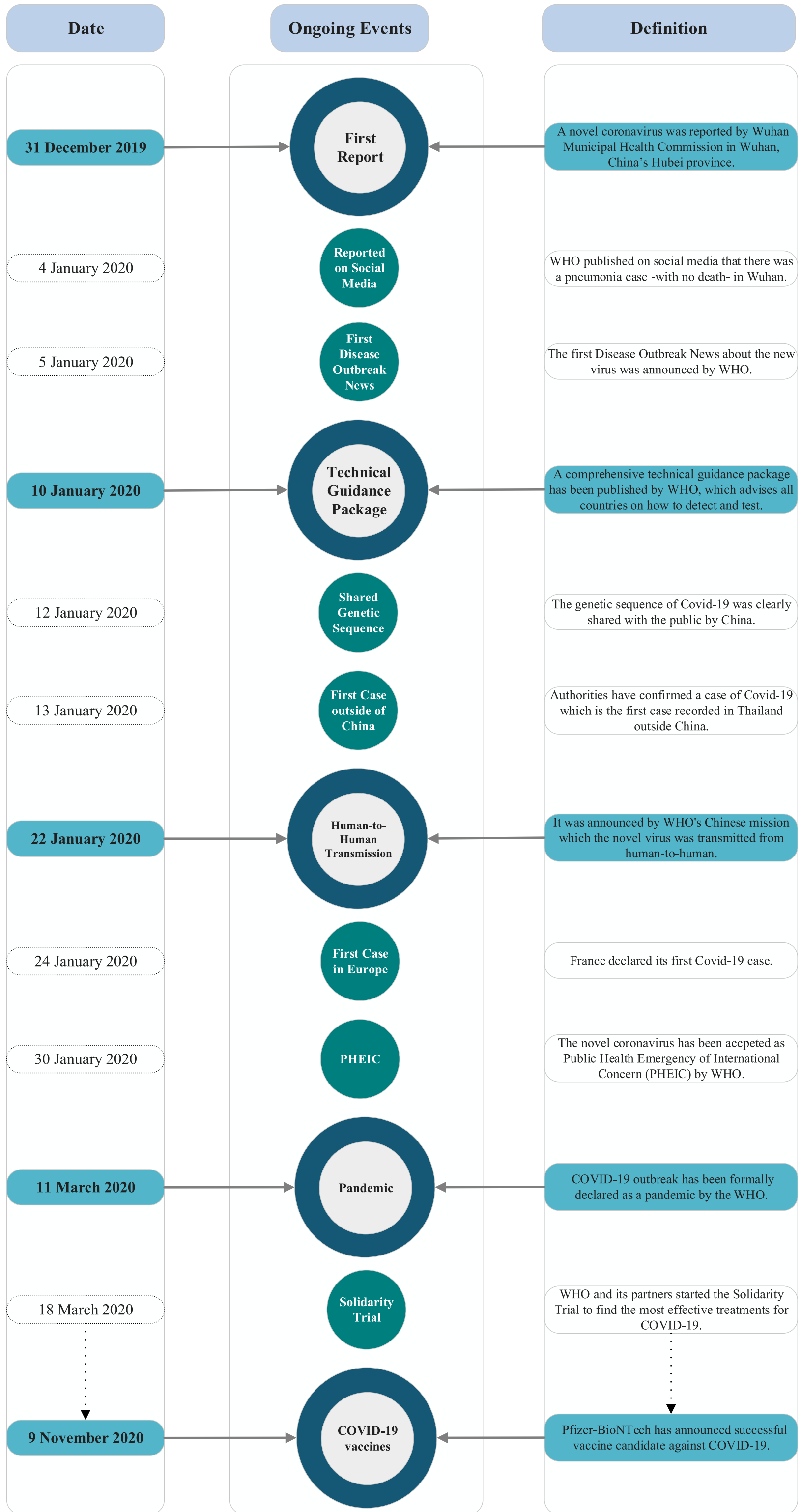

Fig. 1. The key milestones in the spread of COVID-19.

of COVID-19 tests and the number of discharges from the new hospitals have increased compared to the previous day.

COVID-19 arrived at a later date in Turkey on 11th of March 2020 but soon spread to all cities in the country. COVID-19 has affected all 81 provinces of Turkey within a month. The government of Turkey has taken measures to prevent the coronavirus pandemic faster than European countries and other continents. While European countries were beginning to take steps on average 39–54 days after the first case, Turkey has acted in the first few days.[4] The initial preventive measures of COVID-19 taken by the government of Turkey are presented as a timeline in Fig. 2 (Adopted from https://www.aa.com.tr/en/info/infographic/17889).

[4] https://www.aa.com.tr/en/europe/turkey-acted-faster-than-europe-in-tackling-covid-19/1798213

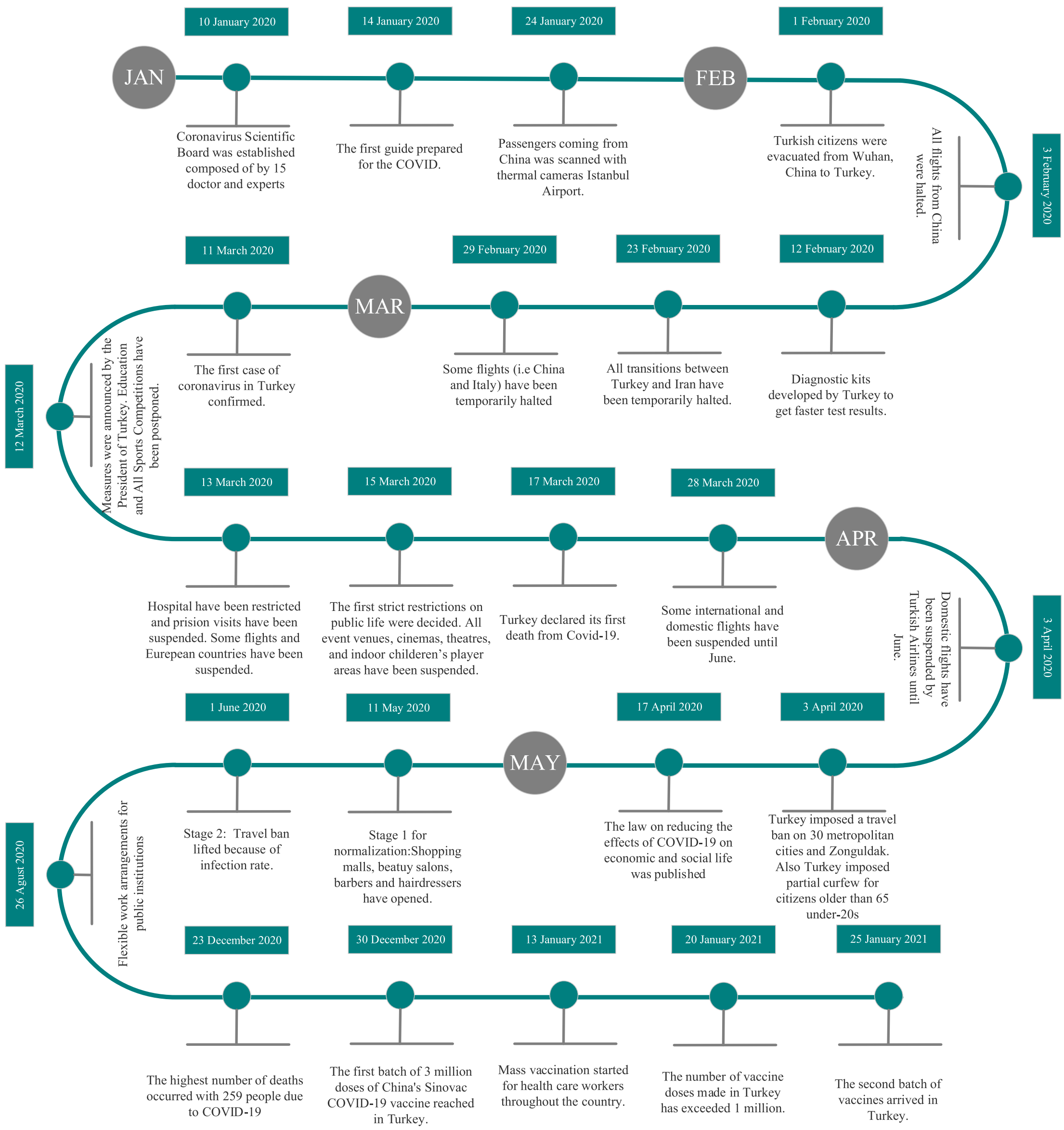

Fig. 2. The timeline considering COVID-19 in Turkey.

Fig. 3 provides information on the first coronavirus cases and their dates for some countries.[5] According to Coronavirus Resource Center of the Johns Hopkins University Statistic Map, as can be seen on Fig. 3, the first case was reported in China, Thailand (the number of cases: 2), USA (the number of cases: 2), Singapore (1), France (2), Canada (1), Germany (1), United Arab Emirates (4), Finland (1), respectively. The first case in Europe was reported in France in 24th of January 2020. The Government of Turkey has recorded its first case of COVID-19 in 11th of March 2020.

4. Data and methods

In this study, the number of cases for 10 different countries are used, predicting from the data obtained from the database. Azerbaijan, India, Indonesia, Italy, Malaysia, Pakistan, Turkey, Spain, United Kingdom and United States are randomly selected countries for the prediction. The first cases examined data, minimum number of reported cases, maximum number of reported cases, mean, standard deviation and number of reported cases are given in Table 3. According to the table, the spread of the virus seems too uncertain in the US and India while Azerbaijan and Malaysia seem to have lower cases and the rate of spreading when comparing with other countries. These values show us whether a robust decision methodology is proposed or not.

In the second phase, the number of patients and deaths are predicted for Turkey. The data is provided by the Republic of Turkey Ministry of Health.[6] Basic statistics regarding the analyzed data are given in Table 4. Accordingly, the high standard deviation indicates that the rate of disease spread over time has a large uncertainty. When the number of patients is examined, it can be observed that the increase in the number of cases may have negative effects on the continuity of service in the health system.

The methodology in the study is summarized in Fig. 4. Accordingly, the prediction of the parameters will be done more accurately using an adaptive prediction model that will provide the best prediction based on historical dataset. In such a pandemic, where the degree of uncertainty is extremely high, the number of cases, patients, deaths and intensive care unit patients increase or decrease rapidly due to many different parameters. In the proposed system, the artificial neural

[5] https://coronavirus.jhu.edu/map.html

[6] covid19.saglik.gov.tr

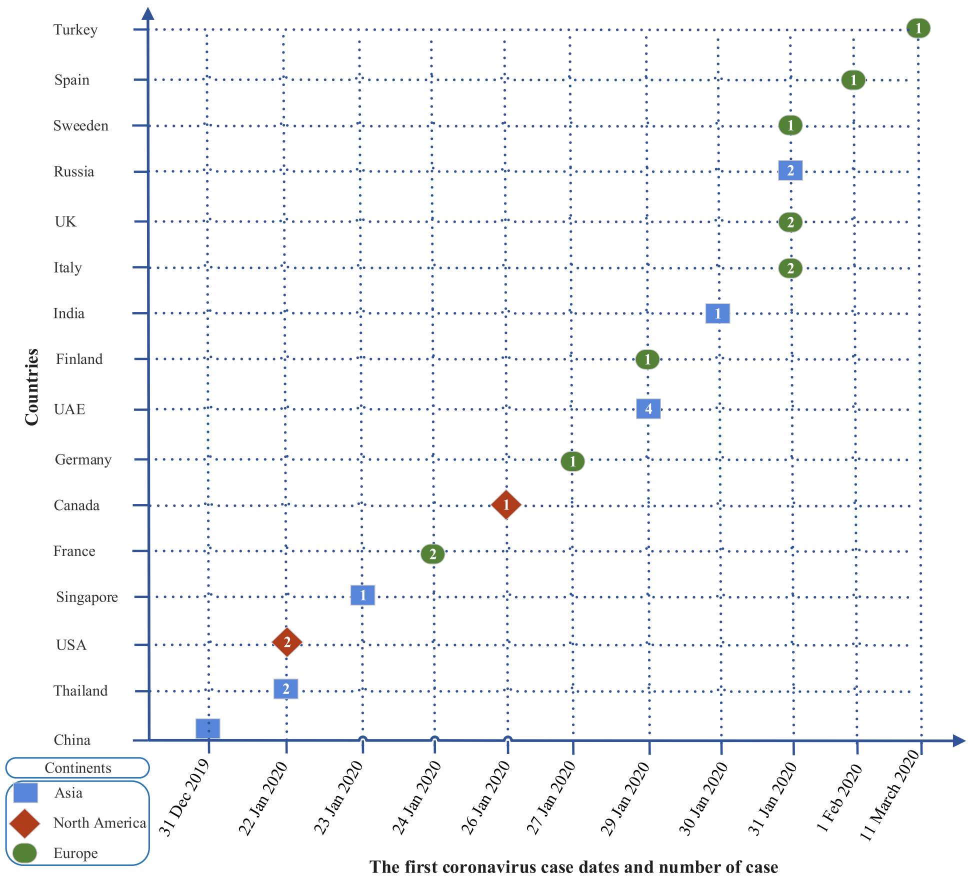

Fig. 3. First COVID-19 cases by date and countries.

Table 3
Description of the dataset.

Country	First date	Min	Max	Mean	Std. Dev.	N
Azerbaijan	03/01/2020	0	4451	639.29	1043.66	385
India	01/30/2020	0	97894	29217	27262.1	398
Indonesia	03/02/2020	0	14518	3801	3454.43	384
Italy	01/31/2020	0	40902	8454.23	9922.94	398
Malaysia	01/25/2020	0	5728	835.49	1178.39	398
Pakistan	02/25/2020	0	12073	1611.33	1438.13	390
Turkey	03/11/2020	0	44506	5816.26	7935.69	375
Spain	02/01/2020	0	93822	8091.51	13685.1	398
United Kingdom	01/31/2020	0	68053	10144.8	13663	424
United States	01/22/2020	0	300295	75029.5	68303	397

Table 4
Description of the dataset.

Statistic	The number of daily patients	The number of daily deaths
Min	0	14
Max	7.381	259
Mean	2.086	82
Std. Dev.	1.581	70
N	319	303

network models and deep learning algorithms are dynamically examined with the test performance and re-training is designed in case of deviation from the predictions. Thereby, a robust decision support system can be developed for different countries.

Model A

In the model A, an artificial neural network model based on different learning algorithms is developed. Artificial neural networks provide tools that basically simulates the activity of the human brain to make predictions [40]. The ANN provides models to make predictions on the complex structures without using specific assumptions [41]. The network structure of the neural network model application (in case of one hidden layer) is given in Fig. 5. According to Fig. 5, the output j can be calculated using Eq. (1) in case of one hidden layer.

$$Y_j = g_j\left(\sum_{k=1}^{K} f_k\left(\sum_{i=1}^{10} W_{ik} * X_i + B_{1k}\right) + B_{2j}\right) \tag{1}$$

where g_j and f_k are the transfer functions of jth output and kth hidden neurons respectively. X_i and Y_j denote the input and output values of the network which are the number of cases in this study. W_{ik} denotes the weights between input neuron i and hidden neuron k. Finally, B_{1k} and B_{2j} are the bias values of hidden neuron k and output neuron j, respectively.

The ANN should be trained regarding the problem dataset. Therefore, the network is divided into three parts which are training, test and validation datasets. The application needs to predict next periods; therefore, the last 10% of the overall dataset is divided as the test dataset. The rest of dataset is divided as training and validation

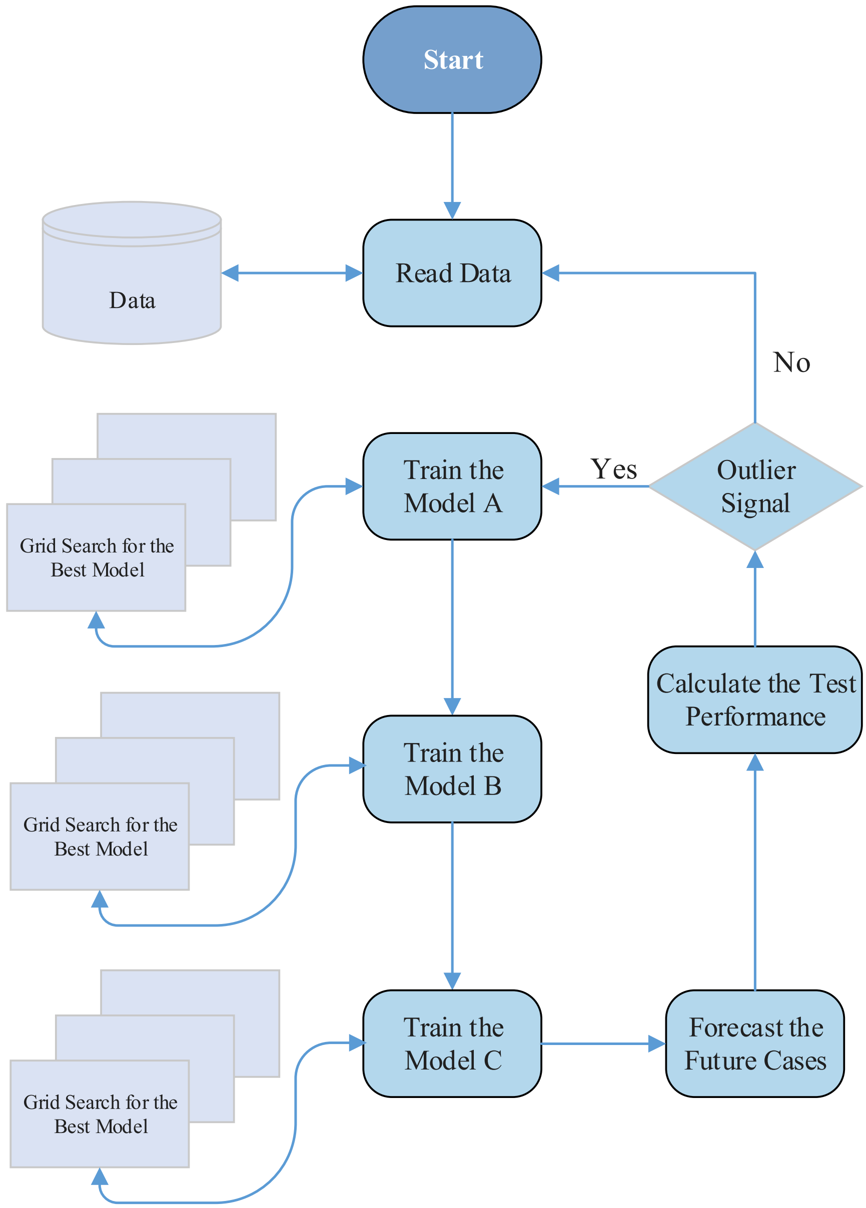

Fig. 4. Adaptive training mechanism of the study.

dataset randomly (ratios are assumed as 90% for training and %10 for validation).

During the training procedure, the main aim is to find the best W_{ik}, B_{1k} and B_{2j} values which provide the minimum of the mean squared error value. The calculation of the mean squared error (MSE) of the output j is given in Eq. (2) where T_{lj} is the target value of output j in sample l and the N is the number of samples are used as training dataset. The training is over when the number of user-defined number of epochs are reached.

$$MSE_j = \frac{\sum_{l=1}^{N} \left(Y_{lj} - T_{lj}\right)^2}{N} \tag{2}$$

There are different training functions, transfer functions and user-defined parameters influencing the prediction performance. Therefore, a grid search mechanism is applied to find the best formulation of the ANN regarding the dataset. In the grid search, the Levenberg–Marquardt [42], the BFGS quasi-Newton backpropagation [43] and resilient backpropagation [44] algorithms are tried as training functions of the proposed ANN. Similarly; linear, positive linear, log-sigmoid and hyperbolic transfer functions are selected as transfer functions. It is undertaken that the proposed ANN having one or two hidden layers with the number of hidden neurons which are tried as 2, 5 and 10 according to try–error results. Finally, the number of epochs is selected as 1000, 1500 and 2000.

Model B

Deep learning and time series artificial neural network algorithms are beneficial to find the best prediction method in time series as studied in Zhu et al. [45]. Model B includes one of the deep learning methods called Recurrent Neural Networks (RNNs) and Long Short Term Memory (LSTM) method. The LSTM extends the neural networks considering the memory states using LSTM cells as depicted in Fig. 6. A general LSTM cell includes memory cell state controlled by input, forget and output gates; therefore, it may remember or forget the information [46]. The LSTM cell structure is depicted in Fig. 7. $\int$ denotes the activation function, x denotes the input variable, C is the cell state, and H is the hidden state. The LSTM also needs training as ANN finds the best parameter values, the same training and test datasets are used with the model A to make a fair comparison. In the training, Adam algorithm which is a gradient descent algorithm is applied to the network [47].

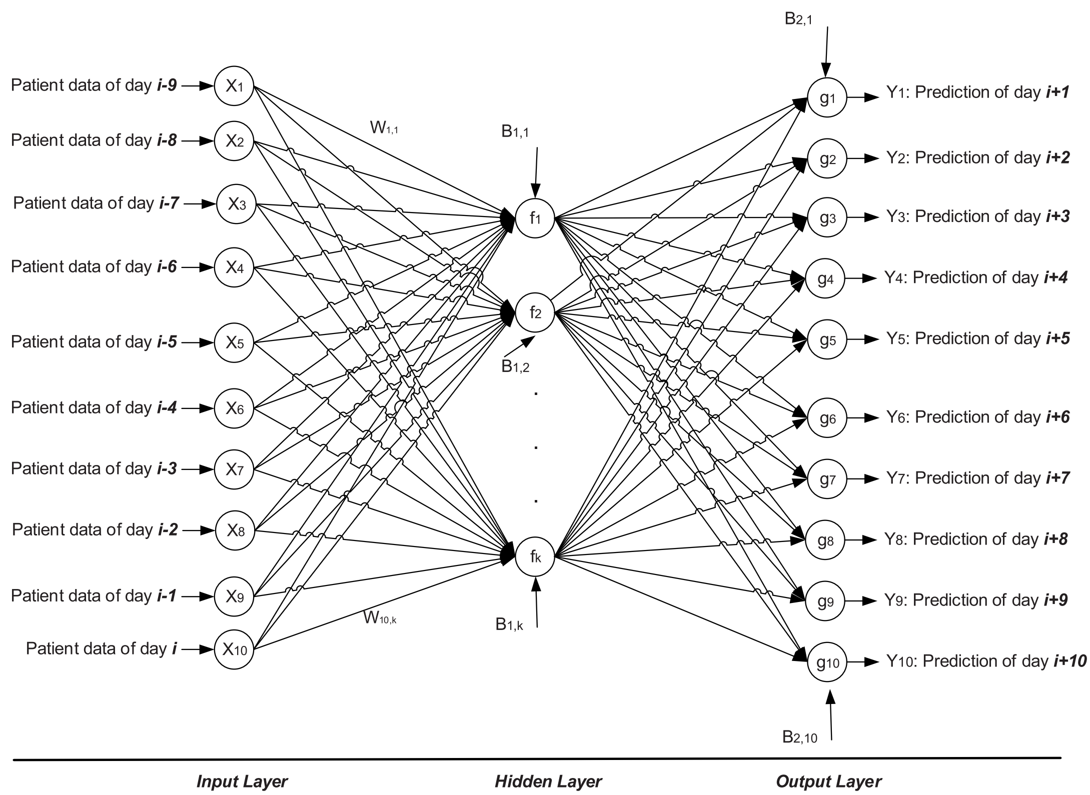

Fig. 5. Model A architecture.

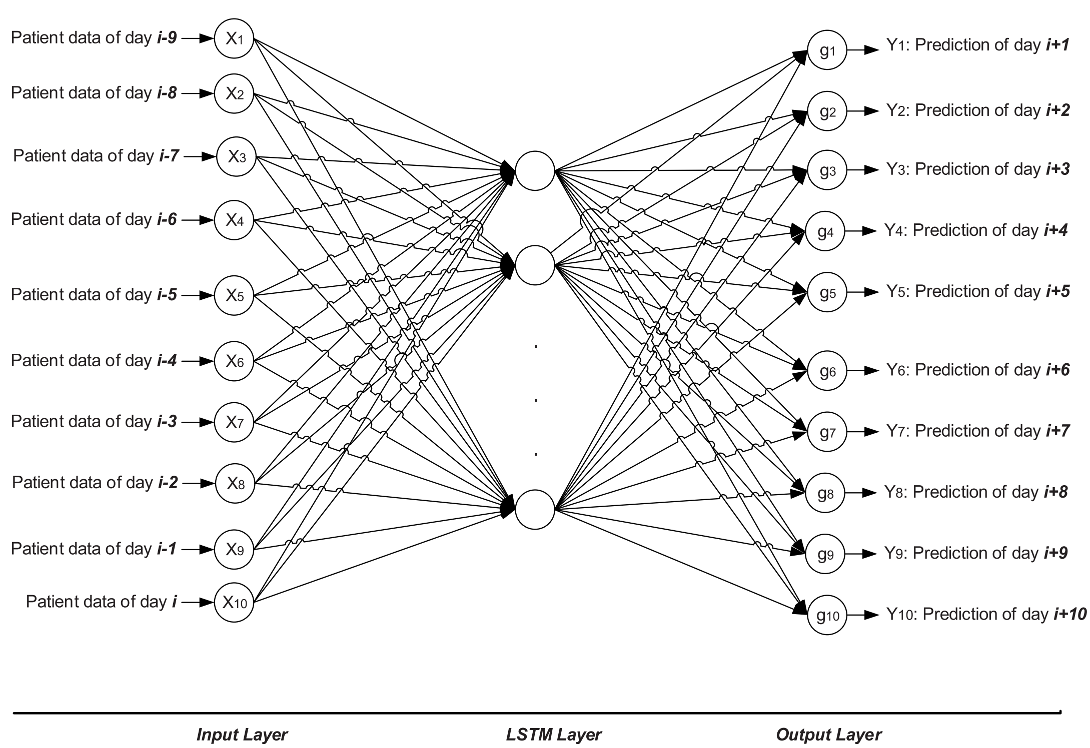

Fig. 6. Model B architecture.

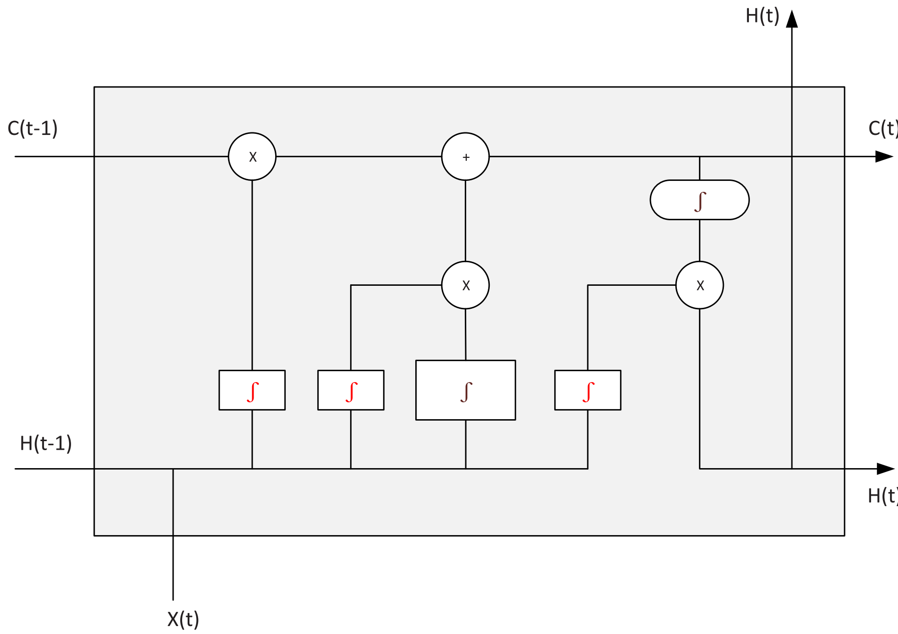

Fig. 7. LSTM cell architecture.

where $\int$ denotes the activation function, x denotes the input variable, C is the cell state, and H is the hidden state.

Model C

Model C has similar multi-layer perceptron to the model A. The only difference between the model A and model C is the learning algorithm. Model C uses Adam algorithm to train neural network architecture depicted in Fig. 5. Similar to model A and B, the same training and test datasets are used in model C to make a fair comparison.

Performance Analysis

The accuracy of the proposed methodologies is evaluated by using statistical performance criteria. The test performance is obtained via mean absolute percentage error (MAPE) values of the test dataset. If the MAPE values of the test dataset is greater than allowable limits (0.2), the model is re-trained to find the best prediction mechanism. The MAPE calculation of the output j is given in Eq. (3).

$$MAPE_j = \frac{\sum_{l=1}^{N}\left(\frac{\left|Y_{lj}-T_{lj}\right|}{T_{lj}}\right)}{N} * 100 \tag{3}$$

5. Findings

In this section, the results obtained from three models to predict the number of cases are given in Section 5.1. The best algorithm is applied to predict the number of patients and deaths in Section 5.2. Model A is programmed with Matlab software package and model B and C are coded with Python programming language having Tensor flow framework.

5.1. Prediction results of reported cases

The models are trained with a grid-search methodology by which the experiments rely on different numbers of learning rates (from 0.001 to 0.5 with 0.01 increments), neurons (1, 5, 10 and 20), epochs (100, 500, 1000 and 5000) and batches (only for deep models as 25, 50 and 100). 5-fold cross validation method is applied for each experiment and MSE values are used as performance function. According to the results, the learning rate and epochs are selected as 0.01 and 1000 for all models. The number of neurons is selected as 10 for model A, 20 for model B and 5 for model C.

When the 10 days prediction results for all countries are examined as given in Table 5, it can be observed that noteworthy results are obtained, especially for India, Indonesia, Malaysia, Pakistan, and Turkey. It seems that good and acceptable predictions are generated for other countries as well. Since training is done using the grid-search approach in creating the prediction models, the prediction models created make good predictions according to different countries. When evaluating the models, the best estimates are usually made with models A and C, although model B gave better results for some countries. In general, when these three methods were compared, the significance value was calculated as 0.423 (F = 0.888) in the ANOVA test. Accordingly, it can be said that all three methods give statistically similar results. When the dataset information in Table 5 is evaluated, it is seen that model B produces more effective results for countries with a higher mean number of cases and higher variances.

Fig. 8 shows the MAPE values of models A, B, and C for the test dataset during the next ten days based on the average of all countries' predictions. Accordingly, although it was observed that all three models gave close results, it was observed that model B generally had higher MAPE values. The best model for predicting the next first day is model B. On the next seventh and eighth days, the best method was determined as model A, and for all other days, the best prediction model was determined as model C. The proposed decision support system will give the best prediction result among these three models to provide predictions adaptively. According to the prediction errors, it is seen that the MAPE values for the next seven days are lower than 30%, which indicates that these prediction models give quite beneficial results. After the eighth day, it was evaluated that the prediction methods gave reasonable results. Finally, the prediction error is increased during the predicted periods.

It is hard to compare the proposed methods with the existing literature due to the prediction length. The proposed methods predict next ten days confirmed cases to make a policy for decreasing of spread. The proposed methods appear to predict daily confirmed cases with lower MAPE values for India and United States when compared with the study of Dash et al. [27] which predicts 90 day future values. They predict the daily confirmed cases with MAPE values are 17.74 for India and 20.27 for United States while Model A having 13.85 MAPE value for India and Model B having 18.57 MAPE value for United States.

Table 5
MAPE values of ten-days reported cases' prediction.

Country	Model A		Model B		Model C	
	Training MAPE	Test MAPE	Training MAPE	Test MAPE	Training MAPE	Test MAPE
Azerbaijan	42.78	27.5	41.09	33.79	42.1	27.48
India	23.43	13.85	29.87	30.66	22.45	15.09
Indonesia	20.57	16.47	19.18	19.09	18.77	15.83
Italy	30.63	27.32	39.1	37.38	27.34	26.65
Malaysia	89.56	15.95	95.32	15.17	91.9	17.06
Pakistan	36.42	14.61	33.04	18.94	33.54	12.96
Turkey	51.77	11.76	28.05	22.95	32.22	17.12
Spain	63.58	41.41	40.7	26.14	56.97	36.1
United Kingdom	60.77	39.28	64.06	34.62	58.64	37.58
United States	23.16	29.48	26.19	18.57	21.2	25.85

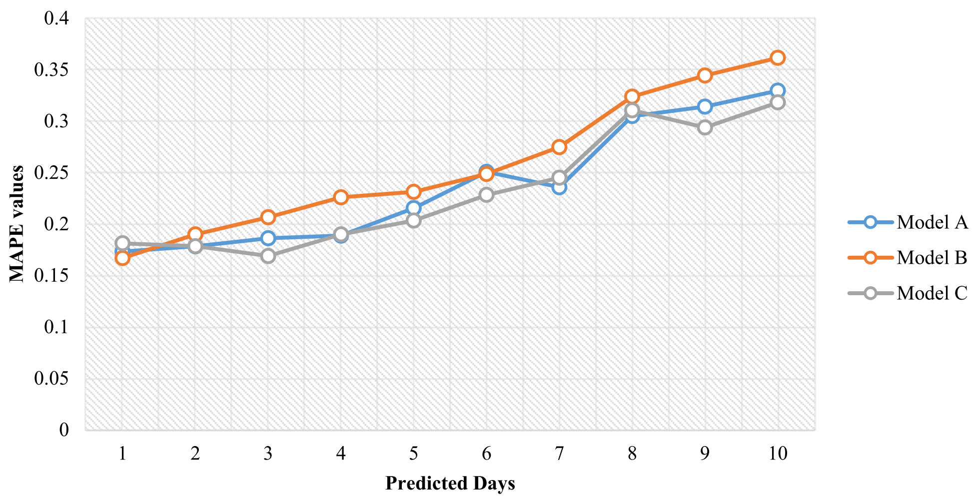

Fig. 8. Prediction performance of models.

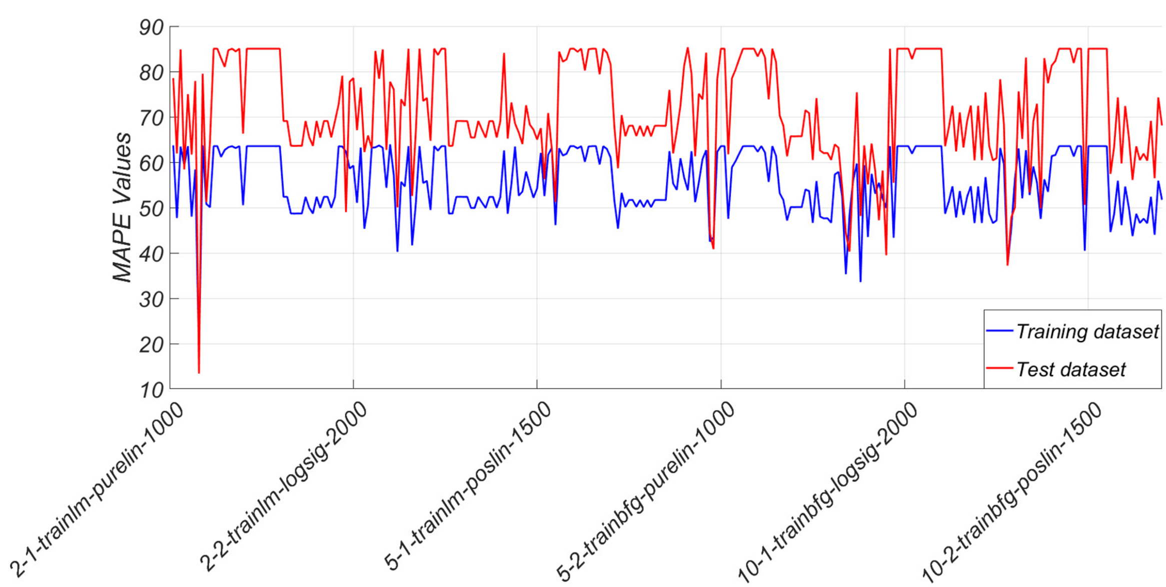

Fig. 9. Grid search results for the daily number of patients.

5.2. Prediction results of the number of patients and deaths

In the second section of the results, the number of daily patients and deaths are examined. As a result of the grid search analysis carried out by the artificial neural network model to obtain the number of daily patients shown in Fig. 9, the best model was obtained by the single layer artificial neural network when there are 2 hidden neurons, when Levenberg–Marquardt learning algorithm and linear transfer function with 1500 epochs are used. The model training and test performance are shown in Fig. 10. According to the results, the MAPE values are increased with the increase of projected day. Unsurprisingly, there are different number of factors that influence the variation of the number of patients; therefore, it is somewhat difficult to predict the future number of patients accurately. The model having at most 22% MAPE value on the training dataset; at the same time, all of the test MAPE values

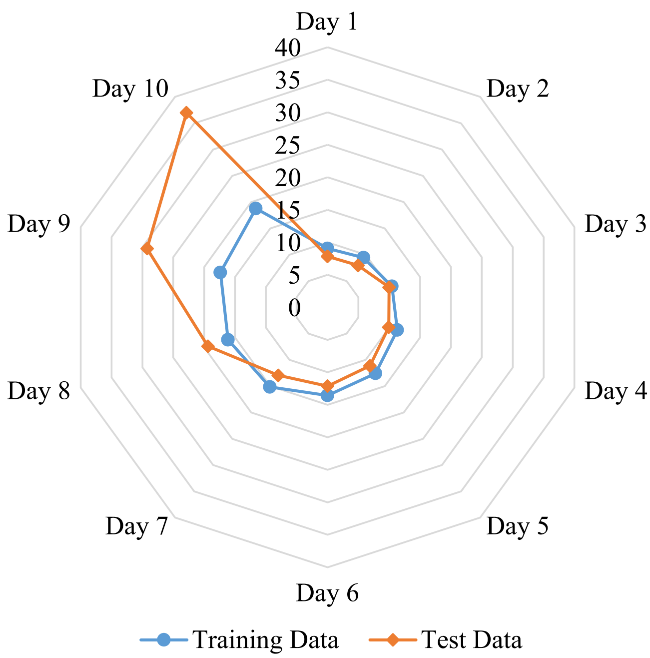

Fig. 10. MAPE values of the prediction of daily number of patients.

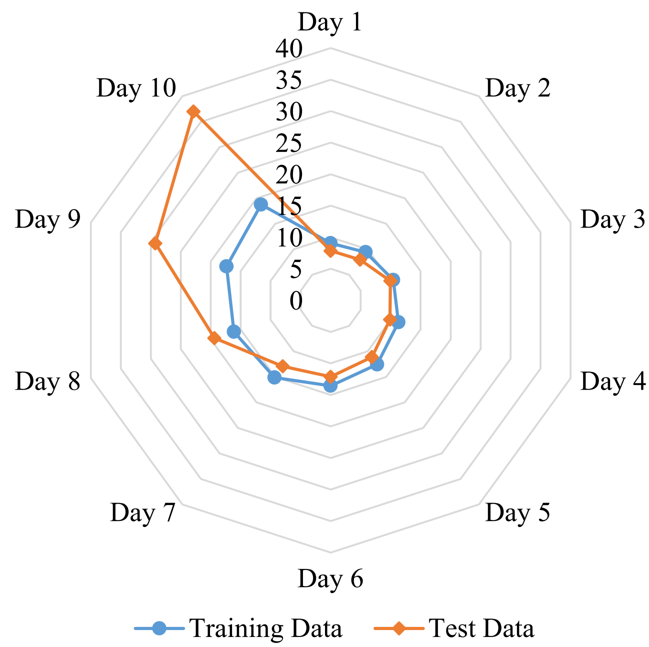

Fig. 12. MAPE values of the prediction of daily number of fatality.

are lower than 15%. It provides a reasonable test performance for the prediction of daily patients.

Another important parameter is the prediction of the number of daily fatality. The accurate prediction of the number of daily fatality in the coming days will ensure that more appropriate policies regarding the spread rate of the disease can be set out. According to the grid search results which are presented in Fig. 11, it was seen that the artificial neural network model with a single hidden layer, 10 hidden neurons, Levenberg–Marquardt learning algorithm and positive linear transfer function gave the best result with 1000 epochs. According to the best model results shown in Fig. 12, the MAPE values for all future days during the training dataset were determined below 20% . In the test phase, it was observed that the MAPE values were less than 20% except for the next 9th and 10th days. In general, it can be interpreted that as the forecast period gets longer, the prediction errors increase, but acceptable results can be obtained for the next 8 days.

6. Conclusions and discussion

In this study, three prediction models basis on artificial neural networks and other deep learning algorithms is designed to forecast the number of COVID-19 cases and fatality using artificial intelligence method. Ten different country including Italy, United Kingdom, United States and Spain which are having widely spread of COVID-19 are considered to analyze the performance of the prediction algorithm. The case data of COVID-19 between the start of disease in the countries and 23 January 2021 is analyzed. The performance of the proposed methods is compared according to the MAPE criteria. The proposed methodology predicts next ten days thereby a policy for decreasing the spread can be done by the policymakers using the proposed methodologies.

As aforementioned, three models are compared to each other for different countries. The models seem slightly closer prediction performances when the mean and standard deviation of the MAPE is

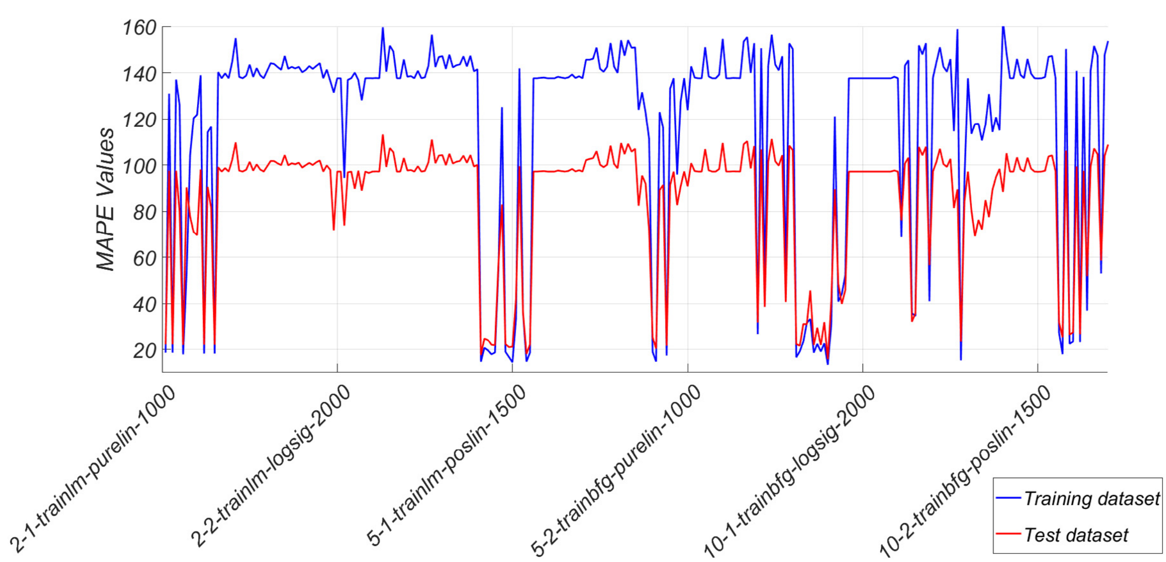

Fig. 11. Grid search results for the number of daily fatality.

considered. However, model A and C predicts lower test MAPE values than the model B. The average MAPE values of the model A, B and C in test dataset are 23.763, 25.731 and 23.172, respectively. The minimum MAPE values are 11.76, 15.17 and 12.96, respectively. As a result of this comparison, model A and C can be beneficial for prediction while model B has closer performance with them.

This study can be expanded to be used for other diseases so that the healthcare systems respond more effectively during an outbreak or a pandemic. One limitation of this study is that the proposed adaptive prediction approach for the number of daily patients and fatality considers the past time data. Different parameters such as the number of cases per day, the number of intensive care unit patients may also be considered. Furthermore, different parameters may be effective for the spread of COVID-19. In this context, more accurate results may be achieved with more detailed datasets and the analysis of most effective parameters. The mutations that can influence the spread of COVID-19 can be included in these models. Different prediction methods, ANFIS etc. may also be added to the proposed system.

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.

References

[1] Q. Zhao, M. Meng, R. Kumar, Y. Wu, J. Huang, Y. Deng, L. Yang, et al., Lymphopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A systemic review and meta-analysis, Int. J. Infect. Dis. (2020).

[2] M. Karlsson, T. Nilsson, S. Pichler, The impact of the 1918 Spanish flu epidemic on economic performance in Sweden: An investigation into the consequences of an extraordinary mortality shock, J. Health Econ. 36 (2014) 1–19.

[3] N. Petrosillo, G. Viceconte, O. Ergonul, G. Ippolito, E. Petersen, COVID-19, SARS and MERS: Are they closely related? Clin. Microbiol. Infect. 26 (6) (2020) 729–734.

[4] S. Mallapaty, What the cruise-ship outbreaks reveal about COVID-19, Nature 580 (7801) (2020) 18.

[5] S. Tuli, S. Tuli, R. Tuli, S.S. Gill, Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing, Internet of Things (2020) 100222.

[6] P. Wang, J.A. Lu, Y. Jin, M. Zhu, L. Wang, S. Chen, Statistical and network analysis of 1212 COVID-19 patients in Henan, China, Int. J. Infect. Dis. (2020).

[7] K. Liu, Y. Chen, R. Lin, K. Han, Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients, J. Infect. (2020).

[8] M.A. Turk, S.D. Landes, M.K. Formica, K.D. Goss, Intellectual and developmental disability and COVID-19 case-fatality trends: TriNetX analysis, Disabil. Health J. (2020) 100942.

[9] Z. Yang, Z. Zeng, K. Wang, S.S. Wong, W. Liang, M. Zanin, J. He, et al., Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions, J. Thorac. Dis. 12 (3) (2020) 165.

[10] A.S.R. Srinivasa Rao, J.A. Vazquez, Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey when cities and towns are under quarantine, Infect. Control Hosp. Epidemiol. (2020) http://dx.doi.org/10.1017/ice.2020.61.

[11] R. Vaishya, M. Javaid, I.H. Khan, A. Haleem, Artificial intelligence (AI) applications for COVID-19 pandemic, Diabetes Metab. Syndr. Clin. Res. Rev. (2020) http://dx.doi.org/10.1016/j.dsx.2020.04.012.

[12] G.S. Randhawa, M.P.M. Soltysiak, H. El Roz, C.P.E. de Souza, K.A. Hill, L. Kari, Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study, PLoS One 15 (2020) e0232391, http://dx.doi.org/10.1371/journal.pone.0232391.

[13] R. Madurai Elavarasan, R. Pugazhendhi, Restructured society and environment: A review on potential technological strategies to control the COVID-19 pandemic, Sci. Total Environ. (2020) http://dx.doi.org/10.1016/j.scitotenv.2020.138858.

[14] V.K.R. Chimmula, L. Zhang, Time series forecasting of COVID-19 transmission in Canada using LSTM networks, Chaos Solitons Fractals (2020) http://dx.doi.org/10.1016/j.chaos.2020.109864.

[15] J.S. Cobb, M.A. Seale, Examining the effect of social distancing on the compound growth rate of COVID-19 at the county level (United States) using statistical analyses and a random forest machine learning model, Public Health 185 (2020) 27–29, http://dx.doi.org/10.1016/j.puhe.2020.04.016.

[16] A.A. Ardakani, A.R. Kanafi, U.R. Acharya, N. Khadem, A. Mohammadi, Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks, Comput. Biol. Med. 121 (2020) 103795, http://dx.doi.org/10.1016/j.compbiomed.2020.103795.

[17] M. R, B. A, S. K, COVID-19 outbreak: Tweet based analysis and visualization towards the influence of coronavirus in the world, GEDRAG Organ. Rev. 33 (2020) http://dx.doi.org/10.37896/GOR33.02/062.

[18] R. Sujath, J.M. Chatterjee, A.E. Hassanien, A machine learning forecasting model for COVID-19 pandemic in India, Stoch. Environ. Res. Risk Assess 34 (2020) 959–972, http://dx.doi.org/10.1007/s00477-020-01827-8.

[19] A. Mollalo, K.M. Rivera, B. Vahedi, Artificial neural network modeling of novel coronavirus (COVID-19) incidence rates across the continental United States, Int. J. Environ. Res. Public Health 17 (4204) (2020) http://dx.doi.org/10.3390/ijerph17124204.

[20] C. Shen, A. Chen, C. Luo, J. Zhang, B. Feng, W. Liao, Using reports of symptoms and diagnoses on social media to predict COVID-19 case counts in Mainland China: Observational infoveillance study, J. Med. Internet Res. 22 (2020) e19421, http://dx.doi.org/10.2196/19421.

[21] H. Kang, L. Xia, F. Yan, Z. Wan, F. Shi, H. Yuan, H. Jiang, D. Wu, H. Sui, C. Zhang, D. Shen, Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning, IEEE Trans. Med. Imaging (2020) 1, http://dx.doi.org/10.1109/TMI.2020.2992546.

[22] G. Pinter, I. Felde, A. Mosavi, P. Ghamisi, R. Gloaguen, COVID-19 pandemic prediction for hungary; A hybrid machine learning approach, Mathematics (2020) http://dx.doi.org/10.3390/math8060890.

[23] S. Vaid, A. McAdie, R. Kremer, V. Khanduja, M. Bhandari, Risk of a second wave of COVID-19 infections: Using artificial intelligence to investigate stringency of physical distancing policies in North America, Int. Orthop. (2020) http://dx.doi.org/10.1007/s00264-020-04653-3.

[24] E. Fayyoumi, S. Idwan, H. Aboshindi, Machine learning and statistical modelling for prediction of novel COVID-19 patients case study: Jordan, Int. J. Adv. Comput. Sci. Appl 11 (2020) 122–126, http://dx.doi.org/10.14569/IJACSA.2020.0110518.

[25] A.J. Aljaaf, T.M. Mohsin, D. Al-Jumeily, M. Alloghani, A fusion of data science and feed-forward neural network-based modelling of COVID-19 outbreak forecasting in IRAQ, J. Biomed. Inform. 118 (2021) 103766.

[26] M. Alodat, Using deep learning model for adapting and managing COVID-19 pandemic crisis, Procedia Comput. Sci. 184 (2021) 558–564.

[27] S. Dash, C. Chakraborty, S.K. Giri, S.K. Pani, Intelligent computing on time-series data analysis and prediction of COVID-19 pandemics, Pattern Recognit. Lett. (2021).

[28] A.A. Elhag, T.A. Aloafi, T.M. Jawa, N. Sayed-Ahmed, F.S. Bayones, J. Bouslimi, Artificial neural networks and statistical models for optimization studying COVID-19, Results Phys. 25 (2021) 104274.

[29] M.A. Elleuch, A.B. Hassena, M. Abdelhedi, F.S. Pinto, Real-time prediction of COVID-19 patients health situations using artificial neural networks and fuzzy interval mathematical modeling, Appl. Soft Comput. 110 (2021) 107643.

[30] N.N. Hamadneh, M. Tahir, W.A. Khan, Using artificial neural network with prey predator algorithm for prediction of the COVID-19: The case of Brazil and Mexico, Mathematics 9 (2) (2021) 180.

[31] L.R. Kolozsvári, T. Bérczes, A. Hajdu, R. Gesztelyi, A. Tiba, I. Varga, J. Zsuga, et al., Predicting the epidemic curve of the coronavirus (SARS-CoV-2) disease (COVID-19) using artificial intelligence: An application on the first and second waves, Inform. Med. Unlocked 25 (2021) 100691.

[32] K.N. Nabi, M.T. Tahmid, A. Rafi, M.E. Kader, M.A. Haider, Forecasting COVID-19 cases: A comparative analysis between recurrent and convolutional neural networks, Results Phys. 24 (2021) 104137.

[33] S. Namasudra, S. Dhamodharavadhani, R. Rathipriya, Nonlinear neural network based forecasting model for predicting COVID-19 cases, Neural Process. Lett. (2021) 1–21.

[34] Z.E. Rasjid, R. Setiawan, A. Effendi, A comparison: Prediction of death and infected COVID-19 cases in Indonesia using time series smoothing and LSTM neural network, Procedia Comput. Sci. 179 (2021) 982–988.

[35] A.H. Shamman, A.A. Hadi, A.R. Ramul, M.M.A. Zahra, H.M. Gheni, The artificial intelligence (AI) role for tackling against COVID-19 pandemic, Mater. Today: Proc. (2021).

[36] M. Shawaqfah, F. Almomani, Forecast of the outbreak of COVID-19 using artificial neural network: Case study Qatar, Spain, and Italy, Results Phys. (2021) 104484.

[37] X. Tang, Z. Li, X. Hu, Z. Xu, L. Peng, Self-correcting error-based prediction model for the COVID-19 pandemic and analysis of economic impacts, Sustainable Cities Soc. (2021) 103219.

[38] G. Toğa, B. Atalay, M.D. Toksari, COVID-19 prevalence forecasting using autoregressive integrated moving average (ARIMA) and artificial neural networks (ANN): Case of Turkey, J. Infect. Public Health (2021).

[39] S.N. Zisad, M.S. Hossain, M.S. Hossain, K. Andersson, An integrated neural network and SEIR model to predict COVID-19, Algorithms 14 (3) (2021) 94.

[40] M.M. Hamed, M.G. Khalafallah, E.A. Hassanien, Prediction of wastewater treatment plant performance using artificial neural networks, Environ. Model. Softw. 19 (10) (2004) 919–928.

[41] Z. Zhang, A gentle introduction to artificial neural networks, Ann. Transl. Med. 4 (19) (2016) 370, http://dx.doi.org/10.21037/atm.2016.06.20.
[42] M.T. Hagan, M.B. Menhaj, Training feedforward networks with the Marquardt algorithm, IEEE Trans. Neural Netw. 5 (6) (1994) 989–993.
[43] J.E. Dennis, R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, NJ, 1983.
[44] M. Riedmiller, H. Braun, A direct adaptive method for faster backpropagation learning: The RPROP algorithm, in: Proceedings of the IEEE International Conference on Neural Networks, 1993, pp. 586–591.
[45] J. Zhu, P. Wu, H. Chen, J. Liu, L. Zhou, Carbon price forecasting with variational mode decomposition and optimal combined model, Physica A 519 (2019) 140–158.
[46] X. Song, Y. Liu, L. Xue, J. Wang, J. Zhang, J. Wang, Z. Cheng, et al., Time-series well performance prediction based on long short-term memory (LSTM) neural network model, J. Pet. Sci. Eng. 186 (2020) 106682.
[47] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.