Predicting the status of COVID-19 active cases using a neural network time series

Peyman Almasinejad[1], Amin Golabpour[2], Fatemeh Ahouz[3], Mohammad Reza Mollakhalili Meybodi[1], Kamal Mirzaie[1], Ahmad Khosravi[4], Marzieh Rohani-Rasaf[5], Azadeh Bastani[3]

[1]Department of Computer Engineering, Maybod Branch, Islamic Azad University, Maybod, Iran
[2]Department of Health Information Technology, School of Allied Medical Sciences, Shahroud University of Medical Sciences, Shahroud, Iran
[3]Department of Computer Engineering, School of Engineering, Behbahan Khatam Alanbia University of Technology, Behbahan, Iran
[4]Ophthalmic Epidemiology Research Center, Shahroud University of Medical Sciences, Shahroud, Iran
[5]Department of Epidemiology, School of Public Health, Shahroud University of Medical Sciences, Shahroud, Iran

Article Info

ABSTRACT

The design of intelligent systems for analyzing information and predicting the epidemiological trends of the disease is rapidly expanding because of the coronavirus disease (COVID-19) pandemic. The COVID-19 datasets provided by Johns Hopkins University were included in the analysis. This dataset contains some missing data that is imputed using the multi-objective particle swarm optimization method. A time series model based on nonlinear autoregressive exogenou (NARX) neural network is proposed to predict the recovered and death COVID-19 cases. This model is trained and evaluated for two modes: predicting the situation of the affected areas for the next day and the next month. After training the model based on the data from January 22 to February 27, 2020, the performance of the proposed model was evaluated in predicting the situation of the areas in the coming two weeks. The error rate was less than 5%. The prediction of the proposed model for April 9, 2020, was compared with the actual data for that day. The absolute percentage error (AE) worldwide was 12%. The lowest mean absolute error (MAE) of the model was for South America and Australia with 3 and 3.3, respectively. In this paper, we have shown that geographical areas with mortality and recovery of COVID-19 cases can be predicted using a neural network-based model.

Corresponding Author:

Amin Golabpour
Department of Health Information Technology, School of Allied Medical Sciences, Shahroud University of Medical Sciences
Shahroud University of Medical Sciences and Health Services, Hafte Tir Square, Shahrood, Iran
Email: a.golabpour@shmu.ac.ir

1. INTRODUCTION

On December 8, 2019, the Chinese government reported the death of one patient and the hospitalization of 41 others with an unknown etiology in Wuhan [1], [2]. This cluster initiated the novel coronavirus disease (COVID-19) respiratory disease pandemic. While early cases of the disease were linked to the wet market, the human-to-human transmission had led to the widespread outbreak of the virus through China [3]. On January 30, 2020, the World Health Organization (WHO) announced the emergence of COVID-19 as a public health emergency with international concern (PHEIC) [4].

By March 16, 2020, WHO had reported the COVID-19 statistics in China and outside of China. However, since March 17, 2020, because of widespread prevalence on all continents, the number of

confirmed cases and deaths on each continent was separately expressed [4]. According to the 49th COVID-19 weekly epidemiological update by WHO released on 20 July 2021, globally, COVID-19 weekly case incidence increased with an average of around 490,000 cases reported each day. As of 18 July 2021, the global number of confirmed cases was 190,169,833 [5].

Because of the widespread and growing prevalence of COVID-19 across the world, several works have examined different aspects of the disease. Most of these include identifying: i) the source of the virus and its gene sequences analysis [6], [7], ii) analysis of patient information [8], iii) analysis of the first cases in the countries involved [9]–[11], iv) methods of virus detection [12]–[15], v) evaluation of treatment methods [16], and vi) estimating the extent of transmission [17]. Artificial intelligence has an important role in changing the medical care paradigm and can predict various diseases states [18]–[23]. Thus, scores of research have been done in the past year in areas related to the COVID-19 pandemic. The most common topics are: i) the development of health care robots to prevent direct contact of medical personnel with COVID-19 patients [24], [25]; ii) monitoring of public places to determine the distance between persons or identification of people with high temperature [26]–[28]; iii) forecasting the spread of COVID-19 [29]–[31]; iv) automatic diagnosis of patients with COVID-19 [32]; and v) predicting confirmed, recovered, or death cases [33], [34].

Fang *et al.* [35] proposed a methodology called group of optimized and multisource selection (GROOMS), which is an ensemble of 5 groups of prediction methods. They also proposed a new version of polynomial neural network (PNN) called "PNN with corrective feedback (PNN+cf)" that includes two extra pieces of information: i) lagged data and ii) training errors from past iterations of model training to predict the epidemic at an early stage. The authors conducted an experiment on epidemic data from Chinese health authorities from January 21 to February 3 to evaluate the initial stage of the COVID-19 epidemic. A time-series of 14 instances about the suspected cases was run through the GROOMS method for 6-days ahead of the forecast. They compared the result of the model to the other nine available methods and claimed that the PNN+cf method with 136,547 root-mean-square-error (RMSE) was better than the other methods. The RMSE of the other methods was reported from 138.042 to 1744.5256.

Corona tracker team proposed a susceptible-exposed-infectious-recovered (SEIR) model based on the queried data in their website and made the 240-day prediction of COVID-19 cases in and out of China, started on 20 January 2020 [36]. They predicted that the outbreak would reach its peak on May 23, 2020, and the maximum number of infected individuals will be 425.066 million globally. The authors predicted that it would start to drop around early July 2020 and reach under 10,000 on 14 Sep 2020. Given the information available now, these predictions were far from what really happened around the world.

Wang *et al.* [37] constructed a COVID-19 prediction model using the improved long short-term memory (LSTM) deep learning method with a rolling update mechanism based on the epidemical data provided by Johns Hopkins University. The trends of the epidemic in 150 days ahead were modeled for Russia, Peru, and Iran. Pointing to the importance of preventive measures which would be taken by the government to reduce the spread of COVID-19, the authors estimated that the number of positive cases per day in Iran by mid-November 2020 will reach less than 1,000, however, it did not happen.

Zawbba *et al.* [38] proposed a regression model based on the multilayer perceptron (MLP) to predict the COVID-19 spread for the coming months in nine countries: Italy, the United States, China, Japan, Iran, Egypt, Algeria, Kenya, and Cote d'Ivoire. For each country, they used the number of confirmed cases, the number of deaths, average age, average weather temperature, Bacille Calmette-Guérin (BCG) vaccination, and Malaria treatment. The model was first trained on Chinese data, which were collected from the European Centre for Disease Prevention and Control (ECDC) from 29th December 2019 to 13th December 2020. Then, for the other eight countries, data were downloaded from 22nd January 2020 to 13th December 2020 from the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). The best and worst RMSE for confirmed cases were 105.94 and 77,822.38 in Cote d'Ivoire and the United States, respectively. In the case of predicting the dead cases, the best RMSE was reported 0.91 in Cote d'Ivoire, and the worst one was 792.07 in the USA.

Ahouz and Golabpour [29] introduced a new representation structure of the COVID-19 dataset. By dividing the data set regions into three groups based on the maximum number of confirmed cases of COVID-19 per day and using the least-squares classification algorithm, the authors developed models for predicting the incidence of COVID-19 between March 30, 2020, and April 12, 2020, for each group. The accuracy of the model in predicting the number of COVID-19-approved cases worldwide is reported to be 98.45%.

A review of COVID-19 research shows that most of the predictive models utilized several months of COVID-19 information [20], [34], [38], [39]. However, at the very beginning of pandemics such as COVID-19, due to the lack of information about the factors affecting the recovery or death of patients, design models with appropriate accuracy based on limited non-clinical information is very important. These models

are useful in the containment of the threat and help healthcare administrators to make effective timely decisions in controlling the spread of the disease in addition to reducing community anxiety.

In this study, using less than two-month prevalence data, we proposed a neural network-based time series model for predicting the recovery or death of COVID-19 patients based on general information of each region including latitude, longitude, date, and number of confirmed/recovered/death cases. Based on the datasets provided by Johns Hopkins University [40], we present a new arrangement of the data for the optimal use of outbreak trend information in neighboring regions. In addition, using the information from each region, we predict the final status of COVID-19 cases in each region in the next day and month.

2. METHOD

2.1. Dataset

COVID-19 epidemiological data are compiled by the Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE) [31]. The data are provided in three separate datasets for confirmed, recovered, and death cases since January 22, 2020, and are updated daily. In each of these datasets, there is a record (row) for each geographic area. The variables in each dataset are province/state, country/region, latitude, longitude, and then incremental dates from January 22, 2020. For each geographic area, the value of each date indicates the cumulative number of confirmed/recovered/death cases from January 22, 2020.

In this study, data from the COVID-19 dataset from January 22 to March 9, 2020, entered into the analysis. This information includes the number of confirmed, recovered, and death cases in 265 different geographical areas in 47 days. According to the input requirements of the proposed model, we changed the data representation in the dataset so that instead of three separate datasets for the three groups of confirmed, recovered, and death cases, only one dataset containing the information of all three groups was arranged. In this new dataset, each record (or row) of the dataset contains information about the number of confirmed, recovered, or deaths per day for each geographic area. As a result, the variables in this new dataset are province/state, country/region, latitude, longitude, date (which specifies a specific date), cases (which indicates the number of confirmed, recovered, or death cases), and type (which specifies the type of confirmed, recovered, or death cases). This structure was suggested by Krispin [41].

This rearranged dataset contains 3,436 records and 7 variables that include information on COVID-19 cases from 111 countries and 265 different geographic regions around the world. There are 113,583 confirmed, 3,996 death, and 62,512 recovered cases in the dataset. Pre-processing runs on the dataset before training the proposed model. The dataset is first examined for noise data. Then, the missing data were investigated, and it was found that the data were recorded with a delay, and 56 records out of 256 were missed. To impute the missing data the following procedure was used:

First, the variable is sorted according to the amount of missing data. Put the variables that have the least missing at the beginning, and the variables that have the most missing at the end. Then, with the help of data mining algorithms, a classification model is generated whose independent variables are the variables that do not exist in the missing data, and its dependent variable is a variable with the lowest rate of missing. Then, with the help of the data mining algorithm, the missing data of the dependent variable are imputed. The dependent variable is then added to the set of independent variables, and the next variable is selected as the dependent variable according to the number of missing data, and the data mining algorithm is applied again. This process continues until there is no variable with the missing data. The selection of data algorithms is performed by the multi-objective particle swarm optimization algorithm. Repeat the above process several times so that no change is made to the data. At the end of this process, all the data are filled in, and there are no missing data.

Then, depending on the need of the learning algorithm used in the proposed model, the values of some variables are changed to another format. If necessary, some variables are merged, and new variables are added to the dataset. There are 24 negative values in the cases which are invalid, so we removed them from the dataset. As a result, the number of records decreased to 3,412.

In the province/state, there were 901 missing data out of 3,436 because the information of some countries was generally reported in that country and not for a particular province. Therefore, we imputed the missing data with the name provided in the country region column. We then assigned a unique numeric code to each of these 265 regions. This new code is called code_zone and will be added to the dataset instead of the country and state columns. After applying these preprocessing steps, the resulting dataset is called the alpha dataset.

The records in the alpha dataset were sorted by the value of the date variable. Since in time series models, the algorithm detects the time series pattern, the date variable is removed from the dataset. This leads to the construction of the beta dataset. Therefore, after preprocessing steps, the beta dataset includes 3,412 records with 5 variables (code_zone, latitude, longitude, cases, and type).

Finally, the beta dataset is compiled for two separate two-class prediction models, namely recovery prediction and death prediction. To predict recovered cases, records with "recovered" value in the type variable are considered as class 1, and other records are considered class 0. To predict death cases, records with "death" value in type are considered as class 1, and other records are considered as class 0.

Descriptive statistics such as mean and standard deviation were used to describe the variables. Kruskal Wallis test was used to compare the mean of longitude and latitude variables in the confirmed, death, and recovered groups. This test is to examine the significance of the relationship between latitude and longitude with mortality and recovery. If the test is approved, the latitude and longitude can be used as metadata in the model.

2.2. Constructing the time series prediction model

Because of the presence of metadata in the dataset, an algorithm that utilizes this information to construct time series is preferable. The nonlinear autoregressive exogenous (NARX) neural network is one of these methods [42]. The output of the NARX network at time t+1 is given using (1):

$$y(t) = F\big(y(t-1), y(t-2), \ldots, y(t-n_y), x(t), x(t-1), x(t-2), \ldots, x(t-n_y)\big) \qquad (1)$$

where $F(.)$ is the mapping function of the neural network, $y(t-1), y(t-2), \ldots, y(t-n_y)$ are the true past outputs of this series, called the desired outputs. And $x(t), x(t-1), x(t-2), \ldots, x(t-n_y)$ are the inputs of the NARX which are called exogenous inputs; these metadata are externally determined and influence the desired output of the series. nx is the number of input delays, and ny is the number of output delays [42]. In the proposed model, we aim to determine which areas will experience the death or recovery of COVID-19 cases in the next day. This model is general and does not depend on a specific area. For this reason, we do not use the present value of $x(t)$. Thus, the future value of the time series $y(t)$ is predicted from the past values of $x(t)$ and the actual past values of the time series, $y(t)$.

Figure 1 illustrates the preparation steps of the NARX inputs. In Figure 1, n represents the number of records, $\hat{y}$ and type are the predicted and actual outputs, respectively, which in the death prediction model can be dead (1) or not-dead (0), and in the recovery prediction model can be recovered (1) or not-recovered (0). In addition, latitude, longitude, cases, and their types on day t-d to day t-1 are the exogenous inputs of the network. Given n record information up to time t-1, the proposed model attempts to predict whether COVID-19 active cases will recover (die) at time t in a geographic area. It should be noted that for each geographic area in the dataset, the past information of that area is used to predict its $\hat{y}(t)$. In the model, input and output delays are represented by d or the delay factor.
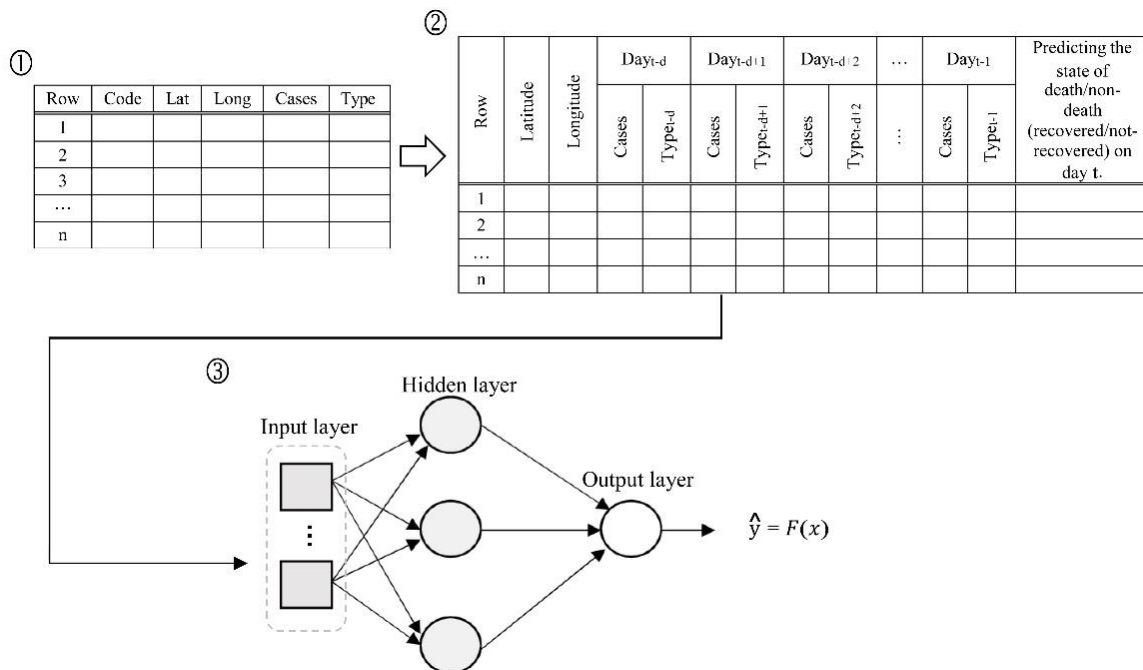

Figure 1. The preparation steps of the NARX input

The delay factor determines how many past observed data of a particular area will participate in predicting the one-step-ahead situation of that area. If the delay is one, it means the model uses only one previously observed point to predict the type of cases at time t. Similarly, if it is k, by using k past observed data points, the model predicts the future at time t. The larger the k, the more information is used to build the model, thus increasing the accuracy of the model as well as its complexity.

In neural network-based time series models, inputs and structure of networks are of great importance. The most important factors are: i) determining the number of neurons in the hidden layers, ii) the learning algorithm for adjusting the weights of the network, and iii) determining the amount of past-observed data to predict the future. In this study, to avoid over-complexity of the network, the maximum number of neurons in hidden layers and the maximum delay are predetermined. For each learning algorithm to adjust network weights, the number of different neurons in the hidden layer is evaluated. For each different number of hidden neurons, different values of d are examined, and the network is trained on the training set. After training, the performance of the network is evaluated using an evaluation dataset, and the weights of the network are updated. Figure 2 shows the steps of the proposed model.

Because the latitude and longitude of each geographic area are entered as metadata in the model, the weights of the neural network are affected by the coordinates of all geographic areas. This allows us to train just one neural network for all areas in the dataset, instead of training one network for each geographic area. After completing the training and evaluating the model, it is even possible to enter the information of new areas into the network and predict the death or recovery status of those areas without the need for retraining. The final time series prediction model is evaluated on the test set, and the performance of the model is evaluated based on sensitivity, specificity, and accuracy. Finally, considering all combinations of hidden neurons and different delays, the best predictive model is reported.
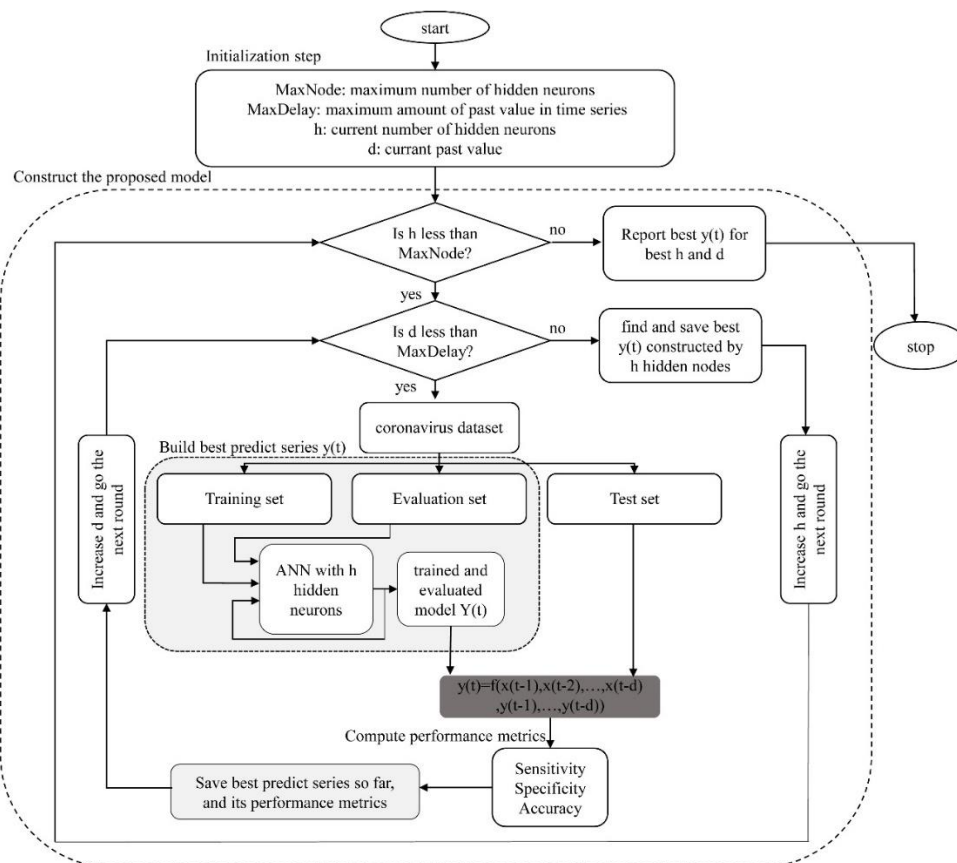

Figure 2. The steps of the proposed method

2.3. Predicting the status of active cases

Active cases of COVID-19 are those who have not yet died or recovered. Therefore, the difference between the total numbers of confirmed cases and the total number of dead or recovered cases in the Beta

dataset are considered active cases. After finding the best models for predicting death and recovery, these models will be used to predict the status of active cases.

To do this, a new test set containing 265 records will be created for each of the 265 different regions in the dataset. For each record, information about code_zone, Latitude, Longitude, and the number of active cases is provided. In addition, we will add date information to this new dataset to predict the situation of each region in the next month. The value of the date variable for all 265 geographic areas is set to one month after the last date in the Alpha dataset. Then, the best proposed models for predicting death and recovery will be trained using the Alpha dataset, which contains the date variable. At this point, the entire dataset is used for training and evaluation. After training, the models are run on the new test dataset. If both death and recovery models are assigned a record to the negative class (class label 0), the record type is set to be active. Otherwise, the type is defined depending on the output of the models, i.e. recovered, death, or both. The model predicts what the status of active cases will be in each geographic area in the next month. Depending on the number of areas in which patients' status is active, dead, or recovered, the number of areas in which confirmed cases of COVID-19 will be expected is calculated according to (2):

$$Confirmed = Active + Death + Recovered \qquad\qquad (2)$$

Finally, to see how well the model predicts areas with COVID-19 death, improved, or confirmed worldwide in the next month, the model's prediction results are compared with the actual data. The criteria are absolute error (AE), mean absolute error (MAE), and absolute percentage error (APE). For each type of case, the percentage of its occurrence is obtained by dividing the number of areas of that type by the sum of the total number of areas of all three types.

The experimentation platform is Intel® Core ™ i7-8550U CPU @ 1.80GHz 1.99 GHz CPU and 12.0 GB of RAM running 64-bits OS of MS Windows 10. The SPSS version 15 was used for descriptive and statistical analysis. The pre-processing and model construction have been implemented in MATLAB.

3. RESULTS
3.1. Descriptive analysis

Data as of March 9, 2020, showed 113,583 confirmed cases, 3,996 deaths, and 62,512 recovered cases. Figure 3 shows the distribution of confirmed, recovered, and death cases by latitude and longitude. The mean distance from the equator in all three groups was significant. The confirmed cases were farther from the equator than other groups, and those recovered were closer to the equator ($P<0.001$). The highest mean distance from prime meridian was related to the recovered group ($P<0.001$). Kruskal Wallis test showed that all groups were significantly different in latitude, and the recovered group was significantly different from other groups in longitude as shown in Table 1.
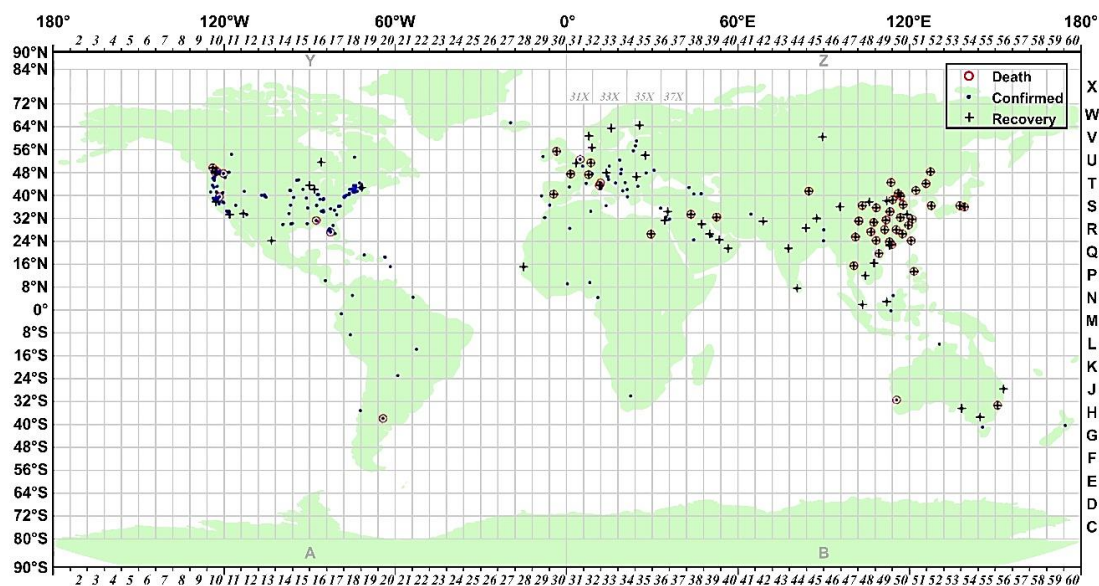

Figure 3. Distribution of recovered, death, and confirmed cases of coronavirus dataset over the world

Table 1. Descriptive analysis of COVID-19 dataset [31] using Kruskal Wallis test

Attribute name	Confirmed Mean±SD	Death Mean±SD	Recovered Mean±SD	Chi-Square	SIG
Latitude	33.30±05.89	32.88±04.51	31.19±03.28	30.633	P<0.001
Longitude	96.24±37.47	95.88±36.05	109.18±16.51	175.627	P<0.001

3.2. Model construction

In this study, scaled conjugate gradient backpropagation (SCG) [43] and Levenberg-Marquardt backpropagation (LM) [44] algorithms were used to train and adjust the weights of the NARX neural network. With a step length of 5, the number of neurons in the hidden layer and the range of delay are varied from 5 to 25 and 5 to 50, respectively. Each model for death and recovery prediction was trained separately on the relevant dataset (interval 5 is obtained by trial and error) [42]. Each dataset is divided into 55% training set, 15% evaluation set, and 30% test set. Since the dataset has a natural temporal order, these values are selected in the dataset with the same order. Thus, the training dataset contains the first 55% of the records, the evaluation set contains 15% of the next records, and the test set contains 30% of the last data records. As a result, data from January 22 to February 27, 2020, were used to build the model (training+evaluation datasets), and data from February 28 to March 9, 2020, were used to test the model. To do this, the "divideblock" function is selected as the neural network

Table 2 and Table 3 show the results of the proposed LM-based model on the test set to predict recovery and death for all combinations of hidden layer neurons and delay factors, respectively. The results are accepted or rejected based on two criteria: the result is discarded: i) if the performance of the model on the test set is better than that on the training set or ii) the model accuracy is less than 0.75%. Accordingly, values written in gray in the tables indicate unacceptable values. Given these two criteria, there is no valid case among the different modes of the proposed SCG-based model for predicting recovered cases. However, there is a single valid mode among the different SCG-based models for predicting deaths. This model is built of 25 nodes in the hidden layer and information from 5 previous records (delay). The accuracy, sensitivity, and specificity were 96.27%, 73.5%, and 98.39%, respectively.

Table 2. The result of the proposed method for prediction of recovered cases using LM

Learning algorithms			Number of past information									
			5	10	15	20	25	30	35	40	45	50
			Levenberg-Marquardt backpropagation number of hidden neurons									
5	Sensitivity	Training	95.54	97.17	92.61	96.96	96.20	95.87	95.54	94.56	96.19	97.17
		Test	96.74	96.38	96.38	96.38	96.38	96.01	95.65	95.65	95.65	95.27
	Specificity	Training	97.68	97.21	97.61	98.84	97.27	98.30	98.57	97.55	97.07	96.39
		Test	92.07	98.24	98.64	97.39	98.48	97.91	84.17	79.41	87.78	64.29
	Accuracy	Training	96.86	97.19	95.68	98.11	96.86	97.36	97.40	96.40	96.73	96.69
		Test	93.33	97.73	98.02	97.11	97.90	97.39	87.37	83.96	90.00	73.03
10	Sensitivity	Training	95.76	95.98	95.22	97.07	96.52	95.54	98.04	95.76	98.04	96.96
		Test	96.74	96.38	96.38	96.74	96.74	96.01	95.65	95.65	94.93	95.64
	Specificity	Training	97.48	97.48	98.77	99.05	98.09	98.84	99.32	95.91	97.07	96.05
		Test	98.92	98.51	98.91	89.03	86.60	92.21	95.52	75.18	72.30	55.14
	Accuracy	Training	96.82	96.90	97.40	98.28	97.49	97.57	98.83	95.85	97.44	96.40
		Test	98.33	97.93	98.22	91.14	89.40	93.27	95.56	80.91	78.67	66.56
15	Sensitivity	Training	95.76	95.76	96.30	96.41	96.52	96.85	96.08	96.95	95.97	95.22
		Test	96.74	96.38	96.74	96.74	96.38	96.01	95.29	94.93	94.93	95.64
	Specificity	Training	99.18	97.41	97.82	98.30	99.25	98.98	97.89	97.07	98.16	99.11
		Test	98.52	98.78	87.74	87.24	99.03	80.25	98.32	95.63	98.58	98.43
	Accuracy	Training	97.86	96.77	97.24	97.57	98.20	98.16	97.19	97.03	97.32	97.61
		Test	98.04	98.13	90.20	89.85	98.30	84.62	97.47	95.43	97.55	97.64
20	Sensitivity	Training	95.65	96.09	96.85	94.67	95.43	97.07	97.82	98.91	97.50	94.57
		Test	96.74	96.38	96.74	96.38	96.38	95.65	95.65	95.29	94.20	95.27
	Specificity	Training	99.32	98.36	97.82	99.18	98.98	98.02	99.25	97.00	99.32	99.39
		Test	93.68	98.38	81.06	99.31	97.93	95.41	82.07	77.43	95.88	90.14
	Accuracy	Training	97.91	97.49	97.44	97.44	97.61	97.65	98.70	97.74	98.62	97.53
		Test	94.51	97.83	85.35	98.51	97.50	95.48	85.86	82.44	95.41	91.59
25	Sensitivity	Training	96.41	95.87	96.20	95.33	96.85	95.98	97.28	96.74	94.99	97.72
		Test	96.74	96.38	96.74	96.38	96.38	96.38	95.29	94.57	94.57	94.55
	Specificity	Training	99.52	98.36	98.98	99.73	99.45	91.34	98.71	98.91	99.18	98.98
		Test	98.39	98.51	89.92	96.43	97.79	80.95	80.53	97.32	83.66	89.29
	Accuracy	Training	98.32	97.40	97.91	98.03	98.45	93.13	98.16	98.07	97.57	98.49
		Test	97.94	97.93	91.78	96.42	97.40	85.23	84.65	96.55	86.73	90.77

Table 3. The result of the proposed method for prediction of death cases using LM

Learning algorithms			Number of past information									
			5	10	15	20	25	30	35	40	45	50
			Levenberg-Marquardt backpropagation number of hidden neurons									
5	Sensitivity	Training	70.18	76.61	88.89	71.93	75.44	65.88	68.82	71.76	75.29	70.79
		Test	75.86	85.06	47.13	59.77	71.26	50.57	64.37	67.82	79.52	54.43
	Specificity	Training	98.83	98.29	97.74	98.60	98.38	98.74	98.47	98.60	98.46	98.14
		Test	98.50	99.35	99.24	98.04	99.23	99.23	99.11	99.44	98.66	97.88
	Accuracy	Training	96.77	96.73	97.11	96.69	96.73	96.40	96.36	96.69	96.77	96.10
		Test	96.57	98.13	94.75	94.73	96.80	94.97	96.06	96.65	97.04	94.36
10	Sensitivity	Training	87.13	75.44	75.44	73.68	72.51	78.24	70.59	82.94	76.44	69.10
		Test	81.61	52.87	42.53	62.07	48.28	65.52	47.13	68.97	54.22	69.62
	Specificity	Training	98.47	98.51	98.83	99.05	99.05	98.87	98.83	98.29	98.51	99.19
		Test	98.93	99.14	98.27	99.02	98.80	97.80	94.02	99.44	99.67	97.21
	Accuracy	Training	97.65	96.86	97.15	97.24	97.15	97.40	96.82	97.19	96.90	96.94
		Test	97.45	95.17	93.47	95.82	94.40	94.97	89.90	96.75	95.82	94.97
15	Sensitivity	Training	71.93	72.51	74.27	74.27	73.68	67.65	91.76	91.18	90.23	75.28
		Test	60.92	64.37	44.83	48.28	52.87	40.23	80.46	85.06	53.01	53.16
	Specificity	Training	98.78	99.05	99.32	98.96	99.01	99.28	98.06	98.96	98.42	99.50
		Test	99.68	99.57	98.81	95.53	99.01	99.01	97.67	96.10	98.66	99.22
	Accuracy	Training	96.86	97.15	97.53	97.19	97.19	97.03	97.61	98.41	97.82	97.70
		Test	96.37	96.55	94.16	91.44	95.00	93.87	96.16	95.13	94.80	95.49
20	Sensitivity	Training	79.53	73.68	70.76	69.59	71.35	65.88	94.12	92.94	83.91	94.38
		Test	80.46	63.22	55.17	65.52	43.68	44.83	62.07	51.72	73.49	75.95
	Specificity	Training	98.29	99.01	98.96	99.28	99.01	99.41	98.92	99.05	98.96	96.24
		Test	98.93	98.71	98.70	97.17	99.67	91.96	97.45	95.43	99.11	96.88
	Accuracy	Training	96.94	97.19	96.94	97.15	97.03	97.03	98.58	98.62	97.86	96.10
		Test	97.35	95.67	94.95	94.43	94.80	87.84	94.34	91.57	96.94	95.18
25	Sensitivity	Training	70.18	73.68	73.10	76.61	78.95	73.53	87.06	89.41	83.91	88.76
		Test	78.16	50.57	72.41	59.77	65.52	58.62	52.87	71.26	45.78	51.90
	Specificity	Training	99.05	99.05	99.19	98.83	99.05	98.65	99.50	99.23	98.92	99.05
		Test	98.61	85.13	98.59	97.71	88.61	99.34	91.14	93.54	98.44	92.30
	Accuracy	Training	96.98	97.24	97.32	97.24	97.61	96.86	98.62	98.53	97.82	98.28
		Test	96.86	82.17	96.34	94.43	86.60	95.78	87.78	91.57	93.98	89.03

To predict the recovered cases, 21 out of 50 possible combinations were accepted among different modes of implementation of the proposed method using the LM learning algorithm. The best combination consists of 15 neurons in the hidden layer and uses the last 25 observations. The accuracy, sensitivity, and specificity of evaluation of this model on the test set were 98.30%, 96.38%, and 99.03%, respectively.

Also, to predict death cases, 27 out of 50 possible combinations were accepted among different modes of implementation of the proposed method using the LM learning algorithm. The best combination consists of 15 neurons in the hidden layer and uses the last 40 observations. Accuracy, sensitivity, and specificity of evaluation of this model in the test set were 95.13%, 85.06%, and 96.10%, respectively.

The figure shows the best period until the neural network reaches the desired level. At this stage, the mean square error for predicting recovered cases using the LM learning method is 0.026. To predict the deaths, this value is 0.048 and 0.037 for LM and SCG learning methods, respectively. Henceforth, the model behavior and error rate remain almost constant and the zigzag mode is not observed in the plot.

3.3. Predicting the status of the active patients

Because the best predictive model of the recovered cases was created using the LM learning algorithm, consisting of 15 hidden layer neurons and 25-time delays, these settings were used to construct a predictive model of the probability of recovery of active cases. The LM-based death prediction model settings consist of 15 hidden layers and 40-time delays and are more sensitive. Thus, this setting was used to construct the predictive model of the probability of death of active cases. The whole Alpha dataset (includes information from January 22 to March 9, 2020) was used to train and evaluate each model with an 85%-15% training-validation partition. The trained model was then run on the new test set containing 265 records. By adding the Date variable to 265 records and setting it to the next month, in this case, April 9, 2020, we can predict the status of active patients in the coming month. The results of the continental implementation are presented in Table 4. For each type of case on each continent, the number of actual and predicted COVID-19 cases and the absolute errors of these two values are reported. The last column shows the MAE of the proposed model in predicting the type of COVID-19 cases on each continent. Accordingly, the lowest MAE of the model is obtained for South America and Australia with 3 and 3.33, respectively. The highest MAE is for North America with 74.67. In predicting the number of areas that will have new confirmed cases of COVID-19on April 9, 2020, the lowest MAE is for Europe with 1 area and the worst for North America with 126 regions. In predicting the number of regions leading to death, the best predictions are for Africa and Asia

with a MAE of 2 and the worst case for Europe with a MAE of 28. To predict the areas where people will recover, the best prediction is for Australia with one area difference and the worst for North America with 79 areas. Globally, the best outcome among the predictions of areas with confirmed, recovery, and death is related to the prediction of death with 13 cases.

Table 4. Prediction of the state of active COVID-19 patients by the proposed model

Row	Continent	Confirmed			Recovered			Death			MAE
		Predicted	Actual	AE	Predicted	Actual	AE	Predicted	Actual	AE	
1	Africa	13.00	26.00	13	6.00	15.00	9	5.00	7.00	2	8.00
2	Asia	86.00	41.00	45	54.00	40.00	14	17.00	19.00	2	20.33
3	Australian	13.00	8.00	5	8.00	7.00	1	5.00	1.00	4	3.33
4	Europe	56.00	55.00	1	34.00	39.00	5	11.00	39.00	28	11.33
5	North America	150.00	24.00	126	87.00	8.00	79	33.00	14.00	19	74.67
6	South America	11.00	11.00	0	5.00	10.00	5	3.00	7.00	4	3.00
	World	329	165	164	194	119	75	74	87	13	84.00

Figure 4 shows the actual and predicted percentage of the areas that will encounter the occurrence of each of the three types of death, recovered, and confirmed cases on April 9, 2020. This figure shows the percentage of occurrence of each type on each continent. In predicting areas with confirmed cases, the best accuracy was obtained in Africa and Australia with 100%. In predicting areas with death cases, the best accuracy was obtained in Africa with 93.75%, and the best recovery prediction was achieved for Europe with 95.66%. In addition, Figure 5 shows the distribution of different predicted types of COVID-19 cases in the next month across the world.
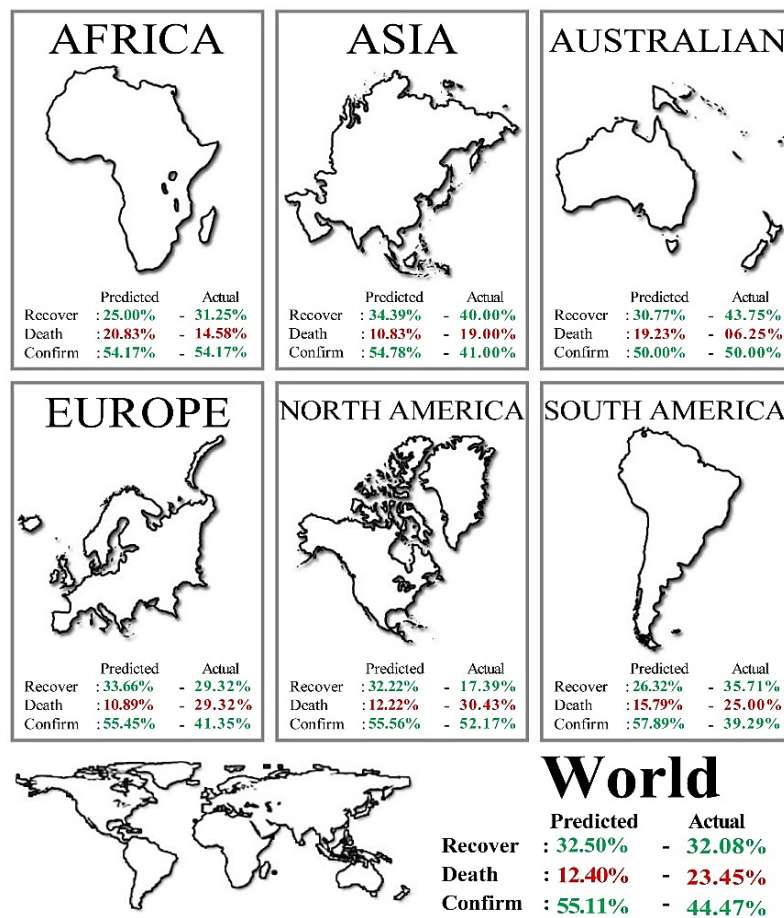
Figure 4. The actual and predicted percentage of occurrences of confirmed, recovered, and death cases around the world on April 9, 2020
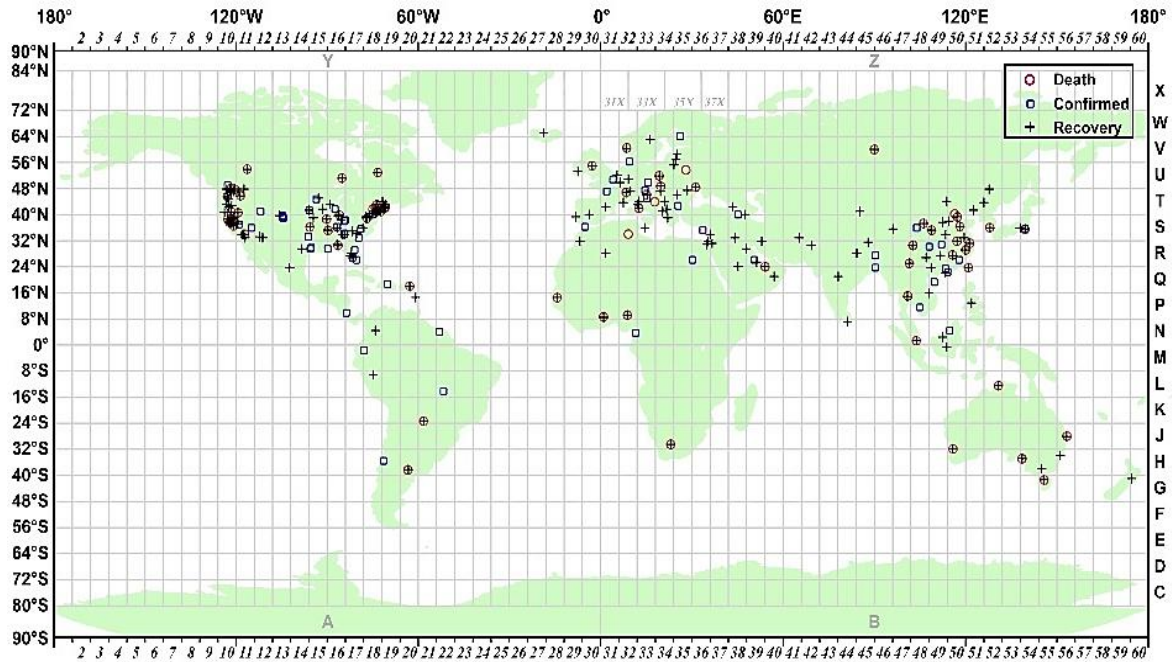
Figure 5. Distribution of death, recovery, and active COVID-19 infected patients across the world

4.   DISCUSSION

In this study, the COVID-19 dataset derived from the JHU CSSE datasets of COVID-19, from January 22 to March 9, 2020, was used. This dataset contains 3,436 records from 265 different geographic regions. Each record stores the information about code_zone, latitude, longitude, cases, and type of COVID-19 cases. The significance of latitude and longitude with type (confirmed-death-recovered) variable with 99% confidence interval were evaluated by the Kruskal Wallis test. There was a significant relationship between latitude and longitude and the status of covid-19 patients (P<0.001). Therefore, we entered these features into the prediction models.

Neural networks are known as an accurate and powerful tool for solving complex and nonlinear problems and predictive models [45]. Thus, a time-series prediction model was designed using a neural network algorithm. This is an efficient method for predicting the status of confirmed, death, and recovered cases of COVID-19.

After training the models on data from January 22 to February 27, 2020, the models were tested on data from February 28 to March 9, 2020. The sensitivity, specificity, and accuracy of the model on the test set were 96.38%, 99.03%, and 98.3%, respectively, to predict the recovery. The sensitivity was 85.06%, specificity was 96.1%, and accuracy was 95.13% for mortality prediction. These results indicate that this model is very suitable for predicting COVID-19 status.

The strength of the proposed model is that it does not require retraining to predict mortality or recovery of affected areas in the short term. Since the training of neural network models is time-consuming, eliminating the training time allows us to predict the situation of the next day of the regions with high speed and accuracy by providing daily information. In this study, we evaluated the proposed model as a pilot to predict the state of the regions for 12 consecutive days. The accuracy was higher than 95% in both models of death and recovery prediction. Daily prediction of death and recovery status of affected areas is important in the management of medical staff and timely measures.

In the second part of the prediction, information about one month after the selected period, i.e. April 9, 2020, for each continent, is expressed in terms of the number and percentage of infected areas. The results of the proposed model were compared with the actual data in the updated JHU CSSE datasets of COVID-19. In Africa, for all three types of recovered, deaths, and confirmed cases, the proposed method had maximum and minimum prediction error of 13 and 2 geographical areas, respectively. In Asia, for all three types of recovered, deaths, and confirmed cases, the proposed method had maximum and minimum prediction errors of 45 and 2, respectively. In Australia, for all three types of recovered, deaths, and confirmed cases, the proposed method has maximum and minimum prediction errors of 5 and 1, respectively. In Europe, for all three types of recovered, deaths, and confirmed cases, the proposed method had maximum and minimum prediction errors of 28 and 1, respectively. In North America, for all three types of recovered, deaths, and

confirmed cases, the proposed method had maximum and minimum prediction errors of 126 and 19, respectively. In South America, for all three types of recovered, deaths, and confirmed cases, the proposed method had maximum and minimum prediction errors of 5 and zero, respectively. The mean absolute percentage error of all three types is 4.17% in Africa, 9.19 % in Asia, 8.65% in Australia, 12.29% in Europe, 12.14% in North America, and 12.41% in South America. According to Figure 4, after aggregating the results of all continents, for all three types of confirmed, death, and recovered in the world, the proposed model had a maximum prediction error of 164 geographical areas and a minimum of 13 geographical areas. The maximum absolute percentage error for the world is 11.05%.

Another limitation of this study is the use of data from all countries involved in COVID-19, while each country has its own protocol for testing and identifying patients. However, in general, this is the only global dataset for COVID-19 that has been used in other studies [17], [29], [34], [46]–[48]. Also, in the proposed model, the past information of each country has been used to predict the COVID-19 status of that country, and this reduces the mentioned limitation. It is suggested that if the area is specifically defined, variables such as temperature and humidity, weather conditions, and population density of the area should be used in creating the model.

5. A DATA SHARING STATEMENT

Dataset is public and available on Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE). Novel coronavirus (COVID-19) cases [49].

6. CONCLUSION

COVID-19 has been dramatically spreading around the world during an epidemic, the speed of information gathering and information dissemination is crucial to the containment of the threat. The mortality or recovery of patients in a region/country or continent should be predicted because it helps with timely action and medical decisions. In addition, resource management will be more effective if that the trend of mortality or recovery in an area is predicted in the next week or two. Since epidemiological models such as SIR are not able to accurately predict mortality and recovery of COVID-19 cases, we presented a more complex model based on machine learning methods using the COVID-19 Cases dataset provided by Johns Hopkins University. Therefore, due to the time-series nature of COVID-19 data, a neural network-based time-series method is presented to predict the mortality or recovery status of COVID-19 cases in different geographical areas around the world. In addition, we used the proposed model to estimate the status of COVID-19 active cases in the next month.

Although almost the same preventive measures have been taken in almost all areas infected with COVID-19, in some areas the death or recovery of patients is very different. The prediction of recovery or death of patients affects the decision of the authorities given the burden of the disease on people's anxiety and the economy. This study almost comprehensively analyzes COVID-19in terms of: i) predicting the final status of patients (recovery or death) and ii) the effect of the longitude and latitude of the infected areas on the final status of active cases separately for each region and continent. Our results indicate that the model might be generalized in similar pandemics considering that the proposed model does not depend on the clinical characteristics of the disease and predicts the condition of patients based on the disease pattern in a region and other infected areas around the world.

The importance of designing such models can be considered for several reasons: i) at the beginning of epidemics, the outbreak does not occur simultaneously all over the world so the information about the outbreak is available for all regions. Therefore, if there are models that can predict the conditions of an area in the coming weeks based on the disease behavior pattern in neighboring areas, preventive measures can be taken to prevent the spread of the disease more effectively and ii) due to the unknown nature of the disease and the factors affecting it, the existence of models that can predict the final status of patients based on non-clinical information can help administrators to manage the allocation of resources and medical staff in affected areas, and thus increasing the quality of services and reducing the community anxiety. The high accuracy of the proposed model in predicting the recovery and death in the next 2-week indicates that the proposed model can be used in the event of other pandemics in the future and can guide planning and resource allocation for prevention, treatment, and palliative care.

REFERENCES

[1] J. Nkengasong, "China's response to a novel coronavirus stands in stark contrast to the 2002 SARS outbreak response," *Nature Medicine*, vol. 26, no. 3, pp. 310–311, Jan. 2020, doi: 10.1038/s41591-020-0771-1.

[2] B. Saman, M. M. A. Eid, and M. M. Eid, "Recently employed engineering techniques to reduce the spread of COVID-19 (corona virus disease 2019): a review study," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 1, pp. 277–286, Apr. 2021, doi: 10.11591/ijeecs.v22.i1.pp277-286.

[3] K. Roosa *et al.*, "Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020," *Infectious Disease Modelling*, vol. 5, pp. 256–263, 2020, doi: 10.1016/j.idm.2020.02.002.

[4] Eurosurveillance Editorial Team, "Note from the editors: world Health Organization declares novel coronavirus (2019-nCoV) sixth public health emergency of international concern," *Eurosurveillance*, vol. 25, no. 5, Feb. 2020, doi: 10.2807/1560-7917.ES.2020.25.5.200131e.

[5] WHO, "Weekly epidemiological update on COVID-19," *World Health Organization*, 2021. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---20-july-2021 (accessed Mar. 01, 2021).

[6] W. Ji, W. Wang, X. Zhao, J. Zai, and X. Li, "Cross-species transmission of the newly identified coronavirus 2019-nCoV," *Journal of Medical Virology*, vol. 92, no. 4, pp. 433–440, Apr. 2020, doi: 10.1002/jmv.25682.

[7] D. Paraskevis, E. G. Kostaki, G. Magiorkinis, G. Panayiotakopoulos, G. Sourvinos, and S. Tsiodras, "Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event," *Infection, Genetics and Evolution*, vol. 79, Apr. 2020, doi: 10.1016/j.meegid.2020.104212.

[8] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *The Lancet*, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.

[9] J. Y. Kim *et al.*, "The first case of 2019 novel coronavirus pneumonia imported into korea from wuhan, china: Implication for infection prevention and control measures," *Journal of Korean Medical Science*, vol. 35, no. 5, 2020, doi: 10.3346/jkms.2020.35.e61.

[10] S. Bernard Stoecklin *et al.*, "First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020," *Eurosurveillance*, vol. 25, no. 6, Feb. 2020, doi: 10.2807/1560-7917.ES.2020.25.6.2000094.

[11] M. Giovanetti, D. Benvenuto, S. Angeletti, and M. Ciccozzi, "The first two cases of 2019-nCoV in Italy: Where they come from?," *Journal of Medical Virology*, vol. 92, no. 5, pp. 518–521, Feb. 2020, doi: 10.1002/jmv.25699.

[12] V. M. Corman *et al.*, "Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR," *Eurosurveillance*, vol. 25, no. 3, Jan. 2020, doi: 10.2807/1560-7917.ES.2020.25.3.2000045.

[13] N. Zhang *et al.*, "Recent advances in the detection of respiratory virus infection in humans," *Journal of Medical Virology*, vol. 92, no. 4, pp. 408–417, Apr. 2020, doi: 10.1002/jmv.25674.

[14] Y. Yang, D. Zhang, C. Zhou, H. Huang, and R. Wang, "Value of lung ultrasound for the diagnosis of COVID-19 pneumonia: a protocol for a systematic review and meta-analysis," *BMJ Open*, vol. 10, no. 8, Aug. 2020, doi: 10.1136/bmjopen-2020-039180.

[15] C. K. Wong *et al.*, "Artificial intelligence mobile health platform for early detection of COVID-19 in quarantine subjects using a wearable biosensor: protocol for a randomised controlled trial," *BMJ Open*, vol. 10, no. 7, Jul. 2020, doi: 10.1136/bmjopen-2020-038555.

[16] H. B. Mehta, S. Ehrhardt, T. J. Moore, J. B. Segal, and G. C. Alexander, "Characteristics of registered clinical trials assessing treatments for COVID-19: a cross-sectional analysis," *BMJ Open*, vol. 10, no. 6, Jun. 2020, doi: 10.1136/bmjopen-2020-039978.

[17] H. Nishiura *et al.*, "The extent of transmission of novel coronavirus in Wuhan, China, 2020," *Journal of Clinical Medicine*, vol. 9, no. 2, Jan. 2020, doi: 10.3390/jcm9020330.

[18] J. B. Long and J. M. Ehrenfeld, "The role of augmented intelligence (AI) in detecting and preventing the spread of novel coronavirus," *Journal of Medical Systems*, vol. 44, no. 3, Mar. 2020, doi: 10.1007/s10916-020-1536-6.

[19] B. McCall, "COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread," *The Lancet Digital Health*, vol. 2, no. 4, pp. 166–167, Apr. 2020, doi: 10.1016/S2589-7500(20)30054-6.

[20] C.-S. Yu *et al.*, "A COVID-19 pandemic artificial intelligence–based system with deep learning forecasting and automatic statistical data acquisition: development and implementation study," *Journal of Medical Internet Research*, vol. 23, no. 5, May 2021, doi: 10.2196/27806.

[21] H. B. Harchegani, A. Farahi, H. M. Shirazi, A. Golabpour, and P. Almasinejad, "Using genetic algorithm for prediction of information for cache operation in? mobile database," in *2009 Second International Workshop on Knowledge Discovery and Data Mining*, Jan. 2009, pp. 148–151, doi: 10.1109/WKDD.2009.201.

[22] R. Torshizi *et al.*, "Altered expression of cell cycle regulators in adult T-Cell leukemia/lymphoma patients," *Reports of biochemistry and molecular biology*, vol. 6, no. 1, pp. 88–94, Oct. 2017.

[23] A. Z. M. Kootiani, M. Doostari, A. Golabpour, and M. Broujerdian, "Differential power analysis in the smart card by data simulation," in *Proceedings - 2008 International Conference on MultiMedia and Information Technology, MMIT 2008*, Dec. 2008, pp. 817–821, doi: 10.1109/MMIT.2008.192.

[24] N. E. Budiyanta, L. Wijayanti, W. Widjaja Basuki, H. Tanudjaja, and V. B. Kartadinata, "The development of healthcare mobile robot for helping medical personnel in dealing with COVID-19 patients," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 3, Jun. 2021, doi: 10.11591/ijeecs.v22.i3.pp1379-1388.

[25] N. A. Hidayat, P. Megantoro, A. Yurianta, A. Sofiah, S. A. Aldhama, and Y. A. Effendi, "The application of instrumentation system on a contactless robotic triage assistant to detect early transmission on a COVID-19 suspect," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 3, pp. 1334–1344, Jun. 2021, doi: 10.11591/ijeecs.v22.i3.pp1334-1344.

[26] A. Singh, V. Jindal, R. Sandhu, and V. Chang, "A scalable framework for smart COVID surveillance in the workplace using Deep Neural Networks and cloud computing," *Expert Systems*, May 2021, doi: 10.1111/exsy.12704.

[27] I. Ahmed, M. Ahmad, J. J. P. C. Rodrigues, G. Jeon, and S. Din, "A deep learning-based social distance monitoring framework for COVID-19," *Sustainable Cities and Society*, vol. 65, Feb. 2021, doi: 10.1016/j.scs.2020.102571.

[28] H. M. Haglan, A. S. Mahmoud, M. H. Al-Jumaili, and A. J. Aljaaf, "New ideas and framework for combating COVID-19 pandemic using IoT technologies," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 3, pp. 1565–1572, Jun. 2021, doi: 10.11591/ijeecs.v22.i3.pp1565-1572.

[29] F. Ahouz and A. Golabpour, "Predicting the incidence of COVID-19 using data mining," *BMC Public Health*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12889-021-11058-3.

[30] D. A. Gomez-Cravioto, R. E. Diaz-Ramos, F. J. Cantu-Ortiz, and H. G. Ceballos, "Data analysis and forecasting of the COVID-19 spread: a comparison of recurrent neural networks and time series models," *Cognitive Computation*, Jun. 2021, doi: 10.1007/s12559-021-09885-y.

[31] C. Kerdvibulvech and L. (Luke) Chen, "The power of augmented reality and artificial intelligence during the covid-19 outbreak," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12424, Springer International Publishing, 2020, pp. 467–476.

[32] A. W. Reza, M. M. Hasan, N. Nowrin, and M. M. Ahmed Shibly, "Pre-trained deep learning models in automatic COVID-19

diagnosis," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 3, pp. 1540–1547, Jun. 2021, doi: 10.11591/ijeecs.v22.i3.pp1540-1547.

[33] M. Masum, H. Shahriar, H. M. Haddad, and M. S. Alam, "r-LSTM: time series forecasting for COVID-19 confirmed cases with LSTMbased framework," in *2020 IEEE International Conference on Big Data (Big Data)*, Dec. 2020, pp. 1374–1379, doi: 10.1109/BigData50022.2020.9378276.

[34] K. E. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji, and T. M. Brenza, "Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and long short-term memory (LSTM) cells," *Chaos, Solitons and Fractals*, vol. 146, May 2021, doi: 10.1016/j.chaos.2021.110861.

[35] S. J. Fong, G. Li, N. Dey, R. Gonzalez-Crespo, and E. Herrera-Viedma, "Finding an accurate early forecasting model from small dataset: a Case of 2019-nCoV novel coronavirus outbreak," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 1, 2020, doi: 10.9781/ijimai.2020.02.002.

[36] A. B. H. Fairoza *et al.*, "Coronatracker: world-wide COVID-19 outbreak data analysis and prediction," *Bulletin of the World Health Organization*, vol. 98, no. 3, Mar. 2020, doi: 10.2471/blt.20.255695.

[37] P. Wang, X. Zheng, G. Ai, D. Liu, and B. Zhu, "Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran," *Chaos, Solitons & Fractals*, vol. 140, Nov. 2020, doi: 10.1016/j.chaos.2020.110214.

[38] H. M. Zawbaa *et al.*, "A study of the possible factors affecting COVID-19 spread, severity and mortality and the effect of social distancing on these factors: machine learning forecasting model," *International Journal of Clinical Practice*, vol. 75, no. 6, Jun. 2021, doi: 10.1111/ijcp.14116.

[39] H. Pham, "Predictive modeling on the number of covid-19 death toll in the United States considering the effects of coronavirus-related changes and covid-19 recovered cases," *International Journal of Mathematical, Engineering and Management Sciences*, vol. 5, no. 6, pp. 1140–1155, Dec. 2020, doi: 10.33889/IJMEMS.2020.5.6.087.

[40] "Novel coronavirus (COVID-19) cases data," *Johns Hopkins University Center for Systems Science and Engineering*, 2020.

[41] R. Krispin, "Coronavirus," 2020. https://github.com/RamiKrispin/coronavirus (accessed Mar. 01, 2020).

[42] Z. Boussaada, O. Curea, A. Remaci, H. Camblong, and N. Mrabet Bellaaj, "A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of the daily direct solar radiation," *Energies*, vol. 11, no. 3, Mar. 2018, doi: 10.3390/en11030620.

[43] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, no. 4, pp. 525–533, Jan. 1993, doi: 10.1016/S0893-6080(05)80056-5.

[44] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994, doi: 10.1109/72.329697.

[45] C. L. Giles, S. Lawrence, and A. C. Tsoi, "Noisy time series prediction using recurrent neural networks and grammatical inference," *Machine Learning*, vol. 44, no. 1–2, pp. 161–183, 2001, doi: 10.1023/A:1010884214864.

[46] F. Ahouz and A. Golabpour, "Predicting the COVID-19 prevalence rate using data mining," *BMC Public Health*, Apr. 2020, doi: 10.21203/rs.3.rs-21247/v1.

[47] M. Adebiyi and O. O. Olugbara, "Binding site identification of COVID-19 main protease 3D structure by homology modeling," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 21, no. 3, pp. 1713–1721, Mar. 2021, doi: 10.11591/ijeecs.v21.i3.pp1713-1721.

[48] K. E. ArunKumar, D. V. Kalaga, C. M. Sai Kumar, G. Chilkoor, M. Kawaji, and T. M. Brenza, "Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: auto-regressive integrated moving average (ARIMA) and seasonal auto-regressive integrated moving averag," *Applied Soft Computing*, vol. 103, May 2021, doi: 10.1016/j.asoc.2021.107161.

[49] "Coronavirus (COVID-19) cases and deaths," *World Health Organization*, 2020. https://data.humdata.org/dataset/coronavirus-covid-19-cases-and-deaths (accessed Mar. 01, 2020).

BIOGRAPHIES OF AUTHORS

Peyman Almasinejad received his Bachelor of Software Engineering from the Islamic Azad University, Iran in 2006 and Master of Software Engineering from Payame Noor University (PNU), Tehran, Iran in 2009. Then, He worked in a software company. Currently, he is working as an instructor at Department of Computer Engineering and Information Technology, Payame Noor University (PNU), Iran. He can be contacted at email: p.almasinejad@pnu.ac.ir.

Amin Golabpour I received my PhD in medical informatics from Ferdowsi University of Mashhad, Mashhad, Iran, and now I am a director of Health Informatics Technology group of School of Allied Medical Sciences department at Shahroud University of Medical Sciences, Shahrour, Iran. I received two master's degree in software engineering and artificial intelligence. He can be contacted at email: a.golabpour@shmu.ac.ir.

Fatemeh Ahouz ⓘ 🔍 SC Ⓟ, I received my M.Sc. in artificial intelligence (AI) from Shahid Chamran University of Ahvaz, Iran. I've been working as a faculty lecturer at Behbahan Khatam Alanbia University of Technology, Behbahan, Iran since 2012. In addition to teaching fundamental courses of computer engineering and soft computing, I've been conducting some research in deep/machine learning and optimization methods especially in extracting rules from data. I am interested in practical aspects of AI in engineering and medical applications. Currently my research works are oriented toward medical informatics. She can be contacted at email: ahouz@bkatu.ac.ir.

Mohammad-Reza Mollakhalili-Meybodi ⓘ 🔍 SC Ⓟ received his M.Sc. degree in Computer Engineering from Amirkabir University of Technology, Tehran, Iran, in 2004, and the B.Sc. degree in Computer Engineering from Shahid Beheshti University in 2001. He joined the faculty of Computer engineering Department at Islamic Azad University (Maybod Branch) in 2004. He obtained his Ph.D. degree in Computer Engineering in 2014. His research interests include Communication Networks, Learning Systems, Algorithms, Graph Theory, Stochastic Optimization, Software Defined Networks and Soft Computing. He can be contacted at email: Mollakhalili@maybodiau.ac.ir.

Kamal Mirzaie ⓘ 🔍 SC Ⓟ obtained his B.Sc. in Computer Engineering from Iran University of Science and Technology (IUST) in 2003 and his M.Sc. in Computer Engineering from University of Isfahan (UI) in 2005. He obtained his Ph.D. In 2011, he is currently working as an Assistant Professor at the Department of Computer Engineering, Maybod Branch, Islamic Azad University, and Maybod, Iran. His research interests include cognitive science, soft computing, medical data mining, parallel processing, image processing, and pattern recognition. He can be contacted at email: k.mirzaie@maybodiau.ac.ir.

Ahmad Khosravi ⓘ 🔍 SC Ⓟ his M.S. and Ph.D. of Epidemiology, Center for Health Related Social and Behavioral Sciences, he is Associate Professor at Shahroud University of Medical Sciences (Iran). He can be contacted at email: khosravi2000us@yahoo.com.

Marzieh Rohani-Rasaf ⓘ 🔍 SC Ⓟ her M.S. in 2012, and Ph.D. in Epidemiology in 2019 at Shahid Beheshti University of Medical Sciences (Iran). Currently, she is Assistant Professor and Head of Student Research Committee at Shahroud University of Medical Sciences (Iran). She can be contacted at email: rohani@shmu.ac.ir.

Azadeh Bastani ⓘ 🔍 SC Ⓟ received the B.Sc. and M.Sc. degrees in Computer Engineering from Isfahan University, Isfahan, Iran, and Iran University of Science and Technology, Tehran, Iran, in 2000 and 2003, respectively. She has been IT expert in Kish University from 2004 to 2008. Since 2008, she has been with the Behbahan Khatam Alanbia University of Technology, Behbahan, Iran, where she is currently Instructor of the Computer Engineering Department. Her current research interests include digital image processing, machine learning, and artificial intelligence systems. She can be contacted at email: bastani@bkatu.ac.ir.