ORIGINAL PAPER

COVID-19 prediction analysis using artificial intelligence procedures and GIS spatial analyst: a case study for Iraq

Bashar Moneer Yahya[1] · Farah Samier Yahya[2] · Rayan Ghazi Thannoun[1]

Abstract

The prediction of diseases caused by viral infections is a complex medical task where many real data that consists of different variables must be employed. As known, COVID-19 is the most dangerous disease worldwide; nowhere, an effective drug has been found yet. To limit its spread, it is essential to find a rational method that shows the spread of this virus by relying on many infected people's data. A model consisting of three artificial neural networks' (ANN) functions was developed to predict COVID-19 separation in Iraq based on real infection data supplied by the public health department at the Iraqi Ministry of Health. The performance efficiency of this model was evaluated, where its performance efficiency reached 81.6% when employed four statistical error criteria as mean absolute percentage error (MAPE), root mean square error (RMSE), coefficient of determination (R^2), and Nash-Sutcliffe coefficient (NC). The severity of the virus's spread across Iraq was assessed in a short term (in the next 6 months), where the results show that the spread severity will intensify in this short term by 17.1%, and the average death cases will increase by 8.3%. These results clarified by creating spatial distribution maps for virus spread are simulated by employing a Geographic Information System (GIS) environment to be used as a useful database for developing plans for combating viruses in Iraq.

Keywords COVID-19 · Prediction · Artificial neural networks · Geographic Information System, Geospatial analysis

Introduction

Coronavirus disease 2019 (COVID-19) is an infectious disease with predominantly respiratory symptoms triggered by a severe acute respiratory syndrome. Coronavirus 2 (SARS-CoV-2) that originated in Wuhan (Province of Hubei, China) in December 2019 (Lai et al. 2020; Lauer et al. 2020) and has spread globally, except for Antarctica recorded in all countries. World Health Organization (WHO) announced COVID-19 a pandemic on March 11, 2020. After the first documented cases from Wuhan, China, after 2019

(Sohrabi et al. 2020) to August 2020 (the time of writing the paper), the number of globally identified cases of COVID-19 is more than 20 million cases and more than 700,000 cumulative deaths (Url-1). Symptoms of COVID-19 range from fever, fatigue, upper respiratory symptoms like cough, shortness of breath, and loss of smell and taste in the majority of cases that may progress to more severe symptoms like pneumonia, acute respiratory distress syndrome (ARDS), and multiorgan failure in some cases (Zu et al. 2020). SARS-CoV-2 is transmitted primarily by direct person to person contact through respiratory droplets. The dangers of the spread of the COVID-19 pandemic at the local and global levels have become a concern that haunts all humanity. To face this pandemic, Artificial Intelligence Technology (AIT) and Geographic Information Systems (GIS) might be used, where these technologies have proven their efficiency in previous studies in cases related to the health field.

Artificial Intelligence Technology (AIT) is a branch of computer science that employs pre-designed computer algorithms to recognize patterns, analyzes the various types of data, and it can also estimate the different results depending on available input data (Liao et al. 2015; He et al. 2019; Wang

✉ Bashar Moneer Yahya
   bashar1974@uomosul.edu.iq

   Farah Samier Yahya
   fsy@uomousul.edu.iq

   Rayan Ghazi Thannoun
   rayan.ghazi@uomosul.edu.iq

[1] Remote Sensing Center, Mosul University, Mosul, Iraq

[2] College of Medicine, Mosul University, Mosul, Iraq

et al. 2019). Based on different artificial intelligent neural networks' (ANNs) approaches, many researchers have been investigating medical data to predict several scenarios related to infectious diseases like predicting the rapid spread and contaminated area (Steiner et al. 2020; Naudé 2020). In the field of combating clinical diseases or highly infectious diseases such as COVID-19 that caused by small organisms such as bacteria or viruses, and due to the accessibility of relevant clinical data, AIT has mostly played a vital role in medical decision-making (Peiffer-Smadja et al. 2019; Fitzpatrick et al. 2020; Stebbing et al. 2020; Yang et al. 2020). Therefore, to address this epidemic problem early, suitable applications of artificial intelligent forecasting procedures were established as crucial plans to facing this virus.

Developing AIT models might act as a crucial aid for detecting the course of the disease's spread. By this technology, an appropriate strategic decision in real-time can be made to identify risks for combating COVID-19 that can be defined as a natural disaster which caused great harm to humanity. The Geographic Information System (GIS) environment is a powerful tool that plays a vital role in analysis, assessment, as well as supplying a quick mitigation current and plans in disaster management and high-risk environmental phenomena like a flood, desertification, or the spread of infectious diseases, etc. GIS techniques represented by spatial analysis and geographic mapping may be critical methods for any kinds of scientific application like in public health, for instance, in disease detection, mapping, and diagnosis (Kandwal et al. 2009; Lyseen et al. 2014). For example, mapping with GIS may further chart the regional spread of disease, possible risk factors, and effective diagnostic and preventive methods. Concerning the spatial analysis of these data, disease risks, outbreak levels over time and space, and infection hotspots may be calculated (Alamo et al. 2020; Murad and Khashoggi 2020).

Since there are no specific treatments or vaccine yet, the preventive measures are the crucial part of dealing with the COVID-19 pandemic aiming to flatten the curve of peak cases; this will reduce the burden on health services from being overwhelmed and direct medical efforts to deal with critical cases until the availability of vaccine or specific treatment. The preventive measures include social distancing, washing hands, wearing masks, quarantine, and stay at home orders. In this study, three Artificial Intelligence Neural Networks (AINN) functions named as Radial Basis-Function (RBF), Fuzzy Cluster-Means (FCM), and Non-linear Autoregressive-Network with Exogenous Inputs (NARX) have been constructed together in one model to foretell the rapid spread rate of COVID-19 in Iraq. The proposed model was also integrated with the Geographic Information System (GIS) environment to find the spatial distribution of this disease in the whole of Iraq and detect which provinces have significantly affected by COVID-19.

Material and methods

Study area and data availability

Iraq is one of the Middle East countries. Six countries, Iran, Kuwait, Saudi Arabia, Jordan, Syria, and Turkey, have bordered with Iraq and a small stretch of the Arab Gulf. Iraq has eighteen governorates, and Baghdad governorate is Iraq's capital city, as seen in Fig. 1. Iraq presently consists of 18 governorates; the population of Iraq is proportional to 0.52% of the world's total population, where the current number of the Iraqi population is about 40,278,833 people based on the latest update of the United Nations database.

The timeline for the spread of the COVID-19 pandemic in Iraq began on February 24, 2020, when the Iraqi Ministry of Health announced the first confirmed case of COVID-19 infection in Najaf Province southern of Iraq. The patient was placed in quarantine at Al-Hakim Hospital, and nine people who touched the patient were quarantined in the same hospital. Then the chronology accelerated its spread in all Iraqi governorates, especially in the capital Baghdad, Najaf, and Kirkuk. This study's used data was collected from the Iraqi Ministry of Health (the daily epidemiological situation of recorded infections of the emerging coronavirus in Iraq). The World Health Organization (WHO) (Url-1) and the world meter website (Url-2) which provide users with useful charts and figures regarding the COVID spread all over the world. A total of 38,000 patients were diagnosed with COVID-19 in Iraq diagnosed from February 25, 2020, to July 15, 2020. In this study, the included dataset was divided into three parts according to the available data, as illustrated in Fig. 2, namely the total cases, the total healed, and the total deaths cases where these datasets have been prepared for training and validation processes as an essential step for design the proposed model.

Proposed COVID-19 prediction model

The study's main objective was to develop an AIT-based predictive model by choosing suitable ANNs procedures to forecast the prognostic factors for the rapid spread of COVID-19 in Iraq. The developed model's methodology was divided into five stages as preparing data, selecting ANNs' functions, modeling processes, statistical model criteria, and model window interface design. Three ANNs' functions described in Fig. 3 due to their main characteristic and their mathematical representation were selected. MATLAB workspace was used in this study because of its convenient computing environment ability to design structure, statistical analysis, estimate performance accuracy, etc. for variant types of artificial neural networks (MATLAB R 2018a, 2018).
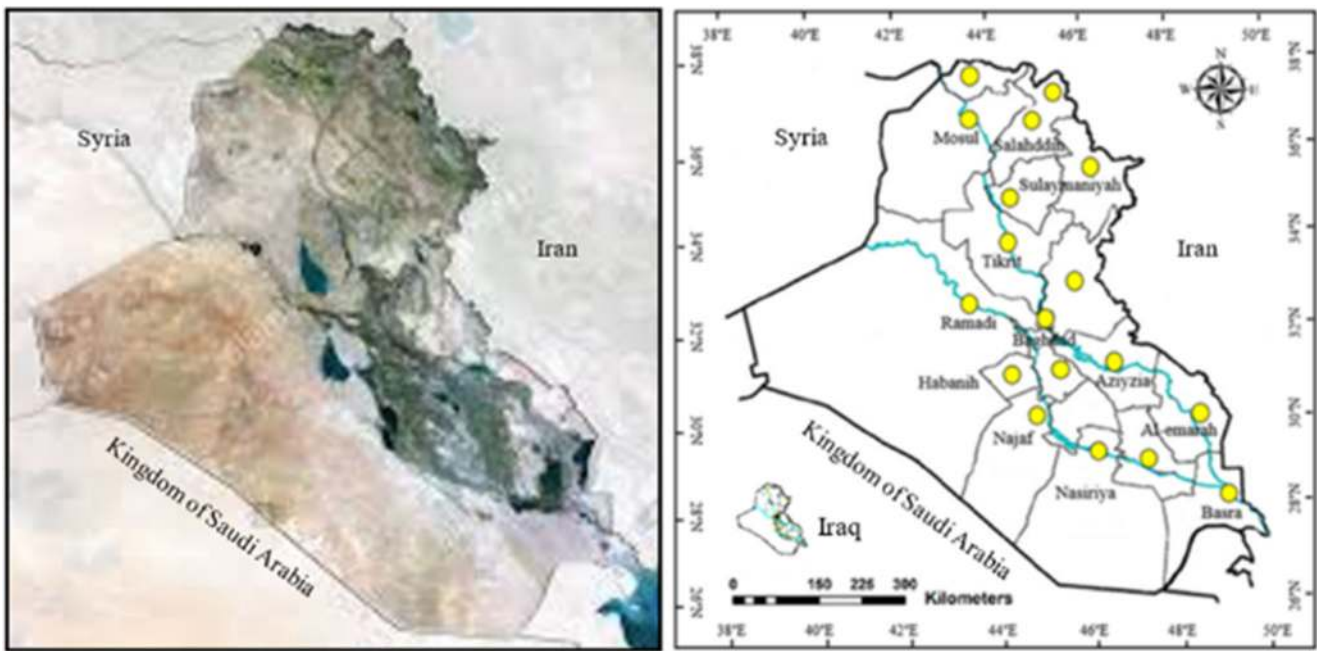
Fig. 1 General view of the study area

Preparing data

Preparing data is a fundamental stage in developing artificial intelligence models (Zhou et al. 2010). Therefore, data organization, imputation reduction, discretization, and normalization were mainly performed to get high-performance models (Zhang et al. 2003). The total used data set has been organized and imported into MATLAB workspace. Out of the 38,000 cases, 70% of them (26,600) were used in the training process that is randomly chosen, 15% (5700) were used to measure the popularization of the designed networks, and the last 15% (5700) of the total dataset used as a new data for invalidation process as data have not been considered with or dealing before.

Selecting AINN functions

RBF, FCM, and NARX, classified as three ANNs' functions, were selected and used to develop a COVID-19 prediction model. These procedures are conformed as powerful ANNs' functions through their mathematical ability for data analyses and predict results when they are involved in developing artificial intelligent models (Mirmozaffari 2019; Yahya and Seker 2019; Ahmad et al. 2020).

First, RBF functions are simply a class of functions. In principle, they could be employed in any nonlinear model and neural networks (single or multilayer) (Uysal 2016). RBF functions can be defined as a multilayer feeding forward neural network used as a robust tool offering many advantages

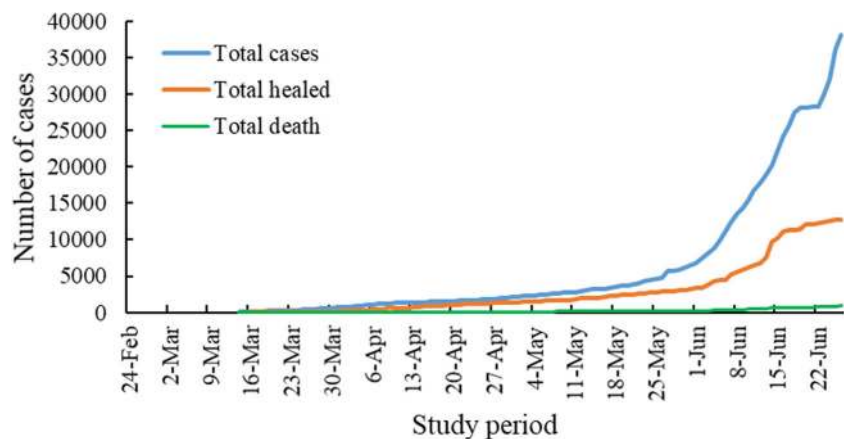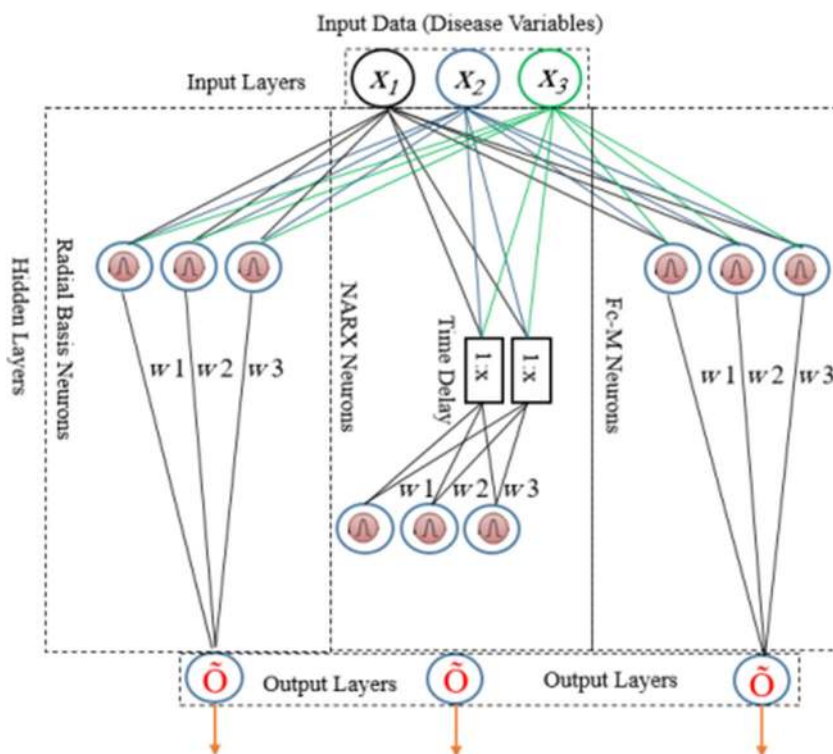Fig. 2 A total of 38,000 patient's dataset of COVID-19 in Iraq

in nonlinear modeling systems. Such functions were commonly used in many research applications such as time series forecasting, pattern recognition, or speech processing, where they can be mathematically described as (Rahmati and Tatar 2019)

$$y_{(t)} = \sum_{j=1}^{n} w_i \cdot \varphi_{j+w_0} \qquad (1)$$

where, $y_{(t)}$ represent RBF output of, w_i the associate weights for every RBF function, Φ_j represents the essential functions set, and w_o is the bias weight of output. A nonlinear function named Gaussian function represents an activation function in the hidden layers used in this study that can be written as Eq. 2 where x represents the training data, x_i is the mean of the data, and σ represents the Gaussian function width.

$$\varphi_i(x) = \exp\left[-\frac{\|x-x_1\|^2}{2\sigma^2}\right] \qquad (2)$$

For this analysis, the suggested RBF neural network was guided for this study to learn about the centers and the width of the Gaussian function for hidden layers, and then the weights were adjusted by employed using the Linear Regression after the parameters for the Gaussian function at the hidden layers have been identified. The k-means clustering method can deduce Gaussian functions for each attribute using cluster testing instances to exercise the Gaussian

function centers. After the Gaussian function parameters at the secret units have been defined, the weights from these units to the output unit are adjusted using Linear Regression. Second, NARX is a suitable estimation algorithm that used for modeling and predicting time series in moderately nonlinear dynamic systems (Lin et al. 1996, Ruiz et al. 2016); the developed NARX in this study can be utilized for a lot of prediction step by providing the prediction process through recursive use of the input dataset. The NARX model can be described by Eq. 3 as follows:

$$y(d+1) = f_{ANN} = \qquad (3)$$
$$\left(y(d), y(d-1), \ldots y(d-n+1), u(d), u(d-1), \ldots u(d-m+1) + \varepsilon(d)\right)$$

where $y(d + 1)$ is the output of model-predict, f_{ANN} is the function (nonlinear) representing the system manner, $y(d)$, $u(d)$, $\varepsilon(d)$ are input, output, as well as vectors of approximation error in the time of case d, n, and m the order of $y(d)$ and $u(d)$, respectively.

Third, image segmentation, medical diagnosis, clustering, and time series forecasting are the most problems that can be solved using the FCM algorithm (Yokoi et al. 2011). Clustering using FCM function is the operation of grouping feature vectors into self-organizing model categories. Fuzzy C-Means Clustering (FCM) is defined as the unsupervised clustering algorithm depending on the fuzzy concept introduced by Bezdek (1981). FCM may also be categorized inside the algorithm of fuzzy clustering. This function is based on splitting a single dataset or time series into several clusters,

each cluster having different attributes. The fuzzy weights were determined in this analysis, depending on the reciprocal distance used to minimize the cumulative weighted mean-square error (Abebe et al. 2000). To accommodate the FCM design procedure, the mathematical steps proposed by (Başkır and Türkşen 2013) can be followed:

Step 1. The FCM algorithm allocates objects to each class by using fuzzy memberships. It splits a set of nth classes into c $(1 < c < n)$ fuzzy clusters with specific centroids. A fuzzy matrix M describes the fuzzy clustering of classes with n rows and c columns in which n is the number of data classes, and c is the number of clusters. The membership function's degree can be indicated by the symbol m_{ij} that initialized randomly, as described mathematically as follows.

$$0 \leq \sum_{i=1}^{n} m_{ij} \leq n \tag{4}$$

$$\sum_{i=1}^{n} m_{ij} = 1, \forall j = 1, \ldots \ldots n \tag{5}$$

Step 2. Define the function of the difference used in FCM is given Eq. 6 to compute dissimilarity.

$$K(M, c1, c2, \ldots \ldots, cp) = \sum_{i=1}^{p} ki = \sum_{i=1}^{p} \sum_{j=1}^{n} m_{ij}^{r} d_{ij}^{2} \tag{6}$$

where m_{ij} ranged between 0 and 1, c_p is the centroid of cluster i, d_{ij} is the Euclidian distance between ith, centroid (c_p) and jth data point, and $m \in [1, \infty]$ is a weighting exponent.

Step 3. Extract centroids (ci) are based on Eq. 7 that represents a minimum of dissimilarity function.

$$cp = \frac{\sum_{j=0}^{n} m_{ij}^{r} d_{ij}^{2}}{\sum_{i=1}^{n} m_{ij}} \tag{7}$$

Step 4. Use Eq. 7 to measure the dimensionality between the data points and the estimated centroids.
Step 5. Compute or update the new partition matrix u using Eq. 8.

$$m_{ij} = \frac{1}{\sum_{k=0}^{c} \left(\frac{d_{ij}}{d_{ki}} \right)^{\frac{2}{m-1}}} \tag{8}$$

Modeling processes

The procedure of developing the proposed model was divided into three stages (stage for each proposed ANN). First, developing RBF neural network to create an RBF structure, the function (newrb) was used where the hidden nodes had been adding until assembly the specified RMSE goal to approximate the function in the training process. To ensure that this neural network works perfectly, three arguments have been taken. Firstly, set up the training process hindrance through configuration the minimum RMSE error where in this study, the RMSE was determined to 0.000001, as well as, the training process will be terminated when reaching this number. Secondly, RBF spread default set up equal to 1.0. Thirdly, to guarantee that the network works satisfactorily without any complexity, the fair number of hidden nodes was carefully chosen by trial and error strategy. However, the most appropriate number of hidden nodes was found to be equal to 140.

Second, creating a NARX neural network, three parameters are entered as input data for the constructed NARX neural network, with one exogenous input parameter and two feedback delays. The developed NARX is structured as two hidden layers with 28 neurons in the first layer and 10 neurons in the second layer. Using the divide block method, the cumulative data was split into two classes. Seventy percent of the total data are deemed a training dataset, and 30% of the total dataset included in validation and testing processes. The original random weight set as 12, and the functions (narxnet) and (close loop) used to trigger the sigmoid's functions as control. Besides, the Levenberg–Marquardt approach was used as a potential solution to network testing by utilizing the (trainlm) function to achieve optimal performance.

Third, developing FCM for designing FCM neural network; first, actual data will be read and pre-processed to implement and prepare the FCM neural network parameters. Genfis3 function is used to build a fuzzy inference system structure (Astakhova et al. 2015). To generate the (genfis3) function, two input parameters must be taken into account as input dataset matrix, matrix dimensionality, and the radii numbers as a vector. This instantly calculates the expected cluster numbers to be three clusters. In this context, those clusters are input only as a starting point into the FCM algorithm. Creating Membership Functions (MF) is where the sum of all MF will be 1.0. The number of MF rules which were produced was 23. The neural network learned using a hybrid learning algorithm to classify single output membership feature parameters where the fuzzy fuzzification method Sugeno-type inference systems were implemented for completion of the training period. The FCM clustering process is iterative. The process stops when the cumulative amount of iterations obtained is less than the minimum quantity specified for the changed, or the target function update for two consecutive iterations. For this neural

network, the total number of iterations tuned as a standard equal to 500 and stopped when it reached 148 iterations.

The performance accuracy of the developing models is a critical part of designing any AIT model; a significant task for a successful design model is to improve any designed model; training, testing, and cross-validating processes play an essential role for a good design model. Various training algorithms can be used for training the ANN, like Backpropagation, Levenberg–Marquardt, Hebbian-based, Bayesian regularization, or One-step secant, for instance (Bello 1992). The development environment in MATLAB workspace offers a wide range of these algorithms that can be chosen according to the proposed neural network. In this study, three training algorithms have been used for training the proposed ANN, where the Backpropagation algorithm was chosen for RBF, Levenberg–Marquardt algorithm was chosen for NARX while Hebbian-based algorithm was chosen for FCM. The testing process is an art of investigating or evaluating the developed model by checking the developed ANN's performance by estimating the testing error. When the testing error decreases, this means the performance of the developed ANN increase and vice versa. Cross-validating is a process for detect overfitting and evaluating the developed models by good training the proposed ANN involved in models structure and based on the available input dataset and evaluating them on the complementary subset of the data.

Statistical model criteria

The best performance is when the correlation is high, and the errors are as small as possible. To ascertain a good performance accuracy for designed ANN models, several error criteria were chosen to test the testing process (Cadenas et al. 2016; Yalur 2019); where the most famous criteria were used in this study, these criteria are Coefficient of Determination (R^2), Root Mean Square Error (RMSE), Nash-Sutcliffe coefficient (CE), and mean absolute percentage error (MAPE) as represented mathematically in Eqs. 9, 10, and 11.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{actual,i}-X_{predict,i}\right)^2}{n}} \qquad (9)$$

$$CE = 1\sqrt{\frac{\sum_{i=1}^{n}\left(X_{actual,i}-X_{predict}\right)^2}{\sum_{i=1}^{n}\left(X_{actual,i}-\overline{X_{predict}}\right)^2}} \qquad (10)$$

$$MAPE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(actual(i)-predicted(i))/actual(i)} \qquad (11)$$

Analysis using GIS environment

One of GIS's famous abilities is providing a quick, comparative view of the area at high risk or hazards through spatial distribution technologies. In this study, Arc View V. 10.4 software was used as a GIS environment, where the Inverse Distance Weighted (IDW) interpolation technique installed in the used software was adopted. This technique represents an algorithm used for interpolating a different kind of data spatially as well as estimating or predicting data (Murugesan et al. 2020). In this part of the study, the spatial distribution of the used infected data (actual data) and predicted results (gained from the developed model) was distributed spatially. This method's outputs consisted of purpose maps that can compare
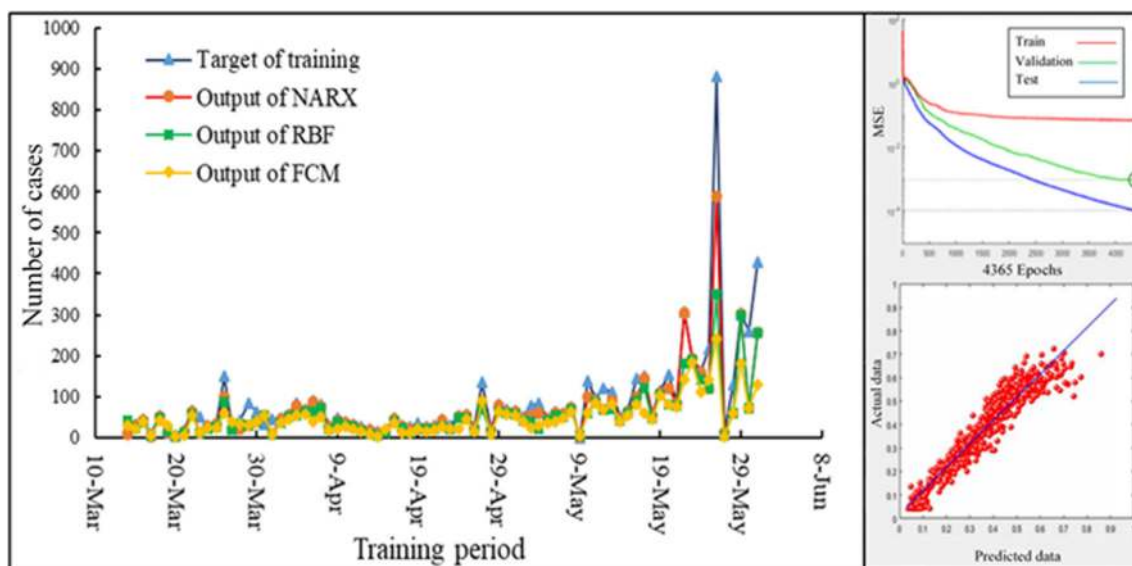

Fig. 4 Performance analysis of model training process.

Table 1 Training process results due to R^2 and RMSE

ANNs	R^2	Av. R^2	RMSE	Av. RMSE	Iteration and epoch
RBF	0.81		0.0084		22 s, reached the epoch 160
NARX	0.86	0.81	0.0013	0.0065	13 s, reached the epoch 185
FCM	0.76		0.0099		50 s, reached the epoch 250

the spread of the COVID-19 disease in the Iraqi governorates between the current period and the predicted period.

GIS environment was employed to prepare short-term spatial distribution maps based on two periods, actual and predicted infected cases for 3 months over Iraq. These spatial distribution maps were created according to the numbers of infected cases in the actual stage compared with the numbers of infected cases gained from the proposed model.

Results and discussion

COVID-19 model implementation

The values discussed in the "Modeling processes" section give acceptable best performance results. The factors that affect the performance of developed RBF, NARX, and FCM were tested which characterized by the learning rate and the iteration time duration. The faster iteration has been accomplished in 23 s, reaching 480 epoch. Figure 4 shows the performance analysis of the model training process according to the three proposed ANNs.

Training stage error analysis

During the training process, the obtained average RMSE shows significant values close to 0, and the average R^2 was found equal to 0.81, as seen in Table 1. Comparatively, RMSE and R^2 for NARX are best than RMSE, and R^2 gained from RBF and FCM means the NARX gives a strong training performance than the RBF and FCM.

Tasting stage error analysis

Table 2 shows that the designed ANNs show performance accuracy ranged from 81.4 to 84.2% according to the estimated performance criteria. Comparatively, the accuracy of NARX is higher than the accuracy of the other designed ANNs.

Model performance accuracy

The best model performance accuracy is gained when the correlation is high, and the estimated errors are as small as possible. To ascertain the model performance accuracy, the tested statistical criteria gave good results where the coefficient of determination (R^2) was 0.895 indicates that there is a perfect positive relationship between the actual and predicted results, and the root mean square error (RMSE) was 0.791, and Nash-Sutcliffe coefficient (CE) was 0.83 indicates that the system has good performance accuracy (Farshid et al. 2021). Furthermore, the AIT has delivered a significant percentage in performance accuracy healthcare and a global public health concern by using different learning approaches and essential data when infection prevention models are developed, where the performance accuracy of developed models ranged between 55 and 98% in some studies (Scardoni et al. 2020). In our study, the mean absolute percentage error (MAPE) between the actual and predicted data was 18.14% that gives a total model performance accuracy equal to 81.6%.

Model productivity results analysis

To estimate the predictive accuracy of the proposed model, the model has executed to predict the infected cases for the current dataset from June 1, 2020, to June 14, 2020, to specify the statistical percentage error as shown in Table 3; this percentage of the statistical error obtained from the correlation between the actual data and the average predicted results of the three ANNs. A small average percentage predicted error was detected equal to 5.74%, where this small error indicates that the proposed model works perfectly.

The model was executed again to predict the infected cases for the 6-month interval from July 15, 2020, to December 15, 2020; the three infected cases predicted time series gained from the three adopted ANNs displayed significant homogeneity with each other, as illustrated in Fig. 5.

Also, according to the comprehensive COVID-19 deaths report for Iraq, the total number reached over than 1450 confirmed deaths; this situation was analysis in the developed

Table 2 Training process results due to R^2, RMSE, and MAPE

Designed ANNs	Performance criteria				
	R^2	RMSE	CE	MAPE	Accuracy
RBF	0.861	0.812	0.74 fair	18.6%	81.4%
NARX	0.905	0.921	0.83 good	15.2%	84.8%
FCM	0.895	0.860	0.81 good	16.2%	83.8%

Table 3 Comparing infected cases results (actual and predicted)

| Predicted infected cases					Statistical error	
Period	RBF	NARX	FCM	Predicted average	Actual dataset	% error
June 1	6460	6451	6675	6528.6	6439	8.98
June 2	6853	6872	6899	6874.6	6868	7.84
June 3	7390	7377	7399	7388.6	7387	4.47
June 4	8182	8172	8177	8177	8168	3.30
June 5	8860	8850	8846	8852	8840	7.75
June 6	9855	9861	9875	9863.6	9846	13.30
June 7	11,125	11,012	11,201	11,112.6	11,098	4.85
June 8	12,345	12,371	12,382	12,366	12,366	3.10
June 9	13,535	13,512	13,589	13,545.3	13,481	3.38
June 10	14,295	14,291	14,282	14,289.3	14,268	2.25
June 11	15,454	15,442	15,462	15,452.6	15,414	3.48
June 11	16,715	16,745	16,736	16,732	16,675	6.20
June 13	17,810	17,799	17,785	17,798	17,770	3.81
June 14	18,982	18,962	18,967	18,970.3	18,950	7.56
					Average error = 5.74%	

model to find the predicted death cases in the next 6 months when the results show that the average death cases increase by 8.3% where according to this result the total expected deaths cases will be 2170 as shown in Fig. 6.

Generally, from observing the model's operation, it was noted that this model gives irrational prediction results until May 17, 2023, then its results start to break down, which means results are irrational. Some of the explanations for this breakdown is the proportional change of standard errors over time. In any case, this breakdown needs to be closely researched, considering some essential factors, such as redesigning the ANNs by increasing the hidden layers and establishing the correct number of neurons. The final version of the developed model is called the presenting stage. Based on this version of the developed model, an application window was configured to be an easy, effective, and executable program interface showing no complications or difficulties. Based on this version, the total infected cases per day, healed cases per day, and death cases per day can be forecasted and illustrated by tables. Furthermore, a specific day can be chosen for forecasting until the day May 17, 2023. Consequently, this version of the upgraded model can display the actual data entered, and the predicted results in several forms like tables and figures.

Infected cases spatial distribution

To obtain the most appropriate map related to Iraq's infected cases, optimum values of these cases are used as input data to the spatial analyst tools in Arc GIS software

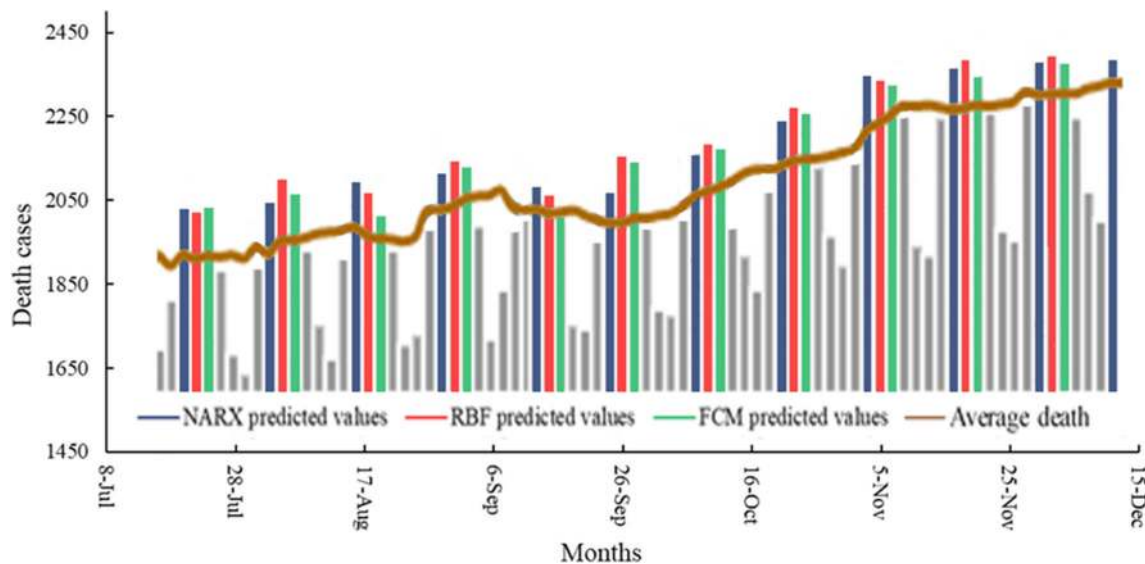**Fig. 5** Predicted infected cases according to the developed model
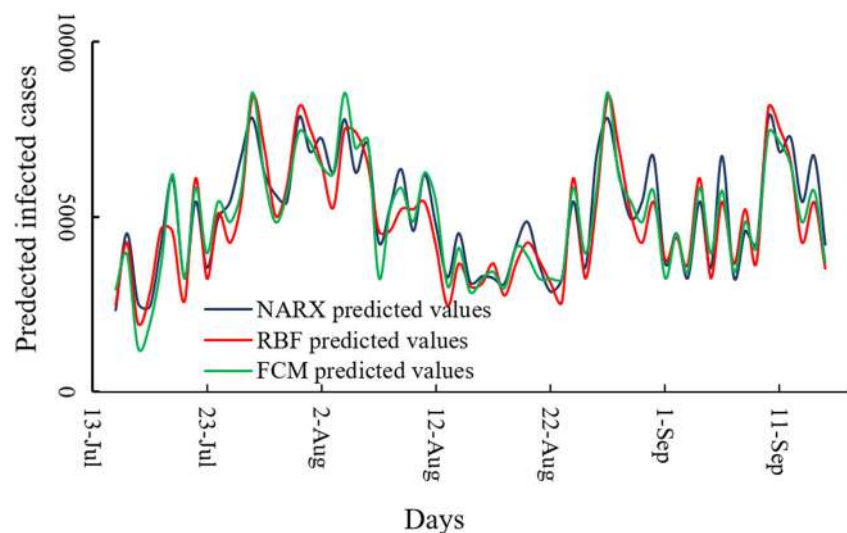
Fig. 6 Death cases, according to the developed model

and then mapped the infected cases. This analysis's output produces a map showing concentrations of the infected cases over Iraq's unit area, Fig. 7. Upon observing the two maps illustrated in Fig. 6, we find a noticeable increase in the spread of the COVID-19 disease in all the Iraqi governorates where the spread level will intensify shortly by 17.1% in Iraq. This matter calls for concern, and several measures must be taken to limit the spread of this disease, where increasing the public awareness in various visual and audio forms to the risks of this disease is one of them.

Conclusion

Developing models based on artificial intelligence procedures in medical applications needs more interest and attention, with the necessity of involving the technical, medical staff in the developing stages, and efficient use to be more clinically useful tools. Due to the increasing prevalence of COVID-19 disease in Iraq, which may get out of control at some point, this study came as an attempt to develop proactive solutions to combat this disease. An AI model was developed throughout this study based on three ANNs; this model achieved 81.6%
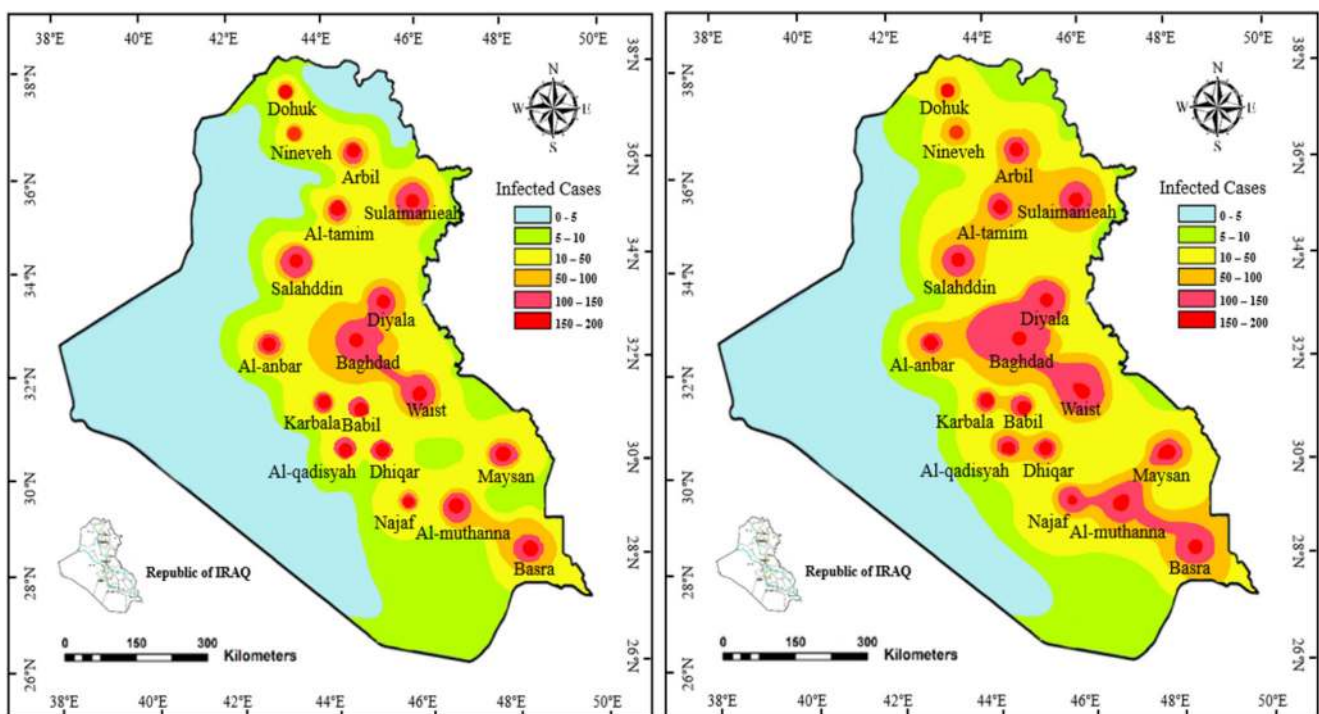

Fig. 7 Short-term spatial distribution maps based on actual and predicted infected cases over Iraq. **a** Actual cases. **b** Predicted cases

as performance accuracy. To make this version of the proposed model a more effective model, the predictive output can improve by continuously collecting infected cases, updating and reviewing data from various remote locations like in districts and remote villages in central the provinces, and selecting and testing more ANNs' tools. This model will provide the possibility to prepare appropriate quarantine places and supplying the necessary medical materials according to forecast infected cases to help the growing number of infected peoples.

The representation of natural phenomena or environmental disasters using GIS technologies is one of the best solutions through which current and plans are put in place to ward off these natural and environmental disasters. In this study, spatial distribution modeling analysis using interpolation IDW integrated with the results gained from the proposed model was performed to illustrate the spread risk of COVID-19 over whole Iraq through a comparison of maps between the actual and predicted infected peoples distributed on all the Iraqi provinces. These maps enable decision-makers in the health and government sectors to anticipate events to develop proactive plans to combat this disease or request the international community's assistance if the health situation gets out of control. This study's procedure can offer more accurate results in future studies when several parameters or information such as age, gender, ethnicity, occupation, pre-existing health conditions, and the local weather tacking into account.

Declarations

Conflict of interest   The authors declare no competing interests.

References

Abebe AJ, Solomatine DP, Venneker RGW (2000) Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events. Hydrol Sci J 45(3):425–436. https://doi.org/10.1080/02626660009492339

Ahmad A, Ghritlahre HK, Chandrakar P (2020) Implementation of ANN technique for performance prediction of solar thermal systems: a comprehensive review. Trends in Renewable Energy 6(1):12–36. https://doi.org/10.17737/tre.2020.6.1.00110

Alamo T, Reina DG, Mammarella M, Abella A (2020) Covid-19: open-data resources for monitoring, modeling, and forecasting the epidemic. Electronics 9(5):827. https://doi.org/10.3390/electronics9050827

Astakhova N, Demidova L, Nikulchev E (2015) Forecasting of time series' groups with application of fuzzy c-mean algorithm. Contemporary Engineering Sciences 8(35):1659–1677. https://doi.org/10.12988/ces.2015.510286

Başkır MB, Türkşen IB (2013) Enhanced fuzzy clustering algorithm and cluster validity index for human perception. Expert Syst Appl 40(3):929–937. https://doi.org/10.1016/j.eswa.2012.05.049

Bello MG (1992) Enhanced training algorithms and integrated training/architecture selection for multilayer perceptron networks. IEEE Trans Neural Netw 3(6):864–875. https://doi.org/10.1109/72.165589

Bezdek, J. C. (1981). Objective function clustering. In Pattern recognition with fuzzy objective function algorithms (pp. 43-93). Springer, Boston, MA. DOI: 10.1007/978-1-4757-0450-1_3

Cadenas E, Rivera W, Campos-Amezcua R, Heard C (2016) Wind speed prediction using a univariate ARIMA model and a multivariate NARX model. Energies 9(2):109. https://doi.org/10.3390/en9020109

Farshid R, Kathryn L, Wenyu O, Alison A, Samantha O, Chaopeng S (2021) Exploring the exceptional performance of a deep learning stream temperature model and the value of streamflow data. Environ Res 16(2). https://doi.org/10.1088/1748-9326/abd501

Fitzpatrick F, Doherty A, Lacey G (2020) Using artificial intelligence in infection prevention. Current Treatment Options in Infectious Diseases 12:1–10. https://doi.org/10.1007/s40506-020-00216-7

He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K (2019) The practical implementation of artificial intelligence technologies in medicine. Nat Med 25(1):30–36. https://doi.org/10.1038/s41591-018-0307-0

Kandwal R, Garg PK, Garg RD (2009) Health GIS and HIV/AIDS studies: perspective and retrospective. J Biomed Inform 42(4):748–755. https://doi.org/10.1016/j.jbi.2009.04.008

Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR (2020) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. Int J Antimicrob Agents 55:105924. https://doi.org/10.1016/j.ijantimicag.2020.105924

Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Lessler J (2020) The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med 172(9):577–582. https://doi.org/10.7326/M20-0504

Liao PH, Hsu PT, Chu W, Chu WC (2015) Applying artificial intelligence technology to support decision-making in nursing: a case study in Taiwan. Health Informatics Journal 21(2):137–148. https://doi.org/10.1177/1460458213509806

Lin T, Horne BG, Tino P, Giles CL (1996) Learning long-term dependencies in NARX recurrent neural networks. IEEE Transactions on Neural Networks 7(6):1329–133. https://doi.org/10.1109/72.548162

Lyseen AK, Nohr C, Sorensen EM, Gudes O, Geraghty EM, Shaw NT, Bivona-Tellez C (2014) A review and framework for categorizing current research and development in health related Geographical Information Systems (GIS) studies. Yearb Med Inform 9:110–124. https://doi.org/10.15265/IY-2014-0008

Mirmozaffari M (2019) Developing an expert system for diagnosing liver diseases. European Journal of Engineering Research and Science 4(3):1–5. https://doi.org/10.24018/ejers.2019.4.3.1168

Murad A, Khashoggi BF (2020) Using GIS for disease mapping and clustering in Jeddah, Saudi Arabia. ISPRS Int J Geo Inf 9(5):328. https://doi.org/10.3390/ijgi9050328

Murugesan B, Karuppannan S, Mengistie AT, Ranganathan M, Gopalakrishnan G (2020) Distribution and trend analysis of COVID-19 in India: geospatial approach. Journal of Geographical Studies 4(1):1–9. https://doi.org/10.21523/gcj5.20040101

Naudé W (2020) Artificial intelligence vs. COVID-19: limitations, constraints, and pitfalls. Ai and Society, 1. https://doi.org/10.1007/s00146-020-00978-0

Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Pantelis G, Lescure FX, Holmes AH (2019) Machine learning for clinical decision support in infectious diseases: a narrative review of current

applications. Clin Microbiol Infect 26:584–595. https://doi.org/10.1016/j.cmi.2019.09.009

Rahmati AS, Tatar A (2019) Application of Radial Basis Function (RBF) neural networks to estimate oil field drilling fluid density at elevated pressures and temperatures. Oil & Gas Science and Technology–Revue d'IFP Energies nouvelles 74:50. https://doi.org/10.2516/ogst/2019021

Ruiz LGB, Cuéllar MP, Calvo-Flores MD, Jiménez MDCP (2016) An application of nonlinear autoregressive neural networks to predict energy consumption in public buildings. Energies 9(9):684. https://doi.org/10.3390/en9090684

Scardoni A, Balzarini F, Signorelli C, Cabitza F, Odone A (2020) Artificial intelligence-based tools to control healthcare associated infections: a systematic review of the literature. Journal of Infection and Public Hhealth 13:1061–1077. https://doi.org/10.1016/j.jiph.2020.06.006

Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, Agha R (2020) World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). Int J Surg 76:71–76. https://doi.org/10.1016/j.ijsu.2020.02.034

Stebbing J, Krishnan V, de Bono S, Ottaviani S, Casalini G, Richardson PJ, Tan YJ (2020) Mechanism of baricitinib supports artificial intelligence-predicted testing in COVID-19 patients. EMBO Molecular Medicine. DOI 10.15252/emmm.202012697

Steiner MC, Gibson KM, Crandall KA (2020) Drug resistance prediction using deep learning techniques on HIV-1 sequence data. Viruses 12(5):560. https://doi.org/10.3390/v12050560

Url-1 <https://covid19.who.int/?gclid=CjwKCAjw4MP5BRBtEiwASfwAL_bHg3-

Url-2 <https://www.worldometers.info/world-population/iraq-population/

Uysal G (2016) Streamflow forecasting using different neural network models with satellite data for snow-dominated regions in Turkey. Procedia Engineering 154:1185–1192. https://doi.org/10.1016/j.proeng.2016.07.526

Wang P, Yao J, Wang G, Hao F, Shrestha S, Xue B, and Peng Y (2019) Exploring the application of artificial intelligence technology for identification of water pollution characteristics and tracing the source of water quality pollutants. Science of the Total Environment, 693, 133440. 10.1016/j

Yahya BM, Seker DZ (2019) Designing weather forecasting model using computational intelligence tools. Appl Artif Intell 33(2):137–151. https://doi.org/10.1080/08839514.2018.1530858

Yalur T (2019) Interperforming in AI: question of 'natural' in machine learning and recurrent neural networks. AI & Soc 35:1–9. https://doi.org/10.1007/s00146-019-00910-1

Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, Liang J (2020) Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease 12(3):165. https://doi.org/10.21037/jtd.2020.02.64

Yokoi S, Takayabu YN, Nishii K, Nakamura H, Endo H, Ichikawa H, Oshima K (2011) Application of cluster analysis to climate model performance metrics. J Appl Meteorol Climatol 50(8):1666–1675. https://doi.org/10.1175/2011JAMC2643.1

Zhang S, Zhang C, Yang Q (2003) Data preparation for data mining. Appl Artif Intell 17(5-6):375–381. https://doi.org/10.1080/08839510390219264

Zhou X, Chen S, Liu B, Zhang R, Wang Y, Li P, Yan X (2010) Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. Artif Intell Med 48(2-3):139–152. https://doi.org/10.1016/j.artmed.2009.07.012

Zu ZY, Jiang MD, Xu PP, Chen W, Ni QQ, Lu GM, Zhang LJ (2020). Coronavirus disease 2019 (COVID-19): a perspective from China. Radiology 200490. 10.1148/radiol.2020200490