Modeling and forecasting number of confirmed and death caused COVID-19 in IRAN: A comparison of time series forecasting methods

Nasrin Talkhi [a], Narges Akhavan Fatemi [b], Zahra Ataei [b], Mehdi Jabbari Nooghabi [b,*]

[a] Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran
[b] Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

A B S T R A C T

Background: The COVID-19 pandemic conditions are still prevalent in Iran and other countries and the monitoring system is gradually discovering new cases every day. Therefore, it is a cause for concern around the world, and forecasting the number of future patients and death cases, although not entirely accurate, helps the governments and health-policy makers to make the necessary decisions and impose restrictions to reduce prevalence.
Methods: In this study, we aimed to find the best model for forecasting the number of confirmed and death cases in Iran. For this purpose, we applied nine models including NNETAR, ARIMA, Hybrid, Holt-Winter, BSTS, TBATS, Prophet, MLP, and ELM network models. The quality of forecasting models is evaluated by three performance metrics, RMSE, MAE, and MAPE. The best model is selected by the lowest value of performance metrics. Then, the number of confirmed and the death cases forecasted for the 30 next days. The used data in this study is the absolute number of confirmed, death cases from February 20 to August 15, 2020.
Results: Our findings suggested that based on existing data in Iran, the suitable model with the lowest performance metrics for confirmed cases data obtained MLP network and the Holt-Winter model is the suitable model for forecasting death cases in the future. These models forecasted on September 14, 2020, we will have 2484 new confirmed and 114 new death cases of COVID-19.
Conclusion: According to the results of this study and the existing data, we concluded that the MLP and Holt-Winter models had the lowest error in forecasting in comparison to other methods. Some models had fitted poorly in the test phase and this is because many other factors that are either not available or have been ignored in this study and can affect the accuracy of forecast results. Based on the trend of data and forecast results, the number of confirmed cases and death cases are almost constant and decreasing, respectively. However, due to disease progression and ignoring the recommendations and protocols of the Ministry of health, there is a possibility of re-emerging this disease more seriously in Iran and this requires more preventive care.

1. Introduction

In late December 2019, a novel virus appeared in Wuhan, China [1], which had an acute effect on the respiratory and it was spreading rapidly [1,2]. The World Health Organization (WHO) introduced this novel virus as SARS-CoV-2 virus, which belongs to the coronavirus family [3].

Some researches and evidence indicate that the main origin of COVID-19 is bats, however, this is not confirmed definitely and needs more investigation and researches [1,3].

This acute infection disease is highly contagious [4]. This virus was declared a global pandemic due to its rapid spread and outbreak in the world [5].

Some of the common symptoms of this disease respiratory issues [1], dry cough [5], fever, chills, difficulty breathing, chest pain [6], pneumonia, etc. [4]. However, as the disease progresses over time, the symptoms in patients are evolving and changing [5].

One of the major problems with this virus is that its incubation period can last up to 14 days and during this period, it can transmit the infection without any symptoms [1,6]. Besides, some people infected with COVID-19 have mild symptoms that look like a common cold or flu [2].

The pandemic has put severe pressure on governments and public health systems [7]. Insufficient medical equipment in hospitals such as beds, ICU beds, staff, ventilators, etc., are some of the major problems

[2,8]. Some other problems that have occurred as a result of the outbreak of this disease and strict quarantine to control it [2,7], are economic and social, affecting the psychological condition of communities, etc. [7].

The occurrence of the above-mentioned problems, and the other hand, issues such as the lack of treatment for this disease so far [2], the dynamic structure of the virus, and its worldwide spread, reveal the need for research on this novel virus and its behavior [2].

Different fields and types of forecasting and modeling are considered. One of these forecasting fields is a model for forecasting the number of cases that will be infected in the future, based on the number of registered confirmed cases. Forecasting the number of future patients, although not entirely accurate, helps the governments and health-policy makers to make the necessary decisions and impose restrictions to reduce prevalence [1].

Also, it is important to forecast future outbreaks, possible mutations of the virus and its spread, and especially the peak time to reduce its severe effects [8]. Forecasting helps decision-makers to prevent and even control the spread of disease by implementing strict and effective policies [2,3,6].

It should be noted that the lack of sufficient information in advance is one of the reasons for the difficulty of forecasting [6], however, it is still an effective policy and guidance for governments to avoid the spread of disease [2,6,8].

Therefore, because statistical and mathematical models that are used to forecast can play an effective role in informing the future trend of the disease [1], in this paper, we applied nine models including NNETAR, ARIMA, Hybrid, Holt-Winter, BSTS, TBATS, Prophet, MLP and ELM model to finding the best model for forecasting numbers of confirmed and death cases, separately, for the 30 next days in Iran.

In the present study, the only available information was the absolute number of confirmed and death cases per day, and other factors were not considered due to unavailability.

The organization of this paper is as: Section 2 explains a brief background of the applied models in this study. Section 3, describes the used data. Section 4 explains the results of this study. Section 5 concludes with a summary of the work performed.

2. Material and methods

In this section, the models used are briefly introduced.

2.1. Neural network auto regression model (NNETAR)

A kind of statistical model is a neural network that it uses in machine learning problems. Neural Network Auto-Regression Model is a kind of neural network and a parametric non-linear model which applied for forecasting problems [9].

In the NNETAR model, forecasting is performed in two phases. For the desired time series, the order of the auto-regressive model is determined in the first phase. In the second phase, the neural network is trained by the training dataset by considering the order of auto-regressive. The number of input nodes or time series lags of the neural network is determined from the order of auto-regressive [9].

In this method, the fitted model with a non-seasonal pattern consists of two components p and k, where p indicates the number of input lags and k indicates the number of hidden neurons. Therefore, this model is presented as NNAR(p, k) form. Also, the fitted model for data with a seasonal pattern is presented as NNAR(p, P, k)[m]. It is similar to ARIMA (p, 0, 0)(P, 0, 0)[m] with nonlinear functions [6].

2.2. Auto-regressive integrated moving average model (ARIMA)

The Box-Jenkins method was proposed by Box, Jenkins [7]. This method includes ARIMA models which are non-stationary time series but they are made stationary with differencing [7].

The auto-regressive integrated moving average (ARIMA) models are one of the most well-known and widely used models in forecasting time series [8]. In the ARIMA models, a linear correlation is considered between the time series and finds patterns of correlation between observations [8]. These models contain three combination models which are the auto-regressive (AR) model and a moving average (MA) model and a white noise process.

A time series y_t follows the auto regressive-moving average (ARMA) model if :

$$y_t = 1^{'} + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + e_t + e_1^{'} e_{t-1} + e_2^{'} e_{t-2} + \ldots + e_q^{'} e_{t-q},$$

where p and q and e_t prefers to auto-regressive part (AR), moving average part (MA), and white noise respectively [10].

The auto-regressive integrated moving average (ARIMA) models are an extension of the ARMA models which is presented by the symbol ARIMA(p, d, q) and it is expressed as follows :

$$\phi_p(B)(1-B)^d y_t = e_0^{'} + e_q^{'}(B)e_t,$$

where p denote orders of auto-regression, q is the order of moving average and d is the number of differencing times. If $d = 0$ then the ARIMA model becomes to ARMA model [10].

2.3. Holt-Winter (HW)

The Holt-Winter forecasting method is an extension of exponential smoothing and applied for univariate time series [8]. This method doesn't need a high data storage and is simple [11]. The HW is suitable for short-term forecasting and uses the maximum likelihood function for estimating parameters [8,11]. There are two Holt-Winter models that use additive or multiplicative models based on the seasonal component [11]. The additive models are applied for a model with a linear trend and with an exponential trend. The Holt-Winters additive model for data with trend and seasonality that don't increase over time is appropriate [8].

Mathematically, the additive model is expressed as follows:

$$\widehat{y}_{t+h/t} = a_t + h*b_t + s_{t-p+1+(h-1)mod(p)},$$

where a_t, b_t, and s_t are expressed as follows:

$$a_t = \alpha(y_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}),$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$$

$$s_t = \gamma(y_t - a_t) + (1-\gamma)s_{t-p}.$$

The multiplicative Holt-Winters forecasting function is expressed as follows:

$$\widehat{y}_{t+h/t} = (a_t + h*b_t)*s_{t-p+1+(h-1)mod(p)},$$

where a_t, b_t, and s_t are expressed as follows:

$$a_t = \alpha(y_t/s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1})$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1},$$

$$s_t = \gamma(y_t/a_t) + (1-\gamma)s_{t-p}$$

where a_t, b_t and s_t, are indicated level, slope, and seasonal of time series at time t, respectively. The p notation indicated the number of seasons in a year. Also, coefficients α, β, and γ are constant and smoothing parameters between zero and one interval. The end h is the forecast horizon [11].

2.4. Hybrid model

There are appropriate functions for ensemble forecasts in R software. In the 'forecastHybrid' package, by default, Forecasts generated from auto.arima(), ets(), thetaf(), nnetar(), stlm(), tbats(), and snaive() can be combined with equal weights. The other weights are based on in-sample errors that introduced by Bates & Granger (1969), or cross-validated weights. Cross-validation is used to evaluate the accuracy of the model and is supported by user-defined models and forecasting functions. Two of the models used in the combination namely, NNETAR, auto.arima have been described Previously [21].

2.5. Bayesian structural time-series (BSTS)

The Bayesian approach based on prior experience (prior distribution) and given data (likelihood function) builds analytical models [12]. The prior distribution and likelihood function are multiplied to make the posterior distribution and this leads to the final Bayesian model [12].

Structural time series models belong to the family of state-space models that are applied for time series data. They can be expressed in terms of a pair of equations

$$y_t = Z_t^T \alpha_t + \varepsilon_t$$

$$\alpha_{t+1} = T_t + \alpha_t + R_t \eta_t$$

The first equation is the observation equation; this equation links the observed data y_t to a latent d-dimensional state vector α_t. The latter equation is the state equation which describes how the latent state evolves through time. The error terms ε_t and η_t are Gaussian and independent of everything else. In these equations, y_t is a scalar observation, Z_t is output vector, T_t is transition matrix, R_t is control matrix. In other words, Z_t, T_t and R_t are structural parameters [12,13].

Structural time-series models are useful and flexible because they are a very large class of models, including all ARIMA models. By using these models can be built time series models for short- and long-term forecasting [13].

2.6. TBATS model

The phrase BATS is abbreviated based on five features including Box-Cox transform, ARMA errors, Trend, and Seasonal components. It is supplemented by $(\omega, \emptyset, p, q, m_1, ..., m_T)$ to presenting the Box-Cox, damping, ARMA(p, q), and Seasonal periods $(m_1, ..., m_T)$ [8,14]. This model is a generalization of the traditional seasonal models with multiple seasonal periods [14].

To make a more parsimonious approach, the trigonometric representation of seasonal components based on the Fourier series is introduced [8,14]. This class of model is called TBATS which the first T notation referred to "trigonometric". This model considers any autocorrelation in the residuals and handles nonlinear attributes in real-time series [14]. Also, it includes a large parameter space with the possibility of better forecasts and it is an efficient estimation procedure totally [8].

2.7. Prophet: automatic forecasting procedure

There is an available forecasting tool called Prophet in R and Python. In fact, the prophet is an additive regression that has a linear trend in piecewise or logistic growth curve trend.

It includes a yearly seasonal component modeled using the Fourier series and a weekly seasonal component modeled using dummy variables. The Prophet is used for business tasks that we deal with on Facebook and has been optimized for this purpose [8].

The method uses a decomposable time-series model consisting of trend, seasonality, and holiday components.

The Prophet depends on the Fourier series to consider seasonality. Thus it creates a more flexible model for periodic effects. Also, to account for holidays, this model requires a predefined list of past and future holiday events [8].

2.8. Multilayer perceptron (MLP)

MLP network is a kind of the main perceptron model [15]. The network architecture is displayed in Fig. 1. MLPs include at least three layers. This model consists of inputs, weights, biases, and an activation function that yields the output [16]. Each input x_i to a neuron, j is multiplied by an adaptive coefficient w_{ij}, called weight, then with a nonlinear activation function (φ) such as sigmoid, hyperbolic tangent, etc. calculate the weighted sum of the inputs as shown in the following Equation:

$$o_i = \varphi \left(\sum_{j=1}^{d} (x_j w_{ij} + b_j) \right)$$

An activation function enables the network to map an input to output, and also the network learns to represented complex data. In other words, from a statistical point of view, MLPs run nonlinear regression [15].

In the output o_i of a neuron in the MLP network, d is the number of the inputs x_j, b_j and w_{ij} are the bias and weights associated with each x_j. In the model training phase, the coefficients or weights of the network are adjusted based on calculating error function and in the next steps, weights are updated based on the learning rate and the error in each iteration. In the final step, all steps are repeated until reaching the number of epochs [16].

2.9. Extreme learning machines (ELM)

The Extreme Learning Machine is a learning algorithm with high speed for the single hidden layer feed-forward neural networks (SLFN) [17]. The ELM network structure is shown in Fig. 2.

This method overcomes the debility of the traditional learning algorithms in the process of learning speed because ELM could be improving the generalization performance and reducing the training time [6]. In other words, ELMs in comparison with traditional learning algorithms tend to reach the smallest training error [6].

The input weights and the hidden layer biases are determined randomly and only the output layer is trained [6,17].

Consider the training sample $\{X, T\} = \{x_i, t_i\}$. Input feature is $X = [x_{i1}, x_{i2}, ..., x_{iN}]$, $i = 1, 2, ..., N$ and output matrix is $T = [t_{j1}, t_{j2}, ..., t_{jN}]$, $j = 1, 2, ..., m$, where n and m are the dimension of the input matrix and the output matrix [6].
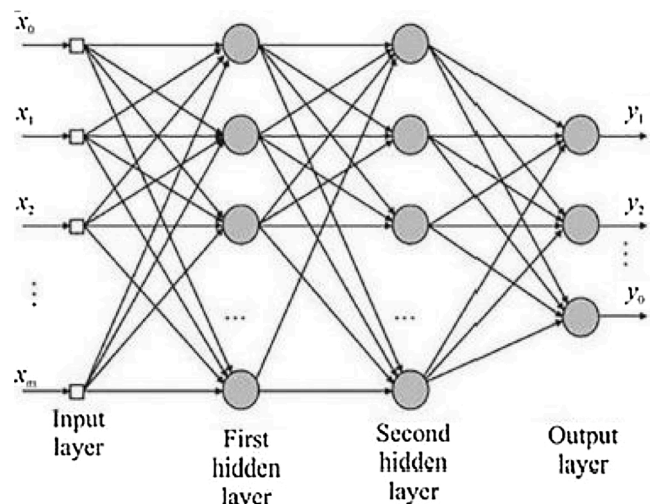

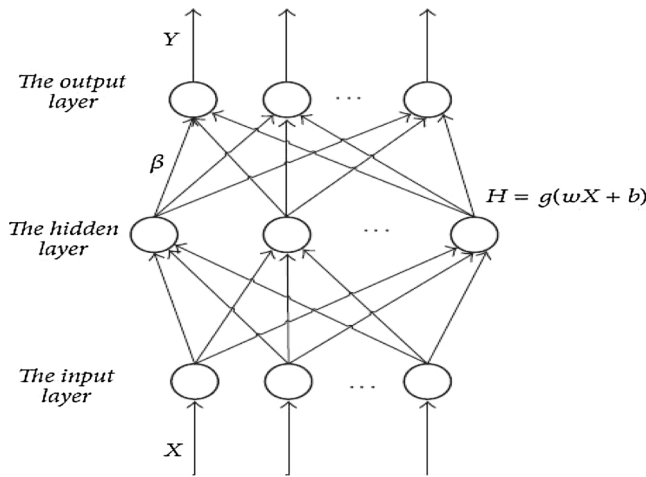**Fig. 1.** The MLP network structure.

Fig. 2. The ELM network structure.

After that weights between the hidden layer and the output layer and bias of the hidden layer neurons set randomly, in the next step the ELM select the network activation function g(x).

Therefore, the output matrix T can be expressed as follows:

$$T = [t_1, t_2, \ldots, t_N]_{m*N}.$$

Each column vector of the output matrix T is as follows:

$$t_j = \sum_{i=1}^{l} \beta_i g(w_i x_j + b_i) \, , \, j = 1, 2, \ldots, N.$$

The above equation can be written in matrix form as:

$$H\beta = T^{`}$$

where $T^{`}$ is the transpose of T and H is the output matrix of the hidden layer. Using the least squares leads to a unique solution and this solution has a minimum-error. Thus, the weight matrix values of β calculate by this approach.

$$\widehat{\beta} = H^{\dagger} T^{`}$$

where $\widehat{\beta}$ is used as the estimated value of b and $H^{\dagger}$ is the Mooree-Penrose generalized inverse of matrix H [6,17].

2.10. Model evaluation

To evaluate the quality or goodness of fit of the used methods in this study, we applied three performance metrics, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) in the training and testing phases. These measures are defined as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \widehat{y}_i)^2},$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \widehat{y}_i|,$$

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \frac{|y_i - \widehat{y}_i|}{y_i} *100\%$$

where y_i is the actual value of time series at time i, and $\widehat{y}_i$ is the forecast value of the time series at time i [1].

3. Data collection and results

In this study, to forecast future behavior of COVID-19, we used the COVID-19 dataset included the absolute number of confirmed, death, and recovered cases caused by the new coronavirus in Iran. The dataset was available on the https://www.worldometers.info/coronavirus/ website and these data were reported daily from February 20 to August 15, 2020, on this site. All data analysis was performed using **R** software version 4.0.2.

In the current study, we intended to find a model for forecasting numbers of confirmed and death cases in the future. The trend of daily confirmed, death, and recovered cases in Iran from February 20 to August 15, 2020, is shown in Fig. 3. To the better presentation of numbers of death cases, we multiple it by 10. In this study, nine different methods were fitted to the data of COVID-19 (confirmed and death cases). We evaluated the performance of methods by training and testing dataset. The first 70 % of data are used as training and the next 30 % data for testing the models. Then, the forecasting quality of the models is evaluated by three metrics RMSE, MAE, and MAPE.

In the training phase, we trained the NNETAR, ARIMA, Hybrid, Holt-Winter, and BSTS models by training data for confirmed and death cases separately. The auto.arima function has been used to fit the ARIMA model to the data. In this way, the best proposed ARIMA model was ARIMA(1, 0, 0) and ARIMA(1, 0, 1) for confirmed and death cases, respectively.

Then, the NNETAR model fits. In this model, input variables are scaled and the obtained model is by first input lag and one hidden node. The next model is a Hybrid model that is a combination of two models, ARIMA and NNETAR. The Hybrid model assigns weight to each of the models.

In order to, there are three approaches, by "equal", "cv.errors" (i.e. Cross Validated errors), and "insample.errors". We implemented this model with two approaches, "equal", "cv.errors" and then we indicated the two relevant models with the symbols "Hybrid-e" and "Hybrid-c", respectively. In the Hybrid-c model, for confirmed cases, the weights were 0.495 for ARIMA and 0.505 for NNETAR, and also for death cases, weights were 0.499 for ARIMA and 0.501 for the NNETAR model.

Next, we train the MLP and ELM models, the number of hidden layers and hidden nodes in each layer were determined by the 5-Fold cross-validation method automatically. Moreover, the activation function was considered a sigmoid function, and the model training was conducted by 20 iterations. In the end, the non-seasonal Holt-Winter model, Bayesian Structural Time-Series model (BSTS), TBATS, and Prophet models are fitted too.

In the testing phase, we forecasted the length of the test data by the training model in the previous phase and compared it with testing data. The performance metrics RMSE, MAE, and MAPE calculated for all of the models in the training and testing phases. These results are reported in Tables 1 and 2 . Also, we showed these results graphically using bar graphs in Fig. 4.

By comparing performance metrics, we concluded that for confirmed cases, except for the Hybrid-e model, other models did not perform well in the test phase. Also, the Holt-Winter model was the best model with the lowest performance metrics for death cases time series data. Therefore, the Hybrid-e and Holt-Winter models are the best models with the lowest performance metrics to forecasting confirmed and death cases, respectively.

After determining the best models, we forecasted the future behavior of the time series of confirmed and death cases for the next 30 days using these models. The 30-days COVID-19 forecasting graphs of confirmed and death cases (Fig. 5) were plotted.

The results of the forecast showed which on September 14, 2020, we will have 2484 new confirmed and 114 new death cases of COVID-19. These values for 30 days are reported in the Appendix.
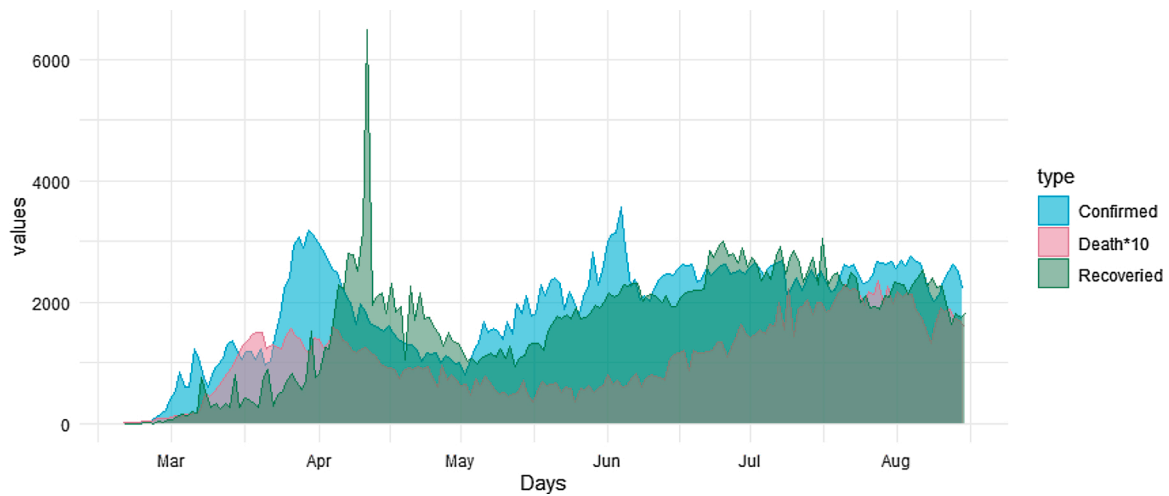
Fig. 3. The trend of Daily of Confirmed, death, and Recovered cases.

Table 1
The results of the models for confirmed cases.

Models	Confirmed Cases					
	Training Data			Testing Data		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
NNETAR(1,1)	255.7547	204.3763	39.566	291.4161	260.1861	10.22983
ARIMA(1,0,0)	231.6003	177.2125	82.10807	561.9214	501.4737	26.62457
Hybrid-e	227.5012	175.0365	21.23171	180.8860	151.9495	6.268913
Hybrid-c	227.4615	175.0335	21.34771	180.8883	151.9539	6.269047
Holt-Winter	233.5451	177.73	13.07673	299.6471	226.3595	9.735324
BSTS	254.8199	195.7948	16.58057	550.1058	455.7354	19.13969
TBATS	225.6698	170.7427	15.62544	217.2329	185.6827	7.394939
Prophet	608.2165	441.5421	311.6574	612.9864	537.7585	22.4437
MLP	**224.4852**	**177.5885**	**24.95336**	**180.2759**	**142.8951**	**5.725628**
ELM	237.8037	190.5021	39.43857	443.9748	405.2195	19.68961

Table 2
The results of the models for death cases.

Models	Death Cases					
	Training Data			Testing Data		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
NNETAR(1,1)	14.14151	10.79158	24.94921	81.83506	75.38808	39.47772
ARIMA(1,0,1)	12.34115	9.318635	23.15612	89.47732	81.7967	84.53056
Hybrid-e	11.85159	8.795046	13.7387	65.13031	58.00313	29.9145
Hybrid-c	11.85194	8.795424	13.73874	65.13291	58.00584	29.91598
Holt-Winter	**12.38061**	**9.435316**	**14.21699**	**35.4963**	**26.75278**	**15.10667**
BSTS	12.86378	9.834921	15.14902	48.90122	41.58697	21.41159
TBATS	12.30943	9.057055	14.30562	42.37191	35.50072	18.09161
Prophet	37.13429	31.7645	175.111	101.7453	97.02142	51.92662
MLP	11.6038	8.513807	14.5441	60.86964	53.39749	27.38357
ELM	12.79517	10.33391	27.59607	87.46979	80.55371	42.1807

4. Discussion

About seven months after the onset of the COVID-19 pandemic, the pandemic conditions are still prevalent in Iran and other countries and the monitoring system is gradually discovering new cases every day. Therefore, it is a cause for concern around the world. The vaccine for this disease has not been definitively discovered yet, and even if it is discovered, there is no guarantee that the primary vaccine will be highly effective [18].

In the absence of vaccines or antiviral drugs for COVID-19, effective non-pharmacological interventions, such as personal protection and social distancing, etc., are critical to controlling the pandemic [19,20].

Because statistical and mathematical models that are used to forecast can play an effective role in informing the future trend of the disease, in this paper, we applied nine models including NNETAR, ARIMA, Hybrid, Holt-Winter, BSTS, TBATS, and Prophet model to find the best model for forecasting numbers of confirmed and death cases, separately, for the 30 next days in Iran. After fitting these models to data, we compared models together by RMSE, MAE, and MAPE measures.

In respect of obtained results in the training and testing phase, the best model with the best performance (lowest RMSE, MAE, and MAPE) and precision for confirmed and death cases was Hybrid (by equal weight) and Holt-Winter models, respectively. These models forecast which on September 14, 2020, we will have 2484 new confirmed and
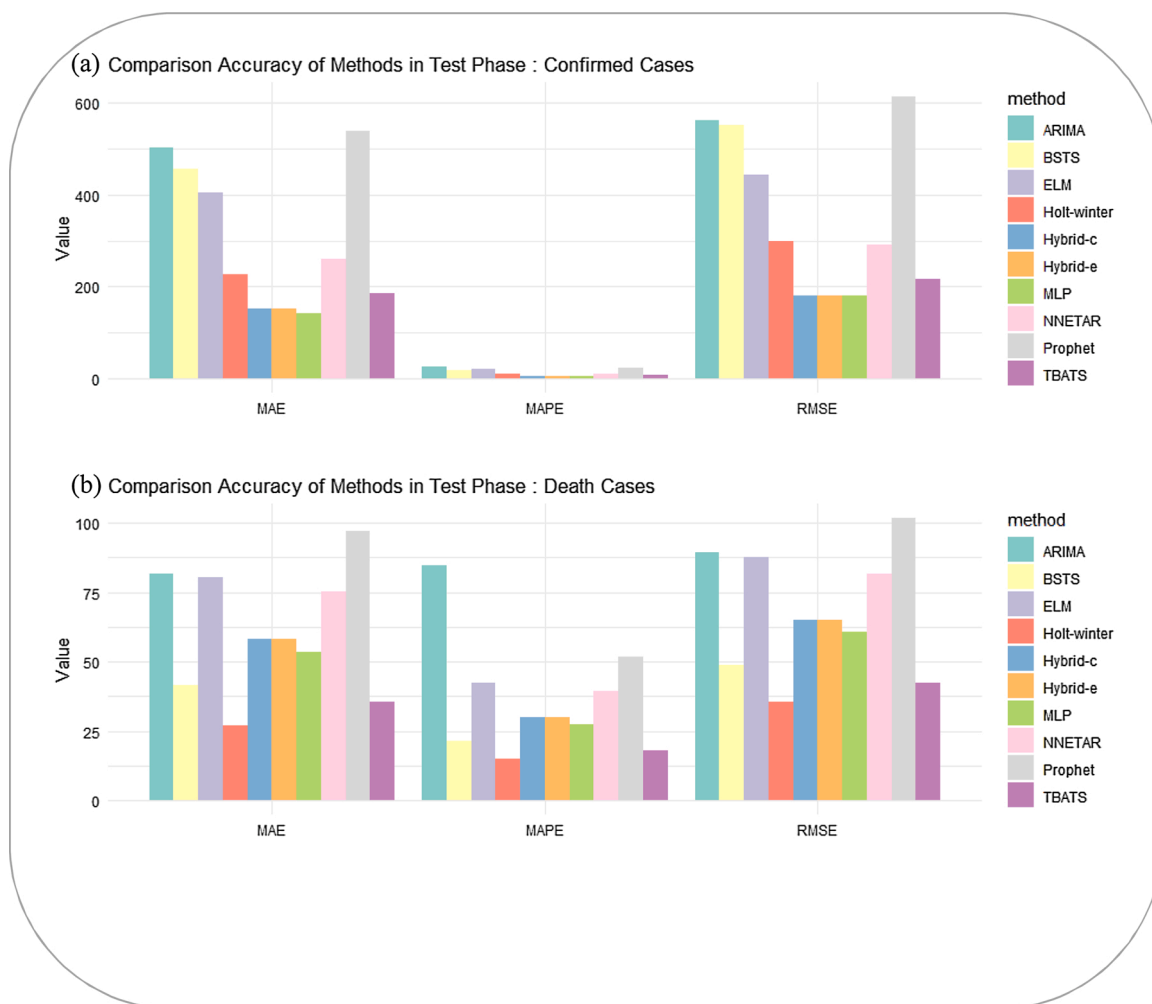
Fig. 4. The comparison of the performance metrics models for the confirmed and death in the test phase.

114 new death cases of COVID-19.

All models except the Hybrid model (about confirmed cases) performed poorly in the test phase compared to the training phase, but we used the test phase results to select the best model. This is because many other factors are either not available or have been ignored in this study and can affect the accuracy of forecast results.

It should be noted that in this 9 model, we used limited and available data including the number of cases and the number of deaths. While for the accuracy of the forecasts, other predictor variables affect the increase or decrease in the number of cases or the number of deaths that were not considered.

These Factors such as age, gender [22,23], other chronic diseases [24], environmental factors, quarantine [3], guidelines, and decisions implemented by governments to reduce the incidence of disease [6], cultural and social issues, health policies, preventive restrictions [7], may have a significant impact on newly infected cases, while we do not take them into account in the forecasting process.

Another issue is that the exact cause and trend of the epidemic are not yet known obviously, while for a more accurate prediction, the actual situation must be considered [4].

On the other hand, lack of diagnostic kits at the beginning of the pandemic, the presence of infected but asymptomatic individuals who have not been diagnosed [3], the duration and severity of restrictions such as social distancing [7], or other factors such as changes in air temperature, humidity, and even air quality that occur during the pandemic period are effective and will affect the results of the forecast [8]. These factors are a kind of limitation for forecasting and studies

about COVID-19, and to achieve the most accurate results, they must be considered, which can be of future study.

Other studies have been confirmed in this field. For example, Moftakhar et al. [3] used two ANN and ARIMA models to forecast the number of new cases during 30 days in Iran, and by comparing the results of the two models, proposed the ARIMA model as a more accurate method for forecasting [3]. Besides, Yang et al. [4] used the ARIMA models to forecast the number of new cases and deaths in Italy too, based on data from Hubei, China [4].

Pantoh et al. [6]. proposed the MLP model for forecasting cases in South Korea if other factors affecting the cases are not considered. The MLP model has been proposed as a suitable model for forecasting the number of confirmed, recorded, and fatal cases using cumulative data in this country [6]. The findings of our research revealed that MLP is a suitable model for forecasting the number of confirmed cases. Therefore, the finding of the current study about the confirmed cases corresponds with the findings of Pantoh et al. [6] study.

Also, Yonar et al. [7] used some curve estimation models, Box-Jenkins (ARIMA) and Brown/Holt linear exponential smoothing methods to forecast the number of patients in the coming days, based on available data. They chose Germany, United Kingdom, France, Italy, Russian, Canada, Japan, and Turkey, for their studies [7].

In another study, Papastefanopoulos et al. [8] used six statistical models to estimate the percentage of active cases for the total population, starting from May 4 for the next 7 days in 10 countries. The models they used are ARIMA, the Holt-Winters additive model (HWAAS), TBAT, Facebook's Prophet, Deep AR, and N-Beats. Their ten selected countries,
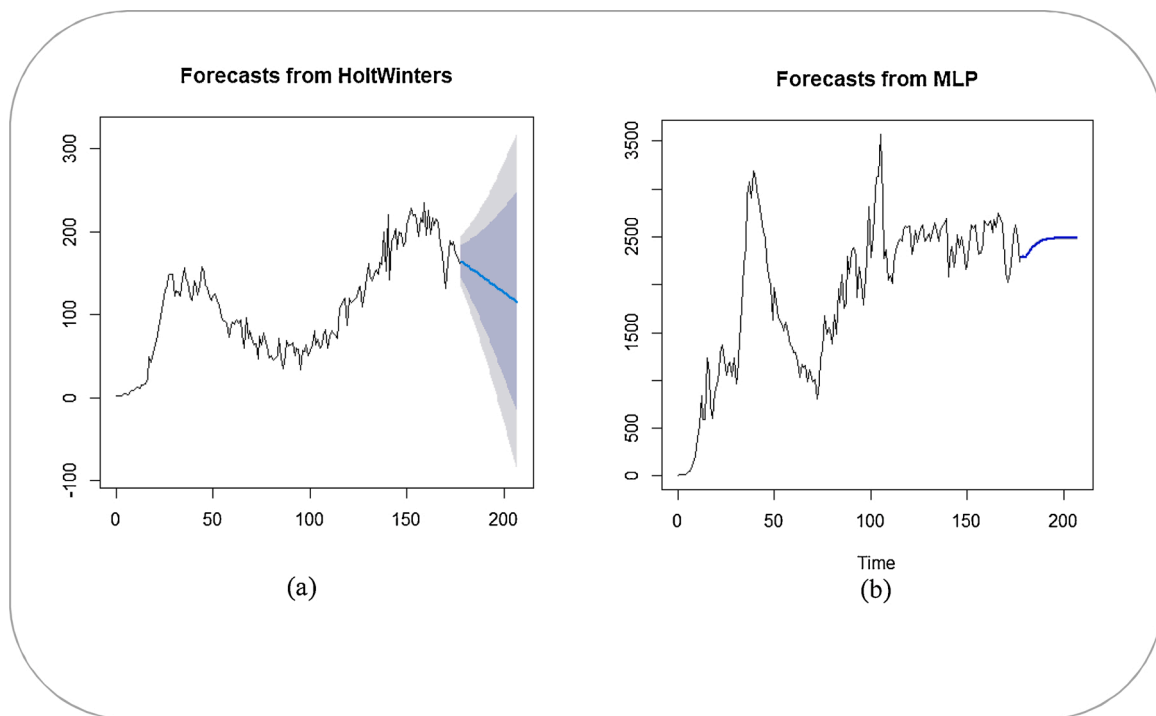
Fig. 5. Forecasting future of the time series for (a) confirmed cases by MLP model (b) death cases by Holt-Winter model.

the USA, UK, Italy, Spain, Russian, France, Turkey, Germany, Iran, and Brazil, were the countries with the highest number of confirmed cases [8].

In the previous studies, we did not find a study that compares all the models in this article, and also, no study was found that used the hybrid model in the "forecastHybrid" Package for forecasting COVID-19.

5. Conclusion

The purpose of this study is to model the COVID-19 data and find the best model for forecasting the behavior of this disease in the future. For this purpose, nine models such as NNETAR, ARIMA, Hybrid, Holt-Winter, BSTS, TBATS, Prophet, MLP, and ELM network, which are used for forecasting, were fitted to COVID-19 data. Based on the findings of this study, it was concluded that the better model with less forecasting error on unseen data of confirmed cases is the MLP network and therefore, it can forecast the confirmed cases in the future more accurately than other models. Also, to forecasting the death cases, the Holt-Winter model has a lower forecasting error and can be used to forecasting death cases.

Based on the trend of data and forecast results, the number of confirmed and death cases are almost constant and decreasing, respectively. However, given that the disease is still progressing and ignoring

the recommendations and protocols of the Ministry of Health i.e. stopping the approach of applying strict government restrictions and policies such as closing schools, stopping business and travel, etc., there is a possibility of more prevalence and re-emerging of this disease more seriously in Iran.

CRediT authorship contribution statement

Acknowledgments

Declaration of Competing Interest

The authors have no actual or potential conflicts of interest related to this manuscript.

Appendix A

Forecasted days	Forecasted confirmed cases by MLP model	Forecasted death cases by Holt-Winter model
2020−08−16	2287.313	164.1993
2020−08−17	2292.207	162.4906
2020−08−18	2282.491	160.7819
2020−08−19	2319.29	159.0733
2020−08−20	2346.384	157.3646
2020−08−21	2368.514	155.6559
2020−08−22	2394.127	153.9472
2020−08−23	2414.632	152.2385
2020−08−24	2431.233	150.5298

(*continued*)

Forecasted days	Forecasted confirmed cases by MLP model	Forecasted death cases by Holt-Winter model
2020−08-25	2445.142	148.8211
2020−08-26	2455.391	147.1124
2020−08-27	2463.584	145.4037
2020−08-28	2469.989	143.695
2020−08-29	2474.524	141.9863
2020−08-30	2477.602	140.2776
2020−08-31	2480.011	138.5689
2020−09-01	2481.69	136.8602
2020−09-02	2482.852	135.1515
2020−09-03	2483.637	133.4428
2020−09-04	2484.151	131.7341
2020−09-05	2484.481	130.0254
2020−09-06	2484.685	128.3167
2020−09-07	2484.806	126.608
2020−09-08	2484.874	124.8993
2020−09-09	2484.908	123.1906
2020−09-10	2484.921	121.4819
2020−09-11	2484.924	119.7732
2020−09-12	2484.92	118.0645
2020−09-13	2484.914	116.3558
2020−09-14	2484.907	114.6471

References

[1] M.A.A. Al-Qaness, A.A. Ewees, H. Fan, M. Abd El Aziz, Optimization method for forecasting confirmed cases of COVID-19 in China, J. Clin. Med. 9 (3) (2020).

[2] Z. Ceylan, Estimation of COVID-19 prevalence in Italy, Spain, and France, Sci. Total Environ. 729 (2020), 138817.

[3] L. Moftakhar, M. Seif, M.S. Safe, Exponentially increasing trend of infected patients with COVID-19 in Iran: a comparison of neural network and ARIMA forecasting models, Iran. J. Public Health 49 (Supple 1) (2020).

[4] Q. Yang, J. Wang, H. Ma, X. Wang, Research on COVID-19 based on ARIMA modelΔ—taking Hubei, China as an example to see the epidemic in Italy, J. Infect. Public Health (2020).

[5] K.K. Sahu, A.K. Mishra, A. Lal, COVID-2019: update on epidemiology, disease spread and management, Monaldi Arch Chest Dis [Internet] (2020), 2020/04//; 90 (1). Available from: http://europepmc.org/abstract/MED/32297723, https://doi. org/10.4081/monaldi.2020.1292.

[6] R.S. Pontoh, Z S, Y. Hidayat, R. Aldella, N.M. Jiwani, Sukono, Covid-19 modelling in South Korea using a time series approach, Int. J. Adv. Sci. Technol. 29 (7) (2020) 1620–1632.

[7] H. Yonar, A. Yonar, M. Agah Tekindal, M. Tekindal, Modeling and forecasting for the number of cases of the COVID-19 pandemic with the curve estimation models, the box-jenkins and exponential smoothing methods, EJMO 4 (2) (2020) 160–165.

[8] V. Papastefanopoulos, P. Linardatos, S. Kotsiantis, COVID-19: a comparison of time series methods to forecast percentage of active cases per population, Appl. Sci. Basel. 10 (11) (2020) 3880.

[9] D. Sena, N.K. Nagwani, A neural network autoregression model to forecast per capita disposable income, ARPN J. Eng. Appl. Sci. 11 (2016) 13123–13128.

[10] M. Almasarweh, S. Alwadi, ARIMA model in predicting banking stock market data, Mod. Appl. Sci. 12 (11) (2018) 4.

[11] A.M. Awajan, M.T. Ismail, S. Al Wadi, Improving forecasting accuracy for stock market data using EMD-HW bagging, PLoS One 13 (7) (2018), e0199582.

[12] S. Jun, Bayesian structural time series and regression modeling for sustainable technology management, Sustainability 11 (18) (2019) 4945.

[13] K.H. Brodersen, F. Gallusser, J. Koehler, N. Remy, S.L. Scott, Inferring causal impact using Bayesian structural time-series models, Ann. Appl. Stat. 9 (1) (2015) 247–274.

[14] A.M. De Livera, R.J. Hyndman, R.D. Snyder, Forecasting time series with complex seasonal patterns using exponential smoothing, J. Am. Stat. Assoc. 106 (496) (2011) 1513–1527.

[15] S. Kaushik, A. Choudhury, P.K. Sheron, N. Dasgupta, S. Natarajan, L.A. Pickett, et al., AI in healthcare: time-series forecasting using statistical, Neural, Ensemble Archit. 3 (4) (2020).

[16] L. Parhizkari, A. Najafi, M. Golshan, Medium term electricity price forecasting using extreme learning machine, J. Energy Manage. Technol. 4 (2) (2020) 20–27.

[17] J. Lai, X. Wang, R. Li, Y. Song, L. Lei, BD-ELM: a regularized extreme learning machine using biased dropconnect and biased dropout, Math. Probl. Eng. (2020) 1–7.

[18] L. Mounesan, S. Eybpoosh, A. Haghdoost, G. Moradi, E. Mostafavi, Is reporting many cases of COVID-19 in Iran due to strength or weakness of Iran's health system? Iran. J. Microbiol. 12 (2) (2020) 73–76.

[19] K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J.M. Hyman, et al., Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020, Infect. Dis. Model. 5 (2020) 256–263.

[20] S. Eubank, I. Eckstrand, B. Lewis, S. Venkatramanan, M. Marathe, C.L. Barrett, Commentary on Ferguson, et al., Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, Bull. Math. Biol. 82 (4) (2020) 52.

[21] https://cran.r-project.org/web/packages/forecastHybrid/index.html website.

[22] W. Wang, J. Tang, F. Wei, Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China, J. Med. Virol. 92 (4) (2020) 441–447.

[23] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet (London, England) 395 (10223) (2020) 497–506.

[24] A. Tavakoli, K. Vahdat, M. Keshavarz, Novel Coronavirus Disease 2019 (COVID-19): An Emerging Infectious Disease in the 21st Century, BPUMS 22 (6) (2020) 432–450.