Process Safety and Environmental Protection 149 (2021) 399–409

Contents lists available at ScienceDirect

Process Safety and Environmental Protection

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

Efficient artificial intelligence forecasting models for COVID-19 outbreak in Russia and Brazil

Mohammed A.A. Al-qaness[a,*], Amal I. Saba[b], Ammar H. Elsheikh[c], Mohamed Abd Elaziz[d], Rehab Ali Ibrahim[d], Songfeng Lu[e], Ahmed Abdelmonem Hemedan[f], S. Shanmugan[g], Ahmed A. Ewees[h,i]

[a] *State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China*
[b] *Department of Histology, Faculty of Medicine, Tanta University, Tanta 31527, Egypt*
[c] *Department of Production Engineering and Mechanical Design, Faculty of Engineering, Tanta University, Tanta 31527, Egypt*
[d] *Department of Mathematics, Faculty of Science, Zagazig University, Zagazig 44519, Egypt*
[e] *School of Computer Science and Technology, Huazhong University of Science and Technology, China*
[f] *Bioinformarics Core Luxebourg Centre For System Biomdeicine, Luxembourg University, Luxembourg*
[g] *Research Centre for Solar Energy, Department of Physics, Koneru Lakshmaiah Education Foundation, Green Fields, Guntur District, Vaddeswaram, Andhra Pradesh 522502, India*
[h] *Department of e-Systems, University of Bisha, Bisha 61922, Saudi Arabia*
[i] *Department of Computer, Damietta University, Damietta 34517, Egypt*

ARTICLE INFO

Article history:
Received 8 June 2020
Received in revised form 26 October 2020
Accepted 4 November 2020
Available online 13 November 2020

Keywords:
COVID-19
Optimization
Chaotic marine predators algorithm
Forecasting
Artificial intelligence
Russia
Brazil

ABSTRACT

COVID-19 is a new member of the Coronaviridae family that has serious effects on respiratory, gastrointestinal, and neurological systems. COVID-19 spreads quickly worldwide and affects more than 41.5 million persons (till 23 October 2020). It has a high hazard to the safety and health of people all over the world. COVID-19 has been declared as a global pandemic by the World Health Organization (WHO). Therefore, strict special policies and plans should be made to face this pandemic. Forecasting COVID-19 cases in hotspot regions is a critical issue, as it helps the policymakers to develop their future plans. In this paper, we propose a new short term forecasting model using an enhanced version of the adaptive neuro-fuzzy inference system (ANFIS). An improved marine predators algorithm (MPA), called chaotic MPA (CMPA), is applied to enhance the ANFIS and to avoid its shortcomings. More so, we compared the proposed CMPA with three artificial intelligence-based models include the original ANFIS, and two modified versions of ANFIS model using both of the original marine predators algorithm (MPA) and particle swarm optimization (PSO). The forecasting accuracy of the models was compared using different statistical assessment criteria. CMPA significantly outperformed all other investigated models.

1. Introduction

The new Coronavirus disease, COVID-19 has been reported in Wuhan, Hubei province, China, in December 2019 (Mi et al., 2020). Thereafter, Wuhan became the center of that outbreak, which spreads globally with a high prevalence rate. The COVID-19 outbreak was declared by WHO as a pandemic on 11 March 2020. The COVID-19 pandemic has spread rapidly all over the world causes about 41,570,883 infected cases and about 1,134,940 deaths around the world (till 23 October 2020). This pandemic has serious effects on the world economy. Therefore, it has been considered as the most critical universal crisis since the World War-II (Boccaletti et al., 2020).

COVID-19 spreads among humans via contacting infected persons or touching objects contaminated with viral particles (Wu et al., 2020). The incubation period of COVID-19 ranges from one day to fourteen days (with a median of three days) (Kang et al., 2020). The most critical issue related to the spreading of this virus is that it can transmit from undiagnosed infected persons with no symptoms of disease (Cao et al., 2020). The main symptoms of COVID-19 are fever, cough, sore throat, shortness of breath, and headache. There are three main levels of these symptoms; mild,

* Corresponding author.
E-mail addresses: alqaness@whu.edu.cn (M.A.A. Al-qaness), amal.saba@med.tanta.edu.eg (A.I. Saba), ammar_elsheikh@f-eng.tanta.edu.eg (A.H. Elsheikh), abd_el_aziz_m@yahoo.com (M.A. Elaziz), rehab100r@yahoo.com (R.A. Ibrahim), lusongfeng@hust.edu.cn (S. Lu), ahmed.hemedan.001@student.uni.lu (A.A. Hemedan), s.shanmugam1982@gmail.com (S. Shanmugan), ewees@du.edu.eg (A.A. Ewees).

https://doi.org/10.1016/j.psep.2020.11.007
0957-5820/

moderate, and severe depending on the immunity of the person (Ye et al., 2020). Severe symptoms may cause respiratory failure, which needs mechanical ventilation. The worst-case scenario may occur in elderly patients with other chronic diseases such as hypertension, diabetes, cancer, autoimmune and cardiovascular diseases.

The modeling and prediction of the prevalence of COVID-19 outbreak, as well as exploring its epidemiological characteristics, are critical topics that should be seriously investigated (Xu et al., 2020). Subsequently, protection policies and thorough plans could be implemented to control the outbreak of COVID-19 diseases. Mathematical and statistical modeling approaches have been utilized to determine the spreading of that outbreak in different countries. The epidemic curve of the confirmed COVID-19 cases was modeled using the exponential growth approach (Zhao et al., 2020a). The increasing rate of the confirmed cases was estimated to increase 21-fold within less than three weeks. The rate of under-ascertainment of COVID-19 infection was estimated based on the evacuation flights data of Japanese citizens from Wuhan, China, at the end of January 2020 using mean serial interval approach (Nishiura et al., 2020). COVID-19 forecasting and transmission in Marche, Italy, were estimated by employing the R statistics approach. Incidence/projections package as a probability-based prediction model was incorporated with a Poisson regression model to predict the geographical characteristics and transmission dynamics of the virus. An autoregressive integrated moving average approach was employed for forecasting the recovered and confirmed COVID-19 cases in Italy (Chintalapudi et al., 2020). Reasonable prediction accuracy of 93.75% and 84.4% were achieved for the confirmed cases and recovered cases, respectively.

An enhanced autoregressive time series model was developed to estimate the recovered and confirmed cases of COVID-19 in the world by Maleki et al. (2020). The developed model was established based on fitting the model parameters to the reported data of recovered and confirmed cases. The model succeeded in estimating the recovered and confirmed COVID-19 cases with reasonable accuracy as the mean relative percentage error was 1.6% and 0.22% for the recovered and confirmed cases, respectively. A hybrid statistical approach composed of an autoregressive integrated moving average model (ARIMA) integrated with a wavelet model was developed to produce short term forecasting for the COVID-19 daily confirmed cases for five countries, namely, Canada, the United Kingdom, South Korea, France, and India (Chakraborty and Ghosh, 2020). The prediction performance of the proposed model was compared with that of the moving average model and wavelet model. The proposed model had outperformance over the other two models for all investigated countries. RMSE ranged between 50.83–710.46, 68.38–740.06, and 55.25–631.91 for the moving average model, wavelet-based forecasting model, and the hybrid model, respectively. Another hybrid approach composed of an ARIMA incorporated with an additive regression model, generalized autoregressive conditional heteroskedastic, and Holt-Winters exponential smoothing model was developed by Abdulmajeed et al. (2020) to forecast the number of COVID-19 cases in Nigeria. A hybrid model composed of a wavelet decomposition model integrated with and ARIMA was developed to forecast COVID-19 confirmed cases in five different countries, the United Kingdom, the United States, France, Italy and Spain (Singh et al., 2020). The accuracy of the forecasted results obtained by the hybrid model is better than that of the autoregressive integrated moving average model by about 50% in cases of France and the United States and 80% Spain, Italy, and the United Kingdom. In another study, an ARIMA was incorporated with the Alpha-Sutte indicator for forecasting the COVID-19 cases in Spain (Ahmar and del Val, 2020). The proposed hybrid approach had better accuracy compared with a conventional autoregressive integrated moving average model based on different forecasting accuracy measures. Zhao et al. (2020b) proposed a hybrid forecasting approach composed of Metropolis-Hastings parameter estimation algorithm incorporated with Susceptible Exposed Infectious Recovered model, to predict the prevalence of the COVID-19 epidemic in several African countries, namely, Algeria, Egypt, Senegal, Nigeria, Kenya, and South Africa. Santosh (2020) proposed an AI-Driven model to improve healthcare applications to deal with coronavirus outbreak. This model depends on active learning (Bouguelia et al., 2018) and cross-population train/test models.

Apart from statistical-based modeling methods, artificial intelligence-based techniques such as artificial neural networks (ANN) have been proposed as a robust predictive tool to model different engineering systems (Shehabeldeen et al., 2020, 2019; Essa et al., 2020; Babikir et al., 2019; Elaziz et al., 2019a). ANN has a number of advantages over other traditional modeling approaches such as generalization capabilities, handling massive amounts of data, required less conventional statistical training, identifying complex relationships between independent and dependent variables, and identifying the interactions between different variables. Chimmula and Zhang (2020) developed a long short-term memory approach as a well-known artificial recurrent neural network to forecast the COVID-19 outbreak in Canada. The trends of the outbreak, as well as possible stopping time of it, are also predicted. Saba and Elsheikh (2020) employed nonlinear autoregressive artificial neural networks to forecast the outbreak of the COVID-19 epidemic in Egypt. The number of confirmed COVID-19 cases was estimated to be triplicated during May 2020. The results of this model were compared with that of the autoregressive integrated moving average, and the former had a better reasonable absolute percentage error (less than 5%) for all investigated cases. A new adaptive neuro-fuzzy inference system (ANFIS) hybrid model was proposed to forecast COVID-19 confirmed cases in China (Al-qaness et al., 2020a). Flower pollination optimization algorithm integrated with the salp swarm optimization algorithm was employed to enhance the accuracy of the model via determining the optimal values of the model parameters. Another attempt to improve the forecasting accuracy of the ANFIS via integration with a new metaheuristic optimization technique called marine predators optimization algorithm was made by Al-qaness et al. (2020b). The developed algorithm was employed for forecasting the COVID-19 cases in Italy, South Korea, the United States and Iran.

In this study, an improved ANFIS model is proposed for forecasting COVID-19 confirmed cases in Brazil and Russia. The traditional ANFIS is a well-known time series prediction and forecasting method that confirmed its performance in various applications, such as Liu et al. (2017), Xidias and Zissis (2017), Alsharif and Younes (2019). However, traditional ANFIS faces some shortcomings and limitations. Therefore, in the recent decade, optimization approaches had been employed to enhance its performance, such as particle swarm optimization (PSO) (El Aziz et al., 2017), Sine cosine algorithm (SCA) (Al-Qaness et al., 2018), Multi-verse optimizer (MVO) (Al-qaness et al., 2019), whale optimizer algorithm (WOA) (Ewees and Elaziz, 2018), and a hybrid of genetic algorithm (GA) and salp swarm algorithm (SSA) (Elaziz et al., 2019b).

The MPA is a natural inspired optimization method presented recently by Faramarzi et al. (2020) based on the foraging strategies of the ocean predators depending on two procedures, called Lévy and Brownian motion, where the predators select these two strategies for optimal foraging. It has been tested using various optimization tasks, and it showed promising performance compared to other optimization approaches.

In this study, we apply a chaotic MPA to enhance ANFIS performance and avoid the shortcomings of the conventional ANFIS. The proposed CMPA-ANFIS is applied to forecast the number of COVID-19 cases in two countries, namely Russia and Brazil. The motivation of using chaotic maps to improve MPA is their proper-

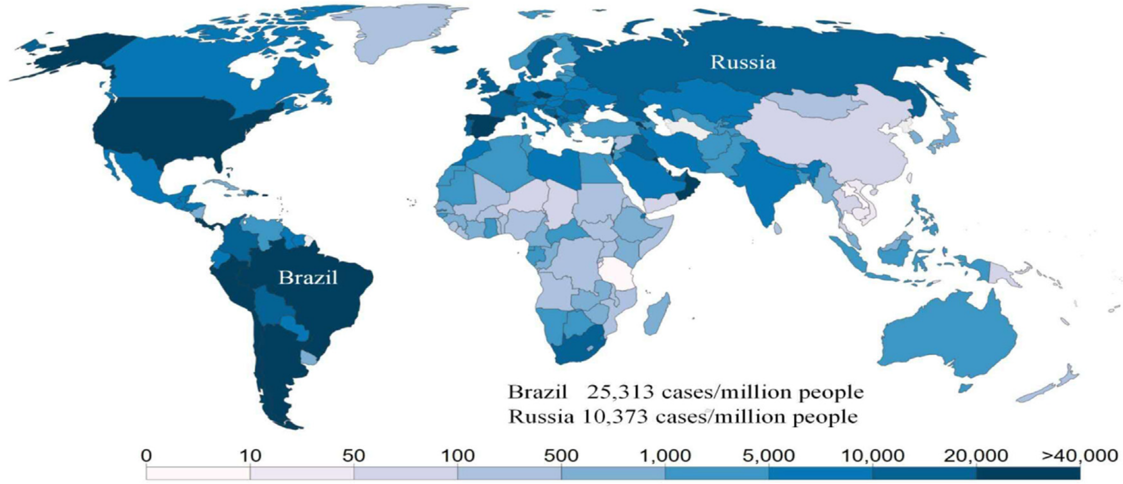

Fig. 1. Total COVID-19 confirmed cases per million people, 24 October 2020.

ties that improve the metaheuristic techniques in general and MPA in specific. The random number used to enhance the exploration ability of the MPA can lead to stuck in local optima and final degradation of the quality of the final output. However, the chaotic maps avoid this problem since they simulate the behavior, for example, dynamical properties, which give a high ability to MPA to produce a solution with suitable diversity in the search domain.

The CMPA-ANFIS for forecasting the COVID-19 dataset starts by preparing the time series of COVID-19 and split them into training and testing. Followed by generating a set of chaotic values using the Tent map and assigned them to the parameter of MPA. Then build a population of solutions that refer to the parameters of ANFIS and assess the quality of each of them using RMSE based on the training set. The best solution is determined, and the solutions are updated using the value of the chaotic map and the operators of MPA. After finishing the training process, the best ANFIS network is used, and the testing set is applied to and compute the final COVID-19 output and compute the performance measure.

Our main contribution can be listed as follows:

1. Forecast COVID-19 confirmed cases in two countries, Russia and Brazil, since these two countries have a serious situation according to the official reports in the current time. Therefore, this model may help these countries' governments to make more strict plans.
2. Present a new short term time series forecasting model based on improving the performance of the conventional ANFIS using an enhanced version of the MPA, called, CMPA.
3. Evaluate the performance of the proposed model using WHO official datasets; more so, we compared the CMPA with existing optimization models to confirm its quality and good performance.

The rest sections of this study are presented as follows. We describe the study area and the collected data in Section 2. Section 3 presents the methods applied in this study. In Section 4, we present the evaluation experiments and comparisons. lastly, in Section 5, we present our conclusion and recommendation.

2. Study areas and dataset

2.1. Study areas

In this study, two COVID-19 hotspot countries have been considered; Russia and Brazil. Russia has the largest area (17.1 million km^2) among all world countries. It is a transcontinental country located in Northern Asia and Eastern Europe. Russia lies between longitudes 19° E and 169° W, and latitudes 41° and 82° N. Russia is bounded to the east and north by the Pacific and Arctic oceans. Russia has boundaries with fourteen countries: Finland, Norway, Estonia, Belarus, Lithuania, Latvia, Georgia, Poland, Ukraine, Kazakhstan, Azerbaijan, Mongolia, North Korea, and China. It has a population of 146.7 million as of 2020, and it is considered as the most populous European country. The enormous size of Russia results in the domination of the continental climate. Russia has an extreme climate with very low temperatures. Healthcare in Russia is regulated by the Russian Ministry of Health. Expenditure on healthcare in Russia as a percentage of gross domestic product is 3.2%, which is lower than that of all other European countries. However, there are only 8 hospital beds and 4.8 doctors per one thousand of the population. The main causes of death in Russia are circulatory and neoplastic diseases.

Brazil is the fifth-largest country in the world and is the largest country in South America, with a total area of about 8.5 million km^2. Brazil lies between longitudes 28° and 74° W, and latitudes 6°N and 34°S. The population of Brazil is more than 211 million, as estimated in May 2020. It is the sixth most populous country in the world. Brazil has highly diversified climate patterns with a wide range of weather conditions. Brazil has six main climatic patterns: tropical, equatorial, oceanic, desert, semiarid, and subtropical. Brazil is bounded to the east by the Atlantic Ocean. It is also bounded by nine different countries: Argentina, Colombia, Venezuela, Uruguay, Paraguay, Peru, Bolivia, Guyana, and Suriname. Healthcare in Brazil is a basic right, according to the Brazilian constitution. Expenditure on healthcare in Brazil as a percentage of gross domestic product is 9.4%. However, there are only 1.9 hospital beds and 2.1 doctors per one thousand of the population. Infectious diseases such as hepatitis B, hepatitis C, tuberculosis, and HIV/AIDS are the main causes of death in Brazil.

Russia and Brazil are considered the two biggest hotspots of COVID-19 infection worldwide, as shown in Fig. 1. Russia has 10,373 cases/million people, while Brazil has 25,313 cases/million people, according to WHO reports on 24 October 2020. Brazil and Russia have the third and fourth ranks after the United States and India in total and daily confirmed COVID-19 cases. Brazil has 5,380,635 total cases and 25,380 daily cases. Russia has 1,487,260 total cases and 7015 daily cases. The reported daily and total cases, as well as the number of deaths as reported by WHO on 25 October 2020 for the most infected countries, are plotted in Fig. 2.

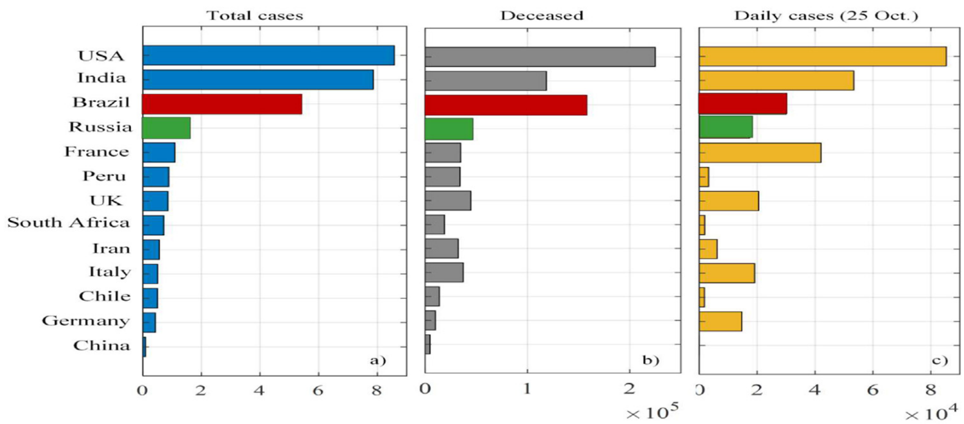

Fig. 2. COVID-19 statistics on 25 October 2020: (a) Total cases; (b) Deceased; (c) Daily cases.

Due to the inefficient funding in the Russian and Brazilian health systems, the prevalence of COVID-19 pandemic may be worth among all European and Latin American countries.

2.2. Data collection

The official COVID-19 reported cases in Russia and Brazil declared by WHO for the period from 26 March to 1 June 2020 was employed to train and test the proposed methods in the current study.

3. Methods

3.1. The basics of the traditional ANFIS model

Generally, the ANFIS generates a mapping between the input and the output using "IF-THEN rules" (which is also known as the "Takagi–Sugeno inference model"). Fig. 3 presents the general structure of the ANFIS, in which the inputs of Layer 1 are represented by x and y. Where O_{1i} is the output of node i, as formulated by the following equations:

$$O_{1i} = \mu_{A_i}(x), \quad i = 1, b, \quad O_{1i} = \mu_{B_{i-2}}(y), \quad i = 3, 4 \tag{1}$$

$$\mu(x) = e^{-\left(\frac{x-\rho_i}{\alpha_i}\right)^2}; \tag{2}$$

where μ is the generalized Gaussian membership function. A_i and B_i are the membership values of μ. In addition, α_i and ρ_i are the premise parameter set.

More so, Eq. (3) illustrates the output of Layer 2:

$$O_{2i} = \mu_{A_i}(x) \times \mu_{B_{i-2}}(y) \tag{3}$$

The output of Layer 3 can be represented as Eq. (4):

$$O_{3i} = \overline{w}_i = \frac{\omega_i}{\sum_{(i=1)}^{2} \omega_i}, \tag{4}$$

where the output of ith nodes from the previous layer is represented by w_i.

Furthermore, Eq. (5) represents the output of Layer 4:

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i) \tag{5}$$

in which f represents the function which combines the parameters and inputs of the networks. r_i, q_i, and p_i represent the consequent parameters of node i.

Eq. (6) represents the output of Layer 5:

$$O_5 = \sum_i \overline{w}_i f_i \tag{6}$$

3.2. Marine predators algorithm (MPA)

This section introduces the basics of the MPA (Faramarzi et al., 2020). Like other metaheuristics, the MPA begins by creating random values based on the search space for a set of solutions, as presented by Eq. (7):

$$X = LB + r_1 \times (UB - LB) \tag{7}$$

where LB is the lower boundary, LB is the upper boundary, and $r_1 \in [0, 1]$ is a random number.

The MPA considers predator and prey as search agents because when a predator searches for prey, the prey also searches for food. Thus, at the end of each population (generation), the matrix of the fittest predators (elite matrix) is updated. Eq. (8) formulates the elite and prey (X) (Faramarzi et al., 2020):

$$Elite = \begin{bmatrix} X_{11}^1 & X_{12}^1 & \ldots & X_{1d}^1 \\ X_{21}^1 & X_{22}^1 & \ldots & X_{2d}^1 \\ \ldots & \ldots & \ldots & \ldots \\ X_{n1}^1 & X_{n2}^1 & \ldots & X_{nd}^1 \end{bmatrix}, \quad X = \begin{bmatrix} X_{11} & X_{12} & \ldots & X_{1d} \\ X_{21} & X_{22} & \ldots & X_{2d} \\ \ldots & \ldots & \ldots & \ldots \\ X_{n1} & X_{n2} & \ldots & X_{nd} \end{bmatrix}, \tag{8}$$

Thereafter, the prey X position is updated using three stages, called high-velocity ratio, unit velocity ratio, and low-velocity ratio. These stages are based on the variant ratio of velocity, with respect to rational velocity amongst the predator and prey. The following subsections describe the details of each stage.

3.2.1. High-velocity ratio

In the high-velocity ratio stage or so-called exploration phase, each predator moves faster than X, which is performed at the first third of the number of iteration (i.e., $\frac{1}{3}t_{\max}$). Thus, the prey can be updated using Eqs. (9) and (10):

$$S_i = R_B \otimes (Elite_i - R_B \otimes X_i), \quad i = 1, 2, \ldots, n \tag{9}$$

$$X_i = X_i + P.R \otimes S_i \tag{10}$$

here, $R \in [0, 1]$ is a vector of uniform random numbers, and $P = 0.5$ is a constant number. The Brownian motion is represented by the random vector R_B. Furthermore, $\otimes$ refers to the element-wise multiplication process.

3.2.2. Unit velocity ratio

In this phase, the predator and prey move in the same space, and these movements simulate the processes of searching for prey and food. Therefore, this refers to changing the status of the marine

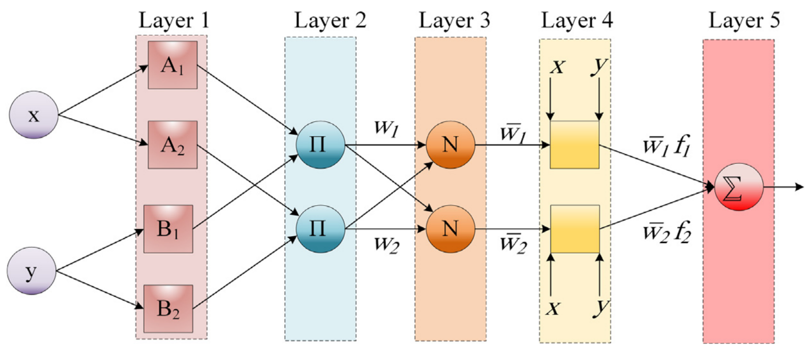

Fig. 3. The ANFIS model structure.

predator algorithm from the exploration phase to exploitation phase. In this stage, both of exploration to exploitation have the same chance to occur. According to Faramarzi et al. (2020), the exploration is implemented using the predator, where the exploitation is implemented by the prey. More so, the Lévy flight represents the prey movement, where Brownian motion represents the predator, as presented in Eqs. (11) and (12), when $\frac{1}{3}t_{\max} < t < \frac{2}{3}t_{\max}$:

$$S_i = R_L \otimes (Elite_i - R_L \otimes X_i), \quad i = 1, 2, \ldots, n \tag{11}$$

$$X_i = X_i + P.R \otimes S_i \tag{12}$$

in which R_L contains random numbers which follow the Lévy distributions. Eqs. (11) and (12) are employed to the front half of the generation, which depicts the exploitation. Where the next half of the generation implements the following equations:

$$S_i = R_B \otimes (R_B \otimes Elite_i - X_i), \quad i = 1, 2, \ldots, n \tag{13}$$

$$X_i = X_i + P.CF \otimes S_i, \quad CF = (1 - \frac{t}{t_{\max}})^{2\frac{t}{t_{\max}})} \tag{14}$$

CF represents a parameter which controls the step size of movements of the predator, where t_{max} is the total number of iterations (generations).

3.2.3. Low-velocity ratio

This stage is the latest optimization process that occurs when the predator's movement is faster than the prey's movement. It represents the exploitation phase where $t > \frac{2}{3}t_{\max}$, as presented by Eq. (13):

$$S_i = R_L \otimes (R_L \otimes Elite_i - X_i), \quad i = 1, 2, \ldots, n \tag{15}$$

$$X_i = X_i + P.CF \otimes S_i, \quad CF = (1 - \frac{t}{t_{\max}})^{2\frac{t}{t_{\max}})} \tag{16}$$

3.2.4. Eddy formation and FADs' effect

Fish aggregating devices (FADs) is one of the environmental factors that may affect the behaviors of the marine predators; therefore, Faramarzi et al. (2020) formulated the effect of FADs as (17):

$$X_i = \begin{cases} X_i + CF[X_{\min} + R \otimes (X_{\max} - X_{\min}) \otimes U & r_5 < FAD \\ X_i + [FAD(1-r) + r](X_{r1} - X_{r2}) & r_5 > FAD \end{cases} \tag{17}$$

Where $FAD = 0.2$, and U are binary solutions, which can be implemented by creating a random solution, which is converted into binary solutions by using threshold 0.2. More so, $r \in [0, 1]$ is a random number, where r_1 and r_2 represent the prey indices.

3.2.5. Marine memory

According to Faramarzi et al. (2020), the memory of the marine predator can remember the good position that it reached. This is performed by comparing the fitness values of each solution with previous fitness values. Then, the memory can save the best fitness value.

4. The proposed CMPA-ANFIS forecasting model

In this section, the developed forecasting COVID-19 model is introduced. The proposed model aims to enhance the prediction of the ANFIS model by estimating its parameters using modified MPA based on the chaotic tent map (CTM). The use of CTM aims to enhance the convergence of the MPA algorithm during the exploration phase, and this is achieved by replacing random number R (in Eqs. (10) and (12)). Then, the improved MPA, called CMPA, is employed to determine the suitable parameters of ANFIS. The full steps of the developed CMPA-ANFIS are discussed in the following.

4.1. Implementation of CMPA-ANFIS model

The CMPA-ANFIS begins by receiving the historical time series of the confirmed cases of COVID-19 (for either Russia or Brazil), then using the Autocorrelation function (ACF) to determine the patterns of the COVID-19 dataset in which the 8-lags is used. The next step is to divide the historical data into a training set (here 75% from data) and a testing set (25%). Followed by determining the parameters of MPA. Then, using the CTM to generate chaotic value as defined in the following equation:

$$R_{k+1} = \begin{cases} \dfrac{R_k}{0.7}, & R_k \prec 0.7 \\ \dfrac{10}{3}(1 - R_k), & R_k \geq 0.7 \end{cases} \tag{18}$$

The next process is to generate a set of N solutions, representing the configuration of ANFIS parameters. Each solution is evaluated by constructing the ANFIS network according to its value and applied the training set to the constructed ANFIS. Then predict the output and compute the Root mean square error (RMSE), which is applied as fitness value, as in Eq. (19):

$$\text{RMSE} = \sqrt{\frac{1}{N_s}\sum_{i=1}^{N_s}(T_i - P_i)^2} \tag{19}$$

in which T and P are the original target, and the predicted output, respectively. N_s refers to the size of the sample.

After that, the best configuration/solution is determined, which has the best fitness value. Then the X solutions will be updated using the operators of CMPA. This is performed by using the value of R that generated using Eq. (18), and replace the random value in Eqs. (10) and (12) during the updating process of MPA. The updating steps are performed till reaching the main stop condition.

Lastly, the testing set is applied to the best configuration of ANFIS parameters and computes the predicted value of the testing set.

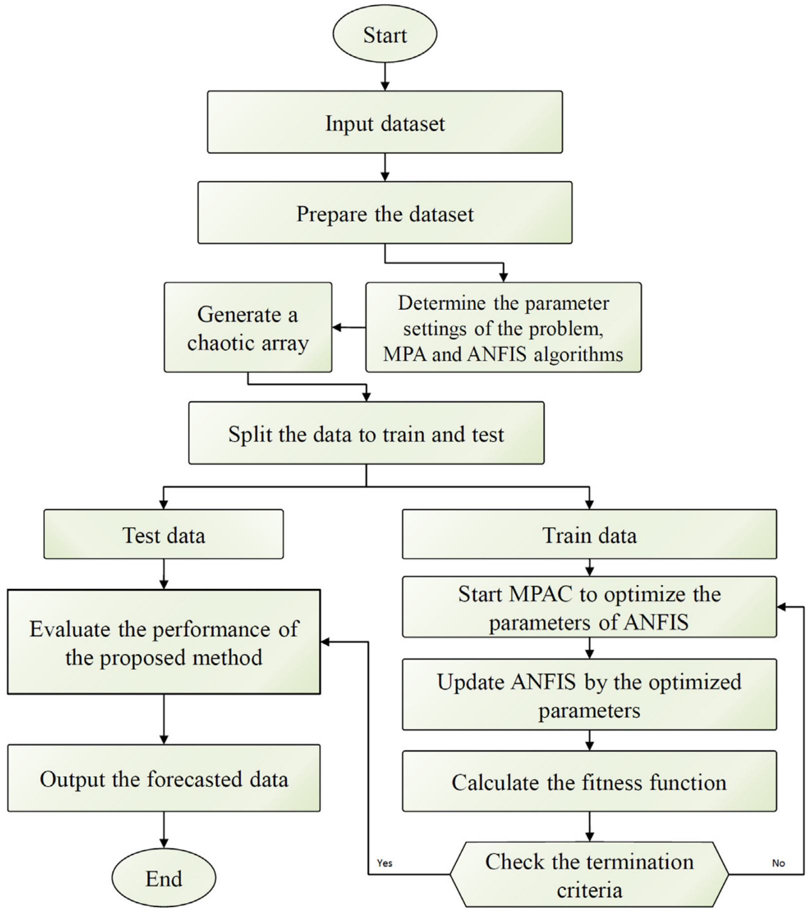

Fig. 4. The proposed CMPA-ANFIS forecasting COVDI-19 model.

Followed by evaluating the performance of this predicted COVID-19 value using different measures. The structure of the CMPA-ANFIS is depicted in Fig. 4.

5. Results and discussion

The reported total COVID-19 confirmed cases in Russia and Brazil for 215 days (26 March–26 October 2020) have been used to train the four proposed models, namely, ANFIS, PSO, MPA, and CMPA. Once the models are trained, they may be used to forecast the cases of further days ahead. First, all models have been trained and tested using the time series data of the reported cases. The reported data by WHO for 151 days was used as a time series data to train the models. 70% of the collected data was used to train the models, and 30% was used to test them to ensure their validity as a forecasting tool. The predicted data during the test stage obtained by all models are compared with the real reported data by WHO. Four statistical criteria have been employed to evaluate the performance of the investigated models: root mean square relative error (RMSRE), mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) (Elsheikh et al., 2019).

The relative percentage errors between the predicted results and of total cases in Russia and Brazil are plotted in Fig. 5. Obviously, CMPA predicted results have less relative percentage error than that of ANFIS, PSO, and CMPA, which indicate the outperformance of CMPA over other investigated models. The outperformance of CMPA over other investigated models may also be observed from qq plots shown in Figs. 6 and 7 for Russia and Brazil, respectively.

Table 1
Statistical evaluation of the developed models.

Country	Model	RMSE	MAE	MAPE	RMSRE	Time
Brazil	ANFIS	24,100	18,272	0.3973	0.0052	–
	PSO	21,182	15,111	0.3293	0.0045	**18.92**
	MPA	21,953	16,658	0.3628	0.0047	35.12
	CMPA	**19,432**	**14,273**	**0.3117**	**0.0042**	34.57
Russia	ANFIS	683	578	0.05041	0.00058	–
	PSO	504	387	0.03286	0.00041	**19.40**
	MPA	515	416	0.03561	0.00042	35.21
	CMPA	**493**	**379**	**0.03223**	**0.00040**	34.43

Bold as in almost published papers is indicating the best acheived results.

It is observed that CMPA predicted results have a better fitting during training and test processes compared with that of other investigated models for both investigated countries.

To compare between the investigated models quantitatively, four statistical criteria (RMSE, MAE, MAPE, and RMSRE) were computed for all predicted results during training and test processes. The computed values of these statistical criteria are plotted in Fig. 8 and tabulated in Table 1. Among all investigated models, CMPA has the lowest RMSE, MAE, MAPE, and RMSRE values of 833, 667, 0.22, and 0.0024 in the case of Russia and 1407, 1073, 0.30, and 0.004 in case of Brazil. ANFIS has the highest RMSE, MAE, MAPE, and RMSRE values of 4029, 3952, 1.18, and 0.0120 in the case of Russia and 2583, 1880, 0.51, and 0.007 in case of Brazil. The low values of RMSE, MAE, MAPE, and RMSRE show the high accuracy of the proposed model, while the high values of them indicate the low accuracy of the model. Therefore, CMPA has the best accuracy among all models.

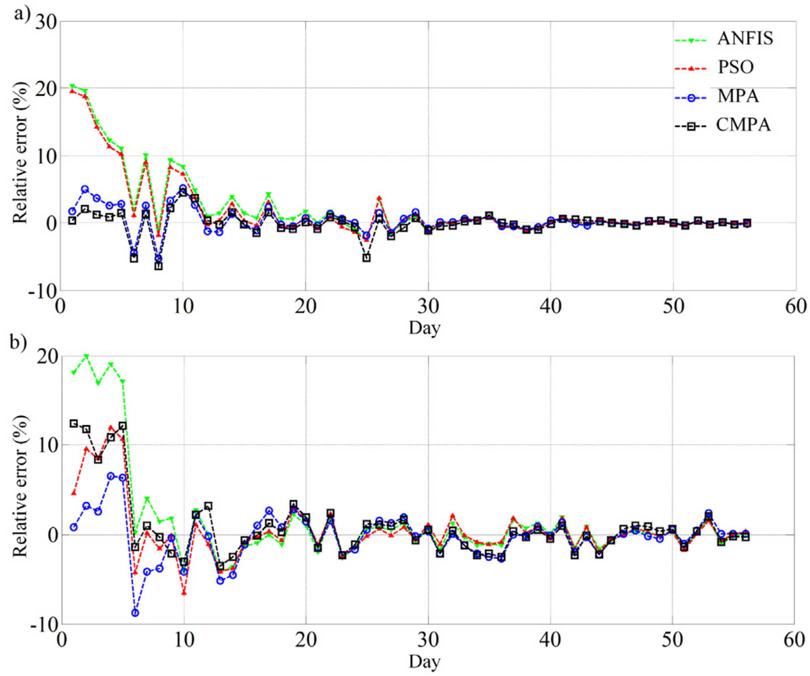

Fig. 5. Relative percentage error between the predicted and real data for the four models: (a) Russia; (b) Brazil.

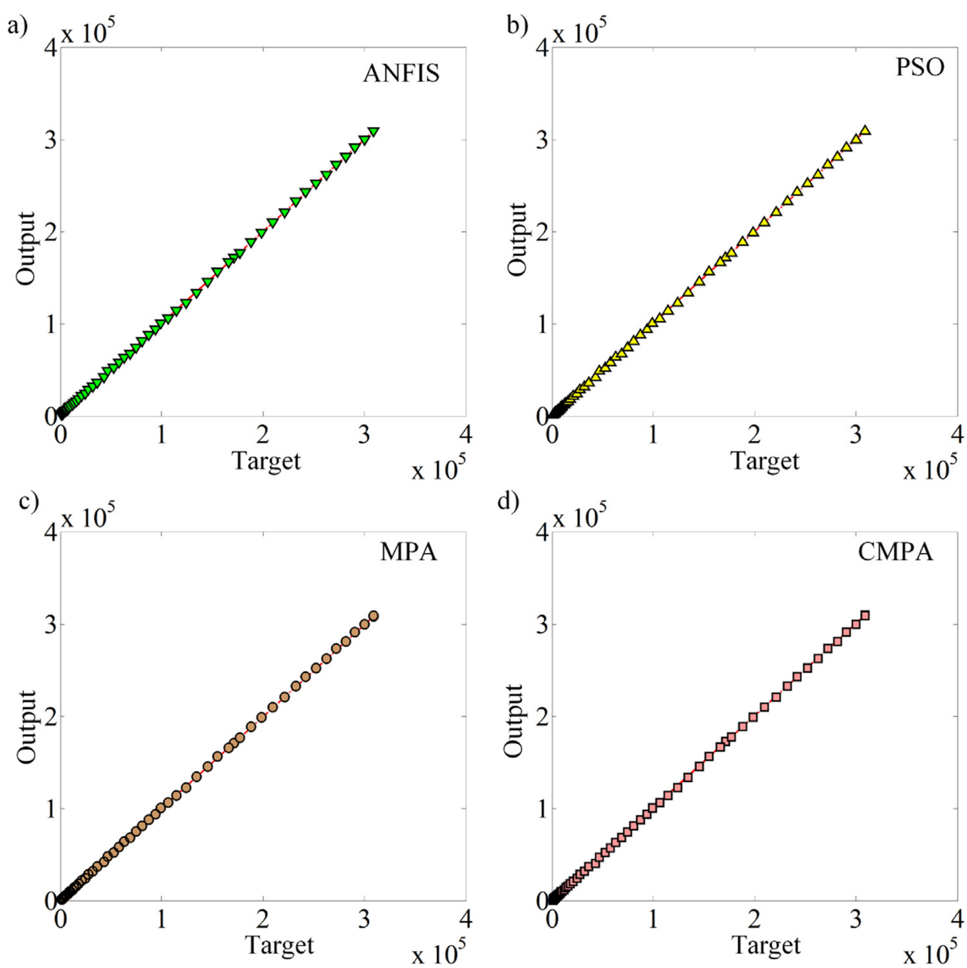

Fig. 6. Fitting of total confirmed cases in Russia data using: (a) ANFIS; (b) PSO; (c) MPA; (d) CMPA.

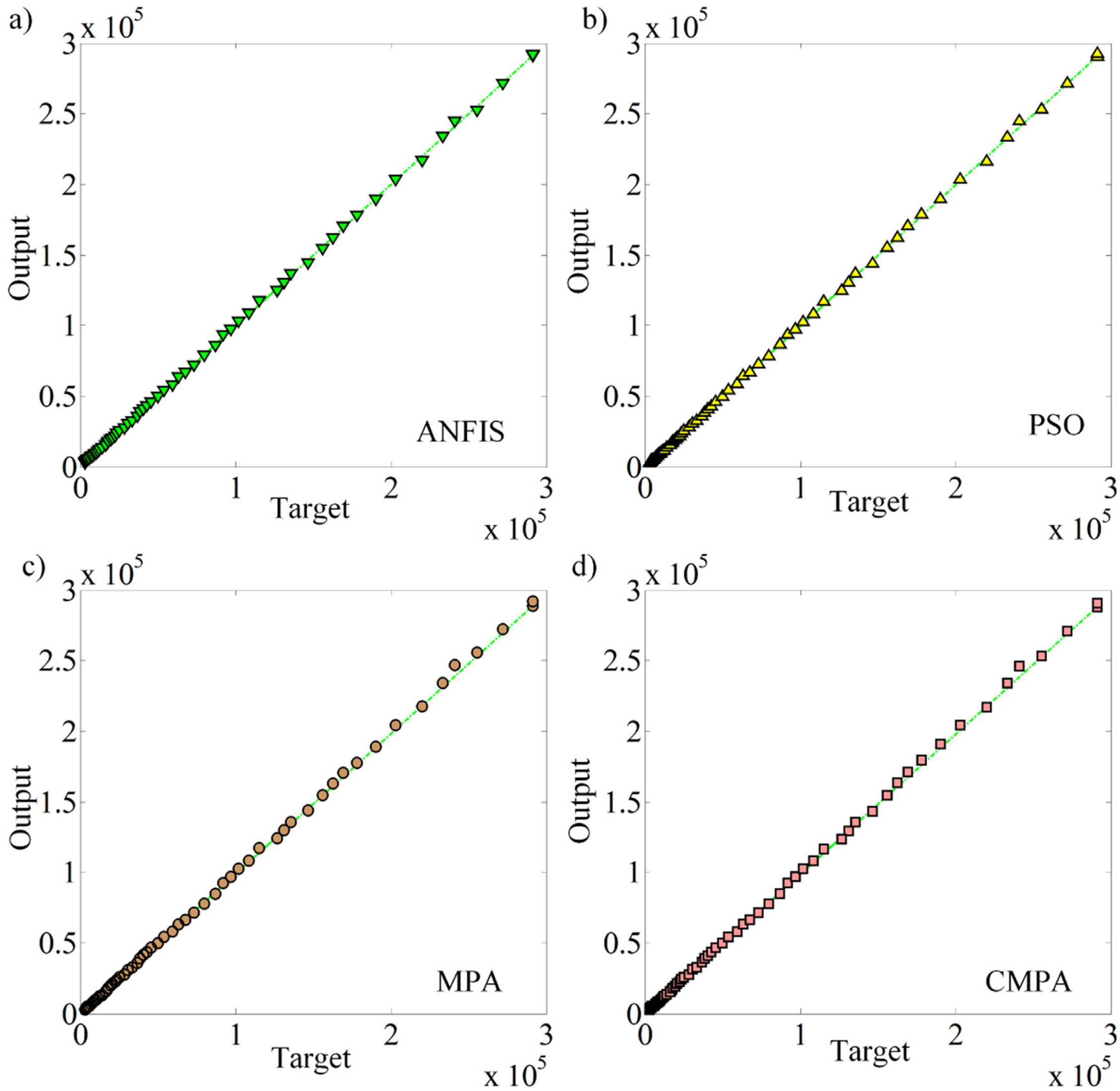

Fig. 7. Fitting of total confirmed cases in Russia data using: (a) ANFIS; (b) PSO; (c) MPA; (d) CMPA.

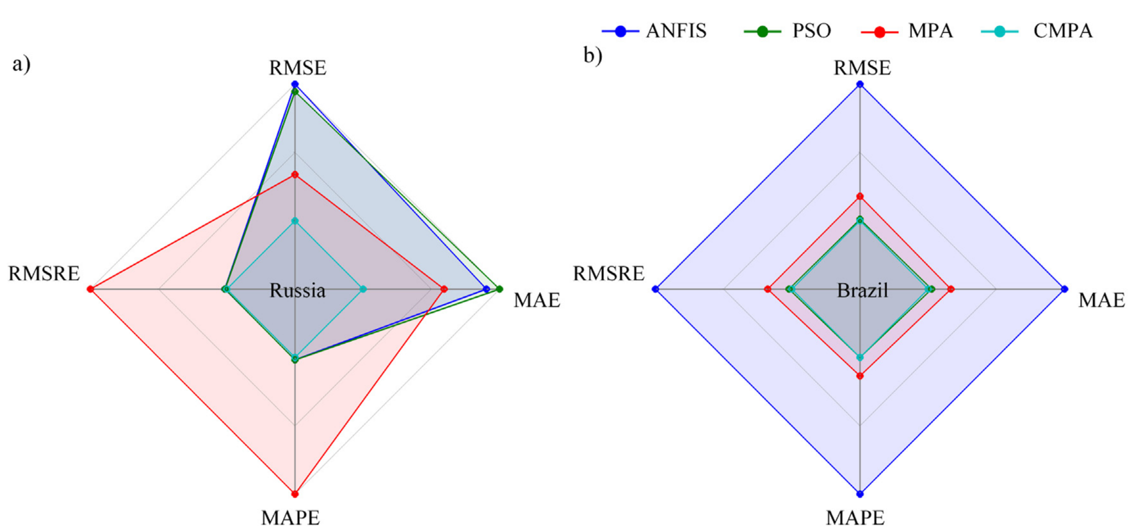

Fig. 8. Assessment criteria of different algorithms for training and test process: (a) Russia results; (b) Brazil results.

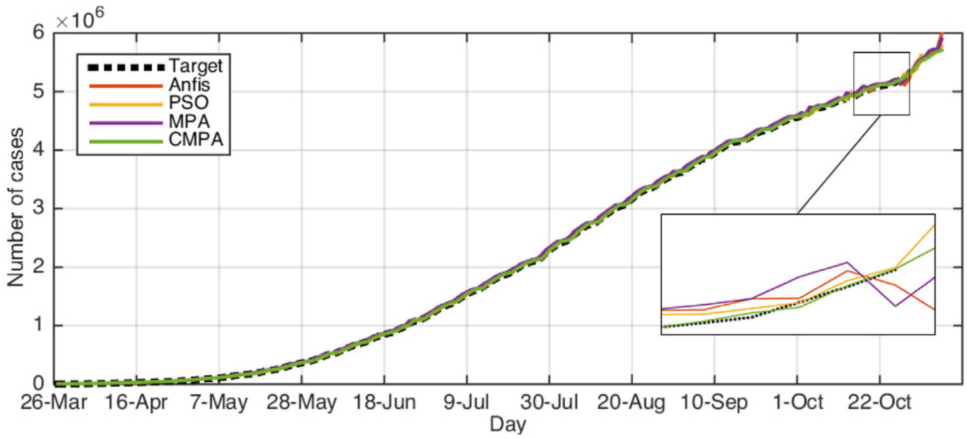

Fig. 9. Forecasted results against the real data for Brazil.

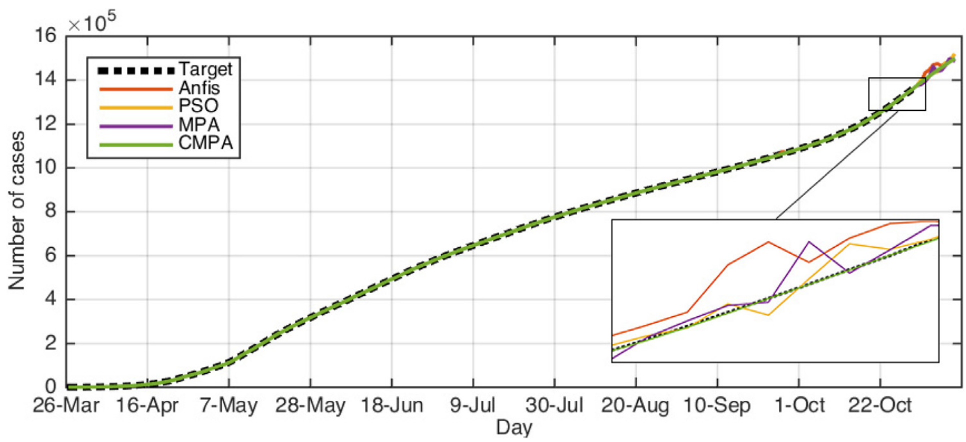

Fig. 10. Forecasted results against the real data for Russia.

Moreover, the proposed CMPA, as a modified version of MPA, has a better accuracy of the conventional one. From the aforementioned discussion, it can be concluded that CMPA has better accuracy in predicting the total cases compared with other models. That is due to the important role of the application of chaotic search in the conventional MPA algorithm.

Fig. 11 depicts the average of all measures for both countries to show the improvement of the proposed method compared to the other methods. From Fig. 11, we can see that the MPA obtained the smallest and best values in all measures followed by the PSO. It also shows that the adding of the chaotic map improves the results of the CMPA compared to its original version (i.e., MPA).

In the aforementioned discussion, the predictive capabilities of the proposed models have been evaluated. Now, we will apply these models to forecast the cases of further days ahead. The forecasting processes have been employed as follows:

- The models were applied to forecast the confirmed cases for 12 days ahead (27 October–7 November) in Russia and Brazil using the experience that the model gained during the training process using the training data from 26 March to 23 August, and testing data from 24 August to 26 October. The forecasted results showed a good agreement with the reported cases by WHO. For the forecasting process, CMPA has the lowest RMSE, MAE, MAPE, and RMSRE values of 493, 379, 0.03223, and 0.0004 in the case of Russia and 19,432, 14,273, 0.3117, and 0.0042 in case of Brazil. These statistical results indicate the better accuracy of CMPA in forecasting the total COVID-19 cases compared with other models.
- Once the outperforms of the proposed CMPA model is verified, it could be applied for forecasting the total COVID-19 cases for longer periods. Thus, the CMPA was applied to forecast the total confirmed cases of COVID-19 in Russia and Brazil for 12 days (27 October to 7 November). The forecasted results are presented in Figs. 9 and 10, respectively. These forecasted results could help decision-makers in Russia and Brazil to modify their policies to face this epidemic.

From the results, it can be concluded that the CMPA has a high ability to forecast the total COVID-19 cases in two hotspot countries, whereas it outperformed the other algorithms in the experiment measures. The main reason for the best results of the CMPA is due to the use of the chaos theory properties such as stochastically intrinsic, ergodicity, and sensitive to initial conditions to enhance the exploration phase of the MPA, which increased its ability to balance between exploration and exploitation phases as well as saved it from getting trap in local optima. In addition, the fast convergence of the CMPA and the simplicity of implementation encourage us to apply it to solve different optimization problems. Although the computation time of the CMPA is slightly better than the original version of MPA, it needs to be improved to be faster such as the PSO algorithm. However, in general, the current computation time of the CMPA is acceptable to work with offline optimization problems.

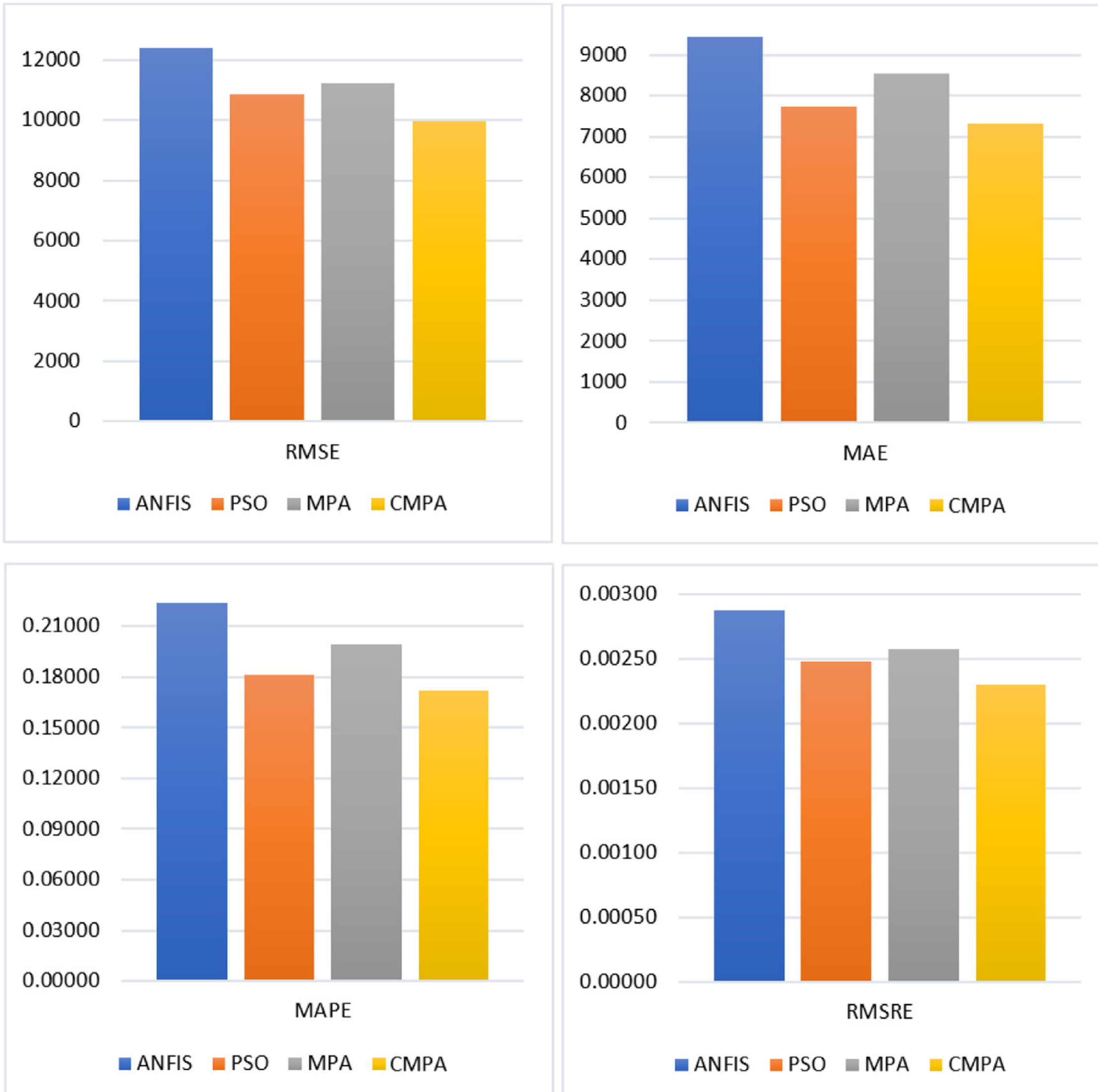

Fig. 11. Average of all measures for both countries.

6. Conclusion

In this paper, four artificial intelligence-based models are employed for forecasting the total COVID-19 cases in two hotspot countries, Russia and Brazil. The reported cases by WHO were utilized as time-series data to train the models. The forecasted results of CMPA have a good accuracy based on different statistical assessment criteria; it has RMSE, MAE, MAPE, and RMSRE of 1407, 1073, 0.3, and 0.004 for Brazil and 833, 667, 0.22, and 0.0024 for Russia, respectively. COVID-19 total cases are expected to be increased by about 40.42% and 17.23% for Brazil and Russia, respectively, during the upcoming two weeks. Thus, the proposed CMPA achieved better performance than the original ANFIS, original MPA, and PSO in all performance measures. The application of chaotic maps improved the performance of the original MPA and enhanced its exploration and exploitation phases. However, the proposed CMPA has a limitation in the computational time, which is not faster than the PSO, since PSO achieved shorter computation time in the experimental tests. Therefore, the proposed models are efficient tools to forecast the total infected people with many days in advance. This will help decision-makers to develop their plans to face this pandemic. The following recommendations should be considered: (1) Instituting total lockdown with more restrictions. (2) Increasing health system capacities. (3) Building more quarantine hospitals. (4) Providing the health care workers with adequate personal protective equipment. (5) Providing better monitoring strategies and COVID tests to the health care workers. (6) Establishing public hand hygiene stations. (7) Preventing social gatherings. (8) Prohibiting of transportation between provinces.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

This work is supported by the Hubei Provincinal Science and Technology Major Project of China under Grant No. 2020AEA011 and the Key Research & Developement Plan of Hubei Province of China under Grant No. 2020BAB100.

References

Abdulmajeed, K., Adeleke, M., Popoola, L., 2020. Online forecasting of covid-19 cases in Nigeria using limited data. Data Brief, 105683.

Ahmar, A.S., del Val, E.B., 2020. Suttearima: short-term forecasting method, a case: Covid-19 and stock market in spain. Sci. Total Environ., 138883.

Al-Qaness, M.A., Elaziz, M.A., Ewees, A.A., 2018. Oil consumption forecasting using optimized adaptive neuro-fuzzy inference system based on sine cosine algorithm. IEEE Access 6, 68394–68402.

Al-qaness, M.A., Abd Elaziz, M., Ewees, A.A., Cui, X., 2019. A modified adaptive neuro-fuzzy inference system using multi-verse optimizer algorithm for oil consumption forecasting. Electronics 8, 1071.

Al-qaness, M.A., Ewees, A.A., Fan, H., Abd El Aziz, M., 2020a. Optimization method for forecasting confirmed cases of covid-19 in china. J. Clin. Med. 9, 674.

Al-qaness, M.A., Ewees, A.A., Fan, H., Abualigah, L., Abd Elaziz, M., 2020b. Marine predators algorithm for forecasting confirmed cases of covid-19 in Italy, USA, Iran and Korea. Int. J. Environ. Res. Public Health 17, 3520.

Alsharif, M.H., Younes, M.K., 2019. Evaluation and forecasting of solar radiation using time series adaptive neuro-fuzzy inference system: Seoul city as a case study. IET Renew. Power Gener. 13, 1711–1723.

Babikir, H.A., Elaziz, M.A., Elsheikh, A.H., Showaib, E.A., Elhadary, M., Wu, D., Liu, Y., 2019. Noise prediction of axial piston pump based on different valve materials using a modified artificial neural network model. Alex. Eng. J. 58, 1077–1087.

Boccaletti, S., Ditto, W., Mindlin, G., Atangana, A., 2020. Modeling and forecasting of epidemic spreading: the case of covid-19 and beyond. Chaos Solitons Fractals.

Bouguelia, M.-R., Nowaczyk, S., Santosh, K., Verikas, A., 2018. Agreeing to disagree: active learning with noisy labels without crowdsourcing. Int. J. Mach. Learn. Cybern. 9, 1307–1319.

Cao, D., Yin, H., Chen, J., Tang, F., Peng, M., Li, R., Xie, H., Wei, X., Zhao, Y., Sun, G., 2020. Clinical analysis of ten pregnant women with covid-19 in wuhan, china: a retrospective study. Int. J. Infect. Dis.

Chakraborty, T., Ghosh, I., 2020. Real-time forecasts and risk assessment of novel coronavirus (covid-19) cases: a data-driven analysis. Chaos Solitons Fractals, 109850.

Chimmula, V.K.R., Zhang, L., 2020. Time series forecasting of covid-19 transmission in Canada using LSTM networks, Chaos. Solitons Fractals, 109864.

Chintalapudi, N., Battineni, G., Amenta, F., 2020. Covid-19 disease outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: a data driven model approach. J. Microbiol. Immunol. Infect.

El Aziz, M.A., Hemdan, A.M., Ewees, A.A., Elhoseny, M., Shehab, A., Hassanien, A.E., Xiong, S., 2017. Prediction of biochar yield using adaptive neuro-fuzzy inference system with particle swarm optimization. 2017 IEEE PES PowerAfrica, 115–120.

Elaziz, M.A., Elsheikh, A.H., Sharshir, S.W., 2019a. Improved prediction of oscillatory heat transfer coefficient for a thermoacoustic heat exchanger using modified adaptive neuro-fuzzy inference system. Int. J. Refrig. 102, 47–54.

Elaziz, M.A., Ewees, A.A., Alameer, Z., 2019b. Improving adaptive neuro-fuzzy inference system based on a modified salp swarm algorithm using genetic algorithm to forecast crude oil price. Nat. Resour. Res., 1–16.

Elsheikh, A.H., Sharshir, S.W., Elaziz, M.A., Kabeel, A., Guilan, W., Haiou, Z., 2019. Modeling of solar energy systems using artificial neural network: a comprehensive review. Sol. Energy 180, 622–639.

Essa, F., Abd Elaziz, M., Elsheikh, A.H., 2020. An enhanced productivity prediction model of active solar still using artificial neural network and Harris Hawks optimizer. Appl. Therm. Eng. 170, 115020.

Ewees, A.A., Elaziz, M.A., 2018. Improved adaptive neuro-fuzzy inference system using gray wolf optimization: a case study in predicting biochar yield. J. Intell. Syst. 29, 924–940.

Faramarzi, A., Heidarinejad, M., Mirjalili, S., Gandomi, A.H., 2020. Marine predators algorithm: a nature-inspired metaheuristic. Expert Syst. Appl., 113377.

Kang, D., Choi, H., Kim, J.-H., Choi, J., 2020. Spatial epidemic dynamics of the covid-19 outbreak in china. Int. J. Infect. Dis.

Liu, J., Wang, X., Lu, Y., 2017. A novel hybrid methodology for short-term wind power forecasting based on adaptive neuro-fuzzy inference system. Renew. Energy 103, 620–629.

Maleki, M., Mahmoudi, M.R., Wraith, D., Pho, K.-H., 2020. Time series modelling to forecast the confirmed and recovered cases of covid-19. Travel Med. Infect. Dis., 101742.

Mi, Y.-N., Huang, T.-T., Zhang, J.-X., Qin, Q., Gong, Y.-X., Liu, S.-Y., Xue, H.-M., Ning, C.-H., Cao, L., Cao, Y.-X., 2020. Estimating instant case fatality rate of covid-19 in china. Int. J. Infect. Dis.

Nishiura, H., Kobayashi, T., Yang, Y., Hayashi, K., Miyama, T., Kinoshita, R., Linton, N.M., Jung, S.-M., Yuan, B., Suzuki, A., et al., 2020. The Rate of Underascertainment of Novel Coronavirus (2019-ncov) Infection: Estimation Using Japanese Passengers Data on Evacuation Flights.

Saba, A.I., Elsheikh, A.H., 2020. Forecasting the prevalence of covid-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. Process Saf. Environ. Prot.

Santosh, K., 2020. Ai-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data. J. Med. Syst. 44, 1–5.

Shehabeldeen, T.A., Elaziz, M.A., Elsheikh, A.H., Zhou, J., 2019. Modeling of friction stir welding process using adaptive neuro-fuzzy inference system integrated with Harris Hawks optimizer. J. Mater. Res. Technol. 8, 5882–5892.

Shehabeldeen, T.A., Abd Elaziz, M., Elsheikh, A.H., Hassan, O.F., Yin, Y., Ji, X., Shen, X., Zhou, J., 2020. A novel method for predicting tensile strength of friction stir welded aa6061 aluminium alloy joints based on hybrid random vector functional link and Henry gas solubility optimization. IEEE Access.

Singh, S., Parmar, K.S., Kumar, J., Makkhan, S.J.S., 2020. Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (arima) models in application to one month forecast the casualties cases of covid-19. Chaos Solitons Fractals, 109866.

Wu, Y., Jing, W., Liu, J., Ma, Q., Yuan, J., Wang, Y., Du, M., Liu, M., 2020. Effects of temperature and humidity on the daily new cases and new deaths of covid-19 in 166 countries. Sci. Total Environ., 139051.

Xidias, E.K., Zissis, D., 2017. Adaptive neuro fuzzy inference system for vessel position forecasting. 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC), 1071–1075.

Xu, H., Yan, C., Fu, Q., Xiao, K., Yu, Y., Han, D., Wang, W., Cheng, J., 2020. Possible environmental effects on the spread of covid-19 in china. Sci. Total Environ., 139211.

Ye, F., Xu, S., Rong, Z., Xu, R., Liu, X., Deng, P., Liu, H., Xu, X., 2020. Delivery of infection from asymptomatic carriers of covid-19 in a familial cluster. Int. J. Infect. Dis.

Zhao, S., Musa, S.S., Lin, Q., Ran, J., Yang, G., Wang, W., Lou, Y., Yang, L., Gao, D., He, D., et al., 2020a. Estimating the unreported number of novel coronavirus (2019-ncov) cases in china in the first half of January 2020: a data-driven modelling analysis of the early outbreak. J. Clin. Med. 9, 388.

Zhao, Z., Li, X., Liu, F., Zhu, G., Ma, C., Wang, L., 2020b. Prediction of the covid-19 spread in African countries and implications for prevention and controls: a case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya. Sci. Total Environ., 138959.