



# A Hybrid Model based on mBA-ANFIS for COVID-19 Confirmed Cases Prediction and Forecast

Sohail Saif<sup>1</sup> · Priya Das<sup>1</sup> · Suparna Biswas<sup>1</sup>

Received: 22 August 2020 / Accepted: 5 January 2021 / Published online: 19 January 2021  
© The Institution of Engineers (India) 2021

**Abstract** In India, the first confirmed case of novel corona virus (COVID-19) was discovered on January 30, 2020. The number of confirmed cases is increasing day by day, and it crossed 21,53,010 on August 9, 2020. In this paper, a hybrid forecasting model has been proposed to determine the number of confirmed cases for upcoming 10 days based on the earlier confirmed cases found in India. The proposed model is based on adaptive neuro-fuzzy inference system (ANFIS) and mutation-based Bees Algorithm (mBA). The meta-heuristic Bees Algorithm (BA) has been modified applying 4 types of mutation, and mutation-based Bees Algorithm (mBA) is applied to enhance the performance of ANFIS by optimizing its parameters. Proposed mBA-ANFIS model has been assessed using COVID-19 outbreak dataset for India and USA, and the number of confirmed cases in the next 10 days in India has been forecasted. Proposed mBA-ANFIS model has been compared to standard ANFIS model as well as other hybrid models such as GA-ANFIS, DE-ANFIS, HS-ANFIS, TLBO-ANFIS, FF-ANFIS, PSO-ANFIS and BA-ANFIS. All these models have been implemented using Matlab 2015 with 10 iterations each. Experimental results show that the proposed model has achieved better performance in terms of Root Mean Square Error (RMSE), Mean Absolute Percentage

Error (MAPE), Mean absolute error (MAE) and Normalized Root Mean Square Error (NRMSE). It has obtained RMSE of 1280.24, MAE of 685.68, MAPE of 6.24 and NRMSE of 0.000673 for India Data. Similarly, for USA the values are 4468.72, 3082.07, 6.1, and 0.000952 for RMSE, MAE, MAPE, and NRMSE, respectively.

**Keywords** Mutation-based Bees Algorithm (mBA) · Adaptive neuro-fuzzy inference system (ANFIS) · COVID-19 · Forecast · RMSE · MAE · MAPE · NRMSE · India

## Introduction

Corona Virus Disease 2019 (COVID-19), led by severe acute respiratory syndrome coronavirus 2 (SARSCoV-2), has become global pandemic within 4 months since it was first reported in China. This virus mainly affects the respiratory system of human being, and it is also distributed among birds, bats, mice and other animals [1, 2]. In December 2019, several people were admitted to hospital due to illness related to respiratory problems and surprisingly these people have visited the seafood market in Wuhan, China [3]. In few days, 79, 355 people were affected by this virus in China by the end of February 2020. Now worldwide humanity is suffering from this deadly virus due to lack of possible cure of this virus. It can easily transmit to human body if someone comes in contact with the droplets produced through cough and sneezing by the infected person. According to the World Health Organization (WHO), a person can be infected if he/she touches the infected person and then touches his/her own eyes, nose or mouth [4].

✉ Suparna Biswas  
mailto:suparna@gmail.com

Sohail Saif  
sohailsaif7@gmail.com

Priya Das  
28priyadas@gmail.com

<sup>1</sup> Department of Computer Science & Engineering, Maulana Abul Kalam Azad University of Technology, Kolkata, WB, India

In India, the first case of COVID was reported on January 30, 2020, and it started increasing gradually. To break the chain of transmission, a complete lockdown was declared by Indian Government; as a result, it seemed to be under control till March 24, 2020, where the total number of confirmed cases and deaths was 511 and 10, respectively. First thousands of the patients had a history of overseas visit, and from them the transmission was started. But now till August 9, 2020, it has been reported 21,53,010 confirmed cases and 38,135 deaths as shown in Table 1. Primarily, most of the people who came in contact with the infected persons were not aware about the symptoms in them.

India has the lowest case fatality rate of 2.04%, and the 68.32% patients have been cured as of August 9 [5]. Around 50% of all reported cases in the country are contributed by six cities—Mumbai, Delhi, Ahmedabad, Chennai, Pune and Kolkata. Surprisingly Lakshadweep is the only region where a single case has not been reported till date. As a preventive measure countrywide lockdown of 21 days was imposed by Indian Government starting from March 24 later on April 14, the lockdown was extended till May 3 which was followed by another two-week extensions. Currently, India has the highest number of confirmed cases in Asia and third highest number of confirmed cases in world after USA and Brazil.

Since this virus is spreading very fast, we require more strict policies and plans; thus, proper techniques are needed to foresee confirmed cases in upcoming days to make proper protection strategy. Here, prediction models play a key role to understand the duration of the epidemic [6–8]. Thus a highly accurate forecasting model is very much important to forecast the number of confirmed cases in upcoming days. Despite the availability of various forecasting models, an investigation is required to improve their performance. To enhance the performance of the

prediction models, machine learning and deep learning models have been extensively used, such as support vector regression (SVR) [9] and artificial neural networks (ANN) [10]. However, these individual models suffer from overfitting and parameter optimization problem. Therefore, to overcome the shortcomings and to increase the prediction accuracy, hybrid models have been introduced. To develop hybrid models, a set of individuals are combined together to overcome the drawbacks of these models to provide better results. This paper presents a novel hybrid forecasting model to determine and foresee the number of expected cases for the next 10 days up to August 20, 2020, based on the previous confirmed case data of India. The proposed model is a hybrid combination of adaptive neuro-fuzzy inference system (ANFIS) [11] and mutation-based Bees Algorithms (mBA) called mBA-ANFIS. Here mBA has been used to optimize the performance of ANFIS by determining the best parameters. The proposed hybrid model found to be more accurate to study the spread of the virus compared to the other prediction models. This can help the government agencies for designing strategies to reduce the spread until specific antivirals or vaccines are found.

## Related Works

In this section, several research works on machine learning- or deep learning-based techniques to foresee the confirmed cases in upcoming days have been discussed.

Sujath et al. [12] proposed machine learning-based techniques to predict the spread of COVID-19 in India. Linear regression (LR), Multilayer perceptron (MLP) and Vector auto regression (VAR) have been applied on COVID-19 dataset to predict the epidemiological example of ailment and pace of COVID-19 cases in India. Main

**Table 1** Date-wise report of confirmed cases and deaths in India

Date of reporting	Confirmed cases	Deaths
09.08.2020	21,53,010	43,379
29.07.2020	15,31,669	34,193
15.07.2020	9,36,181	24,309
01.07.2020	5,85,493	1740
15.06.2020	3,32,424	9520
14.05.2020	78,003	2549
15.04.2020	11,438	377
24.03.2020	492	9
01.03.2020	3	0
30.01.2020	1	0

objective was to predict the potential pattern of COVID-19 effects in India. 80 days' Data on Number of confirmed, death and recovered cases in India have been used in Weka and Orange and LR, MLP and VAR have been employed to predict the future effects of this deadly virus in India. Experimental results show that MLP method outperforms other two methods.

Tomar and Gupta [13] used data-driven estimation methods such as long short-term memory (LSTM) and curve fitting approach to forecast the expected COVID-19 cases for the next 30 days in India. To evaluate the performance error percentage has been obtained which shows less error percentage for curve fitting approach.

Rustam et al. [14] presented a study to show the use of supervised machine learning approach linear regression (LR), least absolute shrinkage and selection (LASSO), exponential smoothing (ES) and support vector machine (SVM) to forecast the number of newly infected cases, deaths and recoveries for the next 10 days. Evaluation parameters such as  $R^2$  score,  $R^2$  adjusted, MAE, MSE, and RMSE have been used to choose the proper model. Experimental results prove that ES performs best compared to other models.

Celestine et al. [15] proposed a fine-tuned Random Forest model boosted by the AdaBoost algorithm to predict the possible outcome of the COVID-19 patients who are under treatment. Patients' location, travel, health and demographic data have been used as input dataset. The proposed model has been compared with decision tree, SVM, Gaussian Naive Bayes in terms of accuracy, F1 score, recall score and precision score. Experimental result shows that Random Forest model achieves prediction accuracy of 94% and F1 score of 0.86 which is highest among other algorithms.

Ardabili et al. [16] presented an analysis of machine learning and soft computing techniques to foresee the COVID-19 spread. Evolutionary algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Grey Wolf Optimization have been combined with multilayered perceptron (MLP) and adaptive neuro-fuzzy inference system (ANFIS). To evaluate the performance of these algorithms, RMSE and  $R^2$  have been calculated using eight mathematical models such as Logistic, Linear, Logarithmic, Quadratic, Cubic, Compound, Power and Exponential. For prediction, COVID-19 dataset of four countries has been employed.

Pinter et al. [17] proposed a hybrid machine learning method to predict COVID-19 cases in Hungary. Adaptive neuro-fuzzy inference system (ANFIS) has been combined with multilayered perceptron-imperialist competitive algorithm (MLP-ICA) to anticipate the time series of infected persons and mortality rate. Dataset contains the daily confirmed cases and deaths from March 4 to April 28,

2020. First 44 days data have been used for training purpose, and rest of the data have been used for validation. Evaluation metrics such as RMSE, MAPE and  $R^2$  have been considered in this work.

Al-qaness et al. [18] described an enhanced ANFIS model to forecast COVID-19 confirmed cases for the next 10 days in China. Flower Pollination Algorithm (FPA), Salp Swarm Algorithm (SSA) and ANFIS have been combined together to develop the hybrid model. FPASSA-ANFIS model has been assessed using COVID-19 outbreak data of China, and it is compared to other hybrid model as well. The proposed model achieved better performance in terms of RMSE, EMSRE, MAPE and  $R^2$ .

Muhammad et al. [19] presented machine learning models to anticipate the recovery of COVID-19-infected persons with the help of epidemiological dataset of South Korea. Machine learning algorithms such as support vector machine, naive Bayes, decision tree, logistic regression, random forest and K-nearest neighbor have been applied on the dataset. Experimental result shows an accuracy of 98.85%, 97.52%, 99.85%, 97.49%, 99.60%, and 98.06%, respectively, where decision tree has the highest accuracy.

Kumar and Hembram [20] proposed a model based on the logistic equation, Weibull equation and Hill equation to predict the infection rates in Italy and China. Data analysis has been done to understand the effect of environmental factors behind the spread of COVID-19. Five cities from each country where the number of infected patients is high have been considered for the experiment, and three environmental factors such as maximum temperature, relative humidity and wind speed have been considered. Experimental results show that there is a negligible relation between wind speed and humidity with the spread of this dangerous virus. Also it has been observed that higher temperatures have a little impact on the spread of this virus. Table 2 shows comparative analysis of experimental results.

Wang et al. [21] used global epidemiological data before June 16, 2020, into the logistic model to fit the cap of epidemic trend. That cap value has been used in FbProphet model which is a machine learning-based time series prediction approach. Curve fitting has been applied, and it has been observed that an estimated 14.12 million people will be infected by the late October and that will be the peak of the outbreak.

Alzahrani et al. [22] have presented a study where they have used four machine learning models such as Autoregressive Model, Moving Average, a combination of both (ARMA) and Autoregressive Integrated Moving Average (ARIMA) to forecast the number of confirmed cases in Saudi Arabia for the next 4 weeks. Experimental shows that ARIMA model performed better than other models. It can be observed from the results that there is continuous

**Table 2** Comparative analysis of experimental results

Authors	Approach	Country	Evaluation metrics			
			RMSE	MAPE (%)	MAE	R <sup>2</sup>
Rustam et al. [14]	ES	Worldwide	16,828.58	–	8867.43	0.98
Ardabili et al. [16]	MLP	Italy	191.27	–	–	0.999
	MLP	China	2318.22	–	–	0.995
	MLP	Iran	391.1	–	–	0.991
	MLP	Germany	55.52	–	–	0.999
	MLP	USA	22.1	–	–	0.999
Pinter et al. [17]	MPL-ICA	Hungary	167.88	23.15	–	0.9971
Al-qaness et al. [18]	FPASSA-ANFIS	China	5779	4.79	4271	0.9645
Elmousalami et al. [25]	SES	Worldwide	4477.72	9.68	–	–
Sahin et al. [26]	FANGBM	Italy	1223	0.9174	–	0.9993
	FANGBM	UK	796	2.813	–	0.9996
	FANGBM	USA	5767	4.8950	–	0.9996

growing of new cases with the number of 7668 cases per day and will reach 127,129 cumulative daily cases within four weeks. In [23], author also used ARIMA model to predict the expected infected patients in India for the upcoming days. To train the model, they have used the data from January 31, 2020, to March 25, 2020, and to test the model they used the data from March 26, 2020, to April 15, 2020. nonlinear autoregressive (NAR) has been compared with their model and experimental results show the R2 value for NAR and ARIMA are 0.97 and 0.95 respectively.

Another recent study [24] presented a hybrid method combining ANFIS with virus optimization algorithm (VOA) to determine the effects of climate-related factors and population density on the spread of the COVID-19 for USA. Experimental observations shows that with the increase in maximum temperature there is a little reduction in infection rate; also it can be concluded from the results that social distancing is very much effective to reduce the infection spread.

## Motivation and Contribution

Artificial Intelligence (AI) can assist us to handle the problem raised because of this pandemic. Deep learning (DL), the current type of AI, works by identifying patterns in training data. Several researches are being made to fight against this deadly virus; out of them forecasting confirmed case is emerging. Accurately predicting the number of cases in upcoming days can help the government to plan the action that needs to be taken, such as to increase the beds, in hospital, etc. Here the numbers of confirmed cases in previous dates are used to foresee the expected

confirmed cases in upcoming days. To do this, various machine learning and mathematical models are available. Though several forecasting models are present, we observed improvement can be done in terms of accuracy of the models. Generally, machine learning-based prediction models suffer from over-fitting and under-fitting problems, and as a result it reduces the accuracy [27]. Performance of machine learning model often depends on the variations in their hyperparameters, which are user-defined and model-specific parameters which control the learning process [27]. Therefore, to overcome the shortcomings and to increase the prediction accuracy hybrid models can be employed. This motivated us to propose hybrid machine learning-based enhanced adaptive neuro-fuzzy inference system (ANFIS) to foresee the number of infected patients for upcoming 10 days in India. ANFIS has been selected since it has provided good results for similar research works such as [28–31]. To train our model, 155 days historical data of confirmed COVID-19 patients in India starting from January 30, 2020, have been used. We have observed that the error percentage in prediction can be decreased if proper parameters for ANFIS are chosen; to do that, Bees Algorithm has been employed. We still found that the error percentage can be minimized if Bees Algorithm is modified. This motivated us to modify the standard Bees Algorithm by applying five types of mutation. The following are the contribution of this paper.

- In this paper, deep learning-based hybrid model which is a combination of adaptive neuro-fuzzy inference system (ANFIS) and Bees Algorithm (BA) has been proposed to predict and foresee the number of COVID-19-affected patients in India for the next 10 days.

•Meta-heuristic algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Differential Evaluation (DE), Firefly Algorithm (FA), Harmony search (HS), Teaching–Learning–Based Optimization (TLBO) and Bees Algorithm (BA) have been used to optimize the ANFIS parameters.

•Hybrid algorithms have been implemented using Matlab, and performance in terms of RMSE, MAPE, MAE and  $R^2$  has been evaluated.

•Based on the experimental result, it has been observed that combination of BA-ANFIS has the lowest error percentage compared to others.

•To minimize the error, modification on standard Bees Algorithm (BA) has been done by applying four different mutation processes and mutation-based Bees Algorithm (mBA) has been developed.

•Proposed model has been applied on the infected COVID-19 patient data of two countries, and finally number of infected patients up to August 20, 2020, has been presented.

**Preliminaries**

In this section, we have discussed about the standard ANFIS and Bees Algorithm in brief.

**Adaptive Neuro-Fuzzy Inference System (ANFIS)**

This is a type of ANN network which is based on Takagi–Sugeno fuzzy inference system. This hybrid model is a combination of fuzzy logic theory and neural networks concept. The two input and one output structure of ANFIS is represented in Fig. 1. The interference system uses IF–THEN rule to map between input and output.

If two inputs are  $x$  and  $y$  and output is  $f$ , then IF–THEN rules can be described as.

RULE 1: If  $x$  is  $A1$  and  $y$  is  $B1$ , then  $f1 = p1x + q1y + r1$ .

RULE 2: If  $x$  is  $A2$  and  $y$  is  $B2$ , then  $f2 = p2x + q2y + r2$ .

An ANFIS model consists of five layers.

Layer 1: Each node of first layer or fuzzification layer of the ANFIS is adaptive node. The output node ( $O_{1k}$ ) of this layer is represented as

$$O_{1p} = \mu A_p(x) \text{ for } p = 1, 2, O_{1p} = \mu B_{p-2}(y) \text{ for } p = 3, 4 \tag{1}$$

$$\mu(x) = e^{-\left(\frac{x-b}{c}\right)^2}$$

where  $\mu$  denotes Gaussian MF.  $\mu A_p$  and  $\mu B_{p-2}$  are fuzzy membership values, and  $b, c$  are parameters of membership function.

Layer 2: This layer uses fuzzy operator (AND) to fuzzify the input. The output node ( $O_{2p}$ ) is represented as

$$O_{2p} = \mu A_p \times \mu B_{p-2}(y)$$

Layer 3: Layer three of ANFIS is also known as normalization layer, as it normalizes the value of second layer which will transfer to the layer four. The output node ( $O_{3p}$ ) of this layer represented as

$$O_{3p} = \bar{W}_p = \frac{W_p}{\sum_{p=1}^2 W_p}$$

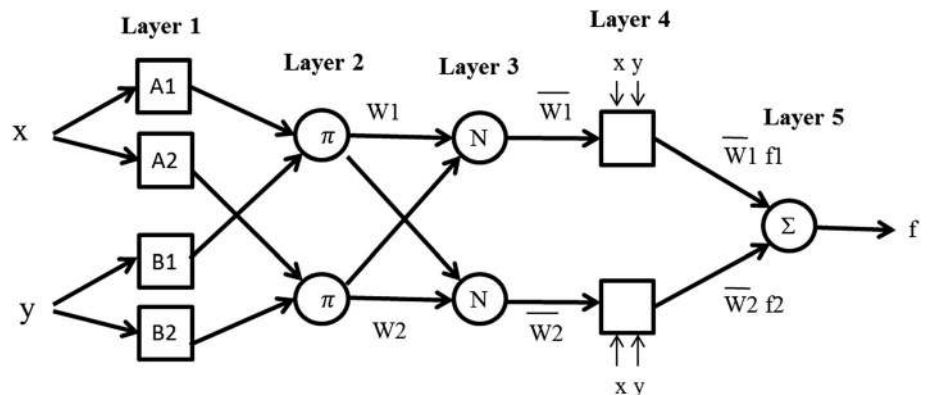
Layer 4: Each node of this layer (de-fuzzification layer) is adaptive. The output ( $O_{4p}$ ) of each node is calculated by product of output of third layer and first-order Sugeno fuzzy model.

$$O_{4p} = \bar{W}_p \times f_p \text{ where } f_p = p_p x + q_p y + r_p$$

Here  $p_p, q_p, r_p$  are the parameter sets.

Layer 5: Layer five contains only one fixed node. Overall output of the ANFIS model is represented as

**Fig. 1** Structure of ANFIS model





$$O_5 = \sum_p \overline{W_p} f_p = \frac{\sum_p \overline{W_p} f_p}{\sum_p W_p}$$

### Bees Algorithm (BA)

The Bees Algorithm (BA) was invented by Pham et al. [32] in 2005 which is a swarm-based algorithm. This algorithm intimates natural food seeking behavior of honey bees. There are mainly two types of bee, scout bee and foraging bee. Scout bees are navigators of food, and foraging bees are the collector of food. Scout bee look for good food source in flower patches. When they found good nectar, they start Waggle dance by which they communicate with other bee and send a message that they has found good food source. This algorithm has neighborhood exploitation search as well as random global exploration search capacity. The parameters of BA are: number of Scout bee ( $n$ ), number of best selected patches (*site*) out of  $n$  recently visited site ( $p$ ), number of elite site out of  $p$  recently visited region ( $e$ ), number of recruited bees for elite site ( $Be$ ), number of recruited bees for ( $p$ - $e$ ) site ( $Bs$ ) and finally the size of each patch ( $r$ ). The first step of the algorithm starts by initializing  $n$  Scout bee. Then fitness value of the swarm is calculated by predefined fitness function. Best  $p$  patches are labeled for neighborhood search. And the size of the initial patch is initialized. Two groups of top bees are recruited for vicinity search to improve exploitation. Few bees ( $n$ - $p$ ) are employed for random search to maintain diversity of this algorithm. After each iteration, patch size is reduced to get more accurate solution. After this step, the fittest bee is selected out of the swarm.

### Proposed Model

A hybrid method of deep learning and meta-heuristic algorithm has been used in this work [33–35]. Here time series method has been used to foresee the expected confirmed cases of COVID-19-infected patients. Adaptive neuro-fuzzy inference system (ANFIS) is a type of artificial neural network which is popular for time series prediction; on the other hand, Bees Algorithm (BA) is an optimization algorithm which has been employed to optimize the parameters of ANFIS. Figure 2 shows the proposed model.

For more diversification, we have modified the standard Bees Algorithm (BA) by introducing mutation process. Standard BA has two sites (elite and selected) for neighborhood search which controls exploitation of the algorithm, and it uses random search for exploration. But this exploration is not enough, and the solution may trap to local optima. To overcome this, mutation operator has been applied for more exploration. Four types of successive

mutation [36, 37] have been applied on the fittest bee, which is found after local and global search. After applying each mutation, if modified solution has better fitness, then it will replace the previous solution; otherwise, the previous solution will not be changed. The benefit of this strategy is the best solution will never be lost and also there is a chance for getting better solution. The proposed mutation-based Bees Algorithm (mBA) uses four different mutation operators [38, 39] which are as follows.

#### Cauchy Mutation

Cauchy mutation is the first stage of mutation that has been applied on the fittest bee.

$$\begin{aligned} \text{modified\_bee}(d) &= \text{bee1} \\ &+ (V \max(d) - V \min(d)) \cdot \text{Cauchy}(o, s) \text{ for } d \\ &= 1, 2, 3, \dots, n \end{aligned} \quad (7)$$

Here  $s$  represents the scale parameter of Cauchy distribution. If *modified\_bee* is better than *bee1*, then *bee1* is replaced by *modified\_bee*. This comparison has been done after each phase of mutation.

#### Opposition-based Mutation as Separate

The second phase of mutation is opposition-based mutation which has been done separately for each dimension. In Cauchy mutation, the fitness of bee has been calculated after changing its entire dimension, but for opposition-based mutation it has been calculated for each dimension.  $\text{modified\_bee}(d) = V_{\min}(d) + V_{\max}(d) - \text{bee1}$  for  $d = 1, 2, 3, \dots, n$  (8)

#### Opposition-Based Mutation as Whole

This is the third stage of mutation where mutation has applied the entire dimension of the candidate solution.

$$\text{modified\_bee} = V_{\min} + V_{\max} - \text{bee1} \quad (9)$$

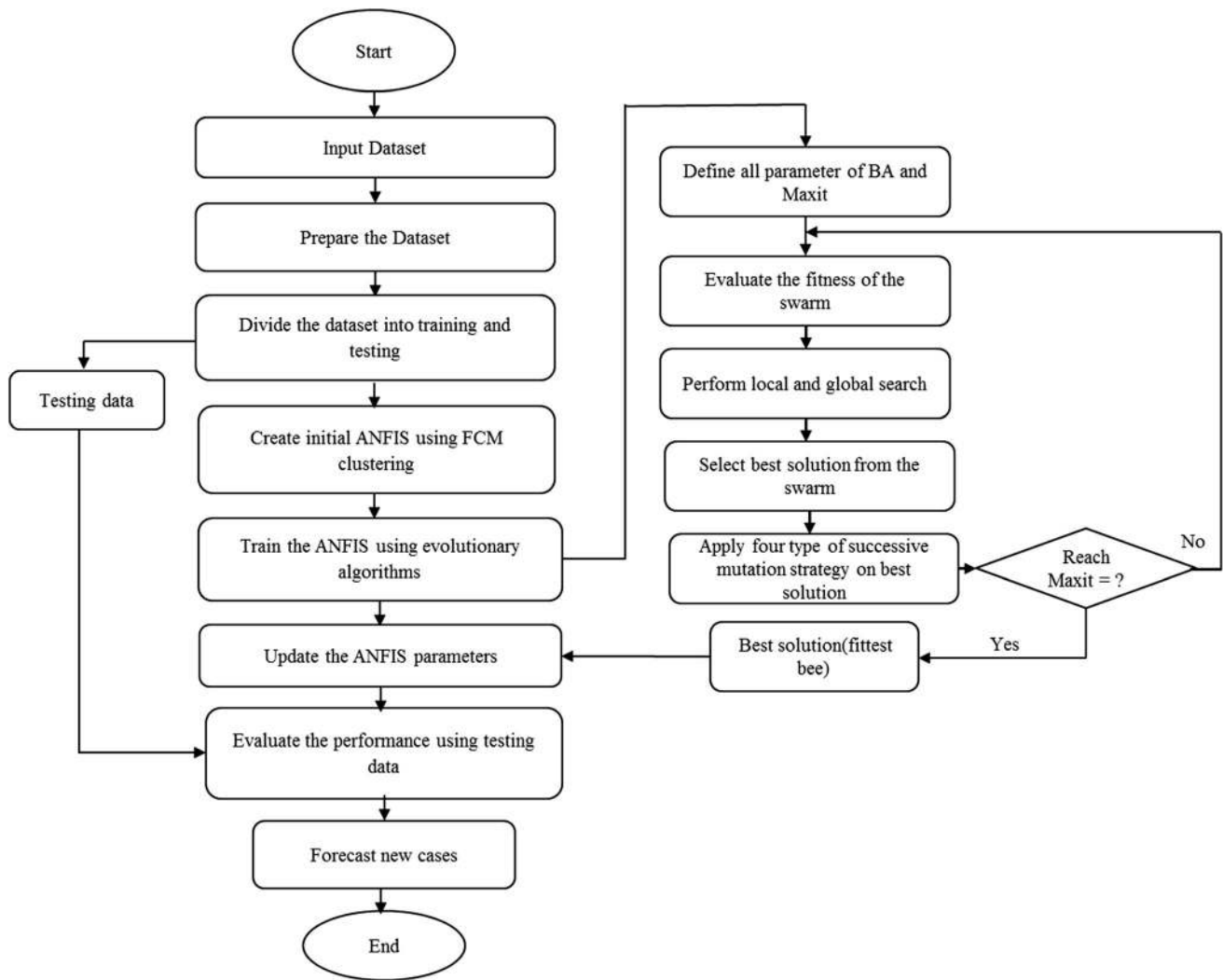
#### DE-Based Mutation

DE-based mutation is the final and last stage of this successive mutation strategy.

$$\text{modified\_bee} = \text{bee1} + F(Va - Vb) \quad (10)$$

where  $Va$  and  $Vb$  are two random bees from the swarm and  $F$  is the scale factor.

ANFIS consists of five layers as the standard ANFIS model, where Layer 1 is the input (previous data of confirmed cases) and Layer 5 is the output layer which produces the forecasted value. During learning, mBA is



**Fig. 2** The proposed mBA-ANFIS model

employed to select the proper weights between layers 4 and 5. In the first step of the model, input data are formatted in time series form and the data are split in 80:20 ratio, where 80% of the data have been used for training purpose and 20% to test the model. Fuzzy c-mean (FCM) technique has been used to define the number of clusters to be constructed in the ANFIS model. The parameters of the ANFIS model have been produced by mBA algorithm and used during training. To evaluate the parameters produced, error between the actual data and predicted data has been calculated using Eq. 11.

$$RMSE = \sqrt{\frac{1}{N_r} \sum_{i=1}^{N_r} (P_i - A_i)^2} \tag{11}$$

where  $N_r$  is number of rows in the dataset,  $P_i$  is the predicted value, and  $A_i$  is the actual value. The smaller error calculated by this objective function depicts good

parameters of ANFIS. After the completion of training phase, testing has been done using the best solution to obtain the final output. Performance of the proposed model has been assessed by comparing real number of confirmed cases with the predicted number of confirmed cases using performance metrics, Finally, mBA-ANFIS gives forecasted value of confirmed cases in India for the next day. Algorithm 1 depicts the proposed mBA-ANFIS algorithm.

**Algorithm 1: Proposed mBA-ANFIS algorithm**


---

```

Split the data into training and testing
Select the numbers and types of input and output MFs
Create initial ANFIS using Fuzzy c-mean (FCM) clustering
Generate random swarm with  $n$  ScoutBee with each having dimension  $d$ 
Calculate  $p, e, Be, Bs$ 
Evaluate the fitness of the swarm
while (termination criteria is not meet) do
    Determine the size of patch size
    Select  $Be$  to perform neighborhood search for top  $e$  site
    Employ  $Bs$  bees for  $(p - e)$  site
    Assign remaining bee  $(n - p)$  to search randomly
    Calculate fitness of all bees
    Select fittest bee (bee1) from current generation
    for  $i = 1$  to  $d$  /*Cauchy mutation*/
         $modified\_bee(d) = bee1 + (Vmax(d) - Vmin(d) \cdot Cauchy(o, s))$ 
    end for
    if ( $f(modified\_bee) < f(bee1)$ )
         $bee1 = modified\_bee$ 
         $f(bee1) = f(modified\_bee)$ 
    end if
    for  $i = 1$  to  $d$  /* Opposition-based mutation separately for each dimension */
         $modified\_bee = Vmin(d) + Vmax(d) - bee1$ 
        if ( $f(modified\_bee) < f(bee1)$ )
             $bee1 = modified\_bee$ 
             $f(bee1) = f(modified\_bee)$ 
        end if
    end for
     $modified\_bee = Vmin + Vmax - bee1$  /* Opposition-based mutation as a whole */
    if ( $f(modified\_bee) < f(bee1)$ )
         $bee1 = modified\_bee$ 
         $f(bee1) = f(modified\_bee)$ 
    end if
     $modified\_bee = bee1 + F(Va - Vb)$  /* DE -based mutation */
    if ( $f(modified\_bee) < f(bee1)$ )
         $bee1 = modified\_bee$ 
         $f(bee1) = f(modified\_bee)$ 
    end if
Generate new swarm with  $n$  ScoutBee
end while
Return fittest bee (Best parameter set of ANFIS)
For testing use ANFIS with best configuration
Calculate the performance of test data using some statistical measure
Forecast number of new COVID-19 confirmed cases

```

---

## Implementation and Result Analysis

In this section, we have discussed about experimental setup, parameters settings of algorithms, dataset description, performance metrics and experimental results.

### Experimental Setup

Proposed mBA-ANFIS model has been implemented using Matlab 2015 on a desktop PC consisting Intel Core i5 Processor with a clock speed of 3.4 GHZ and 4 GB of RAM. Other meta-heuristic algorithms such as GA, PSO,

DE, FF, HS, and TLBO has also been implemented using same platform. Ten independent runs have been performed for each algorithm using population size of 25 and 200 iterations. Parameters configuration for each algorithm is shown in Table 3.

### Dataset Description

COVID-19 dataset contains historical data of confirmed cases for most of the countries of the world. This dataset is updated on regular basis and can be downloaded from [40]. From the dataset, we have chosen India and USA data



**Table 3** Parameters setup

Algorithm	Parameter Value
GA	$Maxit = 200, npop = 25, Pc = 0.7, Pm = 0.3$
PSO	$Maxit = 200, npop = 25, w = 1, C1 = 2, C2 = 2$
DE	$Maxit = 200, npop = 25, pCR = 0.2, F = 0.9$
HS	$Maxit = 200, HMS = 25, HMCR = 0.9, PAR = 0.1,$
FF	$Maxit = 200, npop = 25, alpha = 0.2, gamma = 1,$
BA	$Maxit = 200, n = 25, p = 13, e = 5, Be = 26, Bs = 13,$ $r = 1$
mBA	$Maxit = 200, n = 25, p = 13, e = 5, Be = 26, Bs = 13,$ $r = 1, s = 2, F = 0.1$

samples to apply in our model. India data samples contain the confirmed cases from January 30, 2020, to August 9, 2020, whereas USA data samples contain the number of confirmed cases from January 21, 2020, to August 9, 2020. For both of the countries, 80% of the samples have been used to train the proposed model and rest of 20% has been used for testing purpose. Table 4 shows the dataset samples.

**Performance Metrics**

Proposed model has been evaluated using the following performance metrics.

- Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N_r} \sum_{i=1}^{N_r} (P_i - A_i)^2}$$

- Normalized Root Mean Square Error (NRMSE):

$$NRMSE = \frac{RMSE}{\max(A) - \min(A)}$$

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{N_r} \sum_{i=1}^{N_r} |P_i - A_i|$$

- Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{N_r} \sum_{i=1}^{N_r} \left| \frac{A_i - P_i}{A_i} \right|$$

where  $N_r$  is number of rows in the dataset,  $P_i$  is the predicted value, and  $A_i$  is the actual value. The lowest value of RMSE, NRMSE, MAE and MAPE depicts the best method. Average experimental result in terms of RMSE, MAPE, MAE and NRMSE has been reported in Result Analysis section.

**Experimental Result**

Proposed model has been compared with standard ANFIS model and nine other hybrid models. Each model has been executed 10 times using India and USA datasets. During each run, RMSE, MAE, MAPE and NRMSE have been calculated and average value of 10 runs is reported in Table 5 and 6 for India and USA datasets. It can be observed from both of the tables that mBA-ANFIS has outperformed the other models in all performance metrics for both countries. It has achieved RMSE of 1280.24, MAE of 685.68, MAPE of 6.24 and NRMSE of 0.000673 for India dataset. In this case, BA-ANFIS ranked second

**Table 4** Sample COVID-19 data of India

iso_code	continent	location	date	total_cases	new_cases	total_deaths	new_deaths
IND	Asia	India	2020-01-30	1	1	0	0
IND	Asia	India	2020-01-31	1	0	0	0
IND	Asia	India	2020-02-01	1	0	0	0
IND	Asia	India	2020-02-02	2	1	0	0
IND	Asia	India	2020-02-03	2	0	0	0
IND	Asia	India	2020-02-04	3	1	0	0
IND	Asia	India	2020-02-05	3	0	0	0
IND	Asia	India	2020-02-06	3	0	0	0
IND	Asia	India	2020-02-07	3	0	0	0
IND	Asia	India	2020-02-08	3	0	0	0
IND	Asia	India	2020-02-09	3	0	0	0
IND	Asia	India	2020-02-10	3	0	0	0
IND	Asia	India	2020-02-11	3	0	0	0
IND	Asia	India	2020-02-12	3	0	0	0

**Table 5** Performance evaluation for India COVID-19 dataset

Method	RMSE	MAE	MAPE	NRMSE
ANFIS	1704.591	873.775	24.53	0.000862
GA-ANFIS	1621.3	864.0032	22.78	0.000861
DE-ANFIS	1618.74	856.3195	18.58	0.000789
HS-ANFIS	1602.248	849.1674	10.45	0.000781
TLBO-ANFIS	1532.317	834.6748	15.53	0.000717
FF-ANFIS	1521.143	776.0159	12.49	0.000714
PSO-ANFIS	1319.345	725.1992	8.73	0.000714
BA-ANFIS	1409.775	722.2773	7.54	0.000691
mBA-ANFIS	<b>1280.244</b>	<b>685.6884</b>	<b>6.24</b>	<b>0.000673</b>

**Table 6** Performance evaluation for USA COVID-19 dataset

Method	RMSE	MAE	MAPE	NRMSE
ANFIS	9642.85	7591.01	22.43	0.00204
GA-ANFIS	9199.81	6772.61	19.65	0.00218
DE-ANFIS	9383.78	7419.22	21.43	0.002
HS-ANFIS	7503.62	5321.98	11.48	0.00159
TLBO-ANFIS	7913.39	6769.67	15.43	0.00168
FF-ANFIS	7143.98	4458.142	9.21	0.00154
PSO-ANFIS	5916.74	4036.52	9.45	0.001242
BA-ANFIS	4490.96	3086.67	6.25	0.000955
mBA-ANFIS	<b>4468.72</b>	<b>3082.07</b>	<b>6.1</b>	<b>0.000952</b>

**Table 7** Forecasted confirmed cases of COVID-19 in India by proposed mBA-ANFIS model

Date	Expected new cases
10-08-2020	2,218,059
11-08-2020	2,278,067
12-08-2020	2,339,889
13-08-2020	2,399,825
14-08-2020	2,460,925
15-08-2020	2,520,713
16-08-2020	2,587,512
17-08-2020	2,653,444
18-08-2020	2,721,159
19-08-2020	2,789,407
20-08-2020	2,858,807

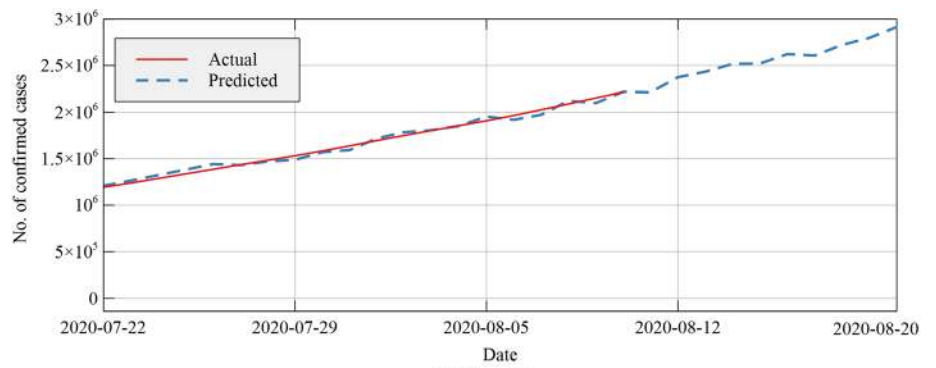
followed by PSO-ANFIS, FF-ANFIS, TLBO-ANFIS, DE-ANFIS, GA-ANFIS and standard ANFIS models. Similarly for USA, the values are 4468.72, 3082.07, 6.1, and

0.000952 for RMSE, MAE, MAPE, and NRMSE, respectively. Here also BA-ANFIS has achieved the second rank, followed by PSO-ANFIS, FF-ANFIS, HS-ANFIS, TLBO-ANFIS, GA-ANFIS, DE-ANFIS and standard ANFIS models.

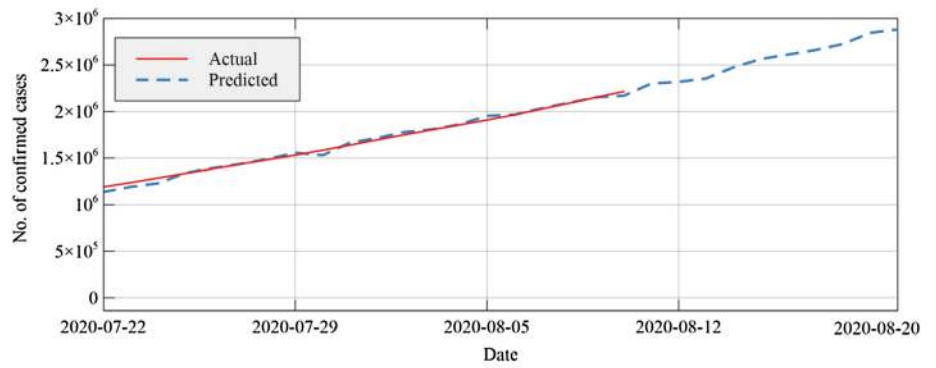
We have graphically plotted the actual and predicted number of confirmed cases generated by these models as well as forecasted value for India. For better representation, we have selected the date range of July 22, 2020, to August 20, 2020, as shown in Fig. 3.

From the above figure, it can be observed that standard Bees Algorithm performs well compared to other meta-heuristic algorithms. The error between the actual value and predicted value is less than other algorithms, but this error can be minimized by proposed model which can be noticed from Fig. 3 (i) where the difference between the actual and predicted is minimal compared to others which makes it the best model among the others. Table 7 shows the forecasting value of expected confirmed cases of COVID-19 in India using proposed model for August 10, 2020, to August 20, 2020.

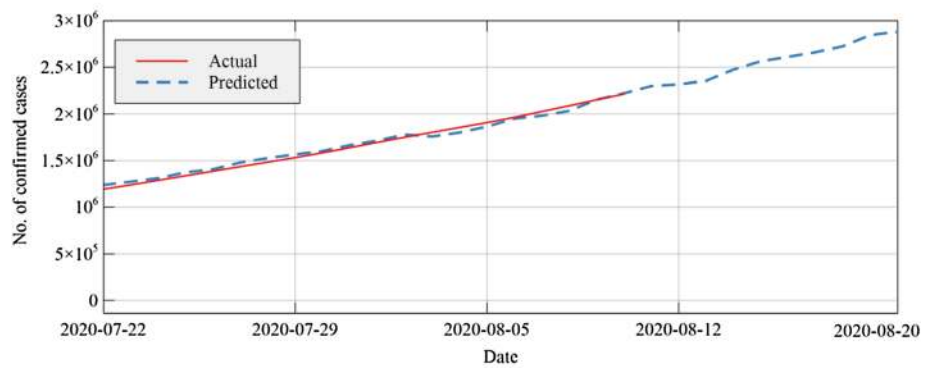
**Fig. 3** Real data (actual) against the forecasted data (predicted) for all models



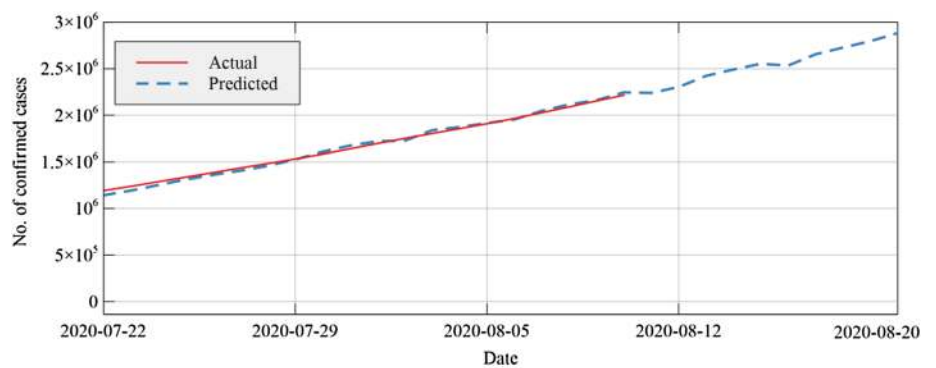
**(a)** ANFIS



**(b)** GA-ANFIS

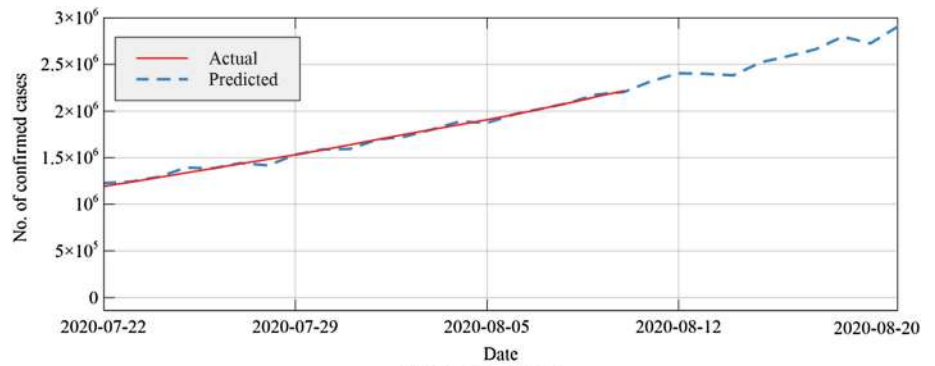


**(c)** DE-ANFIS

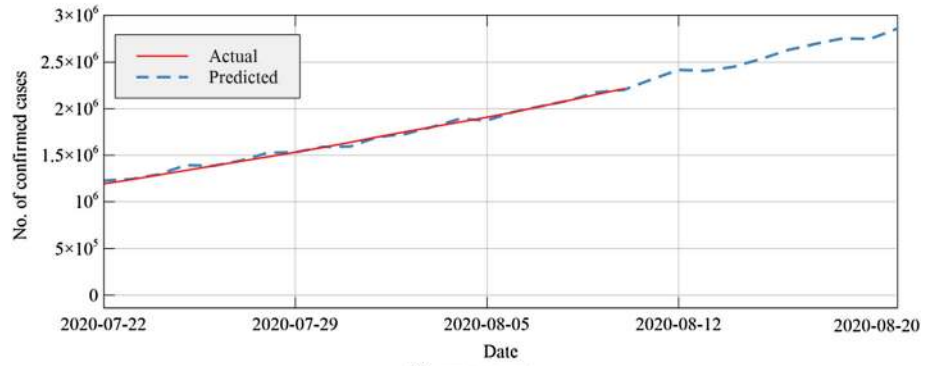


**(d)** HS-ANFIS

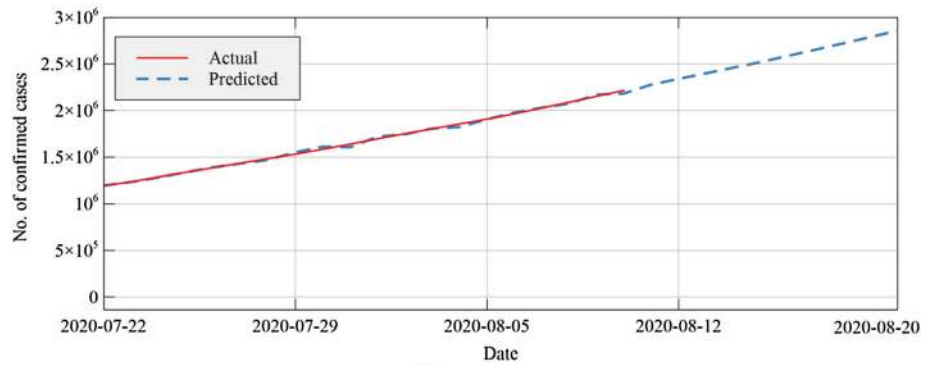
Fig. 3 continued



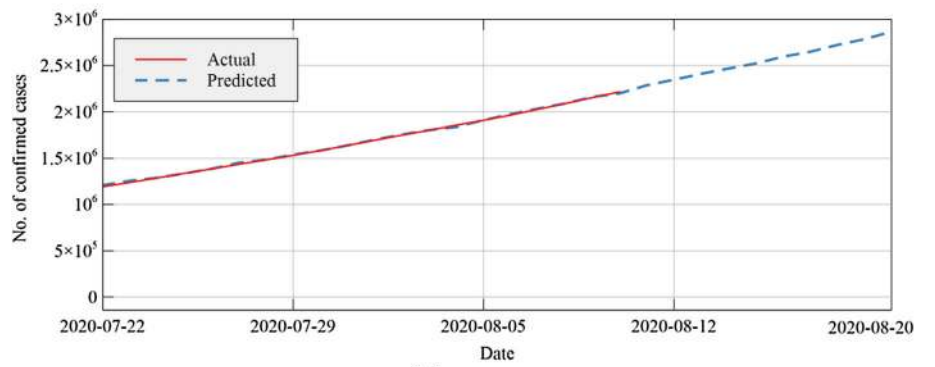
(e) TLBO-ANFIS



(f) FF-ANFIS

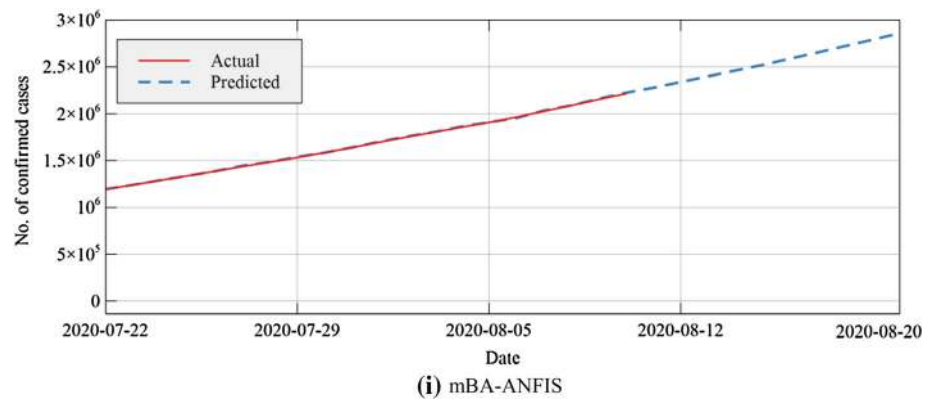


(g) PSO-ANFIS



(h) BA-ANFIS

Fig. 3 continued



## Conclusion

In this paper, a unique hybrid model mBA-ANFIS consisting of mutation-based Bees Algorithm (mBA) and adaptive neuro-fuzzy inference system (ANFIS) has been used to forecast the number confirmed cases. For more diversification, we have modified the standard Bees Algorithm (BA) by introducing mutation process. Four types of successive mutation have been applied on the fittest bee, which is found after local and global search. These modifications improve the performance of the algorithm compared to the standard one. The proposed mBA-ANFIS has been implemented using Matlab and compared with several hybrid algorithms. This model has been applied on COVID-19 dataset for two countries. Experimental results show that the model is very much effective to forecast new cases since it has obtained less RMSE, MAE, MAPE and NRMSE for both of the countries. These values are relatively low compared to other methods. According to the promising results achieved by the proposed model, it can be employed to other forecasting applications.

**Acknowledgements** This work has been carried out with support of grant received from WBDST funded research project on secure remote healthcare with project sanction no. 230(Sanc)/ST/P/S&T/6G-14/2018.

## References

- X.Y. Ge, J.L. Li, X.L. Yang, A.A. Chmura, G. Zhu, J.H. Epstein, J.K. Mazet, B. Hu, W. Zhang, C. Peng et al., Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013)
- L.F. Wang, Z. Shi, S. Zhang, H. Field, P. Daszak, B.T. Eaton, Review of bats and SARS. *Emerg. Infect. Dis.* **12**(12), 1834 (2006)
- A. Anzai, T. Kobayashi, N.M. Linton, R. Kinoshita, K. Hayashi, A. Suzuki, Y. Yang, S. Jung, T. Miyama, A.R. Akhmetzhanov, H. Nishiura, Assessing the impact of reduced travel on exportation dynamics of novel coronavirus infection (COVID-19). *J Clin Med.* Volume. **9**(2), 601 (2020)
- A. Agarwal, A. Lubet, E. Mitgang, S.K. Mohanty, D.E. Bloom, population aging in india: facts, issues, and options. PGDA Working Papers 13216, Program on the Global Demography of Aging (2016)
- “CoVID news by MIB”. Twitter. Retrieved 09 August <https://twitter.com/COVIDNewsByMIB/status/1292308781234233345>
- K. Sarkar, S. Khajanchi, J.J. Nieto, Modeling and forecasting the COVID-19 pandemic in India. *Chaos, Solitons Fractals* **139**, 110049 (2020)
- A.H. Elsheikh, A.I. Saba, M.A. Elaziz, S. Lu, S. Shanmugan, T. Muthuramalingam, R. Kumar, A.O. Mosleh, F.A. Essa, T.A. Shehabeldeen, Deep learning-based forecasting model for COVID-19 outbreak in Saudi Arabia, *Process Safety and Environmental Protection*, 149, 223–233, 2021.
- Michał Wieczorek, Jakub Silka, Marcin Woźniak, Neural network powered COVID-19 spread forecasting model. *Chaos, Solitons & Fractals* **140**, 110203 (2020)
- D. Parbat, M. Chakraborty, A python based support vector regression model for prediction of COVID19 cases in India. *Chaos Solitons Fractals* **138**, 109942 (2020)
- N. Hasan, A methodological approach for predicting COVID-19 epidemic using EEMD-ANN hybrid model. *Internet Things* **11**, 100228 (2020)
- J.S.R. Jang, Anfis: Adaptive-network-based fuzzy inference system. *IEEE Trans. Sys. Man. Cybern.* **23**(3), 665–668 (1993)
- R. Sujath, J.M. Chatterjee, A.E. Hassanien, A machine learning forecasting model for COVID-19 pandemic in India. *Stoch Environ Res Risk Assess* **34**, 959–972 (2020)
- A. Tomar, N. Gupta, Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Sci Total Environ* **728**, 138762 (2020)
- F. Rustam et al., COVID-19 Future Forecasting Using Supervised Machine Learning Models. *IEEE Access* **8**, 101489–101499 (2020)
- I.Celestine, B.A. Kashif, P.Atharva, R.Sujatha, C.Jyotir Moy, P. Swetha, M.Rishita, P.Sofia, J.Ohyun, COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm, *Frontiers in Public Health*, 8 .357 2020.
- S.F. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A.R. Varkonyi-Koczy, U. Reuter, T. Rabczuk, P.M. Atkinson, P.M. COVID-19 Outbreak Prediction with Machine Learning. *medRxiv*, 2020.
- G. Pinter, I. Felde, A. Mosavi, P. Ghamisi, R. Gloaguen, COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach. *Mathematics* **8**, 890 (2020)
- 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**, 674 (2020)

19. L.J. Muhammad, M.M. Islam, S.S. Usman et al., Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients' Recovery. *SN COMPUT. SCI.* Volume. **1**, 206 (2020)
20. J. Kumar, K.P.S.S Hembram. Epidemiological study of novel coronavirus (COVID-19). 2020 arXiv preprint 2003.11376.
21. P. Wang, X. Zheng, J. Li, B. Zhu, Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals* **139**, 110058 (2020)
22. S.I. Alzahrani, I.A. Aljamaan, E.A. Al-Fakih, Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *J. Infect. Pub. Health* **13**(7), 914–919 (2020)
23. F.M. Khan, R. Gupta, ARIMA and NAR based prediction model for time series analysis of COVID- 19 cases in India. *J. Saf. Sci. Resil* **1**(1), 12–18 (2020)
24. A. Behnood, E.M. Golafshani, S.M. Hosseini, Determinants of the infection rate of the COVID- 19 in the U.S. using ANFIS and virus optimization algorithm (VOA). *Chaos, Solitons & Fractals* **139**, 110051 (2020)
25. Elmousalami, Haytham&Hassanien, Aboul. (2020). Day Level Forecasting for Coronavirus Disease (COVID-19) Spread: Analysis, Modeling and Recommendations, arXiv preprint, 2003.07778
26. U. Şahin, T. Şahin, Forecasting the cumulative number of confirmed cases of COVID-19 in Italy, UK and USA using fractional nonlinear grey Bernoulli model. *Chaos, Solitons & Fractals* **138**, 109948 (2020)
27. Y. Peng, M.H. Nagata, An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data. *Chaos Solitons & Fractals* **139**, 110055 (2020)
28. M.A. Al-Qaness, M.A. Elaziz, A.A. Ewees, Oil consumption forecasting using optimized adaptive neuro-fuzzy inference system based on sine cosine algorithm. *IEEE Access* **6**, 68394–68402 (2018)
29. K. Ahmed et al., A hybrid Krill-ANFIS model for wind speed forecasting. in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016. AISI 2016. Advances in Intelligent Systems and Computing* ed. by A. Hassanien, K. Shaalan, T. Gaber, A. Azar, M. Tolba, vol 533, (Springer, Cham, 2017) pp. 365–372
30. M.A. Al-qaness, M. AbdElaziz, A.A. Ewees, X. Cui, A Modified Adaptive Neuro-Fuzzy Inference System Using Multi-Verse Optimizer Algorithm for Oil Consumption Forecasting. *Electronics* **8**, 1071 (2019)
31. Z. Alameer, M.A. Elaziz, A.A. Ewees, H. Ye, Z. Jianhua, Forecasting copper prices using hybrid adaptive neuro-fuzzy inference system and genetic algorithms. *Nat. Resour. Res.* **28**, 1385–1401 (2019)
32. D.T. Pham, A. Ghanbarzadeh, E. Koc, S. Otri, S. Rahim and M. Zaidi The Bees Algorithm. Technical Note Manufacturing Engineering Centre Cardiff University, UK, 2005.
33. D. Karaboga, E. Kaya, Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. *Artif Intell Rev* **52**, 2263–2293 (2019)
34. A. Sarkheyli, A.M. Zain, S. Sharif, Robust optimization of ANFIS based on a new modified GA. *Neurocomputing* **166**, 357–366 (2015)
35. M. Ehteram, H.A. Afan, M. Dianatikhah, A.N. Ahmed, C. Ming Fai, M.S. Hossain, M.F. Allawi, A. Elshafie, Assessing the Predictability of an Improved ANFIS Model for Monthly Streamflow Using Lagged Climate Indices as Predictors. *Water* **11**, 1130 (2019)
36. A. RezaeeJordehi, Enhanced leader PSO (ELPSO): A new PSO variant for solving global optimisation problems. *Appl. Soft Comput.* **26**, 401–417 (2015)
37. C. Li, Y. Liu, A. Zhou, L. Kang and H. Wang. "A fast particle swarm optimization algorithm with cauchy mutation and natural selection strategy." In *International Symposium on Intelligence Computation and Applications*, pp. 334–343. Springer, Berlin, Heidelberg, 2007.
38. M.I. Iqbal, N.K. Khan, M.A. Jaffar, M. Ramzan, and A. Rauf Baig. "Opposition based genetic algorithm with Cauchy mutation for function optimization." In *2010 International Conference on Information Science and Applications*, pp. 1–7. IEEE, 2010.
39. Hinterding, Robert. "Gaussian mutation and self-adaption for numeric genetic algorithms." In *Proceedings of 1995 IEEE International Conference on Evolutionary Computation* 1, 384. IEEE, 1995.
40. COVID-19 data maintained by Our World in Data, <https://covid.ourworldindata.org/data/owid-covid-data.xlsx>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.