A hybrid model based on mBA-ANFIS for COVID 19 confirmed cases prediction and forecast

Sohail Saif (sohailsaif7@gmail.com)
Maulana Abul Kalam Azad University of Technology, West Bengal
Priya Das
Maulana Abul Kalam Azad University of Technology, West Bengal
Suparna Biswas
Maulana Abul Kalam Azad University of Technology, West Bengal

Research Article

Keywords: mutation based bees algorithm (mBA), adaptive neuro-fuzzy inference system (ANFIS), COVID-19, Forecast, RMSE, MAE, MAPE, NRMSE, India

Posted Date: September 25th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-83375/v1

Version of Record: A version of this preprint was published at Journal of The Institution of Engineers (India): Series B on January 19th, 2021. See the published version at https://doi.org/10.1007/s40031-021-00538-0.

A hybrid model based on mBA-ANFIS for COVID 19 confirmed cases prediction and forecast

Sohail Saif[1], PriyaDas[2], Suparna Biswas[3]

[1,2,3]*Department of Computer Science & Engineering*

[1,2,3]*Maulana Abul Kalam Azad University of Technology, WB.*

[1]sohailsaif7@gmail.com

[2]28priyadas@gmail.com

[3]mailtosuparna@gmail.com

Abstract

In India, the first confirmed case of novel corona virus (COVID-19) was discovered on 30 January, 2020. The number of confirmed cases is increasing day by day and it crossed 21,53,010 on 09 August, 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 modelis based on adaptive neuro-fuzzy inference system (ANFIS) and mutation based Bees Algorithm (mBA). ThemetaheuristicBees 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 USAand the number of confirmed cases in next 10 days in Indiahas been forecasted. Proposed mBA-ANFISmodelhas 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 showthat the proposed model has achieved better performance in terms of Root Mean squared 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, 0.000952 for RMSE, MAE, MAPE, NRMSE respectively.

Keywords

mutation based bees algorithm (mBA), adaptive neuro-fuzzy inference system (ANFIS), COVID-19, Forecast, RMSE, MAE, MAPE, NRMSE, India.

1 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, 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 respiratoryproblems and surprisingly these people have visited the seafood market in Wuhan, China [3]. In few days 79,355 peoples was affected by this virus in China by 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].

In India, the first case of COVID was reported on 30-01-2020 and it started increasing gradually.To break the chain of transmission, a complete lockdown was declared by Indian Government as a result it seemed be under control till March 24,2020 where the total number of confirmed case 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 09-08-2020 it has been reported 21,53,010 confirmed cases and 38,135 deaths as shown in Table I.Primarily, most of the people, who came in contact with the infected persons were not aware about the symptoms in them.

Date of Reporting	Confirmed Cases	Deaths
09.08.2020	21,53,010	43379
29.07.2020	15,31,669	34193
15.07.2020	9,36,181	24309
01.07.2020	5,85,493	1740
15.06.2020	3,32,424	9520
14.05.2020	78,003	2549
15.04.2020	11438	377
24.03.2020	492	9
01.03.2020	3	0
30.01.2020	1	0

Table I. Date wise report of confirmed cases and deaths in India

India has the lowest case fatality rate of 2.04% and the 68.32% patients has 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. 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) [6], artificial neural networks (ANN) [7] etc. However, these individual models suffer from over-fitting 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 next 10 days up to August 20,2020based on the previous confirmed case data of India. Proposed model is a hybrid combination of Adaptive neuro-fuzzy inference system (ANFIS) [8] 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.

2 Related Work

In this section, severalresearchworks on Machine Learning or Deep Learning based techniques to foresee the confirmed cases in upcoming days have been discussed.

Sujath et al. [9] 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 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 has been used in Weka and Orange and LR, MLP and VAR has been employed to predict the future effects of this deadly virus in India. Experimentalresults shows that MLP method outperforms other two method.

Tomar and Gupta [10] used data-driven estimation methods such as long short-term memory (LSTM) and curve fitting approach to forecast theexpected COVID-19 cases for 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. [11] 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,RMSE has been used to choose the proper model. Experimental results proves that ES performs best compared to other models.

Celestine et al. [12] 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 has been used as input dataset. 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 archives prediction accuracy of 94% and F1 score of 0.86 which is highest among other algorithms.

Ardabili et al. [13] 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 has been combined with Multi-layered perceptron (MLP) and Adaptive neuro fuzzy inference system (ANFIS). To evaluate the performance of these algorithms,RMSE and R^2 has been calculated using 8 mathematical models such as Logistic, Linear, Logarithmic, Quadratic, Cubic, Compound, Power and Exponential. For prediction, COVID-19 dataset of 4 countries has been employed.

Pinter et al. [14] proposed a hybrid machine learning method to predict COVID-19 cases in Hungary. Adaptive neuro fuzzy inference system (ANFIS) has been combined with multi-layered 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 has been used for training purpose and rest of the data has been used for validation. Evaluation metrics such as RMSE, MAPE and R^2 has been considered in this work.

Al-qaness et al. [15]described an enhanced ANFIS model to forecast COVID-19 confirmed cases for next 10 days in China. Flower Pollination Algorithm (FPA), Salp Swarm Algorithm (SSA) and ANFIS has 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. Proposed model achieved better performance in terms of RMSE, EMSRE, MAPE and R^2.

Muhammad et al. [16] 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 has been applied on the dataset. Experimental result shows an accuracy of 98.85%, 97.52%, 99.85%, 97.49%, 99.60%, 98.06% respectively, where Decision tree has the highest accuracy.

Kumar and Hembram [17] 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. 5 cities from each country where the number of infected patients are high has been considered for the experiment and three environmental factors such as maximum temperature, relative humidity and wind speed has been considered. Experimental results shows 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 has a little impact on the spread of this virus. Table II shows comparative analysis of experimental results.

Authors	Approach	Country	Evaluation Metrics			
			RMSE	MAPE (%)	MAE	R^2
Rustam et al. [11]	ES	Worldwide	16828.58	-	8867.43	0.98
Ardabili et al. [13]	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. [14]	MPL-ICA	Hungary	167.88	23.15	-	0.9971
Al-qaness et al. [15]	FPASSA-ANFIS	China	5779	4.79	4271	0.9645
Elmousalami et al. [18]	SES	Worldwide	4477.72	9.68	-	-
Sahin et al. [19]	FANGBM	Italy	1223	0.9174	-	0.9993
	FANGBM	UK	796	2.813	-	0.9996
	FANGBM	USA	5767	4.8950	-	0.9996

Table II. Comparative analysis of experimental results

3 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 need 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. While reviewing the recent works we could not find sufficient research works to predict confirm cases in India.This motivated us to propose Learning based enhanced Adaptive neuro fuzzy inference system (ANFIS) system 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 [20, 21, 22, 23].To train our model, 155 days historical data of confirmed COVID-19 patients in India starting from 30-01-2020 has 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 type of mutation. 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 effected patients in India for 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) has 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 20-08-2020 has been presented.

4 Preliminaries

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

4.1 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 has been represented in Figure 1. The interference system uses IF-THEN rule to map between input and output.

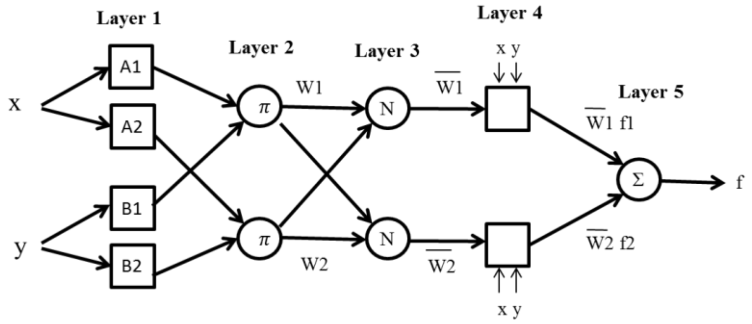


Fig.1. Structure of ANFIS model.

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 represented as

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

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

Where μ denote Gaussian MF. μAp , $\mu Bp - 2$ is fuzzy membership value and b, c are parameters of membership function.

Layer 2: This layer use fizzy operator (AND) to fuzzify the input. The output node (O_{2p}) represented as

$$O_{2p} = \mu A_p \times \mu B_{p-2}(y) \tag{3}$$

Layer3: 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

$$O_{3p} = \overline{W}_p = \frac{W_p}{\sum_{p=1}^{2} W_p} \tag{4}$$

Layer4: 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.

$$O_{4p} = \overline{W}_p \times f_p \quad \text{Where} \quad f_p = p_p x + q_p y + r_p \tag{5}$$

Here p_p, q_p ,$r_p$arethe parameter set

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

$$O_5 = \sum_p \overline{W_p} f_p = \frac{\sum_p W_p f_p}{\sum_p W_p} \tag{6}$$

4.2 Bees Algorithm (BA):

The Bees Algorithm (BA) was invented by Pham et.al [24] 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 patch ($site$) out of n recently visited site is p, number of elite site out of p recently visited region is e, number of recruited bees for elite site is Be, number of recruited bee for (p-e) site is Bs and finally set the size of each patch by 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 is 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, fittest bee is selected out of the swarm.

5 Proposed Model

A hybridmethod of Deep Learning and Meta heuristic Algorithm has been used in this work. 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 typeof artificial neural network which is popular for time series prediction on the other hand Bees Algorithm (BA) is an optimization algorithm which has beenemployed to optimize the parameters of ANFIS. Figure 2 shows the proposed model.

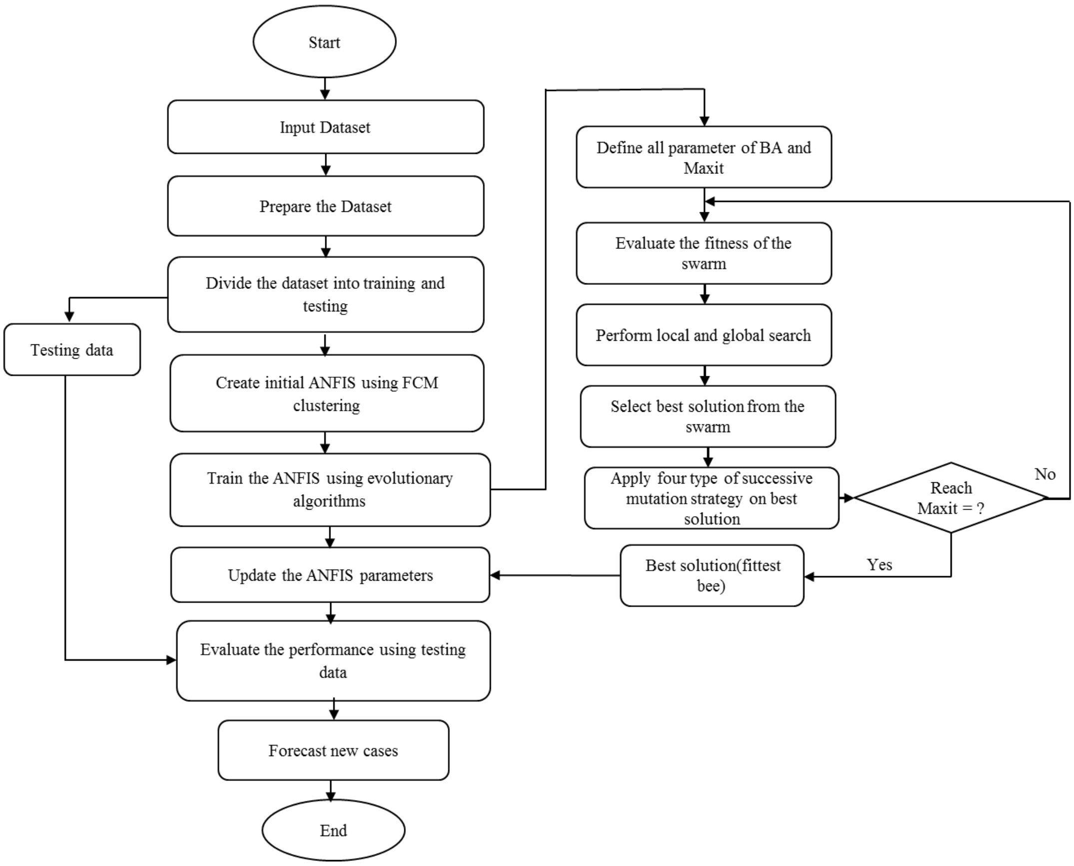


Fig.2. The proposed mBA-ANFIS model.

For more diversification, we have modified the standard Bees Algorithm (BA) by introducing mutationprocess. 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 type of successive mutation [25] has 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 be not changed. The benefit of this strategy is the best solution will never be lost and also there is a chance for getting better solution. Proposed Mutation Based Bees Algorithm (mBA) uses four different mutation operatorswhich are as following.

Cauchy mutation

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

$$modified\_bee(d) = bee1 + (Vmax(d) - Vmin(d).Cauchy(o,s)) \text{for d = 1,2,3....n} \quad (7)$$

Here s represent 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.

$modified\_bee(d) = Vmin(d) + Vmax(d) - bee1$ for d = 1,2,3….n (8)

Opposition-based mutation as whole

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

$$modified\_bee = Vmin + Vmax - bee1 \quad (9)$$

DE –based mutation

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

$$mofified\_bee = bee1 + F(Va - Vb) \quad (10)$$

Where Va and Vb are two random bee from the swarm, 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 employed to select the proper weights between Layer 4 and 5**.**In the first step of the model, input data is formatted in time series form and the data is splitted in 80:20 ratio, where 80% of the data has 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 hasbeen produced by mBA algorithm and that is used during training. To evaluate the parameters produced, error between the actual data and predicted data has been calculated using equation 11.

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

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 withthe predicted number of confirmed cases using performance metrics, Finally, mBA-ANFIS gives foretasted 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*/**
$$modified\_bee(d) = \text{bee1} + (Vmax(d) - Vmin(d).Cauchy(o,s))$$
end for
if$(f(modified\_bee) < f(\text{bee1}))$
$$bee1 = modified\_bee$$
$$f(bee1) = f(modified\_bee)$$
end if
for i = 1 to d **/* Opposition-based mutation separately for each dimension */**
$$modified\_bee = Vmin(d) + Vmax(d) - \text{bee1}$$
if$(f(modified\_bee) < f(\text{bee1}))$
$$bee1 = modified\_bee$$
$$f(bee1) = f(modified\_bee)$$
end if
end for
$modified\_bee = Vmin + Vmax - \text{bee1}$ **/* Opposition-based mutation as a whole */**
if$(f(modified\_bee) < f(\text{bee1}))$
$$bee1 = modified\_bee$$
$$f(bee1) = f(modified\_bee)$$
end if
$modified\_bee = \text{bee1} + F(Va - Vb)$ **/* DE –based mutation */**
if$(f(modified\_bee) < f(\text{bee1}))$
$$bee1 = modified\_bee$$
$$f(bee1) = f(modified\_bee)$$
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

6 Implementation and Result Analysis

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

6.1 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 4GB of RAM. Other meta heuristic algorithms such as GA, PSO, DE, FF, HS, TLBO has also been implemented using same platform. 10 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 III.

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	$Maxit = 200$, $n = 25$, $p = 13$, $e = 5$, $Be = 26$, $Bs = 13$, $r = 1$, $s = 2$, $F = 0.1$

Table III. Parameters setup

6.2 Dataset Description:

COVID-19 dataset which 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 [26]. From the dataset we have chosenIndia and USA data samples to apply in our model. India data samples contain the confirmed cases from 30.01.2020 to 09.08.2020 whereas USA data samples contain the number of confirmed cases from 21.01.2020 to 09.08.2020. For both of the countries 80% of the samples has been used to train the proposed model and rest of 20% has been used for testing purpose.Table IV shows the dataset samples.

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 IV. Sample COVID-19 data of India

6.3 Performance Metrics:

Proposed model has been evaluated using the following performance metrics

•Root Mean Square Error (RMSE):

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

•Normalized Root Mean Square Error (NRMSE):

$$NRMSE = \frac{RMSE}{\max(A) - \min(A)} \tag{12}$$

•Mean Absolute Error (MAE):

$$MAE = \frac{1}{N_r}\sum_{i=1}^{N_r}\left|P_i - A_i\right| \tag{13}$$

•Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{N_r}\sum_{i=1}^{N_r}\left|\frac{A_i - P_i}{A_i}\right| \tag{14}$$

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.

6.4 Experimental result:

Proposed model has been compared with standard ANFIS model and 9 other hybrid models. Each model has been executed 10 times using India and USA dataset. During Each run RMSE, MAE, MAPE and NRMSE has been calculated and average value of 10 runs has been reported in Table V and Table VI for India and USA dataset.It can be observed from both of the tables thatmBA-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 Data. In this case BA-ANFIS ranked second followed by PSO-ANFIS, FF-ANFIS, TLBO-ANFIS, DE-ANFIS, GA-ANFIS and standard ANFIS model. Similarly for USA the values are 4468.72, 3082.07, 6.1, 0.000952 for RMSE, MAE, MAPE, NRMSE respectively. Here also BA-ANFIS has achieved the second rank, followed by PSO-ANFIS, FF-ANFIS, HS-ANFS, TLBO-ANFIS, GA-ANFIS, DE-ANFIS and standard ANFIS model.

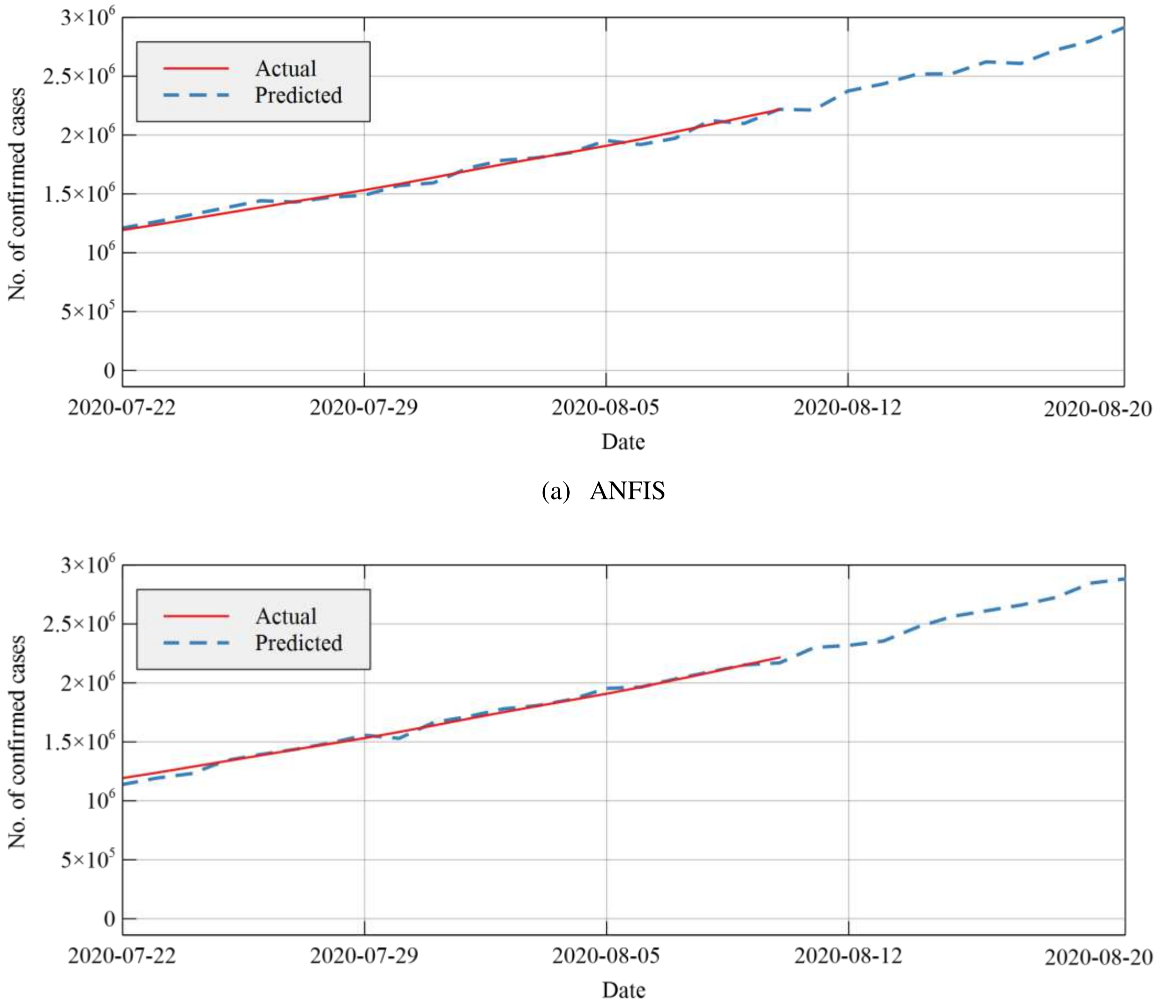
Method	**RMSE**	**MAE**	**MAPE**	**NRMSE**
ANFIS	1704.591	873.775	24.53	0.000862
GA-ANFIS	1621.3	864.0032	22.78	0.000861
DE-ANFIS	1618.74	856.3195	18.58	0.000789
HS-ANFIS	1602.248	849.1674	10.45	0.000781
TLBO-ANFIS	1532.317	834.6748	15.53	0.000717
FF-ANFIS	1521.143	776.0159	12.49	0.000714
PSO-ANFIS	1319.345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

Table V. Performance evaluation for India 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	**4468.72**	**3082.07**	**6.1**	**0.000952**

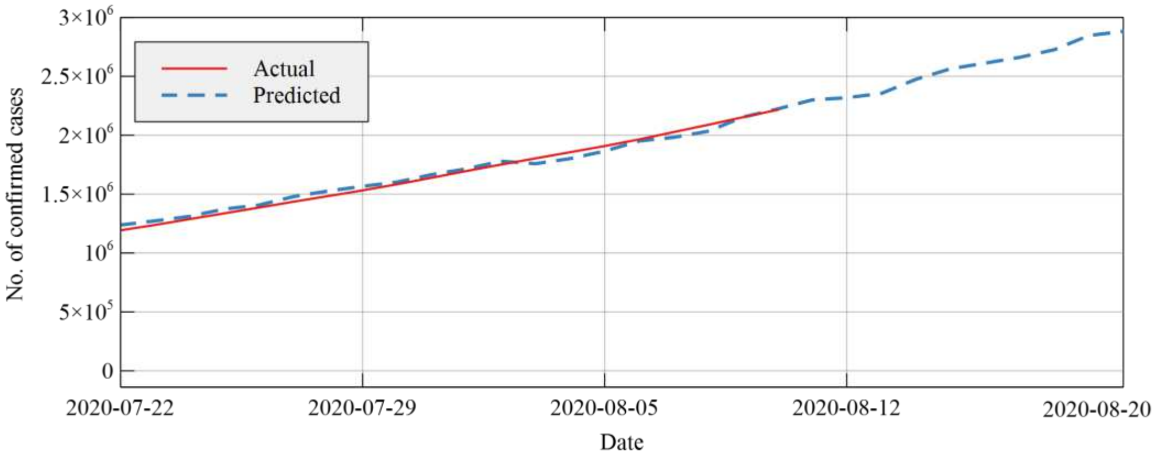
Table VI. Performance evaluation for USA COVID-19 dataset.

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 22.07.2020 to 20.08.2020 as shown in Figure 3.

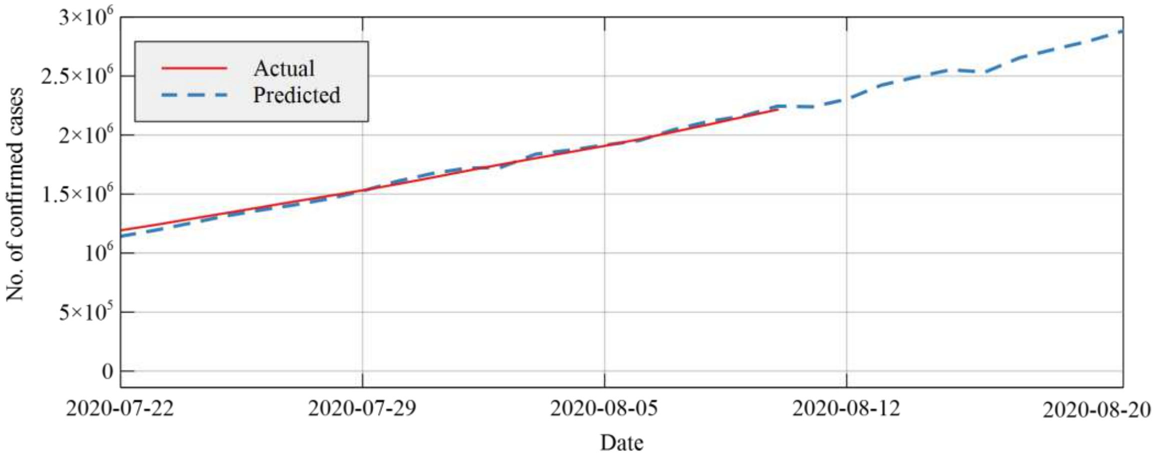


(a) ANFIS

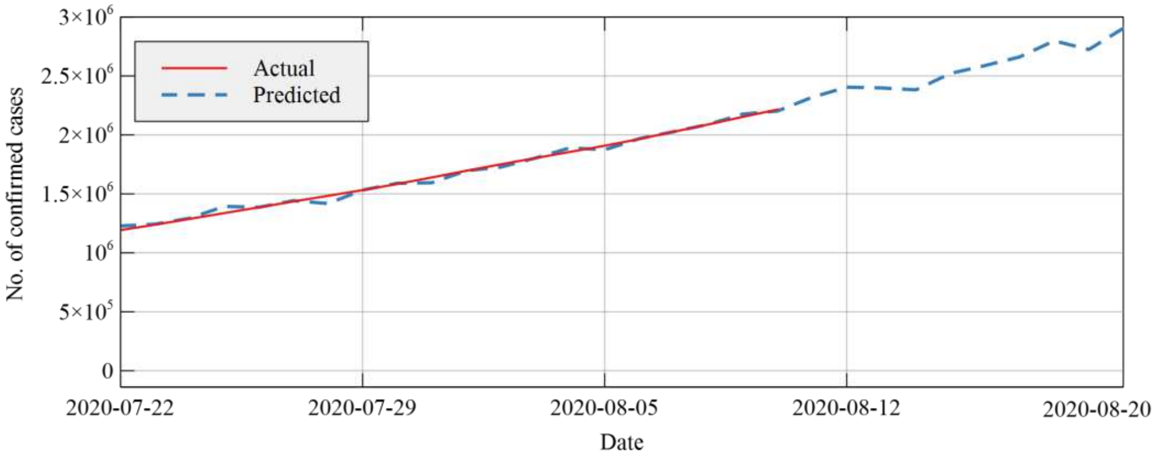
(b) GA-ANFIS



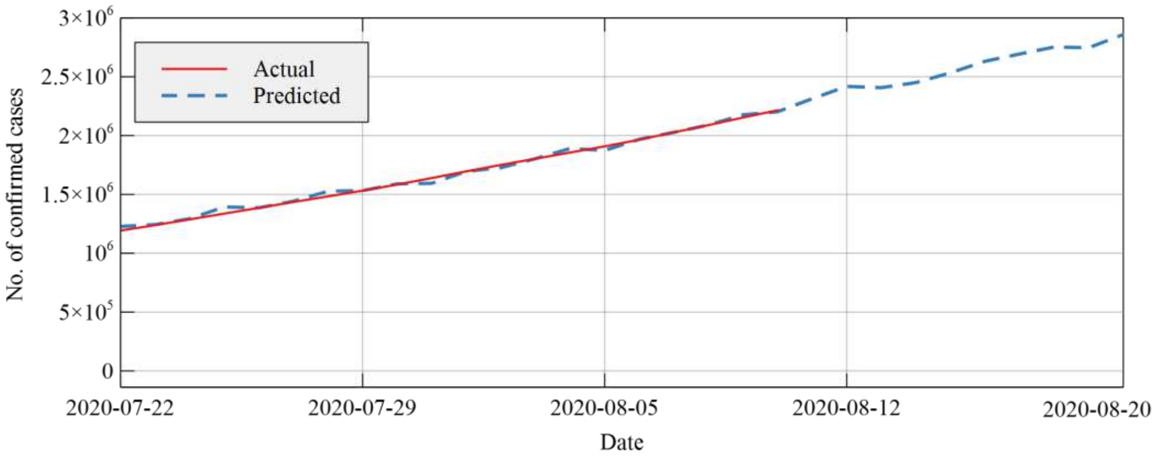
(c) DE-ANFIS



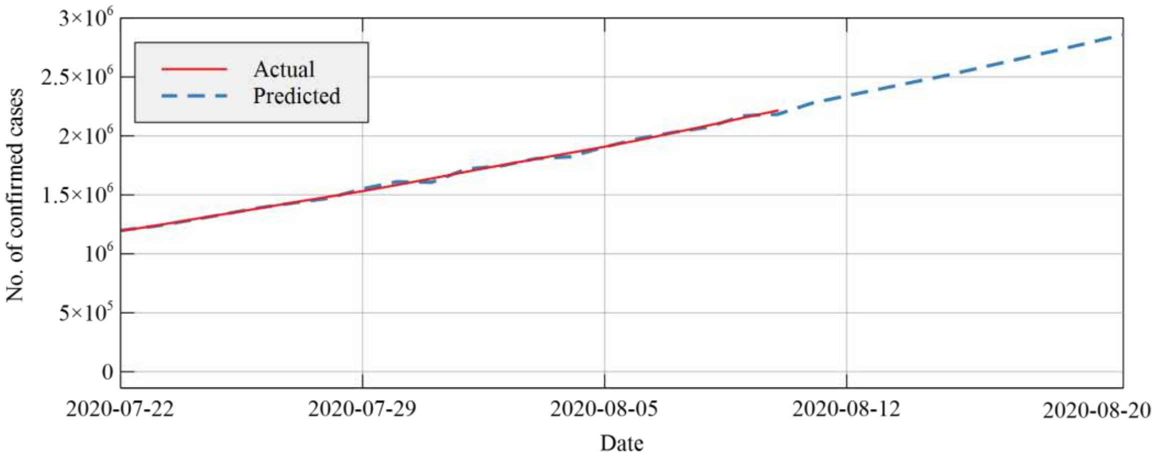
(d) HS-ANFIS



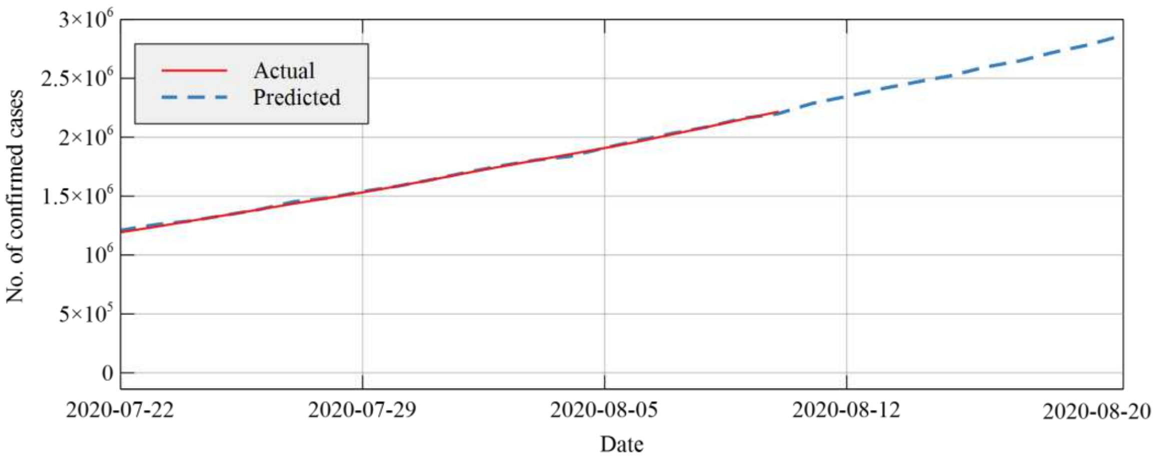
(e) TLBO-ANFIS



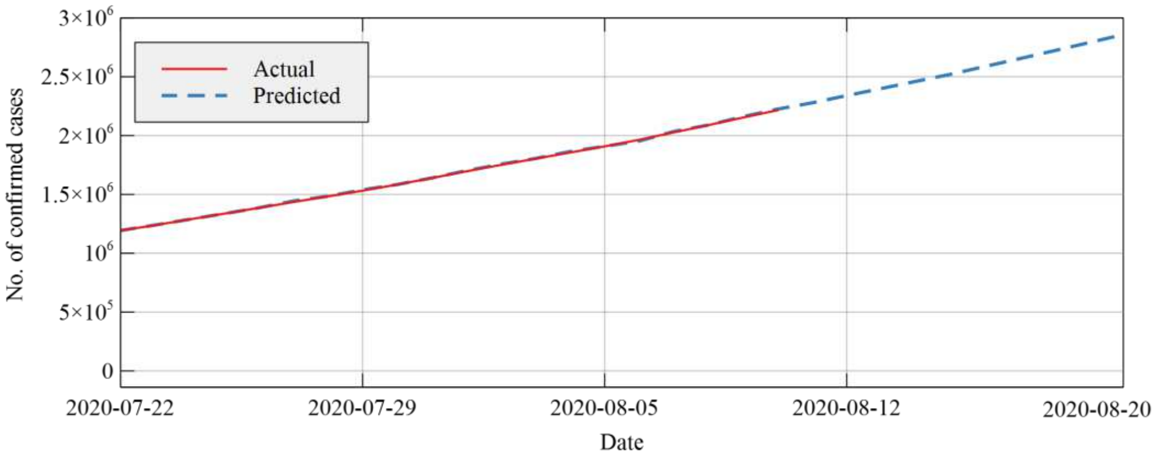
(f) FF-ANFIS



(g) PSO-ANFIS



(h) BA-ANFIS



(i) mBA-ANFIS

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

From the above figure it can be observed 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 Figure 3 (i) where the difference between the actual and predicted is minimal compared to others which makes it best model among the others.Table VII shows the forecasting value of expected confirmed cases of COVID-19 in India using proposed model for 10.08.2020 to 20.08.2020.

Date	**Expected new cases**
10-08-2020	2218059
11-08-2020	2278067
12-08-2020	2339889
13-08-2020	2399825
14-08-2020	2460925
15-08-2020	2520713
16-08-2020	2587512
17-08-2020	2653444
18-08-2020	2721159
19-08-2020	2789407
20-08-2020	2858807

Table VII. Forecasted confirmed cases of COVID-19 in India by proposed mBA-ANFIS model.

7 Conclusion

In this paper, a hybrid model mBA-ANFIS consisting mutation based Bees Algorithm (mBA) and adaptive neuro-fuzzy inference system (ANFIS) has been used to forecast the number confirmed cases. Here standard Bees Algorithm has been modified by adding four types of mutation operator. These modifications improve the performance of the algorithm compared to the standard one. 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 shows 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

[1] Ge, X.Y.; Li, J.L.; Yang, X.L.; Chmura, A.A.; Zhu, G.; Epstein, J.H.; Mazet, J.K.; Hu, B.; Zhang, W.; Peng, C.; et al.Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 2013,503, 535–538.

[2] Wang, L.F.; Shi, Z.; Zhang, S.; Field, H.; Daszak, P.; Eaton, B.T. Review of bats and SARS. Emerg. Infect. Dis. 2006, 12, 1834.

[3]Anzai A, Kobayashi T, Linton NM, Kinoshita R, Hayashi K, Suzuki A, Yang Y, Jung S, Miyama T, Akhmetzhanov AR, Nishiura H (2020) Assessing the impact of reduced travel on exportation dynamics of novel coronavirus infection (COVID-19). J Clin Med 9(2):601.

[4]Agarwal A, Lubet A, Mitgang E, Mohanty S, Bloom DE (2020) Population aging in india: facts, issues, and options, issue 10162,pp 289–311.

[5]"CoVID news by MIB". Twitter. Retrieved 09 August https://twitter.com/COVIDNewsByMIB/status/1292308781234233345

[6]Parbat D, Chakraborty M. A python based support vector regression model for prediction of COVID19 cases in India. Chaos Solitons Fractals.138:109942;2020.

[7]N.Hasan,A Methodological Approach for Predicting COVID-19 Epidemic Using EEMD-ANN Hybrid Model,Internet of Things,Vol. 11,2020

[8]Jang, J.-S. R.,Anfis: Adaptive-network-based fuzzy infer-encesystem.IEEE Transactions on Systems, Man, andCybernetics,23(3), 665–68,1993

[9] Sujath, R., Chatterjee, J.M. &Hassanien, A.E. A machine learning forecasting model for COVID-19 pandemic in India.Stoch Environ Res Risk Assess 34, 959–972,2020.

[10]A.Tomar, N.Gupta,Prediction for the spread of COVID-19 in India and effectiveness of preventive measures,Science of The Total Environment,Vol. 728,2020

[11]F. Rustam et al., "COVID-19 Future Forecasting Using Supervised Machine Learning Models," in IEEE Access, vol. 8, pp. 101489-101499, 2020.

[12]I.Celestine, B.A. Kashif, P.Atharva, Sujatha R., 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, Vol.8, pp.357, 2020.

[13]Ardabili, S.F.; Mosavi, A.; Ghamisi, P.; Ferdinand, F.; Varkonyi-Koczy, A.R.; Reuter, U.; Rabczuk, T.;Atkinson, P.M. COVID-19 Outbreak Prediction with Machine Learning. medRxiv 2020

[14]Pinter, G.; Felde, I.; Mosavi, A.; Ghamisi, P.; Gloaguen, R. COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach. Mathematics, 8, 890,2020

[15]Al-qaness, M.A.A.; Ewees, A.A.; Fan, H.; Abd El Aziz, M. Optimization Method for Forecasting Confirmed Cases of COVID-19 in China. J. Clin. Med., 9, 674,2020

[16]Muhammad, L.J., Islam, M.M., Usman, S.S. et al. Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients' Recovery. SN COMPUT. SCI. 1, 206 (2020)

[17]Kumar J, Hembram KPSS. Epidemiological study of novel coronavirus (COVID-19). 2020 arXiv preprint https ://arXiv:2003.11376

[18]Elmousalami, Haytham&Hassanien, Aboul. (2020). Day Level Forecasting for Coronavirus Disease (COVID-19) Spread: Analysis, Modeling and Recommendations, arXiv preprint, arXiv:2003.07778

[19]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,Vol. 138,2020.
[20]Al-Qaness, M.A.; Elaziz, M.A.; Ewees, A.A. Oil consumption forecasting using optimized adaptive neuro-fuzzy inference system based on sine cosine algorithm. IEEE Access 2018, 6, 68394–68402.
[21]Ahmed, K.; Ewees, A.A.; El Aziz, M.A.; Hassanien, A.E.; Gaber, T.; Tsai, P.W.; Pan, J.S. A hybrid krill-ANFIS model for wind speed forecasting. In International Conference on Advanced Intelligent Systems and Informatics; Springer: Berlin, Germany, 2016; pp. 365–372
[22]Al-qaness, M.A.; AbdElaziz, M.; Ewees, A.A.; Cui, X. A Modified Adaptive Neuro-Fuzzy Inference System Using Multi-Verse Optimizer Algorithm for Oil Consumption Forecasting.Electronics 2019, 8, 1071.
[23]Alameer, Z.; Elaziz, M.A.; Ewees, A.A.; Ye, H.; Jianhua, Z. Forecasting copper prices using hybrid adaptive neuro-fuzzy inference system and genetic algorithms. Nat. Resour. Res. 2019, 28, 1385–1401.
[24]Pham DT, Ghanbarzadeh A, Koc E, Otri S, Rahim S and Zaidi M. The Bees Algorithm. Technical Note, Manufacturing Engineering Centre, Cardiff University, UK, 2005
[25]A. RezaeeJordehi,Enhanced leader PSO (ELPSO): A new PSO variant for solving global optimisationproblems,Applied Soft Computing,Vol. 26,Pages 401-417,2015.
[26]COVID-19 data maintained by Our World in Data, https://covid.ourworldindata.org/data/owid-covid-data.xlsx

Figures

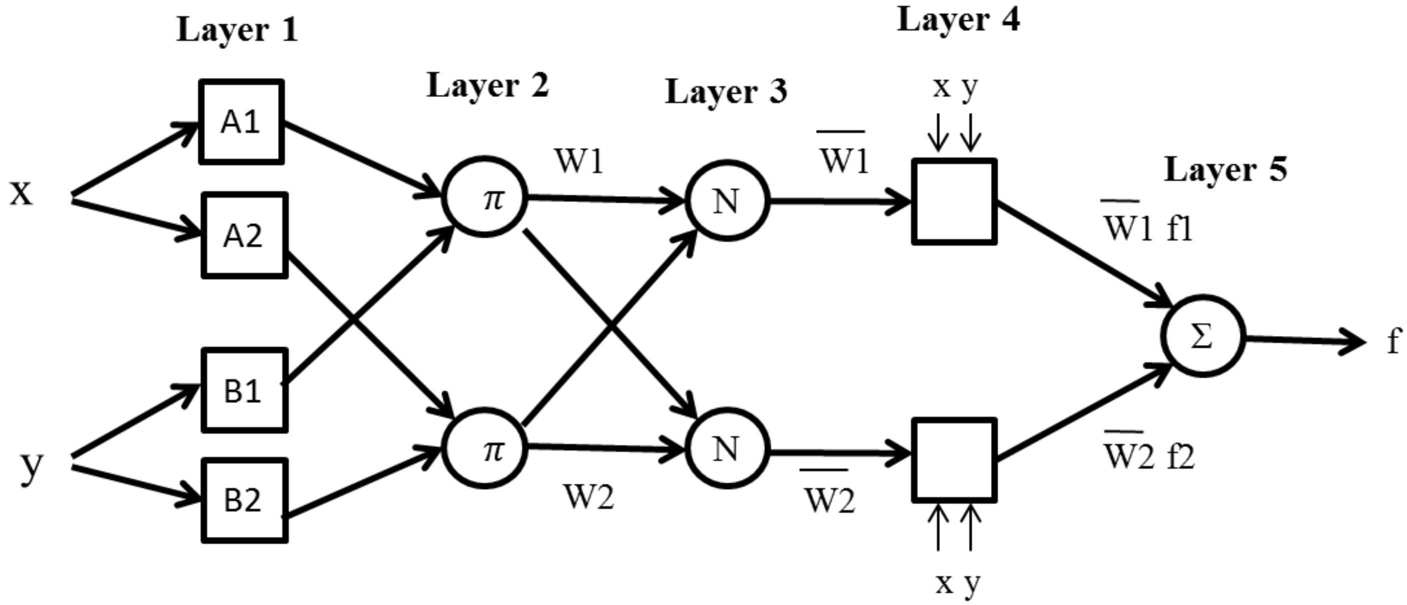


Figure 1

Structure of ANFIS model.

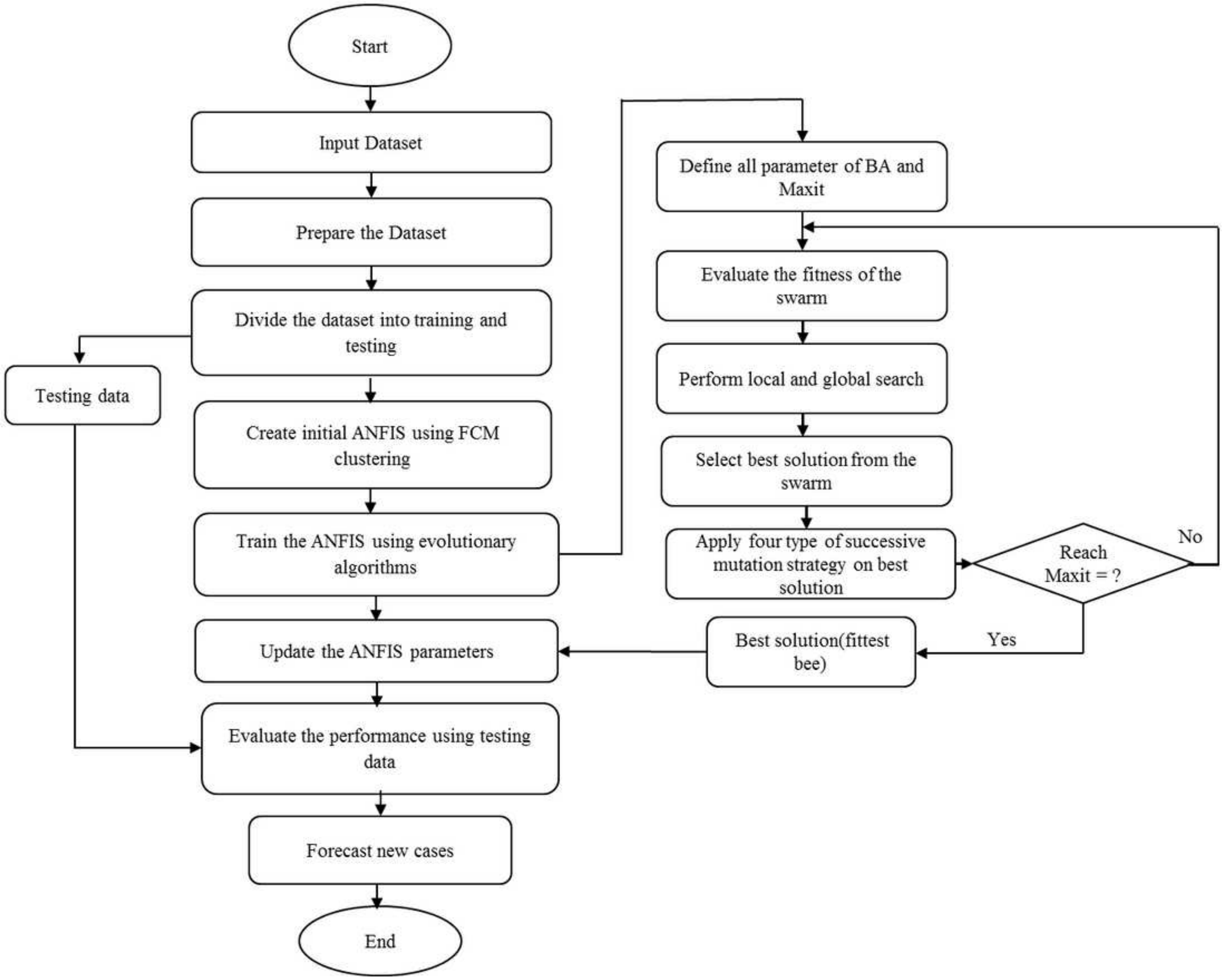


Figure 2

Structure of ANFIS model.

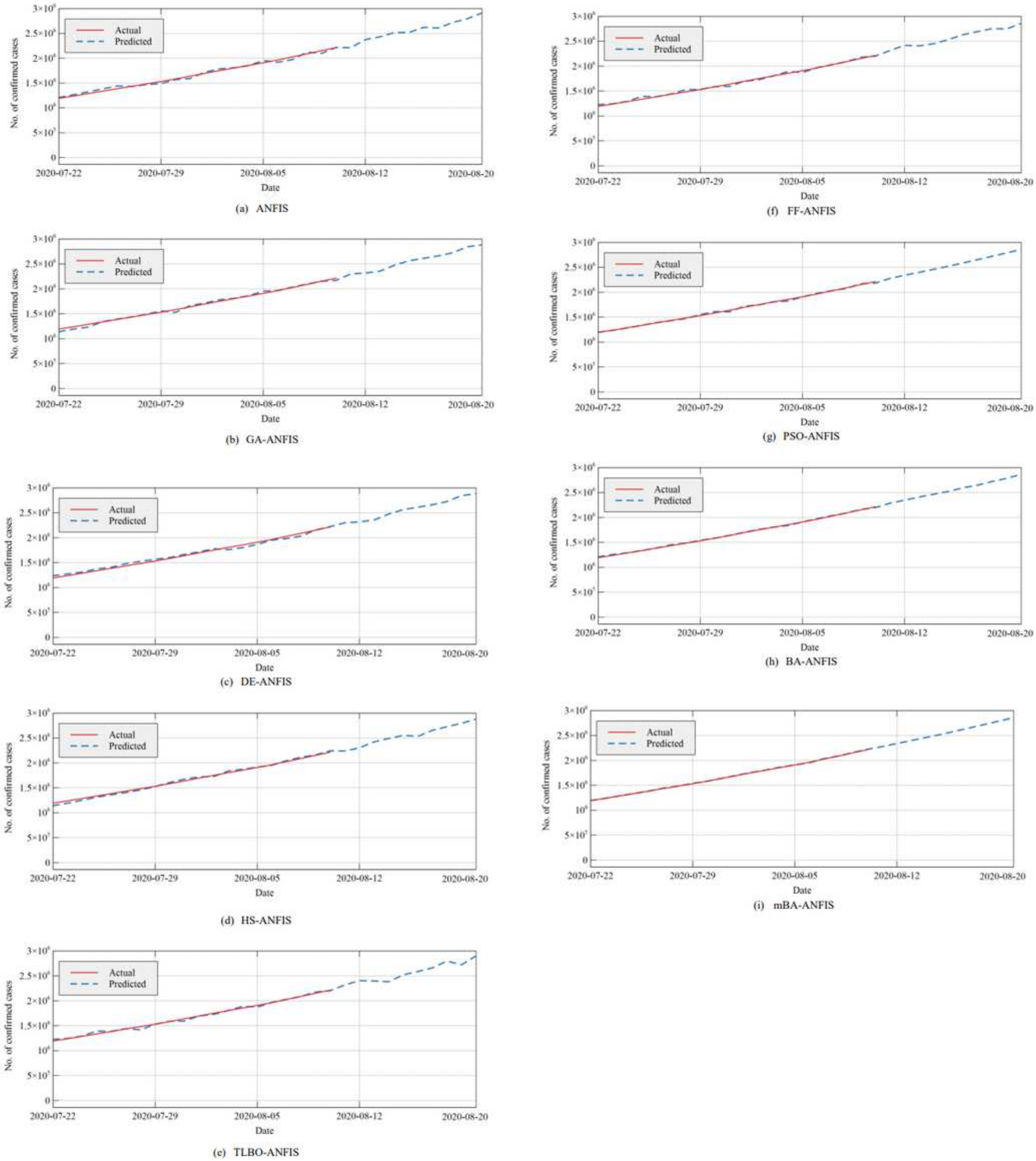


Figure 3

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