COVID-19 Pandemic Prediction for Hungary; a Hybrid Machine Learning Approach

Gergo Pinter, Imre Felde, Amir Mosavi and Pedram Ghamisi

May 2, 2020

COVID-19 Pandemic Prediction for Hungary; a Hybrid Machine Learning Approach

Gergo Pinter [1], Imre Felde [1], Amir Mosavi [2,3,*], Pedram Ghamisi [4]

[1] John von Neumann Faculty of Informatics, Obuda University, 1034 Budapest, Hungary.
[2] Kalman Kando Faculty of Electrical Engineering, Obuda University, 1034 Budapest, Hungary.
[3] Faculty of Civil Engineering, Technische Universität Dresden, 01069 Dresden, Germany
[4] Machine Learning Group, Exploration Division, Helmholtz Institute Freiberg for Resource Technology, Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany; p.ghamisi@hzdr.de
*   Correspondence: a.mosavi@ieee.org

Abstract: For the prediction of the COVID-19 outbreak, several epidemiological models are being used around the world to project the number of infected individuals and mortality rates. Advancing accurate prediction models is of utmost importance to take proper actions. Due to a high level of uncertainty and lack of essential data, the standard epidemiological models have been challenged for delivering higher accuracy for long-term prediction. As an alternative to the susceptible-infected-resistant (SIR)-based models, this study proposes a hybrid machine learning approach to predict the COVID-19 outbreak in Hungary. The hybrid machine learning methods of adaptive network-based fuzzy inference system (ANFIS) and multi-layered perceptron-imperialist competitive algorithm (MLP-ICA) are used to predict the time series of the infected individuals and mortality rate. The models predict that by late May, the outbreak and the total morality will drop substantially. The validation is performed for nine days with promising results, which confirms the model accuracy. It is expected that the model maintains its accuracy as long as no significant interruption occurs. Based on the results reported here, and due to the complex nature of the COVID-19 outbreak and variation in its behavior from nation-to-nation, this study suggests machine learning as a useful tool to model the outbreak. This paper provides an initial benchmarking to demonstrate the potential of machine learning for future research.

Keywords: COVID-19; Coronavirus; SARS-CoV-2; prediction model; machine learning

1. Introduction

Severe acute respiratory syndrome coronavirus 2, also known as SARS-CoV-2, is reported as a virus strain causing the respiratory disease of COVID-19 [1]. The World Health Organization (WHO) and the global nations confirmed the coronavirus disease to be extremely contagious [2,3]. The COVID-19 pandemic has been widely recognized as a public health emergency of international concern [4]. To estimate the outbreak, identify the peak ahead of time, and also predict the mortality rate the epidemiological models had been widely used by officials and media. Outbreak prediction models have shown to be essential to communicate insights into the likely spread and consequences of COVID-19. Furthermore, governments and other legislative bodies used the insights from prediction models to suggest new policies and to assess the effectiveness of the enforced policies [5].

The COVID-19 pandemic has been reported to be extremely aggressive to spread [6]. Due to the complex nature of COVID-19 outbreak and its irregularity in different countries, the standard epidemiological models, i.e., susceptible-infected-resistant (SIR)-based models, had been challenged for delivering higher performance in individual nations. Furthermore, as the COVID-19 outbreak

showed major differences with other recent outbreaks, e.g., Ebola, Cholera, swine fever, H1N1 influenza, dengue fever, and Zika, advanced epidemiological models have been emerged to provide higher accuracy [7]. Nevertheless, due to several unknown variables involved in the spread, the complexity of population-wide behavior in various countries and differences in containment strategies model uncertainty has been reported inevitable [8-10]. Consequently, standard epidemiological models face new challenges to deliver more reliable results.

The strategy standard SIR models is formed around the assumption of transmitting the infectious disease through contacts, considering three different classes susceptible, infected, and recovered [11]. The susceptible to infection (class S), infected (class I), and the removed population (class R) build the foundation of the epidemiological modeling. Note that definition of various classes of outbreak may vary. For instance, R is often referred to those that have recovered, developed immunity, been isolated, or passed away. However, in some countries, R is susceptible to be infected again and there exists uncertainties in allocating R a value. Advancing SIR-based models requires several assumptions. It is assumed that the class I transmits the infection to class S where the number of probable transmissions is proportional to the total number of contacts computed using basic differential equations as follows [12-14].

$$\frac{dS}{dt} = -\alpha SI \tag{1}$$

where I, S, and α represent the infected population, the susceptible population, and the daily reproduction rate, respectively. The value of S in the time-series produced by the differential equation gradually declines. At the early stage of the outbreak, it is assumed that $S \approx 1$ where $\frac{dI}{dt}$ becomes linear. Eventually the class I can be stated as follows.

$$\frac{dI}{dt} = \alpha SI - \beta I, \quad \frac{dR}{dt} = \beta I \tag{2}$$

where β regulates the daily rate of spread. Furthermore, the individuals excluded from the model is computed as follows. Considering the above assumption, the outbreak modeling with SIR is finally computed as follows:

$$I\,(t) \approx I_0\,exp\{(\alpha - \beta)\}. \tag{3}$$

Furthermore, to evaluate the performance of the SIR-based models, the median success of the outbreak prediction is used which is calculated as follows:

$$f = \frac{Prediction}{True\ value}. \tag{4}$$

Several analytical solutions to the SIR models have been provided in literature [15,16]. As the different nations take different actions toward slowing down the outbreak, the SIR-based model must be adopted according to the local assumptions [17]. Inaccuracy of many SIR-based models in predicting the outbreak and mortality rate have been evidenced during the COVID-19 in many nations. The key success of an SIR-based model relies on choosing the right model according to the context and the relevant assumptions. SIS, SIRD, MSIR, SEIR, SEIS, MSEIR, and MSEIRS models are among the popular models used for predicting COVID-19 outbreaks worldwide. The more advanced variation of SIR-d models carefully considers the vital dynamics and constant population [16]. For instance, at the presence of the long-lasting immunity assumption when the immunity is not realized upon recovery from infection, the susceptible-infectious-susceptible (SIS) model was suggested [18]. In contrast, the susceptible-infected-recovered-deceased-model (SIRD) is used when immunity is assumed [19]. In the case of COVID-19 different nations took different approaches in this regard.

SEIR models have been reported among the most popular tools to predict the outbreak. SEIR models through considering the significant incubation period of an infected person reported to present relatively more accurate predictions. In the case of Varicella and Zika outbreaks the SEIR models showed increased model accuracy [20,21]. SEIR models assume that the incubation period is a random variable and similarly to the SIR model, there is a disease-free-equilibrium [22,23]. It should be noted, however, that SEIR models can not fit well where the contact network are non-stationary

through time [24]. Social mixing as key factor of non-stationarity determines the reproductive number R_0 which is the number of susceptible individuals for infection. The value of R_0 for COVID-19 was estimated to be 4 which greatly trigged the pandemic [1]. The lockdown measures aimed at reducing the R_0 value down to 1. Nevertheless, the SEIR models are reported to be difficult to fit in the case of COVID-19 due to the non-stationarity of mixing, caused by nudging intervention measures. Therefore, to develop accurate SIR-based models an in-depth information about the social movement and the quality of lockdown measures would be essential. Other drawback of SIR-based models is the short lead-time. As the lead-time increases, the accuracy of the model declines. For instance, for the COVID-19 outbreak in Italy, the accuracy of the model reduces from $f = 1$ for the first five days to $f = 0.86$ for day 6 [17]. Overall, the SIR-based models would be accurate if firstly the status of social interactions is stable. Secondly, the class R can be computed precisely. To better estimate the class R, several data sources can be integrated with SIR-based models, e.g., social media and call data records (CDR), which of course a high degree of uncertainty and complexity still remains [25-32]. Considering the above uncertainties involved in the advancement of SIR-based models the generalization ability are yet to be improved to achieve scalable model with high performance [33].

Due to the complexity and the large-scale nature of the problem in developing epidemiological models, machine learning has recently gained attention for building outbreak prediction models. ML has already shown promising results in contribution to advancing higher generalization ability and greater prediction reliability for longer lead-times [34-38]. Machine learning has been already recognized a computing technique with great potential in outbreak prediction. Application of ML in outbreak prediction includes several algorithms, e.g., random forest for swine fever [39] [40], neural network for H1N1 flu, dengue fever, and Oyster norovirus [41] [11] [42], genetic programming for Oyster norovirus [43], classification and regression tree (CART) for Dengue [44], Bayesian Network for Dengue and Aedes [45], LogitBoost for Dengue [46], multi-regression and Naïve Bayes for Dengue outbreak prediction [47]. Although ML methods were used in modeling former pandemics (e.g., Ebola, Cholera, swine fever, H1N1 influenza, dengue fever, Zika, oyster norovirus [11,39-48]), there is a gap in the literature for peer-reviewed papers dedicated to COVID-19. Nevertheless, machine learning has been strongly proposed as a great potential for the fight against COVID-19 [49,50]. Machine learning delivered promising results in several aspects for mitigation and prevention and have been endorsed in the scientific community for, e.g., case identifications [51], classification of novel pathogens [52], modification of SIR-based models [53], diagnosis [54,55], survival prediction [56], and ICU demand prediction [57]. Furthermore, the non-peer reviewed sources suggest numerous potentials of machine learning to fight COVID-19. Among the applications of machine learning improvement of the existing models of prediction, identifying vulnerable groups, early diagnose, advancement of drugs delivery, evaluation of the probability of next pandemic, advancement of the integrated systems for spatio-temporal prediction, evaluating the risk of infection, advancing reliable biomedical knowledge graphs, and data mining the social networks are being noted.

As stated in our former paper machine learning can be used for data preprocessing. Improving the quality of data can particularly improve the quality of the SIR-based model. For instance, the number of cases reported by Worldometer is not precisely the number of infected cases (E in the SEIR model), or calculating the number of infectious people (I in SEIR) cannot be easily determined, as many people who might be infectious may not turn up for testing. Although the number of people who are admitted to hospital and deceased wont support R as most COVID-19 positive cases recover without entering hospital. Considering this data problem, it is extremely difficult to fit SEIR models satisfactorily. Considering such challenges, for future research, the ability of machine learning for estimation of the missing information on the number of exposed E or infecteds can be evaluated. Along with the prediction of the outbreak, prediction of the total mortality rate (n(deaths) / n(infecteds)) is also essential to accurately estimate the number of potential patients with the in the critical situation and the required beds in intensive care units. Although the research is in the very early stage, the trend in outbreak prediction with machine learning can be classified in two directions.

Firstly, improvement of the SIR-based models, e.g., [53,58], and secondly time-series prediction [59,60]. Consequently, the state-of-the-art machine learning methods for outbreak modeling suggest two major research gaps for machine learning to address. Firstly, improvement of SIR-based models and secondly advancement in outbreak time series. Considering the drawbacks of the SIR-based models, machine learning should be able to contribute. This paper contributes to the advancement of time-series modelling and prediction of COVID-19. Although ML has long been established as a standard tool for modeling natural disasters and weather forecasting [61-65], its application in modeling outbreak is still in the early stages. More sophisticated ML methods are yet to be explored. A recent paper by Ardabili et al, [50], explored the potential of MLP and ANFIS in time series prediction of COVID-19 in several countries. Contribution of the present paper is to improve the quality of prediction through proposing a hybrid machine learning and compare the results with ANFIS. In the present paper the time series of the total mortality is also included.

The rest of this paper is organized as follows. Section two describes the methods and materials. The results are given in section three. Sections four presents conclusions.

2. Materials and methods

2.1 Data

Dataset is related to the statistical reports of COVID-19 cases and mortality rate of Hungary which is available at: https://www.worldometers.info/coronavirus/country/hungary/. Figure 1 and 2 presents the total and daily reports of COVID-19 statistics, respectively from 4-March to 19-April.
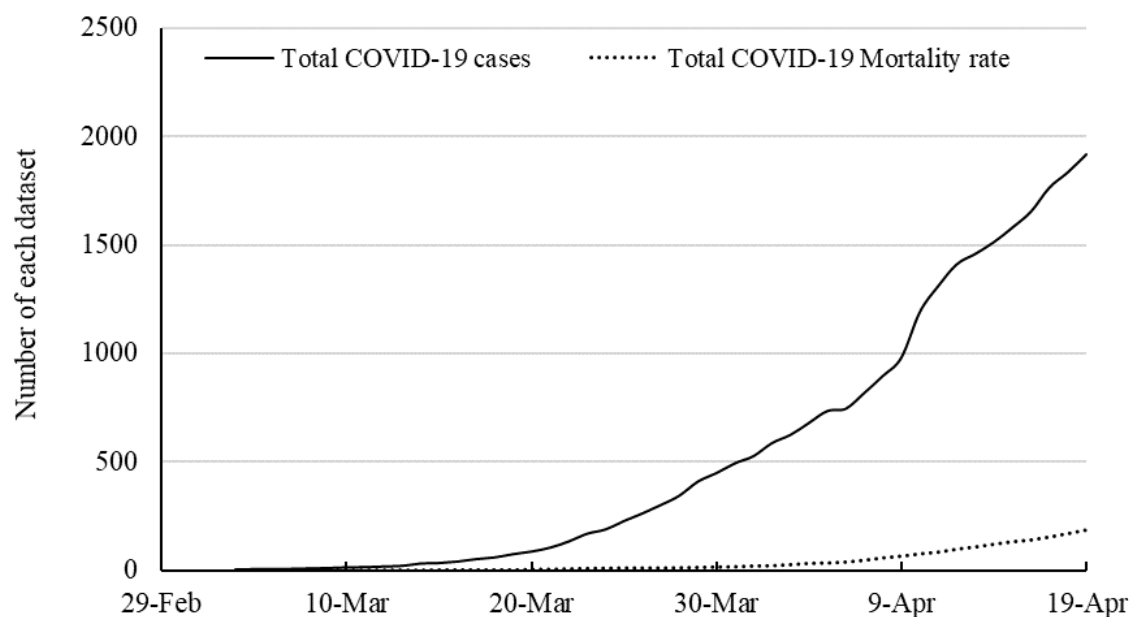

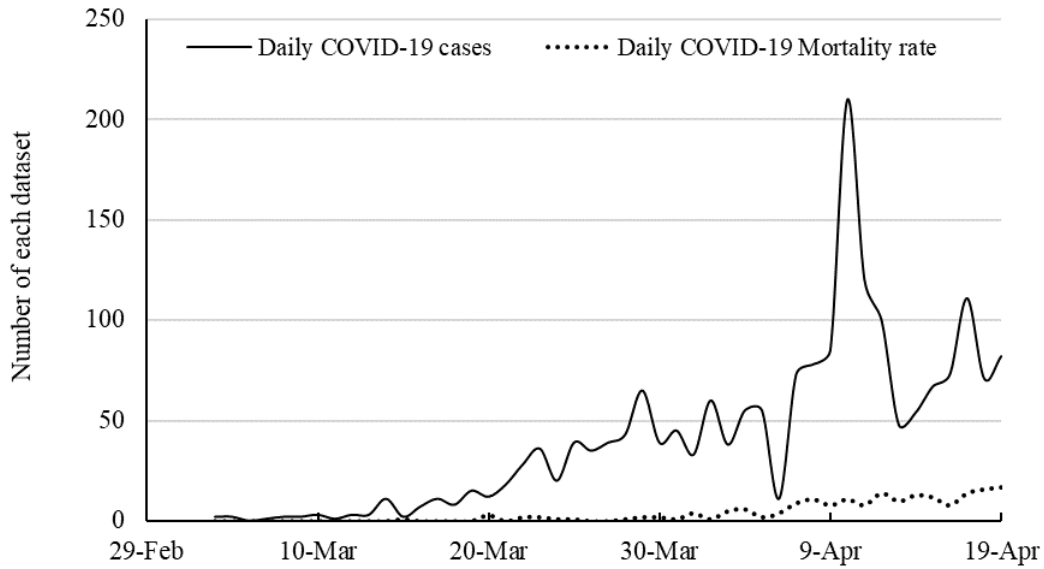Figure 1. Total statistics for the number of cases and mortality rate of COVID-19

Figure 2. Daily statistics for the number of cases and mortality rate of COVID-19

2.2 Methods and modeling strategy

In the present study, modeling is performed by machine learning methods. Training is the basis of these methods as well as many artificial intelligence (AI) methods [66,67]. According to some psychologists, humans or living things interact with their surroundings by trial and error and achieve the best performance to reach a goal. Based on this theory and using the ability of computers to repeat a set of instructions, these conditions can be provided for computer programs to interact with the environment by updating values and optimizing functions, according to the results of interaction with the environment to solve a problem or achieve a specific goal. How to update values and parameters in successive repetitions by a computer is called a training algorithm [68-70]. One of these methods is Neural Networks (NN), according to which the modeling of the connection of neurons in the human brain, software programs were designed to solve various problems. To solve these problems, operational NN such as classification, clustering, or function approximation are performed using appropriate learning methods [71,72]. The training of the algorithm is the initial and the important step for developing a model [73,74]. Developing a predictive AI model requires a dataset categorized into two sections i.e. input(s) (as independent variable(s)) and output(s) (as dependent variable(s)) [75].

In the present study, time-series data have been considered as the independent variables for the prediction of COVID-19 cases and mortality rate (as dependent variables). Time-series dataset was prepared based on two scenarios as described in table 1. The first scenario categorizes the time-series data into four inputs for the last four consequently odd days' cases or mortality rate for the prediction of x_t as the next day's case or mortality rate, and the second scenario categorizes the time-series data into four inputs for the last four consequently even days' cases or mortality rate for the prediction of x_t as the next day's case or mortality rate.

Table 1. Two proposed scenarios for time-series prediction of COVID-19 in Hungary

	Inputs	Outputs
Scenario 1	x_{t-1}, x_{t-3}, x_{t-5} and x_{t-7}	x_t
Scenario 2	x_{t-2}, x_{t-4}, x_{t-6} and x_{t-8}	x_t

In the present study the two robust hybrid methods of ANN algorithm i.e. MLP-ICA and ANFIS have been employed for developing the required models. The dataset is devised into two parts. One part is devoted to training and the second part, i.e., 20-28 April, is devoted to model validation.

2.2.1 Hybrid multi layered perceptron-imperialist competitive algorithm (MLP-ICA)
Multi-layered-perceptron (MLP) is the frequently used ANN method for the prediction and modeling purposes. This technique as a single method provides an acceptable accuracy for prediction tasks in simple and semi-complex dataset. But, in case of doing modeling tasks in complex dataset, there is a need for more robust techniques [76,77]. For this reason, hybrid methods have been growing up [76,77]. Hybrid methods contain a predictor and one or more optimizer [76,77]. The present study develops a hybrid MLP-ICA method as a robust hybrid algorithm for developing a platform for predicting the COVID-19 cases and mortality rate in Hungary. The ICA is a method in the field of evolutionary calculations that seeks to find the optimal answer to various optimization problems. This algorithm, by mathematical modeling, provides a socio-political evolutionary algorithm for solving mathematical optimization problems [78]. Like the all algorithms in this category, the ICA constitutes a primary set of possible answers. These answers are known as countries in the ICA. The ICA gradually improves the initial responses (countries) and ultimately provides the appropriate answer to the optimization problem [78,79].

The algorithms are based on the policy of assimilation, imperialist competition and revolution. This algorithm, by imitating the process of social, economic, and political development of countries and by mathematical modeling of parts of this process, provides operators in the form of a regular algorithm that can help to solve complex optimization problems [78,80]. In fact, this algorithm looks at the answers to the optimization problem in the form of countries and tries to gradually improve these answers during a repetitive process and eventually reach the optimal answer to the problem [78,80]. In the *Nvar* dimension optimization problem, a country is an array of *Nvar × 1* length. This array is defined as follows:

$$country \; = \; [p1, p2, ..., pNvar].\tag{5}$$

To start the algorithm, the country number of the initial country is created ($N_{country}$). To select the N_{imp} as the best members of the population (countries with the lowest amount of cost function) as imperialists. The rest of the N_{col} countries, form colonies belonging to an empire. To divide the early colonies between the imperialists, imperialist owns number of colonies, the number of which is proportional to its power [78,80]. The following figure symbolically shows how the colonies are divided among the colonial powers.
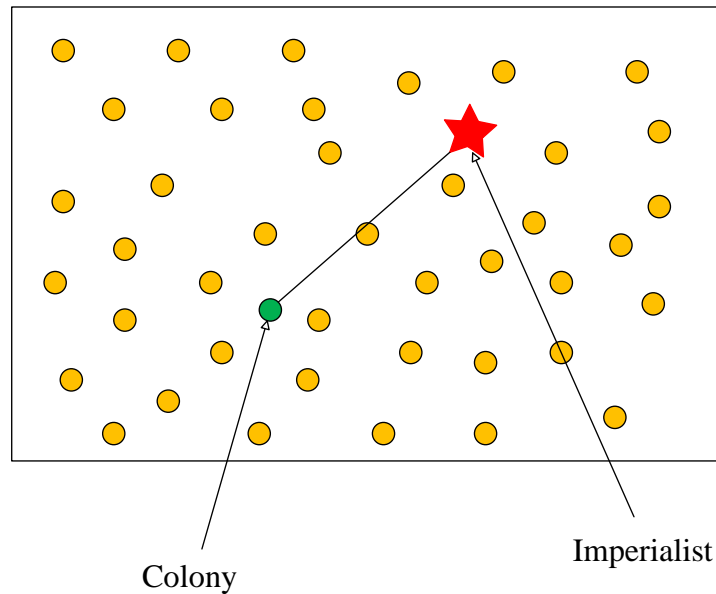
Figure 3. The initial empires generation [78]

The integration of this model with the neural network causes the error in the network to be defined as a cost function, and with the changes in weights and biases, the output of the network improves and the resulting error decreases [81]. The most important factors in training an ANN-ICA method are containing the number of countries, imperialists and decades and the number of neurons in the hidden layer which can be defined by different ways. One of the common ways for defining them is trial and error.

In the present study, inputs and outputs of each scenario where imported to the model. The number of countries, imperialists and decades have been defined by trial and error. For developing MLP model three architectures including 4-10-1, 4-14-1 and 4-18-1 (with 10, 14 and 18 neurons in the hidden layer, respectively).

2.2.2 ANFIS

An adaptive neuro-fuzzy inference system (ANFIS) is a type of artificial neural network based on the Takagi-Sugeno fuzzy system [82,83]. This method was developed in the early 1990s. Since this system integrates neural networks and concepts of fuzzy logic, it can take advantage of the capabilities of both methods. It has nonlinear functions [82,83]. Figure 4 presents the architecture of the developed ANFIS model.
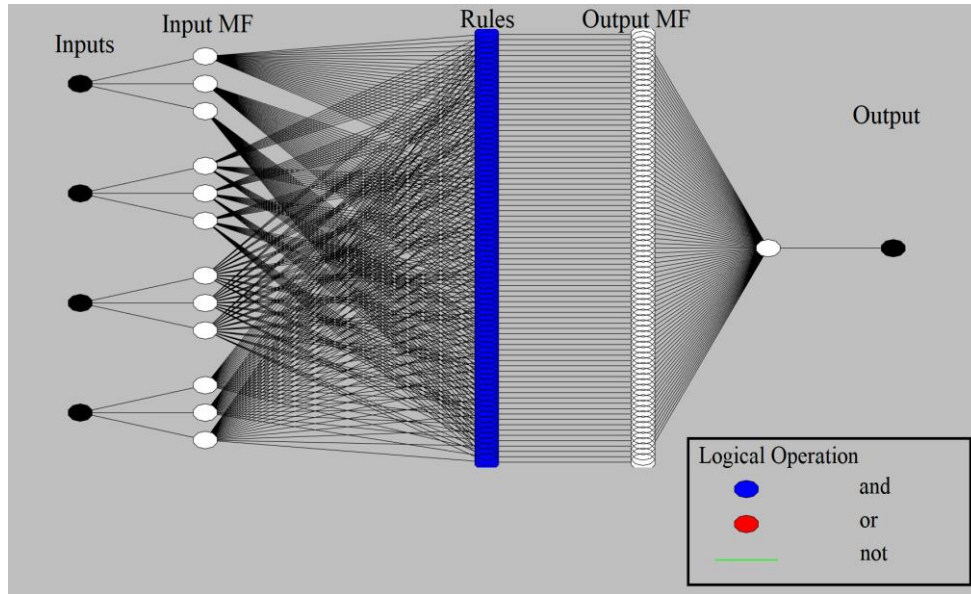
Figure 4. the architecture of the developed ANFIS

As is clear from Figure 4, ANFIS has five main layers [84-86]. The first layer is the inputs layer which takes the parameters and imports them to the model. This layer is also called as the input layer of the fuzzy system. The outputs of the first layer imports to the second layer and carries prior values of membership functions (MFs). Fuzzy rules are concluded from the nodes on second layer related degree of activity. The third layer normalizes the degree of activity of any rules. The fourth layer adopts the nodes and function and produces the outputs [87] and send them to the output layer. The important factors for determining the accuracy of ANFIS is the number and type of MFs, the optimum method and the output MF type [88-92].

ANFIS model was developed by ANFIS toolbox on MATLAB. Input parameters were independent variables of each scenario and the output variable was the number of cases or mortality rate. The ANFIS model was trained with three triangular, trapoizidal and gaussian MFs. This step was performed in order to select the best MF. The output membership function type selected linear type Because of its ability to further reduce of errors. Training of FIS was done with backpropagation optimum method and 0 value of error tolerance.

2.2.3 Evaluation criteria

Evaluations were conducted by determination coefficient, root mean square error and mean absolute percentage error values. These factors compare the target and output values and calculates a score as an index for the performance and accuracy of the developed methods [93,94]. Table 2 presents the evaluation criteria equations.

Table 2. Model Evaluation metrics

Accuracy and Performance Index	
$Determination\ coefficient = \sqrt{\dfrac{N \sum (XY) - \sum (Y) \sum (Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum XY)^2]}}}$	(6)

$$MAPE= \frac{1}{N}|\frac{X-Y}{X}| \times 100 \qquad\qquad (7)$$

$$RMSE= \sqrt{\frac{1}{N}\sum (A-P)^2} \qquad\qquad (8)$$

where N is the number of data, X and Y are, respectively, predicted (output) and desired (target) values.

3. Results

The performance of the proposed algorithm is evaluated using both training and validation data. The training data are used to train the algorithm and define the best set of parameters to be used in ANFIS and MLP-ICA. After that, the best setup for each algorithm is used to predict outbreaks on the validation samples. Worth mentioning that due to the lack of adequate sample data to avoid the overfitting, the training is used to evaluate the model with higher performance.

3.1 Training results

The training step for ANFIS was performed by employing three MF types as described in tables 3 and 4. Table 3 presents the training results of ANFIS for COVID-19 cases and Table 4 presents the training results of ANFIS for COVID-19 mortality rate in Hungary. As is clear, results have been compared using evaluation metrics of RMSE and MAPE. According to Table 3, Gaussian MF type with MF number 3, BP optimum method and linear output MF provided the highest performance by the lowest value of RMSE. On the other hand, as is clear scenario 1 provided the lowest RMSE in comparison with scenario 2 for the selected MF. Therefore, it can be concluded that, scenario 1 is suitable for modeling COVID-19 cases in comparison with scenario 2.

Table 3. ANFIS training results (cases)							
	MF type	No. of MFs	Optimum method	Output MF	Epoch No.	RMSE	MAPE
Scenario 1	Triangular	3	Back-propagation	Linear	233	248.06	38.44
	Trapezoidal	3	Back-propagation	Linear	253	158.59	35.67
	Gaussian	*3*	*Back-propagation*	*Linear*	*388*	*119.42*	*39.95*
Scenario 2	Triangular	3	Back-propagation	Linear	266	250.26	37.75
	Trapezoidal	3	Back-propagation	Linear	206	211.98	33.39
	Gaussian	*3*	*Back-propagation*	*Linear*	*285*	*182.76*	*34.6*

According to Table 4, it can be claimed that Gaussian MF provided the lowest error and highest accuracy compared with other MF types for the prediction of mortality rate. Also it can be claimed that, for the selected MF type, scenario 2 provides the highest performance compared with scenario 1 for the prediction of mortality rate.

			Table 4. ANFIS training results (mortality rate)				
	MF type	No. of MFs	Optimum method	Output MF	Epoch No.	RMSE	MAPE
Scenario 1	Triangular	3	Back-propagation	Linear	204	31.14	47.12
	Trapezoidal	3	Back-propagation	Linear	135	31.09	44.36
	Gaussian	*3*	*Back-propagation*	*Linear*	*198*	*26.42*	*39.86*
Scenario 2	Triangular	3	Back-propagation	Linear	245	32.69	45.56
	Trapezoidal	3	Back-propagation	Linear	155	31.49	41.53
	Gaussian	*3*	*Back-propagation*	*Linear*	*260*	*23.95*	*33.95*

Tables 5 and 6 presents results for the prediction of COVID-19 cases and mortality rate, respectively by MLP-ICA. According to Table 5, MLP architecture with 10 neurons in the hidden layer provided the highest accuracy for the prediction of COVID-19 cases in the presence of both scenarios. But, Scenario 2 provided higher accuracy than scenario 1 according to the lowest RMSE value.

|	Table 5. MLP-ICA training results (cases)					
	No. of countries	No. of decades	No. of initial imperialists	No. of neurons	RMSE	MAPE (%)
Scenario 1	*300*	*40*	*50*	*10*	*37.30*	*23.15*
	300	40	50	14	37.39	23.82
	300	40	50	18	37.48	23.57
Scenario 2	*300*	*40*	*50*	*10*	*37.24*	*23.47*
	300	40	50	14	37.43	23.63
	300	40	50	18	38.29	22.15

According to Table 6, neuron number 14 for scenario 1 and neuron number 18 for scenario 2 provided the highest accuracy in integrating by ICA method in comparison with other architectures. By comparing the evaluation criteria values, it can be concluded that, scenario 1 is more suitable than scenario 2 for the prediction of mortality rate using MLP-ICA.

	Table 6. MLP-ICA training results (mortality rate)					
	No. of countries	No. of decades	No. of initial imperialists	No. of neurons	RMSE	MAPE
Scenario 1	250	55	70	10	1.902	37.1
	250	*55*	*70*	*14*	*1.9*	*36.99*

	250	55	70	18	1.901	36.99
	250	55	70	10	1.902	37.16
Scenario 2	250	55	70	14	1.917	37.18
	250	*55*	*70*	*18*	*1.902*	*37.09*

Figure 3 presents the plot diagram for the selected models according to Table 3 to 6. In these figures, the vertical axis is "Target values" or in another word the actual values and the horizontal axis is the "predicted values" or in another word the output value of the model. In these plots, the dash-line is the 1:1 line. The distance of each point from the dash-line discusses the accuracy of the prediction. Such that, points on dash-line have the minimum error and increases the determination coefficient and points those that have distance from dash line increases the error and reduces the accuracy in accordance with its distance and reduces the determination coefficient.

As is clear from Figure 5, ANN provides a higher determination coefficient (0.9963, 0.9963, 0.9987 and 0.9987, respectively for the prediction of COVID-19 cases and COVID-19 mortality rate in the presence of scenario 1 and scenario (see the Figure 5)) than ANFIS. Also, the ability of ANN in the prediction of COVID-19 mortality rate is higher than that for the prediction of COVID-19 cases. But, ANFIS provides a little difference behavior than ANN. Such that, in ANFIS scenario 1 provides the highest determination coefficient for the prediction of COVID-19 cases but scenario 2 provides the highest correlation coefficient for the prediction of mortality rate (0.9689, 0.934, 0.8909 and 0.9427, respectively for the prediction of COVID-19 cases and COVID-19 mortality rate in the presence of scenario 1 and scenario (see the Figure 5)).
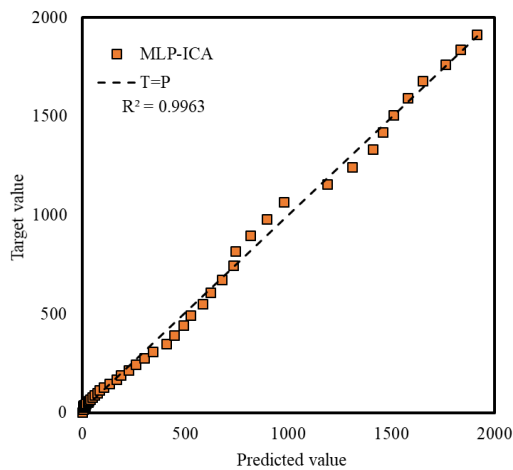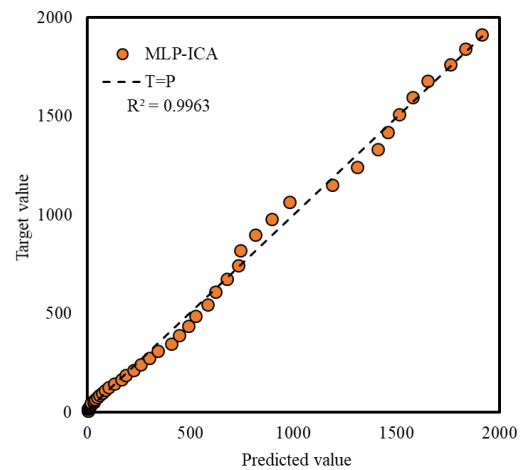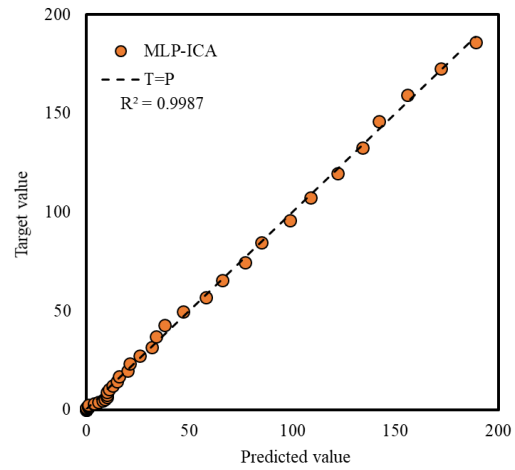

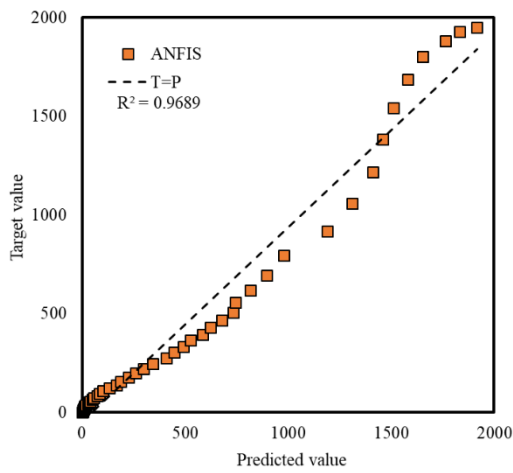Scen. 1 (Cases, MLP-ICA)   Scen. 2 (Cases, MLP-ICA)
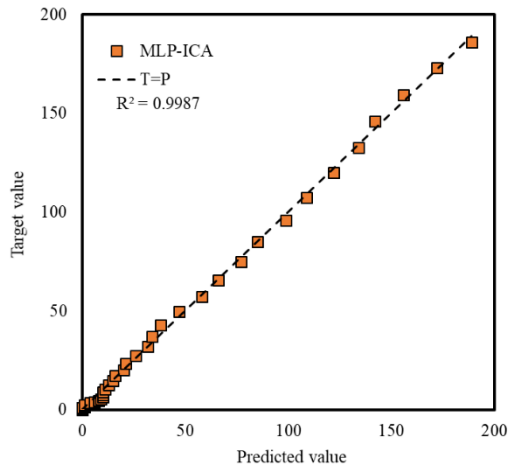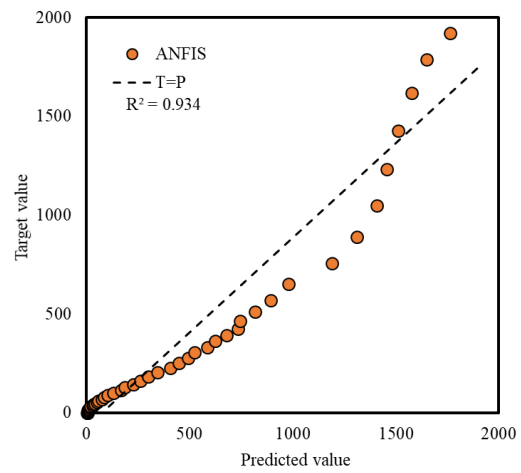
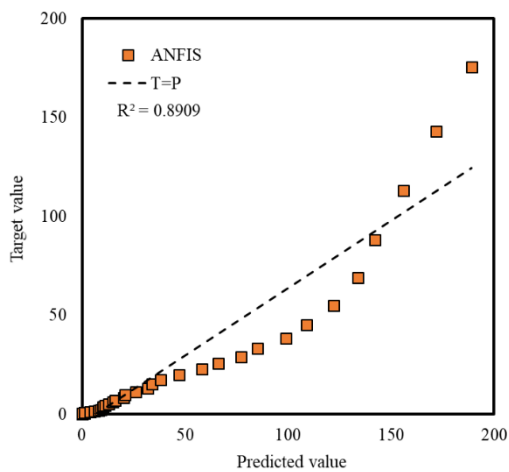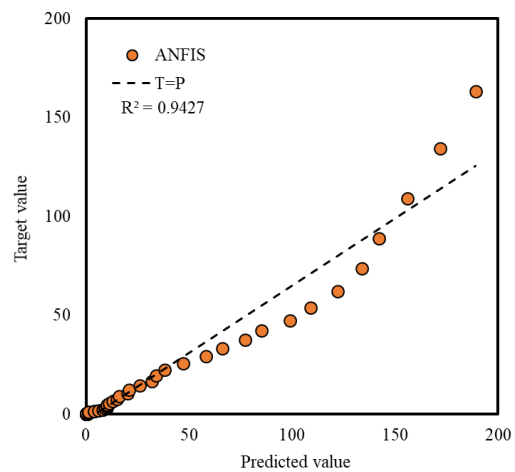Scen. 1 (Mortality, MLP-ICA)　　　　　　　Scen. 2 (Mortality, MLP-ICA)

Scen. 1 (Cases, ANFIS)　　　　　　　　Scen. 2 (Cases, ANFIS)

Scen. 1 (Mortality, ANFIS)　　　　　　　Scen. 2 (Mortality, ANFIS)

Figure 5. Plot diagram for the prediction of COVID-19 cases and mortality rate

Figure 6 presents the deviation from target values for the COVID-19 cases from 4-March to 19-April. According to Figure 6, MLP-ICA in the presence of scenario 2 provides lower deviation from target

value followed by MLP-ICA in the presence of scenario1 than the ANFIS in the presence of both scenarios.
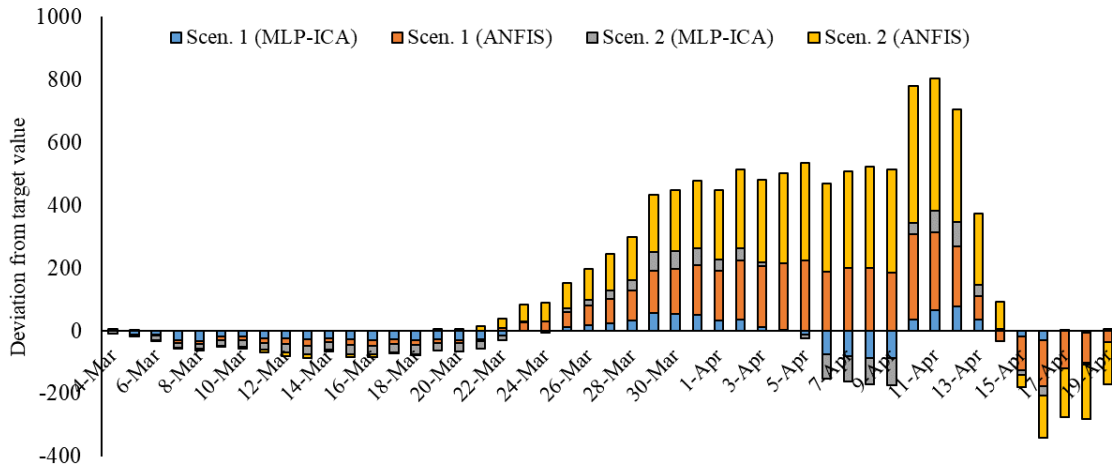

Figure 6. Deviation from target value (Cases)

Figure 7 presents the deviation from target values for the COVID-19 mortality rate from 4-March to 19-April. According to figure 7, MLP-ICA in the presence of scenario 1 provides lower deviation from target value followed by MLP-ICA in the presence of scenario2 than the ANFIS in the presence of both scenarios.
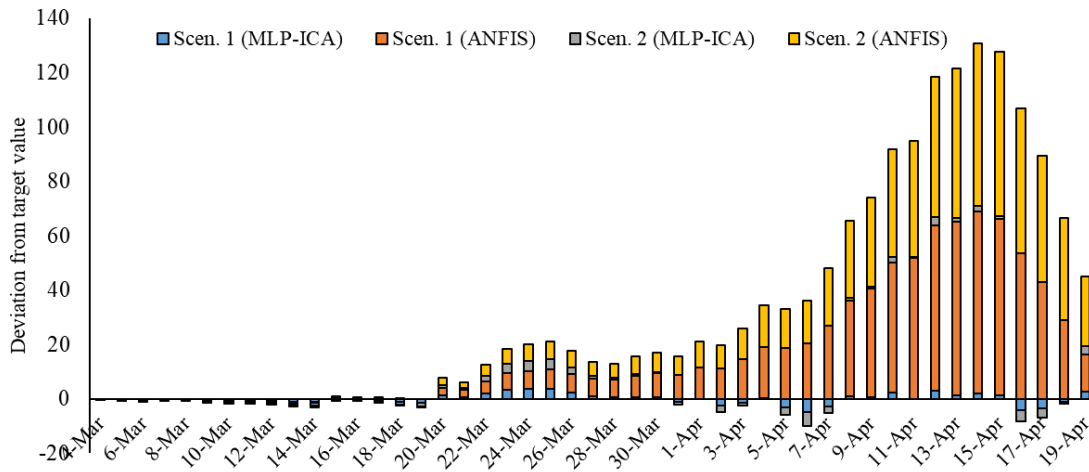

Figure 7. Deviation from target value (mortality rate)

By evaluating the modeling methods in the last sections, it was decided to select the MLP-ICA in the presence of scenario 2 for the prediction of COVID-19 outbreak and MLP-ICA in the presence of scenario 1 for the prediction of COVID-19 mortality rate in Hungary. Predictions were performed in two stages. First stage for total prediction and the second one for daily prediction. Figures 8 and 9 present total cases and total mortality rate, respectively, and Figure 10 and 11 present the daily Prediction of the results from 20-April to 30-July. Each figure has two sections including the reported

statistic (from 24-March to 19-April) and the predicted by the selected model (from 20-April to 30-July).
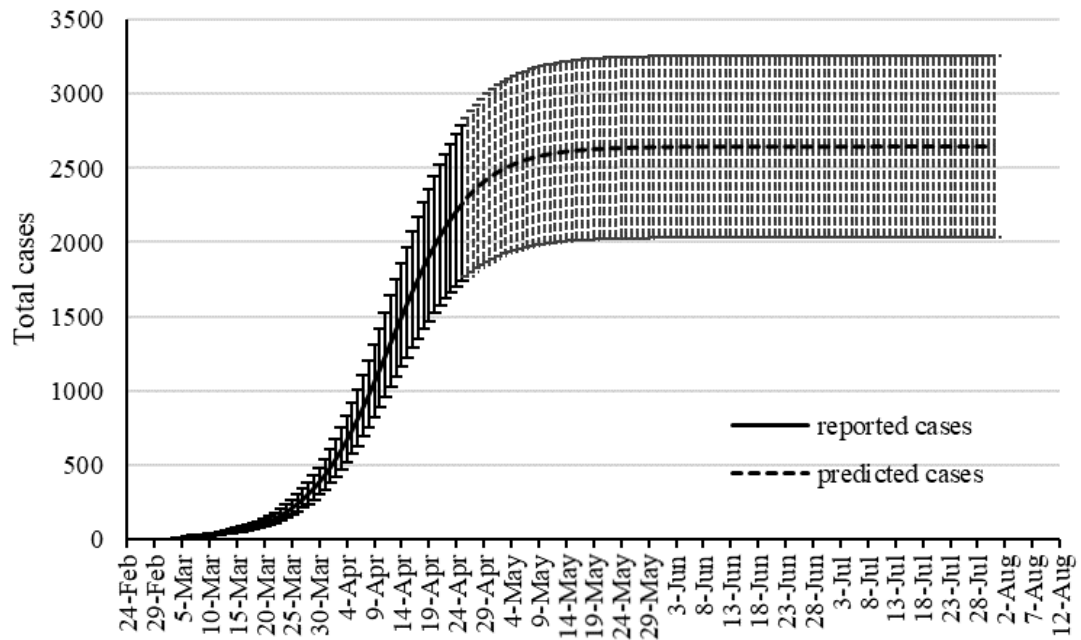

Figure 8. Total outbreak prediction for MLP-ICA
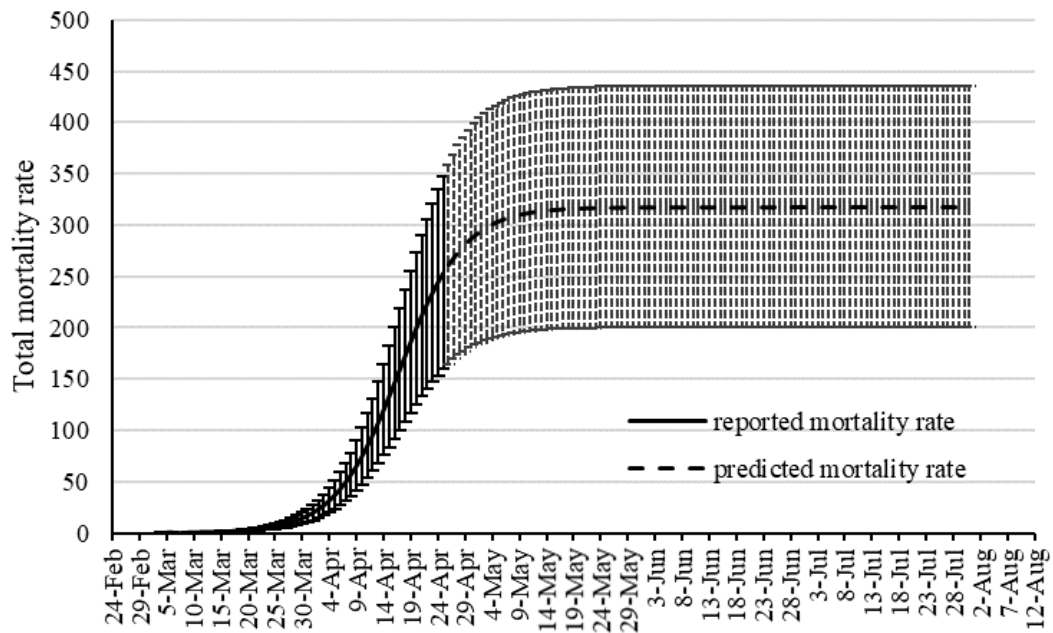
Figure 9. Total mortality rate prediction for MLP-ICA
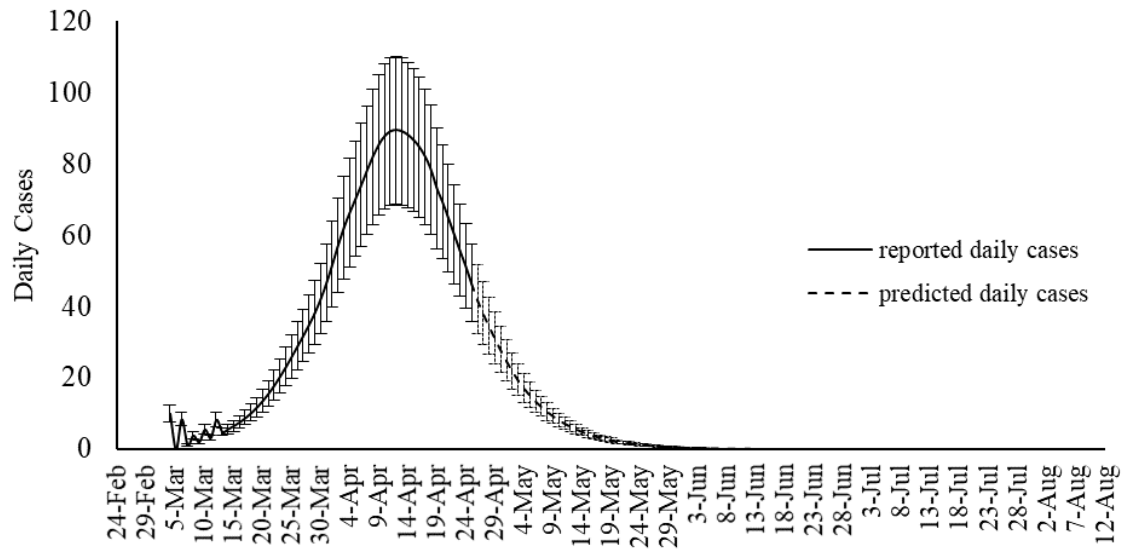
Figure 10. Daily outbreak prediction for MLP-ICA
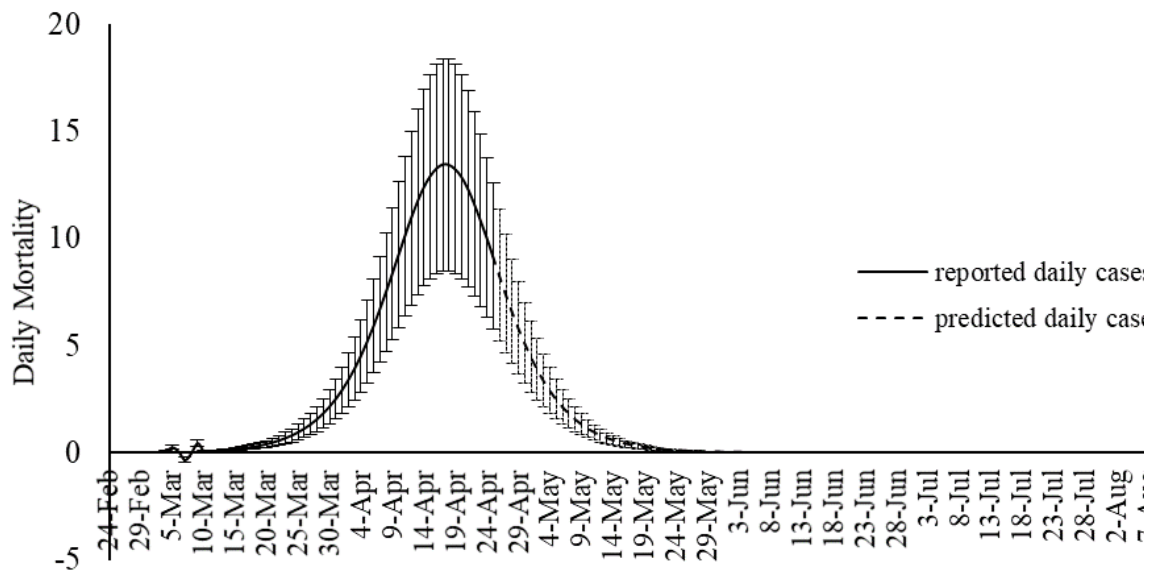
Figure 11. Daily mortality rate prediction for MLP-ICA

3.2 Validation

Table 7. Validation of the models for 9 days

	Total Cases		Total mortality rate	
For 20-28 April	MPL-ICA	ANFIS	MLP-ICA	ANFIS
RMSE	167.88	194.10	8.32	15.25
Determination coefficient	0.9971	0.9563	0.9986	0.7491

Table 7. represents the validation of the MPL-ICA and ANFIS models for the period of 20-28 April. The proposed model of MPL-ICA presented promising values for RMSE and determination coefficient for prediction of both outbreak and total mortality.

4. Conclusions

Although SIR-based models been widely used for modeling the COVID-19 outbreak, they include some degree of uncertainties. Several advancements are emerging to improve the quality of SIR-based models suitable to COVID-19 outbreak. As an alternative to the SIR-based models, this study proposed machine learning as a new trend in advancing outbreak models. The machine learning approach makes no assumption on the pandemic and spread of the infection. Instead it predicts the time series of the infected cases as well as total mortality cases.

In this study the hybrid machine learning model of MLP-ICA and ANFIS are used to predict the COVID-19 outbreak in Hungary. The models predict that by late May the outbreak and the total morality will drop substantially. Based on the results reported here, and due to the complex nature of the COVID-19 outbreak and variation in its behavior from nation-to-nation, this study suggests machine learning as an effective tool to model the outbreak. Two scenarios were proposed. Scenario 1 considered sampling the odd days and Scenario 2 used even days for training the data. Training the two machine learning models ANFIS and MLP-ICA were considered for the two scenarios. A detailed investigation was also carried out to explore the most suitable number of neurons. Furthermore, the performance of the proposed algorithm is evaluated using both training and validation data. The training data are used to train the algorithm and define the best set of parameters to be used in ANFIS and MLP-ICA. After that, the best setup for each algorithm is used to predict outbreaks on the validation samples. The validation is performed for nine days with promising results which confirms the model accuracy. In this study due to the lack of adequate sample data to avoid the overfitting, the training is used to choose and evaluate the model with higher performance. In the future research, as the COVID-19 progress in time and with the availability of more sample data further testing and validation can be used to better evaluate the models.

Both models showed promising results in terms of predicting the time series without the assumptions that epidemiological models require. Both machine learning models, as an alternative to epidemiological models, showed potential in predicting COVID-19 outbreak as well as estimating total mortality. Yet, MLP-ICA outperformed ANFIS with delivering accurate results on validation samples. Considering the availability of a small amount of training data, further investigation would be essential to explore the true capability of the proposed hybrid model. It is expected that the model maintains its accuracy as long as no major interruption occurs. For instance, if other outbreaks would initiate in the other cities, or the prevention regime changes, naturally the model will not maintain its accuracy. For the future studies advancing the deep learning and deep reinforcement learning models is strongly encouraged for comparative studies on various ML models for individual countries.

Nomenclatures

Multi-layered perceptron	MLP	Susceptible-infectious-susceptible	SIS
Adaptive network-based fuzzy inference system	ANFIS	Mean square error	MSE
Call data record	CDR	Artificial intelligence	AI
Classification and regression tree	CART	Artificial neural network	ANN
Evolutionary algorithms	EA	Susceptible-exposed-infected-recovered	SEIR
Imperialist competitive algorithm	ICA	Random Forest	RF
Maternally-derived-immunity-susceptible- exposed-infected-recovered-susceptible	MSEIRS	Maternally-derived-immunity-susceptible-exposed-infected-recovered	MSEIR

Maternally-derived-immunity-susceptible-infected-recovered	MSIR	Susceptible-exposed-infected-susceptible	SEIS
Membership function	MF	Machine learning	ML
Particle swarm optimization	PSO	Neural Networks	NN
Susceptible-infected-recovered	SIR	Root mean square error	RMSE
Susceptible-infected-recovered-deceased	SIRD	Gaussian	Gauss.

References

1. Gorbalenya, A.E.; Baker, S.C.; Baric, R.S.; de Groot Raoul, J.; Drosten, C.; Gulyaeva, A.A.; Haagmans, B.L.; Lauber, C.; Leontovich, A.M.; Neuman, B.W. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* **2020**, *5*, 536-544.

2. Chan, J.F.-W.; Yuan, S.; Kok, K.-H.; To, K.K.-W.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C.C.-Y.; Poon, R.W.-S. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* **2020**, *395*, 514-523.

3. Organization, W.H. Novel Coronavirus (2019-nCoV): situation report, 3. **2020**.

4. Organization, W.H. Coronavirus disease 2019 (COVID-19): situation report, 72. **2020**.

5. Remuzzi, A.; Remuzzi, G. COVID-19 and Italy: what next? *Lancet* **2020**.

6. Ivanov, D. Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *Transp. Res. Part E Logist. Transp. Rev.* **2020**, *136*, doi:10.1016/j.tre.2020.101922.

7. Koolhof, I.S.; Gibney, K.B.; Bettiol, S.; Charleston, M.; Wiethoelter, A.; Arnold, A.L.; Campbell, P.T.; Neville, P.J.; Aung, P.; Shiga, T., et al. The forecasting of dynamical Ross River virus outbreaks: Victoria, Australia. *Epidemics* **2020**, *30*, doi:10.1016/j.epidem.2019.100377.

8. Rypdal, M.; Sugihara, G. Inter-outbreak stability reflects the size of the susceptible pool and forecasts magnitudes of seasonal epidemics. *Nat. Commun.* **2019**, *10*, doi:10.1038/s41467-019-10099-y.

9. Scarpino, S.V.; Petri, G. On the predictability of infectious disease outbreaks. *Nat. Commun.* **2019**, *10*, doi:10.1038/s41467-019-08616-0.

10. Zhan, Z.; Dong, W.; Lu, Y.; Yang, P.; Wang, Q.; Jia, P. Real-Time Forecasting of Hand-Foot-and-Mouth Disease Outbreaks using the Integrating Compartment Model and Assimilation Filtering. *Sci. Rep.* **2019**, *9*, doi:10.1038/s41598-019-38930-y.

11. Koike, F.; Morimoto, N. Supervised forecasting of the range expansion of novel non-indigenous organisms: Alien pest organisms and the 2009 H1N1 flu pandemic. *Global Ecol. Biogeogr.* **2018**, *27*, 991-1000, doi:10.1111/geb.12754.

12. Dallas, T.A.; Carlson, C.J.; Poisot, T. Testing predictability of disease outbreaks with a simple model of pathogen biogeography. *R. Soc. Open Sci.* **2019**, *6*, doi:10.1098/rsos.190883.

13.   de Groot, M.; Ogris, N. Short-term forecasting of bark beetle outbreaks on two economically important conifer tree species. *For. Ecol. Manage.* **2019**, *450*, doi:10.1016/j.foreco.2019.117495.

14.   Kelly, J.D.; Park, J.; Harrigan, R.J.; Hoff, N.A.; Lee, S.D.; Wannier, R.; Selo, B.; Mossoko, M.; Njoloko, B.; Okitolonda-Wemakoy, E., et al. Real-time predictions of the 2018–2019 Ebola virus disease outbreak in the Democratic Republic of the Congo using Hawkes point process models. *Epidemics* **2019**, *28*, doi:10.1016/j.epidem.2019.100354.

15.   Miller, J.C. A note on the derivation of epidemic final sizes. *Bulletin of mathematical biology* **2012**, *74*, 2125-2141.

16.   Miller, J.C. Mathematical models of SIR disease spread with combined non-sexual and sexual transmission routes. *Infectious Disease Modelling* **2017**, *2*, 35-55.

17.   Maier, B.F.; Brockmann, D. Effective containment explains sub-exponential growth in confirmed cases of recent COVID-19 outbreak in Mainland China. *medRxiv* **2020**, 10.1101/2020.02.18.20024414, 2020.2002.2018.20024414, doi:10.1101/2020.02.18.20024414.

18.   Werkman, M.; Green, D.; Murray, A.G.; Turnbull, J. The effectiveness of fallowing strategies in disease control in salmon aquaculture assessed with an SIS model. *Preventive veterinary medicine* **2011**, *98*, 64-73.

19.   Osemwinyen, A.C.; Diakhaby, A. Mathematical modelling of the transmission dynamics of ebola virus. *Applied and Computational Mathematics* **2015**, *4*, 313-320.

20.   Pan, J.R.; Huang, Z.Q.; Chen, K. [Evaluation of the effect of varicella outbreak control measures through a discrete time delay SEIR model]. *Zhonghua Yu Fang Yi Xue Za Zhi* **2012**, *46*, 343-347.

21.   Zha, W.T.; Pang, F.R.; Zhou, N.; Wu, B.; Liu, Y.; Du, Y.B.; Hong, X.Q.; Lv, Y. Research about the optimal strategies for prevention and control of varicella outbreak in a school in a central city of China: Based on an SEIR dynamic model. *Epidemiol. Infect.* **2020**, 10.1017/S0950268819002188, doi:10.1017/S0950268819002188.

22.   Dantas, E.; Tosin, M.; Cunha, A., Jr. Calibration of a SEIR–SEI epidemic model to describe the Zika virus outbreak in Brazil. *Appl. Math. Comput.* **2018**, *338*, 249-259, doi:10.1016/j.amc.2018.06.024.

23.   Leonenko, V.N.; Ivanov, S.V. Fitting the SEIR model of seasonal influenza outbreak to the incidence data for Russian cities. *Russ J Numer Anal Math Modell* **2016**, *31*, 267-279, doi:10.1515/rnam-2016-0026.

24.   Imran, M.; Usman, M.; Dur-e-Ahmad, M.; Khan, A. Transmission Dynamics of Zika Fever: A SEIR Based Model. *Differ. Equ. Dyn. Syst.* **2020**, 10.1007/s12591-017-0374-6, doi:10.1007/s12591-017-0374-6.

25.   Miranda, G.H.B.; Baetens, J.M.; Bossuyt, N.; Bruno, O.M.; De Baets, B. Real-time prediction of influenza outbreaks in Belgium. *Epidemics* **2019**, *28*, doi:10.1016/j.epidem.2019.04.001.

26.   Sinclair, D.R.; Grefenstette, J.J.; Krauland, M.G.; Galloway, D.D.; Frankeny, R.J.; Travis, C.; Burke, D.S.; Roberts, M.S. Forecasted Size of Measles Outbreaks Associated With Vaccination Exemptions for Schoolchildren. *JAMA Netw Open* **2019**, *2*, e199768, doi:10.1001/jamanetworkopen.2019.9768.

27.   Zhao, S.; Musa, S.S.; Fu, H.; He, D.; Qin, J. Simple framework for real-time forecast in a data-limited situation: The Zika virus (ZIKV) outbreaks in Brazil from 2015 to 2016 as an example. *Parasites Vectors* **2019**, *12*, doi:10.1186/s13071-019-3602-9.

28.   Fast, S.M.; Kim, L.; Cohn, E.L.; Mekaru, S.R.; Brownstein, J.S.; Markuzon, N. Predicting social response to infectious disease outbreaks from internet-based news streams. *Ann. Oper. Res.* **2018**, *263*, 551-564, doi:10.1007/s10479-017-2480-9.

29.   McCabe, C.M.; Nunn, C.L. Effective network size predicted from simulations of pathogen outbreaks through social networks provides a novel measure of structure-standardized group size. *Front. Vet. Sci.* **2018**, *5*, doi:10.3389/fvets.2018.00071.

30. Bragazzi, N.L.; Mahroum, N. Google trends predicts present and future plague cases during the plague outbreak in Madagascar: Infodemiological study. *J. Med. Internet Res.* **2019**, *21*, doi:10.2196/13142.

31. Jain, R.; Sontisirikit, S.; Iamsirithaworn, S.; Prendinger, H. Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. *BMC Infect. Dis.* **2019**, *19*, doi:10.1186/s12879-019-3874-x.

32. Kim, T.H.; Hong, K.J.; Shin, S.D.; Park, G.J.; Kim, S.; Hong, N. Forecasting respiratory infectious outbreaks using ED-based syndromic surveillance for febrile ED visits in a Metropolitan City. *Am. J. Emerg. Med.* **2019**, *37*, 183-188, doi:10.1016/j.ajem.2018.05.007.

33. Reis, J.; Yamana, T.; Kandula, S.; Shaman, J. Superensemble forecast of respiratory syncytial virus outbreaks at national, regional, and state levels in the United States. *Epidemics* **2019**, *26*, 1-8, doi:10.1016/j.epidem.2018.07.001.

34. Burke, R.M.; Shah, M.P.; Wikswo, M.E.; Barclay, L.; Kambhampati, A.; Marsh, Z.; Cannon, J.L.; Parashar, U.D.; Vinjé, J.; Hall, A.J. The Norovirus Epidemiologic Triad: Predictors of Severe Outcomes in US Norovirus Outbreaks, 2009-2016. *J. Infect. Dis.* **2019**, *219*, 1364-1372, doi:10.1093/infdis/jiy569.

35. Carlson, C.J.; Dougherty, E.; Boots, M.; Getz, W.; Ryan, S.J. Consensus and conflict among ecological forecasts of Zika virus outbreaks in the United States. *Sci. Rep.* **2018**, *8*, doi:10.1038/s41598-018-22989-0.

36. Kleiven, E.F.; Henden, J.A.; Ims, R.A.; Yoccoz, N.G. Seasonal difference in temporal transferability of an ecological model: near-term predictions of lemming outbreak abundances. *Sci. Rep.* **2018**, *8*, doi:10.1038/s41598-018-33443-6.

37. Rivers-Moore, N.A.; Hill, T.R. A predictive management tool for blackfly outbreaks on the Orange River, South Africa. *River Res. Appl.* **2018**, *34*, 1197-1207, doi:10.1002/rra.3357.

38. Yin, R.; Tran, V.H.; Zhou, X.; Zheng, J.; Kwoh, C.K. Predicting antigenic variants of H1N1 influenza virus based on epidemics and pandemics using a stacking model. *PLoS ONE* **2018**, *13*, doi:10.1371/journal.pone.0207777.

39. Liang, R.; Lu, Y.; Qu, X.; Su, Q.; Li, C.; Xia, S.; Liu, Y.; Zhang, Q.; Cao, X.; Chen, Q., et al. Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data. *Transboundary Emer. Dis.* **2020**, *67*, 935-946, doi:10.1111/tbed.13424.

40. Tapak, L.; Hamidi, O.; Fathian, M.; Karami, M. Comparative evaluation of time series models for predicting influenza outbreaks: Application of influenza-like illness data from sentinel sites of healthcare centers in Iran. *BMC Res. Notes* **2019**, *12*, doi:10.1186/s13104-019-4393-y.

41. Anno, S.; Hara, T.; Kai, H.; Lee, M.A.; Chang, Y.; Oyoshi, K.; Mizukami, Y.; Tadono, T. Spatiotemporal dengue fever hotspots associated with climatic factors in taiwan including outbreak predictions based on machine-learning. *Geospatial Health* **2019**, *14*, 183-194, doi:10.4081/gh.2019.771.

42. Chenar, S.S.; Deng, Z. Development of artificial intelligence approach to forecasting oyster norovirus outbreaks along Gulf of Mexico coast. *Environ. Int.* **2018**, *111*, 212-223, doi:10.1016/j.envint.2017.11.032.

43. Chenar, S.S.; Deng, Z. Development of genetic programming-based model for predicting oyster norovirus outbreak risks. *Water Res.* **2018**, *128*, 20-37, doi:10.1016/j.watres.2017.10.032.

44. Titus Muurlink, O.; Stephenson, P.; Islam, M.Z.; Taylor-Robinson, A.W. Long-term predictors of dengue outbreaks in Bangladesh: A data mining approach. *Infect. Dis. Modelling* **2018**, *3*, 322-330, doi:10.1016/j.idm.2018.11.004.

45. Raja, D.B.; Mallol, R.; Ting, C.Y.; Kamaludin, F.; Ahmad, R.; Ismail, S.; Jayaraj, V.J.; Sundram, B.M. Artificial intelligence model as predictor for dengue outbreaks. *Malays. J. Public Health Med.* **2019**, *19*, 103-108.

46.   Iqbal, N.; Islam, M. Machine learning for dengue outbreak prediction: A performance evaluation of different prominent classifiers. *Informatica* **2019**, *43*, 363-371, doi:10.31449/inf.v43i1.1548.

47.   Agarwal, N.; Koti, S.R.; Saran, S.; Senthil Kumar, A. Data mining techniques for predicting dengue outbreak in geospatial domain using weather parameters for New Delhi, India. *Curr. Sci.* **2018**, *114*, 2281-2291, doi:10.18520/cs/v114/i11/2281-2291.

48.   Mezzatesta, S.; Torino, C.; De Meo, P.; Fiumara, G.; Vilasi, A. A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis. *Comput. Methods Programs Biomed.* **2019**, *177*, 9-15, doi:10.1016/j.cmpb.2019.05.005.

49.   Alimadadi, A.; Aryal, S.; Manandhar, I.; Munroe, P.; Joe, B.; Cheng, X. Artificial Intelligence and Machine Learning to Fight COVID-19. *Physiological genomics* **2020**.

50.   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. *Available at SSRN 3580188* **2020**.

51.   Rao, A.S.S.; Vazquez, J.A. Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey in the populations when cities/towns are under quarantine. *Infection Control & Hospital Epidemiology* **2020**, 1-18.

52.   Randhawa, G.S.; Soltysiak, M.P.; El Roz, H.; de Souza, C.P.; Hill, K.A.; Kari, L. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *bioRxiv* **2020**.

53.   Yang, Z.; Zeng, Z.; Wang, K.; Wong, S.-S.; Liang, W.; Zanin, M.; Liu, P.; Cao, X.; Gao, Z.; Mai, Z. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease* **2020**, *12*, 165.

54.   Barstugan, M.; Ozkaya, U.; Ozturk, S. Coronavirus (covid-19) classification using ct images by machine learning methods. *arXiv preprint arXiv:2003.09424* **2020**.

55.   Apostolopoulos, I.D.; Mpesiana, T.A. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine* **2020**, 1.

56.   Yan, L.; Zhang, H.-T.; Xiao, Y.; Wang, M.; Sun, C.; Liang, J.; Li, S.; Zhang, M.; Guo, Y.; Xiao, Y. Prediction of survival for severe Covid-19 patients with three clinical features: development of a machine learning-based prognostic model with clinical data in Wuhan. *medRxiv* **2020**.

57.   Grasselli, G.; Pesenti, A.; Cecconi, M. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. *Jama* **2020**.

58.   Pandey, G.; Chaudhary, P.; Gupta, R.; Pal, S. SEIR and Regression Model based COVID-19 outbreak predictions in India. *arXiv preprint arXiv:2004.00958* **2020**.

59.   Liu, D.; Clemente, L.; Poirier, C.; Ding, X.; Chinazzi, M.; Davis, J.T.; Vespignani, A.; Santillana, M. A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models. *arXiv preprint arXiv:2004.04019* **2020**.

60.   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**.

61.   Nosratabadi, S.; Mosavi, A.; Duan, P.; Ghamisi, P. Data Science in Economics. *arXiv preprint arXiv:2003.13422* **2020**.

62.   Mosavi, A.; Ozturk, P.; Chau, K.-w. Flood prediction using machine learning models: Literature review. *Water* **2018**, *10*, 1536.

63. Sheikh Khozani, Z.; Sheikhi, S.; Mohtar, W.H.M.W.; Mosavi, A. Forecasting shear stress parameters in rectangular channels using new soft computing methods. *PLoS ONE* **2020**, *15*, e0229731.

64. Lorestani, Y.; Feiznia, S.; Mosavi, A.; Nádai, L. *Hybrid Model of Morphometric Analysis and Statistical Correlation for Hydrological Units Prioritization*; 2516-2314; 2020.

65. Datta, A.; Si, S.; Biswas, S. Complete Statistical Analysis to Weather Forecasting. In *Computational Intelligence in Pattern Recognition*, Springer: 2020; pp. 751-763.

66. Ardabili, S.; Mosavi, A.; Varkonyi-Koczy, A.R. Advances in machine learning modeling reviewing hybrid and ensemble methods. **2019**.

67. Zhang, D., Ashraf, M.A., Liu, Z., Peng, W.X., Golkar, M.J. and Mosavi, A., 2020. Dynamic Modeling and Adaptive Controlling in GPS-Intelligent Buoy (GIB) Systems Based on Neural-Fuzzy Networks. Ad Hoc Networks, p.102149.

68. Mosavi, A.; Ardabili, S.; Várkonyi-Kóczy, A.R. List of Deep Learning Models. *Engineering for Sustainable Future, Lecture Notes in Networks and Systems, Springer Nature Switzerland.* **2019**.

69. Choubin, B.; Moradi, E.; Golshan, M.; Adamowski, J.; Sajedi-Hosseini, F.; Mosavi, A. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of the Total Environment* **2019**, *651*, 2087-2096.

70. Ardabili, S.; Mosavi, A.; Dehghani, M.; Várkonyi-Kóczy, A.R. Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review. In Proceedings of International Conference on Global Research and Education; pp. 52-62.

71. Gundoshmian, T.M.; Ardabili, S.; Mosavi, A.; Varkonyi-Koczy, A.R. Prediction of combine harvester performance using hybrid machine learning modeling and response surface methodology. *Engineering for Sustainable Future, Lecture Notes in Networks and Systems, Springer Nature Switzerland.* **2019**.

72. Ardabili;, S.; Mosavi;, A.; Varkonyi-Koczy;, A. Systematic Review of Deep Learning and Machine Learning Models in Biofuels Research,. *Engineering for Sustainable Future, Lecture Notes in Networks and Systems, Springer Nature Switzerland* **2019**.

73. Nádai, L.; Imre, F.; Ardabili, S.; Gundoshmian, T.M.; Gergo, P.; Mosavi, A. Performance Analysis of Combine Harvester using Hybrid Model of Artificial Neural Networks Particle Swarm Optimization. *arXiv preprint arXiv:2002.11041* **2020**.

74. Nosratabadi, S.; Karoly, S.; Beszedes, B.; Felde, I.; Ardabili, S.; Mosavi, A. Comparative Analysis of ANN-ICA and ANN-GWO for Crop Yield Prediction. **2020**.

75. Sharabiani, V.; Kassar, F.; Gilandeh, Y.; Ardabili, S. APPLICATION OF SOFT COMPUTING METHODS AND SPECTRAL REFLECTANCE DATA FOR WHEAT GROWTH MONITORING. *The Iraqi Journal of Agricultural Science* **2019**, *50*, 1064-1076.

76. Gundoshmian, T.M.; Ardabili, S.; Mosavi, A.; Varkonyi-Koczy, A.R. Prediction of combine harvester performance using hybrid machine learning modeling and response surface methodology. **2019**.

77. Nosratabadi, S.; Mosavi, A.; Keivani, R.; Ardabili, S.; Aram, F. State of the art survey of deep learning and machine learning models for smart cities and urban sustainability. In Proceedings of International Conference on Global Research and Education; pp. 228-238.

78. Atashpaz-Gargari, E.; Lucas, C. Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In Proceedings of 2007 IEEE congress on evolutionary computation; pp. 4661-4667.

79. Lei, D.; Yuan, Y.; Cai, J.; Bai, D. An imperialist competitive algorithm with memory for distributed unrelated parallel machines scheduling. *International Journal of Production Research* **2020**, *58*, 597-614.

80.   Khabbazi, A.; Atashpaz-Gargari, E.; Lucas, C. Imperialist competitive algorithm for minimum bit error rate beamforming. *International Journal of Bio-Inspired Computation* **2009**, *1*, 125-133.

81.   Ahmadi, M.A.; Ebadi, M.; Shokrollahi, A.; Majidi, S.M.J. Evolving artificial neural network and imperialist competitive algorithm for prediction oil flow rate of the reservoir. *Applied Soft Computing* **2013**, *13*, 1085-1098.

82.   Chang, F.-J.; Chang, Y.-T. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. *Advances in water resources* **2006**, *29*, 1-10.

83.   Polat, K.; Güneş, S. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. *Digital Signal Processing* **2007**, *17*, 702-710.

84.   Yilmaz, I.; Yuksek, G. Prediction of the strength and elasticity modulus of gypsum using multiple regression, ANN, and ANFIS models. *International Journal of Rock Mechanics and Mining Sciences* **2009**, *46*, 803-810.

85.   Wali, W.; Al-Shamma'a, A.; Hassan, K.H.; Cullen, J. Online genetic-ANFIS temperature control for advanced microwave biodiesel reactor. *Journal of Process Control* **2012**, *22*, 1256-1272.

86.   Mostafaei, M.; Javadikia, H.; Naderloo, L. Modeling the effects of ultrasound power and reactor dimension on the biodiesel production yield: Comparison of prediction abilities between response surface methodology (RSM) and adaptive neuro-fuzzy inference system (ANFIS). *Energy* **2016**, *115*, 626-636.

87.   Keshavarzi, A.; Sarmadian, F.; Shiri, J.; Iqbal, M.; Tirado-Corbalá, R.; Omran, E.-S.E. Application of ANFIS-based subtractive clustering algorithm in soil Cation Exchange Capacity estimation using soil and remotely sensed data. *Measurement* **2017**, *95*, 173-180.

88.   Yadav, D.; Chhabra, D.; Gupta, R.K.; Phogat, A.; Ahlawat, A. Modeling and analysis of significant process parameters of FDM 3D printer using ANFIS. *Materials Today: Proceedings* **2020**, *21*, 1592-1604.

89.   Naderloo, L.; Alimardani, R.; Omid, M.; Sarmadian, F.; Javadikia, P.; Torabi, M.Y.; Alimardani, F. Application of ANFIS to predict crop yield based on different energy inputs. *Measurement* **2012**, *45*, 1406-1413.

90.   Najah, A.A.; El-Shafie, A.; Karim, O.A.; Jaafar, O. Water quality prediction model utilizing integrated wavelet-ANFIS model with cross-validation. *Neural Computing and Applications* **2012**, *21*, 833-841.

91.   Mohandes, M.; Rehman, S.; Rahman, S. Estimation of wind speed profile using adaptive neuro-fuzzy inference system (ANFIS). *Applied Energy* **2011**, *88*, 4024-4032.

92.   Ekici, B.B.; Aksoy, U.T. Prediction of building energy needs in early stage of design by using ANFIS. *Expert Systems with Applications* **2011**, *38*, 5352-5358.

93.   Faizollahzadeh_Ardabili; Mahmoudi, A.; Mesri Gundoshmian, T. Modeling and simulation controlling system of HVAC using fuzzy and predictive (radial basis function ,RBF) controllers. *Journal of Building Engineering* **2016**, *6*, 301–308.

94.   Ardabili, S.; Mosavi, A.; Mahmoudi, A.; Gundoshmian, T.M.; Nosratabadi, S.; Varkonyi-Koczy, A.R. Modelling Temperature Variation of Mushroom Growing Hall Using Artificial Neural Networks. *Engineering for Sustainable Future, Lecture Notes in Networks and Systems, Springer Nature Switzerland.* **2019**.