

Hybrid Machine Learning Model Coupled with School Closure For Forecasting COVID-19 Cases in the Most Affected Countries

Yildiran Yilmaz¹  Selim Buyrukoglu² 

¹Recep Tayyip Erdogan University, Department of Computer Engineering, Rize, Turkey

²Cankiri Karatekin University, Department of Computer Engineering, Cankiri, Turkey

ABSTRACT

Coronavirus disease (Covid-19) caused millions of confirmed cases and deaths worldwide since first appeared in China. Forecasting methods are essential to take precautions early and control the spread of this rapidly expanding pandemic. Therefore, in this research, a new customized hybrid model consisting of Back Propagation-Based Artificial Neural Network (BP-ANN), Correlated Additive Model (CAM) and Auto-Regressive Integrated Moving Average (ARIMA) models were developed for the purpose of forecast Covid-19 prevalence in Brazil, US, Russia and India. The Covid-19 dataset is obtained from the World Health Organization website from 22 January, 2020 to 6 January, 2021. Various parameters were tested to select the best ARIMA models for these countries based on the lowest MAPE values (5.21, 11.42, 1.45, 2.72) for Brazil, the US, Russia and India, respectively. On the other hand, the proposed BP-ANN model itself provided less satisfactory MAPE values. Finally, the developed new customized hybrid model was achieved to obtain the best MAPE results (4.69, 6.4, 0.63, 2.25) for forecasting Covid-19 prevalence in Brazil, the US, Russia and India, respectively. Those results emphasize the validity of our hybrid model. Besides, the proposed prediction models can assist countries in terms of taking important precautions to control the spread of Covid-19 in the world.

Keywords: Covid-19, Neural network, ARIMA, Time series, Correlated additive model

Article History:

Received: 2021/01/15

Accepted: 2021/05/18

Online: 2021/06/30

Correspondence to: Yildiran Yilmaz,
Recep Tayyip Erdogan University,
Department of Computer Engineering
E-mail: yildiran.yilmaz@erdogan.edu.tr

INTRODUCTION

Humanity has faced an outbreak (Covid-19) first time after the second world war. The origin of Covid-19 is Wuhan, China and it was announced initially in December 2019 [1]. By 6 January 2021, the spreading of this outbreak is quite fast around the world, and more than 87.2 million people have infected over 192 countries [2]. Besides, 1.88 million people have died due to Covid-19. To date, various time-series forecasting models have played an important role in the maintenance of health public systems. Additionally, governments obtain information about the spread and consequences of infectious diseases through these models. In this sense, governments have taken some precautions to reduce the spreading speed of Covid-19 based on the results of these models, such as declare a curfew at regular intervals, quarantines etc. [3]. In the field of case forecasting, various studies have been proposed to predict Covid-19 cases. More detailed information about the proposed studies are available in the literature (Literature Review Section).

In literature, even if the proposed studies have provided promising results, to the best of our knowledge, no work has focused on a hybrid model for predicting Covid-19 cases by exploiting the impacts of school closure and interruption of face to face education in countries with the highest number of Covid-19 cases. Thus, we proposed a new hybrid model combining the ARIMA model, CAM(Correlated Additive Model) and Back-Propagation Artificial Neural Network algorithm (BP-ANN) which is one of the contributions of this study. It proposes a novel correlated additive model (CAM) in contrasts to the traditional hybrid models. In our study, correlation is calculated between the number of the actual case and school closure situation in the US, Brazil, Russia and India. Then, a correlation rate (CoR) is obtained for each country. Finally, CoR is multiplied by the additive model to create CAM. A case study was carried out before we formed the final version of the CAM. In this case, the additive model was divided and multiplied by CoR, respectively. At the end, we decided

to use CAM based on the case study in the way that the additive model is multiplied by CoR. Thus, the purpose of this paper is to build a new hybrid model based on the ARIMA model, CAM and BP-ANN to forecast Covid-19 cases in the US, Brazil, Russia and India. Moreover, no work has focused on the performance comparison of the popular machine learning algorithm, statistical model and hybrid model for prediction of the Covid-19 cases in the US, Brazil, Russia and India.

The aforementioned countries were selected since they are the most affected countries in the world by Covid-19 on 6 January 2021. More detailed information on the dataset is available in Dataset Description Section. We believe that our study helps governments and health organisations in terms of reducing the spreading speed of Covid-19. Performance metrics of these models are compared in terms of Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The contributions are listed below based on the aim of this study.

- In the initial contribution, we proposed a new BP-ANN forecast algorithm for confirmed cases of Covid-19 in the US, Brazil, Russia and India.
- The second contribution lies in developing a hybrid model for predicting covid19 cases by exploiting the impacts of school closure and interruption of face to face education in countries with the highest number of COVID cases.
- The final contribution to this study compares the most popular machine learning algorithm, statistical model and hybrid model, and then evaluates the models forecasting in multi-day ahead in the interval of one, five and ten ahead. Thus, this range of the forecasting time allows us to verify the effectiveness of the proposed forecasting models in different scenarios which helps governments and health societies in terms of reducing the spreading speed of Covid-19.

The structure of the paper is as follows: Literature Review Section provides information about the literature review. The next section introduces materials and methods which can be listed under three headings: dataset description, the theoretical background of models and the evaluation criteria of models. Results Section presents results and then the next section presents the discussion part of this study. Conclusions are drawn in the last section.

LITERATURE REVIEW

Artificial Intelligence (AI) models have been widely used both for time series and machine learning forecasting [4], [5]. A review paper (in the context of AI) is presented in [6] on the basis of Covid-19 analysis. The key aspect of AI in Covid-19 pandemic can be listed as follows: “early

detection and diagnosis of the infection”, “monitoring the treatment”, “contact tracing of the individuals”, “projection of cases and mortality”, “development of drugs and vaccines”, and also, “reducing the workload of healthcare workers” [7]. Covid-19’s data structure is a kind of non-linear data, and its management and analysis can quickly be performed using AI. We believed that AI can be used simply to assist governments in terms of the projection of upcoming cases. In this sense, forecasting can be made with the available data on confirmed cases from governments’ official website or social media etc. The forecasting would help governments about the spread and risk of the virus.

In the field of case forecasting, various studies have been presented on a daily basis. In a study, Moran’s correlation algorithm has been used to analyze the spread of the Covid-19 pandemic [8]. In another study [9], a new forecasting model is presented in order to forecast the number of positive cases in China in the upcoming ten days. Adaptive neural fuzzy inference system is combined with salp swarm and enhanced flower pollination algorithms. In a different study [10], a mathematical model was developed in terms of the short time estimation of the Covid-19 outbreak’s peak time and final size. It has been proposed to evaluate the human to human transmission outside of Wuhan, China. AI and statistical-based methods have been presented in the literature on time series problems. For instance, Artificial Neural Network (ANN) is used to forecast powder dispersion in a complex environment [11]. Moreover, a study has already been developed using an artificial neural network in order to diagnose Covid-19 using medical images [12]. Non-linear data mostly has been used to forecast weather condition, powder dispersion, bacteria population forecast and virus infection in different areas. In this sense, ARIMA is mostly used as a statistical-based method and then its performance is mostly compared to AI methods such as ANN models, neuro-fuzzy inferred systems, hybrid random vector functional link and henry gas solubility optimization [13]–[16]. Many researchers have stated that hybrid models mostly provides better performance compared to single based prediction systems [17]–[22].

In literature, various hybrid models have been proposed to predict Covid-19 cases. Ensemble empirical mode decomposition (EEMD) and artificial neural network (ANN) have been combined to forecast the upcoming Covid-19 cases in a study, [17]. In a different study, autoregressive integrated moving average (ARIMA) and wavelet forecasting models have been incorporated for short-term prediction for Covid-19 cases [18]. Moreover, the number of confirmed case and mortality rate has been predicted based on a hybrid model applying both adaptive network-based fuzzy inference system (ANFIS) and multi-layered perceptron

imperialist competitive algorithm (MLP-ICA) [19]. Discrete wavelet decomposition and ARIMA models have been applied in a hybrid model to forecast the casualties cases of Covid-19 [20]. Also, a new hybrid artificial intelligence (AI) model has been proposed applying the natural language processing (NLP) module and the long short-term memory (LSTM) network in order to predict the Covid-19 pandemic in China [22]. In a study, X-Ray images have been used for the recognition of Covid-19 disease based on a hybrid model including 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique [21].

As highlighted in the Introduction Section, to the best of our knowledge, no work has focused on a hybrid model for predicting Covid-19 cases by exploiting the impacts of school closure and interruption of face to face education in countries with the highest number of Covid-19 cases. Thus, the initial aim of this paper is to build a new hybrid model. The next section presents the background of the proposed new hybrid model.

METHODOLOGY

This section firstly describes the Covid-19 dataset employed in this paper. Subsequently, the theoretical background of data mining models are introduced and the proposed hybrid correlated additive model (CAM) is described. Then, performance metrics such as Root Mean Squared Error, Mean Absolute Error and Mean Absolute Percentage Error for the evaluation of the data mining models are described in this section.

Dataset Description

The Covid-19 prediction model flow chart in this paper is illustrated in Fig. 1. Our work based on time series was carried out to predict the COVID-19 cases in countries with the highest number of cases by January 6, 2021, namely the US, Brazil, Russia and India. From January 22, 2020 to January 6, 2021, Covid-19 cases, which are confirmed daily in the mentioned countries, have been derived from the open datasets provided from the daily reports of the World Health Organization (WHO). The dataset was collected from <https://data.world/datasets/covid-19>. The dataset of 350-day cases in the before-mentioned countries was used for experiments. Also, previous studies have shown that if time series modelling is used for prediction, the infection disease case forecasting may be more effective [23]. Therefore, time series forecasting was used in our methods as the Covid-19 cases prediction problem involve a time component. As mentioned in Introduction Section, the correlation rate is calculated based on the covid data and school status data. School status data represents whether schools are fully open, closed or partially open in the aforementioned countries as illustrated in Fig. 2.

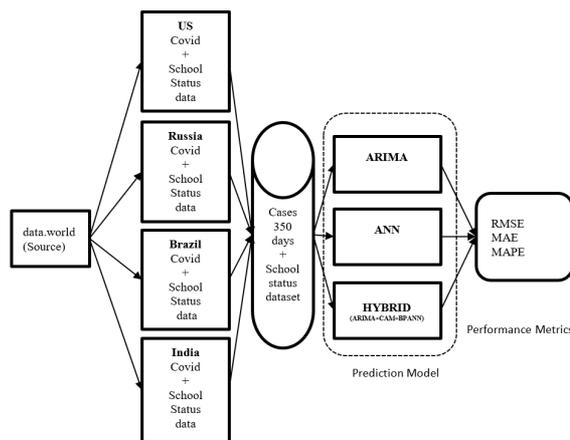


Figure 1. Infectious disease Covid-19 prediction model.

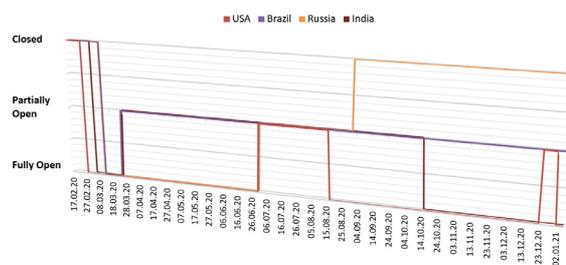


Figure 2. School status of the most affected countries during Covid-19 pandemic

The Theoretical Background of Models

The data mining models implemented in this work and the proposed hybrid correlated additive model (CAM) are described in this section.

The Autoregressive Integrated Moving Average (ARIMA)

ARIMA model was firstly presented by Box and Jenkins towards the end of the 1900s [24]. In time series analysis, the aim is to reveal predictable and significant data to forecast upcoming values of the series [25], [26]. ARIMA is one of the most widely used time series models, and also, random distributions and regular changes are taken into account in the time series. ARIMA is calculated based on three parameters: ARIMA (p,d,q), where p denotes the auto-regression order, d is the differencing degree, and q represents the moving average order [27]. In the process of model identification and parameter estimation, MSE rate was taken into account to select the top three ARIMA models. Then, the best model was selected using performance metrics (RMSE, MAE, MAPE). The best models for each country are highlighted with bold font in Table 1. Note that the Orange DB platform has been used to forecast the number of upcoming cases in the US, Brazil, Russia and India. Detailed information on the performance metrics is available in the the Evaluation Metrics of Models Section.

Back Propagation-Based Artificial Neural Network (BP-ANN)

An artificial neural network (ANN), which is especially capable of solving nonlinear problems, is a mathematical model designed to solve complex problems especially nonlinear problems [11]. In this regard, BP based neural networks (BP-ANN) are one of the most broadly employed neural network types, as they improve the accuracy of predictions [28]. Backpropagation (BP) networks use the method of returning the error amount in a feedforward network to the neurons in the hidden layer, thereby increase the success of its training. Fig. 3 shows backpropagation between neurons in a hidden layer. However, there is no backpropagation in the input and output layers in BP-ANN.

The learning rule of BP-ANN is based on continuously adjusting weights with the steepest descent method [19]. BP-ANN working principles can be described precisely as follows. Firstly, inputs (I) enter into the preconnected path. Then the input is modelled using randomly selected weights (W). Afterwards, the output for each neuron in the input, hidden and output layers are calculated. Subsequently, the network estimates the error (E) in the outputs by subtracting the obtained output (O) from the actual output (A). Finally, the network moves backwards to the hidden layer and modifies the weights in order to decrease the error (E). This process repeats itself until the E is reduced to the below threshold error value. The model with the ultimate optimal weights is created when the sum of the square error of the network reaches the specified minimum threshold.

In this study, a BP-ANN with a single hidden layer is used to forecast the number of Covid-19 cases as it is one of the best training algorithm improving prediction accuracy, consistent with the expression elsewhere [28]. Forecasting is performed with the BP-ANN algorithm developed in C sharp visual studio environment.

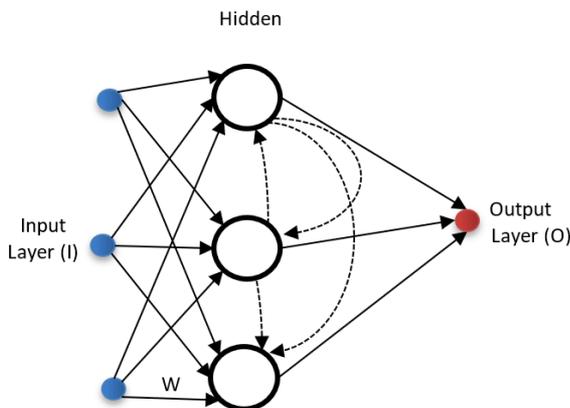


Figure 3. Backpropagation in a hidden layer

Time Series Model

The time series can be defined as a set containing the variable of observation values put in time order. This variable takes different values depending on various reasons over time [29]. Therefore, time series can be expressed as a variable value set in time order. In other words, the series obtained by performing observations at equal intervals for a certain period such as days, weeks, months, and years are called time series. For example, when the observed time series is symbolised by Y_t where $t=1,2,..,T$, then the observations can be interpreted as $Y_1, Y_2,.., Y_t$ [29].

In time series analysis, forecasting is based on analysing past data and using it to make predictions. Consequently the relationship between output (Y_t) and inputs ($Y_{t-1}; Y_{t-2}; ..; Y_{t-p}$) is as follows.

$$Y_t = a_0 + \sum_{j=1}^q a_j g \left(\beta_{0j} + \sum_{i=1}^p \beta_{ij} Y_{t-i} \right) + \varepsilon_t \tag{1}$$

This formula has the following variables: α_j where $(j = 0,1,..,q)$ and β_{ij} stand for the network connection weights; p for the number of input neurons; q for the number of hidden neurons. As the single hidden layer backpropagation network is a commonly employed network model in forecasting applications [28], our network contains a single hidden layer along with input and output layers, connected by links as shown in Fig. 4.

There are 4 input, 12 hidden and 1 output neurons connected together in the introduced Back Propagation-Neural Network architecture (BP-ANN) as depicted in Fig. 4. BP-ANN uses the sigmoid activation function in neurons. The dataset for BP-ANN has created with Covid- 19 cases as inputs (or features) and forecasted value as the output. BP-ANN gives the mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE) values as the evaluation metrics. BP-ANN uses backpropagation to calculate weights in order to train the introduced neural network.

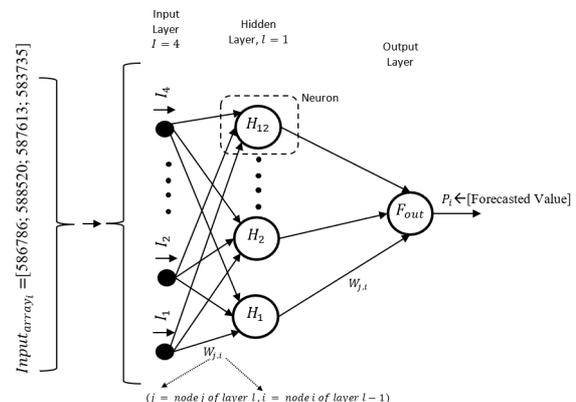


Figure 4. The architecture of the developed ANN

Proposed Hybrid Correlated Additive Model (CAM)

The proposed hybrid model is consisting of two steps. The first step analyses the linear part of the time series with the ARIMA model because ARIMA is capable of modelling the linear components in a time series. Then nonlinear time series are obtained by using ARIMA forecasted values. Once the nonlinear time series is captured, then ANN models the nonlinear components of the specified time series. In other words, the second step models the residuals from the ARIMA by employing ANN, because ANN is better at modelling the nonlinear components in a time series [11]. In order to analyse the time series, two models can be considered such as additive and multiplicative models [23]. Both models can be expressed as follows.

- Additive Model: $a_t = x_t + y_t$
- Multiplicative Model: $a_t = x_t * y_t$

In both formula, a_t stands for the actual value, x_t for linear component and y_t for the nonlinear component.

In the first step, the linear components in the time series i.e. $\{a_t, t=1,2,\dots\}$ are issued to the ARIMA to generate the predicted forecast value i.e. x_t^- . Afterwards, to obtain nonlinear components termed as N_t , actual value a_t is compared with the forecasted value of the linear component x_t^- . The comparison is carried out as follows according to the additive model and multiplicative model.

- In the Additive Model: $N_t = a_t - x_t^-$
- In the Multiplicative Model: $N_t = a_t / x_t^-$

Consequently, a nonlinear time series termed as N_t can be generated based on the above formulas. It was decided that the best method for this investigation was to use the additive model along with the proposed method which multiplies correlation rate value and nonlinear components as described in Formula 2. Through the use of the proposed model, we were able to obtain nonlinear components and at the end, better-predicted forecast values. In our hybrid correlated additive model termed as CAM, we used the additive method then multiplied the results with the correlation rate as follows.

$$N_t = (a_t - x_t^-) * CoR \tag{2}$$

where the correlation rate (*CoR*) is the value that correlates Covid cases in four countries(US, Russia, Brazil and India) to whether schools are fully open, closed or partially open.

In the second step, nonlinear components i.e. $\{N_t, t=1,2,\dots\}$ are given to the ANN as inputs to generate the predicted forecast value of nonlinear components i.e. y_t^- . The combined forecasted value can be generated as follows based on the additive and multiplicative model.

- Additive Model: $a_t = x_t^- + y_t^-$
- Multiplicative Model: $a_t = x_t^- * y_t^-$

Based on the trial and error method, we used the additive model in this experiment as it gives the best results for the Covid19 dataset.

Evaluation Metrics of Models

RMSE, MAE and MAPE can be considered as model evaluation criteria in terms of accuracy [30]. The average magnitude of the error is measured by MAE, and accuracy is measured for continuous variables. On the other hand, RMSE measures the average magnitude of the error and it is considered as the most useful error metric when huge errors are undesirable (Çınaroğlu, 2017). The value of RMSE is always bigger or equals to MAE [30]. MAPE measures the average of the absolute percentage error [32]. Note that models are considered a better fit of the data when RMSE, MAE and MAPE values are low. RMSE, MAE and MAPE are calculated by the following formulas.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - A_i)^2} \tag{3}$$

$$MAE = \frac{\sum_{i=1}^N |P_i - A_i|}{N} \tag{4}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{P_i - A_i}{A_i} \right| \times 100 \tag{5}$$

Three formulas have the following variables: N stands for the number of training data; A_i for the actual value; P_i for the predicted value.

RESULTS

In this section, the performance results obtained from data mining models are illustrated with figures and tables. Different ARIMA models were created using different parameters for each country as highlighted in ARIMA Section. Then the best models were selected based on MAPE rates. The selected best ARIMA models are the ARIMA (0,2,1), ARIMA (2,2,1), ARIMA (0,1,2) and ARIMA (0,2,3) for Brazil, US, Russia and India, respectively. The models were fitted to Covid-19 data well with the smallest $MAPE_{Brazil} = 5.21$, $MAPE_{US} = 11.42$, $MAPE_{Russia} = 1.45$, and $MAPE_{India} = 2.72$ rates (see Table 1).

On the other hand, the proposed new BP-ANN algorithm also provided sophisticated MAPE rates. The rates are as follows: $MAPE_{Brazil} = 8.27$, $MAPE_{US} = 7.45$, $MAPE_{Russia} = 1.7$, and $MAPE_{India} = 4.75$. As seen in Table 1, ARIMA model performances are slightly better than the proposed

Table 1. Comparison of tested models with the proposed hybrid model

Country	Model	RMSE	MAE	MAPE
Brazil	ARIMA (0,2,1)	239299	171873	5.21
	ARIMA (3,1,2)	307737	263087	7.6
	ARIMA (2,1,0)	1526802	1312023	37.94
	ANN	511597	410623	8.27
	Hybrid (ARIMA+CAM+ANN)	215369	154686	4.69
US	ARIMA (1,2,1)	698729	630971	12.27
	ARIMA (3,1,1)	1532416	1326203	23.89
	ARIMA (2,2,1)	649569	583112	11.42
	ANN	410666	114362	7.45
	Hybrid (ARIMA+CAM+ANN)	357542	329522	6.4
Russia	ARIMA (0,1,2)	14244	12011	1.45
	ARIMA (0,2,1)	272771	210771	21.09
	ARIMA (2,1,1)	61897	51191	5.24
	ANN	52817	43065	1.7
	Hybrid (ARIMA+CAM+ANN)	6267	5284	0.63
India	ARIMA (3,2,1)	71582	49265	4.85
	ARIMA (0,2,3)	69635	47327	2.72
	ARIMA (1,2,2)	738307	360436	12.77
	ANN	47667	38182	4.75
	Hybrid (ARIMA+CAM+ANN)	58298	41564	2.25

new BP-ANN algorithm. One possible implication of this is that the MAPE rates between the ARIMA models and BP-ANN is few. Moreover, the proposed new customised hybrid model provided promising results among the ARI-

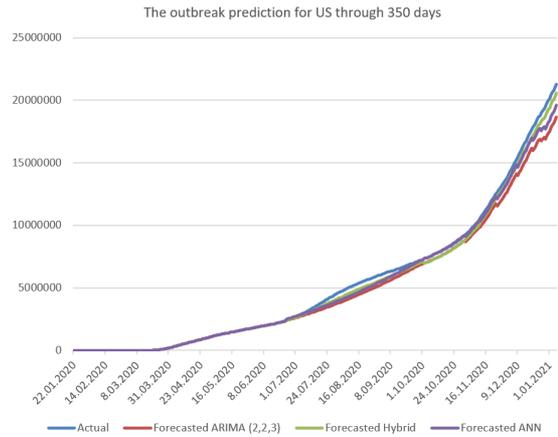


Figure 6. The Covid-19 case prediction for the US

MA and BP-ANN about MAPE rates. They are $MAPE_{Brazil} = 4.69$, $MAPE_{US} = 6.4$, $MAPE_{Russia} = 0.63$, and $MAPE_{India} = 2.25$ as seen from Table 1. Fig. 5 evaluates the three models forecasting in multi-day ahead in the interval of one, five and ten forward. These models give satisfactory prediction results. In addition, it is noted that the proposed new hybrid model can be effectively used for forecasting as a better alternative to the ARIMA and BP-ANN models.



Figure 5. Multi-day ahead predictions of models at one, five and ten forward intervals for four countries.

It is apparent from Fig. 6, 7, 8, 9, that the spread of Covid-19 slightly increases for Brazil, the US, Russia and India. In the same vein, the forecast value is also slightly increasing for both ARIMA, BP-ANN and hybrid models. While the ARIMA model forecasted a closer value to the confirmed case value for Brazil, India and Russia than the proposed BP-ANN model, the BP-ANN model forecasted a closer value to the confirmed case value for the US. Among the aforementioned machine learning models, the proposed new hybrid model forecasted the best values for these countries and gave the least error rates in terms of the evaluation metrics such as RMSE, MAE and MAPE as compared in Table 1.

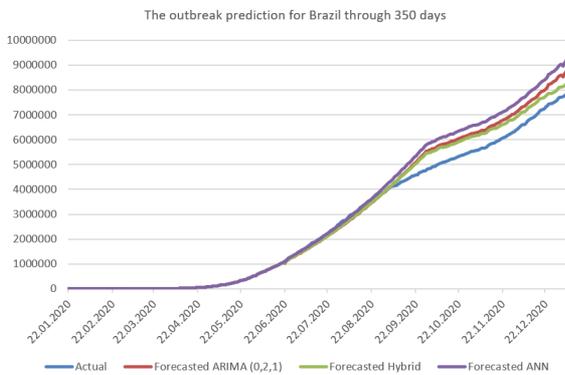


Figure 7. The Covid-19 case prediction for the Brazil

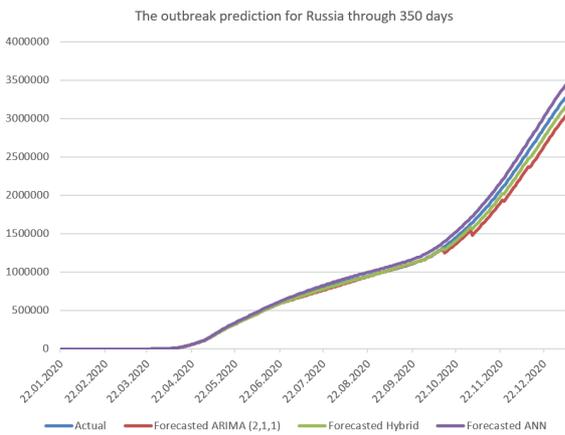


Figure 8. The Covid-19 case prediction for the Russia

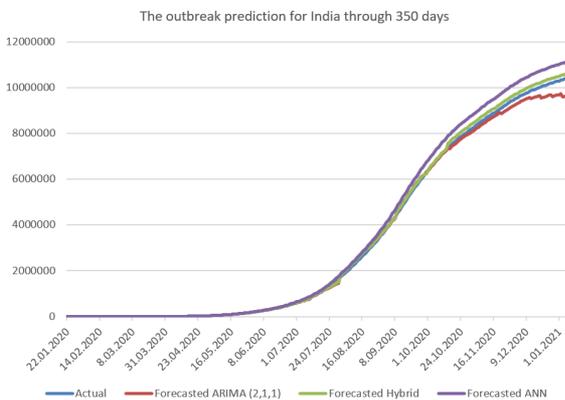


Figure 9. The Covid-19 case prediction for the India

DISCUSSION

ARIMA and BP-ANN models' well-organised construct and satisfactory forecasting performance made them popular time-series forecasting methods [33]. In this research, the current situation of the most affected four countries (Brazil, US, Russia and India) in the world by Covid-19 was highlighted, and then the number of confirmed cases in these countries were predicted with machine learning models for comparison. To the best of our knowledge, this research is the first to combine ARIMA, CAM (Correlated Additive Model), and BP-ANN models to forecast the prevalence of Covid-19 in the US, Brazil, Russia and India.

A great concern of Covid-19 is that medical treatments cannot be sufficient to treat infected patients, even if the US and Russia are the developed and rich countries. Unfortunately, the total of Covid-19 deaths in these four countries has increased dramatically during the Pandemic period. By 6 January 2021, additionally, the US, Brazil, Russia and India four worst-hit countries with 364963, 201242, 59568 and 150222 deaths, respectively. To conclude, life might return to normal as Covid-19 vaccine usage becomes widespread during this outbreak [7]. Otherwise, people may meet a great disaster because of the Covid-19 pandemic.

CONCLUSION

The following conclusions can be drawn from the present study that Covid-19 is a global and pandemic disease that can be spread rapidly; therefore, it threatens all humanity. In this regard, it is vital to predict the number of cases of this disease and to take the relevant measures immediately in countries with a high incidence. For this reason, this work was carried out to lessen the government health burden by forecasting the number of cases in countries with the highest Covid-19 outbreaks. In addition, the aim of this study was to create a more accurate prediction model than conventional models. The proposed hybrid method is an architecture that combines ARIMA, CAM (Correlated Additive Model) and the deep learning Backpropagation algorithm. In this study, two of the most commonly used time series forecasting models, ARIMA and ANN, was also used, and then they were compared with the hybrid model in terms of error rates. The forecasting performance seen in this work reveals that the Hybrid model provides the best results and secondly ARIMA gives slightly better results than the proposed BP-ANN model. We believe that the proposed hybrid model may provide more convincing results than the similar studies presented in Literature Review Section. The reason behind this is that our study proposes a novel correlated additive model (CAM) in contrasts to

the traditional hybrid models presented in Literature Review Section. Consequently, the proposed hybrid model can be used as a Covid-19 prediction application to obtain better forecasting values and take immediate measures.

CONFLICT OF INTEREST

Authors approve that to the best of their knowledge, there is not any conflict of interest or common interest with an institution/organization or a person that may affect the review process of the paper.

AUTHOR CONTRIBUTION

Public data were used in this study. The rest of all sections including conceptualisation, methodology, software, analysis, writing, review and editing were equally organised and performed by Yıldırım Yılmaz and Selim Buyrukoglu.

References

1. Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *J Med Virol.* 2020;92(4):401-402. doi:10.1002/jmv.25678
2. Worldometers.info. COVID-19 CORONAVIRUS PANDEMIC. https://www.worldometers.info/coronavirus/?utm_campaign=CSauthorbio? Published 2020. Accessed August 10, 2020.
3. Rypdal M, Sugihara G. Inter-outbreak stability reflects the size of the susceptible pool and forecasts magnitudes of seasonal epidemics. *Nat Commun.* 2019;10(1). doi:10.1038/s41467-019-10099-y
4. Davis JK, Gebrehiwot T, Worku M, et al. A genetic algorithm for identifying spatially-varying environmental drivers in a malaria time series model. *Environ Model Softw.* 2019;119(February):275-284. doi:10.1016/j.envsoft.2019.06.010
5. Scavuzzo JM, Trucco F, Espinosa M, et al. Modeling Dengue vector population using remotely sensed data and machine learning. *Acta Trop.* 2018;185(October 2017):167-175. doi:10.1016/j.actatropica.2018.05.003
6. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes Metab Syndr Clin Res Rev.* 2020;14(4):337-339. doi:10.1016/j.dsx.2020.04.012
7. Kaur SP, Gupta V. COVID-19 Vaccine: A comprehensive status report. *Virus Res.* 2020;288(August):198114. doi:10.1016/j.virusres.2020.198114
8. Kang D, Choi H, Kim JH, Choi J. Spatial epidemic dynamics of the COVID-19 outbreak in China. *Int J Infect Dis.* 2020;94(January):96-102. doi:10.1016/j.ijid.2020.03.076
9. Al-Qaness MAA, Ewees AA, Fan H, Aziz MA El. Optimization method for forecasting confirmed cases of COVID-19 in China. *Appl Sci.* 2020;9(3). doi:10.3390/JCM9030674
10. Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis.* 2020;20(5):553-558. doi:10.1016/S1473-3099(20)30144-4
11. Lauret P, Heymes F, Forestier S, Aprin L, Pey A, Perrin M. Forecasting powder dispersion in a complex environment using Artificial Neural Networks. *Process Saf Environ Prot.* 2017;110:71-76. doi:10.1016/j.psep.2017.02.003
12. Narin A, Kaya C, Pamuk Z. Department of Biomedical Engineering, Zonguldak Bulent Ecevit University, 67100, Zonguldak, Turkey. arXiv Prepr arXiv200310849. 2020.
13. Babikir HA, Elaziz MA, Elsheikh AH, et al. Noise prediction of axial piston pump based on different valve materials using a modified artificial neural network model. *Alexandria Eng J.* 2019;58(3):1077-1087. doi:10.1016/j.aej.2019.09.010
14. Elaziz MA, Elsheikh AH, Sharshir SW. Improved prediction of oscillatory heat transfer coefficient for a thermoacoustic heat exchanger using modified adaptive neuro-fuzzy inference system. *Int J Refrig.* 2019;102:47-54. doi:10.1016/j.ijrefrig.2019.03.009
15. Elsheikh AH, Sharshir SW, Abd Elaziz M, Kabeel AE, Guilan W, Haiou Z. Modeling of solar energy systems using artificial neural network: A comprehensive review. *Sol Energy.* 2019;180(January):622-639. doi:10.1016/j.solener.2019.01.037
16. Shehabeldeen TA, Elaziz MA, Elsheikh AH, et al. A Novel Method for Predicting Tensile Strength of Friction Stir Welded AA6061 Aluminium Alloy Joints Based on Hybrid Random Vector Functional Link and Henry Gas Solubility Optimization. *IEEE Access.* 2020;8:79896-79907. doi:10.1109/ACCESS.2020.2990137
17. Hasan N. A Methodological Approach for Predicting COVID-19 Epidemic Using EEMD-ANN Hybrid Model. *Internet of Things.* 2020;11:100228. doi:10.1016/j.iot.2020.100228
18. Chakraborty T, Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos, Solitons and Fractals.* 2020;135. doi:10.1016/j.chaos.2020.109850
19. Pinter G, Felde I, Mosavi A, Ghamisi P, Gloaguen R. COVID-19 Pandemic Prediction for Hungary; a Hybrid Machine Learning Approach. 2020.
20. Singh S, Parmar KS, Kumar J, Makkhan SJS. Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19. *Chaos, Solitons and Fractals.* 2020;135:1-8. doi:10.1016/j.chaos.2020.109866
21. Altan A, Karasu S. Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique. *Chaos, Solitons and Fractals.* 2020;140. doi:10.1016/j.chaos.2020.110071
22. Zheng N, Du S, Wang J, et al. Predicting COVID-19 in China Using Hybrid AI Model. *IEEE Trans Cybern.* 2020;50(7):2891-2904. doi:10.1109/TCYB.2020.2990162
23. Chae S, Kwon S, Lee D. Predicting infectious disease using deep learning and big data. *Int J Environ Res Public Health.* 2018;15(8). doi:10.3390/ijerph15081596
24. Sato RC esa. Disease management with ARIMA model in time series. *Einstein (Sao Paulo).* 2013;11(1):128-131.

- doi:10.1590/S1679-45082013000100024
25. He Z, Tao H. Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study. *Int J Infect Dis.* 2018;74:61-70. doi:10.1016/j.ijid.2018.07.003
 26. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data Br.* 2020;29:105340. doi:10.1016/j.dib.2020.105340
 27. Li X, Zhang C, Zhang B, Liu K. A comparative time series analysis and modeling of aerosols in the contiguous United States and China. *Sci Total Environ.* 2019;690:799-811. doi:10.1016/j.scitotenv.2019.07.072
 28. Wang M, Wang H, Wang J, et al. A novel model for malaria prediction based on ensemble algorithms. *PLoS One.* 2019;14(12):1-15. doi:10.1371/journal.pone.0226910
 29. Gamboa JCB. Deep Learning for Time-Series Analysis. 2017. <http://arxiv.org/abs/1701.01887>.
 30. Karasu S, Altan A, Sarac Z, Hacıoglu R. Prediction of Bitcoin prices with machine learning methods using time series data | Zaman serisi verilerini kullanarak makine öğrenmesi yöntemleri ile bitcoin fiyat tahmini. 26th IEEE Signal Process Commun Appl Conf SIU 2018. 2018:1-4. doi:10.1109/SIU.2018.8404760
 31. ÇINAROĞLU S. Sağlık Harcamasını Tahmininde Makine Öğrenmesi Regresyo Yöntemlerinin Karşılaştırılması. *Uludağ Univ J Fac Eng.* 2017;22(2):179-200. doi:10.17482/uumfd.338805
 32. Swamidass PM. Encyclopedia of Production and Manufacturing Management.; 2000. doi:10.1007/1-4020-0612-8
 33. Wang Y, Xu C, Wang Z, Zhang S, Zhu Y, Yuan J. Time series modeling of pertussis incidence in China from 2004 to 2018 with a novel wavelet based SARIMA-NAR hybrid model. *PLoS One.* 2018;13(12):1-23. doi:10.1371/journal.pone.0208404