International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)
ISSN: 2305-0543 (Online), ISSN: 2411-6173 (Print)
Vol. 12 (45), Jul. 2022, pp. 5208-5217

Analysis of Window Sizes in Prediction of Daily COVID-19 Cases using Machine Learning Models

1st* Ahmet Çifci
Department of Electrical-Electronics Engineering
Burdur Mehmet Akif Ersoy University
Burdur, Turkey
acifci@mehmetakif.edu.tr

2rd Muhammer İlkuçar
Department of Management Information Systems
Muğla Sıtkı Koçman University
Muğla, Turkey
muhammerilkucar@mu.edu.tr

***Abstract*—The first case in Turkey of the COVID-19 pandemic was reported on 11 March 2020, with the number of confirmed cases rapidly rising to over 60.000 by April 2021. In the absence of effective treatment, an important tool of outbreak management is the modeling and predicting of the COVID-19 pandemic's future and behavior. The present study considered machine learning (ML) models to predict the daily confirmed cases of the COVID-19 outbreak in Turkey. The daily confirmed cases of COVID-19 data from November 25, 2020, to June 13, 2021, were obtained from the website of the Republic of Turkey, Ministry of Health. The predicted values were explored with a range of window sizes from 2 to 14 and four ML models: Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbor (k-NN), eXtreme Gradient Boosting (XGBoost). The results show that for the test dataset, the performance of window size 6 is better than the other window sizes for the SVM and k-NN models. For the RF and XGBoost models, the performance of window sizes 8 and 10 is better than the other window sizes, respectively.**

Keywords— COVID-19, SVM, RF, k-NN, XGBoost, Turkey

I. INTRODUCTION

In this century (21st), new outbreaks of virus disease (avian influenza, Ebola, Middle East respiratory syndrome (MERS), severe acute respiratory syndrome (SARS), etc.) have occurred [1-3]. The most recent outbreak of virus disease, which was named the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or simply the coronavirus disease 2019 (COVID-19) [4], became a global problem [5, 6]. It first appeared in December 2019, in Wuhan, the capital of China's Hubei province [7].

The COVID-19 pandemic is considered the biggest global rapid risk worldwide because of millions of deaths [8]. The COVID-19 virus spread to almost all countries. Many national governments took emergency measures including quarantine, self-isolation, travel restrictions, using a face mask in public, social distancing, etc. to prevent the spread of the virus. The total number of confirmed cases worldwide exceeded 167 million patients, and the total number of deaths was recorded at about 3.5 million, according to the report dated by May 25, 2021 [9].

Turkey, with its geographically strategic location, is a transit route for people thanks to international airports. Therefore, the spread has been much faster in all of Turkey. The first case of COVID-19 in Turkey was reported on 11 March 2020. Turkey currently (May 25, 2021) has 5.2 million confirmed COVID-19 cases with 46,446 deaths. Turkey ranked fourth globally with the highest daily confirmed cases of COVID-19 with 63,082 data from the Health Ministry on April 16, 2021. Turkey reported 394 COVID-19-related deaths on April 30, 2021, its highest number of deaths since the pandemic began. Turkey entered a nearly three-week full national lockdown to bring case numbers down on April 29, 2021. After a 17-day lockdown which ended on the morning of May 17, the number of daily cases fell below 10,000.

In comparison to past viral outbreaks, the rate of contagious and spread of the virus is very quickly from human to human. Due to its rapid global spread, it is very crucial that qualified research studies should be performed for combating COVID-19. To tackle many aspects of the COVID-19 crisis, researchers in different fields of science have started studying the issue. Epidemiological [10-13], clinical [14-16], pharmaceutical [17-20], statistical [12, 21-23], and mathematical [8, 24-26] studies have conducted to examine the COVID-19 pandemic. Recently, ML methods have been applied to models various COVID-19 related outcomes such as confirmed cases, early diagnosis, and prognosis, patient outcome prediction, tracking and predicting the outbreak, drug discovery, vaccine development, etc. [27, 28]. The epidemic data of COVID-19 in China, Italy, and France between January 22 and March 15, 2020, were analyzed by Fanelli and Piazza [29]. Chimmula and Zhang [30] applied a new forecasting model using long short-term memory (LSTM) networks to forecast future COVID-19 cases in Canada. In [31], Auto-Regressive Integrated Moving Average (ARIMA), Nonlinear Auto regression Neural Network (NARNN), and LSTM models were used for forecasting the COVID-19 cases in some European countries, including Turkey. Peng and Nagata [32] utilized a support vector regression (SVR) mechanism for the prediction of COVID19 cases in the twelve countries most affected by COVID-19, including Turkey. Yeşilkanat [33] investigated the capability of the RF algorithm in terms of estimating the near future case numbers for 190 countries in the world from 23/01/2020 to 17/6/2020. In the study [34], a feed-forward neural network-based prediction model of COVID-19 in Iraq, in the time window March 2, 2020, to August 2, 2020, was proposed. Abbasimehr and Paki [35] studied models based on multi-head attention, LSTM, and convolutional neural network (CNN) with the Bayesian optimization algorithm in COVID-19 forecasting in various countries, including Turkey. Ballı [36] made time series forecasting of COVID-19 using ML methods. Another forecasting of COVID-19 cases was

Article History:
Received Date: Dec. 25, 2021, Accepted Date: Feb. 12, 2022, Available Online: Jul. 01, 2022

presented in [37]. In one very recent study, Arun Kumar et al. [38] proposed models based on recurrent neural networks (RNN) to predict cumulative COVID-19 cases (confirmed, recovered, and deaths).

As the COVID-19 epidemic outbreak continues to propagate globally, the main purpose of this study is to predict the daily confirmed cases of the COVID-19 outbreak in Turkey. Data was collected from the website of the Republic of Turkey, Ministry of Health in the time window of November 25, 2020, to June 13, 2021 [39]. Four different methods of ML namely SVM, RF, k-NN, and XGBoost were used and compared the performances with varying window sizes.

The remainder of this paper is structured as follows. Section 2 describes the dataset and gives the ML methods utilized in this paper. Section 3 discusses the results and shows a performance comparison of models. Finally, section 4 concludes the paper and outlines possible further studies.

II. Materials and Methods

A. Data Description

The COVID-19 data used in this study was obtained from the website of the Republic of Turkey, Ministry of Health [39]. The dataset consists of daily confirmed cases for 201 days from November 25, 2020, to June 13, 2021, in Turkey. Turkey was only reporting symptomatic cases since the day of the first detected case. Turkey started including asymptomatic cases in its daily count on November 25, 2020. For this reason, the starting date of our data was taken as 25 November 2020. Fig. 1 shows the time series of the daily confirmed COVID-19 cases in Turkey. As presented in Fig. 1, the day when the highest daily confirmed cases of COVID-19 appeared as 63.082 is the 143rd day (April 16, 2021). This value corresponds to the peak of the pandemic.

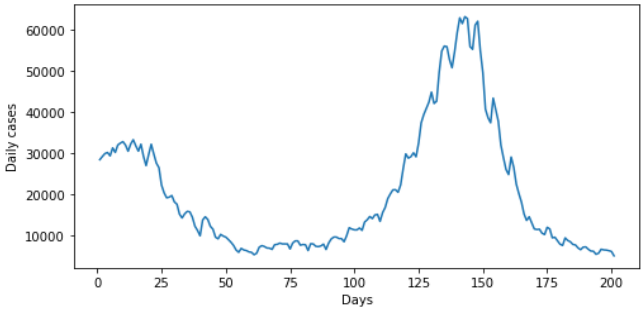


Fig. 1. Daily confirmed cases of COVID-19 in Turkey

B. Machine Learning Models

ML is referred to as a subset of artificial intelligence. It gives computers the capability to learn and improve without being explicitly programmed [40]. In the present study, SVM, RF, k-NN and XGBoost ML models would be used to predict the daily confirmed cases of the COVID-19 outbreak in Turkey.

Support vector machine (SVM)

SVM is a supervised ML method based on statistical learning theory that can be used for classification and regression analysis [41]. The basic concept of the SVM model was given by Cortes and Vapnik in 1995 [42]. The SVM models aim to obtain the optimal separating hyperplane that will separate the classes from each other. In other words, it is to maximize the distance between support vectors belonging to different classes. SVM has different models with different types of kernel functions such as linear, radial basis, sigmoid and polynomial.

Random forest (RF)

RF, proposed by Breiman in 2001, is an ensemble method combining hundreds of decision trees [43]. All of the trees in the RF are constructed independently. Every tree depends on a random vector sampled from the input data, with every other tree in the RF having similar distribution. The predictions are achieved by averaging the results of classification and regression trees using bootstrap followed by aggregation. RF models are popular due to their superior performance and low computational costs.

k-nearest neighbors (k-NN)

k-NN was developed in 1951 by Fix and Hodges [44] and later modified by Cover and Hart [45]. The k-NN algorithm is among the most basic instance-based learning algorithms. In instance-based learning algorithms, the learning process is carried out based on the data held in the training set. A newly encountered instance is classified according to the similarity between the instances in the training set [46]. In the k-NN algorithm, the instances in the training set are specified with n-dimensional numerical features. All training instances are held in an n-dimensional sample space, with each instance representing a point in n-dimensional space. When an unknown instance is encountered, the k instances that are closest to the unknown instance are determined from the training set. The class label of the new instance is assigned according to the majority vote of the class labels of its k nearest neighbors [47]. The closeness between instances is measured by distance measures such as Euclidean, Manhattan, Chebychev, Minkowski, and Hamming.

Extreme gradient boosting (XGBoost)

XGBoost is known as one of the best performing algorithms widely used for supervised learning in ML. It can be used for classification, regression and sorting tasks [48]. It is an efficient implementation of the gradient boosted trees algorithm. The aim of the XGBoost algorithm is to optimize the objective function value [49]. The most important specialty of the XGBoost algorithm is its scalability in all scenarios [50]. The algorithm avoids the overfitting of the model thanks to the feature of adjusting the regularization parameters and achieves a high success rate by controlling the complexity of the trees in the model.

III. Results and Discussion

The flowchart of the proposed system is shown in Fig. 2. The data obtained from the website of the Republic of Turkey, Ministry of Health were harmonized in a comma-separated value file (.csv). The data were examined for extreme, missing, and incorrect values and normalized in the range [0, 1]. Also, the dataset was divided into 80% training set and 20% test set.

Time series models are used to predict future data based on previous data. Sliding window time series analysis was used to prepare the database for predicting process. It is a temporary approximation over the actual value of the time

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

series data [51]. The size of the window and segment increases until it reaches the least error approach. After the first segment is selected, the next segment is selected from the end of the first segment. The process is finished when whole time series data has been segmented. Fig. 3 shows the process of sliding windows.

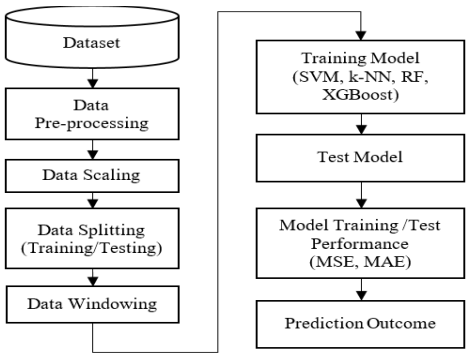


Fig. 2. Flowchart of the proposed system for predicting COVID-19

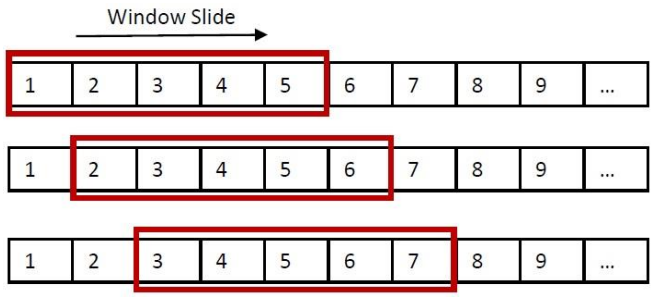


Fig. 3. Process of sliding window

How much previous data should be used in the prediction process affects the performance of the system. Fig. 4 presents the mean squared error (MSE) measures for different ML methods and window sizes.

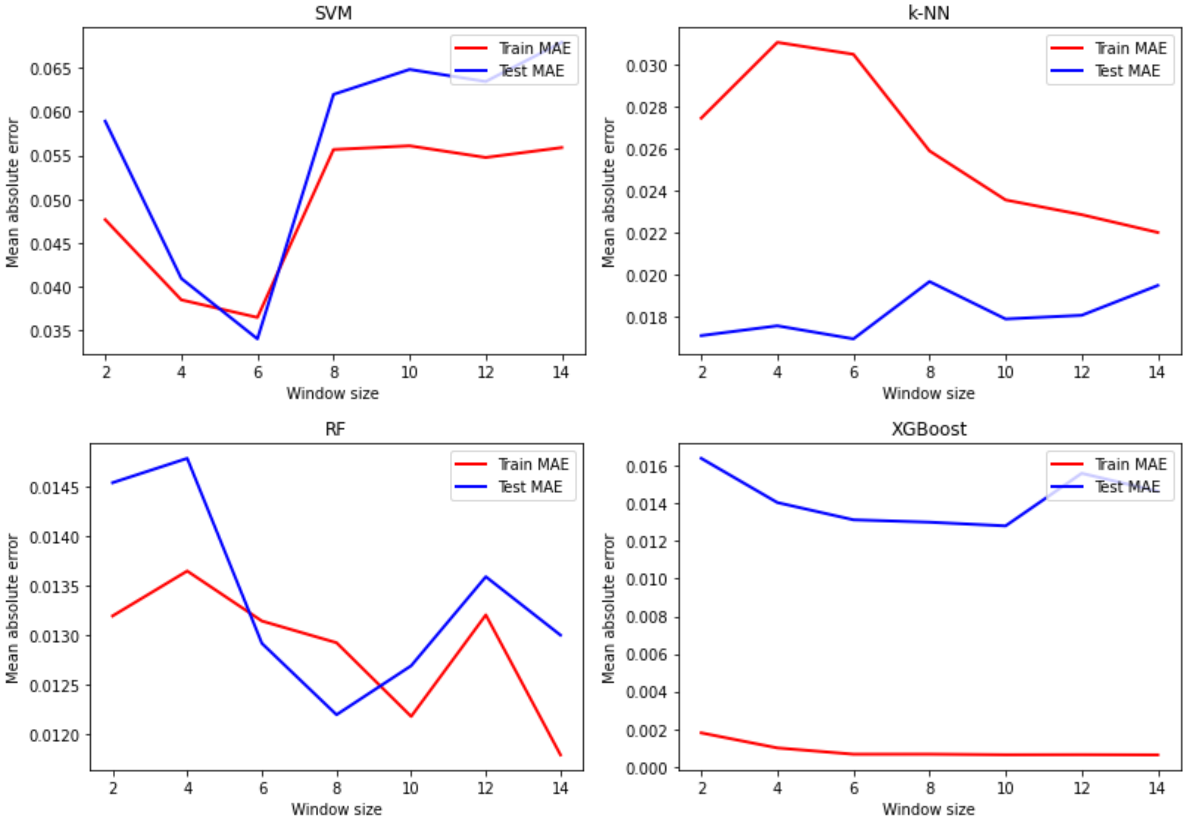


Fig. 4. Performance evaluation on different ML methods and window sizes

As seen in Fig. 4, for the test dataset, a window size of 6 shows the best performance for the SVM model, while a window size of 8 performs better in RF. This shows that the appropriate window size can play an important role in predicting. For this reason, predictions were made using different window sizes in the study, and accordingly, it was tried to find the appropriate window size for each model.

Evaluation metrics have used the quality of a model's performance. In this study, the performances of the models were analyzed using the MSE and mean absolute error (MAE) metrics. MSE and MAE can be calculated by using Eqs. 1 and 2, respectively.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (2)$$

where n, y_i and $\hat{y}_i$ indicate the number of samples, actual value and predicted value, respectively.

Tables 1 and 2 show the performance of each model by different window sizes. The values closer to zero represent the best performance of the evaluated model for the metrics MSE and MAE. According to Table 1, when the MSE and MAE metrics for different window sizes from 2 to 14 are examined, it is seen that there are very low training error values. This indicates successful learning has taken place. As shown in Table 2, when the test error value is taken into account, the error values are very close to each other in all four models. However, k-NN, RF and XGBoost models have

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

lower MAE values than SVM. Among these, the lowest MAE value of 0.01197 belongs to the RF model. Considering the effect of window size value on test performance in the study, SMV and k-NN gave the best performance with window size 6, RF achieved the best performance with window size 8 and XGBoost reached the best performance with window size 10.

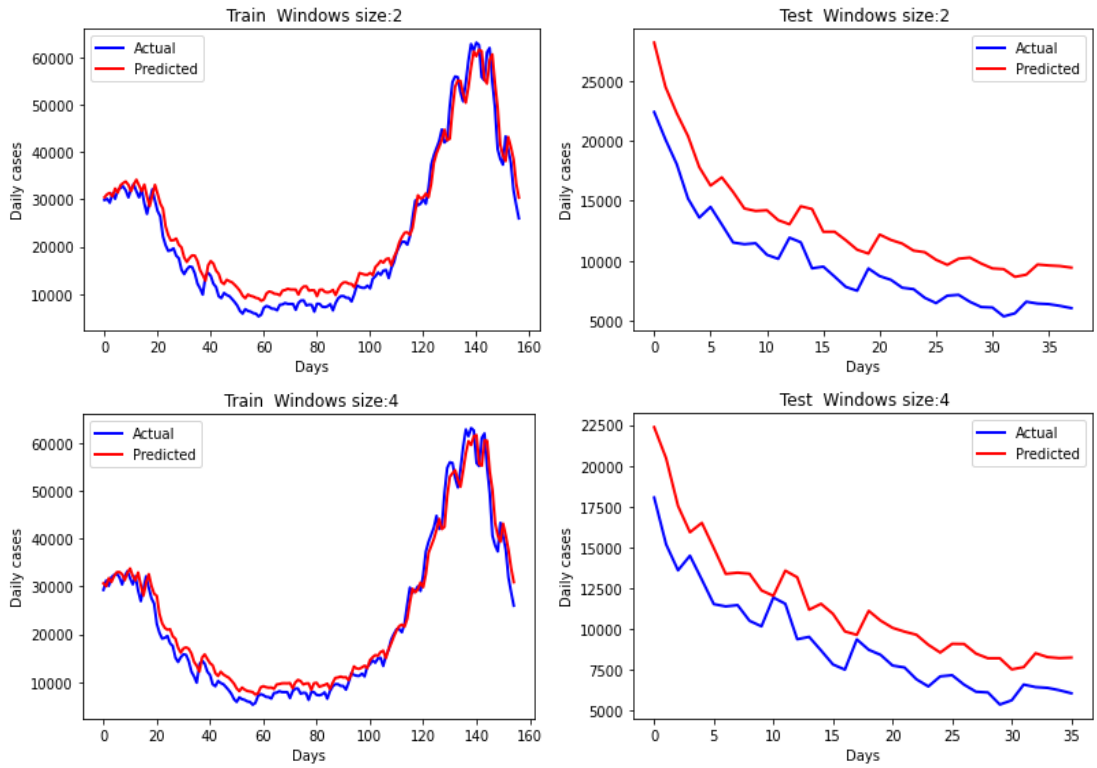
Figures 5-8 show the plot of actual and predicted data with different window sizes using different ML models.

TABLE I. ERROR METRICS FOR THE TRAINING SET OF EACH MODEL BY DIFFERENT WINDOW SIZES

	Training							
	SVM		k-NN		RF		XGBoost	
Window Size	***MSE***	***MAE***	***MSE***	***MAE***	***MSE***	***MAE***	***MSE***	***MAE***
2	0.00294	0.0476	0.00157	0.0274	0.000327	0.01217	0.000001	0.00179
4	0.00216	0.0385	0.00209	0.0311	0.000397	0.01344	0.000002	0.00099
6	***0.00204***	***0.0364***	0.00207	0.0305	0.000473	0.01429	0.000001	0.00066
8	0.00368	0.0556	0.00157	0.0259	0.000378	0.01289	0.000001	0.00066
10	0.00378	0.0561	0.00119	0.0235	0.000374	0.01253	0.000001	0.00063
12	0.00363	0.0548	***0.00118***	***0.0228***	***0.000326***	***0.01198***	0.000001	0.00063
14	0.00374	0.0559	0.00118	0.0230	0.000334	0.01181	***0.000001***	***0.00062***

TABLE II. ERROR METRICS FOR THE TEST SET OF EACH MODEL BY DIFFERENT WINDOW SIZES

	Test							
	SVM		k-NN		RF		XGBoost	
Window Size	***MSE***	***MAE***	***MSE***	***MAE***	***MSE***	***MAE***	***MSE***	***MAE***
2	0.00294	0.0589	0.000574	0.0171	0.000476	0.01525	0.000562	0.01641
4	0.00216	0.0409	0.000702	0.0175	0.000416	0.01492	0.000337	0.01405
6	***0.00204***	***0.0340***	***0.000613***	***0.0169***	0.000346	0.01312	0.000306	0.01314
8	0.00368	0.0619	0.000639	0.0196	***0.000260***	***0.01197***	0.000292	0.01301
10	0.00378	0.0046	0.000546	0.0178	0.000232	0.01215	***0.000267***	***0.01281***
12	0.00439	0.0634	0.000524	0.0180	0.000269	0.01344	0.000346	0.01561
14	0.00487	0.0679	0.000568	0.0195	0.000286	0.01344	0.000300	0.01465



IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

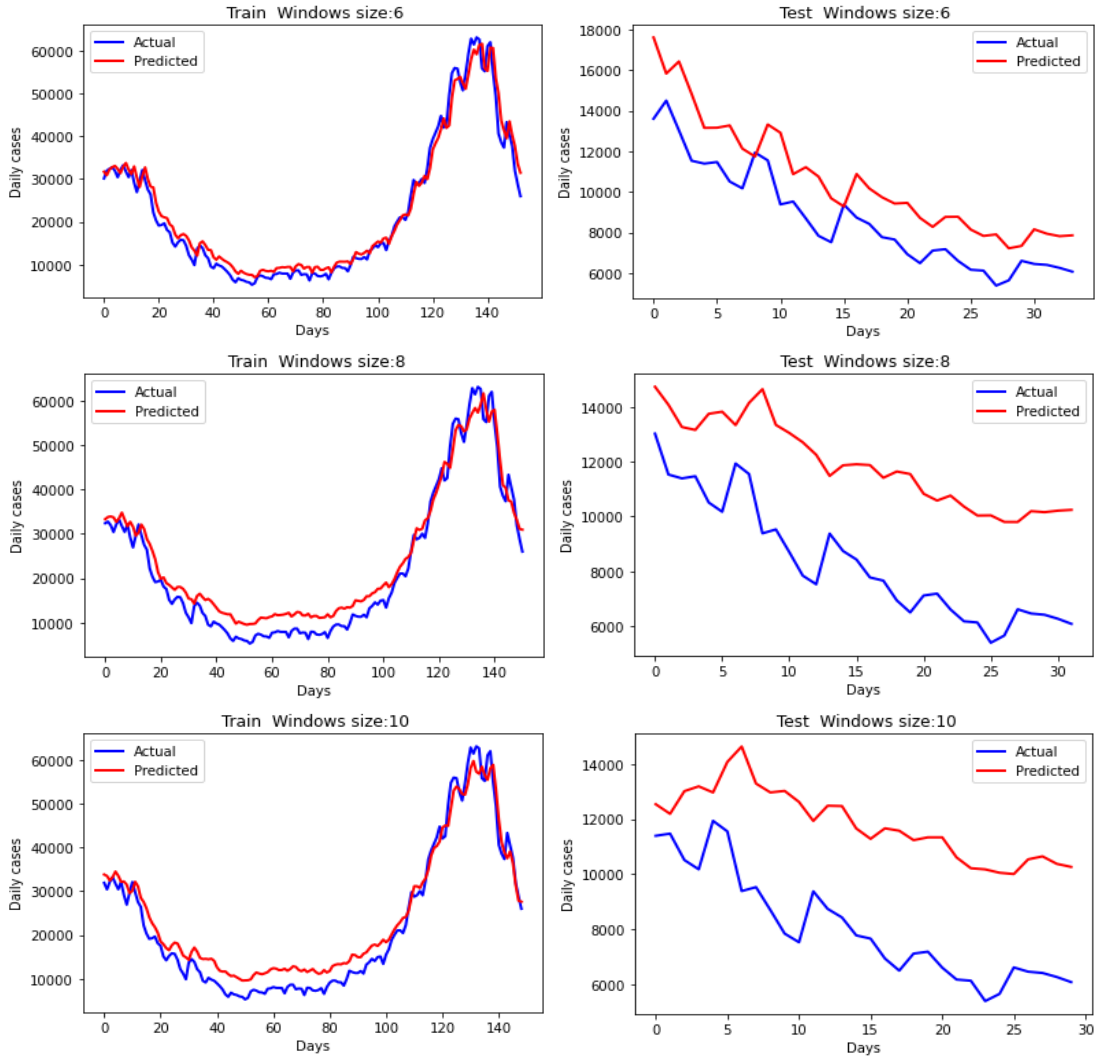
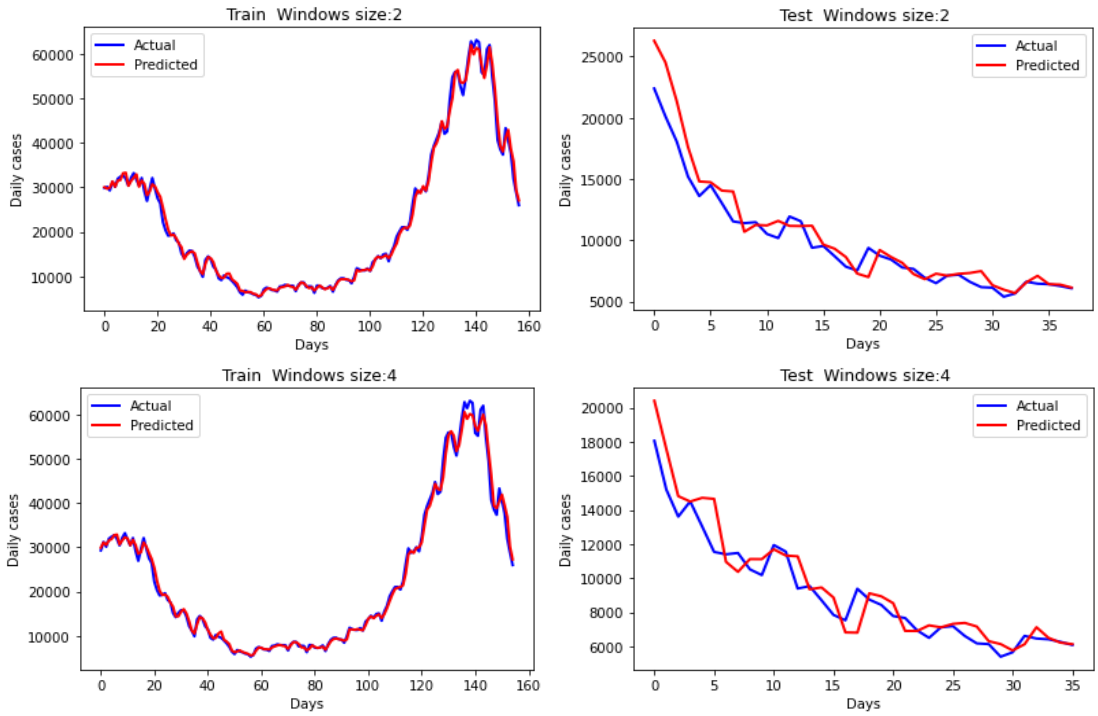


Fig. 5. Prediction of daily cases using the SVM model on different window sizes



IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

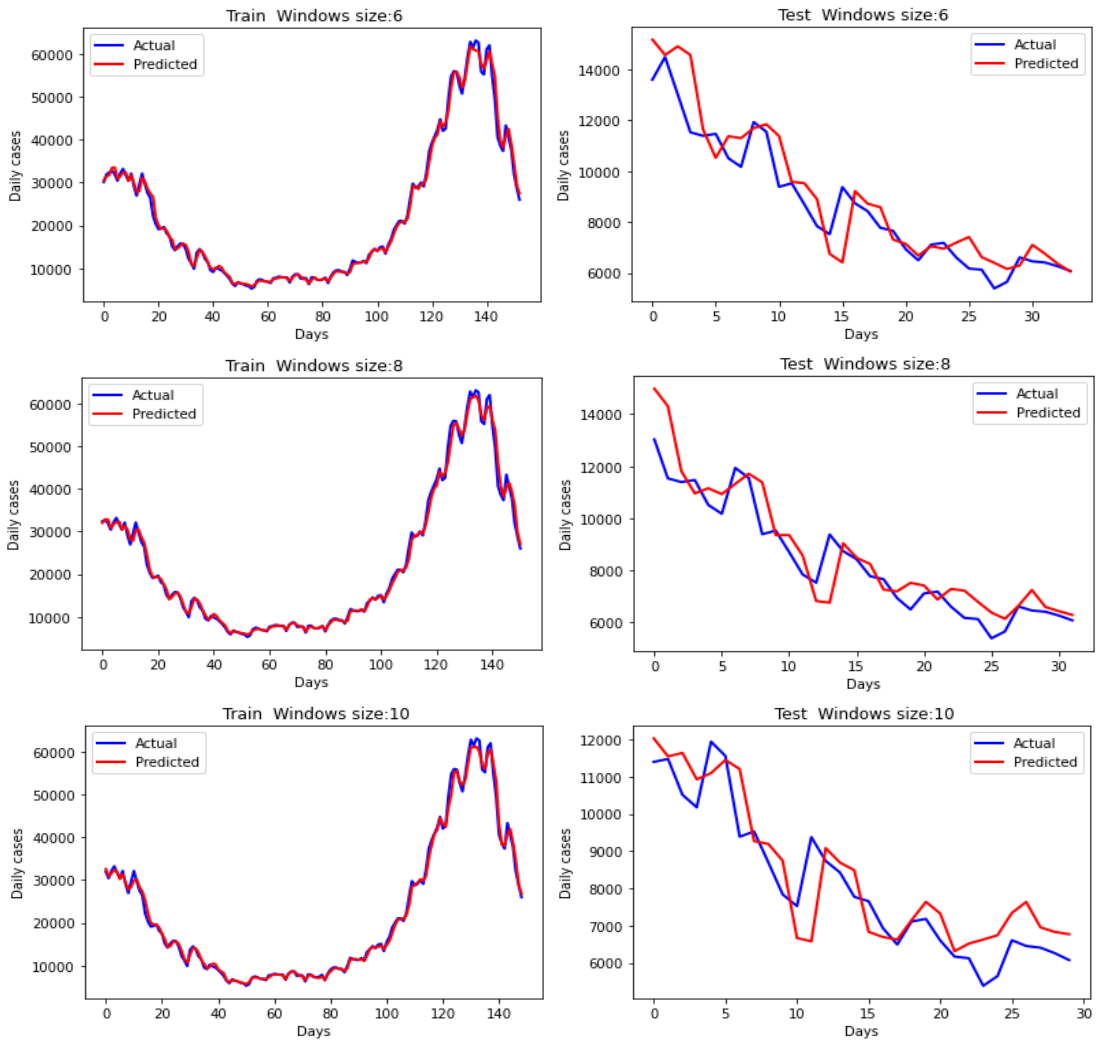
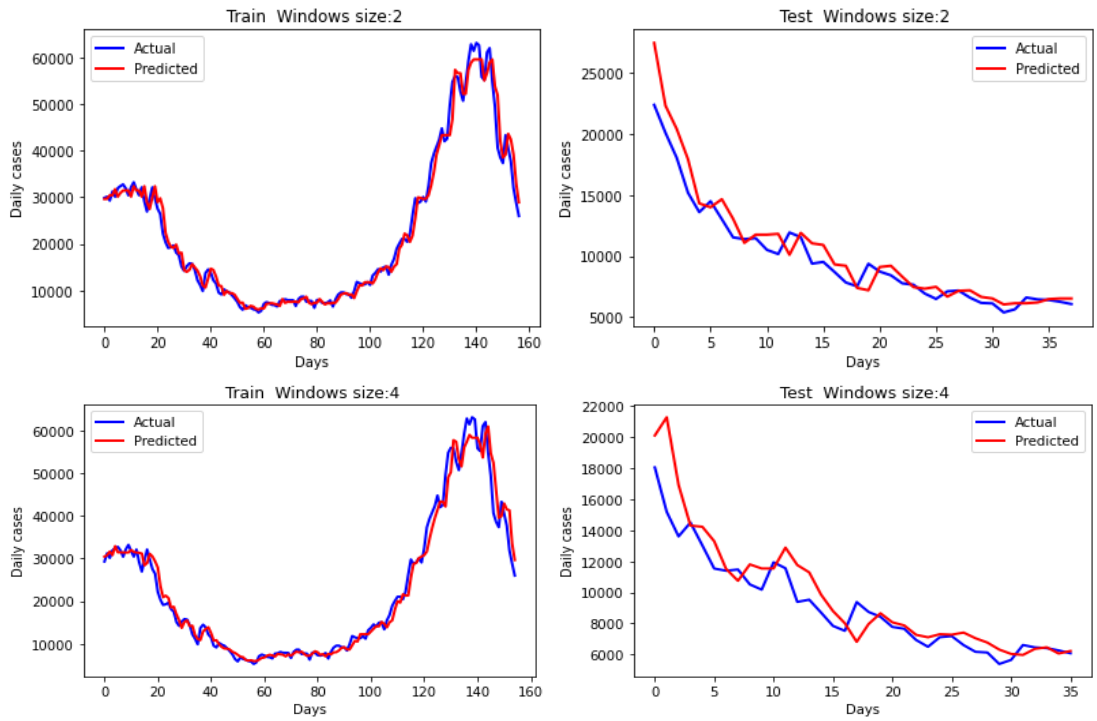


Fig. 6. Prediction of daily cases using the RF model on different window sizes



IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

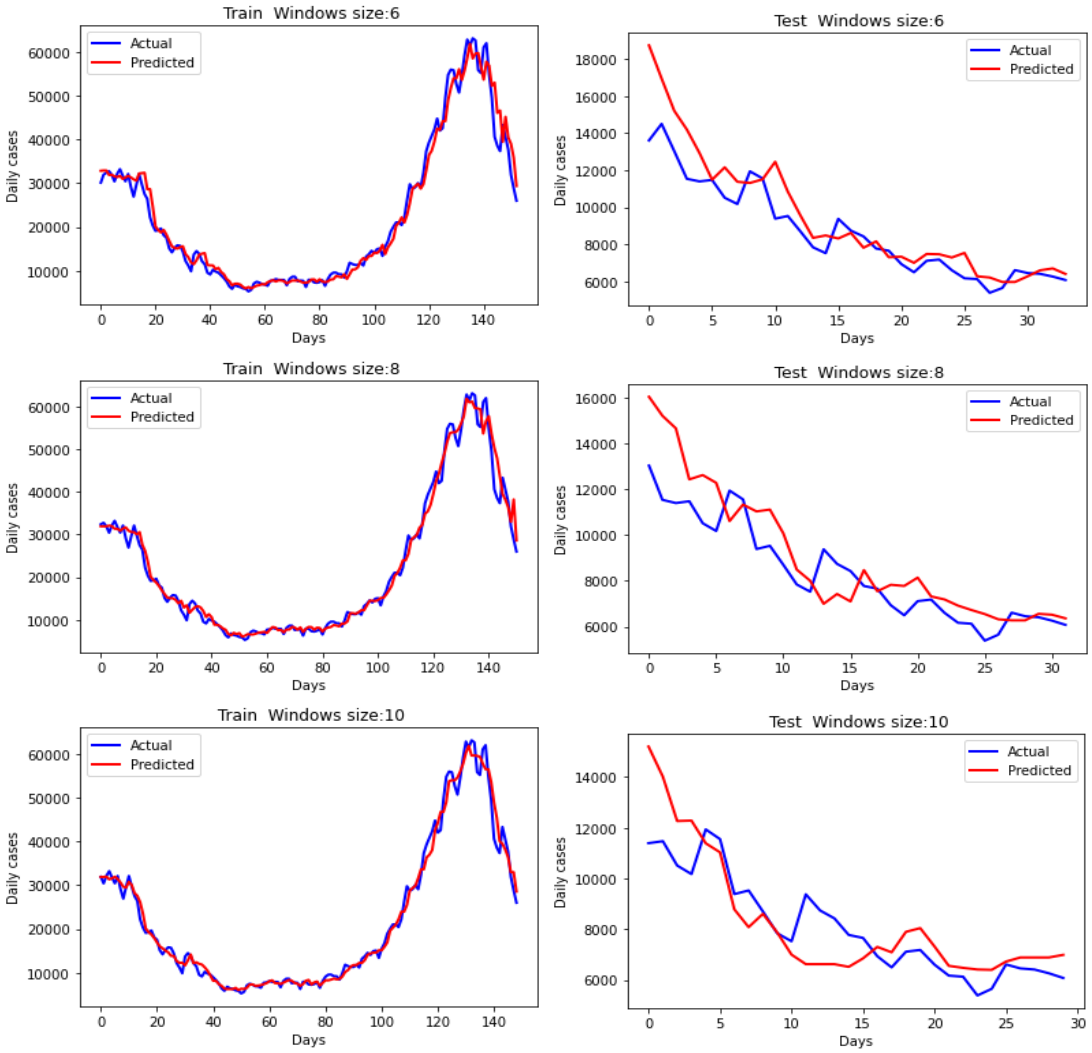
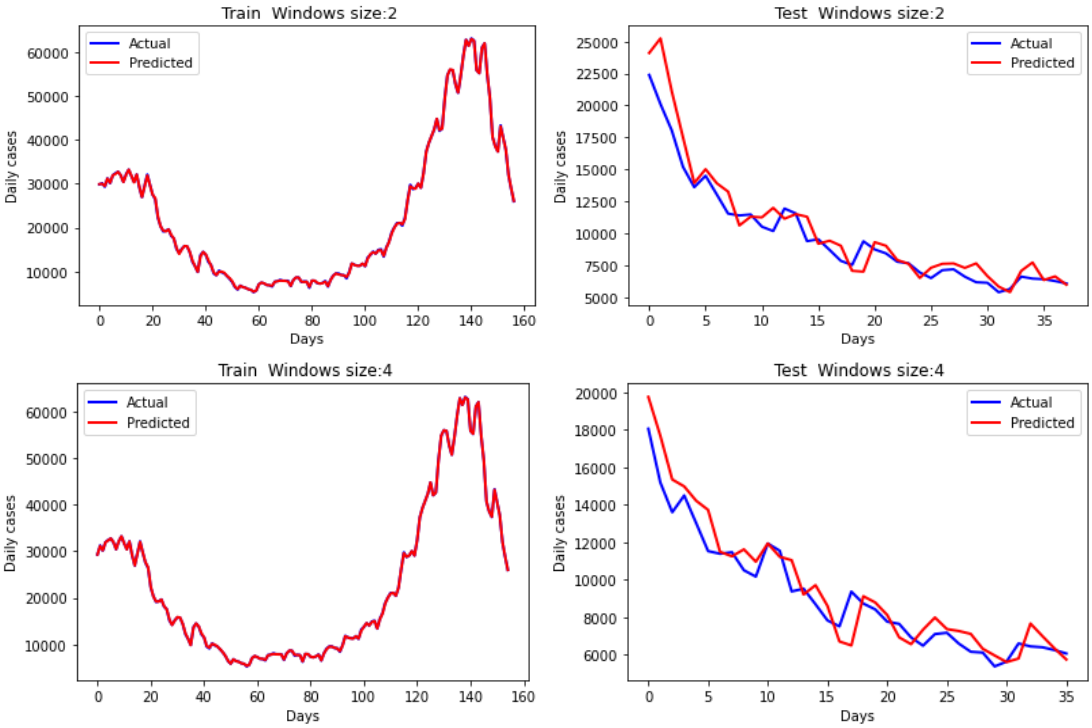


Fig. 7. Prediction of daily cases using the k-NN model on different window sizes



IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

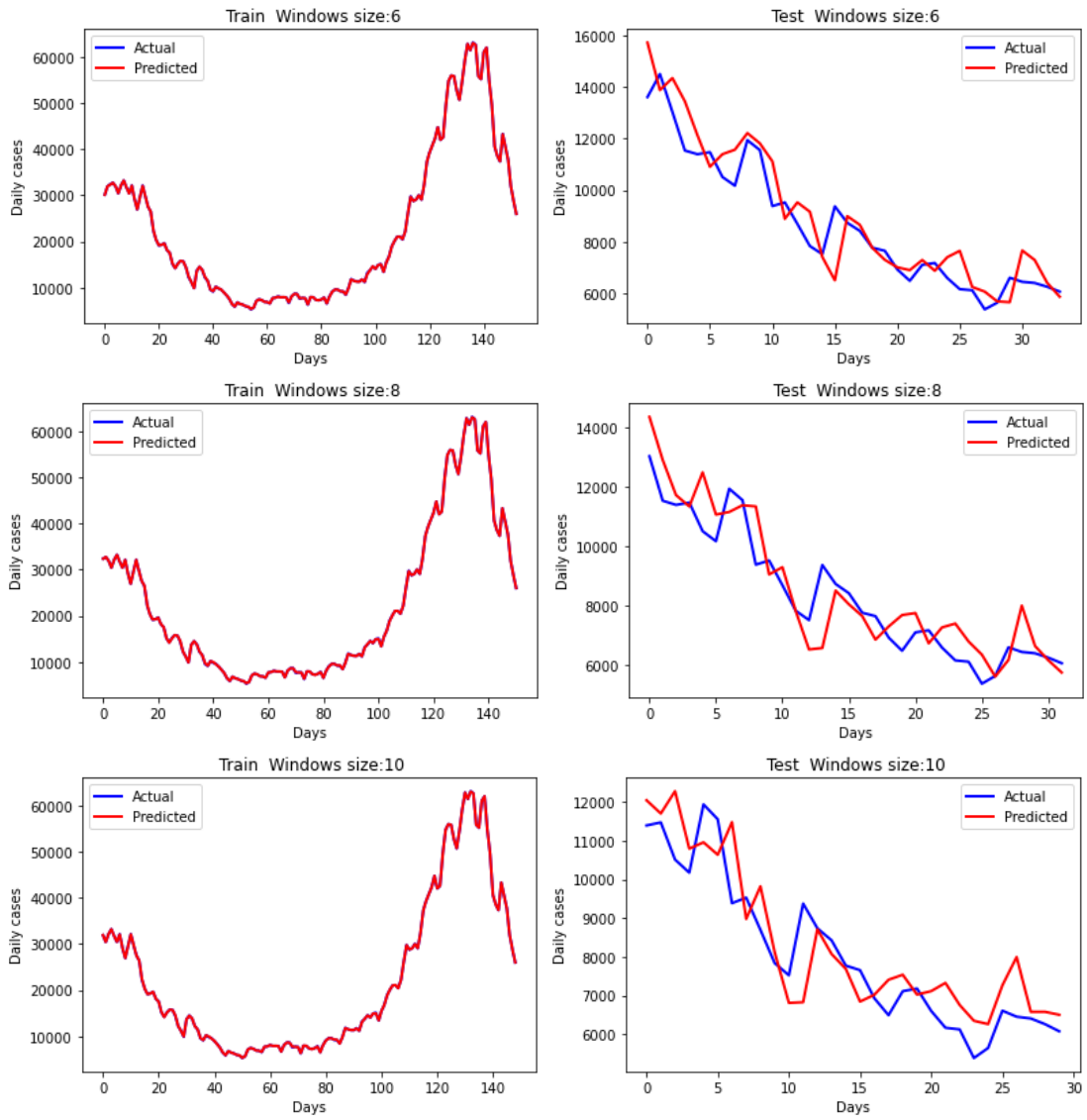


Fig. 8. Prediction of daily cases using the XGBoost model on different window sizes

IV. Conclusion

The COVID-19 pandemic has affected all aspects of society. It has widely spread to almost every country. Many studies on the early prediction of COVID-19 have been ongoing. Different types of predictive models have been applied in various studies. In this study, considering the official data reported by the Republic of Turkey, Ministry of Health, the performance of window sizes on ML models in the task of predicting the number of daily confirmed cases due to COVID-19 was investigated. SVM, RF, k-NN, and XGBoost ML models by different window sizes were employed for this purpose. The findings show that appropriate window size selection is a critical factor in prediction tasks. A poorly selected window size can mislead researchers. As future work, this study can be extended to other countries by using different ML models in a hybrid way. Moreover, different COVID-19 data such as recovered cases, and deaths can be used.

References

[1] F. R. Lashley, "Emerging infectious diseases at the beginning of the 21st Century," Online J. Issues Nurs., vol. 11, no. 1, 2006. https://doi.org/10.3912/OJIN.Vol11No01Man01

[2] E. Pelinovsky, A. Kurkin, O. Kurkina, M. Kokoulina, and A. Epifanova, "Logistic equation and COVID-19," Chaos, Solitons & Fractals, vol. 140, 2020. https://doi.org/10.1016/j.chaos.2020.110241

[3] M. Z. Ahmed, O. Ahmed, A. Zhou, H. Sang, S. Liu, L. Siyu, and A. Ahmad, "Epidemic of COVID-19 in China and associated psychological problems," Asian J. Psychiatr., vol. 51, 2020. https://doi.org/10.1016/j.ajp.2020.102092

[4] A. E. Gorbalenya, S. C. Baker, R. S. Baric, R. J. de Groot, C. Drosten, A. A. Gulyaeva, B. L. Haagmans, C. Lauber, A. M. Leontovich, B. W. Neuman, D. Penzar, S. Perlman, L. L. M Poon, D. V. Samborskiy, I. A. Sidorov, I. Sola, and J. Ziebuhr, "The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2," Nature Microbiology, vol. 5, no. 4, pp. 536-544, 2020. https://doi.org/10.1038/s41564-020-0695-z

[5] World Health Organization. Coronavirus disease (COVID-19) outbreak, https://www.who.int/emergencies/diseases/novel-coronavirus-2019; 2020 [accessed 20 May 2021]

[6] D. Cucinotta and M. Vanelli, "WHO declares COVID-19 a pandemic," Acta Biomedica: Atenei Parmensis, vol. 91, no. 1, pp. 157-160, 2020. https://doi.org/10.1016/10.23750/abm.v91i1.9397

[7] World Health Organization. Novel coronavirus in China, https://www.who.int/csr/don/12-january-2020-novel-coronavirus-China/en/; 2020 [accessed 20 May 2021]

[8] F. Ndaïrou, I. Area, J. J. Nieto, and D. F. Torres, "Mathematical modeling of COVID-19 transmission dynamics with a case study of

Wuhan," Chaos, Solitons & Fractals, vol. 135, 2020. https://doi.org/10.1016/j.chaos.2020.109846

[9] World Health Organization. WHO coronavirus disease (COVID-19) dashboard, https://covid19.who.int/; 2021 [accessed 25 May 2021]

[10] J. L. Izquierdo, J. Ancochea, J. B. Soriano, "Clinical characteristics and prognostic factors for intensive care unit admission of patients with COVID-19: a retrospective study using machine learning and natural language processing," Journal of Medical Internet Research, vol. 22, no. 10, 2020. https://doi.org/10.2196/21801

[11] Q. Ge, Z. Hu, K. Zhang, S. Li, W. Lin, L. Jin, et al., "Recurrent neural reinforcement learning for counterfactual evaluation of public health interventions on the spread of COVID-19 in the world," medRxiv 2020. https://doi.org/10.1101/2020.07.08.20149146

[12] S. Uhlig, K. Nichani, C. Uhlig, and K. Simon, "Modeling projections for COVID-19 pandemic by combining epidemiological, statistical, and neural network approaches," medRxiv 2020. https://doi.org/10.1101/2020.04.17.20059535

[13] S. Ghamizi, R. Rwemalika, L. Veiber, M. Cordy, T. F. Bissyande, M. Papadakis, et al., "Data-driven simulation and optimization for COVID-19 exit strategies," Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 2020, pp. 3434–3442. https://doi.org/10.1145/3394486.34128632020

[14] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. M. U. Din, "Machine learning based approaches for detecting COVID-19 using clinical text data," International Journal of Information Technology, vol. 12, pp. 731-739, 2020. https://doi.org/10.1007/s41870-020-00495-9

[15] M. Nemati, J. Ansary, and N. Nemati, "Machine-Learning approaches in COVID-19 survival analysis and discharge-time likelihood prediction using clinical data," Patterns (N Y), vol. 1, no. 5, 2020. https://doi.org/10.1016/j.patter.2020.100074

[16] J. S. Zhu, P. Ge, C. Jiang, Y. Zhang, X. Li, Z. Zhao, et al., "Deep-learning artificial intelligence analysis of clinical variables predicts mortality in COVID-19 patients," Journal of the American College of Emergency Physicians Open, vol. 1, no. 6, pp. 1364-1373, 2020. https://doi.org/10.1002/emp2.12205

[17] D. Ho, "Addressing COVID-19 drug development with artificial intelligence," Advanced Intelligent Systems, vol. 2, no. 5, 2020. https://doi.org/10.1002/aisy.202000070

[18] Y-Y Ke, T-T Peng, T-K Yeh, W-Z Huang, S-E Chang, S-H Wu, et al., "Artificial intelligence approach fighting COVID-19 with repurposing drugs," Biomedical Journal, vol. 43, no. 4, pp.355-362, 2020. https://doi.org/10.1016/j.bj.2020.05.001

[19] S. Mohanty, M. Harun Al Rashid, M. Mridul, C. Mohanty, and S. Swayamsiddha, "Application of artificial intelligence in covid-19 drug repurposing," Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 14, no. 5, pp. 1027-1031, 2020. https://doi.org/10.1016/j.dsx.2020.06.068

[20] B. R. Beck, B. Shin, Y. Choi, S. Park, and K. Kang, "Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model," Comput Struct Biotechnol J, vol. 18,pp. 784-790, 2020. https://doi.org/10.1016/j.csbj.2020.03.025

[21] A. Banerjee, S. Ray, B. Vorselaars, J. Kitson, M. Mamalakis, S. Weeks, et al., "Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population," Int Immunopharmacol, vol. 86, 2020. https://doi.org/10.1016/j.intimp.2020.106705

[22] K. Zhang, X. Liu, J. Shen, Z. Li, Y. Sang, X. Wu, et al., "Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography," Cell, vol. 181, no. 6, pp. 1423-1433, 2020. https://doi.org/10.1016/j.cell.2020.04.045

[23] A. I. Saba and A. H. Elsheikh, "Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks," Process Safety and Environmental Protection, vol. 141, pp. 1-8, 2020. https://doi.org/10.1016/j.psep.2020.05.029

[24] O. Torrealba-Rodriguez, R. Conde-Gutiérrez, and A. Hernández-Javier, "Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models," Chaos, Solitons & Fractals, vol. 138, 2020. https://doi.org/10.1016/j.chaos.2020.109946

[25] Z. Zhang, A. Zeb, E. Alzahrani, et al., "Crowding effects on the dynamics of COVID-19 mathematical model," Adv Differ Equ, vol. 675, pp. 1-13, 2020. https://doi.org/10.1186/s13662-020-03137-3

[26] A. Saeed Ahmad, S. Owyed, A. H. Abdel-Aty, E. E. Mahmoud, K. Shah, and H. Alrabaiah, "Mathematical analysis of COVID-19 via a new mathematical model," Chaos, Solitons & Fractals, vol. 143, 2021. https://doi.org/10.1016/j.chaos.2020.110585

[27] A. Ahmad, S. Garhwal, S. K. Ray, G. Kumar, S. J. Malebary, and O. M. Barukab, "The number of confirmed cases of COVID-19 by using machine learning: Methods and challenges," Archives of Computational Methods in Engineering, vol. 4, pp. 1-9, 2020. https://doi.org/10.1007/s11831-020-09472-8

[28] A. Arvanitis, I. Furxhi, T. Tasioulis, and K. Karatzas, (2021). "Prediction of the effective reproduction number of COVID-19 in Greece. A machine learning approach using Google mobility data," medRxiv 2021. https://doi.org/10.1101/2021.05.14.21257209

[29] D. Fanelli and F. Piazza, "Analysis and forecast of COVID-19 spreading in China, Italy and France," Chaos, Solitons & Fractals, vol. 134, 2020. https://doi.org/10.1016/j.chaos.2020.109761

[30] V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," Chaos, Solitons & Fractals, vol. 135, 2020. https://doi.org/10.1016/j.chaos.2020.109864

[31] I. Kırbaş, A. Sözen, A. D. Tuncer, and F. Ş. Kazancıoğlu, "Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches," Chaos, Solitons & Fractals, vol. 138, 2020. https://doi.org/10.1016/j.chaos.2020.110015

[32] Y. Peng and M. H. Nagata, "An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data," Chaos, Solitons & Fractals, vol. 139, 2020. https://doi.org/10.1016/j.chaos.2020.110055

[33] C. M. Yeşilkanat, "Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm," Chaos, Solitons & Fractals, vol. 140, 2020. https://doi.org/10.1016/j.chaos.2020.110210

[34] A. J. Aljaaf, T. M. Mohsin, D. Al-Jumeily, and M. Alloghani, "A fusion of data science and feed-forward neural network-based modelling of COVID-19 outbreak forecasting in IRAQ," Journal of Biomedical Informatics, vol. 118, 2021. https://doi.org/10.1016/j.jbi.2021.103766

[35] H. Abbasimehr and R. Paki, "Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization," Chaos, Solitons & Fractals, vol. 142, 2021. https://doi.org/10.1016/j.chaos.2020.110511

[36] S. Ballı, "Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods," Chaos, Solitons & Fractals, vol. 142, 2021. https://doi.org/10.1016/j.chaos.2020.110512

[37] K. N. Nabi, M. T. Tahmid, A. Rafi, M. E. Kader, and M. A. Haider, "Forecasting COVID-19 cases: A comparative analysis between Recurrent and Convolutional Neural Networks," Results in Physics, vol. 24, 2021. https://doi.org/10.1016/j.rinp.2021.104137

[38] K. E. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji, and T. M. Brenza, "Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and long short-term memory (LSTM) cells," Chaos, Solitons & Fractals, vol. 146, 2021. https://doi.org/10.1016/j.chaos.2021.110861

[39] The Republic of Turkey, Ministry of Health. General coronavirus table, https://COVID19.saglik.gov.tr/; 2020 [accessed 02 June 2021]

[40] A. L. Samuel, "Some studies in machine learning using the game of checkers," IBM Journal of Research and Development, vol. 3, no. 3, pp. 210-229, 1959. https://doi.org/10.1147/rd.33.0210

[41] V. N. Vapnik, Statistical learning theory. New York, NY: John Wiley & Sons; 1998.

[42] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, pp. 273-297, 1995. https://doi.org/10.1007/BF00994018

[43] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5-32, 2001. https://doi.org/10.1023/A:1010933404324

[44] E. Fix and Jr J. L. Hodges, Discriminatory analysis-nonparametric discrimination: consistency properties. Technical report, California University, Berkeley, 1951.

[45] T. M. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Trans Inf Theory., vol. 13, pp. 21-27, 1967. https://doi.org/10.1109/TIT.1967.1053964

[46] Mitchell TM. Machine learning. New York, NY: McGraw Hill; 1997.

[47] J. Han, M. Kamber, and J. Pei, Data mining: concepts and techniques. Waltham, MA: Morgan Kaufmann Publishers; 2006.

[48] R. Mitchell and E. Frank, "Accelerating the XGBoost algorithm using GPU computing," PeerJ Computer Science, vol. 3, pp.127–164, 2017. https://doi.org/10.7717/peerj-cs.127

[49] H. Zheng, J. Yuan, and L. Chen, "Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation," Energies, vol. 10, no. 8, pp. 1168–1188, 2017. https://doi.org/10.3390/en10081168

[50] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 13-17 August 2016, pp. 785–794. http://doi.org/10.1145/2939672.2939785

[51] H. S. Hota, R. Handa, and A. K. Shrivas, "COVID-19 pandemic in India: Forecasting using machine learning techniques," Data Science for COVID-19, pp. 503–525, 2021. http://doi.org/10.1016/B978-0-12-824536-1.00030-7

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173