REVIEW ARTICLE

A Review on Machine Learning Approaches in COVID-19 Pandemic Prediction and Forecasting

Nor Nisha Nadhira Nazirun[1], Nashuha Omar[1], Koshelya Selvaganeson[1], Asnida Abdul Wahab[1,2]

[1] School of Biomedical Engineering and Health Sciences, Faculty of Engineering, Universiti Teknologi Malaysia, Skudai 81310, Johor, Malaysia
[2] Medical Devices and Technology Centre (MEDITEC), Universiti Teknologi Malaysia, Skudai 81310, Johor, Malaysia

ABSTRACT

Novel COVID-19 Coronavirus disease, namely SARS-CoV-2, is a global pandemic and has spread to more than 200 countries. The sudden rise in the number of cases is causing a tremendous effect on healthcare services worldwide. To assist strategies in containing its spread, machine learning (ML) has been employed to effectively track the daily infected and mortality cases as well as to predict the peak growth among the states or/and country-wise. The evidence of ML in tackling previous epidemics has encouraged researchers to reciprocate with this outbreak. In this paper, recent studies that apply various ML models in predicting and forecasting COVID-19 trends have been reviewed. The development in ML has significantly supported health experts with improved prediction and forecasting. By developing prediction models, the world can prepare and mitigate the spread and impact against COVID-19.

Keywords: Machine learning, COVID-19, Pandemic, Prediction, Forecasting

Corresponding Author:
Asnida Abdul Wahab, PhD
Email: asnida.aw@utm.my
Tel: +607-5558458/ +6017-4118086

INTRODUCTION

COVID-19 is a contagious disease that originated from a newly discovered coronavirus, subsequently named SARS-CoV-2 (1). Since its first human case was reported in the Wuhan district of China in December 2019, the World Health Organization (WHO) has declared the outbreak as a pandemic in March 2020 due to its rapid spread worldwide in a short period (2). The most common symptoms of COVID-19 are fever, dry cough, fatigue, and mild to severe respiratory complications, which can lead to death if very severe (1). As of January 2021, the novel disease has spread to more than 200 countries, affecting 95,321,880 confirmed cases with 2,058,227 of total death tolls reported to WHO (1). However, it is not known how this global pandemic will peak or diminish; thus, predicting the trend of COVID-19 is of striking significance (3).

Researchers and scientists are working towards finding possible steps to mitigate this virus from spreading further (4). To break the chain, each country has taken immediate action by imposing quarantine, lockdown, and travel ban for months (4). Apart from introducing control measures, researchers are also searching for innovative solutions as a new angle to control the spread of the virus (5). With correct data and integration methods, ML has shown to be a promising technology employed to advance clinical decision support by improving the prediction accuracy for screening diseases (4). It is proven that ML results in better scale-up, faster processing power, reliability, and can even outperform humans in specific healthcare tasks (6,7). In the context of COVID-19, ML can be implemented to handle large amounts of data and intelligently predict the disease spreads (8).

Therefore, this paper focuses on the use of modern ML techniques in predicting and forecasting the number of cases or trends during the outburst to curb the spreading of the virus, improve aid distribution, manage disruption control, and other challenges faced. A review of recent studies which further discusses types of the dataset used, ML models applied, and performances of each proposed model is presented in this paper.

ML IN COVID-19 PREDICTION AND FORECASTING

Wrapping multiple data sources with ML can be very helpful in analysing the growth of infection with community behaviour, especially in forecasting where and when the disease is likely to spread (8). Several predicting methods based on ML have been applied to the time series prediction of COVID-19 development in some severe countries and globally (3). The Centers for Disease Control and Preventions (CDC) prioritises models for a) mortality forecast, b) hospitalisation forecasts, c) COVID-19 pandemic planning scenarios, and d) COVID-19 surge. The models aim to aid in pandemic response by informing decisions about planning, resource allocation, and the need for social distancing (9). Most of the studies conducted different approaches of ML techniques in the short-term forecasting the number of new cases, recovered, and deaths for the following days, weeks, or even months (10–13). Batista et al. (14), Mehta et al. (15), and Pourhomayoun and Shakibi (16) implemented ML in predicting the different risks for COVID-19. As for the maximal number of patients across different locations, it is discussed by Batista (17) and Car et al. (18). At the end of this section, Table 1 summarises the comparison of the ML approaches done by these researchers in the application of COVID-19 prediction and forecasting, which also includes the data size and evaluation rate.

COVID-19 data

Coronavirus is a chain reaction caused by an airborne transmission that can be diagnosed by the reverse transcription-polymerase chain reaction (RT-PCR) test (4). Although the results of this test may take up to 48 hours long, it is typically highly accurate and is used to report the number of daily infected cases (14). Among all the studies reviewed in this paper, the COVID-19 data collected are mostly from the available public dataset, consisting of infected or confirmed, recovered, and death cases. Since COVID-19 is a newly-found virus, thus there is no available data from the previous year. Most researchers stated that their data were taken between specified dates or a period of days during the COVID-19 spread, making the data size indefinite. Studies have also discussed the use of data retrieved from hospitals and laboratories (14,16), while others used the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) (3,11,19). The data extracted may be from different locations or countries, depending on the demographic of the study conducted, such as China (17,20,21), the United States of America (USA) (15,22), Brazil (13,23), Malaysia (10,24) and India (21).

ML methods

Recently, many researchers have discussed the different types of ML models developed for the COVID-19 prediction and forecasting (3). The frequency of ML models implemented in the recently reviewed papers for COVID-19, ranked from the most to the least, is Support Vector Machine (SVM), Logistic Regression, Decision Tree (DT), and Random Forest (RF). Despite that, mathematical modelling of infectious disease, namely Susceptible-Exposed-Infected-Recovered (SEIR), is still widely used to characterise the epidemic peak of COVID-19 (10). The model assigns the population to four main components, including susceptible (S), exposed (E), infected (I), and recovered (R), while calculating the R0 value is the most important aspect of this model. Due to limited data availability, it is assumed that (i) the number of births and deaths remains constant; (ii) $1/\sigma$ is a latent period of disease and $1/\gamma$ is infectious period; and (iii) during the calculation period, the recovered individual was not sick again when describing the spread of COVID-19 by using SEIR (11).

SVM is a type of supervised ML algorithm that works well on both classification and regression (19,22,25). It separates the variables or so-called support vectors using a hyperplane by maximising the margin between two classes, obtained from the distance of decision boundaries as illustrated in Figure 1 (13,25,26). The main advantages of SVM lie in its capability to recognise the predictor non-linear pattern and can improve the forecasting cases based on the history (13,27). In this COVID-19 case, it is advantageous to employ SVM since the epidemic trend is indeed curvy and non-smooth (27) and if the samples are small (13). The majority of authors compared SVM with other ML models, which resulted in Batista et al. (14) and Ribeiro et al. (13) concluding that SVM is the best model regarding the accuracy in all scenarios.

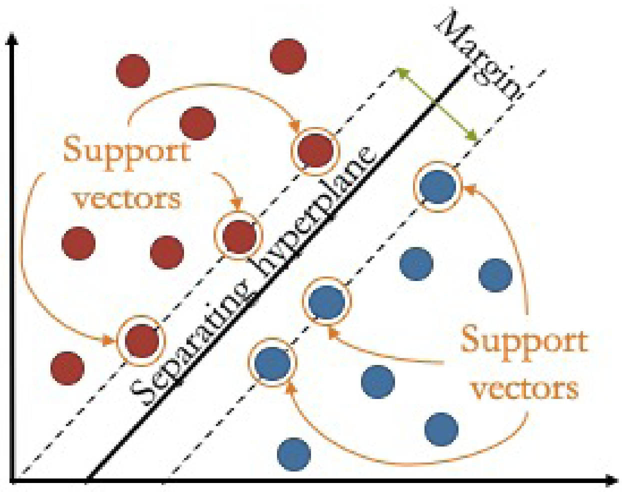

Figure 1: SVM hyperplane

Regression models are statistical sets of ML methods that allow the prediction of the target or continuous outcome variable, which is determined by the value of the predictor or dependent variable/s (11,12). The models have many variants such as linear regression, logistic regression, ridge regression, and polynomial regression (11). In epidemiology, logistic regression is commonly used in the time series regression fitting to predict the likelihood of the occurrence of a certain disease due to its simplicity and efficient calculation (3,20). It uses the sigmoid function to perform predictive analysis based on the relationship between 0/1 or binary dependent variables (25). The logistic growth of COVID-19 is characterised as in Figure 2A, in which the spread starts with a slow increase in growth, then grows fast near the peak of the incidence curve, and latterly a slow growth phase near the end of the outbreak (3).

Another form of the regression model, namely polynomial regression, performs in a curvilinear relationship between the dependent and independent variables, as illustrated in Figure 2B (11,26). In the COVID-19 outbreak data analysis, residuals play a significant role in regression analysis. Yadav (12) used two or more polynomials to analyse COVID-19 data; as a result, fitting a high degree of polynomial to a model can reduce the residuals. If the value of the polynomial degree is lower than the actual, the model is unable to fit properly, and if otherwise, overfitting of the training data will occur (11). Monica and Devi (28) compared the polynomial regression with two other regressors (DT and RF) and proved that it exactly coincides with the actual COVID-19 confirmed cases.

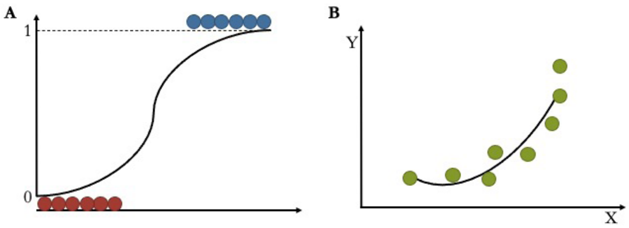

Figure 2: Regression modelling. (A) Logistic regression sigmoid function and (B) polynomial regression curvilinear

DT observes an object's features and trains a model which is represented in the form of a binary tree to predict data in the future (25,28,29). Figure 3A shows that the prediction is made by taking the root node of the binary tree with a single input variable, splitting the dataset based on the variable, and its leaf nodes have resulted as the output variable (25,26). The Gini index function is often used to determine the split or impurity level for the predictions. In a study by Monica and Devi (28), the DT of COVID-19 data has a root node as an observation date, and the output is the number of confirmed cases. Theerthagiri et al. (25), on the other hand, used COVID-19 dataset with two inputs which are taken as age and gender, while the output is whether the patient is recovered or deceased.

As in Figure 3B, RF is a bagging ensemble-based model that combines multiple DT predictors trained using random data samples and feature subsets (13,16,21). It is an efficient and accurate supervised learning method capable of dealing with the randomness of time series (13,26). Every DT has a high variance, but when combined in RF, the resultant variance is low as each DT gets perfectly trained on that particular sample data, and hence the output does not depend on one DT but multiple DTs (28). However, Ribeiro et al. (13) pointed out that the inability of RF to forecast cumulative COVID-19 cases could be due to the model's need for more observations to learn the data pattern effectively.

In the more advanced technology, deep learning (DL) has been shown to have many contributions in predicting and forecasting data trends (30). DL models such as Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN) are usually chosen as COVID-19 deals with time-series data (24,31). Some researchers also have started to prefer hybrid algorithms to improve the prediction and forecasting accuracy of COVID-19 trends (24,32). DL, however, has a drawback when the amount of data available is insufficient due to the requirement of massive labelled data to build an accurate model (26,32). Despite the involvement of many excellent models, Chakraborty et al. (33) stressed that predicting and forecasting COVID-19 is challenging primarily due to seven major factors, including limited availability of data and extreme sources of uncertainty resulting in no gold standard for accurately forecasting the pandemic data.

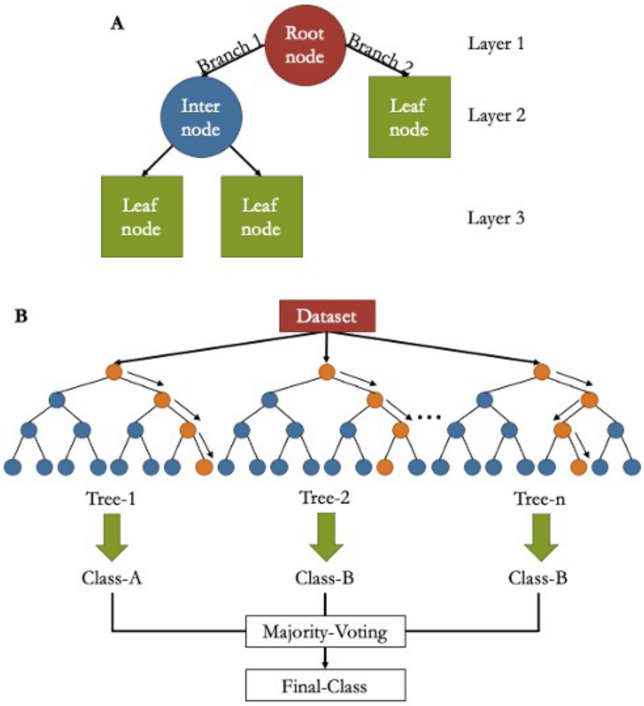

Figure 3: Tree-based algorithm. (A) DT nodes and internodes and (B) RF splitting trees

Model evaluation

Various approaches to the model evaluation were conducted to evaluate the performance of the developed models. For COVID-19 prediction and forecasting, the evaluation metrics are used in regression and binary classification tasks. Confusion matrix (accuracy, sensitivity, and specificity) and area under the receiver operating characteristics (ROC) curve (AUC) are well-known to measure the binary classification or 0/1 prediction. However, most of the studies applied statistical analysis used in regression tasks such as r-squared (R^2) score, mean square error (MSE), and root mean square error (RMSE). Using these metrics, Rustam et al. Field (19) divided the COVID-19 forecasting into three categories: death, recovery, and new confirmed rate to evaluate different ML models for each category. The model that achieved a high R^2 but low MSE and RMSE values indicate a high forecast precision (10). Also, studies stipulated at least three performance criteria to be met in selecting the best optimal model.

R^2 or coefficient of determination score is a statistical measure of regression models. It is used to check the goodness-of-fit of the trained regression model (19,20). As stated by Rustam et al. (19), the R^2 score is used by finding the dispersion of data points around the regression line. It is defined in the range of [0,1], in which the closer the fitting coefficient to a perfect value of 1, the more accurate the prediction model is (18–20). A study by Jia et al. (20) has divided the fitting goodness calculation of COVID-19 cases into three; $R^2(C)$ is for cumulative confirmed cases, $R^2(N)$ is for new confirmed cases, and $R^2(DC)$ is for cumulative deaths. These calculations analysed the prediction done by three different models in different regions. In Wuhan, for example, the $R^2(C)$ obtained were 0.9991 (Logistic model), 0.999 (Gompertz model), and 0.9989 (Bertalanfffy model). Equation 1 shows the formula for calculating the R^2 score.

$$R^2 = \frac{1-(\sum(y_i-\hat{y}_i)^2}{y_i-\bar{y}_i)^2} \tag{1}$$

MSE is another method in evaluating regression models by finding the average of the squared difference between predicted values and actual values (19,25). Squaring the values is essential to remove the negative sign from the value and emphasise larger differences by giving more weight (19). The smaller the value, or the closer the value to 0, shows the closest to finding the best fit line; thus, the better the quality of the results are (19,25,28). Theerthagiri et al. (25) concluded in their study that K-Nearest Neighbor (KNN) resulted in a lower error rate of 0.9 due to the testing dataset was classified by calculating Euclidean distance between the new instance and existing instance. In a study by Rustam et al. (19), Exponential Smoothing (ES) has shown to perform better among all the models with the lowest MSE (662228.72) while Linear Regression and Least Absolute Shrinkage and Selection Operator (LASSO) performed equally well and achieved low MSE values of 840240.11 and 3244066.79, respectively. MSE can be defined as:

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

The square root of MSE can be used to measure RMSE which depicts the inconsistencies among the observed and predicted values (25). It is described as the standard deviation of the prediction errors where it measures how close the actual data points are to the best-fit line (19). A good model shows a smaller value of RMSE. Fong et al. (27) used the evaluation of RMSE for measuring time series forecasting in their study of COVID-19. The winning model which is Polynomial Neural Network (PNN) with corrective feedback (cf) offers the lowest RMSE (136.547) among all. RMSE can be calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

Where y_i is the actual cumulative COVID-19 cases; $\hat{y}_i$ is the predicted cumulative COVID-19 cases; $\bar{y}$ is the average of the actual cumulative COVID-19 cases. With an accuracy of 90% by calculating the R^2 score and RMSE of 53262.68, a study by Monica and Devi (28) has concluded that the polynomial regression model best predicts COVID-19 confirmed cases as compared to DT regression and RF regression. In predicting the recovered and deceased cases of COVID-19, Theerthagiri et al. (25) showed that the MSE value of SVM (0.21) is lower compared to Logistic Regression (0.2146) and DT (0.2466) meanwhile as for RMSE, the error rate for SVM (0.4583) is also very low followed by Logistic Regression (0.4633), and DT (0.4966).

It appears that the model which has outperformed others due to its high accuracy in predicting and forecasting COVID-19 is a polynomial regression model, while other preferred methods include the logistic model and SVM. Batista (17) and Jia et al. (20) stated that the logistic model is better and can be used for the prediction as the COVID-19 growth is characterised similarly to the model. Meanwhile, SVM performed well due to its capability in recognising the predictor of curvy and non-smooth patterns like the COVID-19 trend (13,14). Above all, polynomial regression is an excellent model as Monica and Devi (28) calculated 90% of the R^2 score and 53262.68 RMSE value. Yadav (12), on the other hand, found that a higher degree of polynomial regression produces more prediction accuracy in which the R^2 score and adjusted R^2 of the sixth-degree polynomial regression model were 0.9990 and 0.9989, respectively.

CONCLUSION

Everyone has voiced their concerns on the impact of COVID-19 and has to embrace the 'new norm'. ML has been widely used in developing prediction models for tailoring health providers in medical decision-making. This paper has addressed recent studies that have applied ML in developing predictive models to predict and forecast trends specifically for COVID-19. Data models were mostly implemented by previous studies. The prediction and forecasting models of COVID-19 could be divided into statistical, ML, and DL. In the context of ML, it can be concluded that the regression model has outperformed other models when evaluated

Table I: ML Approaches in Predicting and Forecasting of COVID-19

Reference	Application	Data	Model	Evaluation	Result
(Alsayed et al., 2020)	Prediction of COVID-19 infection rate, epidemic peak, and number of infected cases for the upcoming five days	Available data from WHO & available data of infected cases from 25 January - 05 April 2020 in Malaysia	SEIR, Genetic Algorithm, Adaptive Neuro-Fuzzy Inference System (ANFIS)	RMSE, Normalized RMSE, Mean Absolute Percentage Error (MAPE), R^2 score	ANFIS model show a low NRMSE (0.041); a low MAPE (2.45%); and a high R2 (0.9964)
(Ballı, 2021)	Predict weekly cumulative cases for global, Germany and USA	Data of COVID-19 between 20 January - 18 September 2020 for USA, Germany, and the global obtained from WHO	Linear regression, Multi-Layer Perceptron (MLP), RF and SVM	RMSE, Absolute Percentage Error (APE), MAPE	SVM achieved the best trend as it resulted in the lowest value for RMSE, MAPE, and APE scores
(Batista et al., 2020)	Prediction of positive COVID-19 diagnosis risk	235 adult patients from the Hospital Israelite Albert Einstein in Sro Paulo, Brazil for 17 - 30 March 2020	Neural Network (NN), RF, Gradient Boosting Trees, Logistic Regression and SVM	AUC, Sensitivity, Specificity, Brier score	SVM performed well
(Batista, 2020)	Prediction of epidemic size for China, South Korea and the rest of the world	Data from January 2020 - March 2020	Logistic model and SIR model	RMSE	Logistic model can be used to predict the COVID-19 cases
(Car et al., 2020)	Prediction of the maximal number of patients across all locations in each time unit	Information on infected, recovered, and deceased patients in 406 locations over 51 days	MLP, Artificial Neural Network	R^2 score, Cross validation	AI models can be used in modeling the spread and effect of infectious diseases
(James Fong et al., 2020)	Prediction of the epidemic lifetime to decide on timely and remedial actions	Data between 21 January - 3 Feb 2020	Autoregressive Integrated Moving Average (ARIMA), Exponential, Holt-Winters Addictive, Linear Regression, SVM, fast DT learner, MP5 DT learner, PNN and PNN+cf	RMSE	PNN+cf is a better model
(Majhi et al., 2021)	Prediction of the number of positive cases in India	China's data from 15 January 2020 for training and India's sample for validation until 3 May 2020	Nonlinear regression, DT-based regression, and RF	MAPE	RF (0.02%) outperforms compared to DT (0.18%) and Nonlinear regression (0.24%)
(Monica and Devi, 2020)	Prediction of COVID-19 progress	Dataset consists of 7 fields and 27166 records containing confirmed, deaths and recovered cases of COVID-19	Polynomial regression, DT regressor, RF regressor	RMSE, MSE	Polynomial regression model produces more prediction accuracy
(Gupta et al., 2020)	Prediction of the number of cases for the next 2 weeks	JHU CSSE repository from 30 January - 30 March 2020	SEIR model and Regression (linear and polynomial) model	Root Mean Squared Log Error (RMSLE)	SEIR model (1.52) and Regression model (1.75)
(Jia et al., 2020)	Prediction and analysis the situation of COVID-19 in China	COVID-19 data of confirmed cases and death cases of whole China mainland	Logistic model, Bertalanffy model and Gompertz model	R^2 score	Logistic model is a better model
(Mehta et al., 2020)	Prediction of the county-level risk	US country level using publicly available data	XGBoost	Sensitivity, specificity, AUC, RMSE, hold out validation	Sensitivity (>71%) and specificity (>94%)
(Pourhomayoun and Shakibi, 2021)	Prediction of mortality risk accuracy for the physical and symptom-based features	117,000 laboratory-confirmed COVID-19 patients from 76 countries	SVM, NN, RF, DT, Logistic Regression, KNN	Accuracy, AUC, sensitivity, specificity	NN has the highest accuracy (93.75%)
(Ribeiro et al., 2020)	Short-term forecasting with one, three, and six-days ahead the COVID-19 cumulative cases in Brazilian states	Cumulative confirmed cases of COVID-19 in Brazil until April 18 or 19, 2020	ARIMA, Cubist Regression, RF, Ridge Regression, SVM, Stacking-Ensemble Learning	Mean Absolute Errors (MAE), symmetric MAPE, Improvement Percentage Index (IP)	SVM is the best model regarding accuracy, in all scenarios
(Rustam et al., 2020)	Prediction of number of newly infected cases, deaths and recoveries	2019 Novel Coronavirus Visual Dashboard operated by JHU CSSE	Linear regression, LASSO, SVM, and ES	R^2 score, R^2 adjusted MSE, MAE and RMSE	ES performed best
(Saba et al., 2021)	Ten-days-ahead forecasting of COVID-19 infected and deaths cases under different lockdown types	Confirmed and deaths cases collected from https://github.com/CSSEGISandData for nine countries between 22 January - 30 September 2020	RF, KNN, SVM, DT, polynomial regression, Holt winter, ARIMA, and SARIMA	MAPE, MAE, and RMSE	Impossible to recommend a single approach for all datasets as different datasets exhibited different trends
(Theerthagiri et al., 2020)	Prediction of recovered and deceased cases of COVID-19	COVID-19 dataset contains the patient's details with recovered and deceased status	Logistic Regression, DT, KNN, SVM, MLP	MSE, RMSE, Cohen's Kappa Score	KNN classification algorithm shows lowest error rate (0.19)
(Wang et al., 2020)	Predict of long-time period trend of COVID-19 in global, also in some heavily infected countries	JHU CSSE dashboard at country level from 22 January - 16 June 2020	Hybrid Logistic and FbProphet model	95% confidence interval	The model has valuable advantage in forecasting epidemic trend
(Yadav, 2020)	Prediction of number of cases for the next 6 days	Database of COVID-2019 from 01 March - 11 April 2020.	Quadratic, 3rd degree, 4th degree, 5th degree, 6th degree, and Exponential Polynomial Regression models	Sum of Square Errors (SSE), R^2 score, Degree of Freedom for Error (DFE) adjusted R^2	6th degree polynomial regression model is a very good models

using methods such as statistical analysis. It is shown that ML can be used in modelling the spreading pattern as well as improving the prediction and forecasting of the COVID-19. However, the fight against COVID-19 is still far from over. This pandemic recovery depends not only on the prediction and forecasting alone but also on other strategies that need to be taken by all parties. For the progression of better models in long haul expectation, future exploration ought to be committed to near examinations on different ML models for an individual country. Because of the major contrasts between the episode in different nations, the headway of worldwide models with speculation capacity would not be feasible.

ACKNOWLEDGEMENTS

This work was financially supported by the Universiti Teknologi Malaysia under the Fundamental Research Grant Scheme (FRGS) and the Ministry of Science, Technology, and Innovation (MOSTI) with reference number 5F364.

REFERENCES

1. World Health Organization. WHO Coronavirus Disease (COVID-19) [Internet]. WHO.int. 2020 [cited 2021 Jan 31]. p. 1. Available from: https://covid19.who.int/
2. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg [Internet]. 2020 Apr;76:71–6. Available from: https://www.sciencedirect.com/science/article/pii/S1743919120301977?casa_token=4Q82iMBJ7SIAAAAA:WH5nDSuGXlzqfzjjVkQ4nIOWhQR0p48aA_3zVG9pzm-r5IIT30Uf-ngNBJm4HZCyv4u_24d873o
3. Wang P, Zheng X, Li J, Zhu B. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. Chaos, Solitons and Fractals [Internet]. 2020;139:110058. Available from: https://www.sciencedirect.com/science/article/pii/S0960077920304550?casa_token=eRXEJPKcRcUAAAAA:5fQ8_ozHDmIERAKMyc_ol-vLoHbUndeNHSlrtvx4WHW5N5MBz6UCcFmFLhppCc2_cWpOyT9nzP4
4. Sharma S, Gupta YK. Predictive analysis and survey of COVID-19 using machine learning and big data. J Interdiscip Math [Internet]. 2021 Jan 12;24(1):1–21. Available from: https://doi.org/10.1080/09720502.2020.1833445
5. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. Chaos, Solitons and Fractals [Internet]. 2020;139:110059. Available from: https://www.sciencedirect.com/science/article/pii/S0960077920304562?casa_token=_EogOtpK17AAAAAA:bA4FHv1PU8feFJfhnD96Pgkry4_W5D_2v9fIYV3A_wgMZgRv1Cn1MqD15uC76Rd_R5jvt2r0VdQ
6. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Futur Healthc J [Internet]. 2019 Jun;6(2):94–8. Available from: /pmc/articles/PMC6616181/?report=abstract
7. Pramenković B, Prasko D, Pulo E, Rončević I, Ramić R, Rakovac A. Machine Learning Techniques for Predicting Outcomes of COVID-19 for Patients with preexisting Chronic Diseases. In: International Conference on Medical and Biological Engineering. Springer; 2021. p. 867–82.
8. Tuli S, Tuli S, Tuli R, Gill SS. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. Internet of Things [Internet]. 2020 Sep;11:100222. Available from: https://www.sciencedirect.com/science/article/ pii/S254266052030055X?casa_token=NbZUpSn43u8AAAAA:9G98HZQ14t7EaQhGLw60Q9ScvPUifLiOns0udyqAMBbPCt5fO2Xw1jufDOy56qO17QWAMycrn5g
9. Centers for Disease Control and Preventions. COVID-19 Mathematical Modeling [Internet]. 2020 [cited 2021 May 26]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/mathematical-modeling.html
10. Alsayed A, Sadir H, Kamil R, Sari H. Prediction of epidemic peak and infected cases for COVID-19 disease in Malaysia, 2020. Int J Environ Res Public Health [Internet]. 2020 Jun;17(11):1–15. Available from: https://www.mdpi.com/1660-4601/17/11/4076
11. Gupta R, Pandey G, Chaudhary P, Pal S. SEIR and Regression Model based COVID-19 outbreak predictions in India. medRxiv [Internet]. 2020 Apr; Available from: https://www.medrxiv.org/content/10.1101/2020.04.01.20049825v1.full.pdf+html
12. Yadav RS. Data analysis of COVID-2019 epidemic using machine learning methods: a case study of India. Int J Inf Technol [Internet]. 2020;12(4):1321–30. Available from: https://www.researchgate.net/publication/341660839_Data_analysis_of_COVID-2019_epidemic_using_machine_learning_methods_a_case_study_of_India
13. Ribeiro MHDM, da Silva RG, Mariani VC, Coelho L dos S. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. Chaos, Solitons and Fractals [Internet]. 2020;135:109853. Available from: https://www.sciencedirect.com/science/article/pii/S0960077920302538?casa_token=qH8938fjG1oAAAAA:HWw1YL_ANHHcQCclrVGYDOQq0QzAgmtpBB5SVsmgG-QGejzyNyk8CGqOURTKRXk5H37i31achbQ
14. Batista AF de M, Miraglia JL, Donato THR, Chiavegatto Filho ADP. COVID-19 diagnosis prediction in emergency care patients: A machine learning approach. medRxiv [Internet]. 2020 Apr; Available from: https://www.medrxiv.org/content/10.1101/2020.04.04.20052092v2.abstract
15. Mehta M, Julaiti J, Griffin P, Kumara S. Early stage machine learning–based prediction of US county vulnerability to the COVID-19 pandemic: Machine learning approach. JMIR Public Heal Surveill [Internet]. 2020 Jul;6(3):e19446. Available from: https://publichealth.jmir.org/2020/3/e19446/
16. Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. Smart Heal [Internet]. 2021 Apr;20:100178. Available from: https://www.sciencedirect.com/science/

article/pii/S2352648320300702

17. Batista M. Estimation of the final size of the second phase of the coronavirus COVID 19 epidemic by the logistic model. medRxiv [Internet]. 2020; Available from: https://www.medrxiv.org/content/10.1101/2020.03.11.20024901v2
18. Car Z, Baressi egota S, Anđelić N, Lorencin I, Mrzljak V. Modeling the Spread of COVID-19 Infection Using a Multilayer Perceptron. Comput Math Methods Med [Internet]. 2020;2020:5714714. Available from: https://www.hindawi.com/journals/cmmm/2020/5714714/
19. Rustam F, Reshi AA, Mehmood A, Ullah S, On BW, Aslam W, et al. COVID-19 Future Forecasting Using Supervised Machine Learning Models. IEEE Access [Internet]. 2020;8:101489–99. Available from: https://ieeexplore.ieee.org/abstract/document/9099302
20. Jia L, Li K, Jiang Y, Guo X, Zhao T. Prediction and analysis of Coronavirus Disease 2019. arXiv [Internet]. 2020 Mar; Available from: http://arxiv.org/abs/2003.05447
21. Majhi R, Thangeda R, Sugasi RP, Kumar N. Analysis and prediction of COVID-19 trajectory: A machine learning approach. J Public Aff. 2021;21(4):e2537.
22. Ballı S. Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. Chaos, Solitons & Fractals [Internet]. 2021;142:110512. Available from: https://www.sciencedirect.com/science/article/pii/S0960077920309048
23. de Moraes Batista AF, Miraglia JL, Rizzi Donato TH, Porto Chiavegatto Filho AD. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. medRxiv [Internet]. 2020 Jan 1;2020.04.04.20052092. Available from: http://medrxiv.org/content/early/2020/04/14/2020.04.04.20052092.abstract
24. Alassafi MO, Jarrah M, Alotaibi R. Time series predicting of COVID-19 based on deep learning. Neurocomputing [Internet]. 2022;468:335–44. Available from: https://www.sciencedirect.com/science/article/pii/S0925231221015150
25. Theerthagiri P, I JJ, A UR, Yendapalli V. Prediction of COVID-19 Possibilities using KNN Classification Algorithm. Int J Curr Res Rev [Internet]. 2020;13(06):156–64. Available from: https://assets.researchsquare.com/files/rs-70985/v2_stamped.pdf
26. Saba T, Abunadi I, Shahzad MN, Khan AR. Machine learning techniques to detect and forecast the daily total COVID-19 infected and deaths cases under different lockdown types. Microsc Res Tech [Internet]. 2021/02/01. 2021 Jul;84(7):1462–74. Available from: https://pubmed.ncbi.nlm.nih.gov/33522669
27. James Fong S, Herrera Viedma E, Fong SJ, Li G, Dey N, Crespo RG, et al. Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak. Int J Interact Multimed Artif Intell [Internet]. 2020 Feb;6(1):132–40. Available from: https://digibug.ugr.es/bitstream/handle/10481/64933/ijimai20206_1_15_pdf_18033.pdf?sequence=1&isAllowed=y
28. Monica G, Devi DMB. Using Machine Learning Approach to Predict Covid-19 Progress. Int J Mod Trends Sci Technol [Internet]. 2020;6(8S):58–62. Available from: https://www.ijmtst.com/volume6/issue08s/12.IJMTSTCIET60.pdf
29. Kunjir A, Joshi D, Chadha R, Wadiwala T, Trikha V. A Comparative Study of Predictive Machine Learning Algorithms for COVID-19 Trends and Analysis. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2020. p. 3407–12.
30. Rahimi I, Chen F, Gandomi AH. A review on COVID-19 forecasting models. Neural Comput Appl [Internet]. 2021; Available from: https://doi.org/10.1007/s00521-020-05626-8
31. Wu F, Shu J. Predictions For COVID-19 With Deep Learning Models of Long Short-Term Memory (LSTM). In: Biomedical and Business Applications Using Artificial Neural Networks and Machine Learning. IGI Global; 2022. p. 128–53.
32. Dairi A, Harrou F, Zeroual A, Hittawe MM, Sun Y. Comparative study of machine learning methods for COVID-19 transmission forecasting. J Biomed Inform [Internet]. 2021;118:103791. Available from: https://www.sciencedirect.com/science/article/pii/S1532046421001209
33. Chakraborty T, Ghosh I, Mahajan T, Arora T. Nowcasting of COVID-19 Confirmed Cases: Foundations, Trends, and Challenges BT - Modeling, Control and Drug Development for COVID-19 Outbreak Prevention. In: Azar AT, Hassanien AE, editors. Cham: Springer International Publishing; 2022. p. 1023–64. Available from: https://doi.org/10.1007/978-3-030-72834-2_29