Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/4/2023404

Machine learning for Covid-19: A review on forecasting models

Shichen Su

Department of Physics and Astronomy, University College London, Gower St, London WC1E 6BT, United Kingdom

teresa.su.21@ucl.ac.uk

Abstract. The accurate forecasting of Covid-19 cases remains a challenge even for comprehensive statistical models due to the complex dynamics inherent in largely urbanised populations. This difficulty, coupled with recent developments in artificial intelligence (AI), has led to a burgeoning interest in implementing machine learning (ML) algorithms to aid such predictions. Ever since the outbreak, various ML-based forecasting approaches have been trained and tested to support the design of intervention and healthcare strategies to protect the population. It is thus vital to assess the performance of these algorithms such that their accuracies, along with strengths and weaknesses, are known. This paper compares the forecasting performance of six commonly used ML models for Covid-19 forecasting. A brief analysis of the algorithms' suitability for forecasting is also provided as part of the review. It is found that different algorithms were able to predict with varying degrees of accuracies in different scenarios, and that there is yet to be a definitive algorithm that is best for forecasting the Covid-19 outbreak.

Keywords: Covid-19, artificial intelligence, forecasting models, machine learning, deep learning.

1. Introduction

The novel coronavirus (Covid-19) was first reported in December 2019. Since then, the virus has caused more than 590 million confirmed cases and 6 million deaths worldwide as of August 2022 [1]. Due to Covid-19's ubiquity, researchers around the world have developed numerous models to forecast the trajectory of the pandemic in an attempt to assist the development of intervention and management policies.

The complex nature of transmission dynamics, along with unprecedented access to an almost global dataset provided an opportunity for the scientific community to leverage the power of AI to forecast the outbreak. Regression algorithms in particular, have been implemented widely to predict the number of new infections, mortality rates, recoveries and other pandemic metrics.

This review aims to provide a brief overview of the popular ML-based forecasting methods and compares the accuracies amongst them. Both Vaishya et al. [2], Mohamadou, Halidou and Kapen [3] have reviewed the use of AI during the pandemic and explored its wider application that includes treatment monitoring, contact tracing and development of vaccines. This review aims to contribute to the existing discussion by providing an in-depth overview of the use of AI in forecasting models.

DOI: 10.54254/2755-2721/4/2023404

The rest of the paper first describes the workflow for the training and testing of an ML-based forecasting model, then introduces some of the most commonly used algorithms. Next, the paper compares and discusses the performance of the following models: Absolute Shrinkage and Selection Operator (LASSO), Support Vector Regression (SVR), Exponential Smoothing (ES), AutoRegressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU). This helps the implementation of AI in predicting infectious disease outbreaks, thereby enabling better policy-making and resource mobilization to minimize impacts on societies.

2. Overview of Forecasting Workflow

The construction and testing of ML-based forecasting models begin with gathering relevant datasets from sources such as the GitHub repository hosted by the Centre for Systems Science and Engineering, Johns Hopkins University. The dataset files then undergo cleaning, transformation, reduction and other processes to ensure that they are ready for algorithm training. After being pre-processed, the data is split into the training set and testing set. The split ratio varies across different studies but typically ranges from 70:30 to 80:20, where the majority of the data is used for training. The training set is then fed into one or more of the aforementioned algorithms to develop a trained model. Finally, the performance of each model is evaluated with the testing set. This process is captured in Figure 1.

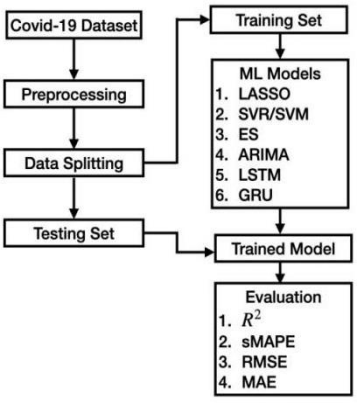

Figure 1. Development and training workflow of forecasting models.

The evaluation parameters used also vary across different studies. The following will be examined in this review:

R^2 score

A value that determines the proportion of variance in the dependent model explainable by the independent variable. It gives an indication of how well real data fits the model. It is calculated by:

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

where $(y_i - \hat{y}_i)$ is the individual prediction error and $(y_i - \bar{y})$ is the difference between the observed value and the mean.

Symmetric Mean Absolute Percentage Error (sMAPE)

A measure of accuracy based on relative errors given by:

$$sMAPE = \frac{100\%}{n}\sum_{t=1}^{n} \frac{|F_t - A_t|}{\frac{|A_t| + |F_t|}{2}} \tag{2}$$

where A_t is the observed value and F_t the predicted value.

DOI: 10.54254/2755-2721/4/2023404

Root Mean Square Error (RMSE)

The standard deviation of prediction errors that measures the distance between the predicted curve and the data points. It is given by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (p - a)^2} \tag{3}$$

where n, p and a are number of observed points, the estimated value and the observed value respectively.

MAE (Mean Absolute Error)

A measure of errors given by the mean of the absolute value of all individual prediction errors:

$$\mathrm{MAE} = \frac{1}{m}\sum_{i}^{m} \left|y_i - \hat{y_i}\right| \tag{4}$$

where $|y_i - \hat{y_i}|$ is the absolute value of the prediction error.

3. Machine Learning Techniques

Machine learning has become an indispensable tool in pattern recognition, utilised in a variety of modelling contexts. It is a subset of AI that enables software to learn and become more proficient at predicting outcomes without explicit parameter setting done by a human. Forecasting makes use of regression algorithms, meaning that they predict continuous values by mapping a predictive correlation between labels and given data points. A regression model relies on gradient descent to minimise the loss function, and, in turn, updates the coefficients of a line to better fit and predict data.

Deep learning is a further subset of machine learning that uses a multilayered structure of algorithms known as an "artificial neural network" (ANN). Each layer consists of nodes that act as neurons to mimic the data processing function of real human brains. The most basic structure of ANN has three parts: an input layer, a collection of hidden layers and an output layer. Each hidden layer passes a set of weighted inputs through an activation function, and outputs it into the following layer until the final output layer is reached. Then, a backpropagation algorithm is used to calculate derivatives of the loss function, and gradient descent is once again used to update the weights and biases of the model. By iterating this process over each layer, the output error is minimised. A deep neural network contains at least two hidden layers, and the more hidden layers there are, the deeper the neural network.

Different algorithms possess different strengths and weaknesses in their prediction, and this paper reviews the performance of both non-deep learning and deep learning algorithms used in forecasting.

3.1. Non-Deep Learning Algorithms

The non-deep learning algorithms that will be reviewed are: Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Regression (SVR), Exponential Smoothing (ES) and AutoRegressive Integrated Moving Average (ARIMA).

LASSO is a standard regression algorithm with regularisation. The shrinkage involves adding a penalty term to the basic least squares fitting approach. The penalty is equal to the absolute magnitude of the coefficient, and severe penalty effectively shrinks coefficients of weaker indicators towards zero, and thereby eliminates them. LASSO thus results in simpler models to avoid overfitting.

SVR is a slightly different regression model based on Support Vector Machine (SVM) and commonly used in predicting distinct values in both linear and non-linear models. The biggest difference between SVM and SVR is that SVR is a regressor whereas SVM is simply a classifier used to predict discrete categorical labels. SVR also differs from other regression models that attempt to minimise error, and instead finds the best fit line by maximising the number of points that lie on a hyperplane that acts

DOI: 10.54254/2755-2721/4/2023404

as the decision boundary. The support vectors are points that are closest to the hyperplane, and are used to determine the line of best fit, which gives the predicted output of the model.

ES is another popular technique that is foundational to many forecasting models. The core idea is to assign varying weights to past observations. More specifically, the weights exhibit exponential decay with the age of observations. This then allows recent observations to have a higher weight and more influence on forecasting, which ensures the reliability of predictions.

Lastly, ARIMA can learn to forecast from non-stationary time series that exhibit trends which change over time. It first makes the time series non-stationary and integrated by differencing, that is, by subtracting the previous value from the current value. The AR part of ARIMA refers to Auto Regression, which is regression of the variable dependent on its own past value (also known as lags). Finally, MA refers to Moving Average, which means that the forecast depends linearly on past values, and that predictions errors also depend linearly on previous errors. ARIMA is thus the combination of AR and MA after being integrated via differencing.

3.2. Deep Learning Algorithms

The deep learning algorithms commonly used in Covid-19 forecasting utilise Recurrent Neural Network (RNN), which is a more complex subset of ANN. In ANN, inputs travel sequentially from the input layer, through the hidden layers and finally to the output layer. In RNN however, the output of processing nodes is saved and fed back into the model recursively to self-learn. If the model gives an incorrect prediction, then the system continues to learn through backpropagation. As RNN remembers each output, its node acts as memory cells, making it useful for time series prediction. However, a major disadvantage of RNN is the Vanishing Gradient problem. This occurs during the backpropagation of time-series data when the gradient of the loss function approaches zero exponentially fast, thereby inhibiting their ability to handle long-term dependencies. LSTM and GRU are two improvements of RNN that overcome this problem, and both are used in Covid-19 forecasting.

LSTM is a special type of RNN that can preserve learning long-term dependencies by having a more sophisticated unit with a cell state and gates to regulate how information travels through the network. The cell state functions as the memory of the network that transfers information throughout the entire sequence, this means that information from the earliest node can be carried to the last node. The gates, on the other hand, can modify this memory by adding and removing information. The gates themselves are also neural networks with an activation function to learn which part of the data to keep and which to forget; this then allows LSTM to learn long-term dependencies within the time series.

Similar to LSTM, GRU is another improved version of RNN with gates to overcome the Vanishing Gradient problem. Rather than using a cell state to remember information, GRU uses a hidden state to transfer information. This information is modified by the update and reset gate, which are two vectors that can be trained to keep information from many time steps ago, and thus preserve long-term dependencies by default.

4. Performance of Models

4.1. Performance of non-Deep Learning Models

LASSO is used in several Covid-19 forecasting studies such as Rustam et al.[4], Bhadana, Jalal and Pathak [5]. In the latter study, it was found that LASSO produced relatively good results when used in a 5-day forecast of new infections in India with a R^2 score of 87.20. This performance was slightly better than the simple linear regression model (LR) that it was built upon, which achieved an R^2 score of 86.90. However, as the increase in total new active cases did not follow a linear trend, polynomial regression (Poly LR) would be more suitable for the forecasting, and was able to provide a much more accurate prediction with an R^2 score of 98.64, along with Poly LASSO performing at a similar level of 98.15. Although the difference in performance between Poly LR and Poly LASSO was small, it signalled a potential limitation of LASSO. In order to avoid overfitting, LASSO arbitrarily drop variables from a group of variables with high collinearity without context, and hence potentially removing relevant

DOI: 10.54254/2755-2721/4/2023404

independent variables. With its focus on producing the best combined prediction, the contribution of individual variable also becomes difficult to investigate and interpret with LASSO.

Nevertheless, besides predictions in India, Poly LASSO also performed well in predicting worldwide new infected cases in Rustam et al., with an R^2 score of 98. Additionally, the study investigated the performance of SVM, which yielded a poor R^2 score of 59. This was not surprising as the fluctuations in the dataset values made it exceptionally challenging for SVM to place an accurate hyperplane between the given values. The SVM-based SVR, however, achieved much better results as a regressor. In Ribeiro et al. [6], SVR performed the best out of six different algorithms that included ARIMA. The models were trained to forecast newly confirmed cases across time horizons of 1, 3 and 6 days ahead in ten Brazilian states. On average, it obtained a sMAPE of 7.95% for the six-day forecast. Its good performance could be attributed to its efficiency in handling small size dataset, and the ability to quickly learn nonlinearities inherent in the time series.

Rustam et al., also found that ES performed exceptionally well, consistently achieving an R^2 score between 98 and 99 in forecasting new confirmed cases, death cases and recoveries globally. Even with limited entries in the time series, it was able to produce accurate predictions. The high accuracy was a result of its compatibility with the nature and size of the data set. As ES weights data records according to their contemporaneity, the spikes in the dataset were less detrimental to its predictions compared to other algorithms.

It is important to note even if identical algorithm were to be implemented with different datasets, the level of performance could vary drastically depending on the nature and size of the dataset. For example, Yonar et al. [7] found that ES had an exceptionally poor accuracy when trained on Japan's Covid-19 data with a R^2 score of merely 46.2. The same model, when applied to Italy, yielded a much higher R^2 score of 89.2. The discrepancy in performance demonstrated difficulties in selecting the optimal ML model, and the process of Covid-19 forecasting in general.

Similar to ES, ARIMA is also a time series model with varying performance levels across different studies. In Alzahrani et al. [8], the research group tested the performance of ARIMA and its constituent models: AR, MA and ARMA (non-integrated ARIMA) algorithms. ARIMA achieved the best R^2 score of 99 as it was able to leverage differencing, autoregression and moving average in a single model. On the other hand, it performed significantly worse when used in a prediction horizon of 6 days in the aforementioned study by Ribeiro et al. It had better forecasts for 1- and 3-days prediction, but the application of these exceptionally short forecasting periods was highly limited.

4.2. Performance of Deep Learning Models

Chimmula and Zhang [9] trained two LSTM models to predict short-term and long-term infections: LSTM model-1 was both trained and tested with a Canadian dataset whilst LSTM model-2 was trained by an Italian dataset and later tested on a Canadian dataset. LSTM model-1 achieved a short-term accuracy of 93.4%, a long-term accuracy of 92.67% and a short-term RMSE error of 34.83. LSTM model-2 on the other hand, had a higher RMSE error of 51.46. This was to be expected as Canada and Italy had different transmission dynamics due to cultural, societal and regulatory factors. The advantage of LSTM over traditional statistical methods stemmed from its ability to select hyperparameters by fitting real-time data and without assumptions, this allowed models to be more readily translated into different scenarios.

In addition, Bi, Fili and Hu [10] tested the performance of GRU against other models for 6-, 12-, 18-, 24- and 30-day periods. It was found that GRU outperformed LASSO for all but the 6-day prediction with a MAE that ranged from 19.62% to 32.72% whilst LASSO's MAE was between 25.71% and 42.85%. Compared to LASSO, GRU was able to give more accurate forecast for longer prediction windows. This could be explained by the complex interactions between the input variables that required models with more intricate and high-level features. The forecasting itself was thus better suited for GRU than linear regression-based models. However, it is important to acknowledge that LASSO has better interpretability than deep learning methods, so it provides great value in basic analysis. Lastly, Zeroual

DOI: 10.54254/2755-2721/4/2023404

et al. [11] found that both LSTM and GRU had much better performance than simple RNN but with poor explained variance due to insufficient training data.

Overall, despite the potential to outperform traditional statistical methods, all of the aforementioned ML-based models required large amounts of data for training, and this tends to be difficult to obtain in the early stages of a pandemic, which is usually when forecasting is needed the most. The deep learning algorithms in particular, also have lower interpretability due to their potential blackbox nature and thus must be employed with caution.

Table 1. Performance of forecasting models.

Evaluation Parameter	Model	Performance	Forecasting Window	Region	Paper
R^2	LR	86.90	5 days	India	[5]
	LASSO	87.20			
	Poly LR	98.64			
	Poly LASSO	98.15			
		98.00	10 days	Worldwide	[4]
	SVM	59.00			
	ES	98.00			
		46.20	2 months	Japan	[7]
		89.20			
	ARIMA	99.00	1 month	Saudi Arabia	[8]
Average sMAPE	SVR	7.95	6 days	Brazil	[6]
	ARIMA	9.13			
RMSE	LSTM Model-1	34.83	2 to 14 days	Canada	[9]
	LSTM Model-2	51.46			
MAE	GRU	19.62% - 32.72%	6 to 30 days	United States	[10]
	LASSO	25.71% - 42.85%			

5. Conclusion

This work reviewed six major machine learning approaches used in Covid-19 forecasting and compared the performance of each. There is no definitive conclusion as to which model would be best suited for Covid-19 forecasting. Each algorithm had different levels of performance in different class predictions, but the extended models typically achieved better performance than the simpler model that they are based on. For example, both LSTM and GRU outperformed basic RNN. Nevertheless, the level of performance varied drastically in different scenarios, and showed that Covid-19 forecast remains a difficult task despite rapid developments in ML algorithms.

There are two major limitations to this review and more broadly speaking, current research on ML-based Covid-19 prediction. To begin, almost all of the simulations were performed in silico by dividing a single dataset into a training and testing set. The models therefore had not undergone external testing by forecasting for a completely independent dataset. The models' accuracy in real-world environments thus could not be validated.

Moreover, this review only compared a limited selection of algorithms, and there are further variations within the reviewed algorithms that were overlooked. For example, different parameter

DOI: 10.54254/2755-2721/4/2023404

combinations of ARIMA were not considered. In addition, due to a lack of consistent experimental procedures and evaluation parameters, the results of comparison and review could not have been definitive.

For future studies and reviews, it is essential that external testing is implemented to assess the accuracy of models in response to completely new sets of data. More simulations on a wider selection of algorithms with consistent methodology and a common set of evaluation parameters are also encouraged to allow more direct and explicit comparisons. Lastly, the effectiveness and accuracy of combined models such as stacking-ensemble learning could also be considered.

References

[1] E. Dong, H. Du, and L. Gardner, 'An interactive web-based dashboard to track COVID-19 in real time', Lancet Infect. Dis., vol. 20, no. 5, pp. 533–534, May 2020, doi: 10.1016/S1473-3099(20)30120-1.

[2] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, 'Artificial Intelligence (AI) applications for COVID-19 pandemic', Diabetes Metab. Syndr. Clin. Res. Rev., vol. 14, no. 4, pp. 337–339, Jul. 2020, doi: 10.1016/j.dsx.2020.04.012.

[3] Y. Mohamadou, A. Halidou, and P. T. Kapen, 'A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19', Appl. Intell., vol. 50, no. 11, pp. 3913–3925, Nov. 2020, doi: 10.1007/s10489-020-01770-9.

[4] F. Rustam et al., 'COVID-19 Future Forecasting Using Supervised Machine Learning Models', IEEE Access, vol. 8, pp. 101489–101499, 2020, doi: 10.1109/ACCESS.2020.2997311.

[5] V. Bhadana, A. S. Jalal, and P. Pathak, 'A Comparative Study of Machine Learning Models for COVID-19 prediction in India', in 2020 IEEE 4th Conference on Information & Communication Technology (CICT), Chennai, India, Dec. 2020, pp. 1–7. doi: 10.1109/CICT51604.2020.9312112.

[6] M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, and L. dos S. Coelho, 'Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil', Chaos Solitons Fractals, vol. 135, p. 109853, Jun. 2020, doi: 10.1016/j.chaos.2020.109853.

[7] H. Yonar, 'Modeling and Forecasting for the number of cases of the COVID-19 pandemic with the Curve Estimation Models, the Box-Jenkins and Exponential Smoothing Methods', Eurasian J. Med. Oncol., 2020, doi: 10.14744/ejmo.2020.28273.

[8] S. I. Alzahrani, I. A. Aljamaan, and E. A. Al-Fakih, 'Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions', J. Infect. Public Health, vol. 13, no. 7, pp. 914–919, Jul. 2020, doi: 10.1016/j.jiph.2020.06.001.

[9] V. K. R. Chimmula and L. Zhang, 'Time series forecasting of COVID-19 transmission in Canada using LSTM networks', Chaos Solitons Fractals, vol. 135, p. 109864, Jun. 2020, doi: 10.1016/j.chaos.2020.109864.

[10] L. Bi, M. Fili, and G. Hu, 'COVID-19 forecasting and intervention planning using gated recurrent unit and evolutionary algorithm', Neural Comput. Appl., May 2022, doi: 10.1007/s00521-022-07394-z.

[11] A. Zeroual, F. Harrou, A. Dairi, and Y. Sun, 'Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study', Chaos Solitons Fractals, vol. 140, p. 110121, Nov. 2020, doi: 10.1016/j.chaos.2020.110121.