



A Study of Data-Driven Methods for Adaptive Forecasting of COVID-19 Cases

Charithea Stylianides¹, Kleantlis Malialis¹, and Panayiotis Kolios¹

KIOS Research and Innovation Center of Excellence,
University of Cyprus, Nicosia, Cyprus

{stylianides.charithea,malialis.kleantlis,kolios.panayiotis}@ucy.com.cy

Abstract. Severe acute respiratory disease SARS-CoV-2 has had a profound impact on public health systems and healthcare emergency response especially with respect to making decisions on the most effective measures to be taken at any given time. As demonstrated throughout the last three years with COVID-19, the prediction of the number of positive cases can be an effective way to facilitate decision-making. However, the limited availability of data and the highly dynamic and uncertain nature of the virus transmissibility makes this task very challenging. Aiming at investigating these challenges and in order to address this problem, this work studies data-driven (learning, statistical) methods for incrementally training models to adapt to these nonstationary conditions. An extensive empirical study is conducted to examine various characteristics, such as, performance analysis on a per virus wave basis, feature extraction, “lookback” window size, memory size, all for next-, 7-, and 14-day forecasting tasks. We demonstrate that the incremental learning framework can successfully address the aforementioned challenges and perform well during outbreaks, providing accurate predictions.

Keywords: incremental learning · data streams · neural networks · time-series forecasting

1 Introduction

The COVID-19 pandemic has caused a massive disruption to society since its emergence in December 2019. An unprecedented number of people were infected, hospitalized and had COVID-19 being their leading cause of death. Moreover, the consequences of the pandemic are still impacting our social and economic ecosystems. Evidently, many countries still impose restrictions and measures

This work was supported by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 739551 (KIOS CoE - TEAMING) and from the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy. It was also supported by the CIPHIS (Cyprus Innovative Public Health ICT System) project of the NextGenerationEU programme under the Republic of Cyprus Recovery and Resilience Plan under grant agreement C1.112.

based on the evolution of the infected population. Hence, effective modelling and prediction of the evolution of the viral load in the society can be of detrimental factor in decision making. By taking proactive measures for closures and lockdowns, restricting public events, health guidelines and vaccine policies, governments can increase their effectiveness and limit transmissibility. A way to capture the spread of the virus is by tracking and predicting the number of positive cases. This constitutes a challenging task because of:

Data non-stationarity. The data exhibit a highly dynamic behaviour, i.e., the data distribution evolves over time [6]. In the COVID-19 case, for instance, there have been many variants of the virus (e.g., Delta and Omicron), as well as many measures which have been imposed (e.g., vaccination and school closure).

Limited data. This refers to the problem of having limited availability of historical data. Evidently most countries reported positive cases on a daily basis which accumulates to a mere 365 data points over the course of a year.

As a result, it is necessary to have an online learning model which is able to adapt to non-stationary environments, and to be incrementally trained from limited data. The contributions of this work are the following.

- The primary focus of this study is on COVID-19 cases forecasting for Cyprus, a European country with a population of around one million.
- We conduct an extensive empirical analysis where we examine the roles of (i) traditional/offline vs online incremental learning; (ii) “look-back” window size; (iii) feature extraction; (iv) memory size; (v) learning (neural network) vs statistical (ARIMA) models. Furthermore, all these are considered in three tasks (next-, 7-, and 14-day forecasting) and we provide a per-wave analysis.

The remainder of the paper is structured as follows. Section 2 discusses work related to ours. Section 3 describes the problem formulation and the incremental learning framework for adaptive forecasting. The experimental setup and results are provided in Sect. 4 and 5 respectively. We conclude in Sect. 6.

2 Related Work

2.1 Compartmental Models

These models, like the well-known Susceptible Exposed Infectious Recovered (SEIR) [11] and any variations of it, split the population into mutually exclusive states that describe a path of infection dynamics through mathematical modelling [15]. For maximum accuracy, studies [4, 10] have deduced parameters describing the transition between states that are time-varying capturing the social changes, medical advancements and non-pharmaceutical interventions during a pandemic [10]. For example, the “DELPHI” model [15] consists of 11 compartments and forecasts detected cases and deaths for about 2 weeks, accounting for government measures and limited population testing. The vast majority of existing work on COVID-19 cases forecasting lie within this domain.

2.2 Data-Driven Methods

The focus of our work is on data-driven methods. Forecasting using data-driven methods can also be successfully achieved through statistical and machine learning methods. Isaac B. et al. [3] compared the performance of a model that combines Convolutional and Long Short Term Memory (LSTM) layers to that of a standard neural network, using a 14-day window of positive cases to predict those of the next seven days both at the regional and national level. Several studies have compared the performance of LSTM to that of other models including Recurrent Neural Networks (RNNs) [2], Gradient Boosting Trees [18] and the statistical model ARIMA [12]. Research includes time series of just confirmed cases for generalizability [3], added features like number of cured patients and deaths [12] and aggregated features [12, 18] for improved accuracy. The superiority of the LSTM is concluded in all the last comparisons. Another comparative study in [25] used LSTM, RNN, Bidirectional LSTM, Gated Recurrent Units (GRUs) and Variational AutoEncoder (VAE) to predict new and recovered cases for the next 17-days where VAE showed the best performance. Encoders of self-attention and recurrent layers, that consider among other factors travelling from each country to predict the spread were also proposed in [13].

The aforementioned methods consider offline learning. Continual or online learning is starting to be used to capture the concept of drift in the spread of COVID-19 and adapt models in real time. The study in [24] evaluates the best number of training samples needed at each time step to minimize the prediction error and, thus, capture drift. Ridge regression is used for predictions of hospitalizations from new cases, severe cases from hospitalizations and deaths from severe cases for the next 7 d using 14-day windows [24]. In [23] an ensemble of regression models predicts 30-day mortality allowing for adaptation of the models at every instance by i) fitting them again on the whole data, ii) fitting them again on just the new instance, or iii) fitting a completely new ensemble on the new data. A linear model with LASSO (least absolute shrinkage and selection operator) penalty [16] and a feed-forward network with autoregressive input (predictions at each time step used for training for the next forecast) [22], were also able to incrementally train and produce 2-day cases predictions [16] and 30-day predictions of hospitalizations and deaths [22], respectively.

2.3 Hybrid

A study [5] has used data from a compartmental model (exposed, infected, recovered and dead population) to evaluate the best lags of each out of time series windows in an ARIMA model, for predicting each of the variables and susceptible population. Based on this, new data is then continuously bootstrapped out of a data stream, predicting and updating incrementally an ensemble of algorithms each time [5]. In [7, 8] incremental learning of a neural network provides 5 parameters (rate of infection during lockdown, time lockdown begins, rate of death, rate of recovery) needed for a Susceptible Infected Recovered Vaccinated Deceased (SIRVD) model. The SIRVD model forecasts monthly trajectories of deaths under different scenarios [7] and monthly total number of cases, active infections and deaths [8].

Algorithm 1 Data-driven framework for adaptive forecasting

Input:
 D : number of days to forecast
 W : “Lookback” window size
 M : Memory / queue size

- 1: Wait W days, observe instance $x^W = \{n^1, \dots, n^W\}$
- 2: Create model $f^W.init()$
- 3: Predict $\hat{y}^{W+1} = f^W.predict(x^W)$
- 4: **for** each time step $t \in [W + 1, W + D - 1]$ **do**
- 5: Get ground truth $y^t = n^t$
- 6: Observe instance $x^t = \{n^t, n^{t-1}, \dots, n^{t-W+1}\}$
- 7: Predict $\hat{y}^{t+1} = f^W.predict(x^t)$
- 8: Initialise memory $q^t = \{\}$
- 9: **for** each time step $t \in [W + D, \infty)$ **do**
- 10: Get ground truth $y^t = n^t$
- 11: Observe $y^t = \{n^t, n^{t-1}, \dots, n^{t-D+1}\}$
- 12: Append example to memory $q^t = q^{t-1}.append((x^{t-D}, y^t))$
- 13: Incremental training $f^t = f^{t-1}.train(q^t)$
- 14: Observe instance $x^t = \{n^t, n^{t-1}, \dots, n^{t-W+1}\}$
- 15: Predict $\hat{y}^{t+1} = f^t.predict(x^t)$

3 Incremental Learning Framework for Adaptive Forecasting

We consider a data generating process $S = \{n^t\}_{t=1}^T$ that provides at each day t a number $n^t \in \mathbb{R}$ of positive cases, from an unknown and evolving probability distribution $p^t(n)$, where $T \in [1, \infty)$. The instances constitute a univariate time series, and n^t corresponds to the number of COVID-19 cases on day t .

To address the temporal aspects of the data, we consider a sliding window of size $W \in \mathbb{Z}^+$, such that, $x^t = \{n^t, n^{t-1}, \dots, n^{t-W+1}\} \in \mathbb{R}^W$ is a W -dimensional vector belonging to input space $X \subset \mathbb{R}^W$. The task is to forecast $D \in \mathbb{Z}^+$ days of the COVID-19 cases, that is, at any day $t > W$ to predict $\hat{y}^{t+D} = \{\hat{n}^{t+D}, \dots, \hat{n}^{t+2}, \hat{n}^{t+1}\} \in \mathbb{R}^D$, a D -dimensional vector belonging to $Y \subset \mathbb{R}^D$.

A regression model f^t receives a new example $x^t \in \mathbb{R}^W$ at time step t and makes a prediction $\hat{y}^{t+D} \in \mathbb{R}^D$, based on a concept $f : X \rightarrow Y$ such that $\hat{y}^{t+D} = f^t(x^t)$. In this study, we will be using neural networks as our regression models, which they have been demonstrated to be effective incremental learners [17, 20, 21]. The loss function used between a prediction $\hat{y}^t \in \mathbb{R}^D$ and ground truth $y^t \in \mathbb{R}^D$ at time t is the Mean Squared Error (MSE) defined as:

$$J^t = MSE(\hat{y}^t, y^t) = \frac{1}{D} \sum_{d=1}^D (\hat{y}_d^t - y_d^t)^2, \quad (1)$$

The model is continually updated using incremental learning, which is defined as the gradual adaptation of a model without complete re-training, that is, $f^t = f^{t-1}.train((x^{t-D}, y^t))$. Learning is performed using incremental Stochastic

Gradient Descent where each neural network weight w is updated according to the formula $w^t \leftarrow w^{t-1} - \alpha \frac{\partial J^t}{w}$, where $\frac{\partial J^t}{w}$ is the partial derivative with respect to w , and α is the learning rate.

Furthermore, we introduce a memory component implemented as a queue q of size M , which stores historical examples. For instance, at time t , it will append to memory the example (x^{t-D}, y^t) , i.e., $m^t = m^{t-1}.append((x^{t-D}, y^t))$. As a result, incremental learning is now performed using $f^t = f^{t-1}.train(q^t)$, and the loss function is defined as the average MSE of all memory examples.

The framework’s pseudocode is shown in Alg. 1. Initially, we wait for W days (Line 1). Subsequently and until day $t < W + D$, we only perform prediction (i.e., forecasting) without any incremental training (Lines 2 - 7). From day $t \geq W + D$ we perform both prediction and incremental learning (Lines 8 - 15).

4 Experimental Setup

4.1 Dataset

Our data consist of reported daily SARS-CoV-2 cases in Cyprus from 15/10/20 to 08/10/22 as they appear in the TESSy platform of the European Center of Disease Prevention and Control (ECDC) in the RESPISURV dataset. Data preprocessing included creating sliding windows of 7, 14, 30 d and removal of daily cases of under 100 for reduced noise and easier learning of the models. Any missing values were imputed with the mean of their row or the previous row and data were normalized by dividing by maximum number of cases. Six periods of interest, referred to as “waves”, are considered in this study and are shown in Fig. 1. The time periods of each wave are as follows: Wave 1: 13/12/20-11/01/21; Wave 2: 04/04/21-03/05/21; Wave 3: 02/07/21-31/07/21; Wave 4: 19/12/21-07/01/22; Wave 5: 17/06/22-26/07/22. Also, we will be referring to the remaining (i.e., non-waves) time period as “normal”.

4.2 Compared Methods

The compared methods follow the same framework shown in Algorithm 1.

MLP. It refers to the standard feed-forward, fully-connected Multilayer Perceptron (MLP) model. In all experiments, its hyper-parameters are: He Normal [9] weight initialisation, the Adam [14] optimisation algorithm, LeakyReLU [19] and ReLU for the hidden and output activation function respectively, the MSE loss function, and mini-batch size of one. The rest of them (architecture, learning rate, regularisation, and number of epochs) slightly vary for each experiment.

ARIMA. The Autoregressive Integrated Moving Average (ARIMA) model is a classical forecasting method. Despite the fact that ARIMA is often considered as a baseline method, it is emphasised that due to the limited availability of historical data, it is actually demonstrated to be very effective particularly during normal and small outbreaks. In all experiments, its hyper-parameters are number of lagged observations for auto-regression: 1, number of times the raw observations are differenced: 0 and moving average window size: 0.

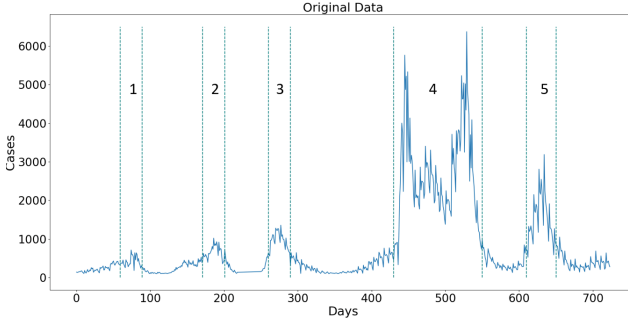


Fig. 1. COVID-19 cases in Cyprus (15/10/20-18/10/22)

4.3 Evaluation Method and Metrics

To evaluate and compare the aforementioned methods, we have been using the following widely adopted metrics for regression forecasts.

MAE. This refers to the Mean Absolute Error (MAE).

MAPE. This refers to the Mean Absolute Percentage Error (MAPE) between actual $y^t \in \mathbb{R}^D$ and predicted $\hat{y}^t \in \mathbb{R}^D$ values as defined below:

$$MAPE(\hat{y}, y) = \frac{100\%}{N} \sum_{d=1}^D \frac{|y_d - \hat{y}_d|}{|y_d|}, \quad (2)$$

For all experiments involving neural networks, we run each one over 10 repetitions and provide the average and standard deviation for both metrics, for an overall time period, as well as during wave and normal periods.

5 Experimental Results

5.1 Role of Incremental Learning

This section compares the performance of MLP using offline learning to that of online incremental learning. For offline learning, the MLP was pre-trained on one month of data and with no further training. For the next-day prediction task, Fig. 2 shows the results of the two paradigms on a daily basis, while Table 1 provides the relevant aggregated metrics. The standard deviation of the error is shown in brackets. The corresponding results for the 7- and 14-day forecasting tasks are shown in Table 2 and Table 3. It is observed that online learning significantly outperforms offline learning in all tasks and all time periods.

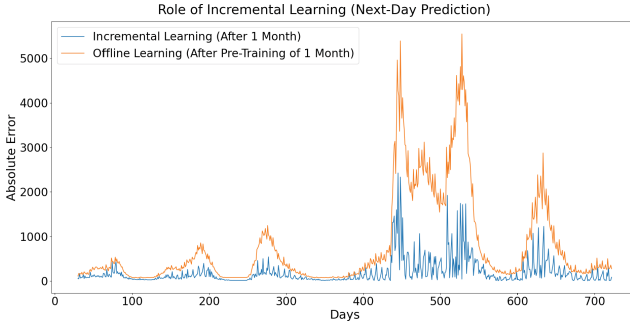


Fig. 2. Online vs offline learning for next-day predictions

Table 1. MLP with online vs offline learning for next-day predictions

	Overall		Waves		Normal	
	MAE	MAPE	MAE	MAPE	MAE	MAPE
Online	186.1 (47.4)	26.3 (14.3)	378.4 (62.5)	25.5 (9.1)	148.2 (50.4)	26.7 (17.9)
Offline	767.9 (405.3)	83.9 (43.8)	1714.1 (934.0)	96.7 (57.1)	588.2 (302.7)	79.9 (40.5)

Table 2. MLP with online vs offline learning for 7-day predictions

	Overall		Waves		Normal	
	MAE	MAPE	MAE	MAPE	MAE	MAPE
Online	342.3 (48.0)	43.7 (5.3)	714.1 (105.8)	46.0 (5.0)	267.9 (36.5)	43.5 (5.5)
Offline	598.7 (59.4)	65.4 (6.3)	1319.8 (140.0)	75.1 (7.3)	461.2 (45.0)	62.7 (6.5)

Table 3. MLP with online vs offline learning for 14-day predictions

	Overall		Waves		Normal	
	MAE	MAPE	MAE	MAPE	MAE	MAPE
Online	524.2 (42.1)	75.9 (4.1)	1011.1 (97.9)	67.9 (4.9)	413.6 (31.6)	74.5 (4.3)
Offline	686.7 (47.2)	83.6 (7.8)	1416.5 (101.4)	88.3 (10.4)	546.3 (35.5)	82.0 (6.9)

5.2 Role of the Sliding Window Size

This section examines the impact of the sliding window size on the performance of MLP using incremental learning. Results are provided in Table 4 and Table 5 for next- and seven-day prediction tasks, respectively.

Using MLP, the 7-day window performs better compared to the 14-day and 30-day ones for all prediction tasks. Regarding average performance across waves, the 30-day window performs best for next-day and 7-day prediction and the 7-day window performs best for 14-day prediction task (not shown here). Normal periods benefit the most from a 7-day window for all prediction tasks. While not shown due to space restrictions, for ARIMA, a 30-day window performs the best on the overall data, wave and normal periods for all tasks.

Table 4. MLP with 7-, 14- and 30-day sliding window (next-day prediction)

	Overall		Waves		Normal	
Window	MAE	MAPE	MAE	MAPE	MAE	MAPE
7	186.1 (47.4)	26.3 (14.3)	378.4 (62.5)	25.5 (9.1)	148.2 (50.4)	26.7 (17.9)
14	343.9 (258.0)	48.2 (28.8)	698.6 (569.1)	42.5 (30.9)	270.0 (201.5)	47.4 (29.9)
30	187.8 (20.8)	28.3 (7.1)	368.5 (23.0)	25.2 (5.2)	154.1 (23.6)	28.6 (8.7)

Table 5. MLP with 7-, 14- and 30-day sliding window (7-day prediction)

	Overall		Waves		Normal	
Window	MAE	MAPE	MAE	MAPE	MAE	MAPE
7	342.3 (48.0)	43.7 (5.3)	714.1 (105.8)	46.0 (5.0)	267.9 (36.5)	43.5 (5.5)
14	378.0 (43.5)	46.5 (4.7)	795.5 (95.5)	48.8 (4.9)	292.6 (32.9)	45.2 (4.6)
30	350.6 (38.9)	46.4 (5.0)	709.2 (81.1)	44.9 (4.7)	281.6 (32.1)	45.9 (5.5)

The better performance of a 7-day window can be attributed to the fewer window days suggesting more recent data, which can increase performance. On the other hand, it is speculated that a 30-day window works best because of the more data and fluctuations considered.

5.3 Role of Feature Extraction

This section describes the role of 20 features in our model, aggregated across a 14-day window. The features are: school closing strictness (mean), public events cancellation strictness (mean), positive cases (min, max), unvaccinated cases (min, median), second dose vaccinated population (min, range), second dose vaccinated cases (mean, median), first dose vaccinated cases (median, mean), weekly deaths (mean), workplace closing strictness (mean), weekly ICU cases (mean), weighted stringency index (median), recovered (s.d.), 70+ aged cases (mean), first dose vaccinated population (median) and 18–24 aged cases (mean).

The results for next- and seven-day prediction tasks are shown in Table 6 and Table 7, respectively. Using the features, MAPE is reduced by 8.3%, 7.2% and 0.2% for overall, wave and normal periods, respectively, for the 7-day prediction. The features seem to be more informative when making later predictions.

5.4 Role of the Memory Size

In this section, the role of the memory size using i) raw data and ii) features is assessed. For these experiments, raw data was used in 7-day windows for 14-day predictions and features were extracted from 14-day windows for 7-day predictions. Window size here is chosen based on best windows for raw data and features, respectively, as stated in Sects. 5.2 and 5.3. The results for the memory use with raw data are shown in Table 8 and with features in Table 9.

In the first case, it is deduced that increasing memory size improves overall, wave and normal periods performance by up to 17%, 12.7% and 16.1%, respectively. Interestingly, in the second case, using memory decreases performance.

Table 6. MLP with raw data vs features (next-day prediction)

	Overall		Waves		Normal	
Data	MAE	MAPE	MAE	MAPE	MAE	MAPE
Raw	186.1 (47.4)	26.3 (14.3)	378.4 (62.5)	25.5 (9.1)	148.2 (50.4)	26.7 (17.9)
Features	211.5 (42.1)	30.5 (11.8)	423.2 (57.1)	29.4 (9.1)	172.0 (43.8)	31.8 (14.1)

Table 7. MLP with raw data vs features (7-day prediction)

	Overall		Waves		Normal	
Data	MAE	MAPE	MAE	MAPE	MAE	MAPE
Raw	342.3 (48.0)	43.7 (5.3)	714.1 (105.8)	46.0 (5.0)	267.9 (36.5)	43.5 (5.5)
Features	284.7 (11.1)	35.4 (2.9)	603.1 (15.4)	38.8 (2.0)	238.7 (13.5)	43.3 (5.5)

Table 8. MLP performance with raw data per memory size (7-day window, 14-day prediction)

	Overall		Waves		Normal	
Memory	MAE	MAPE	MAE	MAPE	MAE	MAPE
1	524.2 (42.1)	75.9 (4.1)	1011.1 (97.9)	67.9 (4.9)	413.6 (31.6)	74.5 (4.3)
30	570.0 (35.7)	98.7 (3.8)	981.1 (93.9)	67.9 (4.1)	464.3 (27.5)	98.2 (4.4)
90	505.9 (50.7)	78.9 (5.8)	917.8 (106.9)	63.4 (5.0)	421.8 (38.9)	82.9 (6.2)
180	428.7 (35.2)	65.5 (3.4)	798.2 (77.2)	57.8 (3.9)	350.3 (25.5)	67.0 (3.2)
240	445.1 (40.4)	65.6 (3.7)	846.8 (90.6)	59.0 (4.3)	349.8 (32.1)	63.9 (4.0)
360	418.7 (46.1)	58.9 (4.3)	820.7 (97.7)	55.2 (5.2)	329.3 (36.6)	58.4 (4.7)

Table 9. MLP performance with features per memory size (14-day window, 7-day prediction)

	Overall		Waves		Normal	
Memory	MAE	MAPE	MAE	MAPE	MAE	MAPE
1	284.7 (11.1)	35.4 (2.9)	603.1 (15.4)	38.8 (2.0)	238.7 (13.5)	43.3 (5.5)
30	513.8 (13.4)	84.9 (0.9)	880.1 (37.3)	56.1 (2.0)	440.2 (11.3)	94.4 (4.4)
90	622.8 (3.7)	132.2 (0.9)	912.2 (5.6)	51.9 (1.4)	581.1 (13.6)	158.2 (7.1)
180	702.7 (2.0)	142.5 (0.6)	1071.6 (2.2)	57.2 (0.6)	580.3 (7.9)	147.6 (4.5)
240	696.7 (3.5)	134.2 (0.9)	1115.2 (5.5)	57.3 (1.4)	537.8 (13.8)	130.1 (6.9)
360	651.0 (2.2)	110.4 (0.6)	1181.4 (2.5)	57.0 (0.6)	496.7 (8.1)	107.7 (4.6)

5.5 Comparative Study

This section aims to compare the best MLP experiments in this study to the traditional forecasting ARIMA method. Results refer to overall, wave and normal periods, as well as each wave. For next-day predictions, they are reported in Table 10 and Table 11, and for 14-day predictions in Table 12 and Table 13. Next-day prediction learning curves for the two models are shown in Fig. 3.

It is observed that for next-day predictions, the neural network outperforms ARIMA at Waves 4 and Wave 5 (Table 11), with MAE of 529.1 (against 557.1) at Wave 4 and MAE of 371.5 (against 385.4) at Wave 5. For 14-day predictions, MLP captures the data distribution shift at Wave 4 (Table 13) better than ARIMA with MAE of 1209 (against 1433.5).

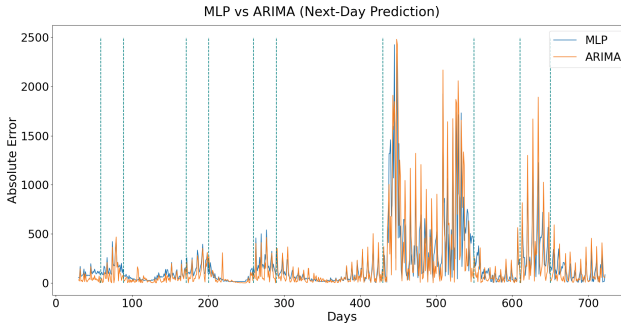


Fig. 3. MLP vs ARIMA (next-day prediction)

Table 10. MLP vs ARIMA (next-day prediction)

	Overall		Waves		Normal	
Model	MAE	MAPE	MAE	MAPE	MAE	MAPE
MLP	186.1 (47.4)	26.3 (14.3)	378.4 (62.5)	25.5 (9.1)	148.2 (50.4)	26.7 (17.9)
ARIMA	176.2	20.3	371.7	20.5	137.6	19.6

Table 11. MLP vs ARIMA per wave (next-day prediction)

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Model	MAE	MAE	MAE	MAE	MAE
MLP	167.3 (131.5)	168.2 (174.4)	206.4 (264.4)	529.1 (13.8)	371.5 (18.2)
ARIMA	103.5	123.5	128.4	557.1	385.4

Table 12. MLP vs ARIMA (14-day prediction)

	Overall		Waves		Normal	
Model	MAE	MAPE	MAE	MAPE	MAE	MAPE
MLP	418.7 (46.1)	58.9 (4.3)	820.7 (97.7)	55.2 (5.2)	329.3 (36.6)	58.4 (4.7)
ARIMA	412.7	42.8	861.6	48.3	322.6	40.9

Table 13. MLP vs ARIMA per wave (14-day prediction)

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Model	MAE	MAE	MAE	MAE	MAE
MLP	247.6 (29.2)	339.8 (50.5)	392.5 (58.8)	1209.0 (153.5)	767.4 (65.4)
ARIMA	142.9	246.3	297.9	1433.5	569.3

6 Conclusions and Future Work

The COVID-19 virus has been acutely affecting millions of people for more than three years. In valuable attempts for prompt government interventions and addressing data non-stationarity and availability, we have conducted an empirical study of data-driven (learning, statistical) methods using incremental training for adaptive forecasting of COVID-19 cases. Some future directions are:

Role of the memory. The impact of memory is unclear. We have demonstrated its effectiveness on the performance of MLP with raw data, however, performance declined for MLP with features. Future work will investigate this.

Statistical models with features. In this study we examine the impact of features in MLP. We plan to use ARIMAX [1] to incorporate features to ARIMA.

Advanced neural architectures. Future work will investigate more complex neural architectures, such as, autoregressive networks and LSTMs.

References

1. Aji, B.S., Rohmawati, A.A., et al.: Forecasting number of Covid-19 cases in Indonesia with arima and arimax models. In: 2021 9th International Conference on Information and Communication Technology (ICoICT), pp. 71–75. IEEE (2021)
2. Alassafi, M.O., Jarrah, M., Alotaibi, R.: Time series predicting of Covid-19 based on deep learning. *Neurocomputing* **468**, 335–344 (2022)
3. Boyd, I., Hedges, D., Carter, B.T., Whitaker, B.M.: Using neural networks to model the spread of COVID-19. In: 2022 Intermountain Engineering, Technology and Computing (IETC), IEEE (2022)
4. Calafiore, G.C., Novara, C., Possieri, C.: A time-varying SIRD model for the Covid-19 contagion in Italy. *Annu. Rev. Control.* **50**, 361–372 (2020)
5. Camargo, E., Aguilar, J., Quintero, Y., Rivas, F., Ardila, D.: An incremental learning approach to prediction models of SEIRD variables in the context of the COVID-19 pandemic. *Health Technol. (Berl.)* **12**(4), 867–877 (2022)

6. Ditzler, G., Roveri, M., Alippi, C., Polikar, R.: Learning in nonstationary environments: a survey. *IEEE Comput. Intell. Mag.* **10**(4), 12–25 (2015)
7. Farooq, J., Bazaz, M.A.: A novel adaptive deep learning model of Covid-19 with focus on mortality reduction strategies. *Chaos Solitons Fractals* **138**(110148), 110148 (2020)
8. Farooq, J., Bazaz, M.A.: A deep learning algorithm for modeling and forecasting of COVID-19 in five worst affected states of India. *Alex. Eng. J.* **60**(1), 587–596 (2021)
9. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034 (2015)
10. IHME COVID-19 Forecasting Team: Modeling COVID-19 scenarios for the united states. *Nat. Med.* **27**(1), 94–105 (2021)
11. Kermack, W.O., McKendrick, A.G., Walker, G.T.: A contribution to the mathematical theory of epidemics. *Proc. Royal Society of London. Series A, Containing Papers of a Math. Phys. Character* **115**(772), 700–721 (1927)
12. Ketu, S., Mishra, P.K.: India perspective: CNN-LSTM hybrid deep learning model-based COVID-19 prediction and current status of medical resource availability. *Soft. Comput.* **26**(2), 645–664 (2022)
13. Kim, M., et al.: Hi-covidnet: Deep learning approach to predict inbound covid-19 patients and case study in south korea. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3466–3473. KDD '20, Association for Computing Machinery, New York, NY, USA (2020)
14. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: *Proceedings of the 3rd International Conference on Learning Representations (ICLR)* (2015)
15. Li, M.L., Bouardi, H.T., Lami, O.S., Trikalinos, T.A., Trichakis, N., Bertsimas, D.: Forecasting COVID-19 and analyzing the effect of government interventions. *Oper. Res.* **71**(4) (2022)
16. Liu, D., et al.: Real-time forecasting of the COVID-19 outbreak in chinese provinces: Machine learning approach using novel digital data and estimates from mechanistic models. *J. Med. Internet Res.* **22**(8), e20285 (2020)
17. Losing, V., Hammer, B., Wersing, H.: Incremental on-line learning: a review and comparison of state of the art algorithms. *Neurocomputing* **275**, 1261–1274 (2018)
18. Luo, J., Zhang, Z., Fu, Y., Rao, F.: Time series prediction of COVID-19 transmission in america using LSTM and XGBoost algorithms. *Results Phys.* **27**(104462), 104462 (Aug2021)
19. Maas, A.L., Hannun, A.Y., Ng, A.Y.: Rectifier nonlinearities improve neural network acoustic models. In: *Proceedings of the 30th International Conference on Machine Learning* (2013)
20. Malialis, K., Panayiotou, C.G., Polycarpou, M.M.: Online learning with adaptive rebalancing in nonstationary environments. *IEEE Trans. Neural Netw. Learn. Syst.* **32**(10), 4445–4459 (2021)
21. Malialis, K., Panayiotou, C.G., Polycarpou, M.M.: Nonstationary data stream classification with online active learning and Siamese neural networks. *Neurocomputing* **512**, 235–252 (2022)
22. Rodríguez, A.: Deepcovid: an operational deep learning-driven framework for explainable real-time Covid-19 forecasting. *Proc. AAAI Conf. Artif. Intell.* **35**(17), 15393–15400 (2022)
23. Tetteroo, J., Baratchi, M., Hoos, H.H.: Automated machine learning for Covid-19 forecasting. *IEEE Access* **10**, 94718–94737 (2022)

24. Uchida, T., Yoshida, K.: Concept drift in Japanese Covid-19 infection data. *Proc. Comput. Sci.* **207**, 380–387 (2022) *knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 26th International Conference KES2022*
25. Zeroual, A., Harrou, F., Dairi, A., Sun, Y.: Deep learning methods for forecasting covid-19 time-series data: a comparative study. *Chaos, Solitons & Fractals* **140**, 110121 (2020)