Sliding window time series forecasting with multilayer perceptron and multiregression of COVID-19 outbreak in Malaysia

Norita Md Norwawi

FACULTY OF SCIENCE AND TECHNOLOGY, UNIVERSITI SAINS ISLAM MALAYSIA, NILAI, NEGER SEMBILAN, MALAYSIA

1. Introduction

Novel coronavirus disease is an infectious disease caused by a coronavirus first detected in Wuhan on Dec. 31, 2019. The World Health Organization (WHO) named the disease coronavirus disease 2019 (COVID-19); the virus is otherwise known as severe acute respiratory syndrome coronavirus 2 [17]. Symptoms of people infected with the COVID-19 virus are mainly, such as coughing, persistent chest pain, and difficulty breathing. However, some infected people become do have symptoms or feel unwell. To control the spread of disease, governments around the globe declared a Movement Control Order (MCO) to break the chain of COVID-19 infections. Compliance with the MCO aims to mitigate the risk for a spike in the pandemic, which is imperative in controlling its spread.

In Malaysia, a country of 31 million people, over 6000 individuals have been infected with COVID-19 and 103 fatalities have been reported. The disease was first reported on Jan. 25 in three travelers from China. The first sporadic case was reported on Mar. 11, which suggests that he did not contract the disease owing to traveling overseas or close contact. Since then, there has been a sharp increase in the trend of people infected with positive COVID-10. As of Apr. 30, 2020, of the 6002 COVID-19 cases reported in the country, 69.49% have been cured and the fatality rate has been 1.7%. Malaysia's strategies in attempting to break the chain of infection through the four phases of the MCO since Mar. 18, 2020 with a 2-week duration for each phase show a remarkable result so far [4,8].

The objectives of this study are to analyze the COVID-19 outbreak in Malaysia from Jan. 2020 until end of Apr. 2020, to assess the trends in the near future, and to estimate the future trend using a sliding window time series forecasting method.

This chapter is organized into five sections. Section 2 discusses some related work with temporal data mining and time series forecasting methods. Section 3 presents the dynamic sliding window technique used in this study. Section 4 discusses the findings of the trend analysis and the sliding window time series forecasting model with temporal data analytics, and Section 5 concludes the chapter.

2. Related work

2.1   Time series forecasting for coronavirus disease 2019

Time series forecasting involves time series data that have a natural temporal ordering. It is usually defined as $y = f(t)$, where time t is an independent variable and where y is the dependent variable whose value depends on time. Time series data are useful for making projections, studying trends, predicting an event, extracting meaningful statistics, and so on.

The primary purpose of time series forecasting is to understand the past and predict the future with the model developed before they are measured. It may be in the univariate or multivariate form, which requires a specific approach to manipulating and deriving solutions. With the outbreak of the coronavirus pandemic around the world, it is important for us to understand the nature of the COVID-19 outbreak and characteristics to search for solutions.

Achintya et al. [2] and Gupta and Pal [13] presented a profile of the COVID-19 situation in India with polynomial fitting parameters and predicted the future trend using the Auto Regressive Integrated Moving Average (ARIMA) model. This projection may help government agencies, businesses, and others make decisions and strategize their defensive tactical solutions in facing the spreading of disease.

On the other hand, the world came to a consensus to lock down countries to mitigate risk. An MCO was declared by most governments to slow the spread of the epidemic. Naomie et al. [9] presented an analysis of the impact of the MCO in lessening the number of infected cases. Ahmadi et al. [3] studied the impact of climatology parameters on the COVID-19 outbreak in Iran using a sensitivity analysis. They found that low values of wind speed, humidity, and solar radiation exposure to a high rate of infection may help in surviving the coronavirus. Moreover, locations with a high population density, intraprovincial movement, and a high humidity rate are more at risk. These are some of the interesting studies conducted to understand the nature of the pandemic and gathering efforts to combat the deadly infectious disease.

2.2   Data mining and artificial intelligence solutions combatting coronavirus disease 2019

With advances in the sophistication of technology, the world is working closely together to overcome this dangerous situation. Thousands of data being transmitted or transported around the globe at such high speed and efficiently shared around the world. In many ways, solutions are rapidly being produced owing to the urgency of the situation for which timeliness is of prime concern. Speed in arriving at practical and realistic solutions is the new norm.

An important technology that has had an important role is artificial intelligence and data analytics. For example, Rao and Vazquez [14] adopted a machine learning algorithm to speed possible case identifications of COVID-19. Abdullah Farid et al. [1] proposed predicting the probability of recurrences in no-recurrence (first-time detection) cases in screening coronavirus disease. In Wuhan, Huang [19] studies a spatiotemporal analysis of the spread of the virus that gives insight into preparing for and mitigating future virus control. They used the logistic and Susceptible, Exposed, Infectious and Removed (SIER) model to analyze and predict future trends.

In this study, an artificial neural network (ANN) is used to identify characteristics of COVID-19 time series data that may have some interesting patterns to be extracted. Hota et al. [5] mentioned that ANN is a favorable technique that can map highly nonlinear input-output data samples, unlike any statistical regression model. They used the sliding window as an estimation over the actual value of time series data. Fig. 29.1 shows the sliding window technique being used in ANN [12]. Vafaeipour et al. [7] adopted a similar technique to predict the wind velocity time series.

The sliding window method is an approach adopted in this study. It is a common method for recording temporal sequences owing to its simplicity and intuitiveness [11].

3. Sliding window technique for temporal data analytics

The sliding window is a method that can be used in the preprocessing phase to restructure data according to a time frame into a supervised classification problem. Given a time series dataset that can be explained with $y = f(x)$, using a sliding window, the data will be restructured in which the previous output or the historical data will be used to predict the next output. For example, for window width 1, the outcome of current time t is based on the output of prior time, $t−1$ as shown in Eq. (29.1):

$$y_t = f(y_{t-1})\tag{29.1}$$

The window width indicates the number of previous time steps. For example, Eq. (29.2) shows the sliding window with width 2:

$$y_t = f(y_{t-1}, y_{t-2})\tag{29.2}$$

For example, given the following data, Tables 29.1a and 29.1b show restructuring of the data for window widths of 1 and 2.

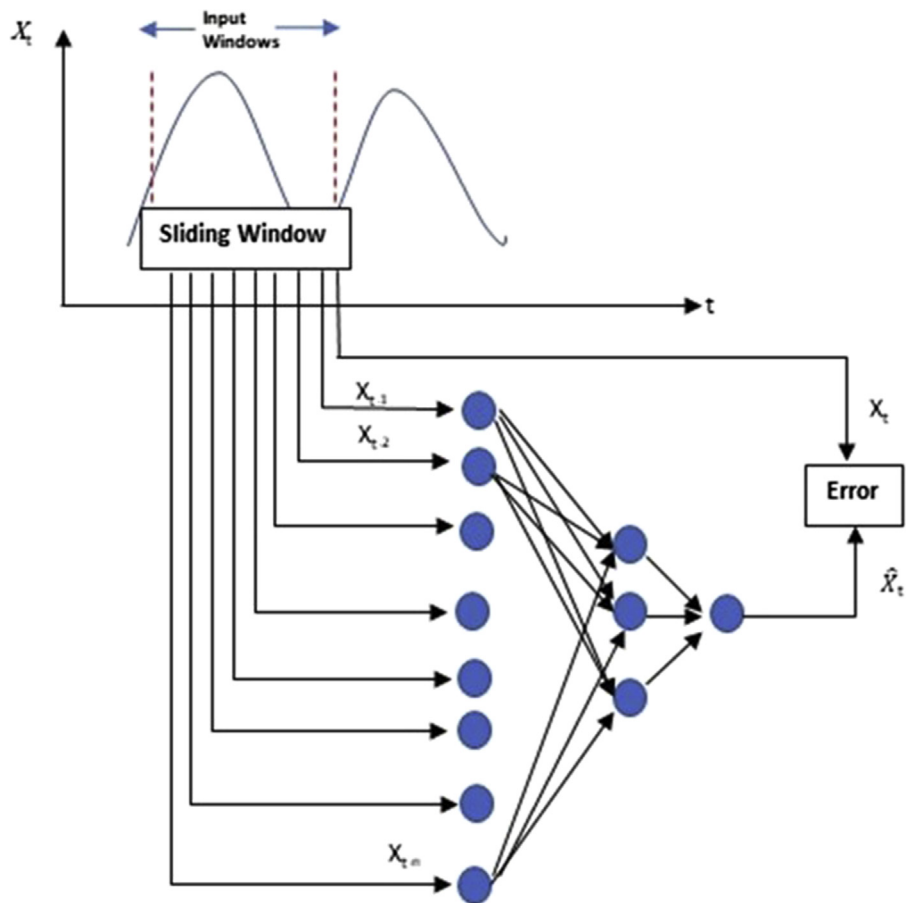**FIGURE 29.1** Sliding window technique in an artificial neural network [14].

Table 29.1a   Restructuring the dataset with window width = 1.

Time	x	Time	x = y_{t-1}	$y_t = f(y_{t-1})$ width = 1
1	2	1	?	2
2	4	2	2	4
3	6	3	4	6
4	8	4	6	8
		5	8	?

Table 29.1b   Restructuring the dataset with window width = 2.

Time	x	Time	$x1 = y_{t-2}$	$x2 = y_{t-1}$	$f(y_{t-1}, y_{t-2})$ width = 2
1	2	1	?	?	2
2	4	2	?	2	4
3	6	3	2	4	6
4	8	4	4	6	8
		5	6	8	?

Time	$x1_t$	$x2_t$	Time	$x1_{t-1}$	$x2_{t-1}$	$y1_t = x1_t$ width = 1	$y2_t = x2_t$ width = 1
1	2	0.30	1	?	?	2	0.30
2	4	0.04	2	2	0.30	4	0.04
3	6	0.23	3	4	0.04	6	0.23
4	8	0.15	4	6	0.23	8	0.15
			5	8	0.15	?	?

The previous examples are for a univariate time series dataset where one variable is varying over time. The sliding window method may also be used for multivariate time series data. Given variable *x1* and *x2* at time *t*, where $y_t = f(x1_t, x2_t)$, we may predict *x2* at time *t*, as shown in Eq. (29.3) and Table 29.2:

$$x2_t = f(x1_{t-1}, x2_{t-1}, x1_t) \tag{29.3}$$

Usually, we have one variable for prediction, but we may also predict more than one output variable, such as in an ANN. We may also predict both variables *x1* and *x2* at time *t*, as shown in Eq. (29.4).

$$x1_t \bullet x2_t = f(x1_{t-1}, x2_{t-1}) \tag{29.4}$$

For example, given two variables in multivariate time series data and the following data, Table 29.3 shows restructuring to predict both variables.

We may also predict more than one next step. This is called multistep forecasting. Table 29.4 shows an example of data restructuring with a sliding window of width = 1 and two-step multiforecasting.

Decisions regarding the window width and forecasting steps with acceptable model performance, good reasoning, and experimentation are needed for the specific dataset.

Table 29.2   Restructuring data with sliding window width 1 for multivariate time series dataset to predict variable *x2*.

Time	$x1_t$	$x2_t$	Time	$x1_{t-1}$	$x2_{t-1}$	$x1_t$	$y_t = x2_t$ width = 1
1	2	0.30	1	?	?	2	0.30
2	4	0.04	2	2	0.30	4	0.04
3	6	0.23	3	4	0.04	6	0.23
4	8	0.15	4	6	0.23	8	0.15
			5	8	0.15	?	?

Table 29.3 Restructuring data with sliding window width 1 for multivariate time series dataset to predict variable x1 and x2.

Time	$x1_t$	$x2_t$	Time	$x1_{t-1}$	$x2_{t-1}$	$x1_t$	$y_t = x2_t$ width = 1
1	2	0.30	1	?	?	2	0.30
2	4	0.04	2	2	0.30	4	0.04
3	6	0.23	3	4	0.04	6	0.23
4	8	0.15	4	6	0.23	8	0.15
			5	8	0.15	?	?

Table 29.4 Example of two-step multiforecasting.

Time	x	Time	x_{t-1}	$y1_t = x_{t-2}$ width = 1	$y2_t = x_{t-1}$ width = 1
1	2	1	?	2	4
2	4	2	2	4	6
3	6	3	4	6	8
4	8	4	6	8	?
		5	8	?	?

In this study, the width of the sliding window or lag that gives the best performance will be discussed and presented.

4. Trend analysis and forecast

The MOC in Malaysia was declared beginning Mar 18. The first phase of the MCO ending on Apr. 1 was extended for another 2 weeks at each phase to the second, third, and then fourth phases until May 12, 2020. Only essential services could continue operations and charges were imposed on those who did to not comply with the MCO. Only one member of the family was allowed to purchase household goods and had to observe social distancing, wear a face mask, and observe hand hygiene. This section will present the trend analysis of the COVID-19 pandemic in Malaysia, especially during the MCO phases.

4.1 The overall data

As of Apr. 30, 2020, there had been 6002 confirmed cases since Jan. 25, 2020; 1729 active patients were still in treatment and quarantine and there was a fatality rate of 1.7% (102 deaths). The rate of recovery was 69.5% and around 3% people who were tested were found to be positive for COVID-19. Table 29.5 shows the situation.

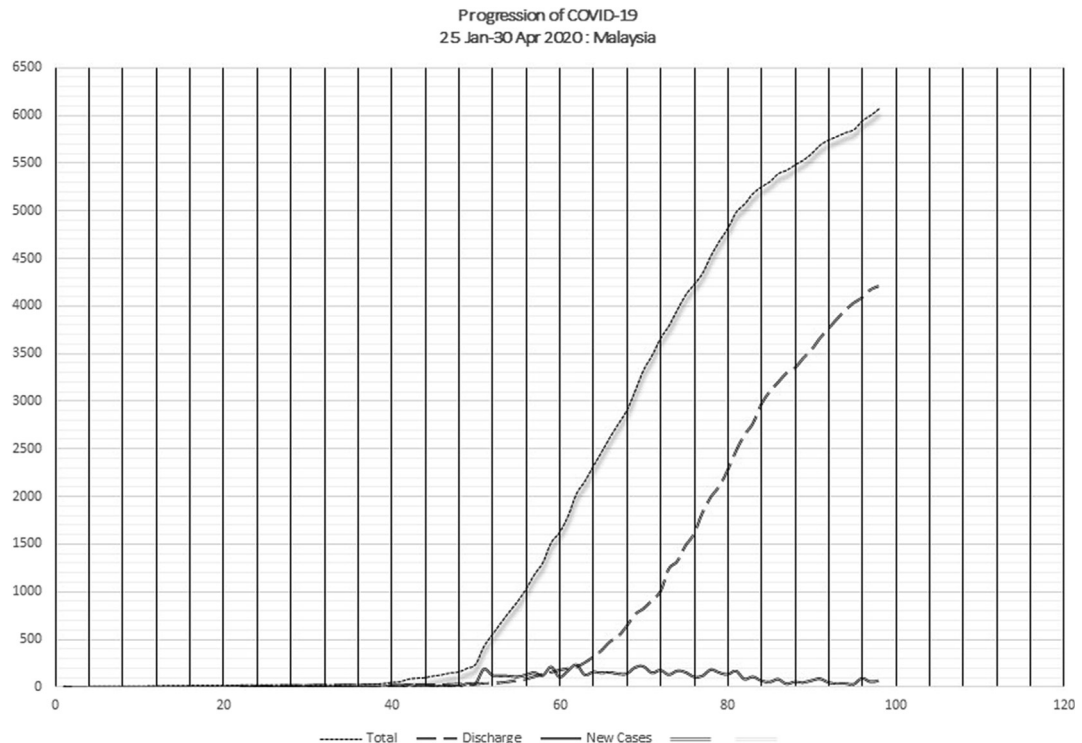**Table 29.5** Coronavirus disease 2019 situation on Apr. 30, 2020.

Confirmed cases	New cases	Active cases	Recovered cases	Deaths
6002	57	1729	4171	102

In terms of sex, 77.4% of COVID-19 patients were was male (79 males and 25 females). This is consistent with statistics around the world [6]. Active cases as stated in Table 29.5 refers to those who were still in quarantine and undergoing treatment at the hospital.

In general, the pandemic situation in Malaysia was based on new, recovered, and active cases and deaths… reported from Jan. 25 until Apr. 30, 2020.

Fig. 29.2 shows an escalating rate of recovery for Malaysian cases and a much lower mortality rate compared with 3.4% declared by the WHO (Worldometer, Mar. 5, 2020). This could be evidence that the MCO manages to achieve the objective of mitigating the risk for the disease and containing its spread to all states of Malaysia with the support of Malaysian citizens and local authorities such as the police, army, and medical frontline workers.

The disease also spread across the country of Malaysia, which has 16 states, to varying degrees, as shown in Table 29.6, which presents cases reported in various states in Malaysia. Based on the statistics, the states of Selangor and Kuala Lumpur has the most

FIGURE 29.2 Overall progression of coronavirus disease 2019 (COVID-19) in Malaysia.

Table 29.6   COVID-19 confirmed cases on April 30, 2020 of states in Malaysia.

State	Perlis	Kedah	Penang	Perak	Selangor	Kuala Lumpur	Putrajaya	Negeri Sembilan
Total cases	18	95	121	253	1455	1264	85	513
%	0.3	1.6	2.0	4.2	23.8	20.5	1.3	8.5
State	Melaka	Johor	Pahang	Terengganu	Kelantan	Sarawak	Labuan	Sabah
Total cases	199	666	297	110	155	509	16	315
%	3.3	11.0	4.9	1.8	2.6	8.4	0.3	5.2

positive cases reported. This may be because of the concentration of people in industrial areas in Selangor and Kuala Lumpur, which are the central places for many organizational headquarters as well the capital of Malaysia.

The progressive spread of the pandemic is shown in Fig. 29.3A–E and occurred in five zones: northern, central, southern, and eastern zones of West Malaysia and East Malaysia.

Fig. 29.3 shows that the Malaysian government also designated some states or district as red, yellow, and green zones, indicating the seriousness of pandemic cases reported.

4.2   Trend of cumulative coronavirus disease 2019 cases in Malaysia

Regarding cumulative cases, the spread of disease has managed to be controlled through the MCO declared by the government and enforcement through the police and army. Only one member of the family, preferably a male, may go out for necessities and must wear a face mask and maintain social distancing in supermarkets, retails shops, banks, and so on.

The projection of cumulative cases may be estimated from the trendline polynomial equation generated through curve-fitting of actual data using the fifth order, as shown in Fig. 29.5. The estimated trend line equation is given by Eq. (29.5), with $R^2 = 0.997$:

$$y = -5E - 06x^5 + 0.0004x^4 + 0.0428x^3 - 3.4011x^2 + 60.782x - 210.83 \qquad (29.5)$$

where y and x represent total cases and days, respectively. The number of total cases by May 12, 2020, which was the end of MCO Phase 4 (i.e., day 109 according to Eq. [29.5]), showed an estimated 965 persons to be infected. The MCO was able to flatten the

Time	$x1_t$	$x2_t$
1	2	0.30
2	4	0.04
3	6	0.23
4	8	0.15

Time	$x1_{t-1}$	$x2_{t-1}$	$x1_t$	$y_t = x2_t$ width = 1
1	?	?	2	0.30
2	2	0.30	4	0.04
3	4	0.04	6	0.23
4	6	0.23	8	0.15
5	8	0.15	?	?

FIGURE 29.3 (A–E) Progression of cumulative confirmed cases according to the geographical region in Malaysia. *COVID-19*, coronavirus disease 2019.
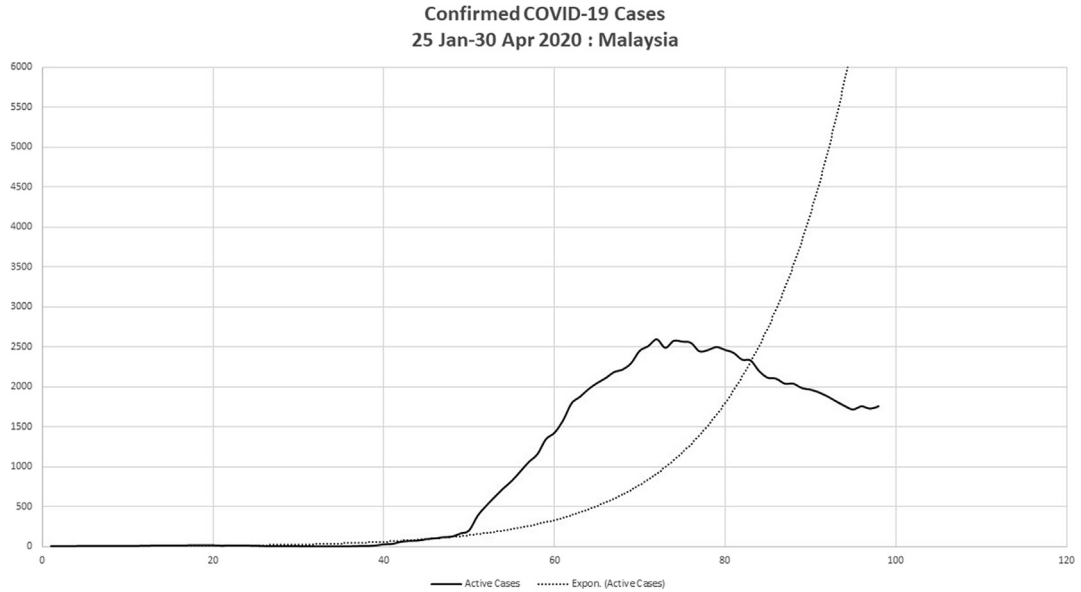
FIGURE 29.4 Cumulative coronavirus disease 2019 (COVID-19) cases from Jan. 25 to Apr. 30, 2020 in Malaysia. *MCO*, Movement Control Order.
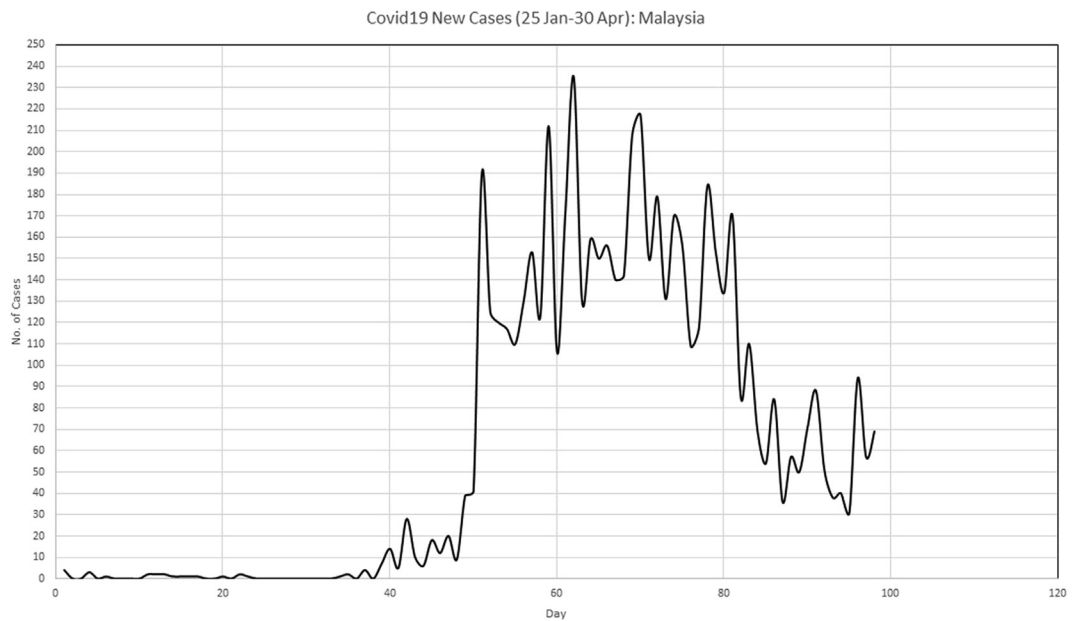
FIGURE 29.5 Pattern of new cases reported. *COVID-19*, coronavirus disease 2019; *MCO*, Movement Control Order.

curve and suppress the spread of the pandemic in an exponential pattern. The projection shows a decline in the spread of positive cases after MCO Phase 4. However, the community still had to observe social distancing and adhere to self-quarantine and heightened hygiene practices.

Fig. 29.6 illustrates the pattern of new cases reported and shows that most new cases were detected during MCO Phases 1 and 2 and slowly reducing locally infected cases. However, the figure also records cases imported from Malaysians who returned from overseas regions such as Indonesia, Iran, and Italy through a special arrangement by the government, and who quarantined for 14 days in a designated location until their tests were clearly negative.

Fig. 29.7 depicts the promising trend of a slowdown in the infection, in which the peak was around the middle of Apr. during MCO Phase 2, as expected by the WHO and the Ministry of Health Malaysia, as reported in the newspaper [15]. Active cases were those found to be positive and undergoing treatment at the hospital:

$$Active\ Cases = Total\ Cumulative\ Cases - Recovered - Death. \tag{29.6}$$

Fig. 29.8 compares new, active, and recovered cases. The trend shows a favorable situation in which fewer than 100 new cases were recorded from Apr. 17 to Apr. 30, 2020.

Regarding the mortality rate, the situation looks promising: there was a lower fatality rate in April (fewer than five cases) and an average of 1.97 people per day. The largest number of deaths recorded was on Apr. 30, seven persons, as depicted in Fig. 29.9. The medical frontlines should be commended for this achievement for their sacrifices.
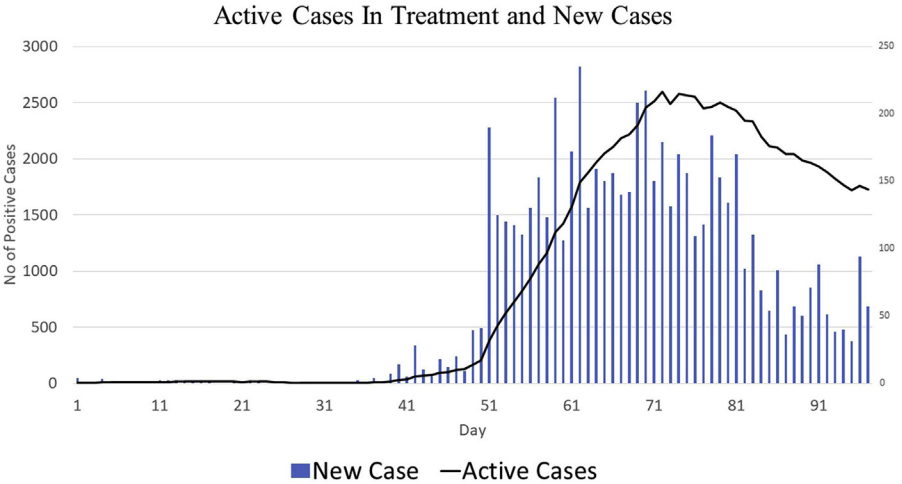

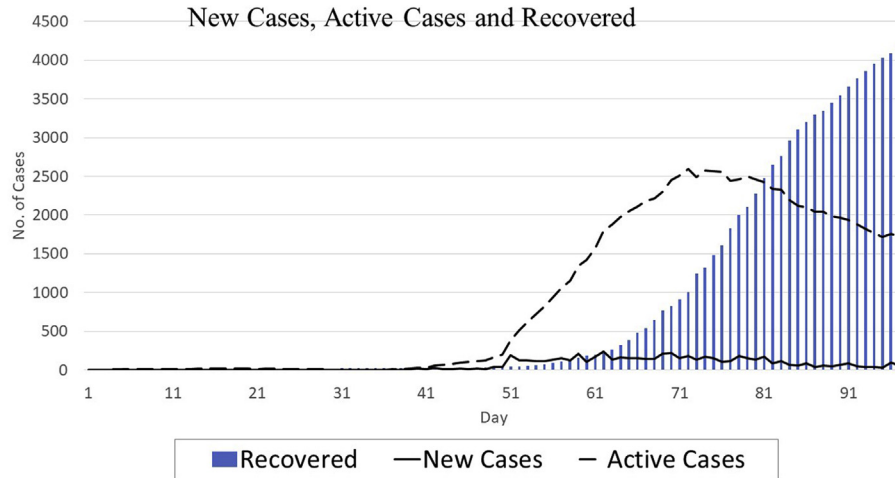FIGURE 29.6 Comparison of active and new cases.

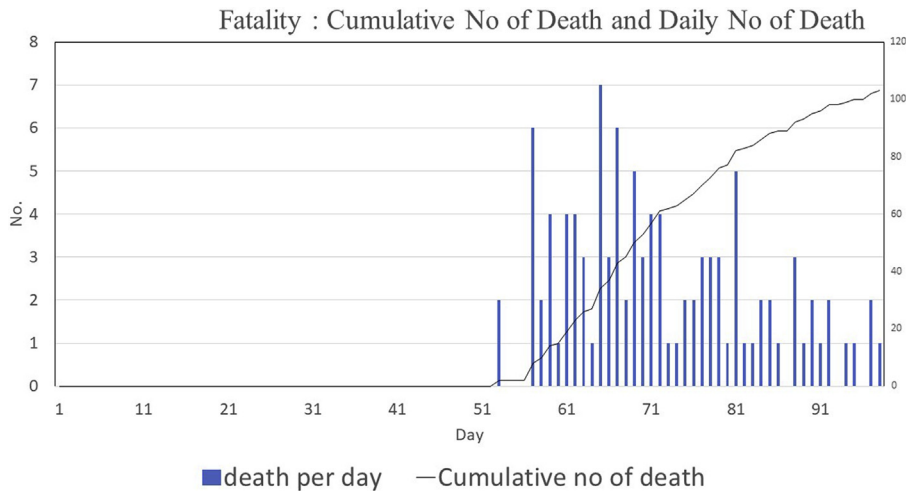**FIGURE 29.7** Comparison of recovered, cases, and active cases.

FIGURE 29.8 Comparison of cumulative deaths and daily death statistics.

4.3   Sliding window time series forecasting for coronavirus disease 2019

As described in Section 3, this study conducts predictive analytics using a sliding window method. The main justifications for selecting this approach are:

a. Time series data are usually analyzed as a single variable varying over time. These data are continuously bonded to each other, meaning that what will happen at
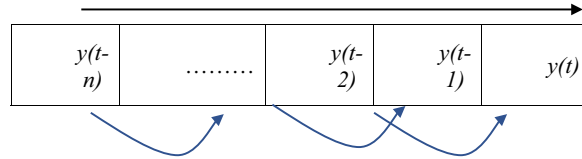
FIGURE 29.9 Sliding window conceptual view.

time t will strongly depend on or be influence by what happened in the previous
time scale such that $y(t) = f(y(t-n))$, where n is the time frame that we are inter-
ested in investigating. Fig. 29.9 illustrates the sliding window concept.

b. There is a time delay from one variable to another. For example, a person who is
infected and his road to recovery will take some duration of time. Thus, data on
new cases will eventually contribute to statistics on the number of patients who
are discharged or die. This time delay must be captured so as not to lose impor-
tant information.

c. Many time series datasets are multivariate in nature. How can we conduct time se-
ries forecasting when we must consider prior events? How do we restructure the
data? How do we run the forecasting algorithm with multivariate datasets?

Two experiments were conducted using (a) the multiple regression technique (b) the
multilayer perceptron, a feedforward ANN with a restructured dataset using a sliding
window approach.

4.4 Data preparation with sliding window representation

In this experiment, the data on which we are focused is the number of recovered COVID-
19 patients from Jan. 25 to Apr. 30, 2020. It is the cumulative number of people infected
who were discharged from the hospital.

The purposes of the experiment are:

a. To identify whether there is a specific time frame that may influence the projection
of future trends;

b. To determine the optimum sliding window width that gives good predictive ana-
lytics performance; and

c. To determine whether it is consistent with the other technique. For this study, we
will compare the result of time series forecasting using multiple regression and
multilayer perceptron with feedforward ANN architecture.

The Microsoft Excel Data Analysis tool was used for this experiment. Tables 29.7 and
29.8 show the data preparation using the sliding window technique. The data need to be
restructured so that they represents a classification problem. However, because the data
are numerals with real values, it is a regression problem.

Table 29.7   Example of data restructuring using a window of width 1, 2, and 3.

x3 = y(t−3)	x2 = y(t−2)	x1 = y(t−1)	y(t)	Actual data
			0	0
		0	0	0
	0	0	1	1
0	0	1	1	1
0	1	1	1	1
1	1	1	2	2
1	1	2	3	3
1	2	3	7	7
2	3	7	8	8
3	7	8		
7	8			
8				

Table 29.8   Restructured data in sliding window format, in which *xi* is the predictor variable for *y(t)*. Table 29.8 shows three sets of data with varying window width.

Sliding window restructured data								
Width = 1		Width = 2			Width = 3			
x1 = y(t−1)	y(t)	x2 = y(t−2)	x1 = y(t−1)	y(t)	x3 = y(t−3)	x2 = y(t−2)	x1 = y(t−1)	y(t)
0	0	0	0	1	0	0	1	1
0	1	0	1	1	0	1	1	1
1	1	1	1	1	1	1	1	2
1	1	1	1	2	1	1	2	3
1	2	1	2	3	1	2	3	7
2	3	2	3	7	2	3	7	8
3	7	3	7	8				
7	8							

Table 29.7 illustrates the restructuring process and Table 29.8 shows the transformed data to be used for predictive analytics.

4.4.1   Sliding window time series forecasting with multiple regression

Malaysia COVID-19 data were gathered from various resources such as the Ministry of Health Malaysia, Department of Statistics Malaysia, Worldometer, and John Hopkins University from Jan. 25 to Apr. 30, 2020. Data collected were the number of cumulative cases from each state in Malaysia, total cumulative cases in Malaysia, active cases, cumulative number of patients who recovered, number of patients in the intensive care unit, number of new cases, and number who died. There were 21 variables in which 16

were states and five statistics were at the national level. In this chapter, we focus only on the number of patients who recovered or were discharged from the hospital. This is an indicator of how well the pandemic was under control. More people discharged show that fewer people are infected and will become infected.

From Jan. 25 to Apr. 30, there were 97 sets of daily data for the COVID-19 situation. The width in window use in the experiment is from 2 to 16 owing to the limitation of the predictor variables of the software for multiple regression. Table 29.9 lists details about the dataset and the corresponding performance of the multiple regression according to the window width with a 95% confidence level. For the regression model with various widths of window, w is statistically significant with an F value < 0.05.

All regression models that use the sliding window perform far better than a typical $y = f(t)$, which consider each type of data as discrete. By using the sliding window, the model indicates that the historical time series or temporal data influence the current outcome owing to continuity in the time line. Upon examining the performance evaluation metrics in this experiment, the window of widths 5 and 14 gave the best performance and were statistically significant. This is also consistent where the P value at every window width is less than 0.05.

Thus, the regression model for the window of width 5 with significant predictor variables is given by Eq. (29.7), where recovery r is:

$$r(t) = 1.2028 * r(t-1) + -0.4239 * r(t-5) \tag{29.7}$$

For the window of width 14:

$$r(t) = 1.00999 * r(t-1) - 0.5743 * r(t-5) + 0.5353 * r(t-7) - 0.6185 * r(t-10) + 0.5469$$
$$* r(t-14) \tag{29.8}$$

The predicted value for recovered cases using a window with widths 5 and 15 is shown in Table 29.10.

Table 29.9 Performance evaluation of the multiple regression model with different window widths.

Window width w	Performance R^2	Observations	Window width w	Performance R^2	Observations
***0	0.654935	97	9	0.999682	88
1	0.998887	96	10	0.999685	87
2	0.999418	95	11	0.999688	86
3	0.999563	94	12	0.999708	85
4	0.99958	93	13	0.999721	84
5	0.999643	92	14	0.999757	83
6	0.999643	91	15	0.999755	82
7	0.999674	90	16	0.999754	81
8	0.999681	89			

Table 29.10   Projected value using the regression model.

Date	Day	w = 5	w = 14
Apr. 30	97	4171	4171
May 1	98	3379.953	3929.629
May 2	99	2388.145	3671.179
May 3	100	1163.322	3391.380

4.4.2   Sliding window time series forecasting with multilayer perceptron with lag

Data with the window of width 14 was chosen because of better performance in predicting future trends compared with the window of width 5. The experiment was conducted using Weka 3.8.3. Data from Excel was converted to CSV format and save into the .arff extension. In Weka, the sliding window can be implemented using a lag. Thus, a sliding window width is the same as the lag size where prior time steps will be used to predict the next time step. Four experiments were carried out (Table 29.11):

a. The original data on the number of patients being discharged without a lag
b. The original data on the number of patients being discharged with a lag with a minimum of 1 and a maximum of 14
c. The original data on the number of patients being discharged with a lag with a minimum of 5 and a maximum of 14 (based on regression model performance)
d. With restructures, discharge data with window width = 14 based on the performance in the regression model.

Based on Table 29.11, the best result with the lowest mean square error uses the original discharge data with the lag set to 5 at the minimum and 14 at the maximum based on the regression model. Table 29.12 lists details of the multilayer perceptron architecture neural network architecture.

Table 29.11   Result of time series forecasting using artificial neural network (ANN).

		Predicted			
Data	Mean square error	May 1	May 2	May 3	May 4
Original data without lag	29.2578	4314.8850	4379.3578	4437.7216	4492.7702
Original data with lag set to minimum of 1, maximum of 14	24.2193	4252.9409	4302.119	4342.7921	4377.3397
Original data with lag set to minimum of 5, maximum of 14	23.4196	4279.4511	4338.5092	4393.4741	4441.0433
Restructured data with window width 14	33.4426	4245.7669	4280.4892	4312.1067	4341.2878

The ANN architecture use the Multilayer perceptron model with lag or sliding window.

Table 29.12   Multilayer perceptron architecture neural network architecture.

Item	Detail	
1	Number of inputs	2
2	Number of lags (minimum)	5
3	Number of lags (maximum)	14
4	Number of transformed input data	23
5	Number of hidden nodes	12
6	Number of outputs	1
7	Activation function	Sigmoid
8	Number of instances	97

5. Discussion

Comparing the performance in terms of the predictive values of discharge cases from May 1–4, 2020, the multilayer perceptron architecture model with a lag of 5–14 performs much better than the regression model in projecting future trends.

The window width of 14 implies that the historical event 2 weeks before the next time scale correlates with the outcome. According to Worldometer [18], on Mar. 12, the current official estimated range for the incubation of COVID-19 was 2–14 days. An evaluation is needed to confirm whether the window width or lag size represents the incubation period of the coronavirus.

The sliding window is a convenient method for conducting time series forecasting. The data need to be restructured to be represented as a temporal classification or regression problem. In time series data, historical data are important in determining future trends. This is also in line with work conducted by Norwawi et al. [10] and Wan Hussain et al. [16], which adopted the sliding window method for time series data for early decision-making for flood emergency management of a water reservoir.

6. Conclusion

This study presented trend analysis on the COVID-19 situation in Malaysia from Jan. 25, 2020 to Apr. 30, 2020. The MCO had an impact on flattening the pandemic curve from spreading exponentially and helped improve the recovery of many infected patients. There was a 1.79% fatality rate, which was at the lower end from of the global rate.

Time series forecasting was also carried out using multiple regression and ANN employing a sliding window approach or lag. ANN is a black box approach that outperformed multiple regression in predictive analytics with these data that were gathered from various resources.

As for future work, a multivariate analysis for the time series data with more than one predictor variables or outcome can be explored. Studying the movement of the epidemic and the relationship among those variables would give more insight into how to fight this pandemic or how to be better situated for preparedness strategies in defending the community and society at large from the spread of disease, especially fatal ones.

Acknowledgments

References

[1] A. Abdullah Farid, G.I. Selim, H.A.A. Khater, A novel approach of CT images feature analysis and prediction to screen for corona virus disease (COVID-19), Preprints (2020) 2020030284, https://doi.org/10.20944/preprints202003.0284.v1.

[2] B. Achintya, D. Bhowmik, J. Mukherjee, Forecast and interpretation of daily affected people during 21 days lockdown due to COVID 19 pandemic in India, medRxiv (2020), https://doi.org/10.1101/2020.04.22.20075572, 2020.04.22.20075572.

[3] M. Ahmadi, A. Sharifi, S. Dorosti, et al., Investigation of effective climatology parameters on COVID-19 outbreak in Iran, Sci. Total Environ. (2020), https://doi.org/10.1016/j.scitotenv.2020.138705.

[4] Department of Statistics Malaysia (DOSM), COVID-19 by State in Malaysia, 2020, https://doi.org/10.1101/2020.03.26.20044511. https://ukkdosm.github.io/covid-19. (Accessed 30 April 2020).

[5] H.S. Hota, R. Handa, A.K. Shrivas, Time series data prediction using sliding window based RBF neural network, Int. J. Comput. Intell. Res. 13 (Number 5) (2017) 1145−1156. ISSN 0973-1873.

[6] M. Koerth, Why Are More Men than Women Dying of COVID-19? Science and Health, John Hopkins CSSE, April 30, 2020. https://coronavirus.jhu.edu/#covid-19-basics.

[7] M. Vafaeipour, M.O. Rahbari, Rosen, A. Fazelpour, P. Ansarirad, Application of sliding window technique for prediction of wind velocity time series, Int. J. Energy Environ. Eng. 5 (2014) 105, https://doi.org/10.1007/s40095-014-0105-5.

[8] Ministry of Health, Malaysia(MOH), COVID-19: Maklumat Terkini, 2020. https://www.moh.gov.my/index.php/pages/view/2274.

[9] S. Naomie, W.H. Chan, S. Mansor, N.E.N. Bazin, S. Amaran, A.A.M. Faudzi, A. Zainal, S.H. Huspi, E.J.H. Khoo, S.M. Shithil, COVID-19 epidemic in Malaysia: impact of lockdown on infection dynamics, medRxiv (2020), https://doi.org/10.1101/2020.04.08.20057463, 2020.04.08.20057463.

[10] N.Md Norwawi, K.R.K. Mahamud, S. Deris, Recognition decision-making model using temporal data mining technique, J. ICT 4 (1) (2005) 37−56. ISSN 1675-414X.

[11] N.Md Norwawi, Computational Recognition-Primed Decision Model Based on Temporal Data Mining Approach in a Multiagent Environment for Reservoir Flood Control Decision (Ph.D. thesis), Universiti Utara, Malaysia, 2004.

[12] C. Paoli, C. Voyant, M. Muselli, M.-L. Nivet, Forecasting of preprocessed daily solar radiation time series using neural networks, Sol. Energy 84 (12) (2010) 2146−2160. https://doi.org/10.1016/j.solener.2010.08.011.

[13] R. Gupta, S.K. Pal, Trend analysis and forecasting of COVID-19 outbreak in India, medRxiv (2020), 2020.03.26.20044511.

[14] A. Rao, J. Vazquez, Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey in the populations when cities/towns are under quarantine, Infect. Contr. Hosp. Epidemiol. (2020) 1−18. https://doi.org/10.1017/ice.2020.61.

[15] The Star, WHO Expects Malaysia's Covid-19 Cases to Peak in Mid-April, April 2 , 2020. https://www.thestar.com.my/news/nation/2020/04/02/who-expects-malaysia039s-COVID-19-cases-to-peak-in-mid-april.

[16] W.I. Wan Hussain, K.M. Ku Ruhana, N.Md Norwawi, Mining temporal reservoir data using sliding window technique, CiiT Int. J. Data Mining Knowledge Eng. 3 (8) (June 2011) 473−478.

[17] World Health Organization, Coronavirus Disease 2019 (COVID-19): Situation Report, 2020, p. 22. (Accessed 30 April 2020).

[18] Worldometer. Covid-19 Coronavirus Pandemic. Available from: https://www.Worldometers.Info/Coronavirus/#Countries. (Accessed 30 April 2020).

[19] R. Huang, M. Liu, Y. Ding, Spatial-temporal distribution of COVID-19 in China and its prediction: a data-driven modelling analysis, J. Infect. Dev. Ctries 14 (3) (2020) 246−253. https://doi.org/10.3855/jidc.12585.