



OPEN

Determining an effective short term COVID-19 prediction model in ASEAN countries

Omar Sharif¹, Md Zobaer Hasan²✉ & Azizur Rahman³

The challenge of accurately short-term forecasting demand is due to model selection and the nature of data trends. In this study, the prediction model was determined based on data patterns (trend data without seasonality) and the accuracy of prediction measurement. The cumulative number of COVID-19 affected people in some ASEAN countries had been collected from the Worldometers database. Three models [Holt's method, Wright's modified Holt's method, and unreplicated linear functional relationship model (ULFR)] had been utilized to identify an efficient model for short-time prediction. Moreover, different smoothing parameters had been tested to find the best combination of the smoothing parameter. Nevertheless, using the day-to-day reported cumulative case data and 3-days and 7-days in advance forecasts of cumulative data. As there was no missing data, Holt's method and Wright's modified Holt's method showed the same result. The text-only result corresponds to the consequences of the models discussed here, where the smoothing parameters (SP) were roughly estimated as a function of forecasting the number of affected people due to COVID-19. Additionally, the different combinations of SP showed diverse, accurate prediction results depending on data volume. Only 1-day forecasting illustrated the most efficient prediction days (1 day, 3 days, 7 days), which was validated by the Nash–Sutcliffe efficiency (NSE) model. The study also validated that ULFR was an efficient forecasting model for the efficient model identifying. Moreover, as a substitute for the traditional R-squared, the study applied NSE and R-squared (ULFR) for model selection. Finally, the result depicted that the prediction ability of ULFR was superior to Holt's when it is compared to the actual data.

The widespread coronavirus disease (COVID-19) began in China (Hubei Province) in December 2019. Multiple phases of the pattern of the COVID-19 disease continue to be the source of infections in numerous countries, frightening to turn out to be a long-term epidemic. The main reason for spreading this virus is that it was not detected early. Researchers have already begun investigating how the pandemic affects global macroeconomic trends¹. However, essential to predict how much the virus is spreading rather than measuring global financial loss.

A former study has shown that biological growth models, especially the sub-epidemic growth model, can reveal the empirical shapes of past epidemics that support producing short-term forecasts of an epidemic model in real-time. These epidemic models are beneficial when the collected data are limited for bringing any ideal information^{2,3}. In addition, the short period of real-time forecasts created from such models could support preparing to allocate the necessary resources, which are crucial to bringing the epidemic under control.

It is known that each communicable disease epidemic shows specific patterns. Seasonal changes and an adaptation of the virus over time are critical reasons for outbreaks showing different patterns in a country or region. Showing different patterns is typically happening at a different scale level concerning time. Generally, pandemic patterns occur in non-linear shapes. Therefore, this characteristic inspires us to create a model to identify such non-linear lively fluctuations⁴. Moreover, the transmission of such transmittable diseases could be defined by analysing the pattern of these non-linear systems.

The main objective of this research is to determine an effective short-term prediction model by utilizing the cumulative number of COVID-19 infected people in four ASEAN (Association of Southeast Asian Nations) countries, i.e., Thailand, Philippines, Singapore, and Indonesia. To the best of our knowledge, no literature has analysed ASEAN countries' to predict short term COVID-19 cases, mainly predicting 1 day. Typically, short-term reliable prediction can assist decision-makers in managing the pandemic efficiently in real time and mitigate

¹Universal College Bangladesh (Monash College), Dhaka, Bangladesh. ²School of Science, Monash University Malaysia, Bandar Sunway, Selangor D. E., Malaysia. ³School of Computing, Mathematics and Engineering, Charles Sturt University, Wagga Wagga, Australia. ✉email: mdzobaer.hasan@monash.edu

any upcoming severe risks. Besides, the study focuses on the four countries for the following motives. In the first place, ASEAN countries, especially these four countries, contribute to a vast global economy. Secondly, ASEAN countries' COVID-19 data is available in Worldometers. Therefore, it is crucial to understand how to predict the COVID-19 spreading rate at some ASEAN countries and take effective measures to control the pandemic. Hence, this paper principally uses three dynamic models to generate 1-day ahead or multiple day's ahead forecasts of the cumulative reported cases of COVID-19 by addressing any seasonal variations. Firstly, the prediction model, Holt's, and prediction parameters are selected. Secondly, it employs Wright modified Holt's method for forecasting irregularity in time spacing. Thirdly, an unreplicated linear functional relationship model (ULFR) is applied to find a more justified forecasting value of dependent and independent variables. This ULFR model is a newly used technique in such a type of prediction analysis. Last but not least, the prediction Holt's model was selected based on data patterns (trend data without seasonality) and the accuracy of prediction measurement with different statistical methods. Particularly, instead of using only the traditional R-squared, this study also used the Nash–Sutcliffe efficiency (NSE), model efficiency factor (MEF) and the R-squared value of ULFR for model selection.

Literature review

Since the outbreak in Wuhan, several worldwide scholars and researchers have reported estimations and predictions for the COVID-19 epidemic in journal publications or on websites^{5–7}. Among the studies, many modelling results have shown a wide range of variations depending on data availability⁸: estimated basic reproduction number varies from 2 to 6, peak time estimated from mid-February to late March, and the total number of infected people ranges from 50,000 to millions. An important question is now, even though predictions are made using transmission models based on either the Susceptible-Infectious-Removed (SIR) or Susceptible-Exposed-Infectious-Removed (SEIR) framework, why is there such a wide variation in model predictions?⁹. The literature review shows that the primary reason for such variation is selecting a small number of data or selecting parameters without justification¹⁰.

Recently applied various models for measuring spreading and forecasting COVID-19 such as Ace-Mod Australian Census-based Epidemic Model (Ace-Mod), SIDR, Fuzzy Clustering, SEIR and DASS-21^{11–13}. Particularly, Roosa et al.¹³ generate forecasts in China using the Richards growth model, a sub-epidemic wave model, and a generalized logistic growth model. Following models have been previously used to forecast outbreaks due to different infectious diseases. Forecasts finding from each model suggest the outbreaks may be nearing extinction in Guangdong and Zhejiang.

Fokas et al.¹⁴ applied a computational approach to computing the impact on the number of deaths globally and Paul et al.¹⁵ predict the disease burden with special emphasis on south Asian countries (India, Bangladesh, and Pakistan) by the SEIR method. Hasan and Siddik¹⁶ Examine the correlation between daily and total COVID-19 cases in Bangladesh by linear relationship model. Petropoulos and Makridakis¹⁷ applied exponential smoothing models to predict the continuation of the COVID-19 by analyzing confirmed cases, deaths, and recoveries of several people. Moreover, Md Hasinur Rahaman Khan and Ahmed Hossain¹⁸ used the Infection Trajectory-Pathway Strategy (ITPS) model to analyze the COVID-19 outbreak situations and predict infections and deaths case from temporal data of confirmed and death cases in Bangladesh. Besides, Fanelli and Piazza¹¹ applied the SIDR model to analyze and forecast COVID-19 spreading in China, Italy, France with a concentration in the variables (Susceptible, Infected, Recovered, Dead, Scheme). Besides, Zhou et al.¹⁹ focused on deep neural network (DNN) and convolutional neural network (CNN) model for behavioural analysis based on heterogeneous health data generated in social media. Furthermore, Mahmoudi et al.²⁰ studied the relationship between the spread of Covid-19 and population size by applying the Fuzzy clustering model in USA, Spain, Italy, Germany, UK, France, and Iran. Even though a few researchers used regression models and Genetic programming for short-term prediction of COVID-19 cases in the different regions due to the small number of data or parameter selection problems, many of those models' outcomes have shown a wide range of dissimilarities^{10,17}. Notably, some studies^{13–17,21} have used various mathematical models to determine the spread of the disease, predict the number of incidences, deal with healthcare missing data, health care facilities in tackling COVID-19 spread. In addition, some researcher uses the SEIR model, Genetic Programming and Regression model for short-term prediction^{22–24}. Moreover, a dynamic model, a generalized logistic growth model, the Richards growth model, and a sub-epidemic wave model were used to generate 5-day and 10-day ahead forecasts of the cumulative reported cases in the provinces of Guangdong and Zhejiang, China¹³. Consequently, selecting the appropriate model and parameter value is key for the forecasting model. Besides, many models are available in the literature to model infectious diseases. A few models have been used primarily for the countries where the number of cases is very high, like China, Italy, Spain, the UK, Germany and the USA. Still, no literature has analysed ASEAN countries' to predict short term COVID-19 cases mainly predicting for 1 day.

Methodology

In forecasting, the popular methods are the average approach, Naïve approach, Drift method, Seasonal naïve approach, Time series methods, Econometric forecasting methods, artificial neural networks. In economics, the widely applied methods fall in Time series methods and Econometric forecasting methods. However, in health science, rarely forecasting models are applied. Unfortunately, COVID-19 teaches us the importance of prediction results to control the spread of the disease. For the model selection, the study first applied the rule of thumb. Then after analyzing data, three models (Holt's method, Wright's modified Holt's method, and unreplicated linear functional relationship model (ULFR)) have been utilized to identify an efficient model for short-time prediction. Afterwards, this study has tested different smoothing parameter by MAPE, MAD, MSE and RMSE to select effective smoothing parameters. Finally, the model is validated by the NSE, MEF, traditional R-squared,

and R-squared (ULFR). Therefore, the present study relies on the three mentioned models that are first applied in health science to forecast the number of COVID-19 incidences in four ASEAN countries based on the actual historical data of August 20, 2020, to September 16, 2020.

Holt's method. To build Holt's method, first, the exponential smoothing technique was projected in the late 1950s²⁵. In addition, the exponential smoothing technique has motivated few of the most practical forecasting approaches. The exponential smoothing method to produce forecasts is defined as weighted averages of former observations. In other words, Holt's is defined as a linear-exponential smoothing. This smoothing model is well known for forecasting data with trends. This model consists of three individual equations that are applied together to create a final forecast result. Among the three equations, the first equation is a smoothing equation that replaces the last period's trend value with the last smoothed value. Besides, the second equation is called the trend equation. The second equation consists of the changes between the last two smoothed values. The final equation consists of level and trend values to find the forecast value. Two parameters are used in Holt's method are called smoothing parameters. One parameter is used for overall smoothing, and the other one is used for trend smoothing. Therefore, another name of holt's method is the double exponential smoothing or trend exponential smoothing model²⁶. It can be expressed by the following three equations:

$$\text{Level equation } Y_t = \alpha X_t + (1 - \alpha)(Y_{t-1} + Z_{t-1}) \quad (1)$$

$$\text{Trend equation } Z_t = \beta(Y_t - Y_{t-1}) + (1 - \beta)Z_{t-1} \quad (2)$$

$$\text{Forecast equation } F_{t,k} = Y_t + KZ_t \quad (3)$$

Herein the smoothing constant and variables are defined below: Y_t Estimate of the level of the series at time t , Z_t Estimate of the trend (slope) of the series at time t , α ($0 \leq \alpha \leq 1$) Smoothing parameter for the level, β ($0 \leq \beta \leq 1$) Smoothing parameter for the trend, X_t Estimate of the period t base level from the current period, $Y_{t-1} + Z_{t-1}$ Estimate of the period t base level based on previous data.

To calculate the optimal forecasting of the Eq. (3), the following optimization technique to minimizing the squared error over all data points: $\text{Min} \sum (y_t - Y_t - KZ_t)^2$.

To calculate Z_t the following two quantities are taken as a weighted average:

1. Estimation of a trend from the current period from the upsurge in the smoothed trend between the periods (t-1) and t.
2. Z_{t-1} , which is the previous estimate of the trend.

To start Holt's method, Y_0 is an initial estimate of the level and another an initial estimate is called Z_0 which is used of the trend. Here, Z_0 equals to previous year's average increase in the time series and Y_0 equals to last observation.

Wright's modified Holt's method. Wright²⁷ introduced a modification of Holt's method for the data with irregularity in time spacing. However, modified Holt's method has some unique characteristics, i.e. it has greater computational efficiency and flexibility in terms of having more smoothing constants. Besides, the modified model has a better performance record with empirical data with missing data or zero data.

The notation and forecasting equation of Wright's modified Holt's method is illustrated below:

$$\text{One period ahead forecast of the next infected number: } f_{n+1} = l_n + m_n(t_n - t_{n-1}) \quad (4)$$

$$\text{Intercept of the trend line at disease infected number } n : l_n = (1 - v_n)[l_{n-1} + m_{n-1}(t_n - t_{n-1})] + v_n x_n \quad (5)$$

$$\text{Slope of the trend line at disease infected number } n : m_n = (1 - u_n)m_{n-1} + \frac{u_n(l_n - l_{n-1})}{(t_n - t_{n-1})} \quad (6)$$

For the Eqs. (5), (6) the variables are defined with following equations

$$u_n = \frac{u_{n-1}}{(d_n + u_{n-1})}, v_n = \frac{v_{n-1}}{(b_n + v_{n-1})}, d_n = (1 - \beta)^{(t_n - t_{n-1})}$$

$$b_n = (1 - \alpha)^{(t_n - t_{n-1})}, u_0 = 1 - (1 - \beta)^q, v_0 = 1 - (1 - \alpha)^q$$

The initializing levelling and trending values are denoted by $l_0 = a$, $m_0 = b$. Moreover, Infected size and period at disease infected number n are denoted by x_n and t_n respectively. Additionally, smoothing constant of intercept and slop are presented by $\tilde{\alpha}$ and β respectively. To initialize, the following constant a and b sequentially represents intercept of trend line and slope of trend line.

Unreplicated linear functional relationship model (ULFR). The linear regression model is a popular model for analysing the dependent and independent variables' relationship. Nevertheless, the relationship

between the variables becomes a fuzzy relationship due to unusual fluctuations with defined variables. Moreover, as²⁸ mentioned in an article, it is unlikely to measure precisely independent variables in all circumstances.

The relationship model for the two variables is introduced following way by considering the above issues. Assume that X and Y are two linearly related unobservable variables. In the Eq. (7) X and Y are defined as a linearly independent variable and target variable respectively.

$$Y_i = \beta_a + \beta_f X_i \quad (7)$$

The parameters value β_a, β_f of the Eq. (7) can be found by least square loss function (minimizing the squared error) over all data points: $\text{Min} \sum (Y_i - \beta_a - \beta_f X_i)^2$.

The linear functional relationship model assigns dependent and independent variables by assuming that both the variables are subject to errors. Let, two variables X_i and Y_i which correspond to random variables x_i and y_i that are observed with errors, d_i and e_i respectively, such that,

$$\left. \begin{aligned} y_i &= Y_i + e_i \\ x_i &= X_i + d_i \end{aligned} \right\} i = 1, 2, \dots, n \quad (8)$$

The following conditions are assumed:

Both of the errors have zero mean: $E(d_i) = E(e_i) = 0$.

The observed errors have constant but different variance $\text{Var}(d_i) = \sigma_d^2, \text{Var}(e_i) = \sigma_e^2, \forall i$.

The errors are uncorrelated, i.e., $\text{Cov}(d_i, d_j) = \text{Cov}(e_i, e_j) = 0, i \neq j$

The errors are uncorrelated, i.e., $\text{Cov}(d_i, e_j) = 0, \forall i, j$

Chang et al.²⁹ termed the defined Eqs. (7) and (8) as ULFR model. In this model, it is assumed that errors d_i and e_i are mutually independent and normally distributed random variables.

When the ratio of the error variance is known, that is $\frac{\sigma_e^2}{\sigma_d^2} = \lambda$ to maximize likelihood estimators of parameters $\beta_a, \beta_f, \sigma_d^2$ and X_i which are derived by differentiating equation *likelihood function* with respect to $\beta_a, \beta_f, \sigma_d^2, X_i$ respectively and equate the result to zero. Subsequently, the equations are simplified with the maximum likelihood estimators to find the parameters. The parameters and variables i.e. $\beta_f, \beta_a, \sigma_d^2$ and X_i are defined as follows (Chang et al.²⁹):

$$\hat{\beta}_a = \bar{y} - \hat{\beta}_f \bar{x} \quad (9)$$

$$\hat{\beta}_f = \frac{(D_{yy} - \lambda D_{xx}) + \{(D_{yy} - \lambda D_{xx})^2 + 4\lambda D_{xy}^2\}^{\frac{1}{2}}}{2D_{xy}} \quad (10)$$

$$\hat{\sigma}_d^2 = \frac{1}{n-2} \left[\sum (x_i - \hat{X}_i)^2 + \frac{1}{\lambda} \sum (y_i - \hat{\beta}_a - \hat{\beta}_f \hat{X}_i)^2 \right] \quad (11)$$

$$\hat{X}_i = \frac{\lambda x_i + \hat{\beta}_f (Y_i - \hat{\beta}_a)}{\lambda + \hat{\beta}_f} \quad (12)$$

where the mean of x and y denoted and defined by $\bar{y} = \frac{\sum y_i}{n}, \bar{x} = \frac{\sum x_i}{n}$.

And the variable D_{xy}, D_{yy}, D_{xx} are defined respectively,

$$D_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}), D_{yy} = \sum (y_i - \bar{y})^2, D_{xx} = \sum (x_i - \bar{x})^2$$

Additionally, Coefficient of determination of ULFR (R_f^2) for $\lambda = 1$.

Proportion of variance: $R_f^2 = \frac{D_r}{D_{yy}}$ and regression sum of square: $D_r = \frac{\hat{\beta}_f (D_{yy} - D_{xx}) + 2\hat{\beta}_f D_{xy}}{1 + \hat{\beta}_f^2}$.

Model selection. In the model selection process, coefficient of determination, NSE and MEF are applied even though the NSE is nearly identical to the coefficient of determination. The main difference is how it is applied³⁰.

Coefficient of determination. The coefficient of determination R^2 is express as the squared value of the coefficient of correlation according to the common method of its definition. Conventionally, when a set of data sets ($o_i, p_i, i = 1, 2, \dots, n$) are obtained, a mathematical model $\hat{p} = f(o_1, o_2 \dots o_m)$ could be formed to predict p based on the observed values of $o_j, j = 1, 2, \dots, m$, where p is defined as outcome variable and values of o_j are termed as observed variables.

The functional form of R-squared is as follows:

Country	N	Range	Minimum	Maximum	Mean	SD
Thailand	28	101	3389	3490	3429.75	30.349
Philippines	28	95,335	177,579	272,914	226,230.07	27,653.601
Singapore	28	1415	56,099	57,514	56,869.36	413.130
Indonesia	28	81,782	147,211	228,993	184,352.93	25,341.732

Table 1. Summary statistics of COVID-19 infected individuals.

$$R^2 = \left(\frac{\sum_{i=1}^n (o_i - \bar{o}_i)(p_i - \bar{p}_i)}{\sqrt{\sum_{i=1}^n (o_i - \bar{o}_i)^2} \sqrt{\sum_{i=1}^n (p_i - \bar{p}_i)^2}} \right)^2$$

Nash–Sutcliffe efficiency. To avoid the shortfall of R-squared, the study also used the NSE factor presented by³¹ to determine the efficiency of the models:

$$NSE = 1 - \frac{\sum_{i=1}^n (o_i - p_i)^2}{\sum_{i=1}^n (o_i - \bar{p}_i)^2}$$

This formula can be applied for linear regression and original data on any model. Besides, NSE values can be negative for the non-linear models. The value of NSE could be between $-\infty$ and 1. Naturally, analysts seek an NSE value close to 1 for the best performance of a model. The negative result of NSE specifies an improper model efficiency. The NSE has an excellent reputation for analysing the efficiency of any model, especially the frequently applied model in hydrology^{32,33}. Besides, this study is also introduced the model efficiency factor (MEF) to give a comprehensive explanation of NSE from a different perspective, i.e.

$$MEF = \sqrt{1 - NSE}$$

Any smaller value of MEF shows the validation of the model. Thus, the values of MEF ranges between 0 and 1, where the smallest value of the MEF shows excellent models' performance. However, the value of MFE zero indicates an error-free model, which is not convincing result³⁴.

Dataset. The data in this research were collected from Worldometers info (<https://www.worldometers.info/coronavirus/>), provided with confirmed cases in ASEAN countries from 20th August to 16th September 2020. There are several reasons behind collecting data from Worldometer. First, there is no reliable free data source for ASEAN countries except Worldometer. Secondly, Worldometer is accredited as the oldest and most reliable data source by the American Library Association (ALA), Johns Hopkins CSSE, Financial Times, The New York Times and Government of UK. Finally, Worldometer is cited as a source of data in thousands of renowned journal articles. The data set in this study included all the ASEAN countries.

Since, trend data are the pre-requisite for better predicting using Holt's method²⁶, therefore some ASEAN countries, Laos, Vietnam, Brunei, Cambodia, Myanmar were excluded from the study were excluded from this study. Besides, we were unable to get the consistent trend pattern of COVID-19 infected cases for these countries. Even though Wright's modified Holt's method can deal with zero values²⁷, the number of zero values was almost all for these countries.. The summary statistics of the data set is given in Table 1. In the table, it is seen that total number data is 28 for all the country. However, standard deviation and maximum and minimum value has huge difference because each country has different number of population.

Result and discussion

Selection of smoothing constant for the prediction model. For obtaining the significant result from Holt's model, it is imperative to determine the best combination of smoothing constant. Therefore, various models such as Ace- Mod (Australian Census-based Epidemic Model)³⁵, neural network-based models³⁶ and others have been employed to access the situation. However, though these models are significant, they are not appropriate for trend and seasonal data²⁵. Table 2 shows the best combinations of smoothing constants for different days in four ASEAN countries. These findings are validated by MAPE (mean absolute percentage error) value as it is mentioned that the MAPE is one of the most popular measures of forecast accuracy³⁷.

Model validation. Several studies on validating the forecasted model of COVID-19 transmission are based on R-squared value only. Nevertheless, our research examined the accuracy and validated our prediction models using R-squared, R-squared (ULFR), NSE, and MEF values. From Table 3, it is noticed that the ULFR model predict the total cases (1-day ahead) with better accuracy compared with Holt's. In addition, the ULFR with a

Day	Thailand (α, β)	Philippine (α, β)	Singapore (α, β)	Indonesia (α, β)
Day 1	(0.2,0.1)	(0.9,0.8)	(0.9,0.7)	(0.3, 0.9)
Day 3	(0.9,0.7)	(0.7,0.7)	(0.7,0.5)	(0.8,0.2)
Day 7	(0.9,0.7)	(0.4,0.9)	(0.6,0.6)	(0.4,0.5)

Table 2. Best combinations of smoothing constants for short term prediction. *Source:* Authors' calculations.

	Thailand			Philippines			Singapore			Indonesia		
	Day 1	Day 3	Day 7	Day 1	Day 3	Day 7	Day 1	Day 3	Day 7	Day 1	Day 3	Day 7
R-squared (Holt's)	0.9940	0.9930	0.6760	0.9990	0.9900	0.9350	0.9920	0.9540	0.6920	1.0000	0.9990	0.9970
R-squared (ULFR)	0.9970	0.9677	0.8331	0.9997	0.9950	0.9595	0.9986	0.9770	0.7621	0.9999	0.9993	0.9983
NSE (Holt's)	0.9875	0.9257	0.5635	0.9997	0.9999	0.9994	1.0000	1.0000	1.0000	0.9999	0.9999	0.9996
NSE (ULFR)	0.9934	0.9959	0.9407	0.9997	1.0000	0.9997	1.0000	1.0000	1.0000	0.9999	1.0000	0.9997
MFE (Holt's)	0.0814	0.2725	0.6606	0.0162	0.0102	0.0251	0.0010	0.0014	0.0032	0.0112	0.0073	0.0188
MFE (ULFR)	0.0814	0.0637	0.2435	0.0159	0.0039	0.0180	0.0010	0.0010	0.0022	0.0111	0.0057	0.0178

Table 3. Results of Statistical analysis for selecting the most efficient model between ULFR and Holt's. *Source:* Authors' calculations. Bold marks indicate significant results at 0.05 level.

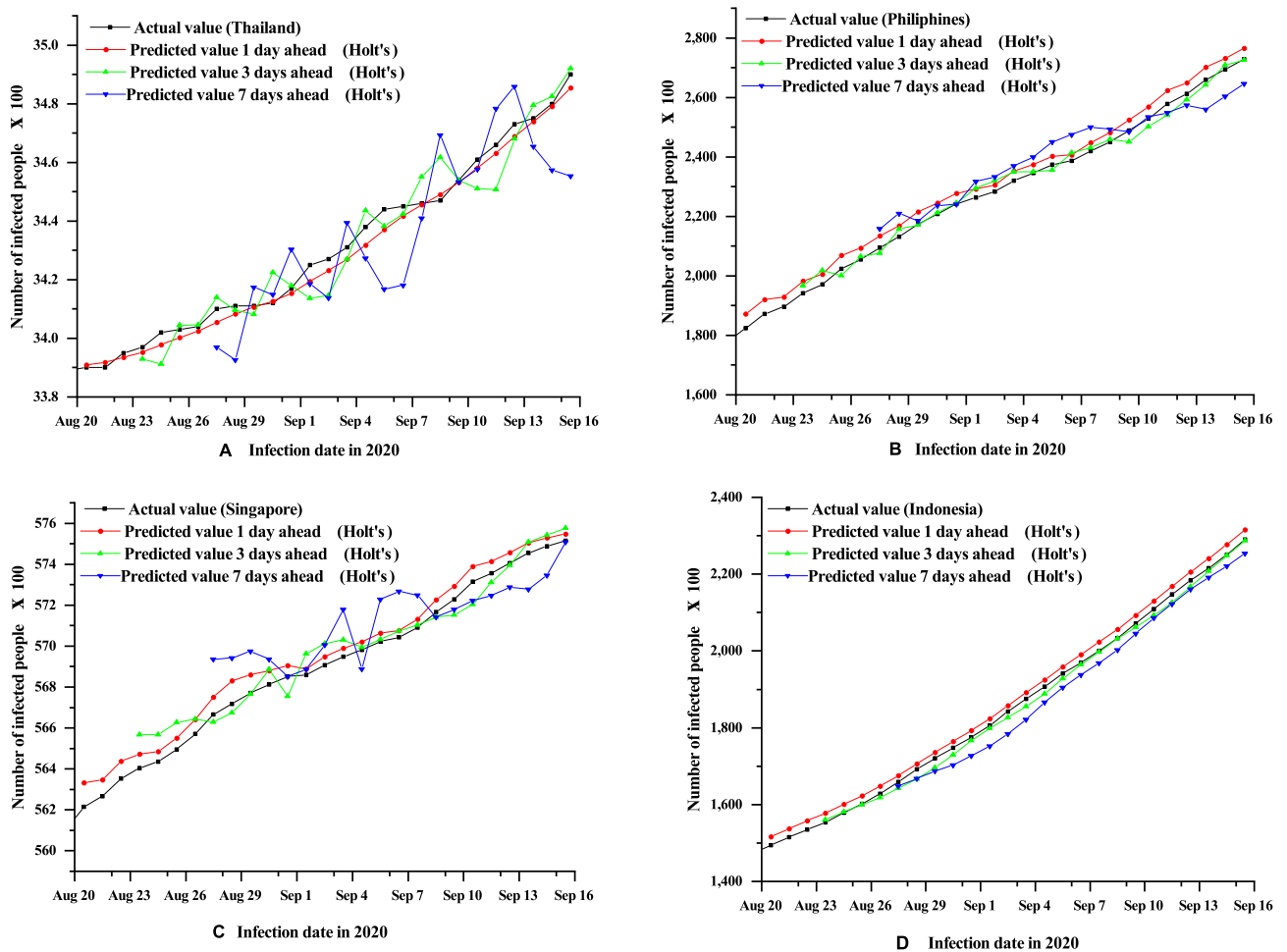


Figure 1. Holt's method short term (1, 3, 7 days) forecasting of COVID-19 cumulative cases for the four countries Thailand (A), Philippines (B), Singapore (C) and Indonesia (D).

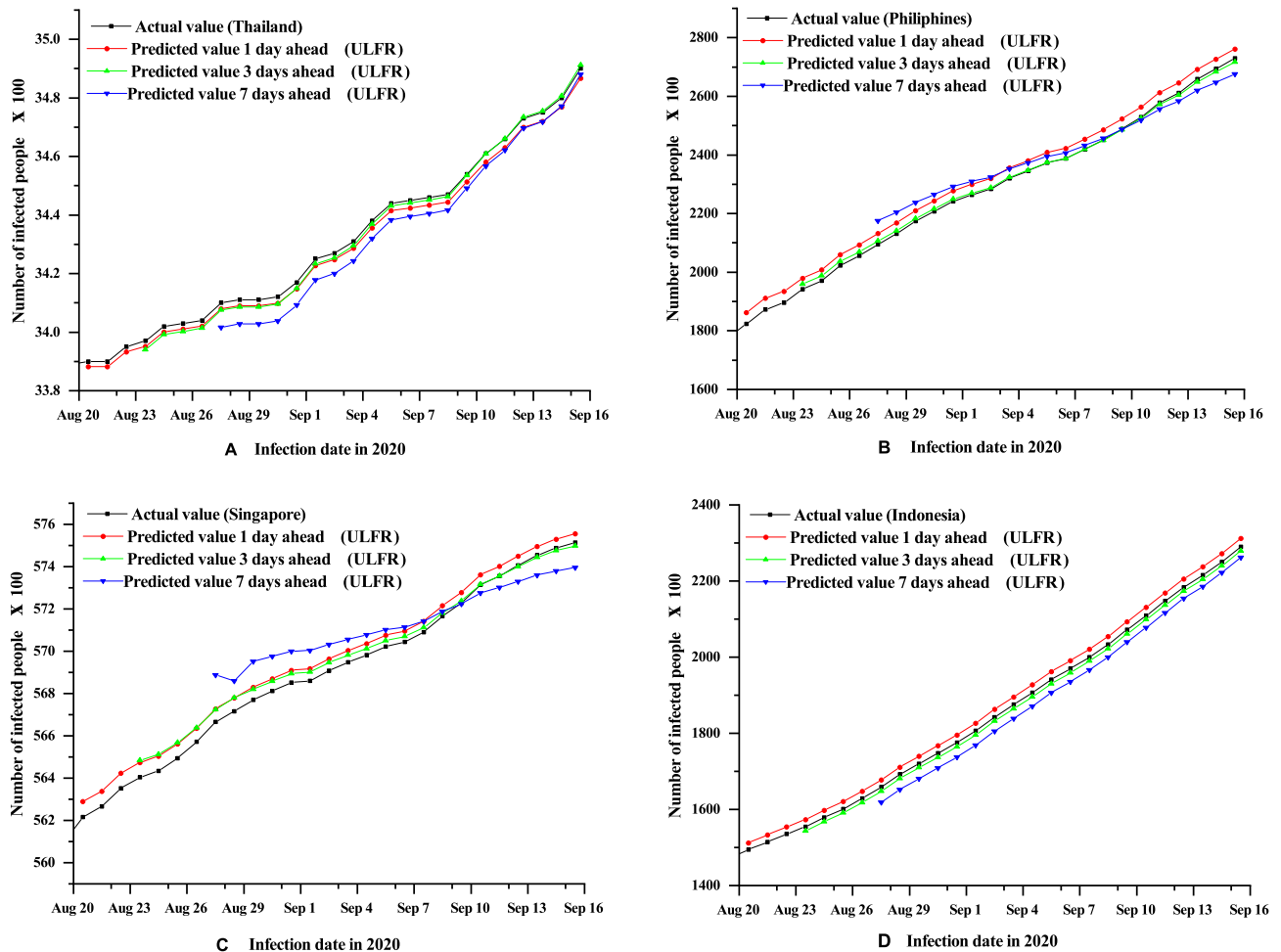


Figure 2. ULFR method short term (1, 3, 7 days) forecasting of COVID-19 cumulative cases for the four countries Thailand (A), Philippines (B), Singapore (C) and Indonesia (D).

1-day ahead prediction had the highest accuracy among the three different days' predictions. Furthermore, it is noticed that Indonesia and Singapore have the highest R-squared and NSE value, which approaches unity. Besides, Indonesia and Singapore show the lowest MEF, which approaches zero. Therefore, to find the forecasting accuracy of the proposed model, this study has been verified by comparing two data sets, i.e., the forecasted data and the reported data. These data sets are applied in the following statistical model to validate the model and maximize the forecasting accuracy. All the statistical evidence shows that ULFR is the best prediction model than Holt's model. Moreover, in the result and discussion section the modified Holt's method (Wright's model) is not included because Holt's and Wright's model showing the same result and it is accepted because there was no missing observations in our studied data³⁸.

Forecasting by Holt's Method and ULFR. In this section, Holt's and ULFR methods have been applied to predict the number of confirmed cases in ASEAN countries, particularly in Thailand, Philippines, Singapore and Indonesia. The goal is to find accurate predictions, which is probably impossible, but to illustrate some general qualitative behaviours which may be observed³⁹. Therefore, Figs. 1 and 2 show the cumulative cases of the forecast for 1, 3 and 7 days ahead by Holt's and ULFR methods respectively. From Fig. 1, it is seen that Holt's method shows better results for the Philippines and Indonesia for 1, 3 and 7 days. However, the forecasting result is not up to the mark for Thailand and Singapore. On the other hand, ULFR in Fig. 2 shows the descent prediction result for all four countries. Therefore, ULFR gives a better prediction result than Holt's. The fluctuating case may partially explain that long term prediction is less significant for both models even though ULFR shows significant results than Holt's. Additionally, it is seen clearly from the Figs. 3, 4 and 5 that short term (1 day) prediction is more accurate than long term prediction. In this case, ULFR shows a perfect prediction result than

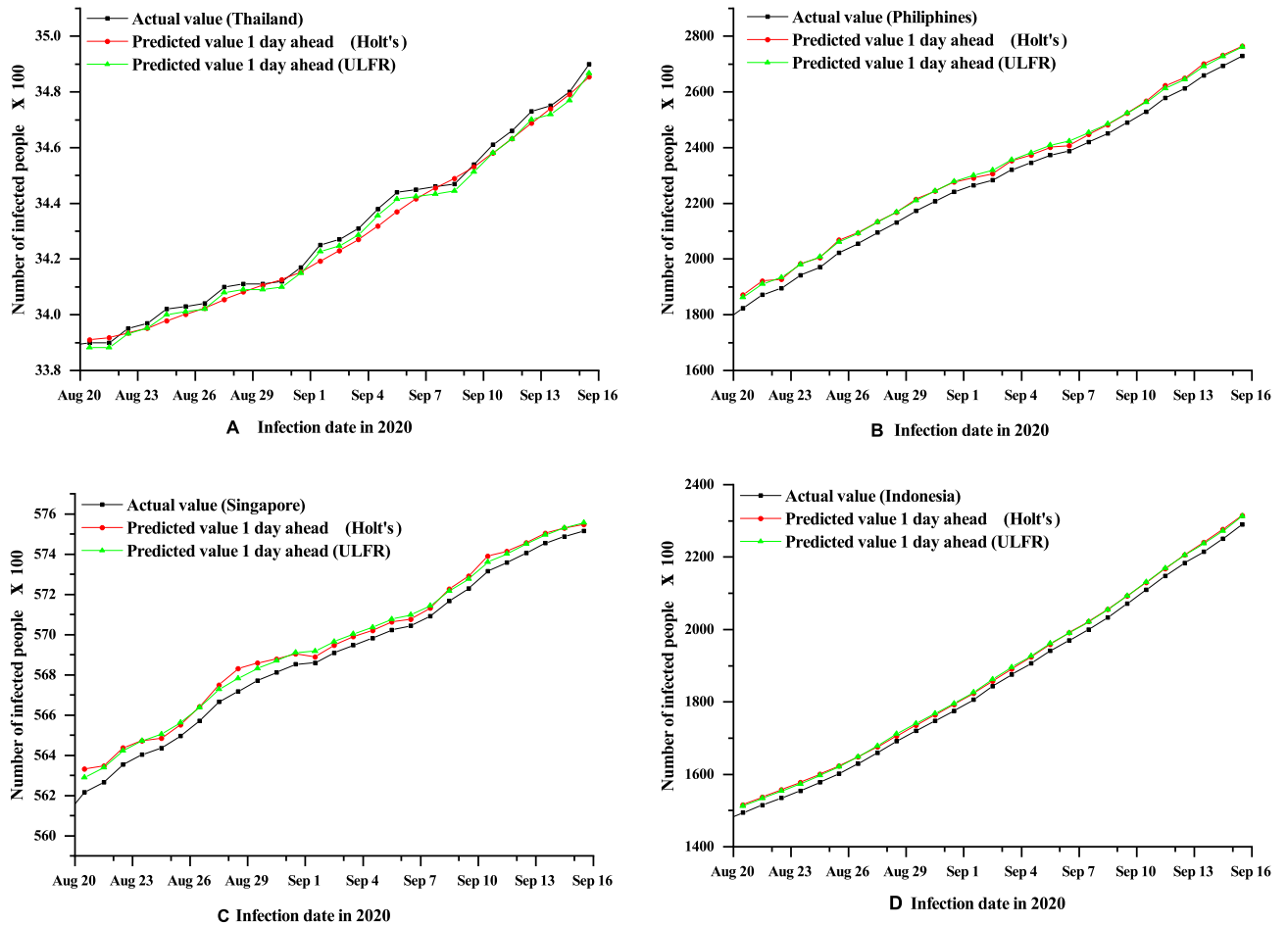


Figure 3. Comparison between Holt's and ULFR method—1 day ahead forecasting of COVID-19 cumulative cases for the four countries Thailand (A), Philippines (B), Singapore (C) and Indonesia (D).

Holt's method. There are studies that investigate the increasing rate at Globally or nationally. For example, Rohit et al.⁴⁰ has mentioned that COVID-19 is increasing globally at a rate of 3% to 5% daily. Sharif et al.¹⁰ commented in their forecasting analysis that the cumulative number of infected cases would be raised to around 500,000 by 2021 in Dhaka division, Bangladesh. However, these studies did not show any justified prediction comparison. But, in our study we have compared original data and predicted data through graphical presentation.

After comparing Figs. 3, 4 and 5 it is clearly visible that day 1 prediction result shows highest accuracy in prediction. Besides, statistical analysis in the Table 3 depicts that highest number of coefficient of determination values of Holt's and ULFR model for 1 day prediction. It is also observed that Fig. 4 depicts better accuracy than Fig. 5 which means 3 days prediction is better than 7 days prediction. However, in a study on deep learning-based forecasting model for COVID-19 outbreak in Saudi Arabia, Brazil, India, Saudi Arabia, South Africa, Spain, and USA⁴¹ found that different countries has different trends. But, that study did not examine all the counties' prediction result with justification. They just showed their accurate result justification by using conventional RMSE, MSE and R^2 value. However, in our study we have analysed prediction result validation by coefficient of determination, NSE and MFE for both Holt's and ULFR result which are presented in the Table 3.

Moreover, Roosa et al.¹³ analyzed daily reported cumulative data of COVID-19 cases of China (Guangdong and Zhejiang) generated from 13 to 23 February 2020 to forecast only 10-day ahead cumulative case by using a generalized logistic growth model, Richards growth model, and a sub-epidemic wave model. Even though the study mentioned that short term prediction, they did not define short term is how much short. In our study we have used deferent regional data and compared the data for different days. In Figs. 1, 2, 3, 4 and 5 it is clearly visible that Philippines and Indonesia forecasting result showing accurate as all the curves in graph B and D are showing perfectly positive regression. However, prediction result of Thailand and Singapore showed moderately positive regression as the prediction result and original result are fluctuated marginally which is visible in the Figs. 1, 2, 3, 4 and 5 C and D section.

In another study the neuro-fuzzy system was used by Al-qaness et al.⁴² by incorporation with marine predators algorithm to predict the COVID-19 cases in Italy, South Korea, Iran, and USA. A high coefficient of determination was obtained for all the forecasted nations; it was 96.48%, 96.96%, 98.74%, and 98.59% for South Korea, Iran, USA, and Italy respectively. On the other hand, in our study a very high coefficient of determination was obtained for day one and day three ahead prediction for all the counties; it was around 99% for all the four

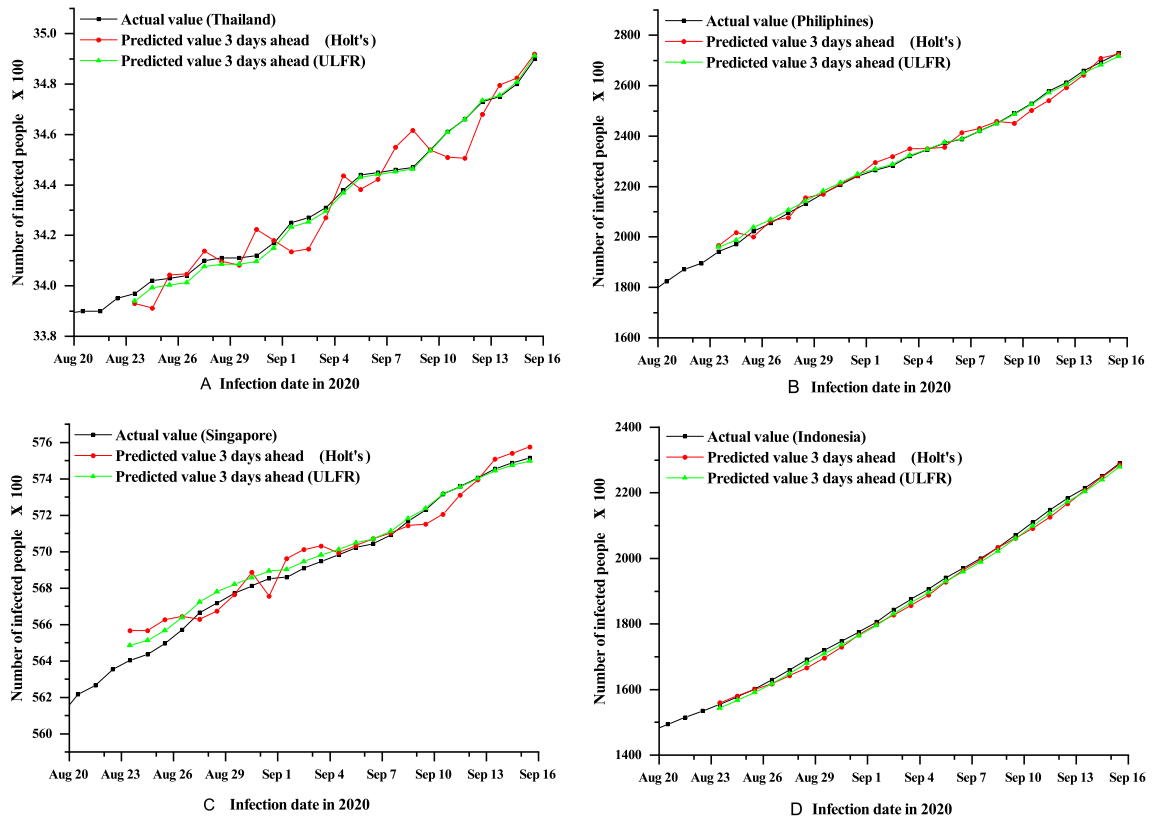


Figure 4. Comparison between Holt's and ULFR method—3 days ahead forecasting of COVID-19 cumulative cases for the four countries Thailand (A), Philippines (B), Singapore (C) and Indonesia (D).

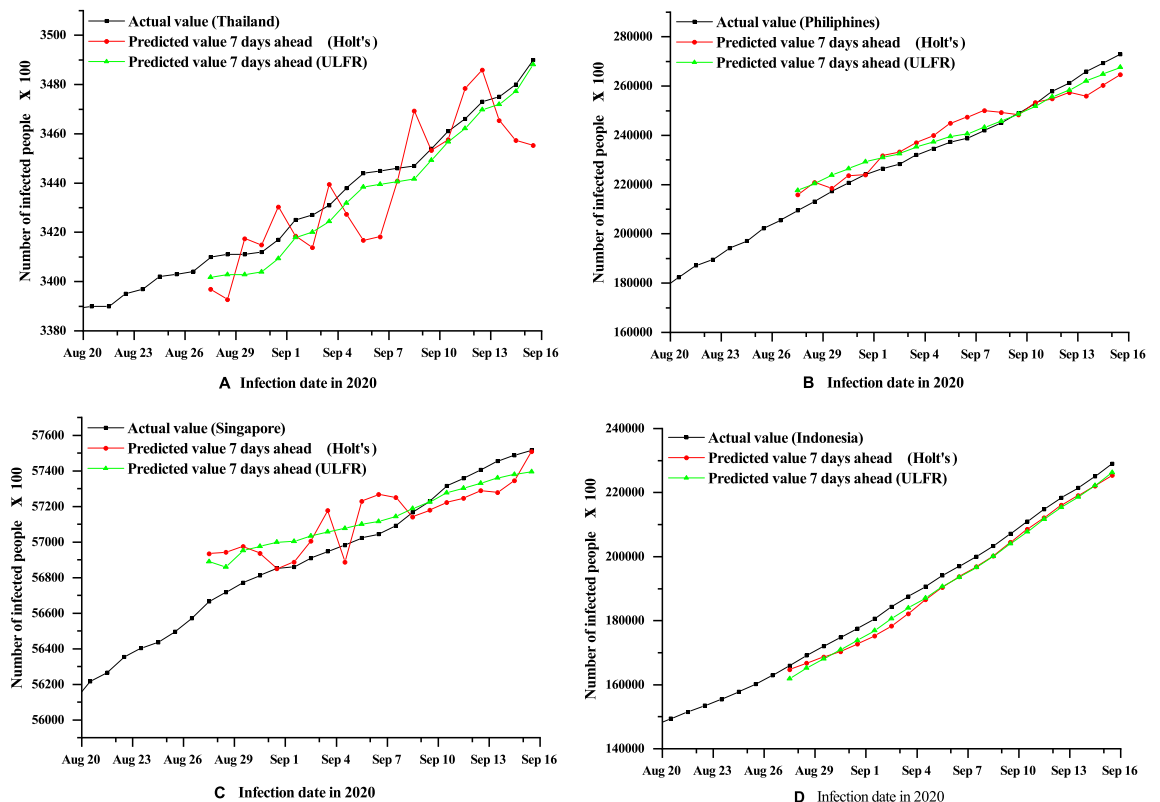


Figure 5. Comparison between Holt's and ULFR method—7 days ahead forecasting of COVID-19 cumulative cases for the four ASEAN countries Thailand (A), Philippines (B), Singapore (C) and Indonesia (D).

countries (Thailand, Philippines, Singapore and Indonesia). However, worst prediction result was seen for 7 days ahead which is around 67% for Thailand.

Conclusion

To sum up, this is the first study to predict the number of infected COVID-19 cases in some ASEAN countries using Holt's and ULFR approaches. Based on our findings, for the short term (1-day prediction), Holt's and ULFR show better results. However, ULFR shows a significant result for 3 and 7 days too. This study's outcome could help ASEAN countries' governments observe the current COVID-19 condition and apply the forecast result to avoid further transmissions. We recommend conducting a follow-up study as many external factors estimate 100% forecast accuracy even though it is impossible to find 100% accurate results. For example, the daily COVID-19 infected cases reported by the government might be lower than the actual number of cases due to delays in publishing test results. Besides, some people would be immune before even screening the COVID-19 test.

There are a few limitations in this study that this study did not include more than 7 days' prediction due to a small data range. Besides, this study could not have all ASEAN countries due to a lack of balance data. Moreover, this study concentrated only on one factor: the number of confirmed cases. Thus, for future study, the next step would be to apply the Holt's and ULFR model to forecast optimal parameters such as the total confirmed cases, total deaths, and total recovered cases for more than 7 days ahead.

Received: 11 November 2021; Accepted: 3 March 2022

Published online: 24 March 2022

References

- McKibbin, W. & Fernando, R. The global macroeconomic impacts of COVID-19: Seven scenarios. *Asian Econ. Pap.* **20**, 1–30 (2021).
- Pell, B., Kuang, Y., Viboud, C. & Chowell, G. Using phenomenological models for forecasting the 2015 Ebola challenge. *Epidemics* **22**, 62–70 (2018).
- Chowell, G. *et al.* Using phenomenological models to characterize transmissibility and forecast patterns and final burden of Zika epidemics. *PLoS Curr.* **31**, 8 (2016).
- Chimmula, V. K. R. & Zhang, L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons Fractals* **135**, 109864 (2020).
- Imai, N., Dorigatti, I., Cori, A., Donnelly, C., Riley, S. & Ferguson, N. Report 2: Estimating the potential total number of novel Coronavirus cases in Wuhan City, China (2020).
- Read, J. M., Bridgen, J. R. E., Cummings, D. A. T., Ho, A. & Jewell, C. P. Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions. MedRxiv (2020).
- Shen, M., Peng, Z., Xiao, Y. & Zhang, L. Modeling the epidemic trend of the 2019 novel coronavirus outbreak in China. *Innov.* **1**, 100048 (2020).
- Cyranoski, D. When will the coronavirus outbreak peak? *Nature* (2020).
- Roda, W. C., Varughese, M. B., Han, D. & Li, M. Y. Why is it difficult to accurately predict the COVID-19 epidemic?. *Infect. Dis. Model.* **5**, 271–281 (2020).
- Sharif, O. *et al.* Analyzing the impact of demographic variables on spreading and forecasting COVID-19. *J. Healthc. Inform. Res.* **6**, 1–19 (2021).
- Fanelli, D. & Piazza, F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons Fractals* **134**, 109761 (2020).
- Escobar, L., Salles, R., Lima, J., Gea, C., Baroni, L., Ziviani, A., *et al.* Evaluating temporal bias in time series event detection methods (2021).
- Roosa, K. *et al.* Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China: February 13–23, 2020. *J. Clin. Med.* **9**, 596 (2020).
- Fokas, A. S., Cuevas-Maraver, J. & Kevrekidis, P. G. A quantitative framework for exploring exit strategies from the COVID-19 lockdown. *Chaos, Solitons Fractals* **140**, 110244 (2020).
- Paul, A., Chatterjee, S. & Bairagi, N. Prediction on Covid-19 epidemic for different countries: Focusing on South Asia under various precautionary measures. Medrxiv (2020).
- Hasan, N. A. & Siddik, M. S. Possible role of meteorological variables in COVID-19 spread: A case study from a subtropical monsoon country, Bangladesh (2020).
- Petropoulos, F. & Makridakis, S. Forecasting the novel coronavirus COVID-19. *PLoS ONE* **15**, e0231236 (2020).
- Khan, M. H. R. & Hossain, A. COVID-19 outbreak situations in Bangladesh: An empirical analysis. MedRxiv (2020).
- Zhou, X., Liang, W., Kevin, I., Wang, K. & Shimizu, S. Multi-modality behavioral influence analysis for personalized recommendations in health social media environment. *IEEE Trans. Comput. Soc. Syst.* **6**, 888–897 (2019).
- Mahmoudi, M. R., Baleanu, D., Mansor, Z., Tuan, B. A. & Pho, K.-H. Fuzzy clustering method to compare the spread rate of Covid-19 in the high risks countries. *Chaos, Solitons Fractals* **140**, 110230 (2020).
- Kong, L., Wang, L., Gong, W., Yan, C., Duan, Y. & Qi, L. LSH-aware multitype health data prediction with privacy preservation in edge environment. *World Wide Web.* 1–16 (2021).
- Kuddus, M. A. & Rahman, A. Analysis of COVID-19 using a modified SLIR model with nonlinear incidence. *Results in Physics* **27**, 104478. <https://doi.org/10.1016/j.rinp.2021.104478> (2021).
- Abdulla, F., Nain, Z., Karimuzzaman, M., Hossain, M. & Rahman, A. A Non-Linear Biostatistical Graphical Modeling of Preventive Actions and Healthcare Factors in Controlling COVID-19 Pandemic. *Int J Environ Res Public Health* **18**(9), 4491. <https://doi.org/10.3390/ijerph18094491> (2021).
- Rahman, A. & Kuddus, M. A. Modelling the Transmission Dynamics of COVID-19 in Six High-Burden Countries. *BioMed Res. Int.* **2021**, 1–17. <https://doi.org/10.1155/2021/5089184> (2021).
- Charles, C. H. Forecasting seasonals and trends by exponentially weighted moving averages. *Int. J. Forecast.* **20**, 5–10 (2004).
- Swamidass, P. M. *Encyclopedia of Production and Manufacturing Management* (Springer, 2000).
- Wright, D. J. Forecasting data published at irregular time intervals using an extension of Holt's method. *Manage Sci.* **32**, 499–510 (1986).
- Fuller, W. A. *Measurement Error Models* (John Wiley. Fuller Measurement Error Model, 1987).
- Chang, Y. F., Choong, W. C., Looi, S. Y., Pan, W. Y. & Goh, H. L. Analysis of housing prices in Petaling district, Malaysia using functional relationship model. *Int. J. Hous. Mark. Anal.* (2019).
- Krause, P., Boyle, D. P. & Båse, F. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* **5**, 89–97 (2005).

31. Nash, J. E. & Sutcliffe, J. V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **10**, 282–290 (1970).
32. Singh, J., Knapp, H. V., Arnold, J. G. & Demissie, M. Hydrological modeling of the Iroquois river watershed using HSPF and SWAT 1. *JAWRA J. Am. Water Resour. Assoc.* **41**, 343–360 (2005).
33. Jain, S. K. & Sudheer, K. P. Fitting of hydrologic models: A close look at the Nash-Sutcliffe index. *J. Hydrol. Eng.* **13**, 981–986 (2008).
34. McCuen, R. H., Knight, Z. & Cutter, A. G. Evaluation of the Nash-Sutcliffe efficiency index. *J. Hydrol. Eng.* **11**, 597–602 (2006).
35. Fethi, M. D. & Pasiouras, F. Assessing bank performance with operational artificial intelligence techniques: A Survey. *Eur. J. Oper. Res.* **204**, 189–198 (2010).
36. Pal, M. & Ayele, Y. Emerging role of foodborne viruses in public health. *Biomed. Res.* **5**, 1–4 (2020).
37. Gibbons, D. L. *et al.* 570 Efficacy, safety and tolerability of MEDI4736 (durvalumab [D]), a human IgG1 anti-programmed cell death-ligand-1 (PD-L1) antibody, combined with gefitinib (G): A phase I expansion in TKI-naïve patients (pts) with EGFR mutant NSCLC. *J. Thorac. Oncol.* **11**, S79 (2016).
38. Hanzák, T. Improved Holt method for irregular time series. WDS. Citeseer pp. 62–67 (2008).
39. Najnudel, J. & Yen, J.-Y. A discussion on some simple epidemiological models. *Chaos, Solitons Fractals* **140**, 110115 (2020).
40. Salgotra, R., Gandomi, M. & Gandomi, A. H. Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. *Chaos, Solitons Fractals* **138**, 109945 (2020).
41. Elsheikh, A. H. *et al.* Deep learning-based forecasting model for COVID-19 outbreak in Saudi Arabia. *Process Saf. Environ. Prot.* **149**, 223–233 (2021).
42. Al-Qaness, M. A. A., Ewees, A. A., Fan, H., Abualigah, L. & Abd, E. M. Marine predators algorithm for forecasting confirmed cases of COVID-19 in Italy, USA, Iran and Korea. *Int. J. Environ. Res. Public Health.* **17**, 3520 (2020).

Author contributions

O.S. and M.Z.H. and A.R. conceived the project concept. O.S. and M.Z.H. cleaned the data and performed the data analysis, model development and interpretation under the supervision of A.R. A.R. refined the analyses results further and conducted model validation with performing specific experimental investigations. O.S. and M.Z.H. wrote the initial draft of this manuscript; A.R. provided critical feedback to improve and complete the manuscript. All authors provided input into revisions and approved the final manuscript and submission for publication.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.Z.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022