Article

Prediction of Epidemic Peak and Infected Cases for COVID-19 Disease in Malaysia, 2020

Abdallah Alsayed [1,*], Hayder Sadir [2], Raja Kamil [1,3,*] and Hasan Sari [4]

[1] Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia

[2] Department of Computer and Wireless Communication, Faculty of Engineering, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia; haydersader@gmail.com

[3] Laboratory of Computational Statistics and Operations Research, Institute for Mathematical Research, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia

[4] College of Computer Science and Information Technology, Universiti Tenaga Nasional, Kajang, Malaysia; hassan_sari@yahoo.com

* Correspondence: eng.abdallah.2013@gmail.com (A.A.); kamil@upm.edu.my (R.K.)

Abstract: The coronavirus COVID-19 has recently started to spread rapidly in Malaysia. The number of total infected cases has increased to 3662 on 05 April 2020, leading to the country being placed under lockdown. As the main public concern is whether the current situation will continue for the next few months, this study aims to predict the epidemic peak using the Susceptible–Exposed–Infectious–Recovered (SEIR) model, with incorporation of the mortality cases. The infection rate was estimated using the Genetic Algorithm (GA), while the Adaptive Neuro-Fuzzy Inference System (ANFIS) model was used to provide short-time forecasting of the number of infected cases. The results show that the estimated infection rate is 0.228 ± 0.013, while the basic reproductive number is 2.28 ± 0.13. The epidemic peak of COVID-19 in Malaysia could be reached on 26 July 2020, with an uncertain period of 30 days (12 July–11 August). Possible interventions by the government to reduce the infection rate by 25% over two or three months would delay the epidemic peak by 30 and 46 days, respectively. The forecasting results using the ANFIS model show a low Normalized Root Mean Square Error (NRMSE) of 0.041; a low Mean Absolute Percentage Error (MAPE) of 2.45%; and a high coefficient of determination (R^2) of 0.9964. The results also show that an intervention has a great effect on delaying the epidemic peak and a longer intervention period would reduce the epidemic size at the peak. The study provides important information for public health providers and the government to control the COVID-19 epidemic.

Keywords: COVID-19; SEIR model; epidemic peak; infection rate; basic reproductive number; ANFIS; GA

1. Introduction

Coronavirus disease (COVID-19) is an infectious disease first reported in China [1]. COVID-19 has been confirmed on 25 January 2020 in Malaysia and currently continues to spread fast in the country, which seriously jeopardizes the lives of elderly people as well as those of any age who experience a serious underlying medical condition [2]. Figure 1 shows the accumulated number of infected cases due to COVID-19 from 25 January to 05 April in Malaysia. It can be observed that the COVID-19 outbreak started to be a pandemic after 27 February, such that more than 98.77% of the total infected cases was reported after this date. This outbreak is mostly attributed to a special religious gathering of more than 15,000 persons between 27 February and 2 March at a local mosque,

which was an infection cluster and the main source of the spike in COVID-19 cases according to the Ministry of Health in Malaysia [3]. The spread of the virus came from the foreign participants who came into Malaysia and participated in the gathering. The sudden increase in the number of infected cases after 12 March is probably due to the fact that infected people without COVID-19 symptoms could significantly spread the infection [4]. Furthermore, the diagnostic tests were initially only made available to those who attended the religious gathering.

Unfortunately, COVID-19 cannot be controlled as there is no proven pharmaceutical-based treatment up to now. However, other behavioral strategies, such as lockdown and movement control of people, can be effective to reduce the number of new cases and delay the epidemic peak. The Malaysian Government has promulgated the restricted activities order on 18 March, which prohibits all mass movements and gatherings across the country, including religious, sports, social, and educational activities. The movement control order was implemented in several stages with the strictness and punishment increasing with each stage to ensure that the public conform to the restrictions. However, exclusions are given to public markets, grocery stores, and convenience stores selling food and essential items. The main public concern is whether the epidemic will continue until August 2020, which would affect the economy, and in particular the tourism plan of "Visit Malaysia 2020" that attracts Middle Eastern and Chinese tourists during the holiday season from June through August. Therefore, the short- and long-term prediction of the COVID-19 epidemic is needed to provide important information for healthcare providers and government that would help them to implement effective intervention measures and policies.
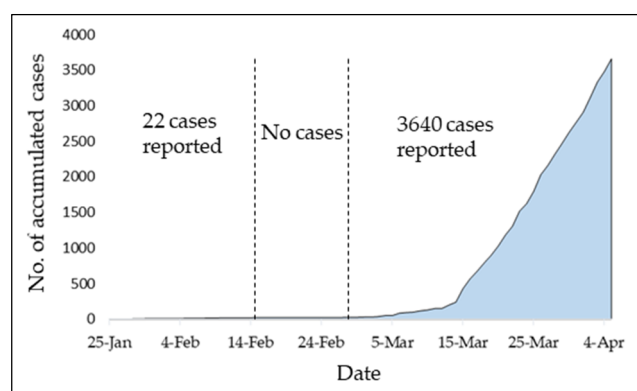
Figure 1. Growth in the total number of infected cases in Malaysia.

Mathematical modeling plays an important role in predicting the epidemic peaks of COVID-19 using real-time historical data [5]. Many statistical and numerical models have been used to predict the COVID-19 outbreaks, such as the Logistic Growth model [6], stochastic Susceptible–Infectious–Removed (SIR) model [7], and Natural Growth model [8]. However, the SEIR (Susceptible–Exposed–Infectious–Recovered) model is still the most widely used to characterize the epidemic peak of COVID-19 in China [9–11], Japan [12], Italy [13], and Iran [14]. Besides, the SEIR model was used to compare the effect of the lockdown of Hubei province on the infection rates in Beijing and Wuhan [15]. On the other hand, in forecasting the number of infected cases for the upcoming few days, the mathematical models are not effective as many parameters should be daily updated and estimated. Thus, the accuracy of short-time forecasting using parametric models may not be high [16].

The infection rate (or transmission rate) parameter provides information on the probability of transmission of COVID-19 from an infectious individual to susceptible individuals [17]. It is one of the two components in the basic reproductive number by which the continuous increase or decrease in the infected cases is decided. In calculating the infection rate, the most common method is the asymptotic statistical theory [18], in which the least-squares method is used to quantify the uncertainty associated with infection rate estimation. However, the least-squares method is subjected to low accuracy that accompanies the estimation of the infection rate. A possible solution is to run the estimation process 10,000 times and then obtain the normal distribution of the infection rate values

Int. J. Environ. Res. Public Health **2020**, *17*, x

3 of 15

with 95% confidence intervals, which would decrease the uncertainty and increase the accuracy as in [12,19]. This method highly increases the time of the estimation process, especially when the range of the hypothesized infection rate is relatively large with a small resolution.

To our knowledge, there are no scientific studies related to the pandemic of COVID-19 spread in Malaysia. Thus, this study is conducted to (1) estimate the infection rate using the Genetic Algorithm (GA); (2) predict the epidemic peak of COVID-19 using the SEIR model, incorporating also the mortality in the population due to COVID-19; and (3) forecast the number of infected cases for the upcoming five days using the Adaptive Neuro-Fuzzy Inference System (ANFIS) predictive model. The available data of infected cases from 25 January to 05 April 2020 in Malaysia was used to calibrate the SEIR model. For forecasting, the data from 22 to 31 March was used to train and test the ANFIS model, while the data from 01 to 05 April was used to validate the ANFIS model.

2. Methods

2.1. SEIR Model for Peak Prediction

The SEIR model that characterizes the epidemic COVID-19 outbreaks is described as follows [20,21]:

$$dS(t)/d(t) = -\beta S(t)I(t),$$

$$dE(t)/d(t) = \beta S(t)I(t) - \alpha E(t),$$

$$dI(t)/d(t) = \alpha E(t) - \gamma I(t) - MI(t), \tag{1}$$

$$dR(t)/d(t) = \gamma I(t),$$

$$dD(t)/d(t) = MI(t)$$

where *S, E, I, R,* and *D* represent the number of susceptible, exposed (not yet infectious), infective, recovered, and death cases given at time $t > 0$. The coefficients β, α, γ, and M denote the infection, onset, removal, and mortality rates. Based on the recent studies related to COVID-19 [22–25], the incubation (α^{-1}) and infectious (γ^{-1}) periods are 5 days and 10 days, respectively. Thus, the α and γ values are 0.2 and 0.1, respectively. The total number of deaths and confirmed cases up to 5 April are 61 and 3662, respectively, and thus the mortality rate M is 0.016 (61/3662). We fixed the unit time to be 1 day and $S + E + I + R + D = 1$, such that each population implies the proportion to the total population. Let assume that there is one infected case recorded at time $t = 0$ among the Malaysian population of $N = 32.6 \times 10^6$ [26]; that is, $X(0) = pNI(0) = 1$, where

$$X(t) = pNI(t), \tag{2}$$

where X is the number of infected cases that are identified at time *t*, and *p* is the identification rate such that we obtain $I(0) = 1/(p \times 32.6 \times 10^6)$. The block diagram of the SEIR model is attached in Appendix A as Figure A1. It is assumed that there are no exposed, recovered, and death cases at $t = 0$, and hence,

$$S(0) = 1 - (E(0) + I(0) + R(0) + D(0)) = 1 - \frac{1}{pN} \tag{3}$$

In Malaysia, the COVID-19 test is mainly performed on those with close contacts to the patients as well as on those with COVID-19 symptoms. We assume that the identification rate is not significantly dependent on the test kit availability as the Malaysian government is able to perform the test for 11,500 persons a day and the average number of daily tested cases is 2500 persons [27]. In [28], 3662 cases are currently confirmed among the tested 43,595 infected cases from 25 January to 05 April. Based on that, *p* is equal to 0.084 (3662/43,595). The basic reproductive number $\mathscr{R}_0$ represents the expected number of secondary cases resulted from an infected individual [29]. It is calculated as the leading eigenvalue of the next generation matrix $G = FV^{-1}$ [30], where

$$F = \begin{bmatrix} 0 & \beta S(0) \\ 0 & 0 \end{bmatrix}, \ V = \begin{bmatrix} \alpha & 0 \\ -\alpha & \gamma \end{bmatrix}, \tag{4}$$

where *F* is a new infection, while *V* represents the transfers of infections from one compartment to another [5]. Then, we obtain

$$\mathscr{R}_0 = \frac{\beta S(0)}{\gamma} = \frac{\beta}{\gamma}\left(1 - \frac{1}{pN}\right) \approx \frac{\beta}{\gamma}. \tag{5}$$

It is obvious that the basic reproductive number only depends on the infection rate (β) and the removal rate (γ). Besides, the influence of the identification rate on $\mathscr{R}_0$ is negligible as the population number (N) is 32.2×10^6. The coefficient parameters of the SEIR model are summarized in Table 1. Note that the estimation of β *and* $\mathscr{R}_0$ values are presented in the next subsection.

Table 1. Coefficient values for the Susceptible–Exposed–Infectious–Recovered (SEIR) model.

Coefficient	Description	Value
α	Onset rate	0.2
γ	Removal rate	0.1
M	Mortality rate	0.016
N	Malaysia population	32.6×10^6
p	Identification rate	0.084

2.2.β. Estimation using GA

In this study, the estimation of the infection rate is accomplished using the Genetic Algorithm [19]. Let us assume $X(t)$ (described in Equation (2)), $t = 0, 1, \ldots, 72$, is the number of daily infected people due to COVID-19 in Malaysia from 25 January to 05 April. We assume $X(t)$ is subjected to the Poisson noise, which reflects the fluctuations of the number of infected cases, so that

$$\dot{X}(t) = X(t) + \varepsilon X(t)^\xi, \text{ Poisson noise} = \varepsilon X(t)^\xi, \tag{6}$$

where $\dot{X}$ is our deterministic model with Poisson noise, while ε is a random variable from a normal distribution with a mean zero and a standard deviation of 1. The ξ is equal to 0.5, such that the variance of the error scales is linear with $X(t)$ and this value refers to the Poisson noise as described by [31]. The classical GA was applied to estimate the β value that minimizes the cost function. The cost function is represented by the sum of squares, as in Equation (7). The β value ranged from the lower bound to upper bound values. The lower and upper bounds of the β value were selected as 0.2 and 0.4, respectively. The minimum cost function $C(\beta'')$ is defined as in Equation (8).

$$C(\beta) = \sum_{t=0}^{72}\left[X(t) - \dot{X}(t)\right]^2, \tag{7}$$

$$C(\beta'') = min_{0.2 \le \beta \le 0.4} C(\beta), \tag{8}$$

The classical GA algorithm was then implemented to find the optimum β values that minimize the cost function using five steps, as follows [32–34]:
a. Population initialization: In order to find a solution to the problem of the cost function, the GA initially creates a number of populations that randomly encodes the chromosomes (individuals). Then, the cost values of the generated population are evaluated.
b. Selection: In this process, each individual identified by its associated cost is ranked and the corresponding individual fitness is selected. According to fitness, the best chromosomes from the population are then selected such that better fitness has a bigger chance to be selected. Subsequently, the solutions selected from one population are implemented to form a new population. This process is motivated by the new population potentially being better than the previous one. The selection process is performed using a certain function that fixes the generation gap. The selected individuals are then recombined.
c. Crossover: To make new offspring (children) for the following iteration, the selected individuals (parents) have to undergo a crossover with a crossover probability. However, if there is no crossover performed, the offspring is an exact copy of the parents.

Int. J. Environ. Res. Public Health **2020**, *17*, x

5 of 15

d.   Mutation: In this process, the information in the chromosomes is randomly modified. The genes occasionally mutate to be converted to novel genes. Based on mutation, it is possible to control the multifariousness of the population as well as to enhance the search capacity of the search scheme.

e.   Evaluation: For each individual, the cost function of the optimization problem is calculated. The stopping criterion of the GA is the number of iterations after which the process is stopped. For each iteration, the β value that has the minimum cost function is recorded. The distribution of the β values is then approximated by a normal distribution with a mean and standard deviation.

The flowchart of the GA for β estimation is demonstrated in Figure 2. The GA parameters are provided in Table 2 and obtained based on the trial and error method. The Optimization Toolbox of the MATLAB® software (MathWorks Inc.) was used to implement and run the GA algorithm.
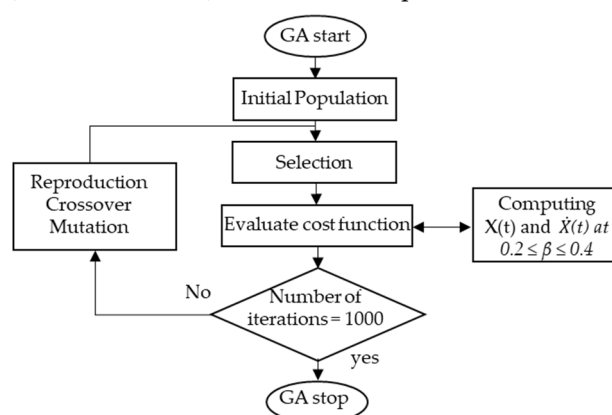
Figure 2. Genetic Algorithm (GA) flowchart for β estimation.

Table 2. GA parameters.

Parameter	Value	Parameter	Value
Population size	200	Mutation rate	0.02
Number of iterations	1000	Mutation percentage	0.9
Crossover percentage	0.95		

2.3. ANFIS for Short-Term Forecasting

ANFIS is a nonparametric model used to solve a nonlinear problem with a small dataset in one framework. It has a powerful hybrid learning capability using an Artificial Neural Network (ANN) and a Fuzzy Logic model to generate an effective processing tool for prediction [35]. The core element of ANFIS is the Fuzzy Interference System (FIS) that is embedded into a framework of adaptive networks that use "IF–THEN" rules to model the behavior of an uncertain system. These adaptive networks contain a number of adaptive nodes connected through directional links. Each adaptive node has a modifiable parameter updated using the fuzzy learning rule aiming to minimize the errors. In this study, the FIS system uses one input x and one output y. The ANFIS model structure is shown in Figure 3. The first order Sugeno fuzzy model with fuzzy "IF–THEN" rules is employed as follows [36]:

$$\text{Rule 1: if } x \text{ is } A_1 \text{ then } y_1 = P_1 x + r_1, \tag{9}$$

$$\text{Rule 2 : if } x \text{ is } A_2 \text{ then } y_2 = P_2 x + r_2. \tag{10}$$

Layer 1 contains the member functions (MFs) of the inputs and generates the input variables for Layer 2. Each node in this layer is adaptive using Equation (11). The MF type used in this study is the Gaussian function, for which 0 and 1 are the lowest and highest values, respectively.

Int. J. Environ. Res. Public Health **2020**, 17, x

6 of 15

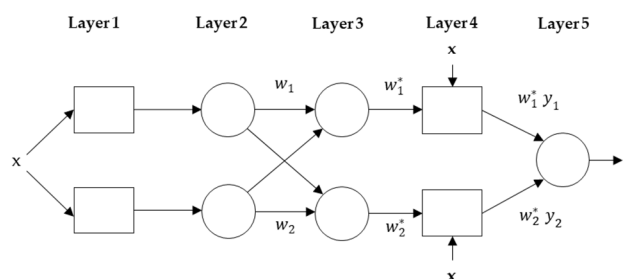$$Q_i = \mu_{Ai}(x), \text{ where } \mu(x) \text{ is MF.} \tag{11}$$

Figure 3. ANFIS structure.

Layer 2 is a membership layer in which the weights of MFs are computed and considered. Input variables of this layer are obtained from the first layer. Noted that, the layer's nodes are fixed nodes. The output of the second layer is a product of all incoming inputs and described as in Eq.12, where w_i represents the weight strength of one rule.

$$w_i = \mu(x)_i \, \mu(x)_{i+1} \text{ and } i = 1,2. \tag{12}$$

In Layer 3 (rule layer), the weight function is normalized and the outputs of this layer are called normalized weights or firing strengths. The normalization is described as:

$$\overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad \text{and } i = 1,2. \tag{13}$$

Layer 4 is the defuzzification layer such that the output from Layer 3 is multiplied with the Sugeno fuzzy rule function as follows:

$$Q_i^4 = w_i^* y = w_i^*(p_i x + r_i), \tag{14}$$

Layer 5 is the output layer in which the inputs and outputs from the previous layer are formulated. Furthermore, this layer converts the results into a crisp output. Thus, all incoming inputs are sum up producing the overall output as follows:

$$Q_i^5 = \sum_i w_i^* y_i = \frac{\sum_i w_i \, y}{\sum_i w_i} \tag{15}$$

Noted that the ANFIS MFs parameters are adjusted (tuned) using the hybrid method of backpropagation and least square techniques [37]. The Neuro-Fuzzy Designer of Matlab® Software (MathWorks Inc.) is used to implement the ANFIS parameters that are summarized in Table 3. In this study, as the number of infected cases is nonlinearly changed from day to day, the ANFIS model is used. The ANFIS model forecasts the numbers of infected cases for the upcoming 5 days based on the numbers of infected cases for the last 10 days. The dataset of 10 days is divided into training (70%) and testing (30%) datasets which are implemented in the ANFIS model. After that, the trained ANFIS model is used to forecast the numbers of cases for the next 5 days. The input and output variables are day number and number of infected cases, respectively.

Table 3. Adaptive Neuro-Fuzzy Inference System (ANFIS) parameters.

Parameter	Method/Value	Parameter	Method/value
Fuzzy structure	Sugeno-type	No. of epochs	300
Rules clustering	Grid partition	Input	Day number
MF type	Gaussian	Output	Infected cases
Optimization method	Hybrid	Output MF	constant

In order to investigate the performance of the ANFIS model, the Root Mean Square Error mean (RMSE), normalized RMSE (NRMSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2) were used as follows [38,39]:

Int. J. Environ. Res. Public Health **2020**, *17*, x

7 of 15

$$\text{RMSE} = \sqrt{\frac{1}{t}\sum_{t=0}^{t}(y_{actual} - y_{estimated})^2}, \text{ and}$$

$$\text{NRMSE} = \frac{RMSE}{y_{max} - y_{min}},$$

(16)

$$\text{MAPE} = \frac{y_{actual} - y_{estimated}}{y_{actual}}$$

(17)

$$R^2 = 1 - \frac{\sum_{t=0}^{t}(y_{actual} - y_{estimated})^2}{\sum_{t=0}^{t}(y_{actual} - y_{average})^2}.$$

(18)

3. Results

3.1. Infection Rate (β) Estimation

GA was applied to estimate the optimum infection rate between *0.2 ≤ β ≤ 0.4* by minimizing the cost function described in Equation (8). Figure 4 depicts the cost values for 1000 iterations. It is observed that the GA searching for the minimum cost value converges to the value of 1.098×10^{-9} at the iteration number 819, which indicates that there is no better cost value than 1.098×10^{-9} based on GA. The optimum *β* values obtained for the entire population size of 200 is shown in Figure 5. The *β* values are approximated by the normal distribution and, subsequently, the infection rate *β* is 0.228 ± 0.013. Based on Equation (5), the basic reproductive number is 2.28 ± 0.13 as *γ* = 0.1.
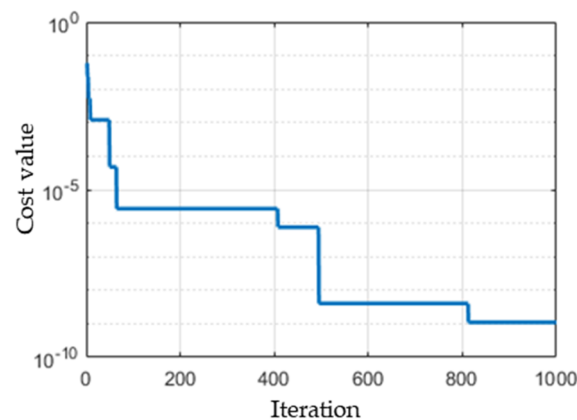

Figure 4. Cost values of 1000 iterations.
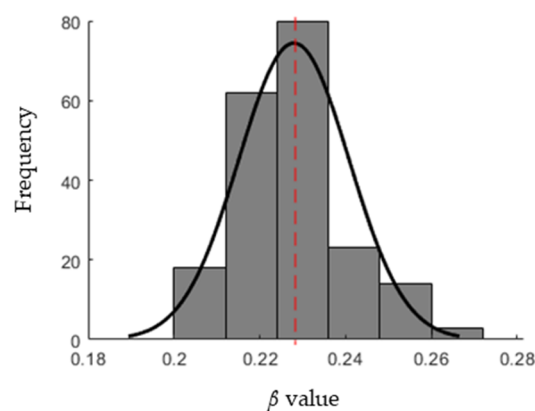
Figure 5. Normal distribution of optimum *β* values. The dotted line represents the mean value.

3.2. Epedimic Peak Prediction

Given that the major outbreak occurs after the second wave, it is assumed that the influence of the number of cases reported before the second wave is negligible in estimating the identification and infection rates. Besides, this assumption is considered due to the absence of the reported numbers related to cases that tested negative during the first wave. The epidemic peak is estimated when a maximum is attained within one year, such that $X(t_{max}) = \max_{0 < t < 365} X(t)$. Based on the current report, the p is around 0.084. Subsequently, Figure 6 shows a one-year behavior of $X(t)$ for the determined infection rate $\beta = 0.228 \pm 0.013$.

It is observed that the epidemic peak may occur between 170 ($\beta = 0.241$) and 200 ($\beta = 0.215$) with an average of 184 ($\beta = 0.228$). This indicate that, starting from 25 January, the predicted epidemic peak is on 26 July ($t = 184$), with deviation from 12 July ($t = 170$) to 11 August ($t = 200$). The COVID-19 pandemic will last until 15 December 2020 ($t = 326$), with the deviation ranging from 22 November 2020 ($t = 303$) to 12 January 2021 ($t = 354$).

Based on the entire period since the COVID-19 onset in Malaysia, the p value ranges from 0.01 to 0.084. Hence, we also estimate the epidemic peak at $p = 0.01$. Figure 7 shows the $X(t)$ over one year for $\beta = 0.228 \pm 0.013$. As seen, the predicted epidemic peak is 19 June ($t = 147$) and the uncertainty range is from 08 June ($t = 136$) to 02 July ($t = 160$). The COVID-19 pandemic will last until 29 September ($t = 249$) with the deviation ranging from 13 September ($t = 233$) to 19 October ($t = 269$). In contrast to the basic reproductive number $\mathcal{R}_0$, it is clear that the epidemic peak and size are responsive to the identification rate p. Furthermore, a lower identification rate leads to a lower number of infected cases, such that the number of infected cases decreases from 2.582×10^5 to 3.077×10^4 at the epidemic peak with $p = 0.01$.
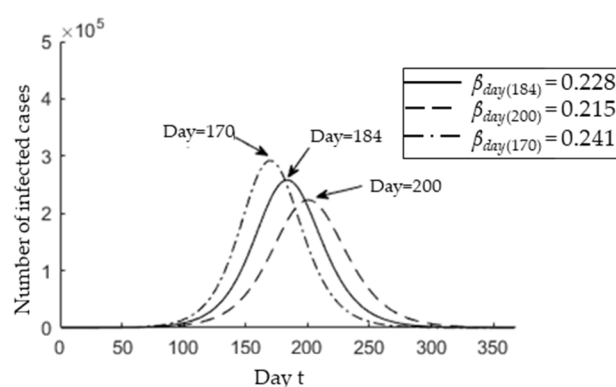
Figure 6. Real-time variation in the number of infected cases identified at time t for $p = 0.084$ and $\beta = 0.215$, $\beta = 0.228$, and $\beta = 0.241$. The text arrows represent the t_{max} for each infection rate.
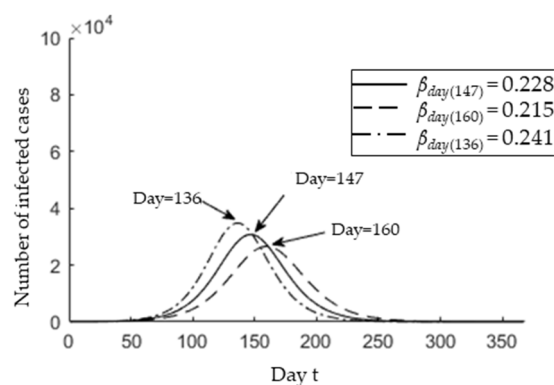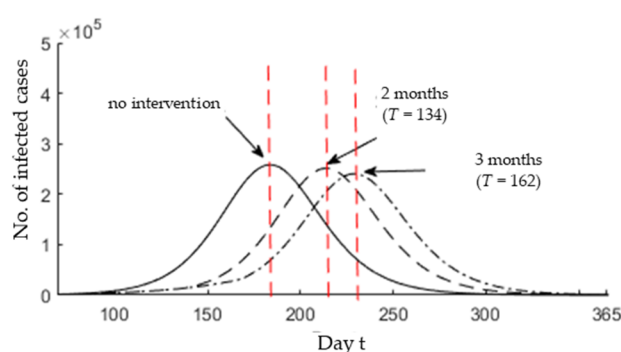

Figure 7. Real-time variation in the number of infected cases identified at time t for $p = 0.01$ and $\beta = 0.215$, $\beta = 0.228$, and $\beta = 0.241$. The text arrows represent the t_{max} for each infection rate.

*3.3. Epidemic Peak after Possible Interventions*

In this subsection, the effect of possible interventions is investigated. In Malaysia, all universities, schools, and workshop places have been closed and most of the social events have been canceled from 17 to 26 March to eliminate the contact risk. However, the government has extended the closure to 14 April as the number of infected cases is still rising by an average of 170 cases per day. Thus, the current governmental effort seems to be limited to contain the COVID-19 up to now.

We assume that the governmental and social efforts can reduce the infection rate $\beta = 0.228$ by 25% of its value ($\beta_{new} = 0.17$) during the period from 05 April ($t = 72$) to the desired day ($t = T > 72$), and we fix p to 0.084 in what follows. Firstly, it is assumed that the intervention is adopted for 2 months; that its, T is equal to 134 (72 + 62). In this situation, the epidemic peak t_{max} is shifted 30 days later from 26 July to 26 August. It is clear that the epidemic size remains relatively unchanged. On the other hand, if the interventions are adopted for three months from 05 April to 04 July ($T = 72 + 92 = 164$), then the epidemic peak t_{max} is moved back from 26 July to 09 September. It can be observed that the epidemic size is significantly reduced. Figure 8 shows the real-time prediction of infected cases that are identified between $t = 0$ and $t = 365$ for no intervention, two months intervention, and three months intervention.

Figure 8. Real-time variation in the number of infected cases ($0 \leq t \leq 365$) for $p = 0.084$. The red dotted lines represent the epidemic peak.

We can also generalize the desired day for possible interventions over $72 < T < 365$, as shown in Figure 9a. It is observed that the epidemic peak t_{max} is linearly delayed as the intervention period increases from $72 \leq T \leq 263$ and then fixed to t_{max} for $T > 263$.


(a) (b)
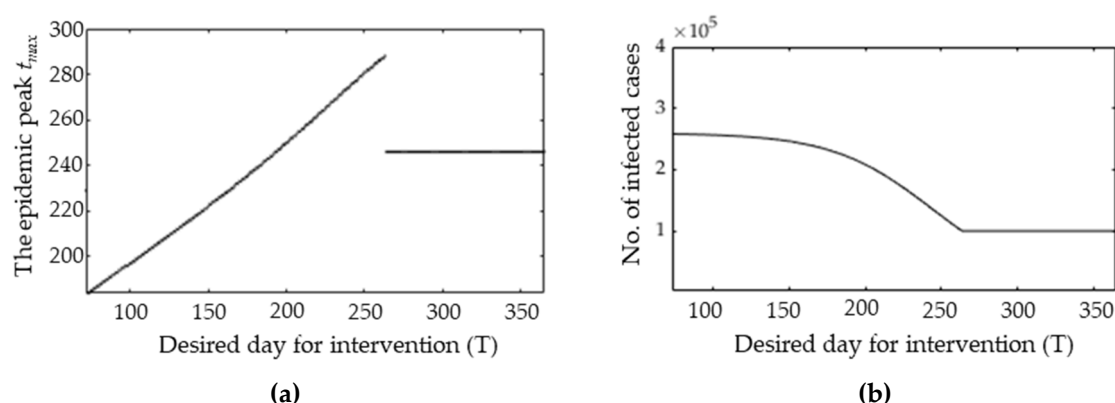
Figure 9. The relationship between the desired day for intervention T and (**a**) the epidemic peak t_{max}; (**b**) the number of infected cases at epidemic peak t_{max}.

The figure also indicates that the interventions have a positive effect to delay the epidemic peak, which may give the government more time to contain the COVID-19 and flatten the curve. Figure 9b shows the relationship between the intervention period (T) and the number of infected cases at the epidemic peaks $X(t_{max})$. It is observed that the number of infected cases is monotonically declined and

Int. J. Environ. Res. Public Health **2020**, *17*, x

10 of 15

fixed as *T* increases. Interestingly, the change in the number of infected cases is rapidly increased for *T* > 72. This implies that an early intervention over a relatively small duration can be effective to reduce the epidemic size and flatten the curve.

3.4. Short-Term Forecasting

The ANFIS model was mainly used to forecast the infected cases for the next five days based on the historical data of 10 days. Firstly, the historical data is randomly split into training and testing datasets according to a 70%:30% ratio to make sure the model is not subjected to overfitting. Figure 10 shows the training and testing errors over the 300 epochs (iterations). Estimated (ANFIS output) and actual infected cases are depicted in Figure 11. Table 4 presents the RMSE, NRMSE, MAPE, and R^2 obtained while training the ANFIS model using the training and testing datasets.
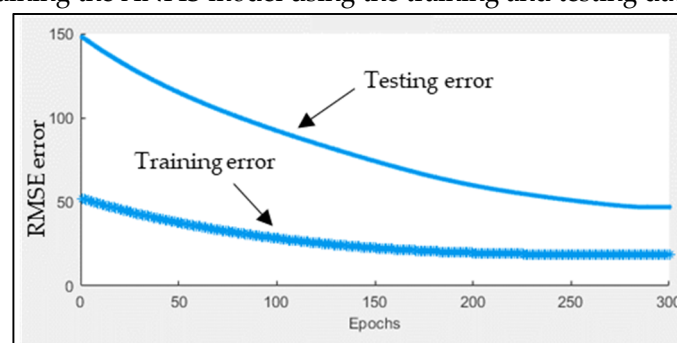
Figure 10. The upper and lower curves represent the training and testing errors, respectively.

Table 4. Performance of the ANFIS model.

Parameter	Training Data	Testing Dataset
RMSE	18.53	46.87
NRMSE	0.012	0.032
MAPE	1.31%	2.79%
R^2	0.9973	0.9998

(a)   (b)
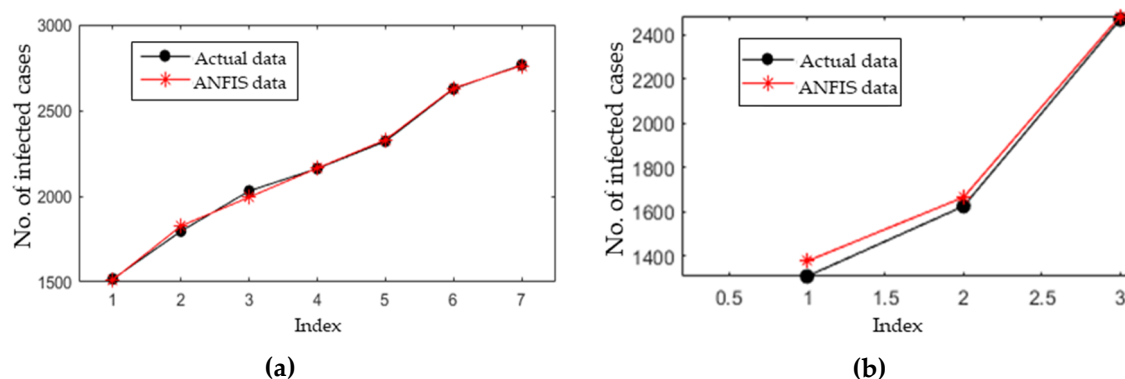
Figure 11. Estimated and actual infected cases using the (**a**) training dataset and (**b**) testing dataset.

Secondly, the developed ANFIS model was then used to forecast the number of infected cases for the next five days. The results of the forecasted and actual number of infected cases are presented in Figure 12. The performance of the ANFIS model to forecast is as follows: the RMSE, NRMSE, MAPE, and R^2 values are 96.8, 0.041, 2.45%, and 0.9964, respectively. These results indicate a very low RMSE, NRMSE, and MAPE, but a high R^2.

Int. J. Environ. Res. Public Health **2020**, *17*, x

11 of 15
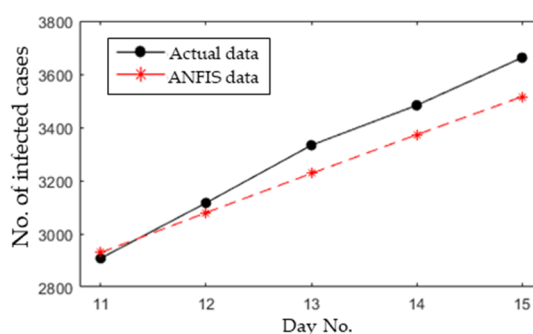
Figure 12. Forecasting results for the next five days.

4. Discussion

This study mainly aims to (1) estimate the infection rate using the GA algorithm; (2) predict the epidemic peak of COVID-19; and (3) forecast the number of infected cases for the upcoming five days based on historical data of the last ten days. First, the confirmed cases from 25 January to 05 April was used to find the coefficient parameters of the SEIR model. Subsequently, the GA was applied to find the infection rate value that minimizes the function of the SEIR model with Poisson noise. As a result, the infection rate is 0.228 ± 0.013. Based on Equation (5), the basic reproductive number $\mathscr{R}_0$ is 2.28 ± 0.13. This value is relatively close to the estimated value by the World Health Organization (WHO), which ranges from 2 to 2.5 for COVID-19 [40]. In addition, this value is not so different from recent estimations: 2.24–3.58 [41], 2.0–3.1 [42], and 2.06–2.52 [43] for COVID-19. However, some studies reported higher $\mathscr{R}_0$ values of 3.28, 2.90, and 3.11, as reported in [44–46], respectively. This bias in estimating the $\mathscr{R}_0$ value is probably attributed to limited available data over a short period and also highly depends on the settings. Furthermore, the estimation of $\mathscr{R}_0$ strongly relies on the estimation method and the validity of the assumptions for some coefficients. Thus, the availability of more data over a long period would provide a more accurate estimation and form a clearer trend.

Secondly, the SEIR model incorporating the mortality in the population due to COVID-19 was used to predict the epidemic peak of COVID-19 in this study. The epidemic peak in Malaysia could be reached late July 2020 and the uncertainty range is from 12 July to 11 August 2020. The results also indicated that the COVID-19 trend in Malaysia will not flatten too quickly. This indication might be consistent with the WHO's statement [47] that COVID-19 is not a seasonal virus and thus will not disappear in the summer, such as the flu. It should be noticed that the epidemic estimation may be subjected to some variability, such that possible big change in social and natural situations would shorten the range of the peak estimation. Besides, the epidemic estimation relies on the mathematical modeling used to describe the epidemic. A complex model with more biological and epidemiological variables is more realistic. However, it requires more model parameters and coefficients to be estimated compared to a simpler one. Therefore, it is important to keep a balance between biological realism and eliminating the variability in the model prediction with a view to increase the reliability of the predictions.

The findings obtained for epidemic peak prediction are as follows: (1) the epidemic size is not affected by the identification rate, which ranges from 0.01 to 0.084 for the total population in Malaysia; (2) a near-future intervention has a great effect to postpone the epidemic peak that would give the government and healthcare providers more time to optimize the medical environment by training more staffs to deal with COVID-19; and (3) a longer period intervention should be taken into account to reduce the epidemic size. Although the Malaysian government has implemented the Movement Control Order (MCO) towards COVID-19 on 18 March throughout the country, the number of daily confirmed cases is still rising with an average of 170 cases for the last two weeks. Besides, more critical cases requiring intensive care units are being recorded. This trend is due to the following possible reasons:

1. The number of people who had contact with COVID-19 patients is enormous, as reported in [48]. This could make the process of tracking and isolating more complex.

Based on the information reported by Chinese medical doctors involved in Wuhan, the critical cases form 10% of the total number of infected people. The early diagnosis and treatment would reduce the flow of COVID-19 patients into the ICU unit [49].

2. Poor experience in treating and managing cases with different levels of infection. For instance, severe cases should be kept under monitoring with intensive care, while mild cases without clear symptoms should be kept with less intensive care in the hospitals. However, patients under investigation should be placed in special isolation outside the hospitals. This kind of management would ease the treating process with the currently available equipment [50].

3. The current MCO implemented in Malaysia is limited to aiding the awareness of the people to the danger of COVID-19. For the first 10 days of the MCO, 60% of the public has obeyed the MCO issued by the government [51]. Thus, more restrictions are needed to enforce the MCO. By increasing the public awareness, the infection rate will be reduced, which would result in decreasing the reproductive number and delaying the epidemic peak.

Lastly, this study provides short-term forecasting for the number of infected cases based on the ANFIS model. The results indicate a high forecast precision is achieved based on the ANFIS model. The ANFIS model achieved 1) an excellent coefficient of determination (R^2 = 0.9964), which is very close to the perfect value of 1; 2) a low NRMSE value (NRMSE = 0.041), which is highly close to the perfect values of 0; and 3) a high MAPE value (MAPE = 2.45%), which is less than 10% [52]. The main motivation behind using the ANFIS model instead of parametric models (e.g., likelihood and Bayesian methods) is that ANFIS is able to achieve a high accuracy using only a few datasets and is easy to be deployed, such that the ANFIS model uses one input as day number, while parametric models require at least four inputs as well as estimation of the coefficients.

This study has some limitations. First, the SEIR model is used with a limited number of cases and COVID-19 is highly infectious; so, the current results of peak estimation are constrained to a limited period and may be changed after inputting a considerable number of infected cases. Secondly, the estimation is based on the available data from the WHO. A possible delay in confirming or reporting could result in an underestimation of $\mathscr{R}_0$. Lastly, the ANFIS model is applicable for short-term forecasting, and so it cannot be used to predict the epidemic peak of COVID-19 as the ANFIS model does not consider the recovered and death rates.

5. Conclusions

As the main public concern in Malaysia is whether the COVID-19 spread will continue for the upcoming few months, we provide here information on predicting the epidemic peak using the SEIR model, estimating the infection rate using the GA algorithm, and short-time forecasting using the ANFIS model. The results related to the epidemic peak show that (1) the epidemic peak could be reached in the period ranging from 12 July to 11 August 2020, and last until the period ranging from 22 November 2020 to 12 January 2021; (2) the identification rate, which ranges from 0.01 to 0.084, does not affect the epidemic size for the total Malaysian population; (3) the influence of the identification rate on the basic reproductive number is negligible; and (4) a near-future intervention may decrease the infection rate, which would lead to a delay the epidemic peak. The results also show that the infection rate is 0.228 ± 0.013, while the basic reproductive number is 2.28 ± 0.13. Furthermore, a high forecasting accuracy is achieved, such that the NRMSE, MAPE, and R^2 values are 0.041, 2.45%, and 0.9964, respectively.

Int. J. Environ. Res. Public Health **2020**, *17*, x

13 of 15

Appendix A

Figure A1. Block diagram of the SEIR model.

References

1. Gallego, V.; Nishiura, H.; Sah, R.; Rodriguez-Morales, A.J. The COVID-19 outbreak and implications for the Tokyo 2020 Summer Olympic Games. *Travel Med. Infect. Dis.* **2020**, 34, 101604.
2. Jiang, F.; Deng, L.; Zhang, L.; Cai, Y.; Cheung, C.W.; Xia, Z. Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). *J. Gen. Intern. Med.* **2020**, 35, 1545–1549.
3. Updates on the Coronavirus Disease 2019 (COVID-19) Situation in Malaysia. Available online: http://www.moh.gov.my/index.php/database_stores/attach_download/337/1378 (accessed on 22 March 2020).
4. Rothe, C.; Schunk, M.; Sothmann, P.; Bretzel, G.; Froeschl, G.; Wallrauch, C.; Zimmer, T.; Thiel, V.; Janke, C.; Guggemos, W. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *New Engl. J. Med.* **2020**, *382*, 970–971.
5. van den Driessche, P. Reproduction numbers of infectious disease models. *Infect. Dis. Model.* **2017**, *2*, 288–303.
6. Roosa, K.; Lee, Y.; Luo, R.; Kirpich, A.; Rothenberg, R.; Hyman, J.M.; Yan, P.; Chowell, G. Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China: February 13–23, 2020. *J. Clin. Med.* **2020**, *9*, 596.
7. Guzzetta, G.; Poletti, P.; Ajelli, M.; Trentini, F.; Marziano, V.; Cereda, D.; Tirani, M.; Diurno, G.; Bodina, A.; Barone, A. Potential short-term outcome of an uncontrolled COVID-19 epidemic in Lombardy, Italy, February to March 2020. *Eurosurveillance* **2020**, *25*, 2000293.
8. Huang, N.E.; Qiao, F. A data driven time-dependent transmission rate for tracking an epidemic: A case study of 2019-nCoV. *Sci. Bull.* **2020**, *65*, 425.
9. Peng, L.; Yang, W.; Zhang, D.; Zhuge, C.; Hong, L. Epidemic analysis of COVID-19 in China by dynamical modeling. *arXiv:2002.06563v1*, **2020**, *1*, 18.
10. Tian, S.; Hu, N.; Lou, J.; Chen, K.; Kang, X.; Xiang, Z.; Chen, H.; Wang, D.; Liu, N.; Liu, D. Characteristics of COVID-19 infection in Beijing. *J. Infect.* **2020**, *4*, 401-406.
11. Qin, L.; Sun, Q.; Wang, Y.; Wu, K.-F.; Chen, M.; Shia, B.-C.; Wu, S.-Y. Prediction of Number of Cases of 2019 Novel Coronavirus (COVID-19) Using Social Media Search Index. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2365.
12. Kuniya, T. Prediction of the Epidemic Peak of Coronavirus Disease in Japan, 2020. *J. Clin. Med.* **2020**, *9*, 789.
13. Rovetta, A.; Bhagavathula, A.S. Modelling the epidemiological trend and behavior of COVID-19 in Italy. *medRxiv* **2020**, doi:10.1101/2020.03.19.20038968.
14. Olfatifar, M.; Houri, H.; Shojaee, S.; Pourhoseingholi, M.A.; Al-Ali, W.; Luca, B.; Ashtari, S.; Shahrokh, S.; Vahedian, A.; Asadzadeh Aghdaei, H. The Required Confronting Approaches Efficacy and Time to Control Iranian COVID-19 Outbreak. *Arch. Clin. Inf. Dis*. **2020**, *15*, e102633.
15. Li, X.; Zhao, X.; Sun, Y. The lockdown of Hubei Province causing different transmission dynamics of the novel coronavirus (2019-nCoV) in Wuhan and Beijing. *medRxiv* 2020, doi10.1101/2020.02.09.20021477

16. Hu, Z.; Ge, Q.; Jin, L.; Xiong, M. Artificial intelligence forecasting of covid-19 in china. Available online: https://arxiv.org/abs/2002.07112 (accessed on 13 May)

17. Li, X.; Xu, B.; Shaman, J. The Impact of Environmental Transmission and Epidemiological Features on the Geographical Translocation of Highly Pathogenic Avian Influenza Virus. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1890.

18. Davidian, M.; Giltinan, D.M. *Nonlinear Models for Repeated Measurement Data*; CRC Press: Boca Raton, FL, USA, 1995; Volume 62.

19. Capaldi, A.; Behrend, S.; Berman, B.; Smith, J.; Wright, J.; Lloyd, A.L. Parameter estimation and uncertainty quantication for an epidemic model. *Math. Biosci. Eng.* **2012**,9 (3), 553-576.

20. Keeling, M.J.; Rohani, P. *Modeling Infectious Diseases in Humans and Animals*; Princeton University Press: Princeton, NJ, USA, 2011.

21. Smirnova, A.; deCamp, L.; Chowell, G. Forecasting epidemics through nonparametric estimation of time-dependent transmission rates using the SEIR model. *Bull. Math. Biol.* **2019**, *81*, 4343–4365.

22. Sun, H.; Qiu, Y.; Yan, H.; Huang, Y.; Zhu, Y.; Chen, S.X. Tracking and Predicting COVID-19 Epidemic in China Mainland. *medRxiv* **2020**, 17. 20.

23. Linton, N.M.; Kobayashi, T.; Yang, Y.; Hayashi, K.; Akhmetzhanov, A.R.; Jung, S.-M.; Yuan, B.; Kinoshita, R.; Nishiura, H. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *J. Clin. Med.* **2020**, *9*, 538.

24. Lauer, S.A.; Grantz, K.H.; Bi, Q.; Jones, F.K.; Zheng, Q.; Meredith, H.R.; Azman, A.S.; Reich, N.G.; Lessler, J. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **2020**, 172. 577-582

25. 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.* **2020**, 5. 271-281

26. Current Population Estimates, Malaysia, 2018–2019. Avaiavle online: https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=155&bul_id=aWJZRkJ4UEdKcUZ pT2tVT090Snpydz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09 (accessed on 20 March 2020).

27. Covid-19: Malaysia to receive new test kit from South Korea. Available online: https://www.thestar.com.my/news/nation/2020/04/05/covid-19-malaysia-to-receive-new-test-kit-from-south-korea (accessed on 12 April 2020).

28. Latest COVID-19 Statistic in Malaysia by MOH. Avialiable online: http://www.moh.gov.my/index.php/pages/view/2019-ncov-wuhan (accessed on 25 March 2020).

29. Jung, S.-M.; Akhmetzhanov, A.R.; Hayashi, K.; Linton, N.M.; Yang, Y.; Yuan, B.; Kobayashi, T.; Kinoshita, R.; Nishiura, H. Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: Inference using exported cases. *J. Clin. Med.* **2020**, *9*, 523.

30. Jones, J.H. *Notes on R0*; Department of Anthropological Sciences: Stanford, CA, USA, 2007.

31. Tuncer, N.; Gulbudak, H.; Cannataro, V.L.; Martcheva, M. Structural and practical identifiability issues of immuno-epidemiological vector–host models with application to rift valley fever. *Bull. Math. Biol.* **2016**, *78*, 1796–1827.

32. Ahmad, F.; Isa, N.A.M.; Osman, M.K.; Hussain, Z. Performance comparison of gradient descent and Genetic Algorithm based Artificial Neural Networks training. In Proceedings of the 2010 10th International Conference on Intelligent Systems Design and Applications, Cairo, Egypt, 29 November–1 December 2010; pp. 604–609.

33. Sorsa, A.; Peltokangas, R.; Leiviska, K. Real-coded genetic algorithms and nonlinear parameter identification. In Proceedings of 2008 4th International IEEE Conference Intelligent Systems, Varna, Bulgaria, 6–8 September 2008; pp. 10-42.

34. Kilinc, M.; Caicedo, J.M. Finding Plausible Optimal Solutions in Engineering Problems Using an Adaptive Genetic Algorithm. *Adv. Civ. Eng.* **2019**, *2019*, 9.

35. Mohammadi, K.; Shamshirband, S.; Kamsin, A.; Lai, P.; Mansor, Z. Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure. *Renew. Sustain. Energy Rev.* **2016**, *63*, 423–434.

36. Rezakazemi, M.; Dashti, A.; Asghari, M.; Shirazian, S. H2-selective mixed matrix membranes modeling using ANFIS, PSO-ANFIS, GA-ANFIS. *Int. J. Hydrog. Energy* **2017**, *42*, 15211–15225.

37. Yi, H.-S.; Park, S.; An, K.-G.; Kwak, K.-C. Algal bloom prediction using extreme learning machine models at artificial weirs in the Nakdong River, Korea. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2078.

Int. J. Environ. Res. Public Health **2020**, *17*, x

15 of 15

38. Mentaschi, L.; Besio, G.; Cassola, F.; Mazzino, A. Problems in RMSE-based wave model validations. *Ocean Model.* **2013**, *72*, 53–58.

39. Piepho, H.P. A coefficient of determination (R2) for generalized linear mixed models. *Biom. J.* **2019**, *61*, 860–872.

40. Coronavirus disease 2019 (COVID-19) Situation Report–46. Avaliable online: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf?sfvrsn=96b04adf_2 (accessed on 20 March 2020).

41. Zhao, S.; Lin, Q.; Ran, J.; Musa, S.S.; Yang, G.; Wang, W.; Lou, Y.; Gao, D.; Yang, L.; He, D. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *Int. J. Infect. Dis.* **2020**, *92*, 214–217.

42. Majumder, M.; Mandl, K.D. Early transmissibility assessment of a novel coronavirus in Wuhan, China. *N Engl J Med.* **2020**, 382. 1199-1207.

43. Zhang, S.; Diao, M.; Yu, W.; Pei, L.; Lin, Z.; Chen, D. Estimation of the reproductive number of Novel Coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *Int. J. Infect. Dis.* **2020**, *93*, 201–204.

44. Liu, Y.; Gayle, A.A.; Wilder-Smith, A.; Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* **2020**. 27. 1-4.

45. Liu, T.; Hu, J.; Kang, M.; Lin, L.; Zhong, H.; Xiao, J.; He, G.; Song, T.; Huang, Q.; Rong, Z. Transmission dynamics of 2019 novel coronavirus (2019-nCoV). **2020**. 1. 919787

46. Read, J.M.; Bridgen, J.R.; Cummings, D.A.; Ho, A.; Jewell, C.P. Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions. *MedRxiv* **2020**, doi:10.1101/2020.01.23.20018549 (preprint).

47. It's a 'False Hope' Coronavirus will Disappear in the Summer like the Flu. Avaliable online: https://www.msn.com/en-us/health/health-news/its-a-false-hope-coronavirus-will-disappear-in-the-summer-like-the-flu-who-says/ar-BB10QrLc (acessed on 12 March 2020).

48. Efforts to Contain Covid-19 no Longer Possible. Avaliable online: https://www.nst.com.my/news/nation/2020/03/575003/efforts-contain-covid-19-no-longer-possible-dr-lee (acessed on 16 March 2020).

49. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). Avalible online: https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf (accessed on 12 March 2020).

50. World Health Organization. *Clinical Management of Severe Acute Respiratory Infection (SARI) When COVID-19 Disease is Suspected: Interim Guidance, 13 March 2020*; World Health Organization: Geneva, Switzerland, 2020.

51. Only 60pc Complied with MCO.; Police may Take Sterner Action. Avaliable online: https://www.malaymail.com/news/malaysia/2020/03/19/ismail-sabri-four-in-10-malaysians-violating-movement-control-order/1848077/ (acessed on 19 March 2020).

52. Gilliland, M. *The Business Forecasting Deal: Exposing Myths, Eliminating Bad Practices, Providing Practical Solutions*; John Wiley & Sons: Hoboken, NJ, USA, 2010; Volume 27.