J. Math. Computer Sci., 29 (2023), 52–59

Online: ISSN 2008-949X

Journal of Mathematics and Computer Science

Journal Homepage: www.isr-publications.com/jmcs

Modelling the fourth wave of Covid-19 pandemic in Egypt

Islam A. Moneim*, Eman I. Abd El-Latif

Department of Mathematics and Computer Science, Benha University, Benha, PO 13518, Egypt.

Abstract

This paper studied the dynamics of Covid-19 in Egypt using machine learning algorithms and the epidemiological model SEIR. Among the machine learning models studied, two models showed promising results (SVR, SEGPR). Predictions of the spread of Covid-19 in the next 70 days conducted using these three approaches. The data of the fourth wave of Covid-19 in egypt taken from the WHO. The statistical measures R-Squared and RMSE used to evaluate the accuracy of these models.

Keywords: Machine learning, Covid-19, SEIR model, prediction, Egypt.

2020 MSC: 91-10, 68Q07, 00A71.

1. Introduction

The machine learning has improved tremendously over the last decade. As a result, machine learning used in a wide range of applications. One of the most recent and important application of machine learning is the study of COVID-19 pandemic. Many researchers used machine learning algorithms to study different problems related to COVID-19 [5, 7, 19]. Forecasting the confirmed cases is one of the important aims of several machine learning models [12]. For examples, k nearest neighbor (kNN), random forest, support vector regression (SVR), cubist regression, and Bayesian neural network are used in Brazil and the USA. Pereira et al., compare statistical and ML approaches to time series forecasting. They used SVR and long short-term memory (LSTM) models to predict the number of infected, deaths, and recovered individuals [18]. The authors also consider exogenous variables such as temperature and precipitation. Numerical experiments produce mixed results with no clear favorite. It can only be noted that VMD improves model performance when the prediction horizon is 6 days ahead. Suvarna et al., [22] used linear regression and SVR to predict the upcoming number of cases and death in India and worldwide. Their results found that the SVR is outgain the linear regression.

In the other hand, SEIR models used to investigate some infectious diseases with an infective latent period such as, HBV, influenza, and AIDS [8] and [23]. In this case, the disease can transmit before symptoms showed up. Influenza is the classic example of a virus that can spread when people have no symptoms at all. We can suppose that these infectious diseases have two ways of infection.

*Corresponding author

Email address: islam.ahmed@fsc.bu.edu.eg (Islam A. Moneim)

doi: 10.22436/jmcs.029.01.05

Received: 2022-01-28 *Revised:* 2022-04-25 *Accepted:* 2022-06-16

Documented reports showed that, the Covid-19 disease transmitted from contaminated persons during their incubation period. Other reports showed that, children may transmit the virus while asymptomatic [2]. The researchers found one child in an infected family had no symptoms, but a further clinical investigation showed that, he had pneumonia and his test for the virus came positive [4].

It has documented that, the reported daily data of new cases and daily deaths in the USA exhibit periodic oscillatory patterns. The analysis of the collected data for COVID-19 infectivity and death rates has revealed oscillatory pattern with a 7-day period in many countries around the world. Some of these countries showed 14-days periodicity in their observed data [16] and [17].

This paper aims to give a more realistic prediction of the dynamics of the new wave of Covid-19. Firstly, some machine learning algorithms and secondly an SEIR model are used to predict the number of new cases of the fourth wave of the new strain of coronavirus in Egypt. The SEIR model used in this paper, assumes that, the disease has two transmission functions $\beta 1(t)$ and $\beta 2(t)$, which represent the contact rate between susceptible and both of the latent and infective individuals respectively. As the reported new cases fluctuates all over the year, the transmission functions $\beta_1(t)$ and $\beta_2(t)$ are considered as periodic functions.

This paper is organized as follows: Section two illustrates the materials and methods. The results are provided in the section three. Sections four presents some conclusions and finally the references end this work.

2. Material and method

This paper is divided into two parts; the first part shows the machine learning and the second part presents the susceptible- exposed-infected-recovered (SEIR). Data are collected from world health organization (WHO) for Egypt over 38 days from 1/8/2021 to 9/9/2021. Figure 1 presents the daily reported number of active cases and Figure 2 shows the death cases in Egypt.

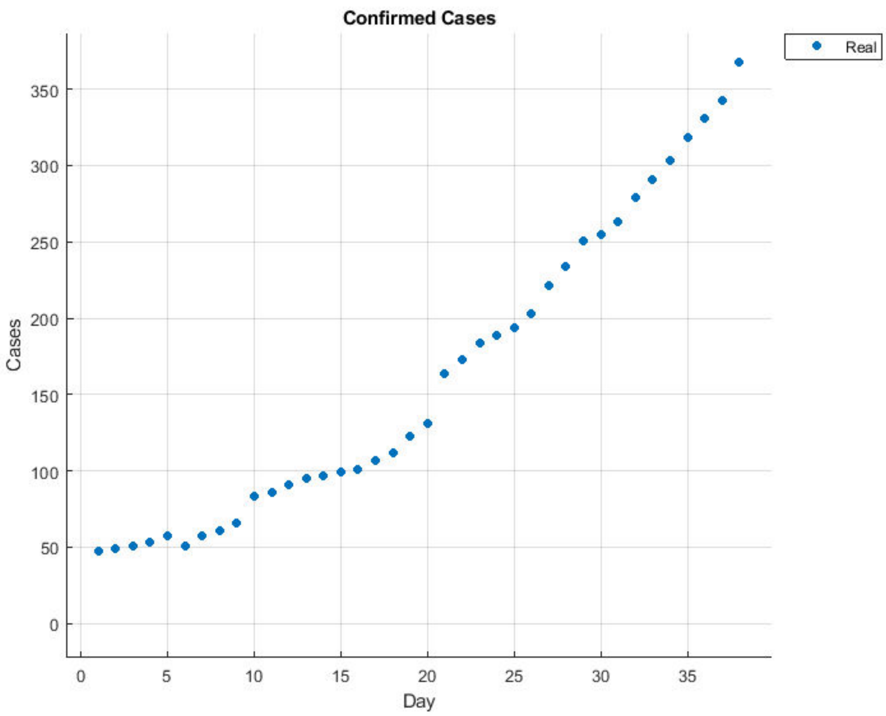

Figure 1: Number of new cases in Egypt.

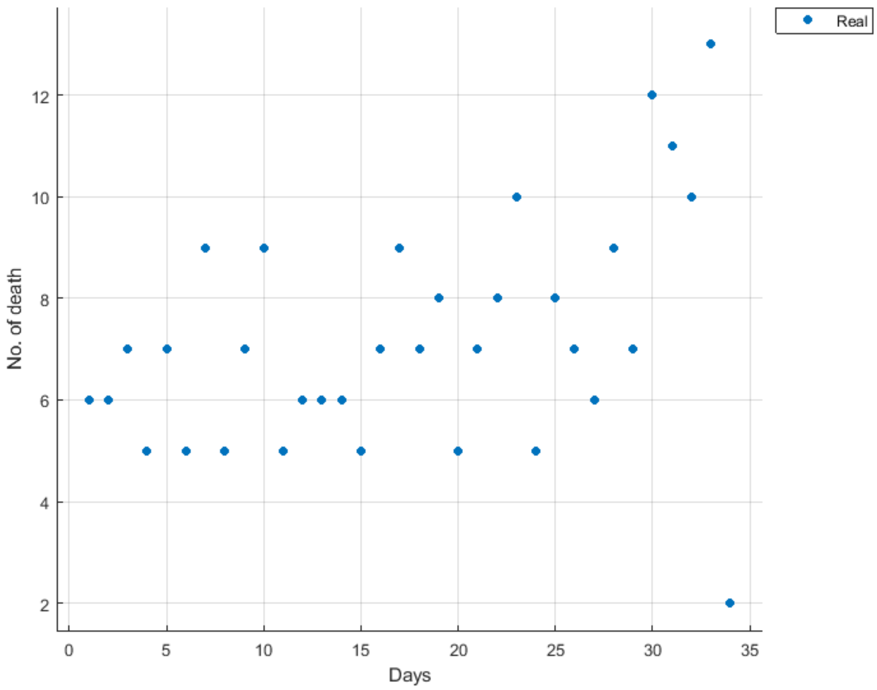

Figure 2: Number of death cases in Egypt.

2.1. *Machine learning*

Machine learning is a branch of artificial intelligence (AI) which makes reliable predictions based on data. When the amount of data available is huge, machine learning can identify patterns and automatically make predictions. Dataset is splitted into training and testing sets. Training sets are used to validate the performance of the model. There are different types of machine learning algorithms used to predict Covid-19.

Gaussian process regression (GPR) is non-parametric methods used widely in machine learning applications because it working well on small datasets and the ability to offer measurements on the predictions [6]. GPR can make prediction using the prior knowledge from training data and then compute the predictive distribution. GPR is used in analyzing data in higher dimensions, regression and statistical modeling. There are many kernel function used in GPR models such as Rational Quadratic GPR, Squared Exponential GPR (SEGPR), Matern 5/2 GPR, and Exponential GPR. SEGPR is matching to the Exponential GPR except that the Euclidean distance is squared. In this research, SEGPR is used in the training a data. Analyzing the results of the model using two evaluation criteria (R-square and RMSE). SEGPR kernel function is calculated by the following formula [24]:

$$\mathrm{SEGPR} = \sigma_f^2 \exp \frac{x^i - x^j}{2l^2},$$

where the length parameter l controls the smoothness of the function and σ_f is the vertical variation. In this research, Support vector regression (SVR) technique and squared exponential gaussian process regression (SEGPR) are used in the prediction.

Support vector regression (SVR) technique and squared exponential gaussian process regression (SEGPR) are used in the prediction. Support vector regression (SVR) is used for the prediction problems and it is based on the support vector machine (SVM) to collect the data points about the hyperplane [3] and [9]. The model generated by SVR depends on the training dataset and the SVR model is expressed by the following equation,

$$f(x) = (z.\phi(x)) + b,$$

where z is the weight vector, b is the bias value, and $\phi(x)$ is the kernel function [11]. There are different types of kernel function such as linear kernel, polynomial kernel, radial basis function (RBF), Gaussian kernel, and sigmoid kernel [21].

3. SEIR model

In this section, susceptible exposed infected removed (SEIR) model is introduced to show how a disease spreads through a population. Following [14] and [13], the SEIR model is described by the following system of nonlinear ordinary differential equations:

$$
\begin{aligned}
\frac{dS}{dt} &= \nu N(1-p) - \beta_1(t)S, E - \beta_2(t)SI - \nu S,\\
\frac{dE}{dt} &= \beta_1(t)SE + \beta_2(t)SI - (\nu+\sigma)E,\\
\frac{dI}{dt} &= \sigma E - (\nu+\gamma)I,\\
\frac{dR}{dt} &= \nu Np + \gamma I - \nu R,
\end{aligned}
$$

where S, E, I and R are the number of susceptible latent infected and recovered individuals respectively at time t. Here γ is the rate at which the infected individuals become recovered. The latent individuals become infectious by the rate σ. The transmission rates are $\beta_1(t)$, between the latent and susceptible individuals, and $\beta_2(t)$, between the infected and susceptible individuals, respectively. The number of populations is a constant N.

4. Evaluation criteria

R-squared (the coefficient of determination) and the root mean square error (RMSE) are statistical measures used in evaluation the prediction models [1]. R-squared decides the proportion of variance in the dependent variable that can be described by the independent variable. R-squared takes values between zero and one. RMSE used to measure the error of a model in predicting data. The RMSE and R^2 are given by the following equations:

$$
RMSE = \sqrt{\frac{\sum_{i=1}^{n}(t_i - p_i)^2}{n}},
$$

and

$$
R^2 = \frac{\sum_{i=1}^{n}(t_i - \overline{t_i})(p_i - \overline{p_i})}{\sqrt{\sum_{i=1}^{n}(t_i - \overline{t_i})^2(p_i - \overline{p_i})^2}},
$$

where t is the measured value, p is the predicted value, $\overline{t_i}$ is the mean of measured value, $\overline{p_i}$ is the mean of predicted value, and n is the total number of data. These statistics measures show how well the data fit the machine learning models. The next step is to find the best model for prediction the dynamics of the Covid-19 in Egypt.

5. Results

Tables 1 and 2 show the calculated values of the RMSE and r-square for each equation fitted to predicted cases using SVM and SEGPR, respectively. In SVM, the cubic equation provided the smallest RMSE and the largest r-square, but the quadratic equation provided the largest RMSE and the smallest R-squared in SEGPR.

Table 1: Accuracy statistics for predicted cases by SVR.

Model Name	Description	R-square	RMSE
Cubic	$c = 0.0002719i^3 + 0.209i^2 + 0.1139i + 54.46$	0.9997	10.93
Quadratic	$q = 0.2249i^3 - 0.1374i + 55.33$	0.997	11.03

Table 2: Accuracy statistics for predicted cases by SEGPR.

Model Name	Description	R-squared	RMSE
Cubic	$c = -0.002333i^3 + 0.3321i^2 - 1.149i + 52.93$	0.999	18.72
Quadratic	$q = 0.1956i^2 + 1.008i + 45.47$	0.9984	23.89

Accuracy statistics for death cases is shown in Table 3 and Table 4. The smallest RMSE and the largest R-squared are provided by using cubic equation fitting.

Table 3: Accuracy statistics for death cases by SVR.

Model Name	Description	R-squared	RMSE
Cubic	$c = 0.0005654i^3 - 0.01894i^2 + 0.1931i$		
	$+5.831$	0.8702	2.099
Quadratic	$q = 0.01074i^2 - 0.2285i + 7.149$	0.8346	2.369

Table 4: Accuracy statistics for death cases by SEGPR.

Model	Description	R-squared	RMSE
Cubic	$c = 5.454\,\mathrm{EXP}(-5)i^3 - 0.0005199i^2 - 0.01315i$	0.4309	2.309
Quadratic	$2 = 0.002343i^2 - 0.05382i + 7.252$	0.425	2.321

Figure 3 and Figure 4 presents the prediction for 75 days in active and death cases respectively.

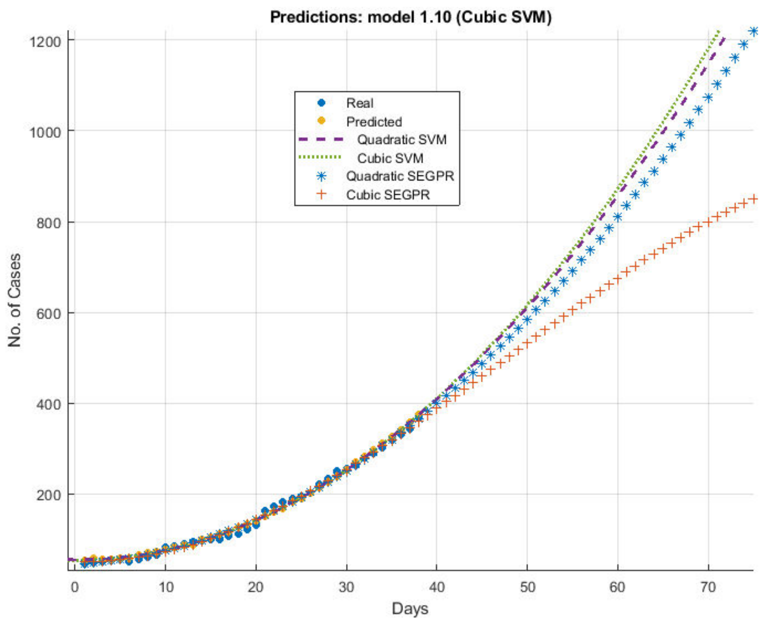

Figure 3: The prediction through 75 days for active cases.

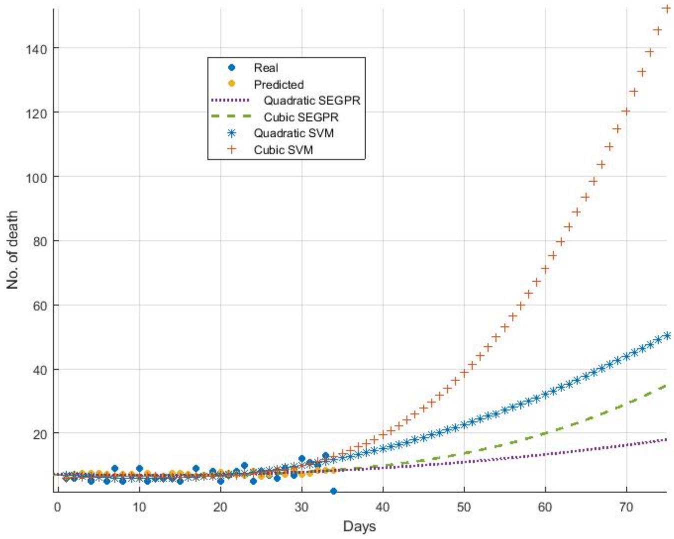

Figure 4: The prediction through 75 days for death cases.

Figure 5 present the results of the SEIR model for prediction the active cases. In this model, we take $\beta_1 = 0.00001$ and $\beta_2 = 0.00003$. We use a constant population size of $N = 100,000,000$ to simulate the spread of Covid-19 in medium population number localized cities. We also supposed that $\nu = 0.02/$ year [15] . The average incubation period varied from 5.1 days to 7.3 days, but most of the reported data show that, incubation period lasts after 14 days in more than 99% of the infected persons in [10]. Thus, we assume that the incubation period will be $\sigma^{-1} = 6$ days. As the mean value of the infectious period of Covid-19 is 3.6 days, then the average infectious period is $\gamma^{-1} = 3.6$ days in our simulations [20].

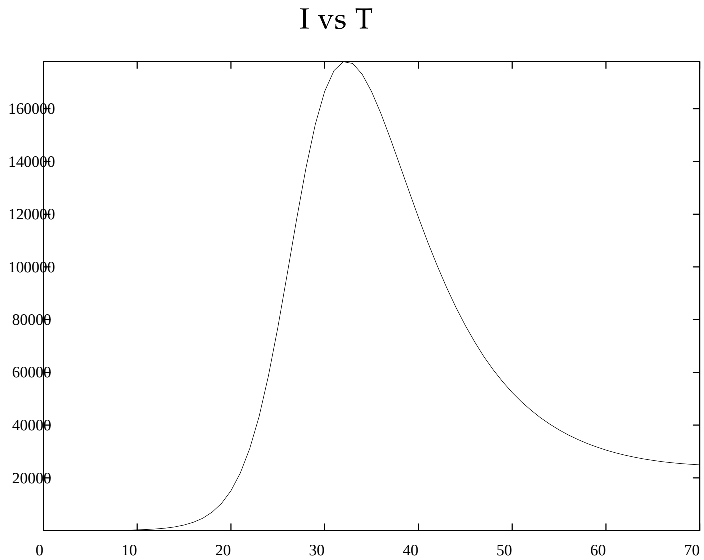

Figure 5: Prediction of confirmed cases based on the SEIR model.

Figure 5 shows that there is a stable one year periodic solution corresponding to relatively large transmission rates, $\beta_1 = 0 : 000005$ and $\beta_2 = 0 : 000003$ for Covid-19 parameter when the contact rates are the sinusoidal functional form stated above. This stable periodic solution means that the Covid-19 will hit again every year even with smaller. Figure 6 plots the predicted result of new cases from SEIR model, cubic equation and quadratic equation respectively. We note that, the cubic equation gives the best accuracy statistics in the first 30 days, as shown as in Table 5.

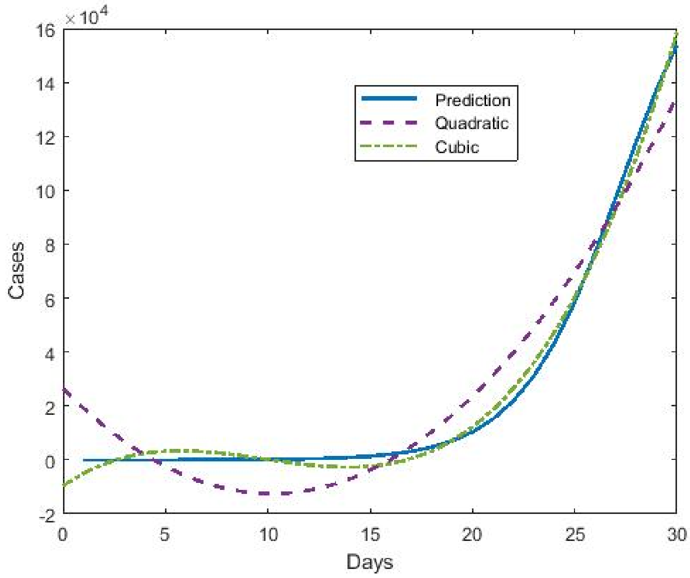

Figure 6: Fitting of predicted cases from the SEIR model.

Table 5: Accuracy statistics for active cases by SEIR.

Model	Description	R-squared	RMSE
Cubic	$c = 22.01i^3 + 648.9i^2 + 5268i - 9633$	0.9949	339.15
Quadratic	$q = 374.6i^2 - 7633i + 2641$	0.9312	1220

6. Conclusions

The parameters of two simple mathematical models, quadratic and cubic were fitted using SVR, SEGPR, and SEIR. The cubic model outdone the quadratic and it is also shown on the training for 30 days. This paper evaluated the applicability of two machine learning models, SVR and SEGPR and an SIER mathematical model for predicting the COVID-19 outbreak.

The progression of the COVID-19 outbreak in Egypt is predicted using three different techniques. These techniques are the SVM, SEGPR and the SEIR epidemiological model. Machine learning models give an alternative to the epidemiological models and showed more accurate and potential results in predicting the dynamics of Covid-19. In the SEIR model, the number of cases reported by WHO for the Egypt situation, is the number of people who are tested, but not the actual number of latent E (which are infectious also). Therefore, the number of infectious people I is not easy to be calculated, as many people may be infectious but not get tested, especially if they showed a mild symptom. Due to the high level of ambiguity and uncertain data, the standard SEIR epidemiological model showed low accuracy for long-term prediction. As shown in Figure 6 , the SEIR represent a fair matching to the real data plotted in Figures 1 within the first 30 days. Unfortunately, Figures 1 and 5 show that the SEIR model failed to give good predictions after the first 30 days. On the other hand, the statistics measures showed that, the machine learning models give a very good prediction for 70 days from the training time.

The results of two ML models (SVR and SEGPR) described a high accuracy for the long-term prediction. With respect to the results obtained our study recommends ML as an effective technique to predict the outbreak of Covid-19.

References

[1] S. F. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A. R. Varkonyi-Koczy, U. Reuter, T. Rabczuk, P. M. Atkinson, *Covid-19 outbreak prediction with machine learning*, Algorithms, **13** (2020), 14 pages. 4

[2] J. Cai, W. J. Sun, J. P. Huang, M. Gamber, J. Wu, G. Q. He, *Indirect Virus Transmission in Cluster of Covid-19*, Emerging Infectious Diseases, **26** (2020), 1343–1345. 1

[3] J. Cervantes, F. Garcia-Lamont, L. Rodriguez-Mazahua, A. Lopez, *A comprehensive survey on support vector machine classification: Applications, challenges and trends*, Neurocomputing, **408** (2020), 189–215. 2.1

[4] J. F. W. Chan, S. Yuan, K. H. Kok, K. K. W. To, H. Chu, J. Yang, F. F. Xing, J. Liu, C.-Y. Yip, R. W.-S. Poon, H.-W. Tsoi, S. K.-F. Lo, K.-Hu. Chan, V. K.-M. Poon, W.-M. Chan, J. Daniel, J.-P. Cai, V. C.-C. Cheng, K.-Y. Yuen, *A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-toperson transmission: a study of a family cluster*, The Lancet, **6736** (2020), 514–523. 1

[5] N. El-Rashidy, S. Abdelrazik, T. Abuhmed, E. Amer, F. Ali, J.-W. Hu, S. El-Sappagh, *Comprehensive survey of using machine learning in the COVID-19 pandemic*, Diagnostics, **11** (2021), 11 pages. 1

[6] S. Hong, Z. Zhou, *Application of Gaussian process regression for bearing degradation assessment*, 6th International Conference on New Trends in Information Science, Service Science and Data Mining (ISSDM2012), **2012** (2012), 644–648. 2.1

[7] S. Kumar, H. Kumar, V. Dutt, P. Dixit, *The role of machine learning in COVID-19 in medical domain: A survey*, J. Recent Inn. Cloud Comput. Virtual. Web Appl., **4** (2020), 13 pages. 1

[8] A. A. Lashari, S. Aly, K. Hattaf, G. Zaman, H. Jung, X.-Z. Li, *Presentation of malaria epidemics using multiple optimal controls*, J. Appl. Math., **2012** (2012), 9 pages. 1

[9] T. P. Latchoumi, L. Parthiban, *Advanced overlap community detection by associative rule mining and multi-view ant clustering*, Int. J. Eng. Tech., **7** (2018), 21–29. 2.1

[10] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, J. Lessler, *The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application*, Ann. Intern. Med., **172** (2020), 577–582. 5

[11] C. J. Lu, C. H. Chang, C. Y. Chen, C. C. Chiu, T.-S. Lee, *Stock index prediction: a comparison of MARS,* BPN *and SVR in an emerging market*, In Proceeding of the IEEE International Conference on Industrial Engineering and Engineering Management, **2009** (2009), 2343–2347. 2.1

[12] Y. Mohamadou, A. Halidou, P. T. Kapen, *A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19*, Applied Intelligence, **50** (2020), 3913–3925. 1

[13] I. A. Moneim, *Seasonally varying epidemics with and without latent period: A comparative Simulation Study*, Math. Med. Biol., **24** (2007), 1–15. 3

[14] I. A. Moneim, *An SEIR Model with infectious latent and a periodic vaccination strategy*, Math. Model. Anal., **26** (2021), 236–252. 3

[15] I. A. Moneim, D. Greenhalgh, *Use of a periodic vaccination strategy to control the spread of epidemics with seasonally varying contact rate*, Math. Biosci. Eng., **13** (2005), 591–611. 5

[16] T. Odagaki, *Self-organized wavy infection curve of COVID-19*, Scientific Reports, **11** (2021), 1–7. 1

[17] T. Pavlicek, P. Rehak, P. Kral, *Oscillatory dynamics in infectivity and death rates of COVID-19*, Msystems, **5** (2020), 7–20. 1

[18] I. Pereira, J. M. Guerin, A. G. S. Junior, G. S. Garcia, P. Piscitelli, A. Miani, C. Distante, L. M. G. Goncalves, *Forecasting Covid-19 dynamics in Brazil: A data driven approach*, Int. J. Environ. Res. Public Health, **17** (2020), 8 pages. 1

[19] I. Rahimi, F. Chen, A. H. Gandomi, *A review on COVID-19 forecasting models*, Neural Comput. Appl., **2021** (2021), 11 pages. 1

[20] J. M. Read, J. R. E. Bridgen, D. A. T. Cummings, A. Ho, C. P. Jewell, *Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions*, medRxiv, **2020** (2020), 12 pages. 5

[21] M. Singla, D. Ghosh, K. K. Shukla, *A survey of robust optimization-based machine learning with special reference to support vector machines*, Int. J. Machine Learning Cyber., **11** (2020), 1359–1385. 2.1

[22] B. Suvarna, B. L. Nandipati, M. N. Bhat, *Support vector regression for predicting covid-19 cases*, Eur. J. Molecular Clinic. Med., **7** (2020), 4882–4893. 1

[23] G. Zaman, S. H. Saker, *Dynamics and control of a system of two noninteracting preys with common predator*, Math. Methods Appl. Sci., **34** (2011), 2259–2273. 1

[24] N. Zhang, J. Xiong, J. Zhong, K. Leatham, *Gaussian process regression method for classification for high-dimensional data with limited samples*, Eighth International Conference on Information Science and Technology (ICIST), **2018** (2018), 358–363. 2.1