

COVID-19 Confirmed Cases Prediction in China Based on Barnacles Mating Optimizer-Least Squares Support Vector Machines

Zuriani Mustaffa¹, Mohd Herwan Sulaiman²

¹Faculty of Computing, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

²Faculty of Electrical & Electronics Engineering Technology, University Malaysia Pahang, Pekan, Malaysia

E-mails: zuriani@ump.edu.my herwan@ump.edu.my

Abstract: The Covid19 has significantly changed the global landscape in every aspect including economy, social life, and many others. After almost two years of living with the pandemic, new challenges are faced by the research community. It may take some time before the world can be declared as totally safe from the virus. Therefore, prediction of Covid19 confirmed cases is vital for the sake of proper prevention and precaution steps. In this study, a hybrid Barnacles Mating Optimizer with Least Square Support Vector Machines (BMO-LSSVM) is proposed for prediction of Covid19 confirmed cases. The employed data are the Covid19 cases in China which are defined in daily periodicity. The BMO was utilized to obtain optimal values of LSSVM hyper-parameters. Later, with the optimized values of the hyper-parameters, the prediction task will be executed by LSSVM. Through the experiments, the study recommends the superiority of BMO-LSSVM over the other identified hybrid algorithms.

Keywords: Barnacles mating optimizer, time series prediction, optimization, machine learning, Meta-Heuristic.

1. Introduction

The year of 2020 has begun with unexpected situation when the Coronavirus hit the citizens of Wuhan, China during end of 2019. Later, the virus which is known as Covid19 spread in no time throughout the whole world. Declared as a pandemic by World Health Organizations (WHO) [1], the virus has caused millions of confirmed cases globally and significantly impacted the healthcare systems. Many countries took drastic action to announce the lockdown from weeks to months, to avoid the spread of the virus. This eventually has given a significant impact to the global economy, education, social and many other sectors. Many companies have chosen to discontinue their operation, the students having trouble to adapt with online learning and even some were left out due to family problem. Not to mention the difficulty that have been going through by the labourers and immigrants. After almost two years,

recent research shows that the virus is now spreading faster than previously reported with the emergence of variants of the virus viz. Delta and Lambda. Many countries are facing with another Covid19 wave, which includes Canada as number of cases spike across many of the countries and makes them now to be in 4th wave. Previously India had a huge strike, which caused the healthcare system to be understaffed. Meanwhile Malaysia is currently going through the hard time when the number of daily cases started to increase in June and hit more than 20000 daily in August.

On the other hand, many countries in the world are now striving to achieve the herd immunity by providing intensive vaccination to the community where the priority is given to the frontliners and elderly. Nonetheless, it is still taking some time before the world can be declared as totally safe. Therefore, proper prevention and precaution steps are important to avoid the sudden increase of such cases. Fig. 1 shows the graph for COVID-19 confirmed cases from 17 March-22 April 2020 while Fig. 2 visualizes the share of people who are fully and partly vaccinated against Covid19.

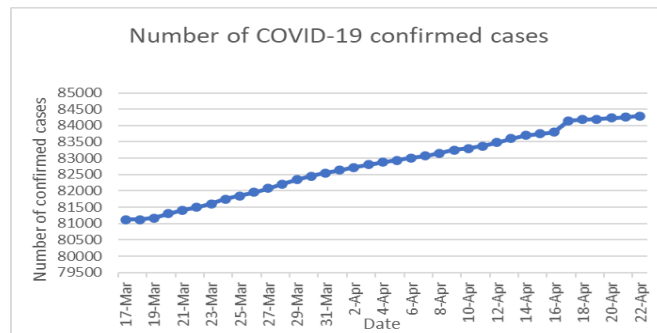


Fig. 1. COVID-19 confirmed cases from 17 March-22 April 2020

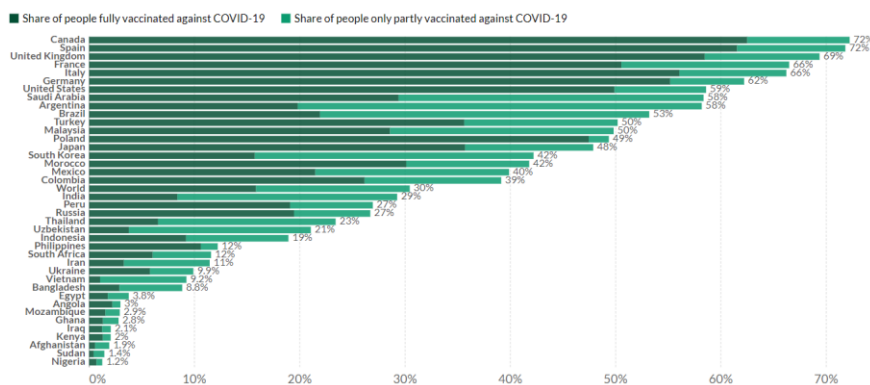


Fig. 2. Share of people vaccinated against COVID-19, August 10, 2021 [2]

Concerning that matter, several safety measures have been emphasized to the public, which include the practice of facemask, social distancing and travelling restriction. Besides that, prediction is one of the actions that can be considered to avert the increase the number of cases. Since years ago, meta-heuristic based prediction models have been favourable among academic community, not only due

to its efficiency [3] but the simplicity offered as well [4]. The application of meta-heuristic based prediction models can be seen widely applied across fields such as in agricultural [5], cloud computing [6], renewable energy [7, 8], tunnelling technology [9], economy [10], energy operation and management [11] and many more. The application of meta-heuristic based methods for prediction also can be applied in various horizon; whether short, medium or long term. Started from the well known Genetic Algorithm (GA) [12], the meta-heuristic later has grown rapidly, and still are progressively developed until today. Basically, meta-heuristic can be classified into four main groups, based on how they are developed. They are Evolutionary Computation (EC), Swarm Intelligence (SI), Physical phenomenon-based algorithm, and human-based algorithm.

In brief, Evolutionary Computation (EC) simulates biological evolution, which includes several processes such as reproduction, mutation, recombination, and selection. Other than GA, Differential Evolution (DE) [13] and Evolutionary Programming (EP) [14] to name a few, are also categorized as Evolutionary Computation. Moving on to the next group, which is Swarm Intelligence algorithm, this algorithm becomes a phenomenon by its uniqueness in mimicking the intelligence behaviour of numerous animals or insects in nature. None of the other Particle Swarm Optimization (PSO) [15] which is among the first SI based algorithm mimics the motion of bird flock or fish school in nature. Besides PSO, this group also consists of Ant Colony Optimization (ACO) [16] which is inspired by the intelligent nature of ants in discovering shortest path to food source, Artificial Bee Colony (ABC) [17] imitates the unique foraging behaviour of honey bees, Grey Wolf Optimizer (GWO) [18] replicates the hunting mechanism of grey wolf (*Canis Lupus*), Moth-Flame Optimizer (MFO) [4] influenced by the flying mechanism of moths in nature to mention a few. These outstanding algorithms undoubtedly dominate the field of nature-inspired **meta-heuristics due to its capability in producing good performance in various fields.**

Meanwhile, Gravitational Search Algorithm (GSA) which is motivated from law of gravity [19], Big Bang Big Crunch Algorithm (BB-BC Algorithm) [20] which is developed based on theories of the evolution of the universe and Black Hole Algorithm (BH Algorithm) [21] which is inspired from black hole phenomenon are several examples of physical phenomenon-based algorithms. Basically, the algorithms that fall under this roof is oriented by various nature and physic law. Finally, Human-based Algorithm incorporated of meta-heuristic algorithms which are motivated by how human characterizes and adapts in life such as in learning and socializing. Example of methods that are categorized under this group are Harmony Search Algorithm (HSA) [22] that is enlightened by the symphonic process of searching for an ideal state of harmony, Volleyball Premier League Algorithm (VPLA) [23] has emerged from the interaction among players in volleyball league and many more.

In this study, a relatively new EC algorithm namely Barnacles Mating Optimizer (BMO) [24] is presented for time series prediction of Covid19 confirmed cases. In literature, a good number of prediction for Covid19 cases have been presented, utilizing various approaches which include a notable machine learning algorithm

namely Support Vector Regression (SVR) for prediction of Covid19 cases in India [25] and also Artificial Neural Network (ANN) for Covid19 prediction in Brazil, Portugal and United States [26]. Meanwhile in [27], a few statistical prediction model have been employed which are based on Exponential Smoothing, ARIMA, SARIMA and SARIMAX. Similar approaches also can be seen in [28]. Nonetheless, the application of statistical techniques in prediction is unsuitable due to the linear structure of the techniques which makes them incapable to capture the nonlinear patterns in the time series data of interest. Even though various methods have been proposed for this issue, however, this issue is still wide open for enhancements and finding better methods is essential due to the critical state of the pandemic.

Enlightened from the unique behaviour of barnacles in mating, the BMO is chosen due to its competitiveness to the other existing meta-heuristic algorithms which includes GA, PSO, MFO and a few others. For the sake of improving the prediction model, rather than applying BMO as an individual, the BMO is hybrid with Least Squares Support Vector Machines (LSSVM). LSSVM has been recognized a powerful machine learning algorithm in solving various tasks which include prediction. By hybridizing the BMO with LSSVM, the prediction model could be avoided from having many optimization parameters which indirectly will reduce the complexity of the prediction model, instead, the BMO is only responsible to automatically optimize two hyper-parameters of interest in LSSVM namely kernel parameter, σ^2 and regularization parameter, γ . Having the optimized values of these two parameters is vital in obtaining good generalization in prediction task. For simulation purposes, the proposed BMO-LSSVM is realized on daily frequency of Covid19 confirmed cases in China. Later, the performance of BMO-LSSVM has been compared against three comparable LSSVM hybrid models namely GWO-LSSVM, MFO-LSSVM and hybrid LSSVM with Cuckoo Search (CS-LSSVM) as well.

The aim if this study is to develop a Covid19 confirmed cases prediction model. In order to achieve the goal, the study is further divided into two objectives as follow:

- i. To optimize the LSSVM hyper-parameters using BMO Algorithm.
- ii. To construct a hybrid BMO-LSSVM that will be able to produce good generalization in prediction and escape from premature convergence.

The remaining of this paper is structured as follow: Section 2 and 3 provide a brief review and mathematical model of BMO and LSSVM, respectively. This is followed by the implemented methodology in Section 4. Section 5 discusses the obtain result while the analysis of the results is provided in Section 6. Finally, Section 7 concludes the study.

2. Barnacles mating optimizer

This section provides an insight of barnacles in nature, followed by the mathematical model of BMO and prediction.

2.1. Barnacles in nature

In nature, barnacles live in the ocean or watery area and are governed by their own uniqueness. Commonly, barnacles are found permanently attached to hard object

such as ships, corals or rocks. The uniqueness of barnacles relies on its penis length where it is probably among the longest among microorganism, relative to its body size. With such feature, barnacles are able to perform mating activity to overcome their immobile condition. Taking the advantage of the distinctive feature possessed by the barnacles, the BMO is formulated as follow:

2.2. Mathematical model of BMO

The following section provides the formulation of BMO [24].

2.2.1. Initialization

Initially, the candidate solutions are represented as in the next equation. This stage involved the generation of candidate solutions which are randomly generated. An evaluation will be executed to get the initial performance of candidates, followed by the sorting phase:

$$(1) \quad X = \begin{bmatrix} x_1^1 & \dots & x_1^N \\ \dots & \dots & \dots \\ x_n^1 & \dots & x_n^N \end{bmatrix},$$

where N indicates the dimension of control variables to be optimized while n represents the population of barnacles which can be considered as the candidate for the solution. The control variables in (1) is subject to the boundaries of the problem at hands, as defined in the next two equations, respectively:

$$(2) \quad \text{ub} = [\text{ub}_1, \dots, \text{ub}_i],$$

$$(3) \quad \text{lb} = [\text{lb}_1, \dots, \text{lb}_i],$$

where **ub** is the **upper bound** and **lb** is **lower bound** of i -th variables.

2.2.2. Selection

The selection process for mating to produce new offsprings is performed randomly. However, it is subject to the parameter that need to be tuned namely pl which represents the range of the barnacles can be mated. Barnacles are classified as hermaphroditic organism which can contribute and receive sperm from others [29]. Nonetheless, for the simplicity in the formulation of the BMO algorithm, it is assumed that at one particular time each barnacle can be fertilized by one barnacle only. In nature, the self-mating process can be occurred, but is is very rare as mentioned in [30]. Therefore, it will not be considered in the mathematical model of BMO.

2.2.3. Reproduction

In BMO, the reproduction of new offspring is based on the Hardy-Weinberg principle according to which in the new offspring created the features or characters from the parents are inherited, as

$$(4) \quad x_i^{N_new} = px_{\text{barnacles_d}}^N + qx_{\text{barnacle_m}}^N,$$

where p indicates the normally distributed pseudo random numbers between the range of $[0, 1]$, $q = 1 - p$, $x_{\text{barnacles_d}}^N$ and $x_{\text{barnacle_m}}^N$ represent the variable of Dad and Mum of barnacles, respectively. It can be said that p and q denote the percentage of characteristic of Dad and Mum that are embedded in the generation of new offspring. Thus, the offspring inherits the features of Dad and Mum based on probability of random number between 0 to 1. Example, let say p if 0.55 (randomly generated), it means that the new offspring inherits 55% of the Dad's feature and 45% from the Mum's feature. It also can be noted that the (4) can be treated as exploitation process of the BMO in obtaining new solution.

In BMO, the exploitation and exploration processes are highly dependent on the value of pl . If the selection of barnacles to be mated is within the range of the Dad's barnacles to be mated with Mum, the exploitation is occurred, i.e., (4) is applied. As described in Section 2.2.2, sperm cast occurs when the selection of barnacles to be mated exceeds the value of pl that has been set initially. The sperm cast process is formulated as

$$(5) \quad x_i^{n-\text{new}} = \text{rand}() \times x_{\text{barnacle_m}}^N,$$

where $\text{rand}()$ denotes the random number generated between the range of $[0, 1]$. The pseudo code of BMO is given in Fig. 3.

```

1. Initialize the population of barnacles  $X_i$ 
2. Calculate the fitness of each barnacle
3. Sorting to locate the best result at the top of the population ( $T$ =the best solution)
4. while ( $I <$  Maximum iterations)
5.     Set the value of  $pl$ 
6.     Selection process
7.     if selection of Dad and Mum= $pl$ 
8.         for each variable
9.             Off spring generation using (4)
10.        end for
11.    else if selection of Dad and Mum> $pl$ 
12.        for each variable
13.            Off spring generation using equation (5)
14.        end for
15.    end if
16.    Bring the current barnacle back if it is out of boundaries
17.    Calculate the fitness of each barnacles
18.    Sorting and update  $T$  if it there is a better solution
19.     $I = I + 1$ 
20. end while
21. Return  $T$ 

```

Fig. 3. Pseudo code of BMO

3. Least squares support vector machines

The LSSVM is a variant of standard Support Vector Machines (SVM) which offers better solution strategy, the formulation of LSSVM for regression task is as follow.

Formally, having a training set of N points $\{x_i, y_i\}, i=1, \dots, N$, where x_i indicates the input values while y_i represents the output values, the objective is to estimate a model based on [31]:

$$(6) \quad y(x) = w^T \varphi(x_i) + b + e_i.$$

In (6), w denotes the weight vector, $\varphi(x_i) \in \mathbb{R}^n$ is the non-linear function which maps the input space into a higher dimensional feature space, b represents the biasness while e_i serves as the error between the actual and predicted output at the i -th training data. Meanwhile, x_i and $y(x)$ designate input and output, respectively. The coefficient vector w and bias term b can be calculated as defined in [31]:

$$(7) \quad \min_{w, b, e} J(w, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^N e_i^2,$$

subject to the equality constraints

$$y_i = w^T \varphi(x_i) + b + e_i, \quad i = 1, 2, \dots, N,$$

where γ is the regularization parameter which is important in balancing the complexity of the LSSVM prediction model.

By applying the Lagrangian multiplier to (7) produces [31]:

$$(8) \quad L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^N \alpha_i \{w^T \varphi(x_i) + b + e_i y_i\},$$

where α_i are Lagrange multipliers. Differentiating (8) with w , b , e_i and α_i , the conditions for optimality is as express in the equations (for $i = 1, 2, \dots, N$) [31]:

$$(9) \quad \begin{aligned} \frac{\partial L}{\partial w} = 0 &\rightarrow w = \sum_{i=1}^N \alpha_i \varphi(x_i), \\ \frac{\partial L}{\partial b} = 0 &\rightarrow \sum_{i=1}^N \alpha_i = 0, \\ \frac{\partial L}{\partial e_i} = 0 &\rightarrow \alpha_i = \gamma e_i, \\ \frac{\partial L}{\partial \alpha_i} = 0 &\rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0, \end{aligned}$$

The following linear equations are obtained by eliminating the w and e_i [31]:

$$(10) \quad \begin{bmatrix} 0 & y^T \\ y & \Omega + I / \gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1_v \end{bmatrix},$$

with $y = [y_1, \dots, y_N]$, $\alpha = [\alpha_1, \dots, \alpha_N]$, I is the identity matrix, and $1_v = [1, \dots, 1]$. The kernel trick is applied as follows [31]:

$$(11) \quad \Omega_{il} = \varphi(x_i)^T \varphi(x_l) = K(x_i, x_l), \quad i, l = 1, \dots, N.$$

The resulting of LSSVM model for regression in (1) becomes to [31]:

$$(12) \quad y(x) = \sum_{i=1}^N a_i K(x, x_i) + b,$$

where a and b are the solutions of (10). Considering the non-linear features of the data set employed, in this study, Radial Basis Function (RBF) kernel is used as it is suitable to handle such data and is able to give good performance in many prediction cases. It is expressed as [31]:

$$(13) \quad K(x, x_i) = e^{-\frac{\|x-x_i\|^2}{2\sigma^2}},$$

where σ^2 is one of the tuning parameter which is related to RBF kernel. The other adjustable parameter of RBF kernel is regularization parameter γ which can be seen in (7).

4. Methodology

This section describes the implemented methodology for this study. It consists of data collection, experiment setup, the structure of BMO-LSSVM and performance evaluation.

4.1. Dataset

For simulation purposes, this study employs historical COVID-19 confirmed cases dataset of China, which is collected from [32]. The dataset which is defined in daily periodicity covers these, reported in Chinese provinces, regions and cities which includes Hubei Province, Guandong Province, Beijing Municipality, and others. In this study, the utilized dataset ranges from 25 January 2020 to 22 April 2020. Sample of dataset is as tabulated in Table 1.

Table 1. Sample of COVID-19 confirmed cases

Date	Number of confirmed cases
25 January 2020	1297
26 January 2020	1985
27 January 2020	2761
28 January 2020	4537
29 January 2020	5997
30 January 2020	7736
31 January 2020	9720
1 February 2020	11821
2 February 2020	14411
3 February 2020	17238

4.2. Experiment setup

This section describes the experiment setup for this study which incorporates input and output, data division for training, validation and testing phases, BMO-LSSVM prediction model and finally performance evaluation criteria.

4.2.1. Input and output

For prediction modelling, in this study, the input utilizes is the number of COVID-19 cases from day 1 to day 6 while the output is to predict the number of confirmed cases from day 7 and onwards. The dataset is divided into training, validation and testing

with a 70-15-15% split. The training set has been utilized to train the prediction model, the validation set for LSSVM hyper-parameters tuning and model selection, while the test set has been employed for evaluating the performance of the final outputs.

4.2.2. BMO-LSSVM

In this study, the BMO is employed to tune the LSSVM hyper-parameters namely γ and σ^2 . Later, the optimized values projected by the BMO will be utilized by LSSVM for prediction of Covid19 confirmed cases. It is worth noting that the capability of LSSVM in prediction is highly dependent on the values of the hyper-parameters. Therefore, it is vital for BMO to be able to produce the optimal values. In this study, the objective function is guided by Root Mean Square Percentage Error (RMSPE). The definition of RMSPE is given Section 4.2.4. The aim of the hybridization is to obtain optimal values for both hyper-parameters of interest that will produce minimum prediction error. The pseudo code of hybrid BMO-LSSVM is given in Fig. 4. By referring to the figure, the LSSVM process is embedded in line 2 and 17, where the evaluation of fitness function is performed based on training and validation by LSSVM.

```

1. Initialize the population of barnacles  $X_i$ 
2. Calculate the fitness of the new solution based on LSSVM's training and validation
3. Sorting to locate the best result at the top of the population ( $T$ =the best solution)
4. while ( $I <$  Maximum iterations)
5.     Set the value of  $pl$ 
6.     Selection process
7.     if selection of  $Dad$  and  $Mum=pl$ 
8.         for each variable
9.             Off spring generation using (4)
10.        end for
11.    else if selection of  $Dad$  and  $Mum>pl$ 
12.        for each variable
13.            Off spring generation using equation (5)
14.        end for
15.    end if
16. Bring the current barnacle back if it is out of boundaries
17. Calculate the fitness of the new solution based on LSSVM's training and validation
18. Sorting and update  $T$  if it there is a better solution
19.  $I = I+1$ 
20. end while
21. Return  $T$ 

```

Fig. 4. Pseudo code of BMO-LSSVM

4.2.3. Properties setting

Prior to running the experimental process, the properties of BMO-LSSVM, CS-LSSVM, GWO-LSSVM and MFO-LSSVM are set, as tabulated in Table 2. The properties involved are size of population, maximum number of iterations, upper

bound and lower bound for LSSVM hyper parameters and pl for BMO-LSSVM. All the properties values are set based on trial-and-error approach. The maximum iteration is set to 100 since the results can converge within 100 iterations. Meanwhile, for population size, the number is adequate to produce the desired results.

Table 2. Properties of prediction techniques utilized

Property	BMO-LSSVM	CS-LSSVM	GWO-LSSVM	MFO-LSSVM
Population size	20	20	20	20
Maximum Iteration	100	100	100	100
pl	13	-	-	-
Upper bound	1000	1000	1000	1000
Lower bound	1	1	1	1

4.2.4. Performance evaluation criteria

In evaluating the time series prediction models, in this study, the prediction performance of BMO-LSSVM and the rest of identified algorithms are measured based on four metrics, namely Root Mean Square Percentage Error (RMSPE), Mean Absolute Percentage Error (MAPE) Median Absolute Percentage Error (MdAPE) and Theil's U. These metrics are defined as in the next four equations, respectively:

$$(14) \quad \text{RMSPE} = \sqrt{\frac{\sum_{n=1}^N \left(\frac{y_n - y(x_n)}{y_n} \right)^2}{N}},$$

$$(15) \quad \text{MAPE} = \frac{1}{N} \left[\sum_{n=1}^N \left| \frac{y_n - y(x_n)}{y_n} \right| \right],$$

$$(16) \quad \text{MdAPE} = \text{median} \left| \frac{1}{N} \left[\frac{y_n - y(x_n)}{y_n} \right] \times 100 \right|,$$

$$(17) \quad \text{Theil'sU} = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N (y_n - y(x_n))^2}}{\sqrt{\frac{1}{N} \sum_{n=1}^N (y_n)^2 + \frac{1}{N} \sum_{n=1}^N (y(x_n))^2}}.$$

Here: $n = 1, 2, \dots, N$; y_n = actual values; $y(x_n)$ = predicted values; N = number of test data. The stated metrics imply that the better performance of prediction models, the metrics would be smaller.

5. Results

This section describes the obtained results upon completing the series of experiments. Table 3 shows the produced results by BMO-LSSVM and the other three identified hybrid algorithms viz. CS-LSSVM, GWO-LSSVM and MFO-LSSVM. From the table, it is highlighted that BMO-LSSVM outperforms the other identified prediction models by producing lowest RMSPE which is 0.0031 when the values of γ and σ^2 are set to 313.769 and 1.7019, respectively. This is followed by CS-LSSVM and MFO-LSSVM which recorded similar RMSPE with 0.0037. Finally, GWO-LSSVM recorded bigger error with 0.0043 of RMSPE. The hyper-parameters values recorded by GWO-LSSVM are $\gamma=243.785$ and $\sigma^2= 5.3622$. Other than RMSPE, the other

metrics namely MAPE, MdAPE and Theil'sU produced by BMO-LSSVM also showed smaller error compared to the other hybrid algorithms. This shows a good agreement between the employed metrics.

Table 3. COVID19 confirmed cases prediction: BMO-LSSVM vs GWO-LSSVM vs MFO-LSSVM vs CS-LSSVM

Algorithm	γ	σ^2	RMSPE (%)	MAPE (%)	MdAPE (%)	Theil'sU
BMO-LSSVM	313.769	1.7019	0.0031	0.282	0.266	0.0015
CS-LSSVM	423.874	6.3026	0.0037	0.349	0.334	0.0018
GWO-LSSVM	243.785	5.3622	0.0043	0.416	0.3983	0.0022
MFO-LSSVM	426.285	6.3312	0.0037	0.353	0.3372	0.0019

The statistical result in graphical presentation is depicted in Fig. 5. From the figure, it is illustrated that the BMO-LSSVM (denoted by straight-crossed line) is able to produce closer prediction values to the target values (blue solid line) in most days during testing phase, compared to the other identified algorithms. During the sudden spike of cases which is day 78, even the prediction value of BMO-LSSVM has been also incapable to produce closer prediction value to the target, nonetheless, the BMO-LSSVM prediction models is still able to portray the pattern compared to the other identified algorithms.

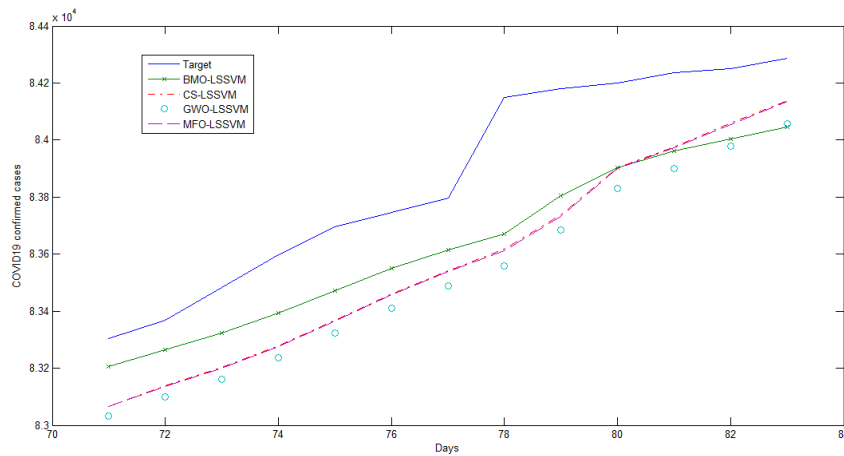


Fig. 5. COVID19 confirmed cases prediction: Comparison between actual values and the identified algorithms

Meanwhile, Table 4 shows the prediction values produced by the prediction models, compared to actual values, from day 77 to 89. The bold font indicates that the prediction values produced by the prediction model is closer to the target values on the respective day. The superiority of BMO-LSSVM is demonstrated when the model is able to produce closer prediction values compared to the other identified models, in most of the days.

Table 4. COVID19 confirmed cases prediction: Target vs BMO-LSSVM vs GWO-LSSVM vs MFO-LSSVM vs CS-LSSVM

Day	Target	BMO-LSSVM	CS-LSSVM	GWO-LSSVM	MFO-LSSVM
77	83305	83205	83067	83032	83065
78	83369	83266	83138	83099	83136
79	83482	83323	83203	83162	83201
80	83597	83393	83278	83236	83275
81	83696	83473	83369	83323	83366
82	83745	83551	83460	83410	83457
83	83797	83614	83543	83488	83540
84	84149	83670	83617	83559	83613
85	84180	83805	83737	83684	83733
86	84201	83904	83903	83829	83899
87	84237	83961	83976	83901	83972
88	84250	84005	84059	83980	84055
89	84287	84047	84139	84056	84135

The convergence performance BMO-LSSVM, CS-LSSVM, GWO-LSSVM and MFO-LSSVM is visualized in Fig. 6. From the figure, it is shown that the BMO-LSSVM converges at a slightly higher value, which is 0.0388, while the convergence value recorded by GWO-LSSVM is 0.0062 lower than the one produced by BMO-LSSVM.

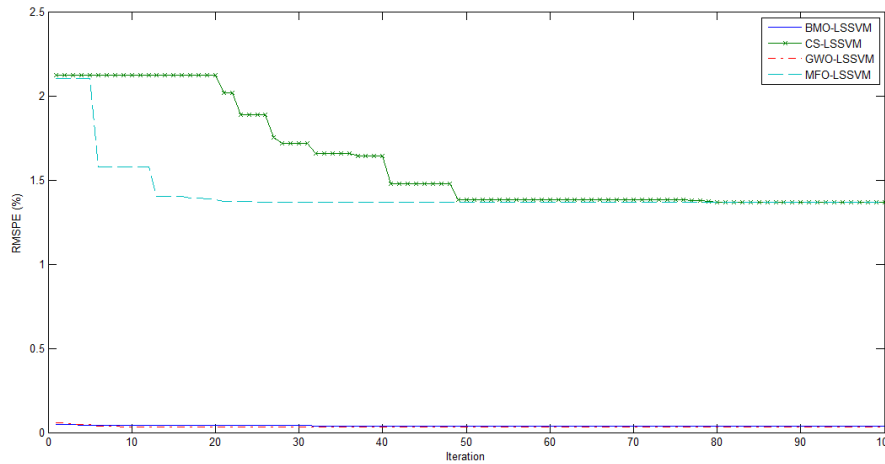


Fig. 6. Comparison on convergence rate: BMO-LSSVM vs. CS-LSSVM vs. GWO-LSSVM vs. MFO-LSSVM

Besides the above performance measurement, another statistical test has been also conducted which is T-test. The obtained results are tabulated in Table 5. The table clearly shows that positive results are obtained for each pair of comparison. This implies that the statistical level of the difference of the means between BMO-LSSVM

and the other three identified hybrid algorithms is significant at 0.05% significance level.

Table 5. Significant test

Algorithms	Sig. (2 tailed)
BMO-LSSVM – CS-LSSVM	0.0181
BMO-LSSVM – GWO-LSSVM	0.0000
BMO-LSSVM – MFO-LSSVM	0.0129

6. Discussion

The obtained results produced by BMO-LSSVM show the capability of the proposed prediction model in yielding closer prediction values, compared to the other identified hybrid algorithms. The results produced show a good agreement in all employed statistical metrics utilized namely RMSPE, MAPE, MdAPE and Theil'sU. Besides the prediction error, the convergence rate show a competitive result as well. This shows the capability of BMO-LSSVM to escape from local minimum. To justify the proposed enhancement, the T-test has been conducted, which showed the positive result. On the other hand, the incapability of the prediction model to predict correct or closer values during sudden spike of cases are one of the challenges in time series prediction. For that matter, enhancing the capability of BMO exploration would be an interesting issue to be addressed, hence better generalization in prediction can be obtained.

7. Conclusions

The Covid19 pandemic is exponentially spreading all over the world rapidly. This study proposes a novel hybrid BMO-LSSVM prediction model for Covid19 confirmed cases in China. The study has employed daily frequency of Covid19 confirmed cases to predict the number of confirmed cases from day 7 and onwards. Comparison has been made against three comparable hybrid algorithms namely CS-LSSVM, GWO-LSSVM and MFO-LSSVM. The practical contribution emphasized in this study is to overcome the limitation of using single BMO as a prediction model where an application of single BMO will cause more optimization parameters to be optimized, which also could cause changes in the algorithm, if the independent factors change. With the hybridization with LSSVM, the prediction model just needs to optimize two hyper-parameters of LSSVM. The findings of the study demonstrate the capability of BMO-LSSVM by its capability in producing lower prediction error over the identified hybrid algorithms. Besides, in prediction rate, the convergence rate produced by BMO-LSSVM is also better. The significant test that has been conducted also prove that the difference of the means between BMO-LSSVM and the other three identified hybrid algorithms is significant.

Acknowledgments: This research was funded by Ministry of Education, Grant Number FRGS/1/2019/ICT02/UMP/02/4. The authors fully acknowledge Faculty of Computing, Universiti Malaysia Pahang (<https://fk.ump.edu.my/index.php/en/>).

References

1. Bullock, J., A. Luccioni, K. H. Pham, C. S. N. Lam, M. Luengo Oroz. Mapping the Landscape of Artificial Intelligence Applications against COVID-19. – *Journal of Artificial Intelligence Research.*, Vol. **69**, 2020, pp. 807-845.
2. Hajrahimi, Z., M. Khashei. Hybrid Structures in Time Series Modeling and Forecasting: A Review. – *Engineering Applications of Artificial Intelligence*, Vol. **86**, 2019, pp. 83-106.
3. Statistics and Research Coronavirus (COVID-19) Vaccinations. 2021.
<https://ourworldindata.org/covid-vaccinations>.
4. Mirjalili, S. Moth-Flame Optimization Algorithm: A Novel Nature-Inspired Heuristic Paradigm. – *Knowledge-Based Systems.*, Vol. **89**, 2015, pp. 228-249.
5. Huang, G., L. Wu, J. Fan, X. Ma, H. Zhou, W. Zeng. Hybrid Extreme Learning Machine with Meta-Heuristic Algorithms for Monthly Pan Evaporation Prediction. – *Computers and Electronics in Agriculture*, Vol. **168**, 2020, No 105115, pp. 1-12.
6. Milan, S. T., L. Rajabion, H. Ranjbar, N. J. Navimipour. Nature Inspired Meta-Heuristic Algorithms for Solving the Load-Balancing Problem in Cloud Environments. – *Computers and Operations Research*, Vol. **110**, 2019, pp. 159-187.
7. Geraili, A., P. Sharma, J. A. Romagnoli. A Modeling Framework for Design of Nonlinear Renewable Energy Systems through Integrated Simulation Modeling and Metaheuristic Optimization: Applications to Biorefineries. – *Computers and Chemical Engineering.*, Vol. **61**, 2014, No 11, pp. 102-117.
8. Ghasemi, A., M. Gheydi, M. J. Golkar, M. Eslami. Modeling of Wind/Environment/Economic Dispatch in Power System and Solving via an Online Learning Meta-Heuristic Method. – *Applied Soft Computing*, Vol. **43**, 2016, pp. 454-468.
9. Zhang, P., H.-N. Wu, R.-P. Chen, T. H. T. Chan. Hybrid Meta-Heuristic and Machine Learning Algorithms for Tunneling-Induced Settlement Prediction: A Comparative Study. – *Tunneling and Underground Space Technology*, Vol. **99**, 2020, No 103383, pp. 1-13.
10. Gocken, M., M. Ozcalici, A. Boru, A. T. Dosdogru. Integrating Metaheuristics and Artificial Neural Networks for Improved Stock Price Prediction. – *Expert Systems with Application*, Vol. **44**, 2016, No February, pp. 320-331.
11. Somu, N., MR G. Raman, K. Ramamritham. A Hybrid Model for Building Energy Consumption Forecasting Using Long Short Term Memory Networks. – *Applied Energy*, Vol. **261**, 2020, No 114131, pp. 1-19.
12. Haupt, R. L., S. E. Haupt. *Practical Genetic Algorithms*. Second Ed. New Jersey, John Wiley & Sons, Inc., 2004.
13. Storn, R., K. Price. Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. – *Journal of Global Optimization*, Vol. **11**, 1997, pp. 341-359.
14. Xin Yao, G. L., Yong Liu. Evolutionary Programming Made Faster. – *IEEE Transactions on Evolutionary Computation*, Vol. **3**, 1999, No 2, pp. 82-102.
15. Kennedy, J., R. Eberhart. Particle Swarm Optimization. – In: *Proc. of International Conference of Neural Networks*, December 1995, pp. 1942-1948.
16. Dorigo, M., T. Stutzle. *Ant Colony Optimization*. Cambridge, MIT Press, 2004.
17. Karaboga, D. *An Idea Based on Honey Bee Swarm for Numerical Optimization*. Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.
18. Mirjalili, S., S. M. Mirjalili, A. Lewis. Grey Wolf Optimizer. – *Advances in Engineering Software*, Vol. **69**, 2014, pp. 46-61.
19. Rashedi, E., H. Nezahabadi-Pour, S. Saryazdi. GSA: A Gravitational Search Algorithm. – *Information Sciences*, Vol. **179**, 2009, pp. 2232-2248.
20. Osman, I. E., K. Erol. A New Optimization Method: Big Bang-Big Crunch. – *Advances in Engineering Software*, Vol. **37**, 2006, No 2, pp. 106-111.
21. Hatamlou, A. Black Hole: A New Heuristic Optimization Approach for Data Clustering. – *Information Sciences*, Vol. **222**, 2013, No 10, pp. 175-184.

22. Kang Seok Lee, Zong Woo Geem. A New Meta-Heuristic Algorithm for Continuous Engineering Optimization: Harmony Search Theory and Practice. – Computer Methods in Applied Mechanics Engineering, Vol. **194**, 2005, No 36-38, pp. 3902-3933.
23. Moghdani, R., S. Khodaram. Volleyball Premier League Algorithm. – Applied Soft Computing, Vol. **64**, 2018, No March, pp. 161-185.
24. Sulaiman, M. H., Z. Mustafa, M. M. Saari, H. Daniyal. Barnacles Mating Optimizer: A New Bio-Inspired Algorithm for Solving Engineering Optimization Problems. – Engineering Applications of Artificial Intelligence, Vol. **87**, 2020, No 103330, pp. 1-13.
25. Parbat, D., M. Chakraborty. A Python Based Support Vector Regression Model for Prediction of COVID19 Cases in India. – Chaos, Solitons & Fractal, Vol. **138**, 2020, No 109942, pp. 1-5.
26. De Oliveira, L. S., S. B. Gruetzmacher, J. P. Teixeira. COVID-19 Time Series Prediction. – Procedia Computer Science, Vol. **181**, 2021, pp. 973-980.
27. Jain, A., T. Sukhdeve, H. Gadia, D. S. P. Sahu, S. Verma. COVID19 Prediction Using Time Series Analysis. – In: Proc. of International Conference on Artificial Intelligence and Smart Systems (ICAIS'21), 2021, pp. 1599-1606.
28. Didem Guleryuz. Forecasting Outbreak of COVID-19 in Turkey; Comparison of Box-Jenkins, Brown's Exponential Smoothing and Long Short-Term Memory Models. – Process Safety and Environmental Protection, Vol. **149**, 2021, No May, pp. 927-935.
29. Barazandeh, M., C. S. Davis, C. J. Neufeld, D. W. Coltman, A. R. Palmer. Something Darwin Didn't Know About Barnacles: Spermcast Mating in a Common Stalked Species. – In: Proc. of Royal Society B Biological Sciences, 2013.
30. Yusa, Y., M. Yoshikawa, J. Kitaura, M. Kawane, Y. Ozaki, S. Yamato, J. T. Høeg. Adaptive Evolution of Sexual Systems in Pedunculate Barnacles. – In: Proc. of the Royal Society B: Biological Sciences, Vol. **279**, 2012, pp. 959-966.
31. Suykens, J. A. K., T. van Gestel, J. de Brabanter, B. de Moor, J. Vandewalle. Least Squares Support Vector Machines. Leuven, Belgium, World Scientific Publishing Co. Pte. Ltd., 2002.
32. Coronavirus Disease (COVID-2019) Situation Reports, 2020.
<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>

Received: 13.08.2021; Second Version: 16.10.2021; Accepted: 29.10.2021