Dengue Outbreak Prediction Using an Improved Salp Swarm Algorithm.

Khairunnisa Amalina Mohd Rosli¹, Zuriani Mustaffa², Yuhonis Yusof³ and Mohamad Farhan Mohamad Mohsin⁴

¹,²Faculty of Computing, Universiti Malaysia Pahang, 26300, Kuantan, Pahang, Malaysia.
³School of Computing, Universiti Utara Malaysia, 06010, UUM Sintok, Kedah, Malaysia.
⁴UUM CAS Student Development and Alumni, School of Computing, Universiti Utara Malaysia, 06010, UUM Sintok, Kedah, Malaysia
khairunnisaamalina@gmail.com

Abstract: Dengue disease is the most common type of disease caused by mosquitoes. It is reported that dengue fever was first recognized in Thailand and Philippines in 1950. According to World Health Organization (WHO), dengue is a viral disease that spread in public environment where the number of dengue cases reported enlarge within 5 years by 1 million from 2.2 million in 2010 to 3.2 million in 2015. Until today, numerous studies by researchers to improve the prediction of dengue fever disease based on Computational Intelligence (CI) methods have been reported. The research includes study using Swarm Intelligence (SI) algorithm. This research study proposed an improved Salp Swarm Algorithm (iSSA) for dengue outbreak prediction. The original SSA will be enhanced by enriching the exploration and exploitation process for the sake of improving the accuracy of dengue outbreak prediction. This will be done by inducing a mutation based on Levy Flight. Later, the iSSA algorithm will be realized on dengue disease dataset. The proposed iSSA will be compared against the original SSA and another CI method known as Grey Wolf Optimization (GWO). Two evaluation indicators known Root Mean Square Error (RMSE) and as Mean Absolute Error (MAE) are proposed in this research study to evaluate the prediction model where the smaller the value obtained, more accurate the prediction model. The result demonstrated that the proposed model produces a better result compare two the other results where the value of MAE and RMSE of the proposed model is smaller compare to other two model.

Keywords— Dengue outbreak prediction, Levy Flight, Salp Swarm Algorithm

1. Introduction
In the last few decades, nature inspired metaheuristic algorithms becoming more popular and dominant in computational intelligence [1]. Computational intelligence and meta heuristic algorithm have evolve and become popular in computer science including machine learning, artificial intelligence and data mining [2]. Many of nature inspired algorithm have been proposed where these algorithms imitate the nature behaviour of animals and they are known as Swarm Intelligence (SI) algorithms [3]. For example, Artificial Bee Colony algorithm (ABC) inspired by the food foraging behaviour of honeybee colony [4]. Next, an algorithm that mimic an ant in nature to find the food searching path is called Ant Colony
Optimization (ACO) [5] and Particle Swarm Optimization (PSO) that simulates the social behaviour flocks of bird [6].

Recently, a relatively new SI algorithm has been proposed, namely Salp Swarm Algorithm (SSA) SSA is introduced in year 2017 by Mirjalili [7]. Exploration and exploitation characteristics are vital in determining the achievement of optimization algorithm where exploration behaviour will give guarantee that the algorithm can explore the search space precisely meanwhile exploitation has the ability to search around the current best solution and select the best solution [8]. SSA algorithm goal is to enhance the balance between exploration-exploitation behaviour where it focuses on particles motion [9][14].

Despite its excellent performance, SSA is still open for an improvement where it is reported that more exploitation than exploration leads to higher convergence speed but the risk of not reaching to the global solution. On the other hand, more exploration leads to slow convergence speed and solutions might not be improved well enough to find the global optimal. These demerits need critical attention as it would demote the prediction result. The main objective of this study is to minimize the objective function where lower the MSE value more promising the performance of the algorithm.

In literature review, there are numerous techniques has been proposed by researchers to improve the prediction of dengue outbreak. Various Computational Intelligence methods have been created to handle the issue of dengue fever disease. In year 2018, a research study to predict dengue outbreak has been proposed [10]. In this study LSSVM is hybrid with new optimization algorithm namely Flower Pollination Algorithm (FPA) is implemented for dengue outbreak prediction in Yogyakarta Indonesia. They highlighted to determine the optimal value of LSSVM hyper-parameter FPA is used as optimization tool. In 2015, a study in predicting Dengue Fever (DF) in Kuala Lumpur by implementing artificial neural network method has been introduced where they used Linear Regression Model [11]. This research study considered mean daily temperature, daily rainfall and EVI as their factors for the proposed model. Cubic Spline Interpolation method is crucial to find a meaningful data for training purpose. Prediction model is used as it can produces model with low variances and deal with irregular observable variables. This research study is using Mean Test Error (MTE) to evaluate the results. The result shows that more training data, it will provide more accurate result.

In addition, quite a number of improvements that involved the inducing of Levy Flight (LF) in SI algorithms have been proposed to enhance the exploration and exploitation. In [12] they proposed an improved PSO by implement Levy Flight (LF) techniques to update the particle’s velocity and improve the capability of global search. A prediction model by implement LF into Cuckoo Search (CS) has been proposed in [13]. The modification of this study is focusing on the information exchange between two best eggs in algorithm or their best solution. Then CS is validated by using test function and compare the performances with GA and PSO. LF random walk is proposed to balance the exploration and exploitation characteristic of ABC [14]. Global optima of unimodal or multimodal functions are decreasing by implement LF random walk. They will find the step size in search space including numbers of scout bees is raised. PSO with Active Velocity Penalty (PSO-AVP) is proposed [15] to penalizes the particles velocities to ensure they are move within the search space to harmonize their exploration and exploitation behaviour.

Due to the efficiency of LF in improving the SI algorithm, this study proposed an improved SSA (iSSA) by embedding the LF function in exploitation and exploration process. Then iSSA later will be realized for dengue outbreak prediction.

This paper is presented as follows: section II will explain a description on SSA continue with the methodology that will be implemented in section III followed with experimental results in section IV and lastly the conclusions of the research study in section V.
2. Salp Swarm Algorithm

SSA algorithm comes from the family of Salpidae known as the family of salps. They have a unique transparent barrel-shaped body that similar to jelly fish including its movement where the water will be pumped through their body as propulsion to move forward. Main foundation of SSA algorithm is because of the swarming behaviour. Salp chain will be form in the deep ocean when a group of salp is combined. Researchers propose this algorithm to solve optimization problem. Leader and followers are the population of SSA where the salp at the front of the chain is known as the leader while the other salps that follow at the back is known as the followers. The followers will follow one another directly or indirectly led by their leader. Some remarks are listed to determine the ability of the proposed SSA algorithm in solving optimization problem [7].

• SSA algorithm will saves the best solution and assigns it to the food source variable to ensure it never gets lost though the whole population worsen.
• SSA algorithm updates the position of the leader salp parallel with the food source to ensure the leader salp can always explores and exploits the space around it.
• SSA algorithm will constantly updates the position of follower salps as they will slowly move towards the leader salp.
• Continuous movement of follower salps can prevent SSA algorithm from changing easily in local optima.
• $c_1$ parameter is reduced adaptively over the course of iterations hence SSA algorithm will explores the search space and then exploits it.
• $c_1$ is the main controlling parameter that will control the exploration and exploitation behaviour.
• SSA algorithm is simple and easy to implement.

3. Methodology

The methodology is divided into four (4) main sections namely data collection, iSSA Prediction Model, Salp Swarm algorithm with Levy Flight Technique and finally evaluation section. For data evaluation, it will discuss about data collection and also data pre-processing process, which is also known as data cleaning to remove noise data and converted the data into understandable data. Next is prediction by using iSSA where iSSA will be applied into dengue disease raw dataset to demonstrate prediction performance. Then, Salp Swarm algorithm with Levy Flight Technique is introduced. Lastly, this section will explain the method used to evaluate the performance of iSSA in dengue outbreak prediction.

3.1. Data Collection

For this study, a real dataset of dengue in Malaysia the data is collected. Dengue dataset is provided by the researchers from Universiti Utara Malaysia. The collected data is from year 2003 until year 2008. There is a total of 11 observations of dengue dataset input. The output variable of this dataset is dengue cases from day 7 and onwards. Input and output variable are important to evaluate the output accuracy. In this study, there are 4 input variables involved for predicting dengue outbreak which are number of dengue cases, rainfall (mm), temperature ($^\circ$C) and humidity (%). These features will be used to assist the model in learning the relationship of the variable with dengue outbreak accuracy. For data analysis, MATLAB language was selected.
Figure 1 explains the methodology of iSSA prediction Model. Firstly, data collection and data pre-process take place to remove meaningless information and redundant data before going to training stage. A good pre-processing technique will improve the efficiency of the result. Then the data will be analyzed by implement into iSSA model to determine whether they have low prediction error or vice versa. Then dengue dataset will be analyzed and undergo evaluation. The input and output of this research study is presented in Table 1.

Table 1. Dengue cases data

| Input       | Output                                      |
|-------------|---------------------------------------------|
| Dengue Case | Dengue cases from day 7 and onwards         |
| Humidity    |                                             |
| Rain        |                                             |
| Temperature |                                             |

1) Training, Validation and Testing: Table 2 shows data proportion that has been set for simulation purposes.


Table 2. Data Proportion for Training, Validation and Testing

| Data Proportion | Training | Validation | Testing |
|-----------------|----------|------------|---------|
|                 | 70%      | 15%        | 15%     |

3.2 iSSA Prediction Model

SSA algorithm begins with approaching the global optimum. They will assign multiple salps with random positions. The fitness of each salp is calculated where the salp with the best fitness function will be elect as the best salp that will chase the food source represented by variable F. In the meantime, $c_1$ parameter is updated. For each dimension, both position of the leader and follower are updated. If any salps goes beyond the search space, it will be brought back on the boundaries. The above steps are executed repeatedly until the end criterion is fulfilled excluding initialization. It is highlighted that the food source will be updated during optimization as the salp chain will find a better solution by exploring and exploiting the search space. The above procedures show that the salp chain modelled can chase a moving food source. Hence, the salp chain can move towards the global optimum that changes over the course of iterations. Figure 2 below show the pseudocode of SSA algorithm:

```
Initialize the salp position $x_i$ ($i = 1, 2, \ldots, n$) considering $ub$ and $lb$
While (end condition is not satisfied)
    Calculate the fitness of search agent (salp)
    $F =$ the best search agent
    Update $c_1$ with equation (1)
        For each salp ($x_i$)
            Update the position of leading salp with equation (2)
        Else
            Update the position of follower salp with equation (3)
    End
End
Amends the salp based on the $ub$ and $lb$
End
Return F
```

Figure 2. SSA pseudocode

\[ x_j^* = \begin{cases} 
  F_j + c_1 \left( (ub_j - lb_j)c_2 + lb_j \right) & c_3 \geq 0 \\
  F_j - c_1 \left( (ub_j - lb_j)c_2 + lb_j \right) & c_3 < 0 
\end{cases} \]  

(1)

\[ c_1 = 2e^{-\frac{4c_2}{u^2}} \]  

(2)

5
Equation (1) is proposed to revise the position of leading salp where $x_j^i$ is the position of leader salp in the $j^{th}$ dimension while $F_j$ is the food source in $j^{th}$ dimension, $u_b^j$ is the upper bound meanwhile $l_b^j$ is the lower bound in $j^{th}$ dimension and $c_1^j$, $c_2^j$, $c_3^j$ are random numbers between $[0,1]$ interval.

In equation (2) $c_i^j$ are crucial parameter that will determine the exploration and exploitation criteria of the algorithm where $l$ is the current iteration and $L$ is the maximum iteration. Equation (3) indicates the position of the follower salp where $x_j^i$ is the position of $i^{th}$ follower in $j^{th}$ dimension. Random quantity that helps in exploration and exploitation of the search space play a crucial role in equation of SSA. In this study, Levy Flight (LF) will be implemented to stabilize the diversity and convergence in SSA. LF has local search and global search where it can be accomplished by coordinate the LF parameters and automatically adjust the step sizes [13]. This proposed technique able to improve the exploitation capability of SSA.

Exploration characteristic main objective is to determine the global optimum by identifying the unknown region of search space. Meanwhile exploitation is the ability to apply the knowledge of previous solution in order to find a better solution. In this study, the main purpose is to improve the performance and search space. It is highlighted to update the previous solutions step size plays an important role. The new solution may exceed the true solution if the step size is large, meanwhile if the step size is small, then the convergence rate may decrease that will result in poor performance.

\[
\chi_j^i = \frac{1}{2} (\chi_j^i + \chi_j^{-1})
\]  

(3)

\[
\sigma_u = \frac{\Gamma(1+\beta)\sin(\pi \beta/2)}{\beta \Gamma[(1+\beta)/2]2^{(\beta-1)/2}}
\]  

(4)

\[
\text{step_size}(\tau) = 0.01 \times s(\tau)
\]  

(5)
\[ x'_i(t + 1) = x_i(t) + \text{step}\_size(t) \times U(0,1) \] (6)

The step size by using Levy distribution for exploitation of search area calculated as in equation (5). Factor of 0.01 is L/100 where L is typical length scale. Equation (6) where \( x'_i \) is individual that will modify the position, \( U(0,1) \) is a random number between [0,1] range and \( \text{step}\_size(t) \times U(0,1) \) is the random walk from levy distribution.

### 3.3 Salp Swarm Algorithm with Levy Flight Technique

The main objective for this research study is to enhance both exploration and exploitation criteria of SSA in order to accomplish better optimization performance. For that matter, Levy Flight technique is introduced to help in exploration consequently of exploitation of the search space. A newly introduced parameter of Improved SSA will enhance its convergence speed and balance the ability of exploitation and exploration. In addition, this proposed algorithm will produce a better performance of the solution. The improved SSA is stated in the equation (7) including the pseudocode of iSSA show in Fig 4 below:

\[
x^f_{j} = \begin{cases} 
F^f_j + \alpha ((\text{ub}_j - \text{lb}_j)\text{Levy} + \text{lb}_j)c_2 
& \geq 0 \\
F^f_j + \alpha ((\text{ub}_j - \text{lb}_j)\text{Levy} + \text{lb}_j)c_2 
& < 0
\end{cases}
\] (7)

\( x^f_{j} \): the position of the first salp in \( j^{th} \) dimension.

\( F^f_j \): the position of the food source in \( j^{th} \) dimension.

\( \text{ub}_j \): is the upper bound.

\( \text{lb}_j \): is the lower bound in \( j^{th} \) dimension.

\( c_1, c_2 \) are random numbers in interval [0,1]

---

**Figure 4. iSSA pseudocode**

```plaintext
Begin
i. Initialize the parameter of SSA (number of salps (s), iteration number (l) and salp position (x))
ii. Initialize a population of salp’s position randomly
iii. Evaluate the fitness of each salp
iv. Set the number of iteration (l)
v. Update \( c_1 \) by equation (1)
vi. For each salp,
   If \( i = 1 \) update the position of leading salp by equation (7)
   Otherwise update the position of follower salp by using equation (5)
   Evaluate the fitness of every salp
   Update \( x^f \)
   Increment i to 1
vii. Repeat step (iv-xi)
viii. Return the best solution and fitness value
End
```
4. Performance Evaluation Matrices
In this paper, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are calculated to forecast the percentage accuracy of dengue outbreak cases. RMSE is the standard deviation of prediction error (residual). Residual are calculated to determine the robustness of the data around the line of best fit. MAE is a quantity used to measure the difference between two continuous variables. The equations are defined as follows:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - p_i)^2}
\]

(8)

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - p_i|
\]

(9)

Based on the above equation, \(y_i\) is the actual value meanwhile \(p_i\) is the prediction value of ith test data obtained from the proposed model and \(n\) is the number of data.

MAE and RMSE are important indicators of dengue prediction results where the smaller values obtained, more accurate the prediction model.

5. Result And Discussion
This section will present the result obtained from the proposed iSSA. The capability of iSSA is compared with the original SSA and another swarm algorithm known as Grey Wolf Optimization (GWO). The performance evaluation of iSSA, SSA and GWO are shown in Table 3.

|                  | iSSA     | SSA      | GWO      |
|------------------|----------|----------|----------|
| MAE              | 1.84E-12 | 2.20E-12 | 3.022E-12|
| RMSE             | 1.357E-6 | 1.484E-06| 1.4125E-06|

From the table, it shows that iSSA recorded the lowest MAE and RMSE value of 1.84E-12 and 1.357E-06 respectively which is the lowest error compared with the original SSA that recorded 2.20E-12 of MAE value and 1.484E-06 of RMSE value. Meanwhile the value of MAE and RMSE value of GWO are 3.022E-12 and 1.4125E-06 respectively. The upper bound value is set to 1000 meanwhile 500 iteration is applied. Based on the result, it shows iSSA is reliable as both value of MAE and RMSE are smaller and has less error.

6. Conclusion
Main objective of this research study was to predict dengue outbreak cases based on the proposed algorithm namely iSSA. By using a real dataset with parameters of number of dengue cases, temperature (°C), rainfall values (mm), and humidity (%). The result achieved by the proposed iSSA was compared with the original SSA and another CI algorithm namely GWO. The performance of these algorithms is guided by MSE and RMSE where the lower the value more reliable the result. The finding of the study shows that the proposed algorithm has a promising result in comparison with the other two algorithms.
Acknowledgment

The author would like to thank to Faculty of Computing (FKOM), Universiti Malaysia Pahang. This study is funded under the UMP Research Grant. Grant No: RDU180366.

References
[1] Xin-She Yang, Su Fong Chien, Tiew On Ting, “Computational Intelligence and Metaheuristic Algorithms with Applications”, The Scientific World Journal Volume 2014.
[2] X. S. Yang, Cuckoo Search and Firefly Algorithm: Theory and Applications, vol. 516 of Studies in Computational Intelligence, Springer, Heidelberg, Germany, 2014.
[3] H.-S. Yang, Engineering Optimisation an Introduction with Metaeuristic Application, 1sr es., John Wiley & Sons, New Jersey, 2010.
[4] D.T. Pham, A. Ghanbarzadeh, E. Koc, S.Otri, S.Rahim, M.Zaidi, The bees algorithm, Technical Note, Manufacturing Engineering Centre, Cardiff University, UK, 2005
[5] N.Zhang, J.Ji, C.Liu,N.Zhong, An ant colony optimization algorithm for learning classification rules in:Proceeding of 2006 IEEE/WIC, 2006, pp. 1034-1037.
[6] J.Kennedy, R.C Eberhart, Particle swarm optimization, in:Proceedings of IEEE International Joint Conference on Neural Networl, vol.4, 195, pp. 1942-1948
[7] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, “Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems,” Adv. Eng. Softw., vol. 114, pp. 163–191, 2017.
[8] C. Suvarna, A. Sali, and S. Salmani, “Efficient heart disease prediction system using optimization technique,” Proc. Int. Conf. Comput. Methodol. Commun. ICCMC 2017, vol. 2018–January, no. ICcmm, pp. 374–379, 2018.
[9] C. Paper, “an Efficient Method To Predict Dengue,” no. October, pp. 12–13, 2015.
[10] Z. Mustaffa, M. H. Sulaiman, F. Emawan, Y. Yusof, and M. F. M. Mohsin, “Dengue outbreak prediction: Hybrid meta-heuristic model,” Proc. - 2018 IEEE/ACIS 19th Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. SNPD 2018, pp. 271–274, 2018.
[11] C. Paper, “an Efficient Method To Predict Dengue,” no. October, pp. 12–13, 2015.
[12] R.Jensi, G.Wiselin Jiji, An enhanced particle swarm optimization with levy flight for global optimization, Applied Soft Computing Elsevier, 2015
[13] Sangita Roy, Chaudhuri S.S, “Cuckoo Search Algorithm using Levy Flight: A Review”, IJ Modern Education and Computer Science, 2013,12,10-15.
[14] Harish Sharma, Jagdish C.B, Arya K.V, Yang Xin-She. “Levy Flight artificial bee colony algorithm”, International Journal of System Science, 2015
[15] Mounir Ben Ghalis, “Particle Swarm Optimization with an improved Exploration-Exploitation Balance”, IEEE, 2008