Prediction of Concrete Compressive Strength from Early Age Test Result Using an Advanced Metaheuristic-Based Machine Learning Technique

D. Prayogo1[a], M.-Y. Cheng2[b], J. Widjaja3[a], H. Ongkowijoyo4[a], and H. Prayogo5[a]

[a]Department of Civil Engineering, Petra Christian University, Indonesia
[b]Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taiwan
E-mail: prayogo@petra.ac.id, myc@mail.ntust.edu.tw, m21414116@john.petra.ac.id, m21414084@john.petra.ac.id, m21415097@john.petra.ac.id

Abstract –

Determining accurate concrete strength is a major civil engineering problem. Test results of 28-day concrete cylinder represent the characteristic strength of the concrete that has been prepared and cast to form the concrete work. It is important to wait 28 days to ensure the quality control of the process, although it is very time consuming. Machine learning techniques are progressively used to simulate the characteristic of concrete materials and have developed into an important research area. This study proposed a comprehensive study using an advanced machine learning technique to predict the compressive strength of concrete from early age test results. In this case, early age test data are being used to get reliable values of the two constants which are required for the prediction. A total of 28 historical cases were used to establish the intelligence prediction model. Obtained results show the performance of the advanced hybrid machine learning technique in predicting the concrete strength with a relatively high accuracy measured by four error indicators. Therefore, the proposed study can offer a high benefit for construction project managers in decision-making processes based on early strength test results.

Keywords –

Concrete; Compressive Strength; Early Strength; Machine Learning; Prediction

1 Introduction

Concrete is a material used in construction that has great versatility and which is used across the globe. Concrete has several advantages, including good compressive strength, durability, workability, construction availability, and low cost. Nevertheless, these benefits have a serious reliance on curing, placing, and the appropriate mix. Strength, within the construction industry, is a major criterion in the selection of a concrete to be used in a specific application. Construction concrete will gain strength throughout an extended period after it has been poured. Concrete's nominal strength has a definition of a sample's compressive strength that is 28 days old.

If there is a faulty mix preparation or mix design on site, the results of the test may indicate failure to reach the required strength, triggering a mandatory repeating of the whole process, and this can be time consuming and costly. For all failures, another 28 days must be awaited, so there has long been a requirement for an ability to estimate, at an early age, the concrete's final strength. Therefore, an appropriate, quick method of predicting concrete strength would be a significant advancement for the industry [1]. An ability to predict the compressive strength of concrete early allows constructers to quickly understand the concrete's probable weaknesses and make a decision to manage a destruction process or continue with construction. Further, to the benefit of both user (or purchaser) and producer, reliably and rapidly predicting the results of a 28-day test would benefit all stakeholders as opposed to waiting the full, conventional, 28 days.

Researchers are eager to explore the behavior of the concrete, therefore predicting concrete strength, and this has been targeted as an active area to be researched. There are multiple recent studies exploring concrete's behavior and the possibility of improving characteristic strength predictions. The studies have revealed that many tests have concentrated on the way that the strength of the concrete is affected by the mix, but only a small number of studies have focused on the relation between early testing (for example, 7 and 14 days) and the full 28-day compressive strength test. Moreover, most of the studies have some limitations; for example, an absence of advanced method for measuring accuracy, no validating techniques involved, or using only a conventional approach.

Machine learning methods are proven to outperform conventional methods because of their excellent learning functions [2-6]. The conventional approaches, such as linear regression and decision tree, are not sufficient to build a satisfactory model in terms of accuracy and computational time. Therefore, this primary research aim is to construct an advanced accurate model to improve early age predictions of concrete strength.

This research proposes a new advanced machine learning technique called Symbiotic Organisms Search-based Support Vector Regression (SOS-SVR). SOS-SVR fuses an accurate prediction technique, Least Squares Support Vector Regression (LS-SVR) [7], and a very promising metaheuristic, Symbiotic Organisms Search (SOS) [8]. The proposed model will be investigated alongside other prediction methods in building an accurate prediction model of concrete strength in correspondence to the early age strength test results.

2 The Proposed Metaheuristic-Based Machine Learning Technique

2.1 Least Squares Support Vector Regression

LS-SVR is a modified version of the support vector regression (SVR) [9]. The LS-SVR is a statistical learning theory that adopts a least squares linear system as a loss function instead of the quadratic program in the original support vector machine (SVM) [10]. The optimization problem and the constraints for LS-SVR can be stated by the following formulation.

Minimize: $J_p(w, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2$ (1)

Subjected to: $y_k = w^T \phi(x_k) + b + e_k, k = 1, \dots, N$ (2)

where $w \in R$ is an undetermined parameter vector; $\phi(\cdot)$ is the nonlinear function introduced by a kernel function for mapping the input space to a high-dimensional feature space; $e_k \in R$ is error variable; γ denotes a regularization constant.

It is worth noting that the objective function includes a sum of squared fitting error and a regularization term. However, when w becomes infinite dimensional, one cannot solve this primal problem. Therefore, it is necessary to construct the Lagrangian and derive the dual problem. The resulting LS-SVR model for function estimation is expressed as:

$$y(x) = \sum_{k=1}^{N} \alpha_k K(x_k, x_l) + b \quad (3)$$

where α_k and b are the solution to the linear system. The kernel function that is often utilized is the radial basis function (RBF) kernel. Description of RBF kernel is given as follows:

$$K(x_k, x_l) = exp\left(\frac{\|x_k - x_l\|^2}{2\sigma^2}\right) \quad (4)$$

where σ is the kernel function parameter.

There has been an increased used of the LS-SVR in many engineering fields. However, the LS-SVR presents a problem in prediction accuracy whenever its parameters are not fine-tuned. The regularization parameter (γ) is a positive cost parameter similar to the C in SVM/SVR, while the kernel parameter (σ) is an additional parameter since LS-SVR utilized the RBF kernel function. These two parameters need to be specified in order to find the best prediction model. Thus, there is a need of metaheuristic algorithm in order to find the most effective combination of LS-SVR parameters. It has been shown that the self-tuned parameter approaches, involving the metaheuristic as the optimizer, have been studied extensively and has produced an increased prediction accuracy [11, 12].

2.2 Symbiotic Organisms Search

In recent years, many researchers proposed a new metaheuristic algorithm to search for optimality in various optimization problems. It has been proven that metaheuristic algorithm can be an effective tool in dealing with challenging problems. Some notably, examples of recent metaheuristic algorithms applications in tackling hard and complex problems can be seen in [13-18].

Another new and powerful metaheuristic algorithm namely symbiotic organisms search (SOS) has been proposed by Cheng and Prayogo [8] to solve the continuous based optimization problems. The experimental results have been shown in the high efficiency of SOS performance in comparison with the conventional metaheuristic techniques including genetic algorithm (GA), particle swarm optimization (PSO), and differential evolution (DE).

The SOS algorithm adopts the interaction behaviour among organisms which are living together in one ecosystem. It gradually guides a population of candidate solutions, denoted as organisms, towards promising regions in the solution space. Each candidate solution, denoted as organism, has a certain fitness value which corresponds to the objective function or goal of the problem. The main steps of the SOS algorithm can be seen as follows.

1: Initialize random population
2: **Do**

3: Mutualism phase
4: Commensalism phase
5: Parasitism phase
6: Preserve the current best solution
7: **Until** termination criteria are met

For more details, the reader can refer to [8]. Since the first publication in 2014, the use of SOS in solving different problems has been significantly increased in various research topics [19-27]. It shows the potential of SOS algorithm as an attractive tool to search for the optimality.

2.3 Symbiotic Organisms Search-Based Support Vector Regression

The symbiotic organisms search-based support vector regression (SOS-SVR) model is a hybrid machine learning system that combines the two different techniques of SOS and LS-SVR. In this system, the LS-SVR acts as a supervised-learning-based predictor to build the accurate input-output relationship of the dataset; and the SOS works to optimize the LS-SVR parameters, the σ and γ parameters.

The SOS-SVR involves eight major steps which are categorized into two phases, beginning with a training phase followed by the testing phase. The whole procedure of SOS-SVR is shown in Figure 1. An explanation of the major steps involved in SOS-SVR is given below:

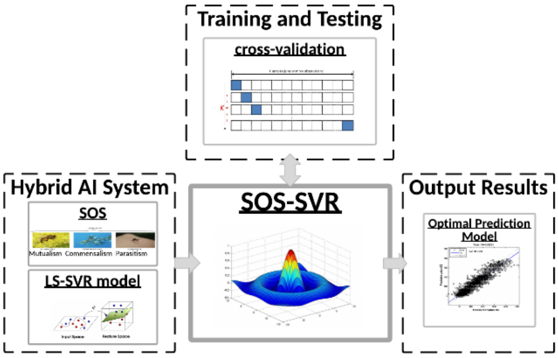

Figure 1. SOS-SVR framework.

1. Hybrid AI System:
   LS-SVR addresses the complex relationship between input and output variables. LS-SVR requires two parameters, σ and γ, to finish the learning process. LS-SVR will gradually obtain the tuning parameters from the optimizer – the SOS algorithm.
   In this hybrid system, SOS is utilized to explore the various combinations of σ and γ parameters to look for the best set of them. SOS utilizes mutualism, commensalism, and parasitism phases to gradually improve the fitness value of each solution.
   The process terminates when the termination criterion is satisfied. While still unsatisfied, the model will proceed to the next iteration. As SOS-SVR uses SOS, the termination criterion used in this study was the SOS iteration number.

2. Training and testing:
   The data for training are obtained from data collection. A fitness function is now developed to evaluate accuracy of the learning system. This function has a correlation with the accuracy of the prediction model. The most accurate prediction model is represented by the combination of σ and γ parameters that produce the best fitness value.
   Instead of randomly splitting the data, it was partitioned into two subsets: learning subset and validation subset. To avoid the sampling bias, the 5-fold cross-validation technique is used in splitting the training data into learning and validation subsets. In this study, the fitness function utilizes mean square error (MSE) of the validation dataset. When the termination criterion is fulfilled, the loop stops. This condition means that the prediction model has identified the input/output mapping relationship with the optimal σ and γ parameters.
   Meanwhile, the data for testing is obtained from data collection. The optimal σ and γ parameters obtained from the training phase are used to establish the prediction model for predicting the testing data.

3. Output Results:
   The optimal σ and γ parameters obtained from the training phase are used to establish the prediction model for predicting the testing data. The accuracy measurement methods are used to measure the performance of the SVR-based prediction model.

3 Experimental Settings

3.1 Data Preparation

The historical database for experiment was obtained from previous literature [28]. In this research, the database is assigned into two group of dataset: (1) the mix proportions with 7-day strength test result and (2) the mix proportions with 14-day strength test result.

Each dataset has a total of 28 records of concrete mix proportion and were used to analyse the behaviour of concrete mixture with strength at the various early ages. Every dataset has 6 input variables and 1 output variable, including the coarse aggregate to cement ratio (CA/C), fine aggregate to cement ratio (FA/C), water to cement ratio (W/C), water (W), coarse aggregate size ratio of 10 mm:20 mm (CAS), 7-day or 14-day strength test result depending on each group (fc-7 or fc-14), and the 28-day strength test result (fc-28), respectively. The characteristics of the variables of each dataset can be seen in Table 1.

Table 1 Statistical description of concrete mix proportion

Variables	Unit	Group	Min	Max	Avg.	Std. dev.
CA/C	-	1, 2	1.4	2.24	1.811	0.256
FA/C	-	1, 2	2.22	2.97	2.591	0.236
W/C	-	1, 2	0.4	0.52	0.460	0.041
W	kg/m3	1, 2	185	190	187.5	2.546
CAS	-	1, 2	1	2	1.464	0.508
fc-7	MPa	1	13.84	26.02	19.025	3.957
fc-14	MPa	2	17.8	35.35	23.659	4.673
fc-28	MPa	1, 2	19.53	37.4	26.972	4.143

3.2 Accuracy Measurement and Parameter Settings

Machine learning is recently employed to predict behavior in many research areas. For the comparison purpose to our proposed method, this research applies five widely used machine learning methods including classification and regression tree (C&R tree), neural network (neural net), regression, support vector machine (SVM), chi-squared automatic interaction detector (CHAID) to model the prediction behavior of concrete strength at various early ages. All aforementioned methods were performed using the IBM SPSS Modeler.

The parameters of the machine learning techniques were set to default for fair comparison. Through trial-and-error, a suitable parameter setting for SOS-SVR is determined as follows: the maximum number of iteration was set to 50; population size was set to 25; and search range for the σ and γ parameters starts from 10^{-5} to 10^{5}.

The effectiveness of state-of-the art machine learning methods can be determined using the accuracy measurement methods. In this research, three measurement methods were adopted – coefficient of correlation (R); root mean squared error (RMSE); mean absolute error (MAE); and mean absolute percentage error (MAPE). The lowest RMSE, MAE, and MAPE values alongside with a highest R value indicate the best model outcome.

3.3 Cross-validation

To obtain the prediction result from each machine learning technique, the dataset must be labelled into two groups of training data and testing data. At first, the training data is used by one method to generate the prediction model. Then, the prediction model is applied to validate the testing data. Once the predicted output is generated, the accuracy measurement is performed to check the deviation between the predicted and actual output.

During the partitioning of training and testing subset, there is a possibility for the bias associated with the random sampling. To minimize the sampling bias, cross-validation technique is utilized in this research. The main goal of the cross-validation technique is to ensure the partitioning process generating the independent and unbiased data subset. The data is then labelled into 5 folds and the performance of each fold is recorded and validated in accordance to each accuracy measurement.

4 Results and Discussions

The training and testing processes are performed for each predictive model. The complete experimental results based on 7-day and 14-day test results can be seen in Table 2 and Table 3, respectively. It can be seen that SOS-SVR has performed better in all measurement categories in the testing dataset.

For training dataset, the SOS-SVR prediction model has produced a decent training accuracy as shown in Table 2. It is worth noting that C&R tree prediction model yielded the highest accuracy on training dataset but the model performs poorly on the testing set. The over-fitting had occurred during the training process of C&R tree. As a result, the generalization ability of C&R tree model to predict the future data is declined drastically. Although the SOS-SVR model did not have the highest accuracy on the training dataset, still it has the highest accuracy on the testing dataset. The metaheuristic-tuned framework helps the SOS-SVR to determine the right balance between learning and generalization paradigm.

Table 2 Comparative experimental results among predictive methods for 7-day strength test result.

Machine Learning Techniques	Average Result			
	R	RMSE (MPa)	MAE (MPa)	MAPE (%)
Training				
C&R tree	0.9965	0.33	0.15	0.55
Neutral net	0.8744	2.06	1.55	5.84
Regression	0.9050	1.71	1.39	5.21
SVM	0.8709	2.18	1.42	5.26
CHAID	0.8618	2.06	1.42	5.14
SOS-SVR	0.8853	1.99	1.30	5.11
Testing				
C&R tree	0.7567	3.55	2.92	10.75
Neutral net	0.8708	2.58	2.19	8.08
Regression	0.8457	3.52	2.92	10.88
SVM	0.8460	2.71	2.26	8.35
CHAID	0.6788	3.77	2.96	10.79
SOS-SVR	0.8793	2.49	2.01	6.96

Table 3 Comparative experimental results among predictive methods for 14-day strength test result.

Machine Learning Techniques	Average Result			
	R	RMSE (MPa)	MAE (MPa)	MAPE (%)
Training				
C&R tree	0.9983	0.22	0.12	0.45
Neutral net	0.8467	2.20	1.58	5.81
Regression	0.9180	1.58	1.21	4.53
SVM	0.8962	2.00	1.21	4.41
CHAID	0.8946	1.78	1.03	3.53
SOS-SVR	0.8765	1.85	1.14	5.00
Testing				
C&R tree	0.7516	4.19	3.35	12.87
Neutral net	0.8215	2.86	2.43	9.69
Regression	0.8289	2.86	2.21	8.63
SVM	0.8499	2.64	2.13	8.43
CHAID	0.8029	3.39	2.82	11.45
SOS-SVR	0.8563	2.51	2.07	7.26

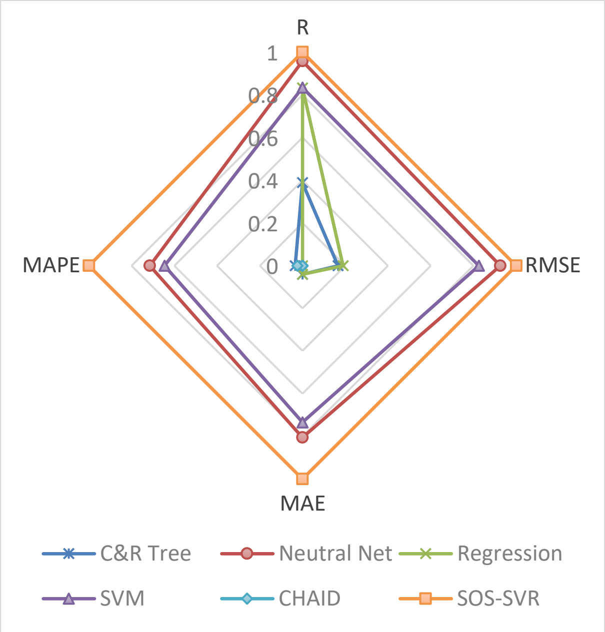

Figure 2. Radar plot of normalized accuracy measures (7-day strength test result)

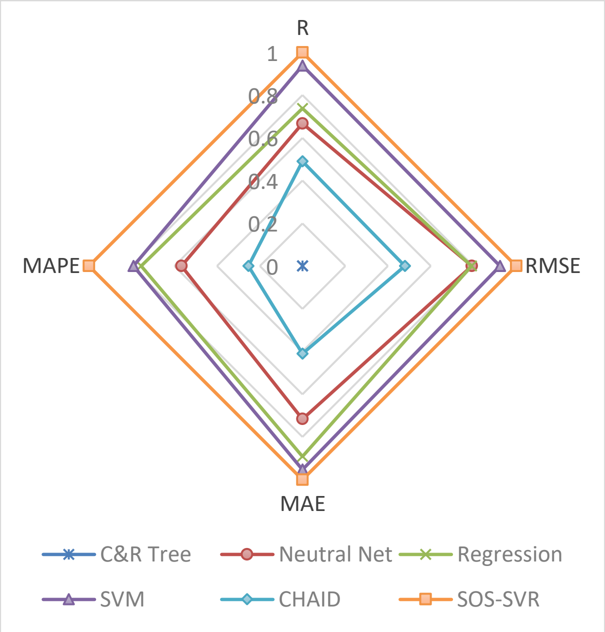

Figure 3. Radar plot of normalized accuracy measures (14-day strength test result)

Figures 2 and 3 further illustrate the comparative performance from each measurement category. Each value from average testing result is normalized into 0 to 1 in which 0 denotes the current worst performance and

1 denotes the current best performance. The outermost line from the figures denotes the best performer. It can be seen that the SOS-SVR prediction model has delivered the best performance in all categories in comparison with other machine learning techniques.

Figures 4 and 5 present a comparison of the actual and predicted values of the concrete strength at 7-day and 14-day test result, respectively. It shows the fit curve between the target and predicted output at training and testing stages of one fold. This further shows the effectiveness of SOS-SVR as the most reliable method for establishing the prediction model in this experiment.

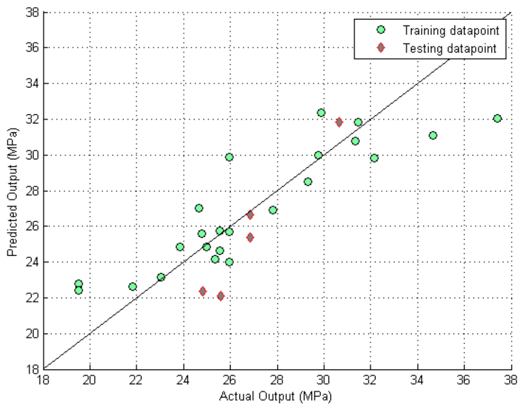

Figure 4. Actual and predicted concrete strength by SOS-SVR prediction model at training and testing stages for one fold (7-day strength test result).

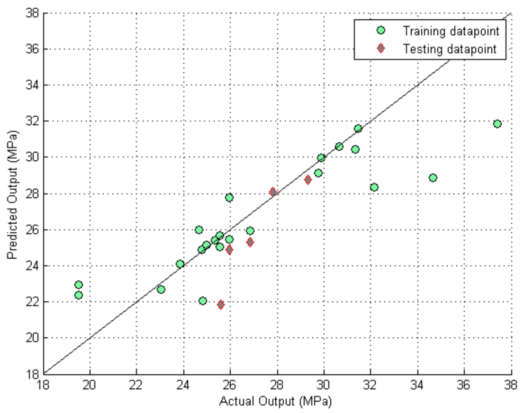

Figure 5. Actual and predicted concrete strength by SOS-SVR prediction model at training and testing stages for one fold (14-day strength test result).

5 Conclusion

This study developed a new prediction method called symbiotic organisms search-based support vector regression (SOS-SVR) to predict the strength of concrete mixtures from the early age test result. To investigate the accuracy of the proposed method, five machine learning techniques were used as benchmarks for the SOS-SVR. The experimental data set was acquired from laboratory test of 28 samples. In this investigation, the proposed predictive techniques were applied to the prepared training and testing datasets generated by 5-fold cross-validation.

The proposed SOS-SVR was further compared for performance outcomes by using four different performance measures (R, RMSE, MAE, and MAPE) to obtain a comprehensive comparison of the applied predictive techniques. The findings showed that the proposed SOS-SVR achieved the best accuracy for all performance measures.

This study presents a significant contribution to address the importance of the problem of early ages prediction of concrete strength. By accurately predicting the concrete behavior, the SOS-SVR assists the users and concrete designers in decision-making processes based on early strength test results. Analytical results indicate that SOS-SVR is the most reliable model for building accurate prediction behavior of concrete mix proportion at various early ages of strength.

Acknowledgement

The authors would express appreciation to the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia for the financial support through the "Bantuan Seminar Luar Negeri 2017" funding program.

References

[1] Kheder G. F., Gabban A. M. A., and Abid S. M. Mathematical model for the prediction of cement compressive strength at the ages of 7 and 28 days within 24 hours, *Materials and Structures*, 36(10): 693, 2003.

[2] Cheng M.-Y., Firdausi P. M., and Prayogo D. High-performance concrete compressive strength prediction using Genetic Weighted Pyramid Operation Tree (GWPOT), *Engineering Applications of Artificial Intelligence*, 29 104-113, 2014.

[3] Cheng M.-Y., Prayogo D., and Wu Y.-W. Novel Genetic Algorithm-Based Evolutionary Support Vector Machine for Optimizing High-Performance Concrete Mixture, *Journal of Computing in Civil Engineering*, 28(4), 06014003, 2014.

[4] Cheng M.-Y., Wibowo D. K., Prayogo D., and Roy A. F. V. Predicting productivity loss

caused by change orders using the evolutionary fuzzy support vector machine inference model, *Journal of Civil Engineering and Management*, 21(7): 881-892, 2015.

[5] Hoang N.-D. and Pham A.-D. Hybrid artificial intelligence approach based on metaheuristic and machine learning for slope stability assessment: A multinational data analysis, *Expert Systems with Applications*, 46 60-68, 2016.

[6] Cheng M.-Y., Prayogo D., Ju Y.-H., Wu Y.-W., and Sutanto S. Optimizing mixture properties of biodiesel production using genetic algorithm-based evolutionary support vector machine, *International Journal of Green Energy*, 13(15): 1599-1607, 2016.

[7] Suykens J. A. K. and Vandewalle J. Least Squares Support Vector Machine Classifiers, *Neural Process. Lett.*, 9(3): 293-300, 1999.

[8] Cheng M.-Y. and Prayogo D. Symbiotic Organisms Search: A new metaheuristic optimization algorithm, *Computers & Structures*, 139 98-112, 2014.

[9] Suykens J. A. K. and Vandewalle J. Least squares support vector machine classifiers, *Neural Processing Letters*, 9(3): 293-300, 1999.

[10] S. K. K., Kim D., S.L.Sekar, and Samui P. Model of least square support vector machine (LSSVM) for prediction of fracture parameters of concrete, *International Journal of Concrete Structures and Materials*, 5(1): 29-33, 2011.

[11] Cheng M.-Y. and Prayogo D. Modeling the permanent deformation behavior of asphalt mixtures using a novel hybrid computational intelligence. In *ISARC 2016 - 33rd International Symposium on Automation and Robotics in Construction*, pages 1009-1015, Auburn, USA, International Association for Automation and Robotics in Construction, 2016.

[12] Cheng M.-Y. and Hoang N.-D. Estimating construction duration of diaphragm wall using firefly-tuned least squares support vector machine, *Neural Computing and Applications*, 1-9, 2017.

[13] Mirjalili S. The Ant Lion Optimizer, *Advances in Engineering Software*, 83 80-98, 2015.

[14] Mirjalili S. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm, *Knowledge-Based Systems*, 89 228-249, 2015.

[15] Civicioglu P. Backtracking Search Optimization Algorithm for numerical optimization problems, *Applied Mathematics and Computation*, 219(15): 8121-8144, 2013.

[16] Cheng M. Y. and Prayogo D. Fuzzy adaptive teaching–learning-based optimization for global numerical optimization, *Neural Computing and Applications*, 1-19, 2016.

[17] Cheng M.-Y. and Prayogo D. A novel fuzzy adaptive teaching–learning-based optimization (FATLBO) for solving structural optimization problems, *Engineering with Computers*, 33(1): 55-69, 2017.

[18] Cheng M.-Y., Prayogo D., Wu Y.-W., and Lukito M. M. A Hybrid Harmony Search algorithm for discrete sizing optimization of truss structure, *Automation in Construction*, 69 21-33, 2016.

[19] Tejani G. G., Savsani V. J., and Patel V. K. Adaptive symbiotic organisms search (SOS) algorithm for structural design optimization, *Journal of Computational Design and Engineering*, 2016.

[20] Tran D.-H., Cheng M.-Y., and Prayogo D. A novel Multiple Objective Symbiotic Organisms Search (MOSOS) for time–cost–labor utilization tradeoff problem, *Knowledge-Based Systems*, 94 132-145, 2016.

[21] Cheng M.-Y., Chiu C.-K., Chiu Y.-F., Wu Y.-W., Syu Z.-L., Prayogo D., and Lin C.-H. SOS optimization model for bridge life cycle risk evaluation and maintenance strategies, *Journal of the Chinese Institute of Civil and Hydraulic Engineering*, 26(4): 293-308, 2014.

[22] Kamankesh H., Agelidis V. G., and Kavousi-Fard A. Optimal scheduling of renewable micro-grids considering plug-in hybrid electric vehicle charging demand, *Energy*, 100 285-297, 2016.

[23] Ayala H., Klein C., Mariani V., and Coelho L. Multi-objective Symbiotic Search Algorithm Approaches for Electromagnetic Optimization, *IEEE Transactions on Magnetics*, PP(99): 1-1, 2017.

[24] Panda A. and Pani S. A Symbiotic Organisms Search algorithm with adaptive penalty function to solve multi-objective constrained optimization problems, *Applied Soft Computing*, 46 344-360, 2016.

[25] Yu V. F., Redi A. A. N. P., Yang C.-L., Ruskartina E., and Santosa B. Symbiotic organisms search and two solution representations for solving the capacitated vehicle routing problem, *Applied Soft Computing*, 52 657-672, 2017.

[26] Cheng M. Y., Prayogo D., and Tran D. H. Optimizing Multiple-Resources Leveling in Multiple Projects Using Discrete Symbiotic Organisms Search, *Journal of Computing in Civil Engineering*, 30(3), 04015036, 2016.

[27] Prayogo D., Cheng M.-Y., and Prayogo H. A Novel Implementation of Nature-inspired Optimization for Civil Engineering: A Comparative Study of Symbiotic Organisms Search, *Civil Engineering Dimension*, 19(1): 36-43, 2017.

[28] Alavi A. H., Ameri M., Gandomi A. H., and Mirzahosseini M. R. Formulation of flow number of asphalt mixes using a hybrid computational method, *Construction and Building Materials*, 25(3): 1338-1355, 2011.