- Word Count: 5542

Plagiarism Percentage

10%

sources:

1 2% match (publications)

Cheng, Min-Yuan, and Nhat-Duc Hoang. "Risk Score Inference for Bridge Maintenance Project Using Evolutionary Fuzzy Least Squares Support Vector Machine", Journal of Computing in Civil Engineering, 2014.

2 1% match (publications)

Dong-Xiao Niu. "Mid-term load forecasting based on dynamic least squares SVMS", 2008 International Conference on Machine Learning and Cybernetics, 07/2008

3 1% match (publications)

Zhiyong Luo. "SVM parameters tuning with quantum particles swarm optimization", 2008 IEEE Conference on Cybernetics and Intelligent Systems, 09/2008

4 1% match (publications)

Vinaixa, M.. "A fuzzy ARTMAP- and PLS-based MS e-nose for the qualitative and quantitative assessment of rancidity in crisps", Sensors & Actuators: B. Chemical, 20050513

5 1% match (Internet from 03-Oct-2009)

http://www.paper.edu.cn/en/downloadpaper.php?serial_number=200908-313&type=1

6 1% match (Internet from 21-May-2016)

http://www.mep.cs.ubbcluj.ro/papers/asphalt_mep.pdf

7 1% match (publications)

Mirzahosseini, M.R.. "Permanent deformation analysis of asphalt mixtures using soft computing techniques", Expert Systems With Applications, 201105

8 < 1% match (publications)

José García-Nieto. "Noiseless functions black-box optimization", Proceedings of the 11th annual conference companion on Genetic and evolutionary computation conference - GECCO 09 GECCO 09, 2009

9 < 1% match (publications)

Cheng, Min-Yuan, Nhat-Duc Hoang, and Yu-Wei Wu. "Hybrid intelligence approach based on LS-SVM and Differential Evolution for construction cost index estimation: A Taiwan case study", Automation in Construction, 2013.

---

10 < 1% match (Internet from 04-Nov-2017)
http://nardus.mpn.gov.rs/bitstream/handle/123456789/5800/BacaninTeza.pdf?isAllowed=y&sequence=1

---

11 < 1% match (Internet from 06-Sep-2017)
http://amslaurea.unibo.it/12262/1/Mattia_Carmine.pdf

---

12 < 1% match (publications)
Alavi, Amir Hossein Ameri, Mahmoud Gando. "Formulation of flow number of asphalt mixes using a hybrid computational method.", Construction and Building Materials, March 2011 Issue

---

13 < 1% match (publications)
Cheng, Min-Yuan, Doddy Prayogo, and Duc-Hoc Tran. "Optimizing Multiple-Resources Leveling in Multiple Projects Using Discrete Symbiotic Organisms Search", Journal of Computing in Civil Engineering, 2015.

---

14 < 1% match (Internet from 25-Oct-2015)
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3274112/

---

15 < 1% match (publications)
Haitao Lv, Ruimin Hu, Jun Chen, Zheng He, Shihong Chen. "Protection Intensity Evaluation for a Security System Based on Entropy Theory", Entropy, 2013

---

16 < 1% match (publications)
Communications in Computer and Information Science, 2014.

---

17 < 1% match (publications)
Jones, Pauline A.. "Mediational behavior and linguistic development [1]", Paper in Linguistics, 1976.

---

18 < 1% match (publications)
Taleizadeh, A.A.. "Meta-heuristic algorithms for solving a fuzzy single-period problem", Mathematical and Computer Modelling, 201109

---

19 < 1% match (Internet from 14-Feb-2017)
http://www.mdpi.com/1424-8220/17/2/307/pdf

---

20 < 1% match (Internet from 07-Dec-2013)
http://ecmj.org.cn/ch/reader/view_abstract.aspx?file_no=201231555588090&flag=1

---

21 < 1% match (Internet from 07-Mar-2016)
http://ethesis.nitrkl.ac.in/1550/1/50609003.pdf

22 < 1% match (Internet from 04-Dec-2016)
http://www.wseas.us/e-library/conferences/digest2003/papers/461-160.pdf

23 < 1% match (publications)
Theodore B. Trafalis. "Learning networks in rainfall estimation", Computational Management Science, 07/2005

24 < 1% match (publications)
ZhuGang Yuan. "Soft sensor for apparent degree of calcination in NSP cement production line", 2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE), 02/2010

25 < 1% match (publications)
Wang, Xiaodong, Meiying Ye, Yimo Zhang, and Yukihiro Ishii. "", Advanced Materials and Devices for Sensing and Imaging II, 2005.

26 < 1% match (publications)
Gao, Ying, Xiaoming Huang, Zhaohui Qi, and Jiupeng Zhang. "Characteristics and Prediction of Permanent Deformation in HMA", Performance Modeling and Evaluation of Pavement Systems and Materials, 2009.

27 < 1% match (publications)
"New Learning Paradigms in Soft Computing", Springer Nature, 2002

paper text:

Prediction of Permanent Deformation in Asphalt Pavements Using a Novel Symbiotic Organisms Search – Least Squares Support Vector Regression Abstract: The prediction of asphalt performance can be very important in terms of increasing service life and performance while saving energy and money. In this study, a new hybrid artificial intelligence (AI) system, SOS-LSSVR, has been proposed

to predict the permanent deformation potential of **asphalt pavement** 26

mixtures. SOS- LSSVR utilizes the symbiotic organisms search (SOS) and the least squares support vector regression (LSSVR), which are seen as a complementary system. The prediction

model can be established from all **input and output data** 24

pairs for LSSVR, while SOS optimizes the system's tuning parameters. To avoid sampling bias and to partition the dataset into testing and training, a cross-validation technique was chosen. The results can be compared to those of previous studies and other predictive methods. Through the use of four error indicators, SOS-LSSVR accuracy was verified in predicting

the permanent deformation behavior of an asphalt mixture. 7

The present study demonstrates that the proposed AI system is a valuable decision-making tool for road designers. Additionally, the success of SOS-LSSVR in building an accurate prediction model suggests that the proposed self-optimized prediction framework has found an underlying pattern in the current database and, thus, can potentially be implemented in various disciplines. Keywords: Asphalt Mixtures; Artificial Intelligence; Permanent Deformation; Least Squares Support Vector Regression; Symbiotic Organisms Search 1. INTRODUCTION Over the past decades, the number of vehicles on the road has significantly increased, causing deformation in pavement. Accumulated traffic load repetition permanently deforms asphalt pavement [1]. The side effects of permanent deformation can be devastating, and include a reduction in the service life of the pavement and the creation of risky conditions for roadway users [2]. Increasing the thickness of asphalt pavement is one possible solution for road designers. However, it is often abandoned due to budget limitations. Predicting the appropriate asphalt mixtures may increase performance and service life at little additional construction cost. However, establishing a model that accurately depicts the relationship between asphalt mixtures and permanent deformation is a complicated task because of the dynamic and complex characteristics of asphalt mixtures. There has been growing interest in the development of artificial intelligence (AI), particularly in predictive techniques due to their excellent learning features [3]. The main idea behind predictive approaches in AI is to develop a prediction model from a collection of input-output data pairs using a specific learning procedure. Once trained, the prediction model can forecast with high accuracy and handle non-linear problems. As a result, much effort has been invested in improving the computational time and accuracy of AI predictive approaches [4-8]. The laboratory dynamic creep test and its flow number are commonly used as indicators that an asphalt mixture has a permanent deformation [9]. In the dynamic creep test result, the flow number can be determined as the load cycle number at which the tertiary deformation begins [10]. The tertiary deformation denotes the phase at which the progressive permanent deformations accelerate and permanent deformations grow rapidly. Determining the flow number within asphalt mixtures requires various AI approaches. For example, Gandomi et al. [11] built prediction models using an AI method called gene expression programming (GEP) by collecting

dynamic creep test samples. Alavi **et al.** [12] utilized **the** 12

genetic programming-simulated annealing (GP/SA) 12

method to build prediction models for asphalt mixtures' performance. Mirzahosseini et al. [13] subsequently used two artificial neural network (ANN) models, multilayer perceptron (MLP) and multi expression programming (MEP), to investigate 4 asphalt pavement performance. These studies showed AI predictive techniques' strong potential to deal with the difficult input-output relationship of asphalt mixtures. Despite the effective performance of the AI approaches that have been reported, previous studies in this area have made limited use of AI techniques. Furthermore, these studies have used only a simple, random division of training and testing sets in the validation process. A more advanced validation method is necessary to eliminate the potential for bias in dividing data points between these two sets. To that end, a serious need exists for more accurate systems in estimating the flow number of asphalt mixtures. The present study proposes an AI system called SOS-LSSVR to predict permanent deformation in asphalt pavement. SOS-LSSVR integrates an accurate prediction technique, least squares support vector regression (LSSVR), with a new nature-inspired optimization technique, symbiotic organisms search (SOS). With the use of radial basis function kernel (RBF), LSSVR is considered an effective AI technique when dealing with prediction problems [14-16]. To improve the modeling performance, LSSVR needs two tuning parameters set correctly: the

regularization parameter (?) and the kernel parameter (?). The selection process of **parameters** can **be** 3

formulated as an optimization problem. As a new nature-inspired algorithm, SOS is considered a powerful and effective continuous-based global optimization method [4]. In previous research, experiments showed that SOS was superior to other nature-inspired techniques [4,17-22]. Nevertheless, the algorithm's capability has not yet been tested in terms of obtaining the best LSSVR parameters. The proposed method is investigated alongside other predictive techniques in terms of its efficacy as a viable prediction model for asphalt mixtures and their permanent deformation potential. The proposed method will use cross-validation, allowing for the validation of the training and testing process. Furthermore, four different measures are employed to judge the accuracy of each prediction model. Obtained results are then compared with those of previous studies. 2. LITERATURE REVIEW 2.1

.Least Squares Support Vector Regression (LSSVR) Considered **an alternative** to **the** 23

support vector machine (SVM), LSSVR is employed for regression analysis and solving the function estimation. Adopting a statistical learning theory, this AI method focuses on replacing the quadratic program with

a least squares **linear system as** its **loss function** [23]. The formulation **of the** optimization **problem** and **the** 14

constraints for LSSVR are shown as follows: Minimize (,) = 12 + 12 ∑=1 2 (1)

Subjected to = () + + , = 1, … , 6 (2) **where ek R** denote slack **variable,** 1

γ > 0 is **a regularization constant, and** 1

() denotes an input mapping to

a higher dimensional feature space. The Lagrangian is 25

given by:

L(w,b,e;?) ? Jp (w,e) ? ??k{ wT ?(xk) ? b ? ek ? yk} N (1) k ?1 where ? k are Lagrange multipliers. The conditions for optimality 3

are given by: ? ?L N ??

?w ?0?w???k?(xk) k ?1 ? ?L ? 0 ? ? ? k ? 0 N ? ?b ? k ?1 (2) ? ?L ? ?ek ? 0 ? ? k ? ?ek , k ? 1,..., N ? ? ?????Lk ? 0 ? wT?(xk) ? b ? ek ? yk ? 0, k ? 1,...N 2

After

elimination of e and w, the following linear system is obtained: ?0 ? 1Tv ? ? ? I / ? ? ? ?b ? ?0 ? ?1v ??? ?? = ?? y?? (3) where y 2

? y1,..., yN , 1v ? [1;...;1] , and ? ? [?1;...;? N] . The following formula represents the kernel function: ? ? ? (xk)T ? (xl) ? K (xk , xl) The resulting LSSVR model may be stated as: () = ∑=1

(,) + (4) (7) where αk and b are the solution to the linear system. The RBF kernel is the 1

most frequently used kernel function. The RBF may be expressed as: (,) = (− ‖2−2‖2)

(8) where σ is the kernel function parameter. While using the RBF kernel, 1

LSSVR needs two parameters:

the regularization parameter () and the kernel parameter (). While the 3

parameter impacts

the smoothness of the regression function, 1

the parameter takes control of all penalties imposed on

data points that deviate from the regression function. 1

2.2.Symbiotic Organisms Search (SOS) Many areas of research employ nature-inspired algorithms for the most complex optimization issues [24,25]. Introduced by Cheng and Prayogo [4], SOS emerges as a newly promising nature-inspired algorithm.

In the search for the optimal global solution, the 4

attempt is to reach promising areas by simulating all symbiotic interactions that move an ecosystem of organisms. Within the ecosystem, each organism receives a certain fitness value that reflects the level of adaptation to the objective. With SOS, the main searching strategy is divided into three phases: mutualism, commensalism, and parasitism. The developed searching strategy simulates the three types of actual symbiotic interactions that occur in the real world. With mutualism, all interactions between organisms are mutually beneficial. The commensalism phase sees one organism benefit, and another have no impact at all. Finally, one organism benefits while the other suffers from parasitism. The detailed structure of the SOS algorithm is explained in Algorithm 1. <Insert Algorithm 1> Since its publication, SOS has been increasingly used in a variety of research fields [18- 20,26-32]. Today, the algorithm has huge potential in the ever-growing search for optimality. 2.3.Hybridization of prediction and optimization approaches In recent years, the collaborative integration between the prediction method and optimization technique has been studied extensively. The prediction methods learn from the given data inputs and outputs until an underlying pattern exists. However, some modeling techniques require advanced parameter settings to produce an acceptable level of accuracy [33]. Many studies have utilized optimization techniques to find suitable parameters so that the prediction methods can determine the complicated input and output relationship and, thus, increase their accuracy. Table 1 summarizes the recent studies for hybridizing the prediction method with the optimization technique. <Insert Table 1> 9 3. THE SYMBIOTIC ORGANISMS SEARCH - LEAST SQUARES SUPPORT VECTOR REGRESSION (SOS-LSSVR) As a hybrid system, SOS-LSSVR integrates two computational intelligence methods with the LSSVR accurately portraying the input/output relationship as a predictor and with the SOS optimizing all LSSVR parameters ensuring the highest level of accuracy. Figure 1 explains the framework of SOS-LSSVR. <Insert Fig. 1> For SOS-LSSVR, there are eight key steps when used across training and testing phases: 1) Training data Training data is used for creating the prediction model. To prevent greater numeric ranges of input variables from dominating the process, the data was normalized into a (0,1) range [34]. 2) LSSVR training With a hybrid system, the complex relationship between output and input variables is addressed by LSSVR. This learning process requires two tuning parameters, γ and ? parameters. Within the boundary range, the parameters are initialized randomly for the first iteration. As the optimizer, SOS simulates the searching for the best tuning parameters, allowing the LSSVR to then build the prediction model with higher accuracy. 3) SOS searching SOS is used to test the many combinations of both parameters that allow for the best set to be found. The generation of the population that best represents the candidate solution allows the search process to begin (consisting of both parameters). Utilizing all three phases – mutualism, commensalism, and parasitism – the fitness value of each solution will gradually improve. 4) Fitness evaluation LSSVR can have a low accuracy when predicting a new and unseen dataset despite its solid performance on all training data. This issue is known as the over-fitting problem [35]. To overcome this problem, the training data was separated into learning subsample and validation subsamples. The learning subsample was used for building the prediction model. The validation subsample has no rule in building the actual model. However, it was used for supporting the generalization capability. To avoid the sampling bias, the

10-fold cross-validation technique **was used to** split **the** 20

training data into smaller subsamples. The prediction model with the highest accuracy is determined based on the combination of two γ and ? tuning parameters that has the lowest error on the validation subsample. An objective function is now developed based on the model accuracy in predicting the validation subsample. The root mean square error (RMSE) is used to represent model accuracy in the objective function, as shown in Eq. 5. S Min Fitness Value = k?1 ? RMSE(validationk) S (5) where S indicates the total number of folds and RMSE(validationk) indicates the value of root mean squared error between the actual and predicted values for the k- th validation subsample. 5) Termination criteria Once the stopping conditions have been met, the process terminates or else proceeds to the next iteration. The total number of SOS iterations was used as the termination criterion. 6) Optimal LSSVR model and parameters As soon as the termination criteria have been met, the loop will come to a complete stop and this suggests the

prediction model has found **the** ideal **input-output mapping relationship** along with **the optimal parameters.** 1

7) LSSVR predicting With the two parameters at the optimum level obtained from the training phase, the prediction model can be established and then used to predict all test data. 8) Testing data Finally, to measure the general accuracy and the prediction performance, the testing data is applied to the trained model. 4. EXPERIMENTAL RESULTS 4.1. Historical Dataset In this study, 118 dynamic creep test samples from a laboratory test were used, and this allowed a prediction to be made of the proposed solution's performance [11,12]. The dataset included 10 different input variables (influencing factors) as well as one output variable. All statistical descriptions of input and output variables are described in Table 2. The historical dataset is

shown in Table 3. <Insert Table 2><Insert Table 3> 17

The dataset was employed for modeling the asphalt pavement performance in [13,12,36,11]. It was revealed that the previous studies have only used a partial amount of all possible input variables. Table 4 lists all previous models that have been employed

to predict the flow number of asphalt mixtures. 6

<Insert Table 4> In Case 1, Alavi et al. [12] and Mirzahosseini et al. [13] used IF3, IF4, IF6, and IF10 as their

input variables. Gandomi et al. [11] employed IF1, IF5, IF6, and IF10 as the input variables in Case 2. Meanwhile, Mirzahosseini et al. [36] utilized IF1, IF3, IF4, IF5, IF6, and IF10 as the input variables in Case 3. 4.2.Experimental Settings To benchmark the performance of SOS-LSSVR, three different widely used predictive techniques were employed, including SVR [37], LSSVR [14], and BPNN [38]. The SVR and LSSVR methods belong to SVM class. Meanwhile, BPNN modifies the ANN by regulating the connection weights and bias values using back-propagation algorithm throughout the training process. This study uses a default set for all parameters to ensure a fair comparison. All parameters for SOS-LSSVR, SVR, LSSVR, and BPNN are listed in Table 5. Four performance measures were used during the evaluation process for AI-based predictive methods throughout this research, as shown in Table 6. To evaluate all predictive methods, these performance measures were used, which allowed for more accurate results and a fairer test all around. <Insert Table 5><Insert Table 6> The historical dataset was now separated into training and testing dataset. Previously, Gandomi

et al. [11], Alavi **et al. [12], and** Mirzahosseni **et al. [13],** 16

used approximately 75% of the dataset for training and 25% of the dataset for testing. To ensure the same proportion of training and testing data as the previous works, four-fold cross-validation was selected. The dataset was split into four folds, which assigns the 3/4 (or 75%) portion of the dataset for training and assigns the remaining portion for validating the prediction model. A total of 4 distinct sets of training and testing data were performed. By using a four-fold cross-validation method for each model, the results were obtained, and they were based on average results for the testing and training datasets. Compared to considered models, cross-validation allowed the best validation capabilities, and this allowed the study to apply all training and testing datasets in both phases. 4.3.LSSVR-SOS Training Process and Prediction Results As mentioned previously, SOS simulates the nature-inspired searching strategies to find the combination of LSSVR tuning parameters that produces the lowest fitness value (training error) during the training process. To ensure that the learning model is accurately generated, the

k-fold cross-validation method was **used. In the** beginning, the **k-fold cross-validation** separates **the** 11

dataset randomly into the

training and testing data. The **training data** 27

is employed to build the prediction model while the testing data is treated as unseen data for verifying the trained model. To avoid the over-fitting, SOS- LSSVR also utilizes

k-fold cross-validation to divide **the** training **data into** the learning **and** 19

validation subsamples 10 times, where each subsample is used as a validation subsample. The learning parameters for the SOS-LSSVM was selected as follows: (1) population size for the decision variables in

SOS is 50; (2) the algorithm stops after 100 iterations; (3) the searching range for γ and ?? tuning parameters begins from 10-10 to 1010 [39]. In this study, every fitness value of LSSVR tuning parameters is determined using the objective function formulated in Eq. 5. The convergence curves of SOS searching were illustrated in Fig. 2. As shown in Fig. 2, the SOS searching improved the fitness value quickly from the starting iteration. The fitness value converged after several iterations, indicating that no further improvement of the fitness value can be obtained. It can be seen that the SOS delivers a great performance as the optimizer in this system. <Insert Fig. 2> Table 7 displays the performance of SOS-LSSVR for each fold on each dataset. Table 8 shows the complete statistical comparative results of the experiment among the predictive methods. These results show that SOS-LSSVR performed better compared to the rest of predictive methods. In each dataset, SOS-LSSVR earned the best score in overall measurement category (R, RMSE, MAPE, and MAE), followed by the BPNN, LSSVR, and SVR. Fig. 3 depicts the performance measures that are described in Table 8. <Insert Table 7><Insert Table 8><Insert Fig. 3> Among three models of dataset, SOS-LSSVR achieved the best overall performance in Case 3. In Case 3, SOS-LSSVR produces the lowest RMSE, MAPE, and MAPE scores of 34.35, 14.47%, and 24.80, respectively, while having the highest R score of 0.9713. To conclude, Case 3, the model with 6 input variables (IF1, IF3, IF4, IF5, IF6, IF10), enables SOS-LSSVR to build the most accurate model for predicting the flow number. 4.4.Comparison with Previous Works Numerous studies have proposed AI methods for estimating the flow number of asphalt mixtures. For further verification, the prediction results of the proposed SOS-LSSVR 17 were compared with those of previous works. Generally, it was not possible

to compare the performance of the proposed method with **the** **21**

previous works because the data divisions for training and testing were different. As discussed previously, this study employed four-fold cross-validation to keep the same proportion of training and testing ratio (75/25) with the previous research. Table 9 summarizes the comparison results between the proposed method and previous works. <Insert Table 9> In Case 1, SOS-LSSVR outperforms GP/SA and MEP in all performance measure categories (R, RMSE, and MAE). MEP has slightly better R and MAE scores compared with GP/SA while GP/SA is better than MEP in terms of RMSE. Overall, error rates (RMSE and MAE) improved by SOS-LSSVR method were 14.0% - 17.4% compared to those of previous methods in this case. Similar to Case 1, the obtained results of the SOS-LSSVR performance are better than those of GEP in every category in Case 2. The error rates of SOS-LSSVR were 13.2% - 16.5% lower than those of GEP. In Case 3, the SOS-LSSVR and ANN produced better performance than LGP. SOS-LSSVR has the best score in terms of RMSE while ANN has the best score in terms of R and MAE. 5. CONCLUSION Permanent deformations in asphalt pavement have become a major issue in road engineering because it creates discomfort and often dangerous situations to the road users. Permanent deformations usually occur after a number of repeated loading cycles, known as flow number, applied to an asphalt pavement. Accurately predicting the flow number is essential for road designers in determining the proper asphalt binder properties. Thus, the present study developed a new predictive method called SOS- LSSVR to model the complex relationship of asphalt mixtures and predict their permanent deformation. The dataset used in this study was obtained from a dynamic creep test containing 118 samples. All proposed predictive techniques used cross-validation through the varying dataset models. As a benchmark, three different predictive methods were used for SOS- LSSVR: SVR, BPNN, and LSSVR. The proposed SOS-LSSVR was compared with other methods through multiple performance measures to build an extensive comparison of the predictive methods. In this study, the SOS-LSSVR is able to achieve better accuracy than all other comparative measures with the BPNN, LSSVR, and SVR achieving the second-, third-, and fourth-best overall accuracies, respectively. Furthermore, the results from SOS- LSSVR are compared with those of past research. It was revealed that the results from SOS-LSSVR outperforms those of previous predictive methods. The present study validates that the new predictive model SOS-LSSVR represents a significant step forward in assisting road designers in addressing the critical problem of permanent deformation in asphalt mixtures. Investigating the selection of relevant input factors of the given dataset represents an interesting direction for further study. Choosing a set of relevant input factors may increase the model performance and reduce the model complexity.

CONFLICT OF INTEREST The authors declare that they have **no conflict of interest.** Algorithm **1.** 15

Pseudo-code of SOS algorithm Input: n: population size UB: upper bound of the solution D: dimensions of problem LB: lower bound of the solution itermax: maximum number of iterations F(X): objective function 1: Generate an initial population X={X1, X2, … , Xn}, and evaluate its fitness 2: Identify the best solution of the initial population Xbest 3: iter = 1 to itermax 4: i = 1 to n 5: / / 6: Choose randomly index ii {1,2, … , n}, which ii ≠ i 7:

BF1 = (1 + round(rand(0,1)) 8: **BF2 = (1 + round(rand(0,1))** 13

9: Xi + Xii Xmutual = (2) 10: j = 1 to D 11:

X′i[j] = Xi **[j] + rand(0,1)** (Xbest **[j]** − BF1 Xmutual **[j])** 10

12: Xii[j] = Xii[j] + rand(0,1) (Xbest[j] − BF2 Xmutual

[j]) 13: 14: F(X′i) < F(Xi) 15: Xi = **X′i** 8

16: 17: F(Xii) < F(Xii) 18: Xii = X′ii 19: 20: / / 21: Choose randomly index ii {1,2, … , n}, which ii ≠ i; 22: j = 1 to D

23: X′i[j] = Xi **[j] + rand(−1,1)** (Xbest **[j]** − Xii **[j])** 18

24:

25: **F(X′i) < F(Xi)** 26: Xi = **X′i** 8

27: 28: / / 29: Choose randomly index ii {1,2, … , n}, which ii ≠ i 30: j = 1 to D 31: rand(0,1) < rand(0,1) 32: Xparasite[j] = Xi[j] 33: 34: Xparasite

[j] = rand(0,1)   (UB[j] − LB[j])   10

+ LB[j] 35:   36:   37: F(Xparasite) < F(Xii) 38: Xii = Xparasite 39:   40: Update the best solution of the current population Xbest 41:   42:   Output: the final best solution of the population Xbest Training 3 SOS searching ?, ? 1 2 4 Training dataset LSSVR train Fitness evaluation 5 Termination criteria NO YES Testing 8 7 6 Prediction Optimal LSSVR model Results Testing dataset LSSVR predict and parameters ? & ? Fig. 1. SOS-LSSVR architecture. Dataset-1 48

Fold 1 46 Fold 2 Fold 3 Fold 4   4

44 fitness value 42 40 38 36 34 32 10 20 30 40 50 60 70 80 90 100 number of iterations Dataset-2 60

Fold 1 58 Fold 2 Fold 3 56 Fold 4   4

fitness value 54 52 50 48 46 44 10 20 30 40 50 60 70 80 90 100 number of iterations Dataset-3 42

Fold 1 Fold 2 40 Fold 3 Fold 4   4

38 fitness value 36 34 32 30 10 20 30 40 50 60 70 80 90 100

number of iterations Fig. 2. Convergence curves of SOS in the   22

training process. Average R Average RMSE SVR LSSVR BPNN SOS-LSSVR SVR LSSVR BPNN SOS-LSSVR

Case 1 Case 1 Case 2 Case 2 Case 3 Case 3 0. 50 0. 60 0. 70 0.   5

80 0.90 1.00 25.0 50.0 75.0 100.0 125.0 150.0 Average MAPE (%) Average MAE SVR LSSVR BPNN SOS-LSSVR SVR LSSVR BPNN SOS-LSSVR

Case 1 Case 1 Case 2 Case 2 Case 3 Case 3 0.   5

0 25.0 50.0 75.0 100.0 125.0 150.0 20.0 40.0 60.0 80.0 100.0 120.0 Fig. 3. Average testing results of the performance measures for the SOS-LSSVR and other methods through cross-validation. Table 1. Summary of recent studies for hybrid prediction-optimization method. Previous Work Description Techniques Prediction of flow number of Prediction: Genetic Programming Gandomi et al. [11] asphalt mixtures Optimization: Simulated Annealing Prediction of flow number of Prediction: Genetic Programming Alavi et

al. [12] asphalt mixtures Optimization: Simulated Annealing Prediction: Radial Basis Function Neural Prediction of Construction Cao et al. [40] Network Cost Index in Taiwan Optimization: Artificial Bee Colony Prediction: Support Vector Machine Prediction of groutability Hoang et al. [6] estimation of grouting process Optimization: Flower Pollination Algorithm Prediction: Least-Squares Support Vector Prediction of fiber-reinforced Chou, Ngo [41] Regression soil Optimization: Smart Firefly Algorithm Prediction: Least-Squares Support Vector Prediction of rainfall-induced Tien Bui et al. [42] Machine shallow landslides Optimization: Differential Evolution Table 2. Input/output variables

and statistical descriptions. Variable Definition Min Max **Average** SD **IF1** 9

Percentage of coarse aggregate (%) IF2 Percentage of fine aggregate (%) IF3

Percentage of filler (%) IF4 **Percentage of bitumen** (%) IF5 **Percentage of** air **voids** 6

(%) IF6

Percentage of voids in mineral aggregate (%) IF7 **Marshall stability** (kN) IF8 **Marshall flow** 6

(mm) IF9

Coarse aggregate to fine aggregate ratio 7

IF10

Marshall stability to flow ratio / Marshal **quotient** 7

Output Flow number 33 81 18 57 1 10 4 7 1.71 8.77 13.20 19.04 2.73 15.3 2.1 4.75 0.58 4.5 0.61 4.81 22 510 57.31 37.15 5.54 5.51 4.54 16.55 10.16 3.50 1.84 2.99 227 14.33 11.31 3.17 0.81 1.52 1.41 2.04 0.62 1.05 0.74 143.97

Table 3. Historical dataset **No. IF1 IF2 IF3 IF4 IF5 IF6 IF7 IF8 IF9 IF10** 9

Output 1 55 2 55 3 55 4 55 5 55 6 55 … … 112 68 113 68 114 68 115 68 116 68 117 68 118 68 38 7 4 38 7 4 38 7 4.5 38 7 4.5 38 7 5 38 7 5 … … … 30 2 6 30 2 6 30 2 6.5 30 2 6.5 30 2 6.5 30 2 7 30 2 7 7.69 16.3 7.52 16.16 5.6 15.45 5.67 15.51 4.55 15.54 4.08 15.12 … … 3.42 16.55 4 17.05 3.46 17.59 3.36 17.51 3.02 17.21 3.36 18.49 2.74 17.97 11.74 3.27 9.49 2.9 11.58 3.4 11.42 3.72 11.38 3.73 12.88 3.8 … … 9.57 3.3

9.71 3.4 9.12 3.48 9.22 3.25 9.55 3.36 9.01 3.51 8.24 3.37 1.4474 1.4474 1.4474 1.4474 1.4474 1.4474 … 2.2667 2.2667 2.2667 2.2667 2.2667 2.2667 2.2667 3.5902 3.2724 3.4059 3.0699 3.0509 3.3895 … 2.9000 2.8559 2.6207 2.8369 2.8423 2.5670 2.4451 260 350 300 310 310 340 … 60 55 50 60 60 50 45 Table 4. Previous models for

Model Previous works No. of input variables List of input variables Case 1 GP/SA [12] 4 IF3, IF4, IF6, IF10 MEP, MLP [13] Case 2 GEP [11] 4 IF1, IF5, IF6, IF10 Case 3 LGP, ANN [36] 6 IF1, IF3, IF4, IF5, IF6, IF10 Table 5. Tuning Parameters of the competing predictive methods. AI method Parameters Setting reference SVR - Regulation parameter C = 1 - RBF kernel parameter ? = 1/N LSSVR - Regulation parameter ? = 1 - RBF kernel parameter ? = 1 BPNN - Training algorithm = Levenberg-Marquardt - Maximum number of iterations = 1000 - Initial ? = 0.01 - ? decrease factor = 0.1 - ? increase factor = 10 - Maximum ? = 1010 - ? searching boundary = 10-8 – 108 SOS-LSSVR - ? searching boundary = 10-5 – 105 - Population size = 25 - Maximum number of iterations = 100 [34] [14] [43] N is number of input variables, ? is learning rate of BPNN Table 6. Performance measures. Performance measure Formula Coefficient of correlation (R) = ∑.′−(∑)(∑′) √(∑2)−(∑)2√(∑′2)−(∑′)2 Root mean squared error (RMSE) = √1 ∑=1(−′)2 Mean absolute percentage error (MAPE) = 1 ∑=1 | −′ | Mean absolute error (MAE) = 1 ∑=1| −′| is the actual value; ′ is the predicted value; and is the number of data samples. Table 7. Summary of the cross-validation results of the proposed SOS-LSSVR over various models. Model No. of Training dataset Testing dataset Optimal Parameters fold R RMSE MAPE (%) MAE R RMSE MAPE (%) MAE ? ? Case 1 Case 2 Case 3 1 0.9723 2 0.9759 3 0.9699 4 0.9804 1 0.9608 2 0.9536 3 0.9702 4 0.9670 1 0.9897 2 0.9866 3 0.9902 4 0.9863 34.10 12.71 31.12 13.14 34.55 13.66 28.07 12.27 40.79 42.48 35.15 36.28 20.47 22.70 20.45 23.77 23.92 21.98 23.93 20.52 20.59 30.40 18.89 30.30 17.34 26.30 17.21 26.87 8.22 14.85 10.98 17.08 7.32 15.01 11.32 17.87 0.9815 0.9533 0.9778 0.9557 0.8952 0.9356 0.9020 0.9435 0.9740 0.9759 0.9670 0.9681 29.30 17.95 47.66 34.12 32.79 12.22 44.99 15.49 58.57 54.82 60.96 51.44 33.93 34.62 33.48 35.35 23.91 34.72 23.04 30.19 30.04 45.26 44.58 44.59 26.11 41.08 20.06 36.46 12.69 22.88 14.85 25.12 14.59 26.20 15.77 24.99 189827.5 11.3 1706709.6 19.4 440099.6 15.3 423084.1 10.8 18.2 12.3 121.0 21.3 127896.4 100000000.0 11418.1 14776685.0 2.6 2.6 2.9 2.3 22.4 92.6 14.6 80.1 37 Table 8. The comparative testing results between SOS-LSSVR with other predictive methods over various models. Model AI methods R RMSE MAPE (%) MAE Best Worst Average Best Worst Average Best Worst Average Best Worst Average Case 1 SOS-LSSVR LSSVR BPNN SVR 0.9815 0.9460 0.9700 0.9249 0.9533 0.9199 0.9385 0.8564 0.9671 0.9312 0.9589 0.9002 29.30 49.59 38.38 135.09 47.66 68.85 53.65 137.50 38.68 60.98 43.56 136.54 12.22 40.65 16.69 108.57 34.12 70.70 32.18 177.44 19.94 51.37 22.17 139.45 Case 2 SOS-LSSVR LSSVR BPNN SVR 0.9435 0.8457 0.9390 0.8156 0.8952 0.6057 0.8181 0.6095 0.9191 0.7497 0.8821 0.7359 Case 3 SOS-LSSVR 0.9759 0.9670 0.9713 LSSVR 0.9491 0.8694 0.9091 BPNN 0.9499 0.8711 0.9241 SVR 0.8707 0.8008 0.8406 51.44 91.33 54.69 130.16 33.48 66.38 43.13 134.04 60.96 104.30 78.41 148.90 35.35 77.30 80.39 151.91 56.45 97.24 67.69 140.64 34.35 69.33 56.71 141.47 20.06 70.05 22.93 106.72 12.69 33.14 24.81 113.71 44.58 114.57 48.40 195.53 15.77 76.21 43.96 163.32 30.20 81.40 32.78 142.96 14.47 57.11 32.33 145.62 23.04 42.25 28.89 112.34 36.46 77.23 40.10 111.68 22.88 51.26 35.34 110.60 34.72 55.57 36.65 121.05 45.26 87.90 60.39 130.19 26.20 61.75 61.97 134.63 27.96 49.86 33.09 117.80 41.85 80.11 50.59 123.14 24.80 55.60 43.42 122.43 Bold text denotes the best performance across the methods 38 Table 9. The average testing results between SOS-LSSVR with previous researches over various datasets. Model AI methods R RMSE MAE Case 1 MEP [13] 0.956 GP/SA [12] 0.948 SOS-LSSVR 0.9671 46.23 46.06 38.68 32.509 33.842 27.96 Case 2 GEP [11] SOS-LSSVR 0.891 0.9191 67.63 56.45 48.218 41.85

Case 3 LGP [36] ANN [36] SOS-LSSVR 0.964 0.974 0.9713 38.44 34.95 34.35 26.442 23.102 24.80 Bold text denotes the best performance across the methods 2 3 5 7 8 10 11 12 13 14 15 16 18 19 20 27 28 29 30 31 32 33 34 35 36 39