参考著作No.15

Measurement
Volume 112, December 2017, Pages 141-149

Prediction of chloride diffusion in cement mortar using Multi-Gene Genetic Programming and Multivariate Adaptive Regression Splines

Nhat-Duc Hoang [a], Chun-Tao Chen [b], Kuo-Wei Liao [c]

Measurement 112 (2017) 141–149

参考著作No.15

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Prediction of chloride diffusion in cement mortar using Multi-Gene Genetic Programming and Multivariate Adaptive Regression Splines

Nhat-Duc Hoang[a,*], Chun-Tao Chen[b], Kuo-Wei Liao[c]

[a] Faculty of Civil Engineering, Institute of Research and Development, Duy Tan University, P809K7/25 Quang Trung, Danang, Viet Nam
[b] Dept. of Civil and Construction Engineering, National Taiwan University of Science and Technology, No. 43, Sec. 4, Keelung Rd., Da'an Dist., Taipei 106, Taiwan
[c] Department of Bioenvironmental Systems Engineering, National Taiwan University, No.1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan

ARTICLE INFO

Keywords:
Chloride diffusion
Cement mortar
Machine learning
Modeling equation
Construction material

ABSTRACT

Chloride-induced damage of coastal concrete structure leads to serious structural deterioration. Thus, chloride content in concrete is a crucial parameter for determining the corrosion state. This study aims at establishing machine learning models for chloride diffusion prediction with the utilizations of the Multi-Gene Genetic Programming (MGGP) and Multivariate Adaptive Regression Splines (MARS). MGGP and MARS are well-established methods to construct predictive modeling equations from experimental data. These modeling equations can be used to express the relationship between the chloride ion diffusion in concrete and its influencing factors. Moreover, a data set, which contains 132 cement mortar specimens, has been collected for this study to train and verify the machine learning approaches. The prediction results of MGGP and MARS are compared with those of the Artificial Neural Network and Least Squares Support Vector Regression. Notably, MARS demonstrates the best prediction performance with the Root Mean Squared Error (RMSE) = 0.70 and the coefficient of determination (R^2) = 0.91.

1. Introduction

The durability of concrete structure has always been a critical concern in port and marine engineering [1]. Among all the factors affecting the reinforced concrete durability, corrosion of reinforcement is often considered as the most influential factor [2–4]. For marine concrete, chloride, which is dissolved in the surrounding environment, gradually penetrates into the structure. Accordingly, steel reinforcements in the structure are subject to corrosion when the chloride content reaches a sufficiently high level [5,6].

In the case of chloride ingression, if no timely maintenance measure is carried out, the diffusion of chloride ion can cause serious consequences for the strength and esthetics of the structure, resulting in the reduction of the service life of structure [7,8]. Noticeably, chloride-induced damage may trigger critical failure of the structure within a relatively short amount of time. Therefore, the study on chloride ion ingression in concrete is of practical need for better ensuring the durability of reinforced concrete structure.

Notably, the ability of ensuring the durability and service life of reinforced concrete structure in marine environment depends upon the accuracy in predicting chloride diffusion in concrete [9]. This prediction of chloride diffusion can help to formulate predictive concrete deterioration model. Based on that, cost-effective strategy can be made regarding the appropriate time of repairing or replacing the degraded structural elements [8].

Conventional prediction approach based on Fick's second law of diffusion is commonly used to estimate the chloride diffusion process [10]. Nevertheless, this traditional formula-based method suffers from severe drawbacks such as the difficulty of parameter estimation [11] and unsatisfactory prediction accuracy [12,13]. The reason is that the dependence between chloride diffusion in concrete and its conditioning factors is inherently complex and time-dependent [11,14]. These facts demand more advanced tools for modeling the phenomenon of chloride ion diffusion in concrete.

Hodhod and Ahmed [15] attempted to model the chloride diffusivity process in high performance concrete with the application of an Artificial Neural Network (ANN). Eskandari, Nik and Eidi [16] constructed an ANN-based inference model for estimating compressive strength of mortar in marine environment. Intelligent models for predicting chloride content in concrete based on an ANN and regression tree have been examined by Asghshahr, Rahai and Ashrafi [17]. Liao, Chen, Wu, Chen and Yeh [12] studied the chloride diffusion in cement mortar by means of the Least Squares Support Vector Regression. Recent applications indicate that advanced machine learning methods

* Corresponding author.
E-mail addresses: hoangnhatduc@dtu.edu.vn (N.-D. Hoang), chuntaoc@mail.ntust.edu.tw (C.-T. Chen), kliao@ntu.edu.tw, liaokuowei@gmail.com (K.-W. Liao).

http://dx.doi.org/10.1016/j.measurement.2017.08.031
Received 25 March 2017; Received in revised form 17 August 2017; Accepted 21 August 2017
Available online 24 August 2017
0263-2241/

provide much better tools for characterizing the chloride diffusion process [18].

Nevertheless, since the chloride ion diffusion in cement mortar is indeed a complex phenomenon, other advanced machine learning approaches should be investigated for tackling with the problem of interest. Moreover, most studies employed learning algorithms that cannot yield explicit model structures. The current study attempts to fill this gap in the literature by examining the possibility of constructing chloride diffusion modeling equations from experimental data. These modeling equations can provide convenient tools for researchers and engineers to express the relationship between the chloride ion diffusion in concrete and its conditioning variables.

The Multi-Gene Genetic Programming (MGGP) and the Multivariate Adaptive Regression Splines (MARS) are selected in this research due to their successful applications in other fields of study [19,20]. Furthermore, to train and validate the prediction models, a data set including 132 cement mortar specimens in simulated marine environment has been collected. The rest of the paper is organized as follows: the second section describes the research method; the experimental setting and results are reported in the next section; the final part provides some conclusions on this study.

2. Research method

2.1. The experimental data set

To establish a data set for constructing and verifying the machine learning solutions, a total number of 132 mortar specimens with different features has been prepared. It is noted that the fresh mixing water is in compliance with the specifications of ASTM C494 – Standard specification for chemical admixtures for concrete [21]. The mortar specimens are made with Portland type I cement, which is complied with the specifications of ASTM C150 – Standard specification for Portland cement [22]. The fine aggregate for the specimens is machine-made sand derived from crushed river rocks. The measured density (saturated surface dry), water absorption (%), and maximum aggregate diameter of the fine aggregate are 2.56, 1, and 4.75 mm, respectively. The sieve analysis of the employed sand is provided in Table 1.

The water-to-cement ratio (w/c) of the mortar is 0.6. It is noted that to mimic marine environment, a 3.5% sodium chloride water solution was employed. This sodium chloride water solution has a sodium chloride content higher than 99.5% and potassium iodate content of 20–35 ppm. Moreover, chloride ion concentration test is carried out via a reagent-grade silver nitrate ($AgNO_3$) solution produced by Fluka.

It is worth noticing that to record the rate of chloride diffusion in the specimens, two groups of specimens were utilized during experimental process. The specimens in the first group (Group 1) were molded in cubic shape with edge = 10 cm, designed to mimic chloride ion diffusion in a beam, column, and wall. The specimens in the second group (Group 2) were cylinders, designed to simulate chloride ion diffusion in an entire concrete structure.

After being de-molded, it is noted that four sides of one-dimensional specimens are sealed by epoxy coat. In case of two-dimensional specimen, only the top and bottom faces of the specimens are sealed. The specimens of the two categories are illustrated in Fig. 1. For sampling purpose, the specimens of Group 1 and Group 2 tests are sliced at an interval of 2 cm. Meanwhile, the reinforced cylinder specimens (the second group) were sliced at an interval of 1.5 cm. It is noted that after sampling, the procedure for determining the chloride ion content in each layer was in compliance with AASHTO T260-97 [23].

Table 1
Sieve analysis result of the employed fine aggregate.

Sieve number	Remaining (%)	Accumulated remaining (%)	Passing (%)
3/8″	0	0	100
Number 4	1.5	1.5	98.5
Number 8	31.7	33.2	66.8
Number 16	25.8	59.1	40.9
Number 30	15.1	74.2	25.8
Number 50	14.3	88.4	11.6
Number 100	11.6	100	0
Sum	100	356.4	
Fineness modulus		3.6	

To estimate the chloride ion concentration (Y) at measured points, the mortar age (X_1), the depth of measured position (X_2), the diffusion dimension (X_3), and the presence of reinforcement (X_4) are employed as chloride ion concentration influencing factors. It is noted that the cases of $X_3 = 0$, $X_3 = 1$, and $X_3 = 2$ denote the status of no reinforcement, steel reinforcement, and aluminum reinforcement, respectively. Table 2 and Fig. 2 summarize the records of all experimental tests in this study and provide the statistical descriptions of the four influencing factors as well as the target output.

2.2. Multi-Gene Genetic Programming (MGGP)

The Genetic Programming (GP) is one of the most influential machine learning methods inspired from real-world biological systems [19]. This learning method automatically generates predictive equations based on the rules of natural genetic evolution. In the task of function approximation, GP is capable of generating explicit prediction equations autonomously without any assumptions about the prior form of the underlying relationship. Unlike conventional regression analysis approaches, GP evolves both the model structure and the model parameters of a prediction model to fit the data set at hand [24].

The Multi-Gene Genetic Programming (MGGP) is a powerful extension of the standard GP. In essence, MGGP is considered to be a weighted linear combination of the outputs obtained from individual GP models. Each individual model is called a gene. A typical MGGP model is illustrated in Fig. 3. With a certain number of genes (*NG*), it is noted that the linear coefficients (w_0, w_1,.., w_{NG}) is computed from the training data set via the ordinary least squares method [25].

The maximum tree depth (*MTD*) and the number of genes (*NG*) should be pre-specified before the model training process. Generally, large values of *MTD* and *NG* can lead to a better modeling performance but the constructed model may suffer from the risk of overfitting. On the other hand, appropriately small values of *MTD* and *NG* can help to establish more compact and comprehensible modeling equations. Similar to the mechanism of the standard GP, during the evolution process, genes of MGGP are created, modified, or deleted via crossover and mutation operators (see Fig. 4).

2.3. Multivariate Adaptive Regression Splines (MARS)

MARS [26] is a novel method for constructing modeling equations from data. This method divides the high-dimensional learning space into sub-ranges of prediction variables and establishes a mapping relationship between those prediction variables and the targeted output variable [27]. MARS uses piecewise linear function for fitting each local model and employs an adaptive approach to determine the final model. According to Freidman [26], MARS can be considered as a generalization of stepwise linear regression or a variant of regression tree with the aim of achieving better modeling capability compared to conventional regression approaches. Evidences of MARS as a powerful machine learning tool are observed in plentiful previous studies [20].

It is noted that a MARS-based model is expressed through a series of simple basis functions which characterizes the relationship between input and output variables. A basis function is shown as follows:

$$b_m(x) = \max(0, C-x) \text{ or } b_m(x) = \max(0, x-C) \quad (1)$$

where b_m denotes a basis function; x is an input variable; C represents a threshold parameter used to divide the original range of x into sub-

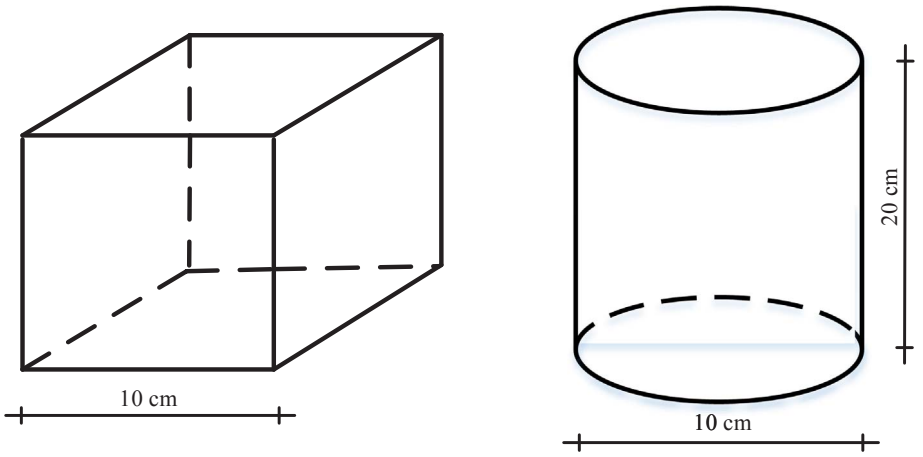


Fig. 1. Illustration of the specimens.

Table 2
Data summary.

Factors	Notation	Min	Mean	Std	Max
Mortar age (day)	X_1	7.00	78.27	51.94	168.00
Depth of measured position (cm)	X_2	0.75	2.69	1.52	5.00
Diffusion dimension	X_3	–	–	–	–
Presence of reinforcement	X_4	–	–	–	–
Chloride ion concentration (mg/g mortar)	Y	0.15	2.68	1.94	8.30

ranges.

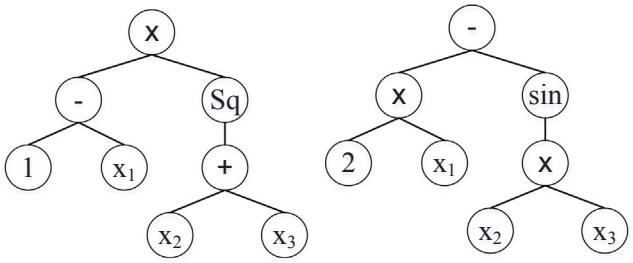
The general form of the model is expressed as follows:

$$f(x) = \alpha_0 + \sum_{m=1}^{M} \alpha_m b_m(x) \tag{2}$$

where $\alpha_0,\alpha_1,\ldots,\alpha_M$ are weighting coefficients of the MARS model; $f(x)$ represents the model output. M denotes the number of weighting coefficients.

The model establishment of MARS is divided into two steps: forward and backward steps. In the first step, basis functions are added into the model so that they can help to reduce the training error; this process terminates when the maximum number of basis function is reached. The second step aims at alleviating overfitting phenomenon by pruning redundant basis functions; each sub-model of MARS is evaluated by the generalized cross-validation (GCV) index [28,29]:

$$GCV = MSE/\left(1-\frac{k+0.5c(k-1)}{n}\right)^2 \tag{3}$$



$$f(x) = w_0 + w_1[(1-x_1)(x_2+x_3)^2] + w_2[(2x_1)-\sin(x_2x_3)]$$

Fig. 3. A typical MGGP model with two genes and three variables.

where MSE stands for mean square error of the model computed with the training data. k denotes the number of basis functions. n represents the number of observations in the training data. c is a penalty coefficient; Friedman [26] and Jekabsons [29] recommend that this parameter should be searched within the range of [2,4].

3. Experimental setting and results

3.1. Experimental setting

The purpose of this section is to construct the modeling equation from experimental data. The two machine learning approaches (MGGP and MARS) are employed to establish the mapping function that determine the relationship between input variables (the mortar age, the depth of measured point, the diffusion dimension, and the presence of reinforcement) and the targeted output (the chloride ion

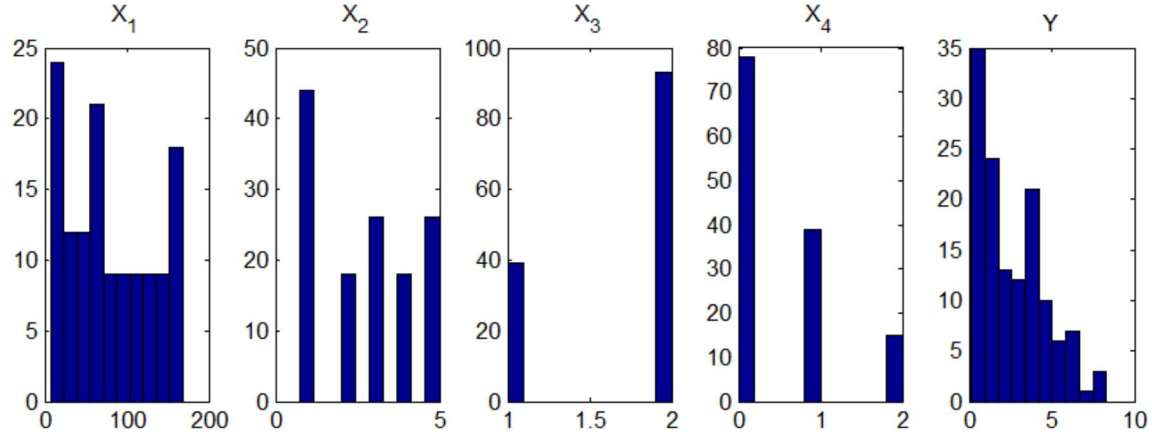


Fig. 2. Histograms of influencing factors.

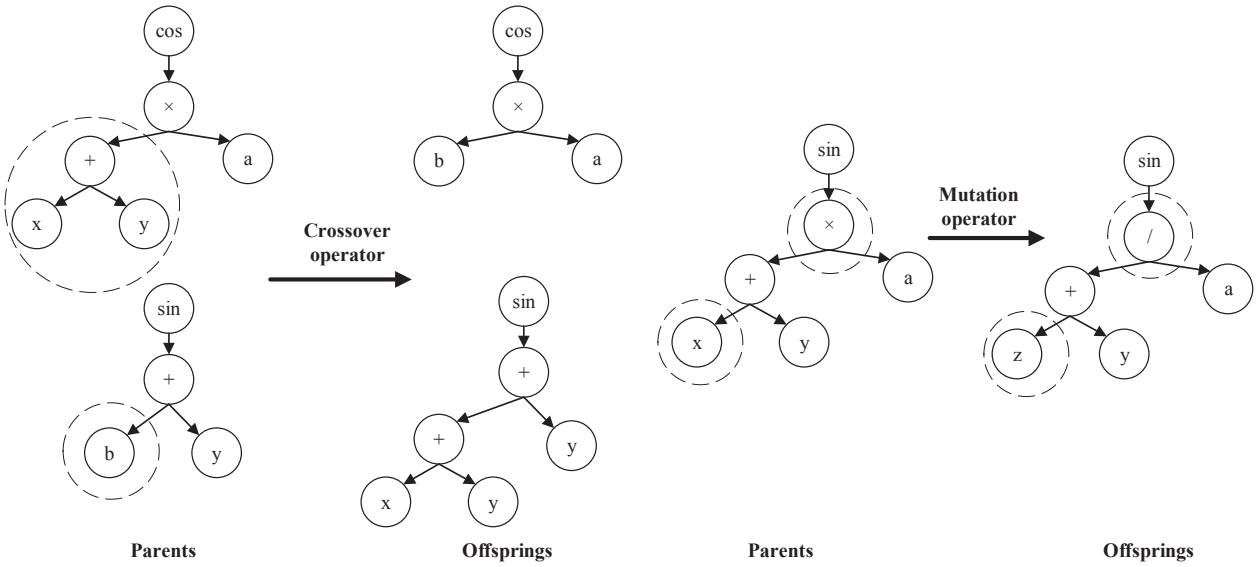


Fig. 4. Typical operations of MGGP.

concentration). It is noted that before the training and prediction phases, the variables within the data set have been normalized by means of the Z-score normalization [30].

In addition, before the data set is employed for model construction, it is beneficial to perform a preliminary investigation on the relevancy of each conditioning variable. At this step, the ReliefF [31] method is employed to carry out the analysis. Based on probability and information theories, ReliefF is capable of detecting conditional dependencies between attributes and provide a unified view on the attribute relevancy in regression problems [32]. This method assigns a weight value for each input variable that express its importance; the higher the weight, the more relevant the input variable. The analysis result is shown in Fig. 5. The ReliefF method has shown that X_1 (the age of mortar) is the most influential factor, followed by X_4 (the reinforcement presence), X_2 (the depth of measure position), and X_3 (the diffusion dimension). Since all weight values of factors are not null, all factors are relevant for characterizing the chloride ion concentration.

It is noted that the chloride diffusion is a highly time-dependent process. Therefore, the chloride diffusion cases in samples collected at earlier ages are used as training data; meanwhile, the chloride diffusion cases in samples recorded at later ages play the role as testing data. In this experiment, the training and testing sets occupy 90% (119 samples) and 10% (13 samples) of the whole data set, respectively.

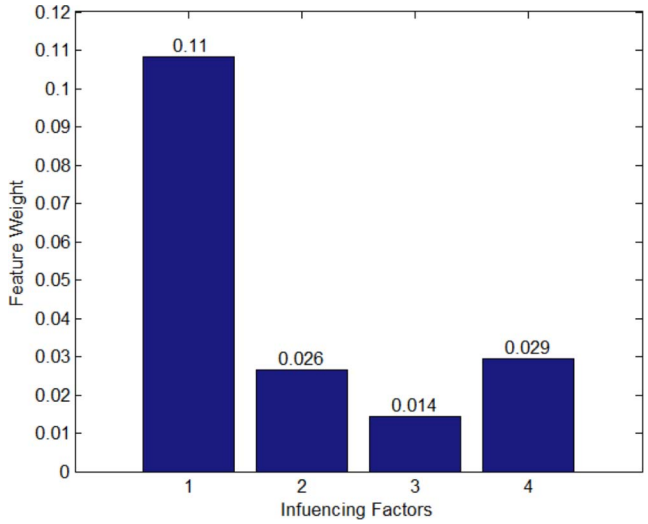


Fig. 5. Weights of factors computed by relieff.

In addition, the Root Mean Squared Error (RMSE) and the coefficient of determination (R^2) are used to quantify the prediction accuracy of each model. RMSE expresses the deviation between the output values actually observed and the output values computed from a trained model. Meanwhile, R^2 shows the proportion of the variability in the output variable explained by the model; and this index demonstrates how well a prediction model regresses the chloride ion concentration on the input variables.

It is noted that in our study, MGGP and MARS are implemented in Matlab environment via the well-developed toolboxes of Searson [25] and Jekabsons [29], respectively. Prior to the training phases of MGGP and MARS, it is required to select suitable free parameters of those models. In the case of MGGP, the most crucial parameter is the number of genes (*NG*). The other model parameters of MGGP are chosen based on the recommendation of Searson [25] and trial-and-error experiments as follows: the population size = 50; the maximum number of generations = 300; the tournament size = 5. In addition, MARS training process necessitates the specification of the maximum number of basis functions (k_{max}) and the penalty coefficient (c). In this study, a ten-fold cross validation process based on the training data set is employed to select the suitable tuning parameters of MGGP and MARS. The most desirable tuning parameters are associated with the model with the lowest average RMSE in the testing phase.

3.2. Experimental results and comparison

This section reports the prediction results of MGGP and MARS. Fig. 6 illustrates the performance of each model corresponding to different values of the tuning parameters. In Fig. 6, the horizontal axis represents the value of the investigated model parameters; meanwhile, the vertical axis denotes the cross-validation based average error (in terms of RMSE) in testing phase. In the case of MGGP (see Fig. 6a), when the number of genes increases from 1 to 5, the model error in testing phases is reduced from 1.47 to 0.86. The MGGP model with more than 5 genes does not show any improvement in prediction accuracy.

In the case of MARS (see Fig. 6b), it is noted that model performance associated with two model parameters, namely the maximum number of basis functions (k_{max}) and the penalty coefficient (c), are investigated. The maximum number of basis functions (k_{max}) starts at 5 and gradually increases with an interval of 5. At each value of k_{max}, the model performance is appraised with each value of the penalty

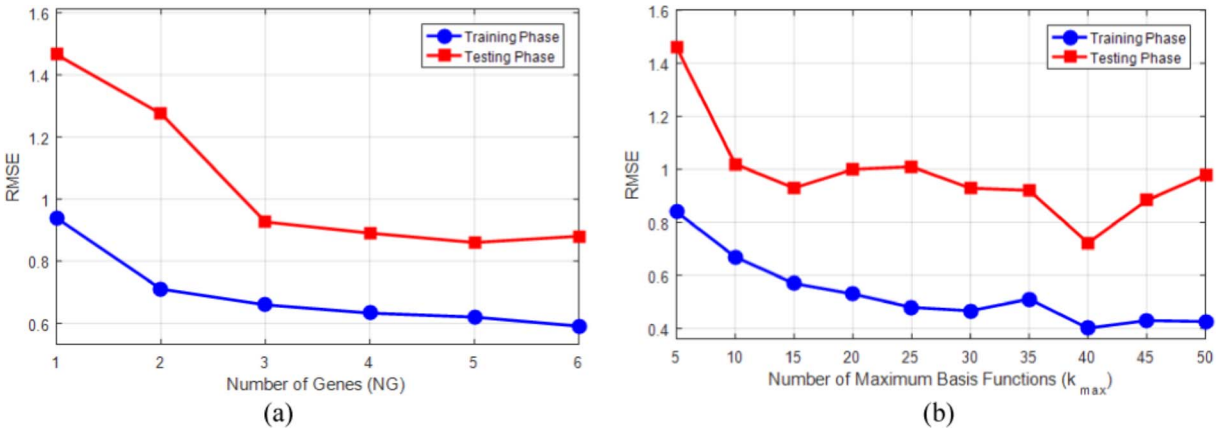


Fig. 6. Model parameter selection outcomes: (a) MGGP and (b) MARS.

Table 3
Basis functions of the MARS model.

Basis function (b)	Formula	Basis function (b)	Formula
1	$\max(0, X_2 + 1.11)$	12	$b_{10} \times \max(0, -X_4 + 0.69)$
2	$\max(0, -1.11 - X_2)$	13	$\max(0, -X_1 - 0.43) \times , X_3 + 1.54)$
3	$\max(0, X_4 - 0.69)$	14	$\max(0, X_1 + 1.23) \times \max(0, X_3 + 1.54) \times \max(0, -X_2 - 1.11)$
4	$\max(0, -X_4 + 0.0.69)$	15	$\max(0, X_1 + 1.23) \times \max(0, X_3 + 1.54) \times \max(0, -X_2 - 0.29)$
5	$\max(0, -0.43 - X_1)$	16	$\max(0, -X_1 + 0.65)$
6	$b_4 \times \max(0, X_1 - 0.65)$	17	$\max(0, X_1 - 0.65) \times \max(0, -X_2 - 0.29)$
7	$b_4 \times \max(0, -X_1 + 0.65)$	18	$\max(0, X_1 - 0.38)$
8	$b_4 \times \max(0, X_2 - 0.20)$	19	$\max(0, X_1 - 1.18)$
9	$b_4 \times \max(0, -X_2 + 0.20)$	20	$\max(0, -X_1 + 1.18)$
10	$\max(0, X_1 + 1.23)$	21	$b_{20} \times \max(0, X_4 - 0.69)$
11	$b_{10} \times \max(0, X_4 - 0.69)$	22	$b_{20} \times \max(0, -X_4 + 0.69)$

Table 4
Result comparison.

Performance		Models			
		LSSVM	MGGP	LM-ANN	MARS
Training	RMSE	0.53	0.61	0.58	0.37
	R^2	0.92	0.90	0.90	0.95
Testing	RMSE	0.89	0.87	0.83	0.70
	R^2	0.86	0.87	0.88	0.91

coefficient within the set of {0.1,0.5,1.0,1.5,⋯,4.0}. The set of parameters corresponding to the lowest average RMSE in the testing phase is selected as the most suitable one. Observed from the figure, when the number of basis functions increase from 5 to 40, the model error in both training and testing phases are reduced. The testing error (RMSE) at this point of MARS model is 0.72 with the average value of the parameter $c = 1.00$. With the numbers of basis functions > 40, we observe increases in the testing error. This fact indicates that the value of k_{max} should be limited to be 40.

When the most suitable number of genes has been identified, the MGGP model is trained with all records of the data set. The resulting prediction equation discovered by a MGGP model is reported as follows:

$$Y = 0.27X_1 - 0.28|X_1 - X_4| + 0.59\exp(-X_2) - 0.28\cos(\exp(X_1 + 2X_4)) - 0.66\cos(\exp(X_{3+2X_4})) - 0.87|X_4| + 0.72 \quad (4)$$

where the model output Y and input variables X_1, X_2, X_3, and X_4 have been defined previously in Table 2.

In addition, with the selected values of parameters ($k_{max} = 40$, $c = 1.00$), the prediction model based on MARS has been identified. It is worth noticing that the final MARS model consists of 22 basis functions; this means that 18 redundant basis functions have been cast out by this algorithm in the second learning stage. The prediction equation found by MARS is reported as follows:

$$\begin{aligned} Y = {} & -20.60 - 0.64 \times b_1 + 6.40 \times b_2 + 12.70 \times b_3 + 10.30 \times b_4 - 1.21 \times b_5 \\ & + 4.02 \times b_6 - 2.95 \times b_7 + 0.18 \times b_8 + 0.17 \times b_9 + 8.40 \times b_{10} \\ & - 6.23 \times b_{11} - 5.33 \times b_{12} - 0.21 \times b_{13} - 0.86 \times b_{14} + 0.15 \times b_{15} \\ & - 2.34 \times b_{16} - 0.73 \times b_{17} + 1.69 \times b_{18} - 7.76 \times b_{19} + 10.80 \times b_{20} \\ & - 5.30 \times b_{21} - 1.93 \times b_{22} \end{aligned} \quad (5)$$

where the basis functions (b) are defined in Table 3.

Furthermore, the prediction results of MGGP and MARS are compared to those of other widely employed machine learning approaches including the Levenberg-Marquardt Artificial Neural Network (LM-ANN) [33] and the Least Squares Support Vector Machine (LSSVM) [34]. These two approaches are powerful tools for nonlinear data modeling and their successful applications have been reported in previous studies [35–39]. In this experiment, the LM-ANN and LSSVM models are executed via the Mathwork' Neural Network Toolbox [40] and the LS-SVMlab Toolbox [41], respectively.

Moreover, to determine an appropriate configuration of LM-ANN (including the number of neurons and the learning rate), model selection processes based on a ten-fold cross validation have been used. The starting number of neurons in the hidden layer is four (which is equal to the number of input variables) and then incrementally increases to 20 neurons. Meanwhile, the learning rate parameters within the set of (0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, and 1) have been employed. The log-sigmoid function is selected as the ANN activation function. Furthermore, the number of training epochs is set to be 5000. Experimental result point outs that an ANN with the number of hidden neurons = 8 and the learning rate = 0.1 achieves the best prediction accuracy. On the other hand, to select the most desirable tuning

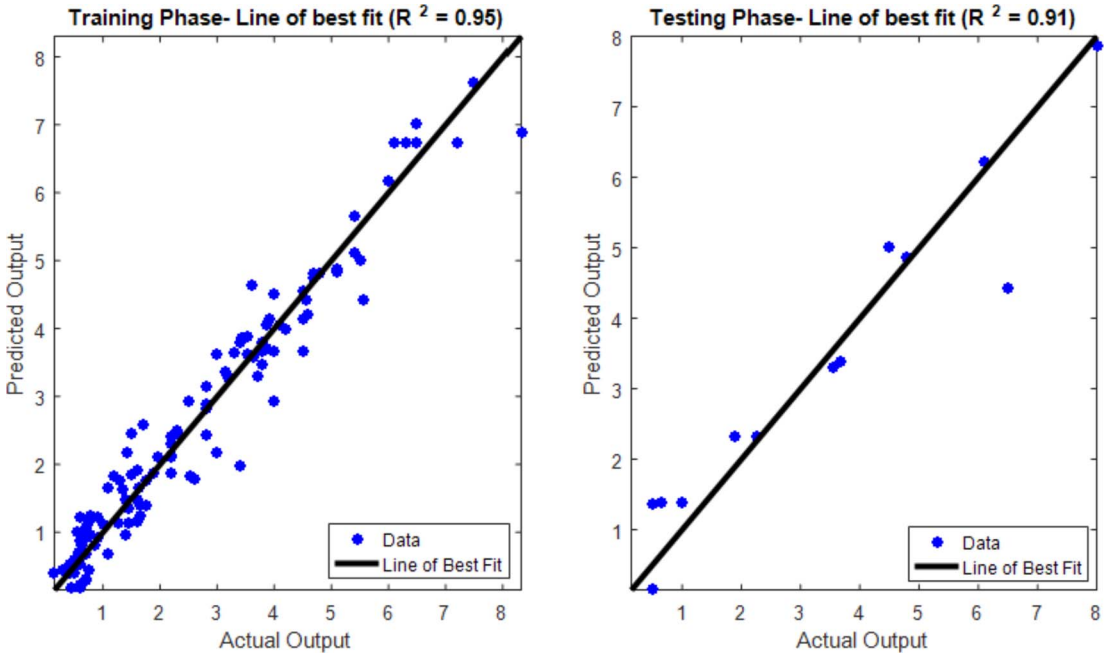


Fig. 7. Prediction results of MARS: Line of best fit.

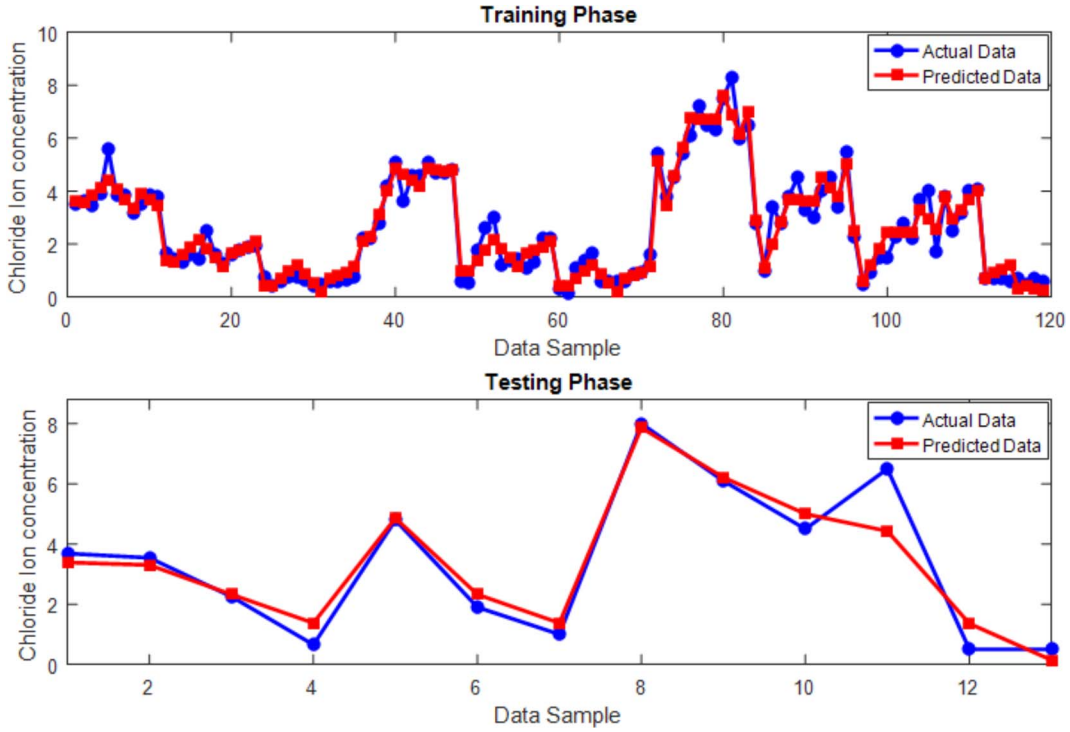


Fig. 8. Prediction results of MARS: actual and predicted observations.

parameters for LSSVM, a metaheuristic algorithm has been employed; the detail of the metaheuristic-based tuning process is provided in the previous work of Pham, Hoang and Nguyen [42].

The detail of the result comparison is summarized in Table 4, which reports the Root Mean Squared Error (RMSE) and the coefficient of determination (R^2) of each model in both training and testing phases. It is noted that the results in Table 4 are average values obtained from ten-fold cross validation processes. Observably, MARS has attained the best prediction performance in the testing phase (RMSE = 0.70 and R^2 = 0.91), followed by LM-ANN (RMSE = 0.83 and R^2 = 0.88), MGGP (RMSE = 0.87 and R^2 = 0.87), and LSSVM (RMSE = 0.89 and R^2 = 0.86). Thus, the prediction performance of MARS is found to be better than those of the two approaches of LSSVM and LM-ANN. In addition, the predictive capability of MGGP is worse than that of LM-ANN and slightly better than that of LSSVM.

These facts demonstrate that machine learning methods including MARS and MGGP are capable of establishing inference models for predicting the chloride ion diffusion in cement mortar. Notably, the formula produced by MGGP seems to be more compact than that yielded by MARS. However, it is evident that MARS has discovered a mapping function that is more accurate than the function found by MGGP. This indicates that the functional mapping between the chloride

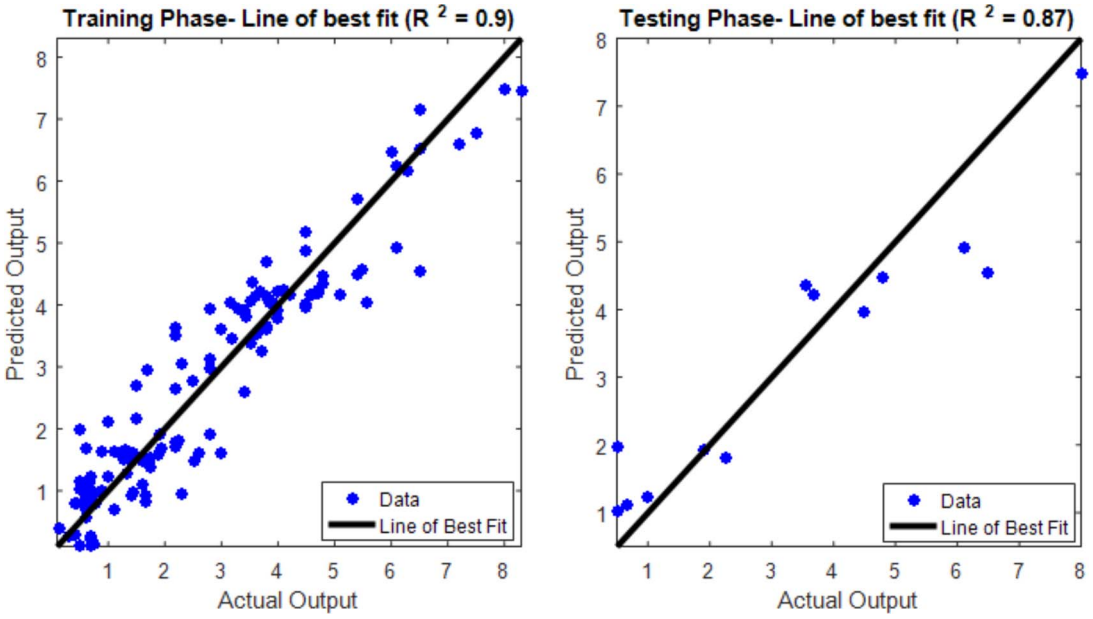


Fig. 9. Prediction results of MGGP: line of best fit.

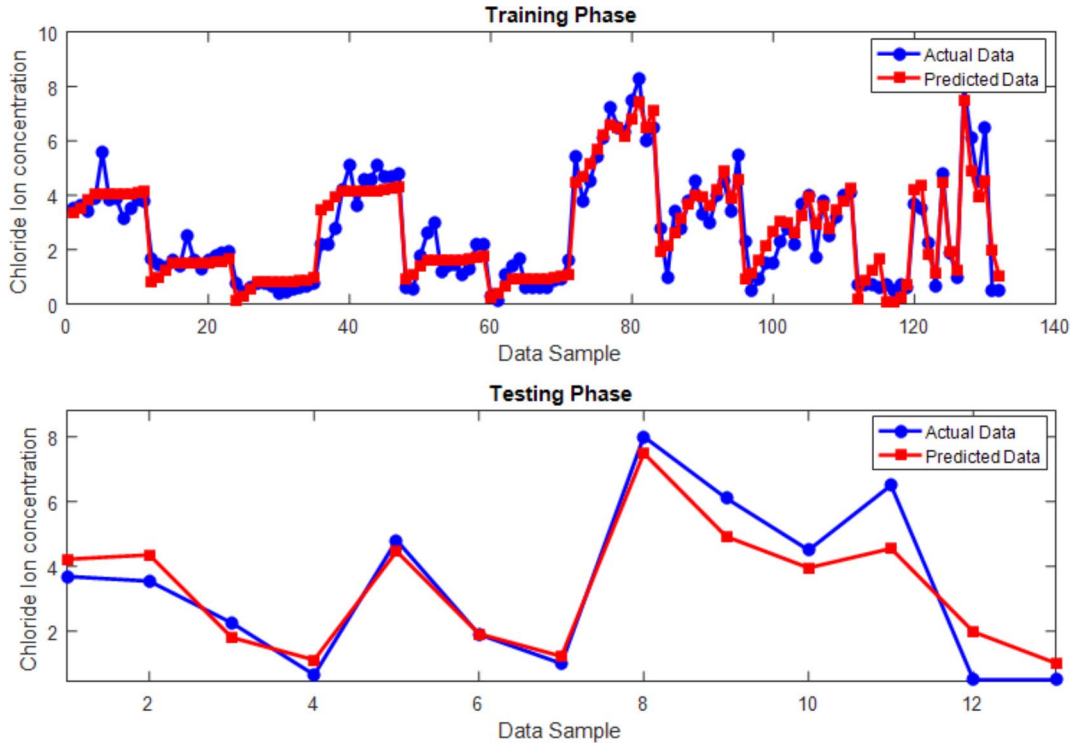


Fig. 10. Prediction results of MGGP: actual and predicted observations.

diffusion in cement mortar and its influencing factors is indeed complex. Therefore, a model with simple structure is incapable of describing such complex functional mapping in a satisfactory manner. The details of the prediction outcomes yielded by MARS and MGGP are illustrated in Figs. 7–10.

Furthermore, illustrative prediction curves of chloride ion concentration corresponding to different values of measurement depth produced by the two models are provided in Fig. 11. These two curves are constructed for the case in which the mortar age = 7, the diffusion dimension = 1, and no reinforcement is used. As can be seen from Fig. 11, there is a good agreement between the prediction curves and the actual values of chloride ion concentration. Another observation is that the model output of MGGP is clearly nonlinear; meanwhile, that of MARS is piecewise linear.

4. Conclusion

This study investigates the possibility of employing machine learning algorithms, including MGGP and MARS, for constructing prediction equations for the modeling of the diffusion of chloride ion in cement mortar. To train and verify these machine learning approaches, a data set containing 132 records of mortar specimens has been collected. The mortar age, the depth of measured point, the diffusion dimension, and the presence of reinforcement have been employed to

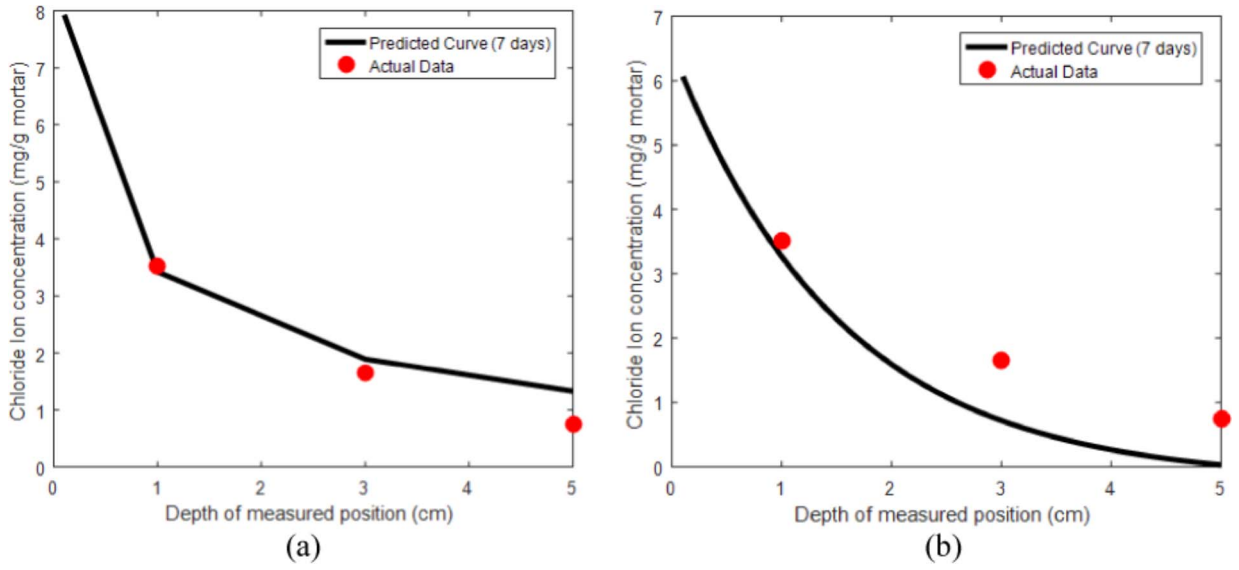


Fig. 11. Chloride ion concentration prediction curves: (a) MARS, (b) MGGP.

characterize the chloride ion concentration in each specimen.

Experimental results point out that both MGGP and MARS can help to establish modeling equation with desirable prediction accuracy; moreover, the prediction equation of MARS shows better modeling outcome than that of MGGP. Result comparison with other benchmark methods including LM-ANN and LSSVR further demonstrates the competitive performances of the two investigated approaches. Accordingly, the prediction equations produced by MARS and MGGP can be helpful to assist decision makers in the design phase of concrete structure used in marine environment. The future extension of the current study may include: (1) conducting more experimental tests with cement mortar to enhance the applicability of current prediction model; (2) investigating other advance machine learning approaches to produce more accurate modeling tools; (3) applying the MGGP as well as MARS for solving other modeling problems in civil engineering.

References

[1] Y. Wang, C.A. Lin, Y. Cui, Experiments of chloride ingression in loaded concrete members under the marine environment, J. Mater. Civ. Eng. 26 (2014).

[2] Z. Ma, T. Zhao, Y. Zhao, Effects of hydrostatic pressure on chloride ion penetration into concrete, Mag. Concrete Res. 68 (2016) 877–886.

[3] E. Ryan, E. Burdette, R. Ankabrandt, R. Nidiffer, B. Buchanan, Comparison of two methods to assess the resistance of concrete to chloride ion penetration, J. Mater. Civ. Eng. 26 (2014).

[4] S. Hong, W.-L. Lai, R. Helmerich, Experimental monitoring of chloride-induced reinforcement corrosion and chloride contamination in concrete with ground-penetrating radar, Struct. Infrastruct. Eng. 11 (2015) 15–26.

[5] S. Caré, E. Hervé, Application of a n-phase model to the diffusion coefficient of chloride in mortar, Transport Porous Med. 56 (2004) 119–135.

[6] H. Zhang, W. Zhang, X. Gu, X. Jin, N. Jin, Chloride penetration in concrete under marine atmospheric environment – analysis of the influencing factors, Struct. Infrastruct. Eng. 12 (2016) 1428–1438.

[7] M. Otieno, H. Beushausen, M. Alexander, Chloride-induced corrosion of steel in cracked concrete – part I: experimental studies under accelerated and natural marine environments, Cement Concrete Res. 79 (2016) 373–385.

[8] X. Shi, N. Xie, K. Fortune, J. Gong, Durability of steel reinforced concrete in chloride environments: an overview, Constr. Build. Mater. 30 (2012) 125–138.

[9] C. Sosdean, L. Marsavina, G. De Schutter, Experimental and numerical determination of the chloride penetration in cracked mortar specimens, Eur. J. Environ. Civ. Eng. 20 (2016) 231–249.

[10] C.L. Page, N.R. Short, A. El Tarras, Diffusion of chloride ions in hardened cement pastes, Cement Concrete Res. 11 (1981) 395–406.

[11] H.-L. Wang, J.-G. Dai, X.-Y. Sun, X.-L. Zhang, Time-dependent and stress-dependent chloride diffusivity of concrete subjected to sustained compressive loading, J. Mater. Civ. Eng. 28 (2016).

[12] K.-W. Liao, C.-T. Chen, B.-H. Wu, W.-L. Chen, C.-M. Yeh, Investigation of chloride diffusion in cement mortar via statistical learning theory, Mag. Concrete Res. 68 (2016) 237–249.

[13] J. Liu, F. Xing, B.Q. Dong, H.Y. Ma, D. Pan, New equation for description of chloride ions diffusion in concrete under shallow immersion condition, Mater. Res. Innov., 18 (2014) S2-265-S262–269.

[14] R. van Noort, M. Hunger, P. Spiesz, Long-term chloride migration coefficient in slag cement-based concrete and resistivity as an alternative test method, Constr. Build. Mater. 115 (2016) 746–759.

[15] O.A. Hodhod, H.I. Ahmed, Developing an artificial neural network model to evaluate chloride diffusivity in high performance concrete, HBRC J. 9 (2013) 15–21.

[16] H. Eskandari, M.G. Nik, M.M. Eidi, Prediction of mortar compressive strengths for different cement grades in the vicinity of sodium chloride using ANN, Procedia Eng. 150 (2016) 2185–2192.

[17] M.S. Asghshahr, A. Rahai, H. Ashrafi, Prediction of chloride content in concrete using ANN and CART, Mag. Concrete Res. 68 (2016) 1085–1098.

[18] W.Z. Taffese, E. Sistonen, Machine learning for durability and service-life assessment of reinforced concrete structures: recent advances and future directions, Automat. Constr. 77 (2017) 1–14.

[19] A.H. Gandomi, S. Sajedi, B. Kiani, Q. Huang, Genetic programming for experimental big data mining: a case study on concrete creep formulation, Automat. Constr. 70 (2016) 89–97.

[20] M.-Y. Cheng, M.-T. Cao, Estimating strength of rubberized concrete using evolutionary multivariate adaptive regression splines, J. Civ. Eng. Manage. 22 (2016) 711–720.

[21] ASTM, C494: Standard Specification for Chemical Admixtures for Concrete, ASTM International, West Conshohocken, PA, USA, 2011.

[22] ASTM, C150 – Standard Specification for Portland Cement, ASTM International, West Conshohocken, PA, USA, 2015.

[23] AASHTO, T260-97: Standard Method of Test for Sampling and Testing for Chloride Ion in Concrete and Concrete Raw Materials, AASHTO, Washington, DC, USA, 2011.

[24] D. Searson, GPTIPS-Genetic Programming & Symbolic Regression for MATLAB, Technical Report, Northern Institute for Cancer Research, Newcastle University, 2009.

[25] D.P. Searson, GPTIPS 2: An Open-Source Software Platform for Symbolic Data Mining, Springer International Publishing, Switzerland, Handbook of Genetic Programming Applications, 2015, pp. 551–573.

[26] J.H. Friedman, Multivariate adaptive regression splines, Ann. Stat. 19 (1991) 1–67.

[27] A. Parsaie, A.H. Haghiabi, M. Saneie, H. Torabi, Prediction of energy dissipation on the stepped spillway using the multivariate adaptive regression splines, ISH J. Hydraul. Eng. 22 (2016) 281–292.

[28] S. Suman, S.K. Das, R. Mohanty, Prediction of friction capacity of driven piles in clay using artificial intelligence techniques, Int. J. Geotech. Eng. 10 (2016) 469–475.

[29] G. Jekabsons, ARESLab: Adaptive Regression Splines toolbox for Matlab/Octave, Technical report, Riga Technical University, available at < http://www.cs.rtu.lv/jekabsons/>, 2016.

[30] E. Kreyszig, Advanced Engineering Mathematics, Fourth ed., Wiley, Hoboken, New Jersey, 1979.

[31] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, Mach. Learn. 53 (2003) 23–69.

[32] M. Robnik-Šikonja, I. Kononenko, An adaptation of Relief for attribute estimation in regression, Machine Learning in: Proceedings of the Fourteenth International Conference (ICML'97), Morgan Kaufmann, 1997, pp. 296–304.

[33] M.T. Hagan, M.B. Menhaj, Training feedforward networks with the Marquardt algorithm, IEEE Trans. Neural Networks 5 (1994) 989–993.

[34] J. Suykens, J.V. Gestel, J.D. Brabanter, B.D. Moor, J. Vandewalle, Least Square Support Vector Machines, World Scientific Publishing Co, Pte. Ltd., Singapore, 2002.

[35] M. Mia, N.R. Dhar, Prediction of surface roughness in hard turning under high

pressure coolant using Artificial Neural Network, Measurement 92 (2016) 464–474.
[36] B.B. Ekici, A least squares support vector machine model for prediction of the next day solar insolation for effective use of PV systems, Measurement 50 (2014) 255–262.
[37] M.-Y. Cheng, N.-D. Hoang, Estimating construction duration of diaphragm wall using firefly-tuned least squares support vector machine, Neural Comput. Appl, Online First, 2017.
[38] S. Chithra, S.R.R.S. Kumar, K. Chinnaraju, F. Alfin Ashmita, A comparative study on the compressive strength prediction models for high performance concrete containing nano silica and copper slag using regression analysis and Artificial Neural Networks, Constr. Build. Mater. 114 (2016) 528–535.
[39] D.T. Vu, N.-D. Hoang, Punching shear capacity estimation of FRP-reinforced concrete slabs using a hybrid machine learning approach, Struct. Infrastruct. Eng. 12 (2016) 1153–1161.
[40] M.H. Beale, M.T. Hagan, H.B. Demuth, Neural Network Toolbox User's Guide, The MathWorks, Inc, 2012.
[41] K. De Brabanter, P. Karsmakers, F. Ojeda, C. Alzate, J. De Brabanter, K. Pelckmans, B. De Moor, J. Vandewalle, J.A.K. Suykens, LS-SVMlab Toolbox User's Guide version 1.8, Internal Report 10-146, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2010.
[42] A. Pham, N. Hoang, Q. Nguyen, Predicting Compressive strength of high-performance concrete using metaheuristic-optimized least squares support vector regression, J. Comput. Civ. Eng. 30 (2015) 06015002.