Journal of Engineering Science and Technology
Vol. XX, No. Y (Year) PPP - QQQ

LINEAR KERNEL SUPPORT VECTOR MACHINES FOR MODELING PORE-WATER PRESSURE RESPONSES

NURADDEEN M. BABANGIDA[1,2], KHAMARUZAMAN W. YUSUF[1], MUHAMMAD R. MUSTAFA[1,*], MOHAMED H. ISA[1]

[1]Department of Civil & Environmental Engineering, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak Darul Ridzuan, Malaysia
[2]Department of Civil Engineering, Bayero University, Kano, Nigeria
*Corresponding author: raza.mustafa@petronas.com.my

Abstract

Pore-water pressure response is vital in many aspects of slope management, design and monitoring. Its measurement however, is difficult, expensive and time consuming. Studies on its predictions are lacking. Support vector machines with linear kernel was used here to predict the responses of pore-water pressure to rainfall. Pore-water pressure response data was collected from slope instrumentation program. Support vector machine meta-parameter calibration and model development was carried out using grid search and k-fold cross validation. The mean square error for the model on scaled test data is 0.0015 and the coefficient of determination is 0.9321. Although pore-water pressure response to rainfall is a complex nonlinear process, the use of linear kernel support vector machine can be employed where high accuracy can be sacrificed for computational ease and time.

Keywords: Pore-water pressure, Support vector machines; Linear kernel; Rainfall infiltration; Malaysia

1. Introduction

Pore-water pressure (PWP), also known as suction in unsaturated soil studies (negative PWP in this sense) is the pressure exerted on soil particles by water within pores of the soil. In soil mechanics PWP is used to analyse the stress state of soils; PWP has been one of the main parameters used in slope design, management and monitoring. In slope studies, especially in the unsaturated zone, it is vital to measure PWP as a rise to extreme levels could trigger failure due to

Nomenclatures

b	Bias
C	Cost parameter
i, j	Vector Indexes
K	Kernel function
n	Number of dimensions
R	Risk Function
r	Rainfall, mm
R^2	Coefficient of determination
t	Time index
U	Pore-water pressure, kPa
$\boldsymbol{w}$	Weight vector
$\mathbf{x}$	An n dimensional real valued data input vector
y	Observed output data
Z	Higher dimension feature space

Greek Symbols

ξ, ξ^*	Slack variables, for error outside the error insensitive zone/tube
α_i, α_i^*	Lagrange multipliers
ε	Width of the insensitive error zone
ϕ	Future mapping function
$\mathbb{R}$	Real Valued Vector of a given n dimensions

Abbreviations

ANN	Artificial Neural Networks
LIBSVM	Library for Support Vector Machines
LkSVM	Linear Kernel Support Vector machines
MSE	Mean Square Error
PWP	Pore-water Pressure, kPa
SVM	Support Vector Machines

reduced cohesion between soil particles. Rise in PWP can easily cause decrease in soil shear strength and dissipation of suction. The mechanism of rainfall-infiltration induced PWP rise and subsequent failure in slopes has been extensively researched and documented in several literatures [1-3]. PWP is readily used in monitoring and establishing thresholds, beyond which a slope may be classified as unsafe [4,5]. Rainfall infiltration is labelled as the main agent of PWP rise. Its mechanism has been extensively investigated in several researches [1, 6-9]. It is thus important that PWP responses to rainfall should be monitored in order to study and manage slopes effectively, most especially unsaturated slopes.

Typically, the measurement or monitoring of PWP response requires setting up an instrumentation program. This is conventionally done using tensiometers to monitor the PWP responses, and rain gauge to record rainfall. Some instrumentation programs require more detailed and extensive monitoring, thus producing more detailed information, often in real time [10].

Collecting information on PWP responses is an expensive endeavour, both resources and time wise. An easier approach is to model or simulate the PWP responses. Most PWP Pore-water studies [1, 2, 11] use finite element approach with the aid of soft-wares such as SEEP/w [12]. However, from the year 2012 several soft computing models of PWP prediction were developed using Artificial Neural Network (ANN). These can be found in the literatures of Mustafa et al. [13,14]. In water resources, ANN has been established as an excellent technique in solving many complex nonlinear problems [15-17].

Even with this success, ANN algorithms are known for over-fitting and converging to local optima. Studies have shown that Support Vector Machines (SVM) performs as good as, or better than ANN [18,19]. SVM not only generally outperforms ANN, it is relatively less complex. It has a clear geometric interpretation of its basic working principles. Furthermore, unlike ANN, its solution is guaranteed to be global. Generally, convergence of solutions to soft-computing techniques require complex and time consuming computations. A linear kernel SVM is known for its relative simplicity and fast convergence. Because of these advantages of SVM (particularly linear kernel). This study aims to evaluate the use of linear kernel support vector machines (LkSVM) to predict the non-linear responses of PWP to rainfall.

2. Support Vector Machines Theory

Support Vector Machines, developed by Vapnik, is a soft computing technique that was initially developed to handle classification problems [20] and later extended to regression problems [21]. It has been established as one of the most robust method in artificial intelligence. SVM originates from statistical learning theory. In its simplest form it tries to find a separating hyperplane, with the maximum of margin, to use as a classifier to classify a given system. A brief explanation of SVM theory is given below. Most of what followed is based on the literature of Kecman [22].

Given a training set $D = \left|\{\mathbf{x}_i, y_i\} : \mathbf{x}_i \in \Re^n, y_i \in \Re, i = 1,2 \ldots l\right|$ typically regression problems may be formulated as the following function;

$$f(\mathbf{x}, \mathbf{w}) = \mathbf{w}\mathbf{x} + b \tag{1}$$

Instead of using traditional loss function to approximate error of the function. Vapnik introduces a loss function called, Vapnik ε-*insensitivity* loss function. This function is insensitive to errors within a certain zone, a ε tube where all predicted errors within the tube are tolerable and are considered as 0. This is defined by;

$$\left|y - f(\mathbf{x}, \mathbf{w})\right|_\varepsilon = \begin{cases} 0, & \left|y - f(\mathbf{x}, \mathbf{w})\right| \le \varepsilon \\ \left|y - f(\mathbf{x}, \mathbf{w})\right| - \varepsilon, & \textit{otherwise} \end{cases} \tag{2}$$

Therefore, the SVM classifier is now formulated as that which minimizes the empirical risk and maximizes the width of the separating hyper-plane. Implementing the use of the ε-insensitive tube raises the question of finding a function that can approximate a given system with ε precision. In reality this is very difficult, and often not possible. Therefore, to tackle this problem, tolerable errors in the form of slack variables ξ were allowed outside the insensitive tube,

and were incorporated into the new constraints as formulated in Eq. (4a) & (b). One then solves the SVM classifier, by minimizing the risk *R* in Eq. (3) using the dual Lagrangian method, subject to the constraints in Eq. (4a) and (b), which yields Eq. (5).

$$R = \frac{1}{2}\|\mathbf{w}\|^2 + C\left(\sum_{i=1}^{m}\xi + \sum_{i=1}^{m}\xi^*\right) \tag{3}$$

$$\left|y - f(\mathbf{x},\mathbf{w})\right|_{\varepsilon} - \varepsilon = \xi \quad \forall \mathbf{x}_{+} \tag{4a}$$

$$\left|y - f(\mathbf{x},\mathbf{w})\right|_{\varepsilon} - \varepsilon = \xi^* \quad \forall \mathbf{x}_{-} \tag{4b}$$

$$L_D = -\varepsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) + \sum_{i=1}^{l}(\alpha_i^* + \alpha_i)y_i - \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^* + \alpha_i)(\alpha_j^* + \alpha_j)x_i \cdot x_j \tag{5}$$

The dual problem in Eq. (5), now is maximized subject to the constraints in Eq. (6a), (b) and (c). The Lagrangian coefficients are then used to determine the optimum values of the weights **w** and the bias *b*.

$$\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) = 0 \tag{6a}$$

$$0 \le \alpha_i \le C \quad \forall i \tag{6b}$$

$$0 \le \alpha_i^* \le C \quad \forall i \tag{6c}$$

2.1. Kernel trick

Often, problems are not linearly separable. For nonlinear cases the use of kernel is employed to map the input space into a higher dimensional feature space, where the data becomes linearly separable, as in Eq. (7). In the dual solution in Eq. (5), the input features are only used as inner products. If one can find a mapping function ϕ, such that $\phi = \Re^n \to Z$ and thus $\mathbf{x} \to \phi(\mathbf{x})$ then one can solve the problem. However, this generally turns out to be an impossible task for high number of features, as such kernels are used to map the input space into higher dimension feature space, without explicitly performing the mapping. A number of kernels exist for use and the basic ones are linear, polynomial, sigmoidal and radial basis function kernel.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi_i(\mathbf{x}) \cdot \phi_j(\mathbf{x}) \tag{7}$$

2.2. Linear kernel

Linear kernel is the simplest of kernels, it is in fact equivalent to no kernel at all, and it can be implemented as in Eq. (8).

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j \tag{8}$$

Linear kernel generally does not perform as good as other kernels when it comes to modelling nonlinearly separable data, but it does have its own merits. For a very large number of features (much more than the number of data set), mapping to higher dimension does not guarantee significantly better results. Thus, the use of linear kernel becomes the most attractive choice. Linear kernels are relatively easier to implement and faster to converge. Training time can increase multiple folds between SVM trained with linear kernel and that trained with other kernels. In systems where traditional SVM training methods failed, decomposition methods, e.g., [23] are often used. This entails using only a small portion of the training data for the model training. With very large number of training sets, even the decomposition method is prone to slow convergence. Again the attractive alternative falls to the use of linear kernels. Linear kernels thus provide convenience for testing and implementing new training algorithms [24]. More so due to its fast convergence, one loses little in first trying linear kernel to ascertain if the accuracy is acceptable.

3. Study Data

Instrumentation program was set up at Universiti Teknologi PETRONAS (UTP), Malaysia. The instrumented slope is approximately 11 m high, 20 m long and about 33^o to the horizontal. Soil test shows the soil at 0.5 depth to be sandy clay (United States department of agriculture- USDA classification), with permeability of 2.78×10^{-5} cm/sec. The slope was instrumented with jet-fill tensiometers with PWP measuring capacity of up to -100 kPa. The tensiometers are fitted with transducers and data logger for automated collection of PWP responses at 30 minutes interval. A raingauge for collecting and measuring rainfall was also installed at the slope toe. The data used herein was collected from a tensiometer installed at the slope crest, at a depth of 0.6 m. Data from such depth is ideal for this study, because so much fluctuations of PWP from external factors is witnessed at such shallow depth. Figure 1 shows the slope geometry and Table 1 shows the statistics of the data used herein.

4. Model Implementation

It is very vital for the analysts to have an in-depth knowledge of the process or system being modelled. Thus, one can make a judicious selection of model parameters whether analytically or otherwise. The ANN models of Mustafa et al. [14] used Eq. (9) as model input. It was obtained after a detailed analysis of cross correlation of PWP and rainfall, and auto correlation of PWP.

$$U_t = f\{U_{(t-1..t-5)}, r_{(t,t-1,t-2)}\} \tag{9}$$

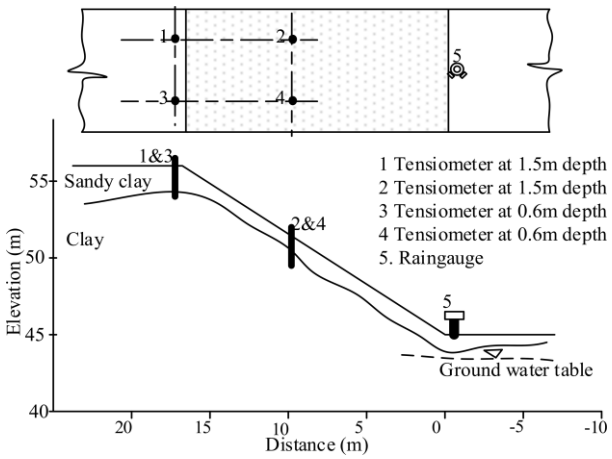


Fig. 1. Schematics of the instrumented slope.

This same input features was used herein. The data was scaled between 0 and 1. Implementation of SVM was conducted by finding the optimum meta-parameters, i.e. cost parameter *C*, the insensitive parameter ε and kernel parameters. In LkSVM, there is no kernel parameter. Grid search and 5-fold cross validation was used in the parameter calibration. The performance was evaluated using the coefficient of determination (R^2) and mean square error (MSE). The implementation of the LkSVM was carried out using LIBSVM [25].

5. Results and Discussion

The LkSVM is applied to the PWP and rainfall data from Table 1. The parameter calibration conducted over five steps of grid search and the subsequent test result is shown in Table 2. The number of support vectors during training was 849 which represent 57 percent of the training data. This is consistent with the ideal range of number of support vectors that ensures no over-fitting [26].

Table 1. Basic data statistics.

Data	Training data 1st-31st Dec '14		Testing data 1st-22nd Nov '14	
Statistics	**PWP**	**Rainfall**	**PWP**	**Rainfall**
N	1488	1488	1056	1056
Max	-5.80	32.50	-5.80	16.00
Min	-12.80	0.00	-12.40	0.00
Mean	-9.72	0.23	-10.47	0.13
Sdev	1.20	1.56	1.03	1.04
Var	1.44	2.44	1.06	1.09
Skew	0.44	12.83	0.94	13.03

N=number of data sets; Min=minimum; Max=maximum; SM=sample mean; SD=sample standard deviation; SK=skewness; VAR=variance.

Table 2. Parameter search and model test results with C=10 and ε=0.01.

	MSE	R^2
Calibration	0.0025	0.9147
Test	0.0015	0.9321

There is a good agreement between the observed and model predicted PWP as evident in Fig. 2. The MSE (scaled data) of the model is only 0.0015 for the test set. The event-based comparison of the observed and predicted PWP response is shown in the scatter plot of Fig. 3 and it has an R^2 of 0.9321. The test results tend to be slightly better than the calibration results, as shown in Table 2. The difference in the two results is due to higher variability of the calibration dataset as reported in Table 1. Some points could not be predicted very well, as they have deviated relatively far away from the perfect fit line in Fig. 3. These points are seen at some peaks in Fig. 2, erring away from the observed record.

Notably the response on the 5th and 13th November have the highest prediction gap of more than 2 kPa. The error in the event of the 13th could be due to an outlier (as no analysis of such was carried out here). Both events could have been influenced by the absence of other PWP influencing factors in the model such as temperature. The event of 5th is preceded by a 30 minutes' rainfall of 15.5 mm (as shown in Fig. 2).

Therefore, in addition to the above explanation, the error could be due to insufficient training data that describes such trend. In the PWP-rainfall time series used as the training data sets, only four rainfall events exceeded 15.5 mm. In all these four cases, different PWP response from the system were recorded. Essentially there are too few high volume rainfall data points in the training data. Thus modelling of points following high rainfall, without enough training data points could result in slight under-prediction. And such is the case with the PWP of the 5th.

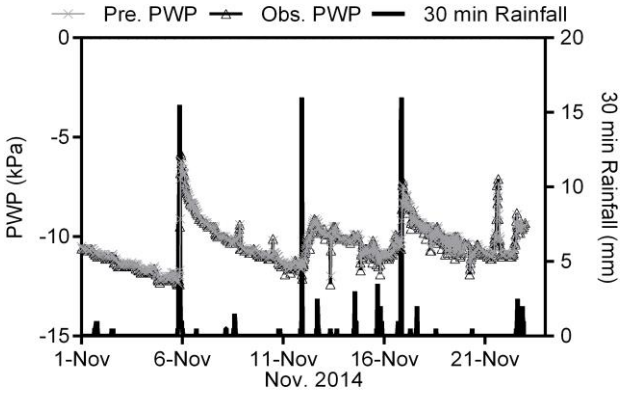


Fig. 2. Comparison of observed (Obs) PWP and Model predicted (Pre) PWP, with 30 minutes rainfall.

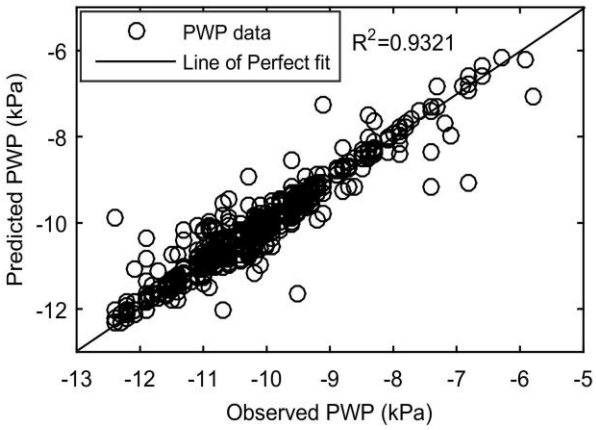
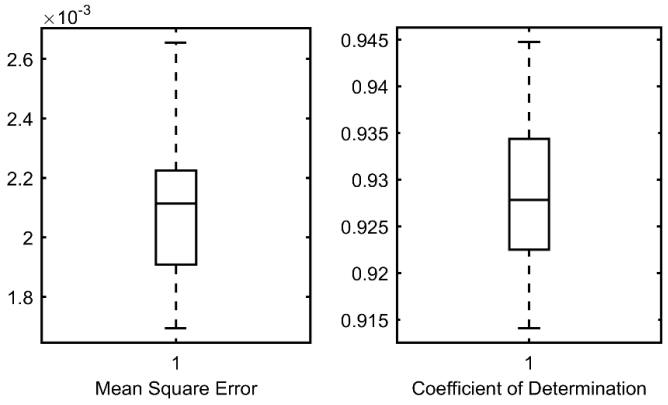


Fig. 3. Scatter plot of observed PWP and model predicted PWP.

Researchers have attempted to model the relationship between PWP and rainfall infiltration using linear predictors, and in many cases the results are not promising [27]. The relationship between PWP and rainfall infiltration has been explained as highly nonlinear and complex in nature [28]. Thus, the few models of PWP simulated it as a nonlinear process. However, the result in this study using a linear kernel SVM, yielded very good results.

In order to evaluate the effect of data skewness and fully assess the model performance. Training and testing samples were selected randomly for 100 runs, thereby creating multiple scenarios of inputs and test data (100 sets). The performance was evaluated based on the average MSE and R^2. The range of the results is shown in Figs. 4(a) and (b). The small range of the MSE and R^2 shows good Model stability. The Model showed consistency in predicting good result, indicating low level of uncertainty. Model R^2 values up to about 0.945 were obtained.



(a) **(b)**

Fig 4. Boxplot indicating range of (a) MSE and (b) R^2 obtained over 100 different simulations.

ANN models in Mustafa et al. [14] were trained with datasets, over three folds, the number of dataset used here. They yielded an R^2 value ranging from 0.97 to 0.99. Although the same accuracy may not have been achieved here; however, an R^2 of 0.9321, with much fewer data sets and ease of implementation is definitely a good proposition. LkSVM has proved to be effective in mapping the complex non-linear nature of PWP response to rainfall. Model miss-predictions are likely to occur even in a model trained with different higher order kernel function.

6. Conclusions

Using present rainfall, few antecedent records of rainfall and few antecedent PWP records, the SVM with linear kernel has successfully mimicked the complex non-linear behaviour of PWP response to rainfall events. Overall there is a good fit between the model predicted PWP and observed PWP. This further demonstrates SVM's robustness, as many linear models have failed in this context. Due to its ease of implementation, the LkSVM model can serve as a rapid approach tool to obtaining one lead time PWP records.

However, the model may sometimes, slightly under-predict peak records of which are preceded by high volume of rainfall. Most especially if there is little data describing such points. This ultimately shows that the model may be susceptible to slight miss-predictions if subjected to interpolations. The model limitations however, could improve by using additional training datasets and also the addition of other PWP influencing factors (e.g. evaporation) in the model input.

Using an SVM with higher order kernel function could generated equal or better result, but when high accuracy can be sacrificed for computational ease and time, then LkSVM provides a very attractive alternative to modelling PWP response to rainfall.

Acknowledgement

This study is financed by Ministry of Education Malaysia under fundamental research grant scheme (FRGS) of cost centre no 0153AB-I61. The first author is thankful to UTP for study sponsorship under the graduate assistantship scheme.

References

1. Gasmo, J.M.; Rahardjo, H.; and Leong, E.C. (2000). Infiltration effects on stability of a residual soil slope. *Computers and Geotechnics*, 26, 145–165.
2. Collins, B.D.; and Znidarcic, D. (2004). Stability analyses of rainfall induced landslides. *Journal of Geotechnical and Geoenvironmental Engineering*, 130(4), 362–372.
3. Yeh, H.F.; and Lee, C.H. (2013). Soil water balance model for precipitation-induced shallow landslides. *Environmental Earth Sciences*, 70(6), 2691–2701.

4. Huang, A.B.; Lee, J.T.; Ho, Y.T.; Chiu, Y.F.; and Cheng, S.Y. (2012). Stability monitoring of rainfall-induced deep landslides through pore pressure profile measurements. *Soils and Foundations*, 52(4), 737–747.
5. Eichenberger, J.; Ferrari, A.; and Laloui, L. (2013). Early warning thresholds for partially saturated slopes in volcanic ashes. *Computers and Geotechnics*, 49, 79–89.
6. Gavin, K.; and Xue, J. (2008). A simple method to analyze infiltration into unsaturated soil slopes. *Computers and Geotechnics*, 35(2) 223–230.
7. Rahardjo, H.; Ong, T.H.; Rezaur, R.B.; Leong, E.C.; and Fredlund, D.G. (2010). Response parameters for characterization of infiltration. *Environmental Earth Sciences*, 60(7), 1369–1380.
8. Tu, X.B.; Kwong, K.L.; Dai, F.C.; Tham, L.G.; and Min, H. (2009). Field monitoring of rainfall infiltration in a loess slope and analysis of failure mechanism of rainfall-induced landslides. *Engineering Geology*, 105(1), 134–150.
9. Rahardjo, H.; Satyanaga, A.; and Leong, E.C. (2013). Effects of flux boundary conditions on pore-water pressure distribution in slope. *Engineering Geology*, 165, 133–142.
10. Rahardjo, H.; Satyanaga, A.; Harnas, F.R.; and Leong, E.C. (2014). Comprehensive instrumentation for real time monitoring of flux boundary conditions in slope. *Procedia Earth Planet Sciences*, 9, 23–43.
11. Kassim, A.; Gofar, N.; Lee, L.M.; and Rahardjo, H. (2012). Pore-water of suction distributions in an unsaturated heterogeneous residual soil slope. *Engineering Geology*, 131, 70–82.
12. I. GEO-SLOPE, (2002). *SEEP/W for Finite Element Seepage Analysis*. GEO-SLOPE International, Ltd., Calgary, Alberta, Canada.
13. Mustafa, M.R.; Rezaur, R.B.; Saiedi, S.; Rahardjo, H.; and Isa, M.H. (2013). Evaluation of MLP-ANN training algorithms for soil pore-water pressure responses to rainfall. *Journal of Hydrologic Engineering*, 18, 50–57.
14. Mustafa, M.R.; Rezaur, R.B.; Rahardjo, H.; and Isa, M.H. (2012). Prediction of pore-water pressure using radial basis function neural network. *Engineering Geology*, 135, 40–47.
15. Valipour, M. (2016). Optimization of neural networks for precipitation analysis in a humid region to detect drought and wet year alarms. *Meteorological Applications*, 100, 91–100.
16. Mustafa, M.R.; Rezaur, R.B.; Saiedi, S.; and Isa, M.H. (2012). River suspended sediment prediction using various multilayer perceptron neural network training algorithms - A case study in Malaysia. *Water Resources Management*, 26, 1879–1897.
17. Valipour, M.; Banihabib, M.E.; and Behbahani, S.M.R. (2012). Monthly inflow forecasting using autoregressive artificial neural network. *Journal of Applied Sciences*, 12(20), 2139.
18. Zakaria, Z.A.; and Shabri, A. (2012). Streamflow forecasting at ungaged sites using support vector machines. *Applied Mathematical Sciences*, 6(60), 3003–3014.

19. Noori, R.; Deng, Z.; Kiaghadi, A.; and Kachoosangi, F.T. (2015). How reliable are ANN, ANFIS, and SVM techniques for predicting longitudinal dispersion coefficient in natural rivers? *Journal of Hydraulic Engineering*, 142(1), 04015039.
20. Vapnik, V.N. (1995). *The nature of statistical learning theory*. New York, New York, USA: Springer.
21. Vapnik, V.N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 10(5), 988–99.
22. Kecman, V. 2001. *Learning and soft computing*. London, England: MIT press.
23. Platt, J.C. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines. *Advances in Kernel Methods: Support Vector Learning*, 208, 1–21.
24. Lin, K.M.; and Lin, C.J. (2003). A study of reduced support vector machines. *IEEE Transactions on Neural Networks*, 14(6), 1449–1459.
25. Chang, C.C.; and Lin, C.J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (ACM TIST)*, 2(3), 1–27.
26. Mattera, D.; and Haykin, S. (1999). Support vector machines for dynamic reconstruction of a chaotic system. In B. Schölkopf, C.J.C. Burges, A.J. Smola (eds). *Advances in Kernel Methods: Support Vector Learning*, MIT Press, 209–241.
27. Rezaur, R.B.; Rahardjo, H.; Leong, E.C.; and Lee, T.T. (2003). Hydrologic behavior of residual soil slopes in singapore. *Journal of Hydrologic Engineering*, 8, 133–144.
28. Karthikeyan, M.; Toll, D.G.; and Phoon, K.K. (2008). Prediction of changes in pore-water pressure response due to rainfall events. in *Unsaturated Soils. Advances in Geo-Engineering*, D. G. Toll, C. E. Augarde, D. Gallipoli, and S. J. Wheeler, Eds. London: Taylor & Francis, 829–834.