

## Evaluation of low degree polynomial kernel support vector machines for modelling Pore-water pressure responses

Nuraddeen Muhammad Babangida<sup>1</sup>, Muhammad Raza UI Mustafa<sup>1</sup>, Khamaruzaman Wan Yusuf<sup>1</sup>, Mohamed Hasnain Isa<sup>1</sup> and Imran Baig<sup>2</sup>

<sup>1</sup>Department of Civil & Environmental Engineering, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak Darul Ridzuan, Malaysia.

<sup>2</sup>Department of Electrical and Computing Engineering, Dhofar University, Salalah, Oman

**Abstract.** Pore-water pressure (PWP) is influenced by climatic changes, especially rainfall. These changes may affect the stability of, particularly unsaturated slopes. Thus monitoring the changes in PWP resulting from climatic factors has become an important part of effective slope management. However, this monitoring requires field instrumentation program, which is resource and labour expensive. Recently, soft computing modelling has become an alternative. Low degree polynomial kernel support vector machine (SVM) was evaluated in modelling the PWP changes. The developed model used pore-water pressure and rainfall data collected from an instrumented slope. Wrapper technique was used to select input features and k-fold cross validation was used to calibrate the model parameters. The developed model showed great promise in modelling the pore-water pressure changes. High correlation, with coefficient of determination of 0.9694 between the predicted and observed changes was obtained. The one degree polynomial SVM model yielded competitive result, and can be used to provide lead time records of PWP which can aid in better slope management.

### 1 Introduction

Climatic factors, such as rainfall and evaporation play an important role in pore-water pressure (PWP) changes, which in turn is vital to slope studies. In tropical region, climatic conditions are typically characterized by variable rainfall, frequent temperature variations and high evapotranspiration. Changes in climatic factors are mostly accompanied by changes in PWP, and sometimes leading to excessively high levels of PWP. High levels or positive PWP levels may lead to failure in slopes [1, 2]. Because of such changes, it is important to monitor PWP changes, particularly to rainfall. Damiano and Mercogliano [3] recommended that for effective monitoring of hydrological behaviour of slope, field in-situ records of PWP amongst other factors should be monitored. Evidently the knowledge of PWP will aid in making objective judgment with regard to slope management.

Pore-water pressure monitoring typically entails field instrumentation of the slope in question [4, 5]. This field instrumentation is expensive, time consuming and requires expertise. Modelling of PWP responses can provide an alternative to the instrumentation. Although it will not fully replace the instrumentation but it can provide an enhancement in that few lead time responses can be obtained. PWP modelling has been carried out using several variations of artificial neural networks

(ANN) [6–10]. Albeit the ANN models boosted a remarkable correlation between observed and predicted records of PWP, ANN algorithm has some flaws that makes it a less preferable modelling algorithm these days. ANN is comparatively more complex, can get stuck in local optima and prone to over-fitting during training. The support vector machines (SVM), unlike ANN, is less complex as it has a simple geometric interpretation and its solution is always global [11]. SVM has been shown to outperform ANN in many engineering fields, as evident in some studies [12–14].

One of the most challenging aspects to the use of SVM is deciding which kernel to use. Different kernels suit different problems. In most cases literature do point the way as to which kernel to use, seldom, this may not be the case, as SVM usage is lacking in some studies. SVM was used by Lamorski et al. [15] in developing a point pedotransfer function, they compared a radial basis function (RBF) kernel and a linear kernel, and found the RBF more suitable to their data. Zhu et al. [16] found the sigmoid kernel most suitable for infrared data. The polynomial kernel has the advantage of behaving as two different kernels in one (polynomial and linear), since at a point during parameter optimization, it can assume the form of linear kernel, if the right parameter search space is selected. Furthermore a low degree polynomial has the advantage of fast training, while preserving the nonlinear capabilities of a nonlinear kernel. This study therefore

aims to evaluate the capability of SVM with low degree polynomial kernel in modelling the nonlinear responses of PWP to rainfall.

## 2 Support Vector Machines Theory

Support vector machines has established itself as a robust method of modelling in the field of machine learning. Unless one has a valid reason for using another method, SVM should be the first technique to be employed.

Support vector machines, developed by Vapnik [17], originates from the frame work of statistical learning theory. Originally applied to classification problems where outputs are hard limited and then later extended to regression problems, and termed as support vector regression (SVR). In SVR, one seek an estimator of some real valued targets. Kecman [18] explain that the problem may be formulated as below. Given a training data sets :  $\mathbf{X}_1 y_1, \mathbf{X}_2 y_2 \dots \mathbf{X}_m y_m : \mathbf{X} \in \mathbb{R}^n, y \in \mathbb{R}$ , where input  $\mathbf{x}$  are  $n$  dimensional vector, and  $y$  are real valued ( $\mathbb{R}$ ) responses. The SVR function that represents the optimal separating hyper-plane may be introduced via linear regression, as shown in Eq. 1.

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

With  $\mathbf{w}$ , being a weight vector and  $b$  a bias. Typically a loss function to measure the approximation error is required. Vapnik introduced (Eq. 2) a loss function of the form;

$$|y - f(\mathbf{x}, \mathbf{w})|_{\epsilon} = \begin{cases} 0, & |y - f(\mathbf{x}, \mathbf{w})| \leq \epsilon \\ |y - f(\mathbf{x}, \mathbf{w})| - \epsilon, & \text{otherwise} \end{cases} \quad (2)$$

This function is insensitive to errors within a certain  $\epsilon$  zone, a zone within which all predicted errors are tolerable and are considered as 0. The objective of the SVM algorithm is to find an estimator of the real valued targets, by minimizing the empirical risk and at the same time maximizing the width of the insensitive zone  $\|\mathbf{w}\|^2$ . This is formulated as Eq. 3, under the constraints in Eq. 4.

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

$$|y - f(\mathbf{x}, \mathbf{w})|_{\epsilon} - \epsilon \leq \xi \quad \mathbf{x} \in \mathbb{R}^n \quad (4a)$$

$$|y - f(\mathbf{x}, \mathbf{w})|_{\epsilon} - \epsilon \leq \xi^* \quad \mathbf{x} \in \mathbb{R}^n \quad (4b)$$

$$\xi, \xi^* \geq 0 \quad (4c)$$

The  $\xi^*$ ,  $\xi$  are slack variables that indicates distance of miss classified points from the insensitive zone,  $C$ , is a regularization parameter that needs to be tuned, along with  $\epsilon$ . The optimization problem in Eq. 3 is solved first by forming a primal variable Langrangian. This is then readily solved in the dual space. A standard optimization algorithm is used to solve the resulting Hessian matrix.

Ultimately, learning results in the Langrange multipliers  $\alpha$ , and the number of free  $\alpha$  is equal to number of support vectors (SV). The obtained  $\alpha$  values are used to find the optimal weights  $\mathbf{w}$  and optimal bias  $b$ , which are in turn used to determine the optimum separating hyper-plane or estimator of Eq. 1.

For nonlinear regression problems, where no linear separating plane exist. Inputs are mapped to higher dimensional space. This is done by use of kernel functions. Kernel maps input features from a given dimension to a higher dimensional feature space. With this, nonlinearly separable problems can be readily solved as linear problems in high dimensional space. Basic kernel functions include, the Linear, Polynomial, Sigmoid and Gaussian. Only basic explanation of SVM formulation is offered here. Kecman [18] provides, detailed explanation on SVM formulations.

### 2.1 Polynomial Kernel

Polynomial kernel, given in Eq. 5, is less frequently used, compared to say RBF kernel. It generally does not give better accuracy. However, low degree polynomial ( $d \leq 3$ ) trained and tested in similar fashion as RBF perform only slightly less, but their training and testing time is much faster, and they can also be trained using linear SVM methods [19]. Additionally, polynomial kernels are found to perform better in natural language predictions [20].

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i \cdot \mathbf{x}_j + r)^d \quad (5)$$

Where,  $\gamma$ ,  $r$  and  $d$ , are kernel parameters that need to be tuned.

## 3 Methodology

### 3.1. Field data

The data used was part of the data from a slope instrumentation program, setup to monitor the PWP responses in a slope. The slope is located in Universiti Teknologi PETRONAS (UTP) in Malaysia. The slope was instrumented with tensiometers (fitted with transducers and data logger) and rain gauge. Part of the data from the crest at 0.6 meter depth was used for this study. Its properties are shown in Table 1. The highest water table around the area was found to be about 1.4 meters below the ground level (slope toe), as monitored from four monitoring wells sited around the slope area. Results from undisturbed sample of the slope material, indicated that the slope soil material is sandy clay in the top one meter, to clay from about two meters.

### 3.2 Input Features Determination

The selection of model representative attributes constitutes great challenge in soft computing modelling. Selecting the optimum number of attributes or input features require a good knowledge of the system to be modelled. The knowledge of the physical system of PWP has been documented in literatures.

**Table 1.** Data Properties

	Training (3577) 30/10/14 – 31/1/15		Test set (2832) 1/2/15 – 31/3/15	
	PWP (kPa)	Rain (mm)	PWP (kPa)	Rain (mm)
<b>Mean</b>	-9.5	0.2	-14.5	0.1
<b>Max</b>	-3.5	46.0	-5.5	26.5
<b>Min</b>	-34.1	0.0	-27.7	0.0
<b>SD</b>	3.3	1.6	3.9	1.2
<b>Var</b>	10.8	2.6	14.9	1.4
<b>Skew</b>	-3.4	16.0	-0.5	15.8

Rahardjo et al. [5] showed that five antecedent records of rainfall are needed to model the PWP responses to rainfall. Mustafa et al. [10], showed that five antecedent PWP records will be needed to develop a good model of PWP. A combination of both of these input features is considered herein. Therefore five antecedent records of each of PWP and rainfall were considered for analysis and determination of the needed features. However, only a subset of these may be relevant.

The state of the art input feature selection involves the use of either, the filter method or wrapper method and to a lesser extent, the embedded method, which has comparatively limited applications [21, 22]. Filter method aims to find the relevant features, outside the framework the algorithm, by evaluating the relevance of the features using a statistical measure. The wrapper method aims to evaluate prediction power of the features within a chosen algorithm. Although filter method is computationally cheaper, it provides a broad selection of features, which are not specific to any algorithm. On the other hand, the wrapper method yields features which are algorithm and data specific. In practice this is the most ideally sought option [23].

### 3.3 Wrapper method of Input determination

The wrapper method requires the selection of a search algorithm and evaluation function. The search algorithm explores the feature search space, using operators tailored to the search engine, which significantly aid in how the search space is manipulated by the search algorithm. The evaluation function, employs the use of a learning scheme, which generally happens to be the algorithm, which the features are to be used within to develop the model in question. A learning scheme is used as the feature evaluation algorithm. By using a learning scheme, K-fold cross validation (CV) may be used to estimate the accuracy of the learning scheme for any selected set of sub features, this way the selected sub features can be evaluated. An evaluation measure such as root mean squared error (RMSE), coefficient of determination ( $R^2$ ), mean absolute error (MAE), etc. are used as accuracy measure of the CV. In essence this serves as the objective function of the search engine.

Several search algorithms, such as evolutionary algorithm (EA), greedy search etc. may be used as the search engine with its objective set as the performance of the evaluation function. The search engine is tasked with

exploring the feature search space. Detailed use of the wrapper method can be found in the literature of [23]. Chandrashekar and Sahin [21] explained some search methods as heuristic search algorithms, methods such as that of genetic algorithm (GA) used as a search engine for feature subsets selection, since these methods do not follow the usual methods of forward or backward selection. They went on to explain that the chromosomes of the GA can be used to represent if a feature is selected or not. With the objective function being set as the performance of the prediction algorithm, which uses the selected chromosomes as input features. In this study such method will be used with GA being employed as the search algorithm. The GA uses the result to determine which amongst the selected feature subsets has less relevance to the predictor algorithm. Upon that the GA creates a better subset by manipulation within it, aided by the following operators: single point crossover, bit-flip mutation, generational replacement with elitism. The wrapper attribute selection will be implemented using Waikato environment for knowledge analysis (WEKA) [24].

### 3.4 SVM model Development

Basic SVM model development may be implemented with four basic steps as follows;

- Identifying the kernel to be used in the model. Although, many studies choose to use the RBF on the basis of being the most common. It is wise to compare the performance of few kernels or consult literature on similar data.
- Identify the kernel parameters to be optimized in addition to the SVM meta-parameters ( $C, \epsilon$ ). Use a suitable optimization technique, such as the GA, Particle swarm optimisation (PSO), grid search etc. to find the optimum of the meta-parameters.
- Use the optimum parameters to train the model on part of the data (training data).
- Test the model on unseen data sets (test set), to evaluate its generalization performance. Appropriate accuracy measures are used for the evaluation. The SVM within the wrapper algorithm, is implemented using WEKA's library SVM [25] and, model development is implemented using library for support vector machines (LIBSVM) [26]

## 4 Results and Discussions

### 4.1 Data Analysis

The wrapper method was employed in this study to determine the optimal subset of input features. Evolutionary algorithm was used as the search engine in the wrapper method. The feature space with 11 sets of features (chromosomes) was exploited through manipulation within the GA to select subsets. SVM was used as the evaluation function, with the data evaluated over fivefold grid search. The performance of the CV was evaluated using root mean squared error (RMSE), and is shown in Table 2.

**Table 2.** Root mean squared error (RMSE) of cross validation results, using different chromosome combinations- as implement with the wrapper algorithm.

	Current Fitness		Global Fitness	
	Best	Least	Best	Least
RMSE	0.0267	0.0272	0.0267	0.1085
CC*	2, 4,7,9	2,3,4,6,7,9	2,4,7,9	11

\* $U_{t-1}=2, U_{t-2}=3, \dots, U_{t-5}=6, r_t=7, r_{t-1}=8 \dots r_{t-5}=12$

Where  $U$  represents PWP, and  $r$  is rainfall. The subscript indexes, are lag positions of the variables. E.g.  $U_{t-1}$  represents one antecedent (30 minutes lag) PWP.

The RMSE of best feature combinations (Chromosome combination—CC), that appears in the fivefold CV showed lowest values for combination of two, four, seven and nine. Thus the result of the wrapper input determination, showed Eq. 6 as an excellent set of parameters for modelling PWP.

$$U_t = f(U_{t-1}, r_{t,t-2}) \quad (6)$$

Although other feature combination yielded similar result to this 3 features, however, they have more number of features and that makes them a less preferable choice for the model build. Fewer features with the same fitness level will ensure a simpler, less complex model.

#### 4.2 SVM Calibration and Testing

The training data was scaled between 0 and 1. Base on the objective of the study, the polynomial kernel was selected. The input features obtained from the wrapper input determination were used as the input for the model development. Grid search with fivefold CV was used to search the optimum model meta-parameters. The grid search was carried out in three stages. First, a coarse grid search, with the following range  $10^{-3} \leq C \leq 10^3$ ,  $10^{-4} \leq \epsilon \leq 10^1$ ,  $10^{-3} \leq \gamma \leq 10^3$ ,  $1 \leq d \leq 3$ , and  $0 \leq r \leq 5$ , was carried out, and then followed by two steps of finer grid search. Table 3 shows the obtained meta-parameters for the model. The combination of parameters from Table 3, gave the best of results during a fivefold CV.

The calibrated parameters from the grid search were used to train the SVM model, which was then tested on the test set. The result of the calibration and testing is shown in Table 3. The results presented in Table 3 are results of analysis carried out using scaled data sets. The polynomial kernel yielded very good results as can be seen from the performance measures in Table 3. A slightly better performance was obtained during the test stages, compared to results from the calibration stages. This can be attributed to data variability. The variance of the data in the test stages is higher than that in the calibration stage, making the training data much smoother, and thus better result in the calibration stage. This is consistent with effects of data variability [27].

**Table 3** Performance metrics (for calibration and testing) of the developed Polynomial kernel Model

$C=1, \epsilon=0.001, \gamma=0.03, r=0, d=1$
--

Calibration (3577 sets)			Test (2832 sets)		
RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
0.0219	0.0115	0.9590	0.0222	0.0101	0.9694

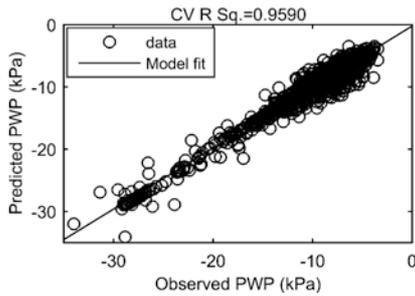
Figure 1. Shows a scatter plot from the calibration of the model. The fivefold CV calibration result was compared with the observed training result. The accuracy measure obtained was good, with R<sup>2</sup> of 0.9845, and low values of MAE and RMSE as well. The model fit line indicates the best fit attained by the model (not the perfect fit line which indicates, perfect agreement), since unlike the perfect fit line, the model fit shows the correlation extent, and what is important is to have all points on the line. The calibration results shows how good the developed Model did in data training. However, in machine learning, a model may yield good approximation result, during training but may fail to yield even close to that result on test data. What is vital is for the model to perform well on unseen data, and that is the generalization power of the model.

Thus the generalization ability of the Model was tested, and the estimation errors are shown in Table 3. The estimation errors obtained are low, and very close to the approximation errors from the fivefold CV. The generalization performance of the model was further brought to light in the individual event based Model predictions, shown in the scatter plot of Figure 2. The result obtained has an R<sup>2</sup> value of 0.9694, indicating not only a strong agreement, but near perfect correlation. This is an indication of good generalization ability of the Model.

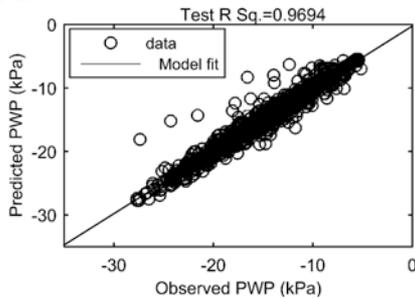
The relationship between the observed and predicted values, is again shown in a better picture, in the time series plot of the two events (shown in Figure 3). The Model mimicked the reality, the complex trend, including several points where a rainfall event has pushed the PWP to high values. Rainfall events push PWP to extreme values, and most often, these are the points that tend to be difficult predict.

Generally in time series modelling, predicted values tend to lag the observed, such lag was not observed here, indicating the robustness of the Model. Nevertheless, some points were difficult to predict, and are unfortunately mis-estimated. Some of these points have a sudden unprecedented high rise in negative PWP. These are considered as outliers, as there was no prior event that could trigger such rise. However, no analysis of outlier is offered in this study. The most pronounced of such points are clearly shown in Figure 2. They have deviated the most from the model fit line, indicating less correlation between the Model predictions and observed records.

Support vector machines with polynomial kernel, has mimicked the nonlinear responses of PWP to rainfall events quite well as indicated by some measures explained above. The results are competitive even in comparison to the neural network model trained with eight input features and nine months of data [6].



**Figure 1.** Scatter plot of the fivefold Cross validation and observed PWP, with Line of fit, indicating model fitness to observed data



**Figure 2.** Scatter plot of the predicted and observed PWP, with Line of fit, indicating model fitness to observed data

## 5 Conclusions

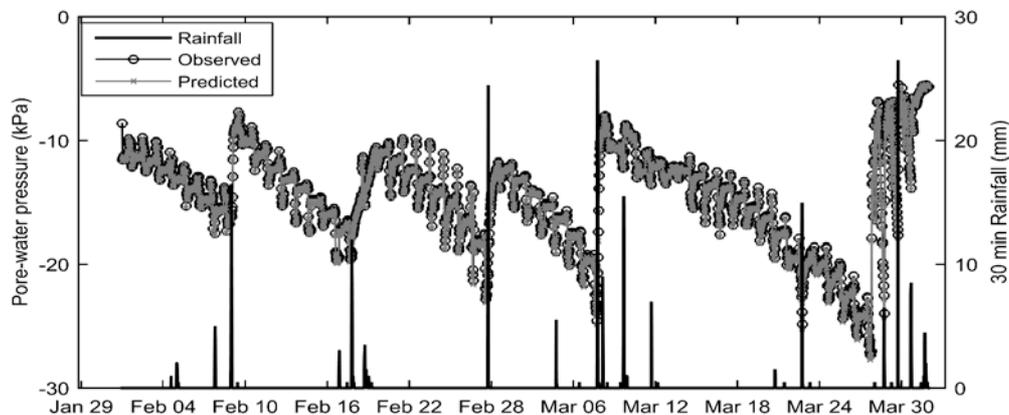
A simple one degree polynomial kernel, has performed well in predicting the complex nonlinear behaviour of PWP, with high correlation recorded between observed and predicted records. This is aided with the excellent choice of input features. As it aided in obtaining a

comparatively simple model (three inputs features), and that is what the low degree polynomial is all about simplicity. The low degree Polynomial kernel SVM model, thus can be readily employed to model PWP responses. However, this by no means constitute polynomial kernel as the best kernel to be used for PWP modelling, as this study did not compare results of the polynomial Model with other kernels. In fact a linear kernel Model, could perform nearly as good, since a linear kernel is equivalent to a polynomial with  $r=0$ ,  $d=1$ , and  $\gamma=1$ . In the calibration of this Model only  $\gamma$  was different (0.03). Therefore the developed model is quite close to a linear kernel model. Thus there is the need to check other basic kernels to ensure the best choice.

In addition to its proven advantages, such as fast implementation and training, the result of the developed Model are competitive and the low degree polynomial kernel SVM, with just one degree can serve as an excellent option in modelling PWP responses.

## 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.



**Figure 3.** 30 minutes time series plot, showing comparison between observed and Model predicted PWP. With the Model prediction closely following the observed, including records immediately after rainfall event

## References

1. E. E. Alonso, N. M. Pinyol, and a. Yerro, "Mathematical Modelling of Slopes," *Procedia Earth Planet. Sci.*, vol. **9**, pp. 64–73 (2014)
2. H. F. Yeh and C. H. Lee, "Soil water balance model for precipitation-induced shallow landslides," *Environ. Earth Sci.*, vol. **70**, no. 6, pp. 2691–2701 (2013)
3. E. Damiano and P. Mercogliano, "Potential Effects of Climate Change on Slope Stability in Unsaturated Pyroclastic Soils," in *Landslide Science and Practice*, vol. **4**, C. Margottini, P. Canuti, and K. Sassa, Eds. Berlin: Springer (2013)
4. H. Rahardjo, A. Satyanaga, and E. C. Leong, "Characteristics of Pore-Water Pressure Response in Slopes During Rainfall," in *3dr Asian Conference on Unsaturated soils-UNSAT-ASIA*, pp. 493–498. (2007)
5. H. Rahardjo, E. C. Leong, and R. B. Rezaur, "Effect of antecedent rainfall on pore-water pressure distribution characteristics in residual soil slopes under tropical rainfall," *Hydrol. Process.*, vol. **22**, no. 4, pp. 506–523, (2008)
6. M. R. Mustafa, R. B. Rezaur, H. Rahardjo, M. H. Isa, and A. Arif, "Artificial neural network modeling for

- spatial and temporal variations of pore-water pressure responses to rainfall Artificial neural network modeling for spatial and temporal variations of pore-water pressure responses to rainfall,” *Adv. Meteorol.*, vol. **2015**, pp. 1–39, (2015)
7. M. R. Mustafa, R. B. Rezaur, M. H. Isa, and H. Rahardjo, “Estimation of soil pore-water pressure variations using a thin plate spline basis function,” in *WIT Transactions of the Built Environment*, vol **137** pp. 615–624 (2014)
  8. M. R. Mustafa, R. B. Rezaur, S. Saiedi, H. Rahardjo, and M. H. Isa, “Evaluation of MLP-ANN Training Algorithms for Modeling Soil Pore-Water Pressure Responses to Rainfall,” *J. Hydrol. Eng.*, vol. **18**, no. January, pp. 50–57, (2013)
  9. M. R. Mustafa, R. B. Rezaur, M. H. Isa, S. Saiedi, and H. Rahardjo, “Effect of antecedent conditions on prediction of pore-water pressure using artificial neural networks,” *Mod. Appl. Sci.*, vol. **6**, no. 2, pp. 6–15, (2012)
  10. M. R. Mustafa, R. B. Rezaur, H. Rahardjo, and M. H. Isa, “Prediction of pore-water pressure using radial basis function neural network,” *Eng. Geol.*, vol. **135–136**, pp. 40–47, (2012)
  11. C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” *Data Min. Knowl. Discov.*, vol. **2**, no. 2, pp. 121–167, Jun. (1998)
  12. R. Noori, Z. Deng, A. Kiaghadi, and F. T. Kachoosangi, “How Reliable Are ANN , ANFIS , and SVM Techniques for Predicting Longitudinal Dispersion Coefficient in Natural Rivers?,” *J. Hydraul. Eng.*, pp. 1–8 (2015)
  13. W. Zhao, T. Tao, and E. Zio, “Parameters tuning in support vector regression for reliability forecasting,” *Chem. Eng. Trans.*, vol. **33**, pp. 523–528 (2013)
  14. Z. A. Zakaria and A. Shabri, “Streamflow Forecasting at Ungaged Sites Using Support Vector Machines,” *Appl. Math. Sci.*, vol. **6**, no. 60, pp. 3003–3014 (2012)
  15. K. Lamorski, C. Sławiński, F. Moreno, G. Barna, W. Skierucha, and J. L. Arrue, “Modelling soil water retention using support vector machines with genetic algorithm optimisation,” *Scientific World Journal*, vol. **2014** (2014)
  16. B. Zhu, Z. Cheng, and H. Wang, “A kernel function Optimisation and Selection Algorithm Based on Cost Function Maximization,” in *IEEE International Conference on Imaging Systems and Techniques (IST), 2013*, no. **2**, pp. 259–263 (2013)
  17. V. N. Vapnik, *The Nature of Statistical Learning Theory*, vol. **8**, no. 6. New York, New York, USA: Springer, (1995)
  18. V. Kecman, *Learning and Soft Computing*. London, England: MIT press, (2001)
  19. Y. Chang, C.-J. Hsieh, K.-W. Chang, M. Ringgaard, and C. Lin, “Training and Testing Low-degree Polynomial Data Mappings via Linear SVM,” *J. Mach.*, vol. **11**, pp. 1471–1490, (2010)
  20. S. Yaman and J. Pelecanos, “Machines for Speaker Veri fi cation,” *IEEE Signal Process. Lett.*, vol. **20**, no. 9, pp. 901–904, (2013)
  21. G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Comput. Electr. Eng.*, vol. **40**, no. 1, pp. 16–28, (2014)
  22. S. Maldonado and R. Weber, “A wrapper method for feature selection using Support Vector Machines,” *Inf. Sci. (Ny)*, vol. **179**, no. 13, pp. 2208–2217, (2009).
  23. R. Kohavi and R. Kohavi, “Wrappers for feature subset selection,” *Artif. Intell.*, vol. **97**, no. 1–2, pp. 273–324, (1997).
  24. H. Mark, F. Eibe, H. Geoffrey, P. Bernhard, R. Peter, and I. H. Witten “The WEKA Data Mining Software: An Update,” *SIGKDD Explorations*, vol **11**, no 1 (2009)
  25. Y. EL-Manzalawy and V. Honavar, "WLSVM : Integrating LibSVM into Weka Environment," (2005). <http://www.cs.iastate.edu/~yasser/wlsv>
  26. C. C. Chang, C. J. Lin "LIBSVM: A Library for Support Vector Machines," *ACM Trans Intell Syst Technol (ACM TIST)* vol **2**, pp. 1–27 (2011)
  27. P. A. Kagoda, J. Ndiritu, C. Ntuli, and B. Mwaka, “Application of radial basis function neural networks to short-term streamflow forecasting,” *Phys. Chem. Earth, Parts A/B/C*, vol. **35**, no. 13–14, pp. 571–581, (2010)