

# Prediction of swelling pressure of soil using artificial intelligence techniques

Sarat Kumar Das · Pijush Samui ·  
Akshaya Kumar Sabat · T. G. Sitharam

Received: 22 January 2009 / Accepted: 3 November 2009 / Published online: 20 November 2009  
© Springer-Verlag 2009

**Abstract** The swelling pressure of soil depends upon various soil parameters such as mineralogy, clay content, Atterberg's limits, dry density, moisture content, initial degree of saturation, etc. along with structural and environmental factors. It is very difficult to model and analyze swelling pressure effectively taking all the above aspects into consideration. Various statistical/empirical methods have been attempted to predict the swelling pressure based on index properties of soil. In this paper, the computational intelligence techniques artificial neural network and support vector machine have been used to develop models based on the set of available experimental results to predict swelling pressure from the inputs; natural moisture content, dry density, liquid limit, plasticity index, and clay fraction. The generalization of the model to new set of data other than the training set of data is discussed which is required for successful application of a model. A detailed study of the relative performance of the computational intelligence

techniques has been carried out based on different statistical performance criteria.

**Keywords** Expansive soil · Swelling pressure · Artificial neural network · Support vector machine

## Introduction

Expansive soil and bedrock underlie more than one-third of world's land surface. Each year, damage to buildings, roads, pipelines, and other structures by expansive soils is much higher than the damage caused by floods, hurricanes, tornadoes, and earthquakes combined (Jones and Holtz 1973). The estimated annual cost of damage due to expansive soils is \$1,000 million in the USA, £150 million in the UK, and many billions of pounds worldwide (Gourley et al. 1993). However, as the hazards due to expansive soils develop gradually and seldom present a threat to life, they have received limited attention, despite their severe effects on the economy. Much of the damage related to expansive soils is not due to a lack of appropriate engineering solutions but to the non-recognition of expansive soils and expected magnitude of expansion early in land use and project planning. The damage to foundation on expansive soil can be avoided/minimized by proper identification, classification, quantification of swell pressure, and provision of an appropriate design procedure. Swelling potential of clayey soil is a measure of the ability and degree to which such a soil might swell if its environments were changed in a definite way. Hence, the expansive soil is classified based on its potential for swelling. However, there is not a definite expression of swell potential for classification of expansive soils (Nelson and Miller 1992). Holtz (1959) referred to swell potential

---

S. K. Das (✉)  
Department of Civil Engineering, National Institute  
of Technology, Rourkela 769008, India  
e-mail: saratdas@rediffmail.com; sarat@nitrrkl.ac.in

P. Samui  
Department of Civil Engineering,  
Tampere University of Technology, Tampere, Finland  
e-mail: pijush.phd@gmail.com

A. K. Sabat  
Department of Civil Engineering, KIIT University,  
Bhubaneswar 751024, India  
e-mail: akshayasabat@rediffmail.com

T. G. Sitharam  
Indian Institute of Science, Bangalore 560012, India  
e-mail: sitharam@civil.iisc.ernet.in

as the volume change of air-dried undisturbed sample, whereas Seed et al. (1962) defined it as change in volume of a remolded sample. Though factors like clay content, Atterberg's limits, and mineral types are found to affect the swelling potential, the available literature presents contradicting results. McCormack and Wilding (1975) found clay content to be reliable in predicting swelling potential for soil dominated by illite, whereas according to Yule and Ritchie (1980) and Gray and Allbrook (2002), there is no relationship between clay percentage and soil swelling. The cation exchange capacity (CEC), saturation moisture, and plasticity index (PI) are also important indices for estimation of swelling potential (Gill and Reaves 1957). Parker et al. (1977) concluded swell index and PI as superior to other indices for swelling potential. El-Sohby and El-Sayed (1981) observed that parameters like initial water content, type of clay mineral, initial dry density, clay content, and type of coarse grained fraction are major controlling factors for the swelling pressure of soil.

The swelling pressure depends upon various soil parameters such as mineralogy, clay content, Atterberg's limits, dry density, moisture content, initial degree of saturation, etc. along with structural and environmental factors. The parameters are interrelated in a complex manner, and it is difficult to model and analyze effectively taking all the above aspects into consideration. However, it can be measured easily with relevant data pertaining to soil, structure, and environment. So various statistical/empirical methods have been attempted for predicting the swelling pressure based on index properties of soil (Mowafy and Bauer 1985; Mallikarjuna 1988; Das 2002). Recently, Erzin and Erol (2004) presented regression equation for prediction of swelling pressure of Bentonite–Kaolinite clay mixture. However, these regression methods are developed based on the total available data and have not been tested with new data set.

The ANNs are becoming more reliable than statistical method due to their special attributes of identifying complex system when the input and output are known from either laboratory or field experimentation. Kayadelen et al. (2009) presented a neuro-fuzzy model for prediction of swelling potential of compacted soil. The biggest challenge in successful application of ANN is when to stop training. If training is insufficient then the network will not be fully trained, whereas if training is excessive then it will memorize the training pattern or learn noise. When the numbers of data points are scanty the training set is driven to a very small value, but when new data are presented to the network the error is too large, which is known as overfitting. The network needs to be equally efficient for new data during testing or validation, which is called as generalization. There are different methods for generalization like early stopping and cross validation (Basheer 2001; Shahin et al.

2002; Das and Basudhar 2006). In case of early stopping criteria the error on the validation/testing set is monitored during the training process and the training is stopped when the error on the testing set begins to rise. In cross validation an independent test set is used to assess the performance of the model at various stages of learning. However, this method is not suitable if data points are scanty. The 'learning' or 'training' process in ANN, in general, is a nonlinear optimization of an error function. The aim of the training is to minimize the error function to get the optimized weight vectors. The most commonly used error function is the mean squared error (MSE) function. The error associated with weights and sigmoid function is a highly non-linear optimization with many local minima. Local and global optimization methods are carried out for finding out the weight vectors. The steepest descent algorithm and Levenberg–Marquardt (LM) algorithm which are gradient search algorithms are mostly used in ANNs applied to geotechnical engineering problems (Das 2005). As the characteristics of traditional nonlinear programming based optimization method are initial point dependent, the results obtained using backpropagation algorithm are sensitive to initial conditions (weight vectors) (Shahin et al. 2002). The use of global optimization algorithms like genetic algorithm (GA) and simulated annealing for training of ANN, though being widely used in other fields of engineering (Morshed and Kaluarachchi 1998), in geotechnical engineering use of GA for training ANN is limited (Goh 2002; Goh et al. 2005). In recent past another heuristic global optimization, called differential evolution (DE), introduced by Storn and Price (1995) is being used successfully in aerodynamic shape optimization and mechanical design.

The support vector machine (SVM) is an emerging machine learning technology where prediction error and model complexity are simultaneously minimized. It provides a new, efficient novel approach to improve the generalization performance due to which the method has received much attention compared to other artificial intelligence techniques. In many applications the results of SVMs are found to equally good if not better than ANN, but they are computationally efficient and produce an actual mathematical function. This has been used for classification problem and non-linear regression estimations to different engineering problems and also in some geotechnical engineering problems (Goh and Goh 2007; Pal 2006; Samui 2008).

With the above points in view, in the present study ANN models using DE algorithm and Bayesian regularization algorithm for training process are used for prediction of swelling pressure of soil from the inputs: natural moisture content ( $w_n$ ), dry density ( $\gamma_d$ ), liquid limit (LL), plasticity index (PI), and clay fraction (CF). Similarly, different SVM models are also developed based on different kernel

functions. The statistical performance criteria like correlation coefficient ( $R^2$ ), coefficient of efficiency ( $E$ ), overfitting ratio, maximum absolute error (MAE), average absolute error (AAE), and root mean square error (RMSE) are used to evaluate different ANN and SVM models. The obtained ANN parameters (weights and biases) are interpreted to identify important input parameters.

### Methodology

#### Artificial neural network (ANN)

Artificial neural network is an artificial intelligence system/alterative statistical tool inspired by the behavior of human brain and nervous system. A typical structure of ANN consists of a number of processing elements or neurons that are usually arranged in layers: an input layer, an output layer, and one or more hidden layers. The input from each processing element in the previous layer is multiplied by an adjustable connection weight ( $w_{ji}$ ). At each neuron, the weighted input signals are summed and a threshold value (bias) ( $b_j$ ) is added. The combined input ( $I_j$ ) is then passed through a nonlinear transfer function  $\{f(\cdot)\}$  to produce the output of processing element. The adjustable connection weights and biases are obtained by ‘learning’ or ‘training’ process, which is nonlinear optimization of an error function. This is equivalent to the parameter estimation phase in conventional statistical models. In the present study, the ANN models are trained with differential evolution and Bayesian regularization method and are defined as DENN and BRNN, respectively. The use of DENN is not reported and use of BRNN is limited in geotechnical engineering (Goh et al. 2005; Das and Basudhar 2008). The results are compared with that obtained from commonly used Levenberg–Marquardt trained neural networks (LMNN) to discuss the prediction efficiency of the networks. The above neural network models have been developed using MATLAB tool boxes (MathWork 2001). A brief description of the Bayesian regularization and differential evolution neural network is presented here for completeness.

#### Bayesian regularization method

In case of back propagation neural network (BPNN) the error function considered for minimization is the mean square error (MSE). This may lead to overfitting due to unbounded values of weights. The other method, called as regularization, in which the performance function is changed by adding a term that consists of mean square error of weights and biases is given below.

$$MSEREG = \gamma MSE + (1 - \gamma) MSW \tag{1}$$

where MSE is the mean square error of the network,  $\gamma$  is the performance ratio and

$$MSW = \frac{1}{n} \sum_{j=1}^n w_j^2. \tag{2}$$

This performance function will cause the network to have smaller weights and biases thereby forcing networks less likely to be overfit. The optimal regularization parameter  $\gamma$  is determined through Bayesian framework (Demuth and Beale 2000) as the low value of  $\gamma$  will not adequately fit the training data and high value of it may result in overfitting. The number of network parameters (weights and biases) that are being effectively used by the network can be found out by the above algorithm. The effective number of parameters remains the same irrespective of the total number of parameters in the network. In the present study it has been applied using MATLAB routine and it is a built-in function in MATLAB (Math Inc 2001). The above combination works best when the inputs and targets area scaled in the range  $[-1, 1]$  (Demuth and Beale 2000).

#### Differential evolution neural network

The training of the feed-forward neural network using differential evolution optimization is known as differential evolution neural network (DENN) (Ilonen et al. 2003). The DE optimization is a population-based heuristic global optimization method. Unlike other evolutionary optimizations, in DE the vectors in current populations are randomly sampled and combined to create vectors for next generation. The real-valued crossover factor and mutation factor govern the convergence of the search process. The details of DENN are available in Ilonen et al. (2003).

#### Support vector machine

Recently, SVM has been used to solve non-linear regression estimation and time series prediction by introducing  $\varepsilon$ -insensitive loss function (Mukherjee et al. 1997; Muller et al. 1997; Vapnik 1995, 1998). The SVM implements the structural risk minimization principle (SRMP), which has been shown to be superior to the more traditional empirical risk minimization principle (ERMP) employed by many of the other modeling techniques (Osuna et al. 1997; Gunn 1998). SRMP minimizes an upper bound of the generalization error, whereas ERMP minimizes the training error. In this way, it produces the better generalization than traditional techniques.

In this section, a brief introduction is presented on the construction process of SVM for regression problems.

There are three distinct characteristics of SVM when it is used to estimate the regression function. First of all, SVM estimates the regression using a set of linear functions that are defined in a high-dimensional space. Second, SVM carries out the regression estimation by risk minimization where the risk is measured using Vapnik’s  $\epsilon$ -insensitive loss function. Third, SVM uses a risk function consisting of the empirical error and a regularization term which are derived from the SRM principle. In SVM, high generalization performance is achieved by minimizing the sum of the training set error and a term that depends on the Vapnik–Chervonenkis (VC) dimension. This study uses the SVM as a regression technique by introducing a  $\epsilon$ -insensitive loss function. The  $\epsilon$ -insensitive loss function ( $L_\epsilon(y)$ ) can be described in the following way:

$$L_\epsilon(y) = 0 \text{ for } |f(x) - y| \leq \epsilon \text{ otherwise } L_\epsilon(y) = |f(x) - y| - \epsilon \quad (3)$$

This defines a  $\epsilon$  tube (Fig. 1) so that if the predicted value is within the tube, the loss is zero, while if the predicted

point is outside the tube, the loss is the magnitude of the difference between the predicted value and the radius,  $\epsilon$ , of the tube. Assume that the training dataset consists of  $l$  training sample  $\{(x_1, y_1), \dots, (x_l, y_l)\}$  where  $x$  is the input and  $y$  is the output. A problem with learning is related to choosing a function that predicts the actual response  $y$  as closely as possible, with a precision of  $\epsilon$ . For this study,  $x = [w_n, \gamma_d, LL, PI, CF]$  and  $y = [SP]$ . Let us assume a linear function

$$f(x) = (w \cdot x) + b \quad (4)$$

where  $w \in R^n$  and  $b \in r$ ;  $w$  = an adjustable weight vector;  $b$  = the scalar threshold;  $R^n$  =  $n$ -dimensional vector space; and  $r$  = one-dimensional vector space.

The main aim of SVM is to find a function  $f(x)$  that gives a deviation  $\epsilon$  from the actual output ( $y$ ), which is, at the same time, as flat as possible. Flatness in the case of Eq. 4 means that one seeks a small  $w$ . One way of obtaining this is by minimizing the Euclidean norm  $\|w\|^2$  (Smola and Scholkopf 2004). The convex optimization problem thus involves

$$\begin{aligned} \text{Minimize : } & \frac{1}{2} \|w\|^2 \\ \text{Subjected to : } & y_i - (\langle w \cdot x_i \rangle + b) \leq \epsilon, \quad i = 1, 2, \dots, l \\ & (\langle w \cdot x_i \rangle + b) - y_i \leq \epsilon, \quad i = 1, 2, \dots, l \end{aligned} \quad (5)$$

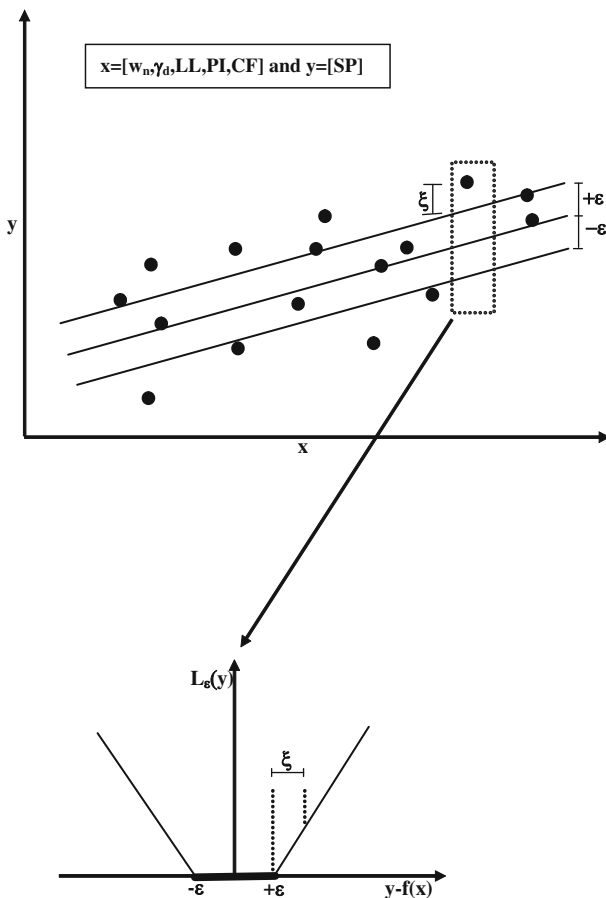
The best regression line is defined by minimizing the following cost function:

$$\begin{aligned} \text{Minimize : } & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{Subjected to : } & y_i - (\langle w \cdot x_i \rangle + b) \leq \epsilon + \xi_i, \quad i = 1, 2, \dots, l \\ & (\langle w \cdot x_i \rangle + b) - y_i \leq \epsilon + \xi_i^*, \quad i = 1, 2, \dots, l \\ & \xi_i \geq 0 \text{ and } \xi_i^* \geq 0, \quad i = 1, 2, \dots, l \end{aligned} \quad (6)$$

Hence, the Lagrangian function is constructed from both the objective function and corresponding constraints in Eq. 6 as follows:

$$\begin{aligned} L(w, \xi, \xi^*, \alpha, \alpha^*, \gamma, \gamma^*) = & \frac{\|w\|^2}{2} + C \left( \sum_{i=1}^l (\xi_i + \xi_i^*) \right) \\ & - \sum_{i=1}^l \alpha_i [\epsilon + \xi_i - y_i + \langle w \cdot x_i \rangle + b] \\ & - \sum_{i=1}^l \alpha_i^* [\epsilon + \xi_i^* + y_i - \langle w \cdot x_i \rangle - b] - \sum_{i=1}^l (\gamma_i \xi_i + \gamma_i^* \xi_i^*) \end{aligned} \quad (7)$$

where  $L$  is the Lagrangian and  $\alpha, \alpha^*, \gamma,$  and  $\gamma^*$  are the Lagrangian multipliers. The partial derivative of  $L$  with respect to  $\omega, b, \xi$  and  $\xi^*$  need to be zero to satisfy the saddle point condition.



**Fig. 1** Shows a typical architecture of SVM for swelling pressure (SP) prediction with pre-specified accuracy  $\epsilon$  and slack variable  $\xi$  in support vector regression (Scholkopf 1997)

$$\begin{aligned} \partial_w L = 0 &\Rightarrow w = \sum_{i=1}^l x_i (\alpha_i - \alpha_i^*) \\ \partial_b L = 0 &\Rightarrow \sum_{i=1}^l \alpha_i = \sum_{i=1}^l \alpha_i^* \\ \partial_{\xi} L = 0 &\Rightarrow \sum_{i=1}^l \gamma_i = \sum_{i=1}^l (C - \alpha_i) \\ \partial_{\xi^*} L = 0 &\Rightarrow \sum_{i=1}^l \gamma_i^* = \sum_{i=1}^l (C - \alpha_i^*) \end{aligned} \tag{8}$$

Substituting (8) into (7) yields the dual optimization problem

$$\begin{aligned} \text{Maximize : } & -\varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) + \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ & - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) (x_i \cdot x_j) \\ \text{Subjected to : } & \sum_{i=1}^l \alpha_i = \sum_{i=1}^l \alpha_i^* ; 0 \leq \alpha_i^* \leq C \text{ and } 0 \leq \alpha_i \leq C \end{aligned} \tag{9}$$

The coefficients  $\alpha_i$  and  $\alpha_i^*$  are determined by solving the above optimization problem (Eq. 9). An important aspect is that some Lagrange multipliers ( $\alpha_i, \alpha_i^*$ ) will be zero, implying that these training objects are considered to be irrelevant for the final solution (sparseness). The training objects with nonzero Lagrangian multipliers are called support vectors. These are the objects with prediction errors larger than  $\pm\varepsilon$ . In this way, the value of  $\varepsilon$  determines the amount of support vectors. Obviously, if  $\varepsilon$  is too large, too few support vectors are selected, which leads to a decrease in the final prediction performance. Furthermore, if the Lagrangian multipliers of the support vectors all have different values, it means that one support vector is considered to be more important than another one. So Eq. 4 can be written as

$$f(x) = \sum_{(\text{support vectors})} (\alpha_i - \alpha_i^*) (x_i \cdot x) + b$$

where

$$b = -\left(\frac{1}{2}\right) w \cdot [x_r + x_s]. \tag{10}$$

From Eq. 10 it is clear that  $w$  has been completely described as a linear combination of training patterns. So, the complexity of a function represented by support vectors is independent of the dimensionality of input space and it depends only on the number of support vectors. The entry of the data in inner products is very important because (1) the dimension of the objects does not appear in the problem

to be solved and (2) extension of this linear approach to nonlinear regression can be easily made.

When linear regression is not appropriate, then the input data must be mapped into a high-dimensional feature space through some nonlinear mapping (Boser et al. 1992) (see Fig. 2). In optimization problem expressed in Eq. 9,  $x$  has been replaced by the feature space,  $\Phi(x)$ . So, the optimization problem (Eq. 9) can be written as

$$\begin{aligned} \text{Maximize : } & -\varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) + \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ & - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) (\Phi(x_i) \cdot \Phi(x_j)) \\ \text{Subjected to : } & \sum_{i=1}^l \alpha_i = \sum_{i=1}^l \alpha_i^* ; 0 \leq \alpha_i^* \leq C \text{ and } 0 \leq \alpha_i \leq C \end{aligned} \tag{11}$$

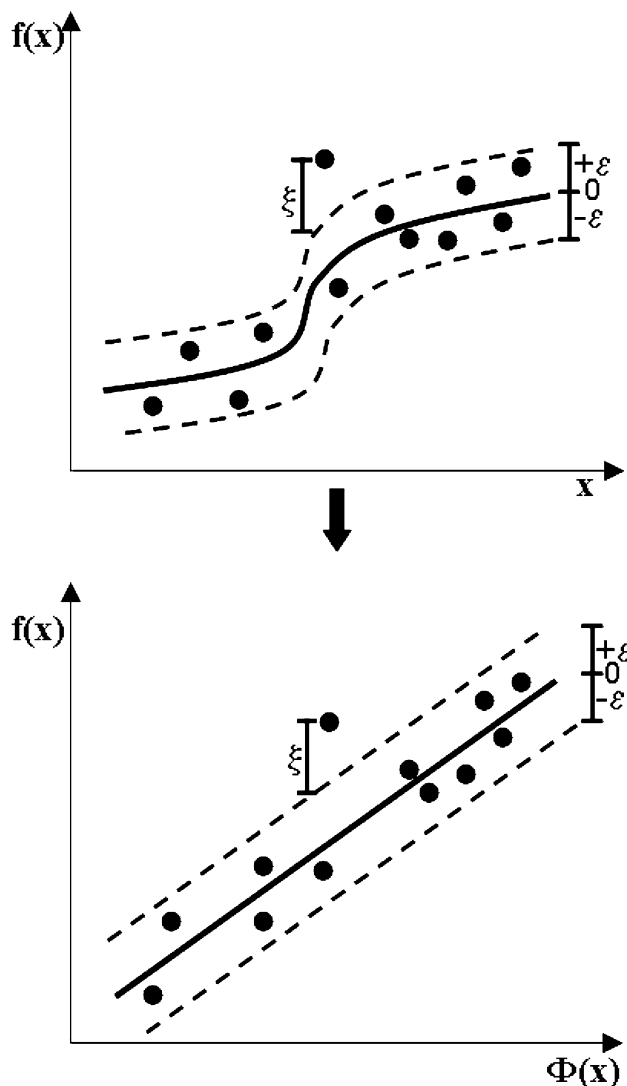


Fig. 2 Concept of nonlinear regression

The concept of a kernel function  $[K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)]$  has been introduced to reduce computational demand (Cristianini and Shwae-Taylor 2000; Cortes and Vapnik 1995). Hence, the optimization problem can be written as

$$\begin{aligned} \text{Maximize : } & -\varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) + \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ & - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) (K(x_i \cdot x_j)) \\ \text{Subjected to : } & \sum_{i=1}^l \alpha_i = \sum_{i=1}^l \alpha_i^* \quad 0 \leq \alpha_i^* \leq C \quad \text{and} \quad 0 \leq \alpha_i \leq C \end{aligned} \tag{12}$$

The introduction of kernels according to Mercer’s theorem (Vapnik 1995) avoids an explicit formation of the nonlinear mapping, makes the dimension of feature space even infinite, and reduces the computational load greatly by enabling the operation in low dimensional input space instead of high-dimensional feature space. Some common kernels, such as a polynomial (homogeneous), polynomial (nonhomogeneous), radial basis function, Gaussian function, and sigmoid functions have been used for non-linear cases. The regression function Eq. 4 has been obtained by applying the same procedure as in the linear case. Figure 3 shows the architecture of a SVM for prediction of swelling pressure. Three kernel functions used in this study are polynomial, radial basis, and spline functions. The SVM models developed using radial

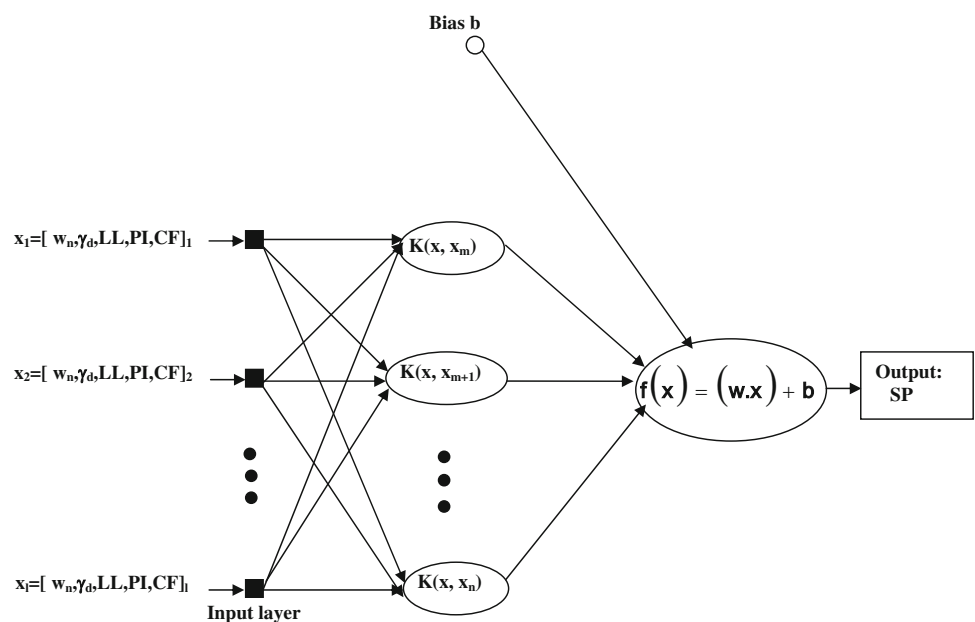
function, polynomial, and spline functions are named as SVM-R, SVM-P and SVM-S, respectively.

**Results and discussion**

The data from various sources available in literature (Aciroyd and Husain 1986; Abdujauwad 1994; Abdujauwad et al. 1998) are taken with input parameters are natural moisture content ( $w_n$ ), dry density ( $\gamma_d$ ), LL, PI, clay fraction (CF), and swelling pressure (SP) as output. The total number of data points considered is 230 out of which 167 are taken for training and 63 are taken for testing. The maximum, minimum, average, and standard deviation for the data used are shown in Table 1 and it can be seen that it covers a wide range of values. The successful application of a method depends upon the identification of suitable input parameters. Table 2 shows the cross correlation between the inputs and output; it can be seen that CF, LL, and PI are found to be important input parameters for predicting swelling pressure.

The results of different ANN models using the above parameters are shown in Table 3. The correlation coefficient ( $R^2$ ) and root means square error (RMSE) are mostly for performance criteria evaluation of ANN models. However,  $R^2$  is a biased parameter and sometimes, higher values of  $R^2$  may not necessarily indicate better performance of the model because of the tendency of the model to be biased toward higher or lower values (Das and

**Fig. 3** Architecture of SVM for swelling pressure (SP) prediction



**Table 1** Parameters of the data considered for the present study

	$w_n$ (%)	$\gamma_d$ (kN/m <sup>3</sup> )	LL	PI	Clay fraction	Swelling pressure (kN/m <sup>2</sup> )
Maximum	63.90	15.70	193.00	165.00	97.00	805.00
Minimum	2.70	1.04	26.00	12.00	19.00	3.00
Average	18.31	9.33	90.74	59.06	41.19	122.61
Standard deviation	9.58	5.89	47.77	43.42	14.04	140.90

**Table 2** Cross correlation between the inputs and output

$w_n$	$\gamma_d$	LL	PI	CF	SP
1.00	-0.15	-0.04	0.02	0.29	-0.19
	1.00	0.19	0.29	-0.44	0.01
		1.00	0.90	0.24	0.37
			1.00	0.10	0.25
				1.00	0.42
					1.00

**Table 3** General performance of different neural network models

ANN models	Training data		Testing data		Overfitting ratio
	$R^2$	$E$	$R^2$	$E$	
DENN	0.91	0.91	0.76	0.75	1.37
BRNN	0.97	0.96	0.81	0.79	2.12
LMNN	0.90	0.90	0.77	0.74	1.43
SVM	0.96	0.96	0.88	0.88	1.40
MLR	0.65	0.65	0.49	0.43	1.39

Basudhar 2006). The coefficient of efficiency ( $E$ ) is also considered. The  $E$  is defined as

$$E = \frac{E_1 - E_2}{E_1} \tag{13}$$

where

$$E_1 = \sum_{t=1}^N (SP_m - \overline{SP_m})^2$$

$$E_2 = \sum_{t=1}^N (SP_p - SP_m)^2 \tag{14}$$

and  $SP_m$ ,  $\hat{SP}_m$  and  $SP_p$  are the measure, average, and predicted swelling pressure, respectively. The  $E$  value compares the modeled and measured values of the variable and evaluates how far the network is able to explain total variance in the data set. The overfitting ratio is defined as the ratio of RMSE for testing and training data, and it defines the generalization. A multilinear regression (MLR) analysis was also done considering input parameters as per Erzin and Erol (2004) and the results are presented in Table 3. The developed ANN models are found to be more efficient compared to MLR in terms  $R^2$  and  $E$  values and

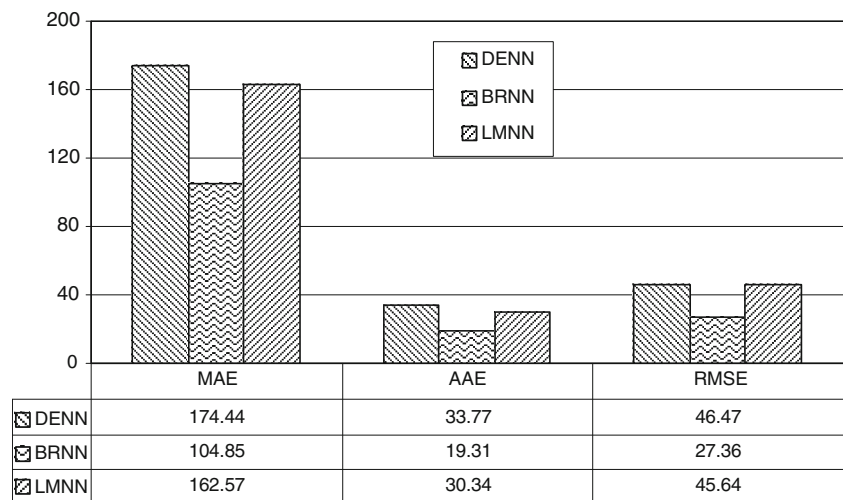
have a strong correlation between predicted and observed values with  $|R| > 0.8$  (Smith 1986). It can also be seen that comparing the values of  $R^2$  and  $E$  values for training and testing data, BRNN is found to better than DENN and LMNN. However, DENN is having good generalization with small overfitting ratio, followed by LMNN and BRNN.

The RMSE value only defined the efficiency of a model overall; however, MAE can reveal the presence of regional areas of poor prediction. Figures 4 and 5 show the value of MAE, AAE, and RMSE for different ANN models for training and testing data, respectively. It can be seen that for training data, BRNN is having lowest values of MAE, AAE, and RMSE. However, for testing data, AAE is comparable for all the methods, but based on MAE and RMSE values BRNN performs better than DENN and LMNN. Hence, based on different statistical performance criteria for the present study it can be concluded that BRNN is better followed by DENN and LMNN. The weights and biases of the final network are presented in Table 4 for BRNN. The weights and biases can be utilized for selection of important input parameters based on its interpretation.

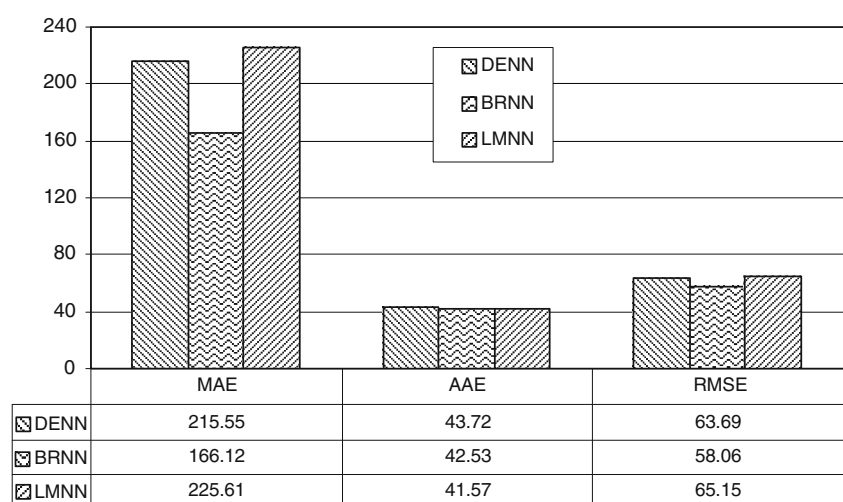
The ANN is considered as a ‘Black box’ system due to insufficient explanations to the weight vectors, but methods like Garson’s algorithm and connection weight approach have been used utilizing the weight vector to identify the important input vectors (Das and Basudhar 2006). Such a study also made here to compare the above two methods in identifying the important parameters. Table 5 shows the ranking of important input parameters as calculated from Garson’s algorithm and connection weight approach with the weights obtained from DENN, BRNN, and LMNN. It can be seen from that the ranking of important input parameters as obtained by Garson’s algorithm and Connection weight approach is different for BRNN and LMNN, whereas for DENN the ranking of first and second parameters are same by both the methods.

The successful application of SVM models depends upon suitable parameters like type of kernel function and the parameters  $C$  and  $\epsilon$  are obtained by trial and error. Table 6 presents the results of SVM models developed. Based on  $R^2$  and  $E$  values SVM model with radial basis kernel function (SVM-R) found to be more efficient compared to models developed with other kernel functions (SVM-P and

**Fig. 4** Comparison of prediction capabilities of ANN models for training data



**Fig. 5** Comparison of prediction capabilities of ANN models for testing data



**Table 4** Weights and biases of BRNN for swelling pressure of soil

Neuron	Weights						Biases	$b_0$
	$w_n$ (%)	$\gamma_d$ (kN/m <sup>3</sup> )	LL (%)	PI (%)	CF (%)	SP (kN/m <sup>2</sup> )	$b_{hk}$	
Hidden neuron 1 ( $k = 1$ )	-0.885	1.107	-6.344	2.770	1.017	3.388	-0.160	-1.773
Hidden neuron 2 ( $k = 2$ )	-0.511	0.455	-2.382	-1.537	0.813	-3.938	-1.337	
Hidden neuron 3 ( $k = 3$ )	0.423	-1.160	-2.868	2.657	-0.199	-1.379	-0.295	

SVM-S). From Table 6, it is clear that SVM model employs 65–75% (radial basis function = 74.85%, polynomial kernel = 65.26% and spline kernel = 66.46%) of the training patterns as support vectors. So, SVM is remarkable in producing an excellent generalization level while maintaining the sparsest structure. Sparseness means that a significant number of the weights are zero (or effectively zero), which has the consequence of producing compact, computationally efficient models, which in addition are simple and therefore produce smooth functions. In SVM, support vectors represent prototypical

examples. The prototypical examples exhibit the essential features of the information content of the data, and thus are able to transform the input data into the specified targets. Figures 6 and 7 show the value of MAE, AAE, and RMSE for different SVM models for training and testing data, respectively. It can be seen that for training data SVM-R is having lowest values of MAE, AAE, and RMSE. In comparison to ANN models SVM-R model is found to be better than all the ANN models. The use of the SRM principle in defining cost function provided more generalization capacity with the SVM compared to the ANN, which uses



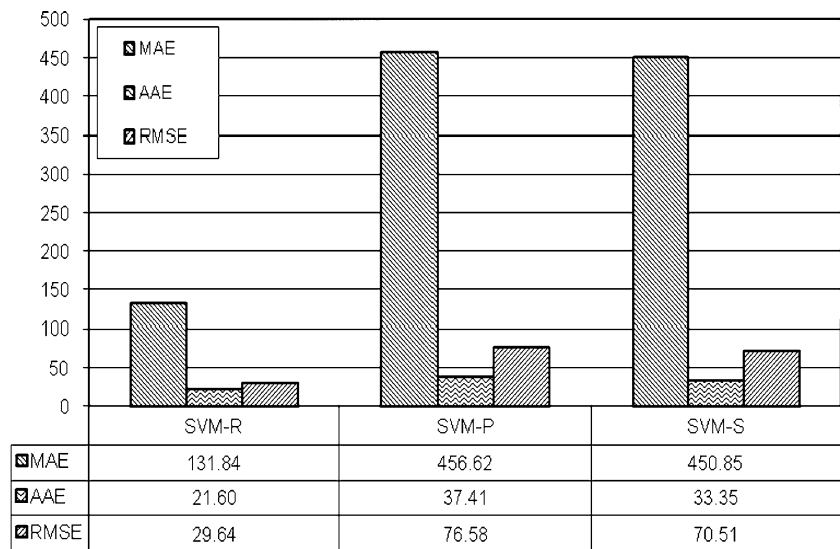
**Table 5** Relative importance of different input parameters

Input parameters	Garson's algorithm			Connection weight approach			Ranking of inputs as per as per SVM-R results	
	Ranking of inputs as per relative importance			Ranking of inputs as per relative importance			S <sub>i</sub> value (%)	Ranking
	DENN	BRNN	LMNN	DENN	BRNN	LMNN		
w <sub>n</sub>	5	5	4	4	4	4	17.80	3
γ <sub>d</sub>	3	3	3	5	3	3	23.95	2
LL	2	1	1	2	2	2	16.40	4
PI	1	2	2	1	1	1	36.65	1
CF	4	4	5	3	5	5	6.50	5

**Table 6** General performance of SVM for different kernels

Kernel	C	ε	Training performance		Testing performance		Number of support vector
			R <sup>2</sup>	E	R <sup>2</sup>	E	
Radial basis function, width (σ) = 2.6 (SVM-R)	20	0.009	0.958	0.958	0.885	0.887	125
Polynomial, degree = 2 (SVM-P)	10	0.01	0.748	0.726	0.425	0.018	109
Spline (SVM-S)	4	0.01	0.792	0.768	0.738	0.657	112

**Fig. 6** Comparison of prediction capabilities of SVM models for training data

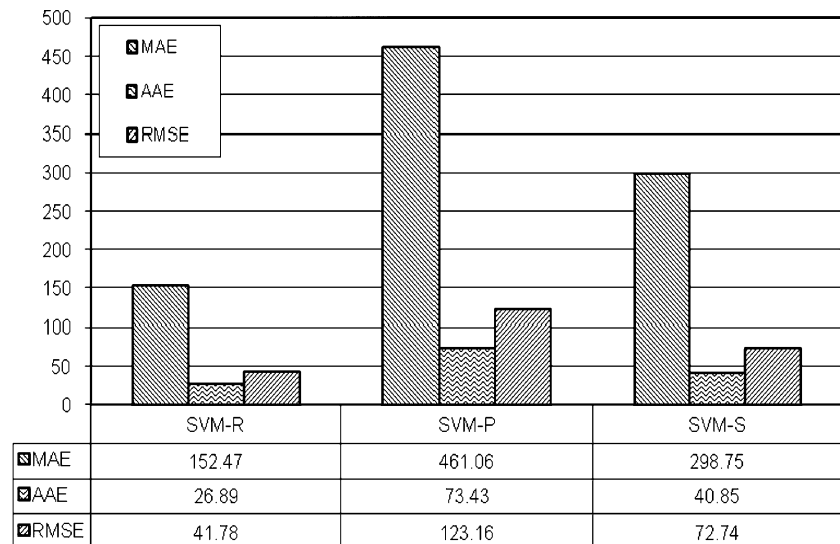


the empirical risk minimization principle. SVM uses only three parameters (radial basis function: σ, C, and ε; polynomial kernel: degree of polynomial, C, and ε). In ANN, there are a larger number of controlling parameters, including the number of hidden layers, number of hidden nodes, learning rate, momentum term, number of training epochs, transfer functions, and weight initialization methods. Obtaining an optimal combination of these parameters is a difficult task as well. Another major advantage of the SVM is its optimization algorithm, which includes solving a linearly constrained quadratic programming function

leading to a unique, optimal, and global solution compared to the ANN. In SVM, the number of support vectors is determined by algorithm rather than by trial-and-error, which has been used by ANN for determining the number of hidden nodes.

In this study, a sensitivity analysis has been carried out to extract the cause and effect relationship between the inputs and outputs of the SVM model. The basic idea is that each input of the model is offset slightly and the corresponding change in the output is reported. The procedure has been taken from the work of Liong et al. (2000).

**Fig. 7** Comparison of prediction capabilities of SVM models and for testing data



According to Liong et al. (2000), the sensitivity ( $S$ ) of each input parameter has been calculated by the following formula:

$$S_i = \frac{1}{N} \sum_i^N \left( \frac{\% \text{Change in output}}{\% \text{Change in input}} \right) \times 100 \quad (15)$$

where  $N$  is the number of data points. The analysis has been carried out on the trained model by varying each of input parameter, one at a time, at a constant rate of 20%. The result of the above analysis is also presented in Table 5. According to the sensitivity analysis using SVM-model, PI is found to be more important parameters followed by  $\gamma_d$  and  $w_n$ , which is similar to ANN analysis using connection weight approach.

## Conclusions

The different ANN techniques and SVM models examined here have shown the ability to build accurate models with high predictive capabilities for prediction of swelling pressure of soil from the inputs: natural moisture content ( $w_n$ ), dry density ( $\gamma_d$ ), liquid limit (LL), plasticity index (PI), and clay fraction (CF). The ANN and SVM models are found to be more efficient compared to MLR analysis. Based on different statistical performance criteria, the Bayesian regularization neural network (BRNN) model is found to be more efficient compared to DENN and LMNN. However, the DENN model is found to better in terms of generalization. The performance of the developed SVM model is better than the developed ANN models. The ranking of important input parameters is found to be consistent as per connection weight approach for the ANN model considered here. However, while using Garson's algorithm the ranking is found to be different for different

ANN models. Developed ANN and SVM models have the advantage that once the model is trained, it can be used as an accurate and quick tool for predicting swelling pressure without a need to perform any manual work such as using tables or charts. Comparison between the ANN and SVM model indicates that SVM model is superior to ANN model for predicting swelling pressure.

## References

- Abdujauwad SN (1994) Swelling behaviour of calcareous clays from the Eastern Provinces of Saudi Arabia. *Q J Eng Geol* 27:333–351
- Abdujauwad SN, Al-Sulaimani GJ, Basunbul IA, Si-Buraim I (1998) Laboratory and field studies of response of structures to heave of expansive clay. *Geotechnique* 48(1):103–121
- Aciroyd LW, Husain R (1986) Residual and lacustrine black cotton soil of North-East Nigeria. *Geotechnique* 36:113–118
- Basheer IA (2001) Empirical modeling of the compaction curve of cohesive soil. *Can Geotech J* 38(1):29–45
- Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifier. In: *The Proceedings of the fifth annual ACM workshop on computational learning theory*, 27–29 July, Pittsburgh, PA, USA
- Cortes C, Vpanik VN (1995) Support-vector networks. *Mach Learn* 20(3):273–297
- Cristianini N, Shawe-Taylor J (2000) *An introduction to support vector machine*. University press, London, Cambridge
- Das BM (2002) *Principles of geotechnical engineering*. Brookes-Cole, New York
- Das SK (2005) Application of genetic algorithm and artificial neural network to some geotechnical engineering problem. Ph.D. thesis submitted to Indian Institute of Technology, Kanpur, India
- Das SK, Basudhar PK (2006) Undrained lateral load capacity of piles in clay using artificial neural network. *Comput Geotech* 33(8):454–459
- Das SK, Basudhar PK (2008) Prediction of residual friction angle of clays using artificial neural network. *Eng Geology* 100(3-4): 142–145

- Demuth H, Beale M (2000) Neural network toolbox. The MathWorks Inc, USA
- El-Sohby MA, El-Sayed AR (1981) Some factors affecting swelling of clayey soils. *Geotech Eng* 12:19–39
- Erzin Y, Erol O (2004) Correlations for quick prediction of swelling pressure. *Electron J Geotech Eng* 9, Bundle F, paper 0476
- Gill WR, Reaves CA (1957) Relationships of Atterberg limits and cation-exchange capacity to some physical properties of soil. *Soil Sci Soc Am Proc* 21:491–494
- Goh ATC (2002) Probabilistic neural network for evaluating seismic liquefaction potential. *Can Geotech J* 39:219–232
- Goh ATC, Goh SH (2007) Support vector machines: their use in geotechnical engineering as illustrated using seismic liquefaction data. *Comput Geotech* 34(5):410–421
- Goh ATC, Kulhawy FH, Chua CG (2005) Bayesian neural network analysis of undrained side resistance of drilled shafts. *J Geotech Geoenviron Eng ASCE* 131(1):84–93
- Gourley CS, Newill D, Shreiner HD (1993) Expansive soils: TRL's research strategy. In: Proceedings of first international symposium on engineering characteristics of arid soils
- Gray CW, Allbrook R (2002) Relationships between shrinkage indices and soil properties in some New Zealand soils. *Geoderma* 108(3–4):287–299
- Gunn S (1998) Support vector machines for classification and regression. Image speech and intelligent systems technical report, University of Southampton, UK
- Holtz WG (1959) Expansive clays—properties and problems. *Q Colo Sch Mines* 54(4):89–117
- Ilonen J, Kamarainen JK, Lampinen J (2003) Differential evolution training algorithm for feed-forward neural network. *Neural Process Lett* 17:93–105
- Jones DE, Holtz WG (1973) Expansive soils—the hidden disaster. Civil engineering. *ASCE* 43(8):49–51
- Kayadelen C, Taskiran T, Gnaydin O, Fener M (2009) Adaptive neuro-fuzzy modeling for the swelling potential of compacted soils. *Environ Earth Sci*. doi:10.1007/s12665-009-0009-5
- Liong SY, Lim WH, Paudyal GN (2000) River stage forecasting in Bangladesh: neural network approach. *J Comput Civ Eng* 14(1):1–8
- Mallikarjuna R (1988) A proper parameter for correlation of swell potential and swell index for remould expansive clays. M. Tech. dissertation submitted to dept. of civil engg, J.N.T. University College of Engg, Kakinada, India
- MathWork, Inc. (2001) Matlab user's manual, version 6.5. The MathWorks, Inc., Natick
- McCormack DE, Wilding LP (1975) Soil properties influencing swelling in Canfield and Geeburg soils. *Soil Sci Soc Am Proc* 39:496–502
- Morshed J, Kaluarachchi JJ (1998) Parameter estimation using artificial neural network and genetic algorithm for free-product migration and recovery. *Water Resour Res* 34(5):1101–1113
- Mowafy MY, Bauer GE (1985) Prediction of swelling pressure and factors affecting the swell behavior of an expansive soils. *Transp Res Rec* 1032:23–28
- Mukherjee S, Osuna E, Girosi F (1997) Nonlinear prediction of chaotic time series using support vector machine. In: Proceedings of IEEE workshop on neural networks for signal processing 7, Institute of Electrical and Electronics Engineers, New York, pp 511–519
- Muller KR, Smola A, Ratsch G, Scholkopf B, Kohlmorgen J, Vapnik VN (1997) Predicting time series with support vector machines. In: Proceedings of international conference on artificial neural networks. Springer-Verlag, Berlin, p 999
- Nelson JD, Miller DJ (1992) Expansive soils: problem and practice in foundation and pavement engineering. Wiley, New York
- Osuna E, Freund R, Girosi F (1997) An improved training algorithm for support vector machines. In: Proceedings of IEEE workshop on neural networks for signal processing 7, Institute of Electrical and Electronics Engineers, New York, pp 276–285
- Pal M (2006) Support vector machine-based modelling of seismic liquefaction potential. *Int J Numer Anal Methods Geomech* 30:983–996
- Parker JC, Amos DF, Kaster DL (1977) An evaluation of several methods of estimating soil volume change. *Soil Sci Soc Am J* 41:1059–1064
- Samui P (2008) Slope stability analysis: a support vector machine approach. *Environ Geol* 56:255–267
- Scholkopf B (1997) Support vector learning. R. Oldenbourg, Munich
- Seed HB, Woodward RJ Jr, Lundgren R (1962) Prediction of swelling potential for compacted clays. *J Soil Mech Found Div Am Soc Civ Eng* 88(SM3):53–87
- Shahin MA, Maier HR, Jaksa MB (2002) Predicting settlement of shallow foundations using neural network. *J Geotech Geoenviron Eng ASCE* 128(9):785–793
- Smith GN (1986) Probability and statistics in civil engineering: an introduction. Collins, London
- Smola AJ, Scholkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14:199–222
- Storn R, Price K (1995) Differential evolution—a simple and efficient adaptive scheme for global optimization over continuous spaces. Technical report TR-95-012. International Computer Science Institute, Berkeley, CA, USA
- Vapnik V (1995) The nature of statistical learning theory. Springer, New York
- Vapnik V (1998) Statistical learning theory. Wiley, New York
- Yule DF, Ritchie JT (1980) Soil shrinkage relationships of Texas vertisols. 1. Small cores. *Soil Sci Soc Am J* 44:1285–1291