Cent. Eur. J. Geosci. • 3(4) • 2011 • 449-461
DOI: 10.2478/s13533-011-0043-1

VERSITA

Central European Journal of **Geosciences**

Machine learning techniques applied to prediction of residual strength of clay

Research Article

Sarat Kumar Das[1]*, Pijush Samui[2] †, Shakilu Zama Khan[3], Nagarathnam Sivakugan[4]

1 *Associate Professor, Civil Engineering Department, National Institute of Technology, Rourkela, India*

2 *Associate Professor, Centre for Disaster Mitigation and Management, VIT University, Vellore-632014, India*

3 *Civil Engineering Department, BIET, B.P. University of Technology, Orissa, India*

4 *Civil & Environmental Engineering, James Cook University, Queensland, Australia*

Received 11 September 2011; accepted 16 November 2011

Abstract: Stability with first time or reactivated landslides depends upon the residual shear strength of soil. This paper describes prediction of the residual strength of soil based on index properties using two machine learning techniques. Different Artificial Neural Network (ANN) models and Support Vector Machine (SVM) techniques have been used. SVM aims at minimizing a bound on the generalization error of a model rather than at minimizing the error on the training data only. The ANN models along with their generalizations capabilities are presented here for comparisons. This study also highlights the capability of SVM model over ANN models for the prediction of the residual strength of soil. Based on different statistical parameters, the SVM model is found to be better than the developed ANN models. A model equation has been developed for prediction of the residual strength based on the SVM for practicing geotechnical engineers. Sensitivity analyses have been also performed to investigate the effects of different index properties on the residual strength of soil.

Keywords: landslides • residual strength • index properties • prediction model • artificial neural network • support vector machine

1. Introduction

Stability of natural slopes or landslides depends upon the shear strength parameters of clay, which varies significantly between the peak and residual states. This is especially true in sensitive clays where there is a substantial reduction in shear strength from a peak to a residual state. At a residual state, the clay has undergone a large strain and hence remoulded, with previous bonds broken and prior fabric destroyed. As a result, in a residual state the clay exhibits no cohesion. Further, the friction angle (ϕ_r) is substantially less than the friction angle at a peak state (ϕ_p), resulting in lower shear strength. In geotech-

*E-mail: saratdas@rediffmail.com
†E-mail: pijush.phd@gmail.com, Tel:91-416-2202283, Fax: 91-416-2243092

nical problems involving large deformations such as landslides, it is appropriate that the residual shear strength parameters ($c_r = 0$ and ϕ_r) be used in the analysis.

The residual strength of soils has received considerable attention after Skempton suggested that the stability of reactivated landslides are governed by residual strength [1]. Meshri and Shahien [2] analysed 99 clay slopes and observed that the shear strength of a clay is close to the residual value in most cases and, even for many of the first-time slope failures, it appears that part of the slip surface is in the residual condition.

There had been attempts to develop correlations between the residual friction angle of soils and index properties such as Atterberg limits, and clay fraction. Skempton [1], related the ϕ_r value with the clay fraction (CF). Generally, ϕ_r decreases with liquid limit (LL), and for a given LL and CF, ϕ_r decreases with increase in normal effective stress [3]. Mesri and Cepeda-Diaz [4] presented a correlation between ϕ_r and LL. Colotta *et al.* [5] have given a correlation between ϕ_r and a parameter which is a function of LL, plasticity index (PI) and CF. For sedimentary soil, Stark and Eid [6] observed that the type of minerals and percent of clay govern the value of ϕ_r. Using LL as an indicator of clay mineral, they proposed correlations of ϕ_r with LL for various ranges of CF. Wesley [7] found that for tropical clays ϕ_r can be better related with ΔPI, the deviation from the A-line in Casagrande's classification chart given by:

$$\Delta PI = PI - 0.73(LL - 20) \quad (1)$$

The data points were scattered and the correlation is valid for highly plastic clays where LL > 50. All the above correlations are graphical. Sridharan and Rao [8] observed that while ϕ_r is related to LL, PI and CF, it gives the best correlation with CF.

Using data from soil specimens collected from more than 80 natural disaster areas, Tiwari and Marui [9], proposed a triangular chart to predict the residual friction angle ϕ_r, based on the mineralogical composition. The correlation of ϕ_r is with the liquid limit, the plasticity index, the clay fraction, the specific surface area, and the proportion of clay mineral smectite. The model gave good predictions for the samples they had tested, but underpredicted the 53 samples tested by other researchers. Kaya and Kwong [10] studied soil properties of some active landslides in Hawaii to evaluate correlations between the residual strength and soil indices. The results showed that for the colluvial soils, which are rich in an amorphous phase, there is poor correlation between soil index properties and residual friction angle. Moreover, it is difficult to find out the specific surface area of the clay minerals and the mineral composition of the clay soil samples.

From the above discussions, it was quite clear that ϕ_r is a function of several variables that can be represented by index properties. Nevertheless, most of the relationships are developed in terms of charts. The correlation given by Tiwari and Marui [9] is based on the mineralogical content of the clay but also is presented in graphical form. For a practicing geotechnical engineer, finding out the specific surface and mineralogical content is not always handy. Hence, it is highly desirable to develop a correlation or method where ϕ_r can be expressed in terms of all the relevant variables. Das and Basudhar [11] used artificial neural network modelling to predict the residual friction angle of clay with a high degree of accuracy. But the above study is limited to tropical soil of a specific region only. Hence, there is an acute need to develop a suitable and efficient method applicable to soils of different origins.

In the present study three ANN models (BRNN, LMNN and DENN) and three SVM models (SVM-G, SVM-P, SVM-S) have been developed to predict the residual friction angle of clay based on the index properties of soil (LL, PI, CF and ΔPI). The statistical performance criteria, such as overfitting ratio, maximum absolute error, average absolute error and root mean square error, are used to evaluate the quality of predictions that come from the different ANN and SVM models. A model equation based on the above analysis is also presented to show correlation, which can be determined using spreadsheet inputs. The degree of importance of the input parameters are identified using SVM and ANN models.

2. Methodology

In the present study artificial intelligence techniques ANN and SVM have been used to examine the effects of the index properties on the residual friction angle. The salient features of the ANN and SVM models are presented here briefly for completeness.

2.1. Artificial Neural Network (ANN)

Artificial Neural Network modelling is an artificial intelligence system which is an alterative statistical tool inspired by the behaviour of the human brain and nervous system. It has been applied successfully to a wide range of geotechnical problems. In the present study the ANN models are developed using two different training algorithms, Bayesian Regularization Neural Network (BRNN) and Differential Evolution Neural Network (DENN), and are compared with the commonly used Levenberg-Marquardt Neural Network algorithms (LMNN).

2.1.1. Bayesian Regularization Method (BRNN)

In the case of using a Back Propagation Neural Network (BPNN), the error function considered for minimization is the Mean Square Error (MSE). This may lead to overfitting due to unbound values of weights. In a Bayesian regularization method, the performance function is changed by adding a term that consists of the mean square error of weights and biases as:

$$\text{MSEREG} = \gamma\,\text{MSE} + (1-\gamma)\,\text{MSW} \tag{2}$$

where MSE is the mean square error of the network, γ is the performance ratio and MSW is the mean square of weights and defined as:

$$\text{MSW} = \frac{1}{n}\sum_{j=1}^{n} w_j^2 \tag{3}$$

Where w_j is the weight.
This performance function will cause the network to have smaller weights and biases, thereby reducing the likelihood of overfitting. The optimal regularization parameter γ is determined through a Bayesian framework [12], as the low value of γ will not adequately fit the training data and a high value of it may result in overfit. The number of network parameters (weights and biases) effectively used by the network can be determined by the above algorithm. The effective number of parameters remains the same irrespective of the total number of parameters in the network. The present study was carried out using the built-in functions in MATLAB [13]. Demuth and Beale [12] suggested that the above combination works best when the inputs and targets area are scaled in the range [-1, 1].

2.1.2. Differential Evolution Neural Network (DENN)

The training of the feed-forward neural network using differential evolution optimization is known as a Differential Evolution Neural Network (DENN) [14]. The DE optimization is a population-based heuristic global optimization method. Unlike other evolutionary optimization techniques, in DE the vectors in current populations are randomly sampled and combined to create vectors for the next generation. The real valued crossover factor and mutation factor govern the convergence of the search process. The details of DENN are available in the literature [14]. The results of an ANN model trained with DENN and BRNN are compared with that from the commonly used LMNN to discuss the quality of the prediction of the different networks.
LMNN is the most widely used ANN method and has been extensively used in geotechnical engineering [11, 15, 16] and is not elaborated here.

2.2. Support Vector Machine

Support Vector Machine (SVM) originated from the concept of statistical learning theory pioneered by Boser *et al.* [17]. Our study uses the SVM as a regression technique by introducing an error (ε) ε-insensitive loss function. In this section, a brief introduction of how to construct an SVM for regression problems is presented. Further details can be found in many publications [17–21]. There are three distinct characteristics when SVM is used to estimate the regression function: the type of kernel function, the optimum capacity factor C, and the optimum error insensitive zone ε. Consider a set of training data$\{(x_1, y_1), \ldots, (x_l, y_l)\}$, $x \in R^n$, $y \in r$, where x is the input, y is the output, R^N is the N-dimensional vector space and r is the one-dimensional vector space. The inputs considered are LL, PI, CF and ΔPI. The output of the SVM model is ϕ_r. In other words, $x = [\text{LL}, \text{PI}, \text{CF}, \Delta\text{PI}]$ and $y = \phi_r$.
The ε-insensitive loss function can be described in the following way.

$$L_\varepsilon(y) = 0 \text{ for } |f(x) - y| < \varepsilon \text{ otherwise } L_\varepsilon(y) = |f(x) - y| - \varepsilon \tag{4}$$

This defines an ε tube such 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 equal to the absolute value of the deviation minus ε. The main aim in SVM is to find a function f(x) that gives a deviation of ε from the actual output and at the same time is as flat as possible. Let us assume a linear function

$$f(x) = (w \cdot x) + b,\ w \in R^n,\ b \in r \tag{5}$$

where, w = an adjustable weight vector and b = the scalar threshold.
Flatness in the case of 5 means a small value of w. One way of obtaining this is by minimizing the Euclidean norm$||w||^2$. This is equivalent to the following convex optimization problem:
Minimize: $\frac{1}{2}\,||w||^2$
Subjected to: $y_i - (\langle w.x_i \rangle + b) \leq \varepsilon$, i = 1, 2,..., l

$$(\langle w.x_i \rangle + b) - y_i \leq \varepsilon, i = 1, 2, \ldots, l \tag{6}$$

The above convex optimization problem is feasible. Sometimes, however, this may not be the case, or we also may want to allow for some errors. This is analogous to the "soft margin" loss function [22] that was used in SVM by Cortes and Vapnik [19]. The parameters ξ_i, ξ_i^* are slack variables that determine the degree to which samples with error more than ε be penalized. In other words, any error smaller than ε does not require ξ_i or ξ_i^* and hence

does not enter the objective function because these data points have a value of zero for the loss function. The slack variables (ξ_i, ξ_i^*) have been introduced to avoid infeasible constraints in the optimization problem below:
Minimize: $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$
Subjected to: $y_i - (\langle w \cdot x_i\rangle + b) \leq \varepsilon + \xi_i$, i = 1, 2,...,l
$(\langle w \cdot x_i\rangle + b) - y_i \leq \varepsilon + \xi_i^*$, i = 1, 2,...,l

$$\xi_i \geq 0 \text{ and } \xi_i^* \geq 0, i = 1, 2, \ldots, l \tag{7}$$

The constant C ($0 < C < \infty$) determines the trade-off between the flatness of f and the amount up to which deviations larger than ε are tolerated [23]. This optimization problem 7 is solved by Lagrangian Multipliers [21], and its solution is given by:

$$f(x) = \sum_{i=1}^{nsv}(\alpha_i - \alpha_i^*)(x_i \cdot x) + b \tag{8}$$

where $b = -\left(\frac{1}{2}\right) w\cdot[x_r + x_s]$
α_i and α_i^* are the Lagrangian multipliers; and *nsv* is the number of support vectors. An important aspect is that some Lagrangian multipliers (α_i, α_i^*) will be zero, implying that these training objects are considered to be irrelevant for the final solution (*cf.* matrix sparseness). The training objects with non-zero Lagrangian multipliers are called support vectors.
When linear regression is not appropriate, then input data have to be mapped into a high dimensional feature space through some nonlinear mapping technique [17]. The two steps in this exercise are, firstly, carrying out a fixed nonlinear mapping of the data onto the feature space and, secondly, carrying out a linear regression in the high dimensional space. The input data are mapped onto the feature space by a map Φ. The dot product given by $\Phi(x_i)\cdot\Phi(x)$ is computed as a linear combination of the training points. The concept of a kernel function $[K(x_i, x) = \Phi(x_i)\cdot\Phi(x)]$ has been introduced to reduce the computational demand [18, 19]. So, equation 8 becomes:

$$f(x) = \sum_{i=1}^{nsv}(\alpha_i - \alpha_i^*)K(x_i \cdot x) + b \tag{9}$$

Some common kernels, such as homogeneous polynomial expressions, non-homogeneous polynomial expressions, radial basis functions, Gaussian functions and sigmoid functions, and their combinations, have been used for non-linear cases. The details about the SVM and its implementation in geotechnical engineering have been presented in Das *et al.* [24].

3. Database and Preprocessing

In the present study, databases obtained from landslide areas, slope failure areas, debris flow areas, and volcanic eruption areas, available in the literature [7, 9, 10] have been considered. The ϕ_r values in the above studies were determined using a laboratory ring shear test, and the average friction angle was considered. The 137 records used in this study include the index properties of soil (LL, PL, PI, CF and ΔPI) and output the residual friction angles. Out of 137 data points, 96 (70%) randomly selected samples were used for training and 41 (30%) data points were used for testing and are presented in Table 1. The soils studied herein have a broad range of geologic and geographic origins as reflected in the wide range and large standard deviation values.

4. Results and Discussion

4.1. Artificial Neural Network Results

Different ANN models were tried using different combinations of the above input variables and the developed models are compared in terms of a correlation coefficient (R). Two of the successful ANN models and their corresponding R values are presented in Table 2. Model 1 with all four input parameters LL, PI, CF and ΔPI as the inputs has the best correlation with ϕ_r values. Model 2, using CF and ΔPI as the two inputs as proposed in the literature [11] gave lower R values in training and testing, suggesting an inferior fit. The models were also compared in terms of a coefficient of determination (E) and the trend in E is similar to the R values. This poor performance of Model 2 as per the literature [11] may be due to the fact that the model was developed for the tropical soil only. This also focuses on the importance of developing ANN models to consider databases for soils of different origin. The weights and biases of the final network are presented in Table 3 for BRNN, LMNN and DENN. The interpretation of the weights and biases to find out the importance of the input parameters and their relationships with the outputs will be discussed later. The above parameters can also be used for presentations of the model equation [11]. The performance of BRNN, LMNN and DENN for the training and testing data set is also shown in Figures 1, 2 and 3, respectively. It can be seen that there is substantial scatter in the data, on both sides of the line of equality, irrespective of the type of ANN model.

4.1.1. Sensitivity Analysis

Sensitivity analysis is of utmost concern in selecting important input variables. Different sensitivity analyses with

Table 1. The data considered for the present study.

TRAINING DATA									
INPUTS				OUT PUT	INPUTS				OUT PUT
LL	PI	D PI	CF	ϕ_r	LL	PI	D PI	CF	ϕ_r
75.00	38.00	-2.15	30.20	15.00	100.00	68.30	9.90	42.80	10.00
82.00	39.10	-6.16	32.20	16.00	120.00	65.50	-7.50	42.00	10.90
78.00	38.00	-4.34	30.10	16.00	108.00	66.20	1.96	41.20	14.70
66.70	21.90	-12.19	17.50	22.00	58.00	29.20	1.46	21.30	19.20
71.00	35.00	-2.23	30.20	14.00	57.00	17.00	-10.01	10.00	12.70
72.30	37.10	-1.08	31.10	14.00	57.00	23.00	-4.01	2.80	10.40
76.00	36.00	-4.88	30.80	16.00	61.00	24.00	-5.93	18.00	18.10
68.10	34.60	-0.51	28.80	12.00	55.00	27.00	1.45	6.50	17.60
84.00	52.90	6.18	45.20	11.00	82.00	51.00	5.74	31.00	15.20
69.20	31.00	-4.92	26.30	12.00	84.00	31.00	-15.72	22.00	14.00
81.40	43.90	-0.92	37.50	11.00	54.00	19.00	-5.82	10.00	12.90
78.20	25.80	-16.69	20.50	12.00	56.00	28.00	1.72	5.40	25.70
75.50	25.60	-14.92	21.50	12.00	49.00	9.00	-12.17	5.00	26.70
77.20	49.50	7.74	38.20	10.00	41.00	6.00	-9.33	2.90	28.90
66.00	31.00	-2.58	27.70	12.50	83.00	54.00	8.01	51.00	11.00
55.00	18.00	-7.55	19.50	18.80	41.20	29.90	14.42	10.00	28.70
86.30	47.10	-1.30	28.20	10.00	46.80	30.30	10.74	10.00	23.10
89.00	41.90	-8.47	37.50	10.00	67.70	27.50	-7.32	17.00	14.40
73.00	35.30	-3.39	22.10	10.10	72.30	29.90	-8.28	17.00	17.20
64.00	34.10	1.98	19.20	12.70	55.60	38.60	12.61	12.00	8.30
68.00	38.20	3.16	28.20	9.80	56.80	34.30	7.44	16.00	7.80
69.00	40.00	4.23	27.20	9.80	36.00	10.00	-1.68	10.00	29.00
51.00	19.00	-3.63	21.80	10.70	35.00	8.00	-2.95	6.00	29.00
71.50	36.40	-1.20	21.20	19.00	26.00	8.00	3.62	5.00	31.00
55.80	32.60	6.47	20.50	18.00	77.20	20.70	-21.06	16.20	25.60
70.00	38.00	1.50	24.00	8.90	57.00	17.30	-9.71	15.20	25.80
65.00	34.00	1.15	22.00	10.70	57.30	12.70	-14.53	11.20	23.40
61.00	15.00	-14.93	20.00	16.20	36.00	5.00	-6.68	9.00	23.30
53.00	26.00	1.91	14.00	19.30	45.00	19.00	0.75	13.00	25.50
91.30	50.50	-1.55	32.80	10.00	47.00	9.00	-10.71	6.50	28.00
94.60	62.40	7.94	33.20	10.00	36.40	9.60	-2.37	0.40	28.40
69.00	38.00	2.23	27.50	9.80	31.10	11.10	3.00	1.00	24.90
62.00	29.00	-1.66	19.90	12.80	54.00	23.50	-1.32	2.40	24.40
59.00	19.00	-9.47	20.00	17.60	48.60	16.60	-4.28	4.30	30.10
96.20	48.00	-7.63	32.20	12.00	42.70	16.10	-0.47	4.20	29.00
94.70	59.20	4.67	33.50	12.00	32.00	4.50	-4.26	1.80	23.60
63.00	32.00	0.61	24.80	17.10	35.00	6.00	-4.95	2.30	21.10
71.00	38.00	0.77	26.20	11.20	36.00	5.50	-6.18	3.90	21.40
LL	PI	D PI	CF	ϕ_r	LL	PI	D PI	CF	ϕ_r
93.00	9.00	-44.29	22.00	35.00	94.00	60.00	5.98	50.00	12.60
39.00	19.00	5.13	40.00	25.60	161.00	132.00	29.07	64.00	6.90
64.00	35.00	2.88	30.00	13.00	41.00	20.00	4.67	38.00	28.70
59.00	31.00	2.53	43.00	8.10	129.00	89.00	9.43	91.00	8.10
113.00	46.00	-21.89	22.00	34.00	165.00	46.00	-59.85	65.00	39.00
213.00	46.00	-94.89	77.00	39.00	113.00	91.00	23.11	52.00	5.50
59.00	29.00	0.53	50.00	9.20	22.00	10.00	8.54	14.00	31.20

TRAINING DATA									
INPUTS				OUT PUT	INPUTS				OUT PUT
LL	PI	D PI	CF	ϕ_r	LL	PI	D PI	CF	ϕ_r
89.00	60.00	9.63	45.00	14.00	77.00	37.00	-4.61	35.00	19.30
63.00	20.00	-11.39	48.00	8.00	72.00	43.00	5.04	57.00	9.90
151.00	68.00	-27.63	28.00	18.10	24.00	12.00	9.08	20.00	30.20
82.00	54.00	8.74	57.00	8.40	73.00	49.00	10.31	48.00	7.10
62.00	38.00	7.34	64.00	9.50	82.00	49.00	3.74	70.00	11.10
63.00	37.00	5.61	51.00	7.30	79.00	58.00	14.93	57.00	10.80
62.00	36.00	5.34	46.00	7.80	71.00	43.00	5.77	55.00	7.10
38.00	10.00	-3.14	13.00	25.00	89.00	50.00	-0.37	57.00	13.10
29.00	13.00	6.43	26.00	25.30	34.00	18.00	7.78	28.00	23.80
71.00	31.00	-6.23	34.00	13.50	80.00	40.00	-3.80	52.00	20.00
95.00	61.00	6.25	59.00	9.40	65.00	33.00	0.15	52.00	8.70
57.00	33.00	5.99	50.00	9.40	88.00	45.00	-4.64	52.00	15.90
59.00	36.00	7.53	51.00	7.10	24.00	12.00	9.08	20.00	30.20
26.00	6.00	1.62	32.00	10.10	42.00	23.00	6.94	14.00	24.40
31.00	12.00	3.97	32.00	12.10	118.00	83.00	11.46	76.00	7.40
66.00	42.00	8.42	53.00	8.00	41.00	25.00	9.67	28.00	22.10
63.00	35.00	3.61	63.00	9.20	93.00	61.00	7.71	60.00	7.00
89.00	47.00	-3.37	72.00	10.00	52.00	19.00	-4.36	20.00	27.50
97.00	50.00	-6.21	59.00	9.90	62.00	36.00	5.34	46.00	8.20
54.00	21.00	-3.82	39.00	26.00	58.00	32.00	4.26	52.00	10.70
85.00	58.00	10.55	50.00	6.60					

Table 2. Different ANN models and their statistical performance.

Models	ANN models	Correlation coefficient (R)		RMSE		MAE		AAE	
		Training	Testing	Training	Testing	Training	Testing	Training	Testing
Model 1 (Inputs: LL, PI, CF, ΔPI)	BRNN	0.888	0.738	5.130	7.000	10.567	14.465	2.886	3.612
	LMNN	0.877	0.722	3.81	5.189	10.780	13.231	3.063	3.759
	DENN	0.865	0.728	4.001	5.005	11.350	13.502	3.182	3.717
Model 2 (Inputs: CF, ΔPI)	BRNN	0.770	0.558	5.056	6.238	17.577	19.358	3.740	4.270
	LMNN	0.694	0.562	5.700	6.437	16.538	16.342	4.385	4.472
	DENN	0.787	0.622	4.884	6.552	16.232	21.700	3.681	4.563

a correlation matrix between inputs and output, including Garson's algorithm, and the Connection Weight Approach [11, 16] are presented in Tables 4, 5, 6.

From Table 4 it can be seen that ϕ_r is highly correlated to PI, followed by ΔPI and CF, as signified by the cross correlation values of -0.563, -0.481 and -0.458, respectively. The sensitivity analysis for the BRNN model with Garson's method and the Connection Weight Approach is also given in Table 4. The CF is found to be the most important input parameter for both Garson's approach (relative importance 30.26%) and the Connection Weight Approach (S_j = -2.76). The negative S_j values imply that PI, ΔPI and CF are inversely related to ϕ_r. However, the relative importances of the other three inputs are different for the two different approaches. The sensitivity analysis for the DENN model with Garson's method and the Connection Weight Approach is given in Table 5. With the Connection Weight Approach, PI should be the most important input parameter for both (S_j = -5.04), followed by LL (S_j = 2.04) and CF (S_j = -1.34). It was also observed that CF, ΔPI and PI are inversely related to ϕ_r. However, with Garson's algorithm, LL is found to be the most important input, followed by PI, CF, and ΔPI. Hence, it can be concluded that inferences drawn from the DENN model

Table 3. Connection weights and biases for the BRNN model.

ANN method	Neuron	Weights (w_{ik})					Biases	
		LL	PI	CF	ΔPI	ϕ_r	b_{hk}	b_0
BRNN	Hidden neuron 1 (k=1)	-2.019	-1.298	-0.287	-0.415	1.698	-2.527	1.417
	Hidden neuron 2 (k=2)	-0.244	-1.318	-1.627	-0.730	-1.597	-1.129	
	Hidden neuron 3 (k=3)	1.365	-0.381	-2.031	-1.046	2.398	-0.216	
LMNN	Hidden neuron 1 (k=1)	-1.009	1.227	-0.732	0.546	0.826	1.924	1.151
	Hidden neuron 2 (k=2)	-1.200	-0.409	1.887	1.330	-1.981	0.118	
	Hidden neuron 3 (k=3)	-2.175	-0.516	0.836	0.014	1.227	-2.529	
DENN	Hidden neuron 1 (k=1)	-2.086	-0.227	1.233	-0.130	1.171	-2.613	0.000
	Hidden neuron 2 (k=2)	-2.633	1.219	1.710	0.876	-1.199	0.000	
	Hidden neuron 3 (k=3)	1.154	-2.885	-0.638	-0.037	1.149	3.977	

Table 4. Relative Importance of different inputs as per Garson's algorithm and connection weight approach as per BRNN Model weights.

Parameters	Pearson's correlation		Garson's algorithm		Connection weight approach	
	Correlation values	Ranking	Relative importance (%)	Ranking of inputs	S_i value	Ranking of inputs
LL	-0.254	4	28.26	2	0.24	4
PI	-0.563	1	24.60	3	-1.01	3
CF	-0.458	3	30.26	1	-2.76	1
ΔPI	-0.481	2	16.88	4	-2.05	2

Table 5. Relative Importance of different inputs as per Garson's algorithm and connection weight approach as per DENN Model weights.

Parameters	Garson's algorithm		Connection weight approach	
	Relative importance (%)	Ranking of inputs as per relative importance	S_i value	Ranking of inputs as per relative importance
LL	40.71	1	2.04	2
PI	28.77	2	-5.04	1
CF	24.55	3	-1.34	3
ΔPI	5.98	4	-1.25	4

Table 6. Relative Importance of different inputs as per Garson's algorithm and connection weight approach as per LMNN Model weights.

Parameters	Garson's algorithm		Connection weight approach	
	Relative importance (%)	Ranking of inputs as per relative importance	S_i value as per connection weight approach	Ranking of inputs as per relative importance
LL	38.33	1	-1.12	4
PI	19.32	3	1.19	3
CF	27.85	2	-3.32	1
ΔPI	14.50	4	-2.17	2

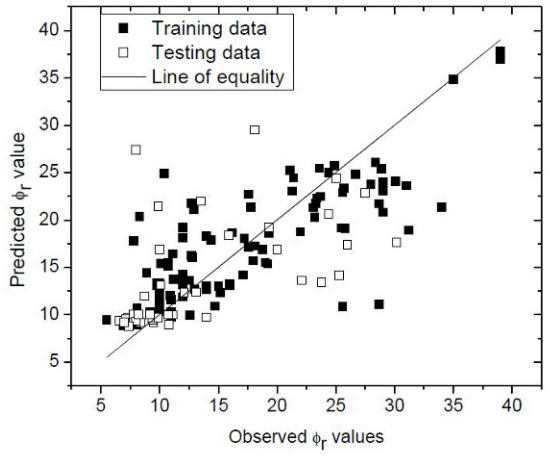


Figure 1. The predicted and observed values for the Model 1 using BRNN.

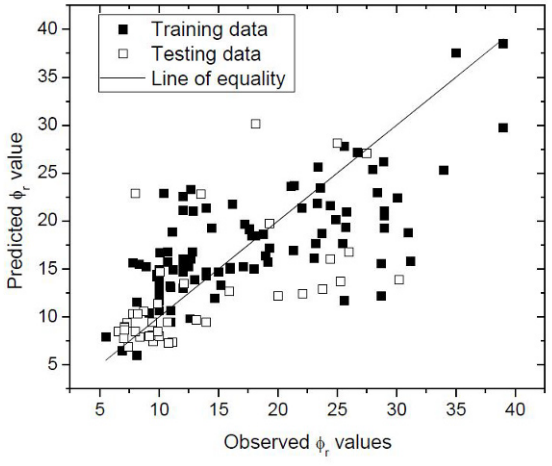


Figure 2. The predicted and observed values for the Model 1 using LMNN.

from both of the approaches are not in compliance with the physical phenomena. The sensitivity analysis for the LMNN model with Garson's method and the Connection Weight Approach is given in Table 6. With the Connection Weight Approach, similar to results from the BRNN model, CF is found to be the most important input parameter for both (S_j = -3.32), followed by ΔPI and PI. It was also observed that CF, DPI and LL are inversely related to ϕ_r. However, with Garson's algorithm, LL is found to be the most important input, followed by CF, PI and ΔPI. Hence, it can be concluded that inferences drawn from the Connection Weight Approach is more similar to physical phenomena. Similar observations have been made by

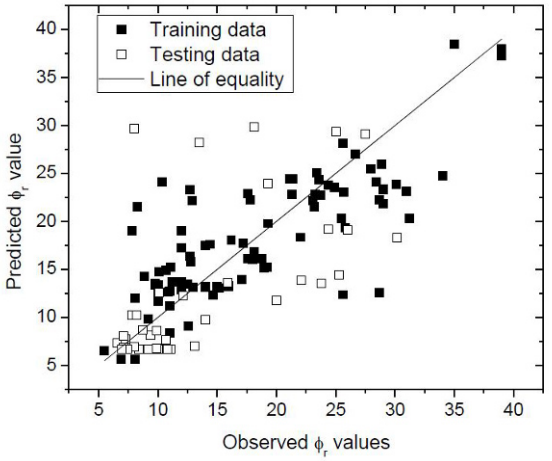


Figure 3. The predicted and observed values for the Model 1 using DENN.

Das and Basudhar [11, 16] and Olden *et al.* [25]. This may be due to the fact that in case of the Connection Weight Approach, actual weights are considered, whereas in the case of Garson's algorithm, absolute values of the weights are considered.

4.2. Results of SVM Method

The performance of the SVM model depends upon the type of kernel function, the optimum capacity factor C, and the optimum error insensitive zone ε [24]. With different kernel functions, different combinations of C and ε values were tried in order to arrive at the best performance for training data, and the final C and ε values are presented in Table 7. Two types of models were developed based on the combination of input parameters; Model 1, with input parameters as LL, PI, ΔPI and CF and Model 2, with ΔPI and CF. Similarly, with kernel functions in the form of a radial (Gaussian) basis function, a polynomial function and a spline kernel function, the SVM models discussed herein are denoted as SVM-G (Gaussian), SVM-P (Polynomial) and SVM-S (Spline), respectively. A detailed parametric study was conducted and the final SVM parameters corresponding to different kernel functions and corresponding R values are shown in Table 7. In order to evaluate the capabilities of the SVM models, each model was validated with testing data that were not part of the training data set. As discussed earlier, although the R values for training data are better (R = 0.965) for SVM-P, with testing data Model 1 is found to be more efficient when SVM-G is used. The training and testing data for Model 1 with different kernel functions are presented in Figure 4. Simi-

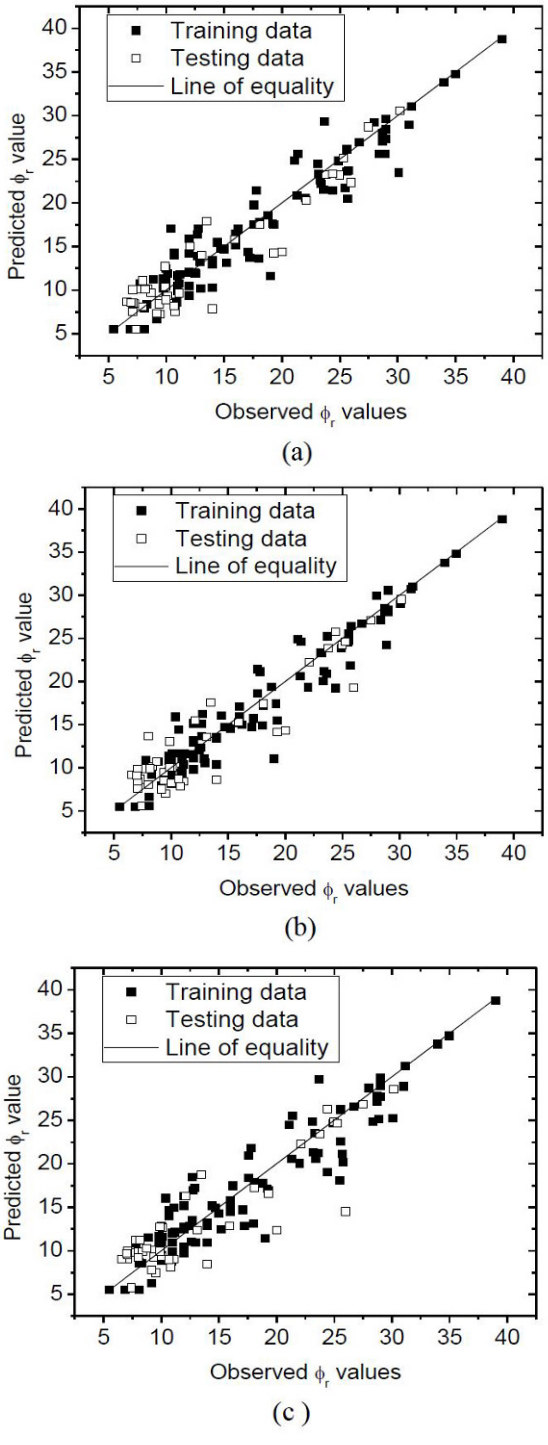


Figure 4. The predicted and observed values for Model 1 using **(a)** SVM-G, **(b)** SVM-P and **(c)** SVM-S.

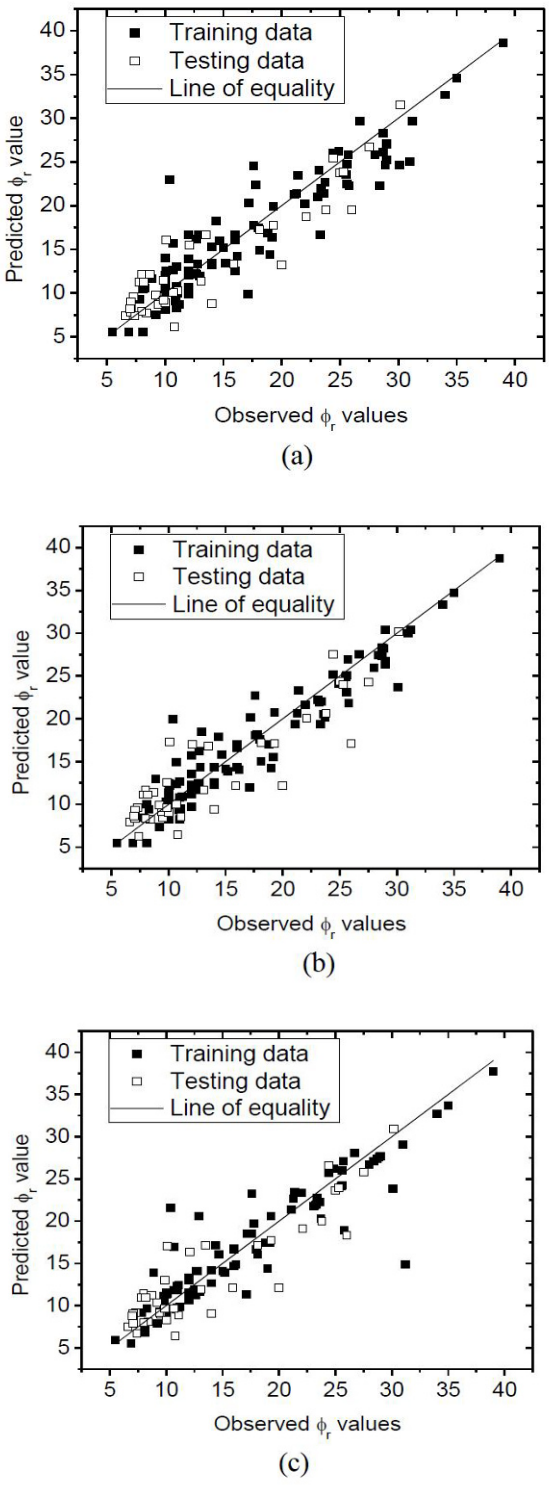


Figure 5. The predicted and observed values for Model 2 using using **(a)** SVM-G, **(b)** SVM-P and **(c)** SVM-S.

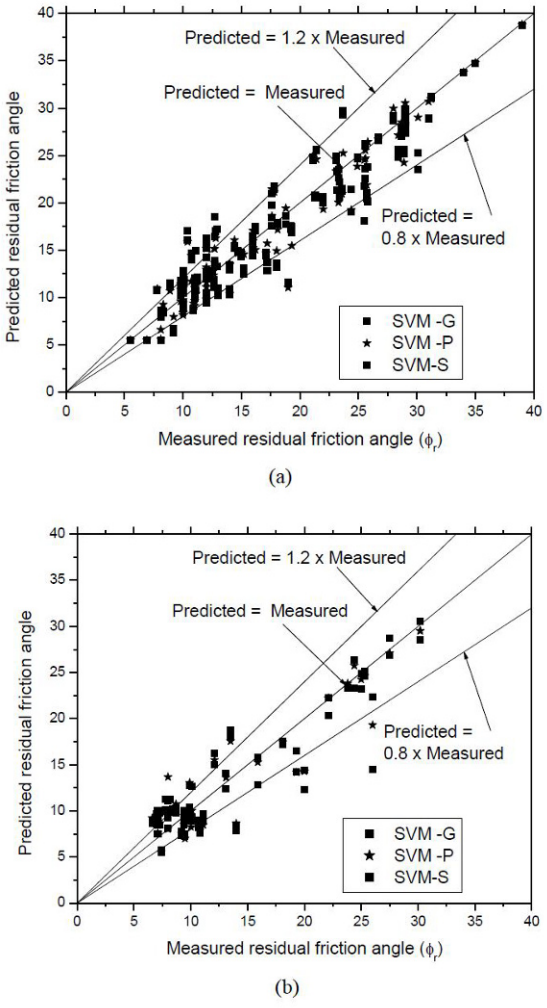


Figure 6. The variation of the data points for **(a)** training and **(b)** testing data using different SVM models for Model 1.

larly Figure 5 shows the data points for Model 2. It can be seen that, compared to ANN models, there is considerably less scatter in the data both for Model 1 and Model 2.

The scatter in data points for training and testing data using the three different SVM models for Model 1 was examined separately in Figure 6. It can be seen that for Model 1 the predicted data points are within 80% of the observed values, both for training and testing data. The efficacy of the developed model was judged from its performances on testing data, the variation of observed and predicted values for all three models when the testing data was used is presented in Figure 7. In comparison to other SVM models, more data points as per SVM-G models are closer to the observed values, demonstrating its slight superiority. Similarly, Figure 8 shows the varia-

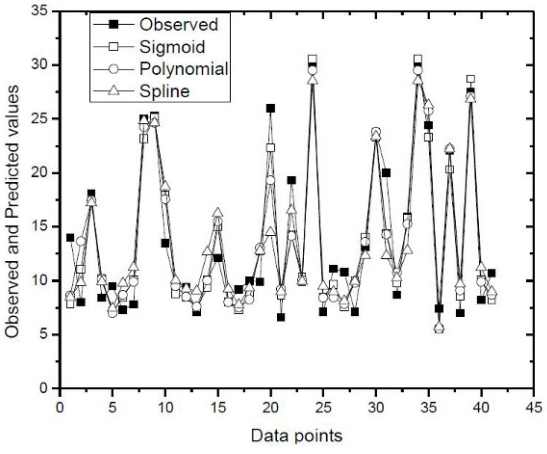


Figure 7. The variation of the observed and predicted values using different SVM models for Model 1.

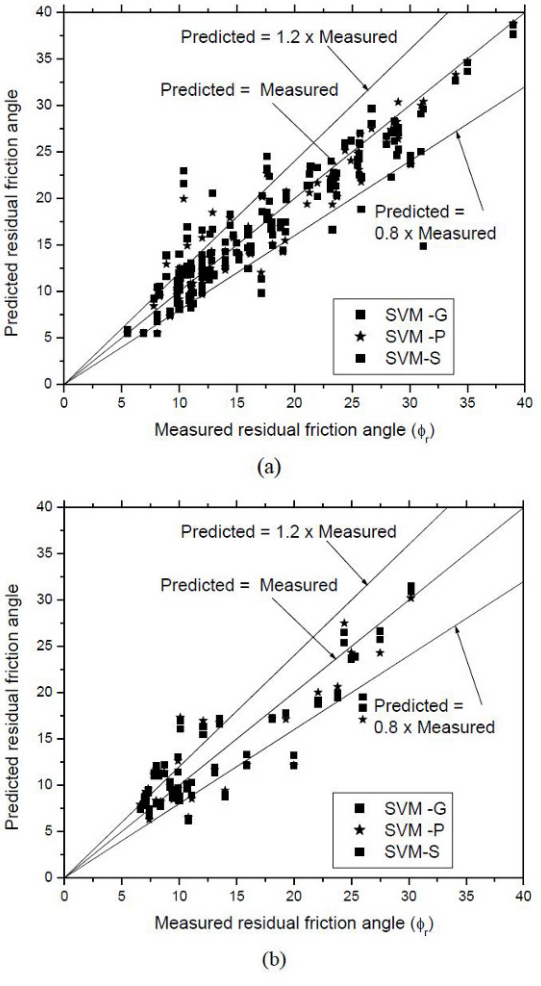


Figure 8. The variation of the data points for **(a)** training and **(b)** testing data using different SVM models for Model 2.

Table 7. Performance of different SVM models with SVM parameters.

Model	Type of kernel function	Correlation coefficient (R)		Coefficient of determination (E)		C	ε	Number of support vectors
		Training performance	Testing Performance	Training performance	Testing Performance			
Model I (LL, PI, CF and DPI as inputs)	Gauss, width=0.4	0.954	0.945	0.910	0.894	100	0.02	80
	Polynomial (3)	0.965	0.933	0.929	0.871	60	0.03	79
	Spline	0.939	0.903	0.881	0.815	50	0.01	78
Model II (CF and DPI as inputs)	Gauss, width=0.1	0.930	0.925	0.864	0.851	10	0.001	90
	Polynomial (2)	0.955	0.902	0.910	0.812	40	0.008	92
	Spline	0.927	0.911	0.858	0.828	60	0.004	93

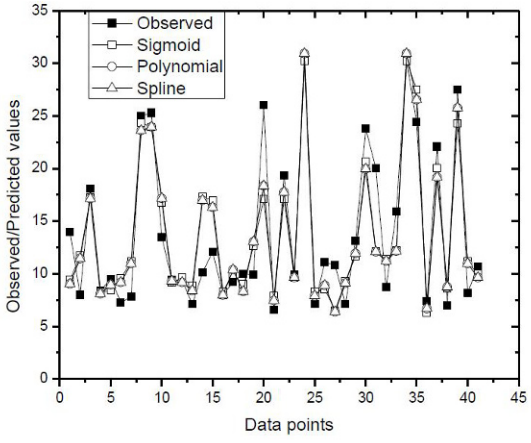


Figure 9. The variation of the observed and predicted values using different SVM models for Model 2.

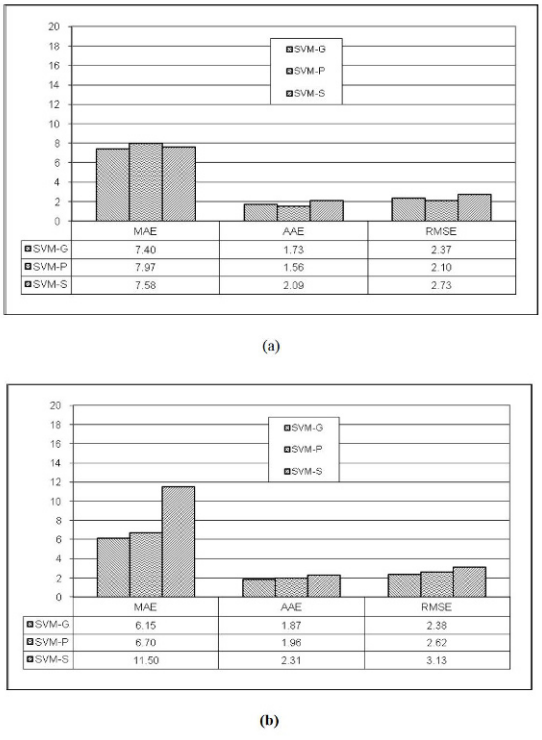
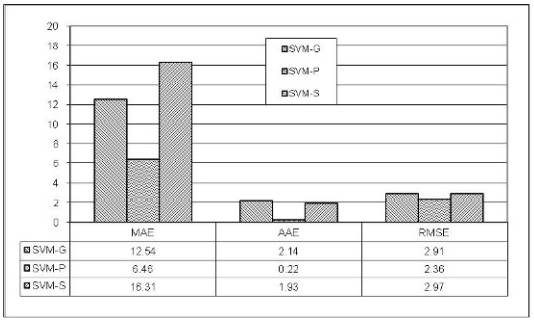


Figure 10. Comparison of prediction capabilities of SVM for Model 1 using **(a)** training data and **(b)** testing data.

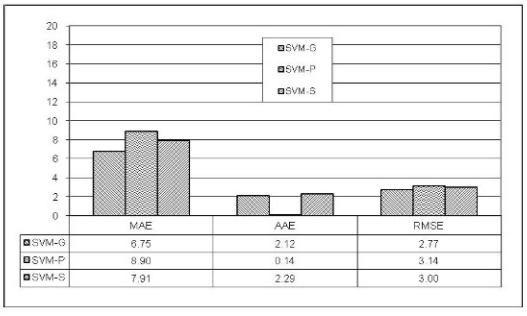
tion of predicted and observed data points using Model 2. The variation of observed and predicted ϕ_r values for the testing data using Model 2 is shown in Figure 9. It can be seen that, here the data points are more scattered particularly in the ϕ_r values of 10◦ - 16◦ in comparison to Model 1. This may be due to the inadequacy of the model for the whole range of data. Hence Model 1 is preferred over Model 2.

Similarly, other statistical parameters like Maximum Absolute Error (MAE), Average Absolute Error (AAE) and Root Mean Square Error (RMSE) are used to compare Model 1 and Model 2 using different SVM models. The error values for Model 1 are shown in Figure 10 and those for Model 2 in Figure 11. As the statistical performances of the models are different for the training and testing datasets, the 'best' model is selected based on the performances of the model for the new (testing) dataset. Based on the MAE, AAE and RMSE values of the testing data (Figure 10b), Model 1 using SVM-G was found to be the 'best' model based on AAE. Similarly from the Figure 11, for Model 2, SVM-G was found to be the 'best' model. Thus, another important aspect of this paper is the presentation of the results of SVM models in terms of an equation which can be used by the professional for simple spread sheet calculations.

Model 1 using SVM-G gives better performance than Model 2. The following equation can be developed for



(a).



(b)

Figure 11. Comparison of prediction capabilities of SVM for Model 2 using **(a)** training data and **(b)** testing data.

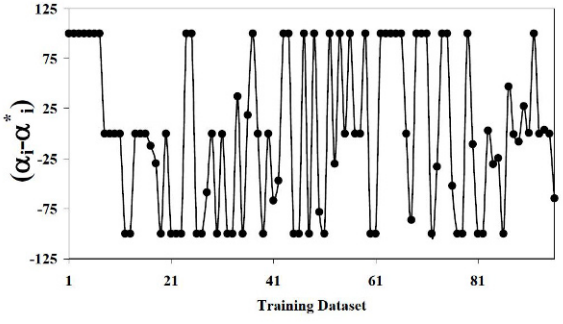


Figure 12. Values of $(\alpha_i - \alpha_i^*)$ obtained as per SVM analysis for Model 1.

the prediction of ϕ_r from Equation 8 by substituting b=0,$K(x_i, x) = \exp\left(-\frac{|x-x_i|^2}{2\sigma^2}\right)$ and s = 0.4.

$$\phi_r = \sum_{i=1}^{80} (\alpha_i - \alpha_i^*) \exp\left(-\frac{|x - x_i|^2}{0.32}\right) \qquad (10)$$

The values of $(\alpha_i - \alpha_i^*)$ are given in Figure 12.
It may be mentioned here that the input data for all other values $(\alpha_i - \alpha_i^*)$ are zero, except those for the support vectors. This model equation highlights the importance of the support vectors, and of their number.

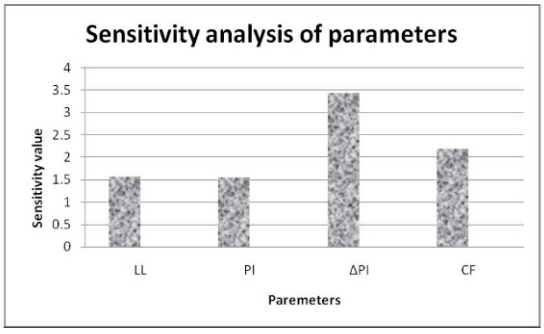


Figure 13. Sensitivity analysis of input parameters as per SVM-G for the Model 1.

4.2.1. Sensitivity Analysis /Selection of Important Input Variables

The sensitivity analysis of each input parameter for the SVM model is carried out as per the following formula:

$$S(\%) = \frac{1}{N} \sum_{j=1}^{N} \left(\frac{\%\text{ change in output}}{\%\text{ change in input}} \right)_j \times 100 \qquad (11)$$

where, N is the number of data points [25]. The trained model is fed with a varying input, keeping the other inputs constant. Analysis was carried out by varying each input parameter, one at a time, at a constant rate of 20%, and the result is shown in Figure 13. It can be seen that ΔPI is the important parameters followed by CF, LL and PI and this matches well with the previous findings, such as those of Wesley [7].

5. Conclusions

In geotechnical problems involving large deformations such as landslides, soils get remoulded and it is more appropriate to analyse them on the basis of their residual shear strength than on their peak shear strength. Even in many of the first-time slope failures, soil along parts of the failure surface is at a residual strength. This study discusses the correlation of residual friction angle of clay (ϕ_r) with the index properties of soil, and considers soils of different origins and a wide range of parameter values using ANN and SVM analyses.
Based on different ANN models, Model 1, with LL, PI, CF and ΔPI as the inputs, has the best correlation with ϕ_r

values. Based on correlation coefficient (R) values, the performances of developed ANN models (BRNN, LMNN and DENN) were found to be almost equally efficient. Based on different sensitivity analyses for the ANN models, it was observed that the important inputs were found to vary depending upon the ANN model and the sensitivity analysis used. But the Connection Weight Approach is more effective in drawing conclusions regarding correlations of inputs with outputs, corresponding to physical phenomena.

While using SVM models, based on the training and testing performances of R, MAE, AAE and RMSE values, Model 1 using SVM-G was found to be more efficient. It was also observed that for Model 1 using SVM, the predicted data points are within 80% of the observed values, both for training and testing data. Hence, it can be concluded that SVM model is more efficient in predicting the residual friction angle of clay in comparison to the ANN models. Using sensitivity analysis it was observed that ΔPI is the most important parameter, followed by CF, LL and PI. A model equation was presented for the developed SVM model, which can be used by the professionals for simple spread sheet calculations.

References

[1] Skempton A.W., The long term stability of clay slopes. *Geotechnique,* 1964, 14, 77-101

[2] Mesri G., Shahien M., Residual Shear Strength Mobilized in First-Time Slope Failures. *J. Geotech. Geoenviron. Eng.*, 2003, 129, 12–31

[3] Bowles J.E., Foundation analysis and design. McGraw-Hill International Edition, Singapore, 1988

[4] Mesri G., Cepeda-Diaz A.F., Residual strength of clays and shales. *Geotechnique,* 1986, 36, 269-274

[5] Colotta T., Cantoni R., Pavesi U., Robert E., Moretti P.C., A correlation between residual friction angle, gradation and index properties of cohesive soil. *Geotechnique*, 1989, 39, 343-346

[6] Stark T.D., Eid H.T., Drained residual strength of cohesive soils. *J. Geotech. Geoenviron. Eng.*, 1994, 120, 856-871

[7] Wesley L.D., Residual strength of clays and correlations using Atterberg limit. *Geotechnique*, 2003, 53, 669-672

[8] Sridharan A., Rao P.R., Discussion: Residual strength of clays and correlation using Atterberg limits. *Geotechnique,* 2004, 54, 503-504

[9] Tiwari B., Marui H., A new method for the correlation of residual shear strength of the soil with mineralogical composition. *J. Geotech. Geoenviron. Eng.*, 2005, 131, 1139-1150

[10] Kaya A., Kwong J.K.P., Evaluation of common practice empirical procedures for residual friction angle of soils: Hawaiian amorphous material rich colluvial soil case study. *Eng. Geol.*, 2007, 92, 49–58

[11] Das S.K., Basudhar P.K., Prediction of residual friction angle of clays using artificial neural network. *Eng. Geol.*, 2008, 100, 142-145

[12] Demuth H., Beale M., Neural Network Toolbox. The Math Works Inc., USA, 2000.

[13] MathWork Inc., Matlab User's Manual. Version 6.5. Natick MA, 2001

[14] Ilonen J., Kamarainen J.K., Lampinen J., Differential Evolution training algorithm for feed-forward neural network. *Neural Processing Letters*, 2003, 17, 93-105

[15] Juang C.H., Elton D.J., Prediction of collapse potential of soil with neural networks. *Trans. Res. Record*, 1997, 1582, 22–28

[16] Das S.K., Basudhar P.K., Undrained lateral load capacity of piles in clay using artificial neural network. *Comput. and Geotech.*, 2006, 33, 454-459

[17] Boser B.E., Guyon I.M., Vapnik V.N., A training algorithm for optimal margin classifiers. In: Haussler D (Ed.) 5th Annual ACM workshop on COLT. ACM, Pittsburgh, 1992 , 144–152

[18] Cristianini N., Shawe-Taylor J., An introduction to support vector machine. University Press, London, Cambridge, 2000

[19] Cortes C., Vapnik V.N., Support vector networks. *Mach. Learn.*, 1995, 20, 273–297

[20] Gualtieri J.A., Chettri S.R., Cromp R.F., Johnson L.F., Support vector machine classifiers as applied to AVIRIS data. In: The summaries of the 8th JPL airborne earth science workshop, 1999

[21] Vapnik V.N., Statistical learning theory. Wiley, New York, 1998

[22] Bennett K.P., Mangasarian O.L., Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, 1992, 1, 23-34

[23] Smola A.J., Scholkopf B., A tutorial on support vector regression. *Stat. Comput.*, 2004, 14, 199–222

[24] Das S.K., Samui P., Sabat A.K., Sitharam T.G., Prediction of swelling pressure of soil using artificial intelligence techniques. *Environmental Earth Science*, 2010, 61, 393-403

[25] Olden J.D., Joy M.K., Death R.G., An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Eco. Model*, 2004, 178, 389-397