Civil Engineering Infrastructures Journal, 50(1): 35 – 49, June 2017
Print ISSN: 2322-2093; Online ISSN: 2423-6691
DOI: 10.7508/ceij.2017.01.003

Bearing Capacity of Shallow Foundations on Cohesionless Soils: A Random Forest Based Approach

Kohestani, V.R.[1*], Vosoughi, M.[1], Hassanlourad, M.[2] and Fallahnia, M.[3]

[1] M.Sc., Faculty of Engineering, Imam Khomeini International University, Qazvin, Iran.
[2] Assistant Professor, Faculty of Engineering, Imam Khomeini International University, Qazvin, Iran.
[3] Assistant Professor, Department of Architecture and Energy, Faculty of Architecture and Urbanism, University of Art, Tehran, Iran.

Received: 03 Jun. 2015; Revised: 05 Feb. 2017; Accepted: 01 Mar. 2017

ABSTRACT: Determining the ultimate bearing capacity (UBC) is vital for design of shallow foundations. Recently, soft computing methods (i.e. artificial neural networks and support vector machines) have been used for this purpose. In this paper, Random Forest (RF) is utilized as a tree-based ensemble classifier for predicting the UBC of shallow foundations on cohesionless soils. The inputs of model are width of footing (B), depth of footing (D), footing geometry (L/B), unit weight of sand (γ) and internal friction angle (ϕ). A set of 112 load tests data were used to calibrate and test the developed RF-based model. The used data set consists of 47 full-scale observations and 65 small-scale laboratory footing load tests. To demonstrate the efficiency of proposed RF-based model, the results are compared with some popular classical formulas that are most commonly used for determining the UBC. The results show the efficiency and capabilities of the proposed RF-based model as a practical tool in evaluating the UBC of shallow foundations in a fast and accurate way.

Keywords: Artificial Intelligence, Decision Tree, Random Forest (RF), Shallow Foundations, Ultimate Bearing Capacity.

INTRODUCTION

Since the Ultimate Bearing Capacity (UBC) of shallow foundations plays a crucial role in design of many structures of small to medium size, it has always been interesting for geotechnical engineers. A Shallow foundation is a load carrying structure with depth-to-width ratio less than or equal to four. Terzaghi (1943) was the first researcher to propose a theory for estimating the UBC of shallow foundations. After Terzaghi, many researchers such as Meyerhof (1963), Hansen (1970) and Vesic (1974) have offered theories to estimate the UBC.

Models are mostly validated using the model-scale footing test. Several studies have focused on how to reduce the scale effects in extrapolation the experimental results to full-scale footings (e.g. De Beer, 1965; Yamaguchi et al., 1977). Tatsuoka, et al. (1991) focused on how the UBC is affected by particle size in model-scale footing tests. Results of large-scale footing tests on dense

* Corresponding author E-mail: vrkohestani@gmail.com

sand indicate that the shearing strains are considerably varies along the slip line. Therefore, the bearing capacity formulas where using the maximum value of friction angle, ϕ_{max} , are generally overestimate the bearing capacities of prototype (Padmini et al., 2008). Therefore, the actual model-scale footing test results produced different values than theoretical equations and special consideration should be used when applying model-scale footing test results to the behavior of full-scale foundations. For this reason, an alternative method is required to ensure better estimates of actual bearing capacity.

Intelligent systems are usually used to model complex relationships between inputs and outputs or find patterns in available data. Artificial intelligence (AI) based methods are able to capture the inherent nonlinearity and complex interaction between the involved variables of the problem of UBC. These methods can be trained to learn the relationship between the mechanical properties of soil and foundation geometry with the foundation bearing capacity, requiring no prior knowledge of the form of the relationship. Recently, to tackle the problem of UBC of shallow foundations, different types of AI-based methods have been employed by different researchers. Artificial Neural Network (ANN); Fuzzy Inference System (FIS), Adaptive Neuro Fuzzy Inference System (ANFIS), Ant Colony Optimization (ACO), Genetic Programming (GP), Weighted Genetic Programming (WGP), Soft-Computing Polynomials (SCP), Support Vector Machine (SVM) and Relevance Vector Machine (RVM) have been successfully adopted for the estimating the UBC of shallow foundations on soil (Kalinli et al., 2011; Padmini et al., 2008; Samui, 2012; Shahnazari and Tutunchian 2012; Tsai et al., 2013). These studies all agreed that soft computing approaches are more accurate compared to analytical formulas. However, these methods are not very transparent and also the modeling process is complicated. Most AI-based models are black box models where the relationship between input and output parameters are not accessible to the users and the models usually require an extensive trial-and-error procedure for setting the model parameters (Hassanlourad et al., 2014).

In this paper a novel AI-based method is proposed for prediction of UBC of shallow foundations on cohesionless soils using Random Forests (RF). RF is an straightforward ensemble learning technique developed by Breiman (2001) where utilizes a combination of a large set of decision trees (DTs) for approximation of complex nonlinear systems. In the last decade, there has been a growing trend in the use of decision tree algorithms such as C4.5 decision tree and Classification and Regression Tree (CART) for the modeling and approximation of complex nonlinear systems (Ardakani and Kohestani, 2015; Barzegari et al., 2015)

RFs have been frequently used in both regression and classifier form (Kohestani et al., 2015; Kohestani et al., 2016). However, in this study for the first time the RF is utilized for predicting the UBC of shallow foundations on cohesionless soils. To demonstrate the efficiency of proposed RF-based model, the results are compared with some popular classical formulas that are most commonly used for determining the UBC.

Decision Trees and Random Forest

Decision trees (DTs) are widely used effective technique for solving both regression problems where the output is a continuous value and classification problems here the output is a nominal value. To provide a briefer and easier-to-understand explanation of how DTs work, we provide a

description of the classification version of the tree.

Classification tree is a binary tree where each node except for leaves consists of decision rules that determine which branch to go through next. The leaves of the tree contain a specific class. When estimating the class of an input vector, the tree is traversed from its root to the leaves.

The decision rules in individual nodes can be seen as functions. These functions are often referred to as *split functions* or *weak learners* (Bańczyk et al., 2011). They are defined as

$$f(x,\theta): \chi \times \tau \rightarrow 0,1 \tag{1}$$

where x is a given vector from the input set χ and θ are the parameters of a test function from the set τ. This function performs mapping to the false and true values in order to determine which branch of the tree should be selected for continuation.

There is a large number of split functions that can be used (Criminisi et al., 2012). Mores split functions are frequently used, such as the general *oriented hyperplane*, *conic learner* and other both linear and non-linear functions (Criminisi et al., 2011).

Finding the best parameters for split function is an optimization problem in which we seek to find the parameters that minimize the classification error. This optimization is an essential task for building a DT. Error is proportional to the probability of selecting a particular class during random selection from the set of vectors that were assigned to the same group based on the split function.

In order to describe the quality of split using a specific split function the *information gain* where uses the concept of entropy from information theory is utilized in this study. Information gain is given by

$$G(S) = -\sum_{i=0}^{NC} p(c_i)\, log\big(p(c_i)\big) \tag{2}$$

where $p(c_i)$ is the probability of selecting a vector of class i from the vector set S and NC is the number of classes. The optimization goal is to maximize the improvement obtained by the split based on the following formula:

$$I = G(S) - \sum_{i \subset \{L,R\}} \frac{|S_i|}{S} G(S) \tag{3}$$

In a similar way, we can define the entropy for continuous output values. To construct a decision tree, it is necessary to define the conditions that determine the tree's growth termination point. This condition can be, for example, the fact that after the last split there are only vectors belonging to a single class in the current branch. This condition is not flawless, because it can lead to a phenomenon known as overfitting, a situation in which the tree loses its ability to generalize. A better condition may be, for example, a limit for the information gain that was obtained with the last split. The final class of a leaf is then determined by the maximum likelihood of selecting the class from the set of vectors belonging to that branch. At the same time we are able to read from the leaf the probability that the class was determined correctly (Breiman, 1996). An example of a DT, namely the CART in its classification mode, is pictured in Figure 1.

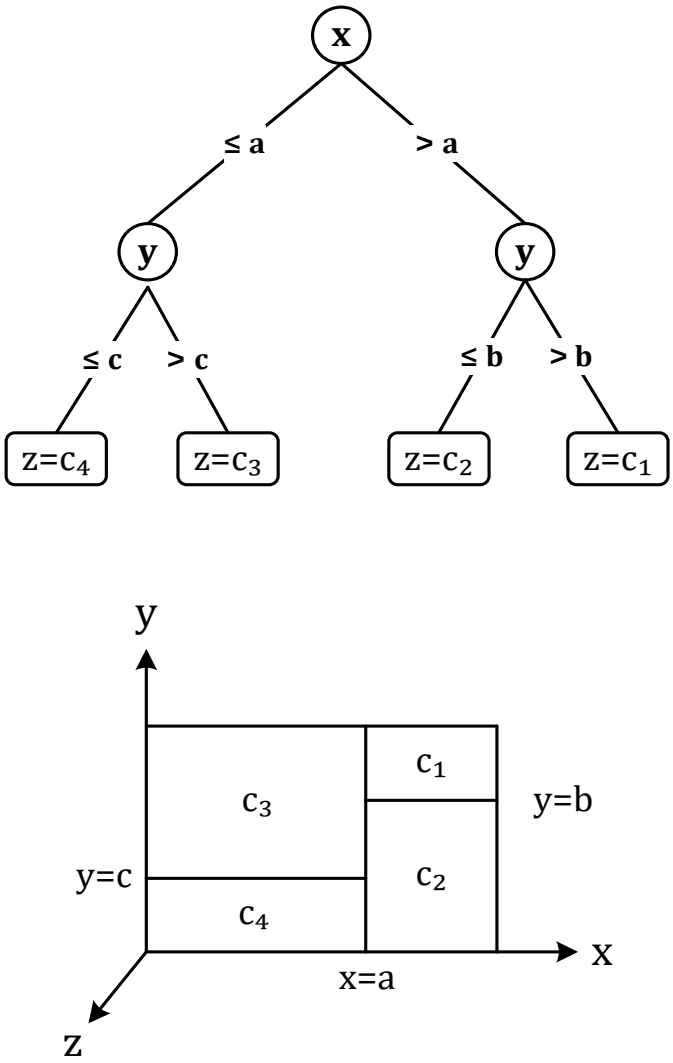


Fig. 1. An example of a decision tree for classification, a) binary decision tree, b) feature space partitioning

DTs algorithms are quite transparent and also do not need optimization of model and internal parameters. However, DTs are rather sensitive to small perturbations in the learning set. It has been demonstrated that this problem can be mitigated by applying bagging (**B**ootstrap **agg**regat**ing**) (Breiman, 1996). Random Forest (RF), as a relatively new pattern recognition method, is a combination of the random subspace method proposed by Ho (1998) and bagging. In this method based on a particular kind of learning strategy "ensemble learning" generate many predictors and average their results. RF can be applied for three purposes: classification, regression and unsupervised learning (Liaw and Wiener, 2002).

In the learning stage, a lot of predictors are generating by selecting a random set of variables and a random sample from the total dataset and then the results are averaged for each tree. RF is very user-friendly in the sense that it has only two parameters: $ntree$, the number of trees in the forest and $mtry$, the number of variables in the random subset at each node, and is usually not very sensitive to their values (Liaw and Wiener, 2002).

In this study the regression RF is utilized to determine the amount of UBC. The algorithm is briefly summarized as follows

(for more detailed information readers are refer to Breiman (2001)).

(1) Given a standard training set X of size n, bagging generates $ntree$ new training sets X_i, each of size n', by sampling from X uniformly and with replacement. If $n' = n$, then for large n the set X_i is expected to have the fraction $(1 - {}^{1}/_{e}) \approx 63.2\%$ of the unique examples of X, the rest being duplicates. This kind of sample is known as a bootstrap sample. The elements not included in X_i are referred to as out-of-bag data (OOB) for that bootstrap sample.

(2) For each bootstrap sample X_i an unpruned regression tree is grown. The tree growing algorithm used in RF is CART. At each node, rather than choosing the best split among all variables as done in classic regression trees, $mtry$ variables are randomly selected and the best split is chosen among them.

(3) New data (out-of-bag elements) are predicted by averaging the predictions of the $ntree$ trees, as explained below.

Out-of-bag elements are used to estimate an error rate, called the out-of-bag (OOB) estimate of the error rate (ERR_{OBB}), as follows:

i. At each bootstrap iteration, the out-of-bag elements are predicted by the tree grown using the bootstrap samples X_i.

ii. For the ith element (y_i) of the training dataset X, all the trees are considered in which the i th element is out-of-bag. On average, each element of X is out-of-bag in one-third of $ntree$ iterations. On the basis of the random trees an aggregated prediction g_{OOB} is developed. The out-of-bag estimate of the error rate is computed as $ERR_{OOB} = (1/ntree)\sum_{i=1}^{ntree}[y_i - g_{OOB}(X_i)]^2$.

The ERR_{OOB} help prevent overfitting and can also be used to choose an optimal value of $ntree$ and $mtry$, by selecting $ntree$ and $mtry$ that minimize ERR_{OOB}. Therefore, we first chose the optimal values of $ntree$ and $mtry$ which minimize ERR_{OOB} and then we proceeded to develop the Random Forest model. Figure 2 presents a general architecture of RF, where i is the number of trees in RF and $(q_u)_1$, $(q_u)_2$ and $(q_u)_i$ are output trees.

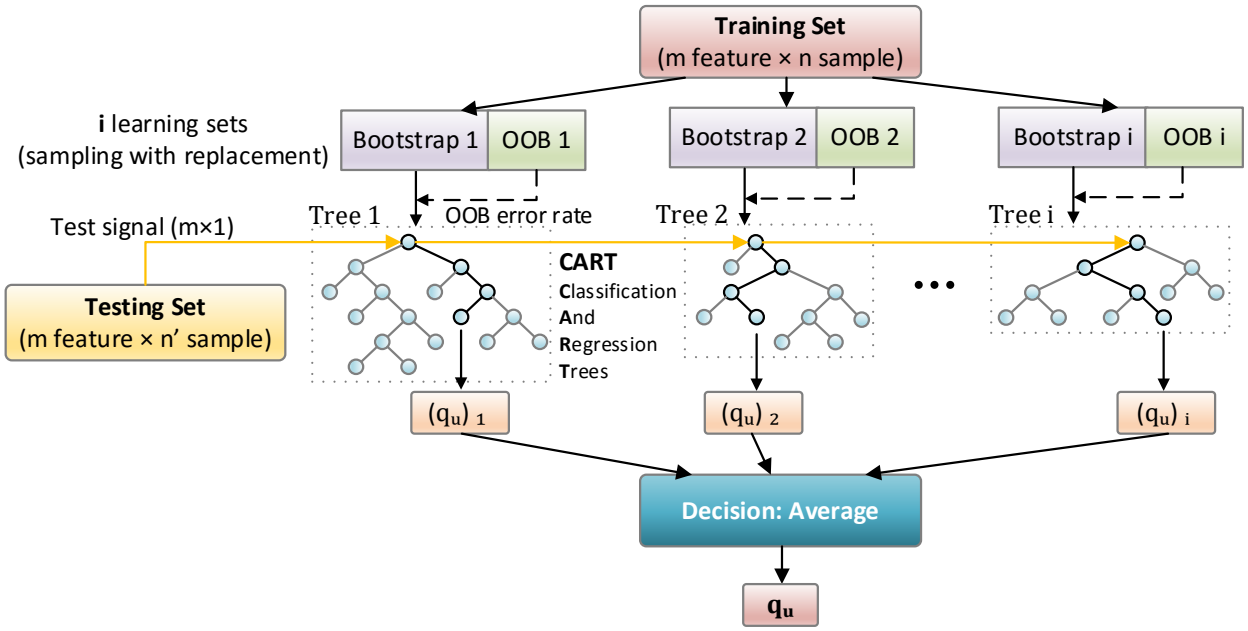


Fig. 2. A general architecture of a random forest for q_u prediction

Theoretical Background of Ultimate Bearing Capacity

Terzaghi (1943) considered a general shear failure and suggested the first semi-empirical equation for UBC. Another investigation was carried out in 1963 by Meyerhof. Considering the shear resistance of the assumed failure surface, Meyerhof (1963) suggested the following equation for centric loading without foundation inclination:

$$q_u = \gamma D N_q F_{qs} F_{qd} + \frac{1}{2}\gamma B N_\gamma F_{\gamma s} F_{\gamma d} \qquad (4)$$

where q_u is UBC of shallow foundations on granular soil (kPa), γ is soil density (kN/m^3), D is depth of the foundation (m), B is width of the foundation (m), $N_q = e^{\pi tan\phi} tan^2\left(45 + \frac{\phi}{2}\right)$, ϕ is angle of internal friction (degree), $N_\gamma = \left(N_q - 1\right)tan(1.4\phi)$,

$$F_{qs} = F_{\gamma s} = \begin{cases} 1 + 0.2K_p\frac{B}{L} & , \ \phi > 10 \\ 1 & , \ \phi = 0 \end{cases}$$

$$F_{qd} = F_{\gamma d} = \begin{cases} 1 + 0.1\sqrt{K_p}\frac{D}{B} & , \phi > 10 \\ 1 & , \phi = 0 \end{cases}$$

Ever since Meyerhof proposed his equation for UBC, it has been tested and modified by other investigators (Hansen, 1970; Vesic, 1974) that are most commonly used for determining the UBC.

Development of RF Model

Model Inputs and Output

In order to obtain accurate prediction of shallow foundation bearing capacity, an understanding of the factors affecting shallow foundation behavior is required. Foundation geometry and mechanical properties of soil are well-established, key parameters in the calculation of UBC of shallow foundation in many of the published methods. A reliability analysis indicated that the key factors were the width of the foundation (*B*), depth of the foundation (*D*), length of the foundation (*L*), unit weight of soil (γ), internal friction angle (ϕ) (Foye et al., 2006). Therefore, the input variables used in this study for model construction are the *B*, *D*, *L*, γ and ϕ. The UBC of the foundation (q_u) is the output variable of the model.

The Data Used for Model Development

A database of previous experimental results was built for training and testing the model. (Kalinli et al., 2011; Padmini et al., 2008). The collected database which is summarized in Table 1 consists of 112 samples of rectangular, square and strip footings of different sizes with centric loading on cohesion less beds of various densities. Among them, only 47 are from load tests on large-scale footings and the rest the reported from scaled model experiments. For more details on the test measurements readers are referred to (Kalinli et al., 2011).

In pattern recognition procedures (e.g. ANN or SVM) the model construction is generally based on adaptive learning over a number of cases. The performance of the trained model is then evaluated using an independent testing data set (Koohestani et al., 2016). Since the ERR_{OOB} is an unbiased estimate of the generalization error, testing the predictive ability of the RF model on an external data set is no needed (Breiman, 2001). However, we preferred to use an independent dataset to perform an external validation of the predictive capabilities of the RF model. Therefore, in this study, from the 112 foundation experiments, 95 were used for to train the model and the remaining 17 tests were used to test the model capability for data generalization randomly.

Table 1. The data used for developing the RF model (Kalinli et al, 2011; Padmini et al., 2008)

B (m)	D (m)	L/B	γ (kN/m^3)	ϕ (°)	q_u (kPa)	B (m)	D (m)	L/B	γ (kN/m^3)	ϕ (°)	q_u (kPa)
0.6	0.3	2	9.85	34.9	270	0.0585	0.058	5.95	17.1	42.5	211
0.6	0	2	10.2	37.7	200	0.094	0.047	6	15.7	34	74.7
0.6	0.3	2	10.2	37.7	570	0.094	0.094	6	15.7	34	91.5
0.6	0	2	10.85	44.8	860	0.094	0.047	6	16.1	37	104.8
0.6	0.3	2	10.85	44.8	1760	0.094	0.094	6	16.1	37	127.5
0.5	0	1	10.2	37.7	154	0.094	0.047	6	16.5	39.5	155.8
0.5	0	1	10.2	37.7	165	0.094	0.094	6	16.5	39.5	185.6
0.5	0	2	10.2	37.7	203	0.094	0.047	6	16.8	41.5	206.8
0.5	0	2	10.2	37.7	195	0.094	0.094	6	16.8	41.5	244.6
0.5	0	3	10.2	37.7	214	0.094	0.047	6	17.1	42.5	235.6
0.52	0	3.85	10.2	37.7	186	0.094	0.094	6	17.1	42.5	279.6
0.5	0.3	1	10.2	37.7	681	0.152	0.075	5.95	15.7	34	98.2
0.5	0.3	2	10.2	37.7	542	0.152	0.15	5.95	15.7	34	122.3
0.5	0.3	2	10.2	37.7	530	0.152	0.075	5.95	16.1	37	143.3
0.5	0.3	3	10.2	37.7	402	0.152	0.15	5.95	16.1	37	176.4
0.52	0.3	3.85	10.2	37.7	413	0.152	0.075	5.95	16.5	39.5	211.2
0.5	0	1	11.7	37	111	0.152	0.15	5.95	16.5	39.5	254.5
0.5	0	1	11.7	37	132	0.152	0.075	5.95	16.8	41.5	285.3
0.5	0	2	11.7	37	143	0.152	0.15	5.95	16.8	41.5	342.5
0.5	0.013	1	11.7	37	137	0.152	0.075	5.95	17.1	42.5	335.3
0.5	0.029	4	11.7	37	109	0.152	0.15	5.95	17.1	42.5	400.6
0.5	0.127	4	11.7	37	187	0.094	0.047	1	15.7	34	67.7
0.5	0.3	1	11.7	37	406	0.094	0.094	1	15.7	34	90.5
0.5	0.3	1	11.7	37	446	0.094	0.047	1	16.1	37	98.8
0.5	0.3	4	11.7	37	322	0.094	0.094	1	16.1	37	131.5
0.5	0.5	2	11.7	37	565	0.094	0.047	1	16.5	39.5	147.8
0.5	0.5	4	11.7	37	425	0.094	0.094	1	16.5	39.5	191.6
0.5	0	1	12.41	44	782	0.094	0.047	1	16.8	41.5	196.8
0.5	0	4	12.41	44	797	0.094	0.094	1	16.8	41.5	253.6
0.5	0.3	1	12.41	44	1940	0.094	0.047	1	17.1	42.5	228.8
0.5	0.3	1	12.41	44	2266	0.094	0.094	1	17.1	42.5	295.6
0.5	0.5	2	12.41	44	2847	0.152	0.075	1	15.7	34	91.2
0.5	0.5	4	12.41	44	2033	0.152	0.15	1	15.7	34	124.4
0.5	0.49	4	12.27	42	1492	0.152	0.075	1	16.1	37	135.2
0.5	0	1	11.77	37	123	0.152	0.15	1	16.1	37	182.4
0.5	0	2	11.77	37	134	0.152	0.075	1	16.5	39.5	201.2
0.5	0.3	1	11.77	37	370	0.152	0.15	1	16.5	39.5	264.5
0.5	0.5	2	11.77	37	464	0.152	0.075	1	16.8	41.5	276.3
0.5	0	4	12	40	461	0.152	0.15	1	16.8	41.5	361.5
0.5	0.5	4	12	40	1140	0.152	0.075	1	17.1	42.5	325.3

B (m)	D (m)	L/B	γ (kN/m^3)	ϕ (°)	q_u (kPa)	B (m)	D (m)	L/B	γ (kN/m^3)	ϕ (°)	q_u (kPa)
1	0.2	3	11.97	39	710	0.152	0.15	1	17.1	42.5	423.6
1	0	3	11.93	40	630	0.08	0	1	17.2	42.8	133
0.991	0.711	1	15.8	32	1773.7	0.15	0	1	17.2	42.8	246
3.004	0.762	1	15.8	32	1019.4	0.05	0	1	17.2	42.8	109
2.489	0.762	1	15.8	32	1158	0.08	0	1	17.1	42.8	130
1.492	0.762	1	15.8	32	1540	0.1	0	1	17.1	42.8	152
3.016	0.889	1	15.8	32	1161.2	0.15	0	1	17.1	42.8	214
0.0585	0.029	5.95	15.7	34	58.5	0.2	0	1	17.1	42.8	266
0.0585	0.058	5.95	15.7	34	70.91	0.25	0	1	17.1	42.8	333
0.0585	0.029	5.95	16.1	37	82.5	0.3	0	1	17.1	42.8	404
0.0585	0.058	5.95	16.1	37	98.93	0.03	0	1	15.89	42	52
0.0585	0.029	5.95	16.5	39.5	121.5	0.04	0	1	15.89	42	92
0.0585	0.058	5.95	16.5	39.5	142.9	0.05	0	1	15.89	42	95
0.0585	0.029	5.95	16.8	41.5	157.5	0.06	0	1	13.2	32	14
0.0585	0.058	5.95	16.8	41.5	184.9	0.06	0	1	14.8	42	72
0.0585	0.029	5.95	17.1	42.5	180.5	0.06	0	1	15.4	42	106

It is worth noting that the ANN literature suggests that the training and test data subsets need to exhibit similar mean value, standard deviation as well as range (Shahin et al., 2004). The statistical characteristics of the test and train subsets are given in Table 2.

Criteria of Evaluation

The performance of the models in predicting can be evaluated using some well-known statistical measures namely coefficient of correlation (CC), coefficient of determination (R^2), root mean square error (RMSE), mean average error (MAE) and mean bias error (MBE) are used. These statistical measures are defined as:

$$CC = \frac{\sum_{i=1}^{n}[(s_i - \bar{s}_i)(c_i - \bar{c}_i)]}{\sqrt{\sum_{i=1}^{n}(s_i - \bar{s}_i)^2(c_i - \bar{c}_i)^2}} \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(s_i - c_i)^2}{\sum_{i=1}^{n}(c_i - \bar{c}_i)^2} \quad (12)$$

$$RMSE = \left(\frac{\sum_{i=1}^{n}(s_i - c_i)^2}{n}\right)^{0.5} \quad (13)$$

$$MAE = \frac{\sum_{i=1}^{n}|s_i - c_i|}{n} \quad (14)$$

$$MBE = \frac{\sum_{i=1}^{n} s_i - c_i}{n} \quad (15)$$

where x_i and y_i denote the predicted and the measured values, respectively, n is the number of measurements and $\bar{x}$ and $\bar{y}$ are the mean of x and y.

Table 2. Statistics of input factors in RF

Parameters	Training				Testing			
	Minimum	Maximum	Mean	StdDev[a]	Minimum	Maximum	Mean	StdDev
B (m)	0.03	3.016	0.375	0.518	0.08	1	0.298	0.25
D (m)	0	0.889	0.145	0.202	0	0.5	0.132	0.151
L/B	1	6	2.704	2.092	1	6	3.459	2.197
γ (kN/m^3)	9.85	17.2	14.395	2.624	11.7	17.2	14.901	2.373
ϕ (°)	32	44.8	38.978	3.43	34	42.8	39.053	3.245
q_u (kPa)	14	2847	415.909	527.895	91.5	1492	326.441	346.412

[a] StdDev refers to the Standard Deviation.

Optimizing the Random Forest Parameters

In this study WEKA (Waikato Environment for Knowledge Analysis) is utilized as a popular machine learning workbench for implementing the RF algorithm. WEKA is developed in java and introduced by Waikato University, New Zealand (Witten and Frank, 2005). In order to determine the optimal value of RF model parameters, $ntree$ and $mtry$, a trial error procedure was employed. The default $ntree$ value is 500. Increasing the number of trees would lead to converge the error rate to a specific limit. Therefore, over-fitting will not occur in large RFs (Breiman, 2001). The default $mtry$ value is $[log_2(N) + 1]$ (N is the total number of variables). It is suggested to start with default $mtry$ and then decrease and increase $mtry$ until the minimum error for the OOB data set is obtained. The stopping criterion (minimum error for the OOB data set) is met for $ntree = 520$ and $mtry = 4$ as shown in (Figure 3).

RESULTS

The statistic measures together with the performance of the trained RF for testing and training data sets are given in Table 3 and Figure 4, respectively. Sensitivity analysis is used to find out the relative importance of independent parameters. For this purpose, parameters were excluded (one by one) from the inputs and the models were developed. The error measures are shown in Table 4. This table shows that the UBC is mostly affected by D and the next important parameters used in this sensitivity analysis are ϕ, L/B, B, and γ, respectively.

To evaluate the efficiency of the trained RF model, the results are compared with some popular classical methods suggested in literature (Meyerhof, 1963; Vesic, 1974; Hansen, 1970) for determining the UBC. The comparison was done for the all dataset. Table 5 shows the values of performance indices for the traditional methods and developed model in this paper. The error indicators reveal that the result of the RF model has much higher values of CC and lower errors (RMSE and MAE) in comparison with the theoretical equations. Also, the equation proposed by Meyerhof shows the best performance among the theoretical formulas. The best fit line of estimated versus measured UBC and the corresponding coefficient of determination (R^2) are illustrated in Figure 5. This figure shows that outputs of the utilized theoretical formulas are more scattered than RF-based forecasts. According to the statistical results obtained (Table 5) and scatter plots depicted, these findings corroborate and demonstrate the superior performance of the RF model compared to the theoretical equations used in this study. The performance of the proposed RF model was also compared with the results of the basic Adaptive Neuro Fuzzy Inference System (ANFIS), ANN and Fuzzy Inference system (FIS), Support Vector Machine (SVM), Relevance Vector Machine (RVM), ant colony optimization (ACO) and genetic programming (GP) which is taken from literature. The performance statistics of all models are presented in Table 6. A comparison between Tables 5 and 6 confirm that soft computing approaches are more accurate compared to analytical formulas.

Table 3. Results of the optimum model with respect to training and testing sets

RF model	CC	RMSE (kPa)	MAE (kPa)
Training set	0.9932	65.33	32.12
Testing set	0.9871	66.88	43.69

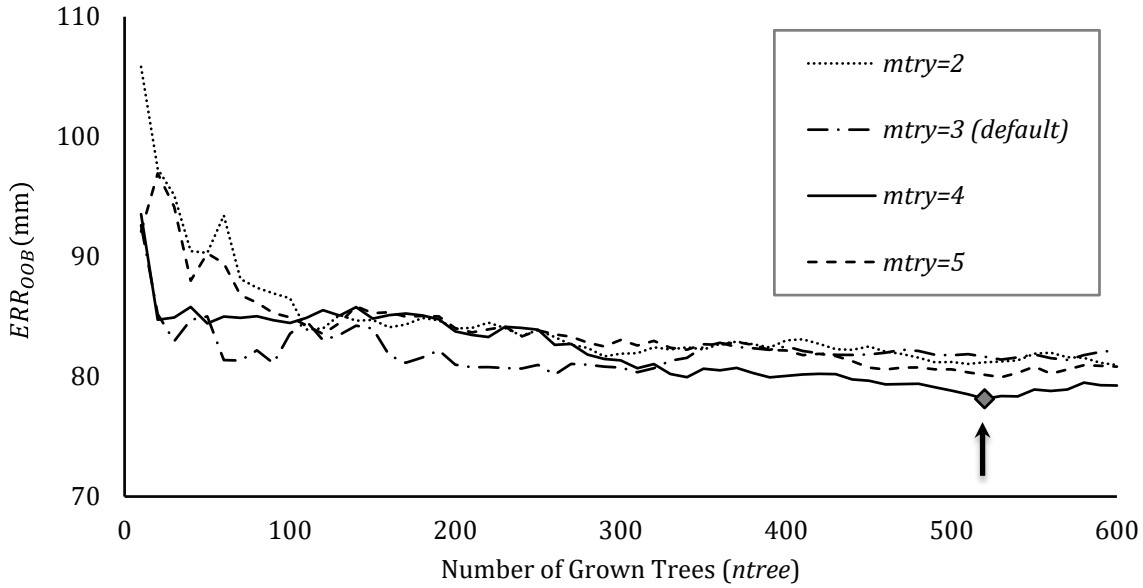


Fig. 3. ERR_{OOB} versus $ntree$ for different $mtry$ (The arrow shows the optimal number of grown tree that produced the least out-of-bag estimate of the error rate)

Table 4. Sensitivity analysis of the governing parameters

Model	Error Statistics		
	CC	RMSE (kPa)	MAE (kPa)
RF with all inputs	0.9932	65.3346	32.1229
RF in the absence of B	0.9903	75.8430	39.6089
RF in the absence of D	0.9221	203.8067	113.2604
RF in the absence of L/B	0.9885	82.1941	37.6184
RF in the absence of γ	0.9914	72.3844	34.1855
RF in the absence of ϕ	0.9870	90.4307	41.5431

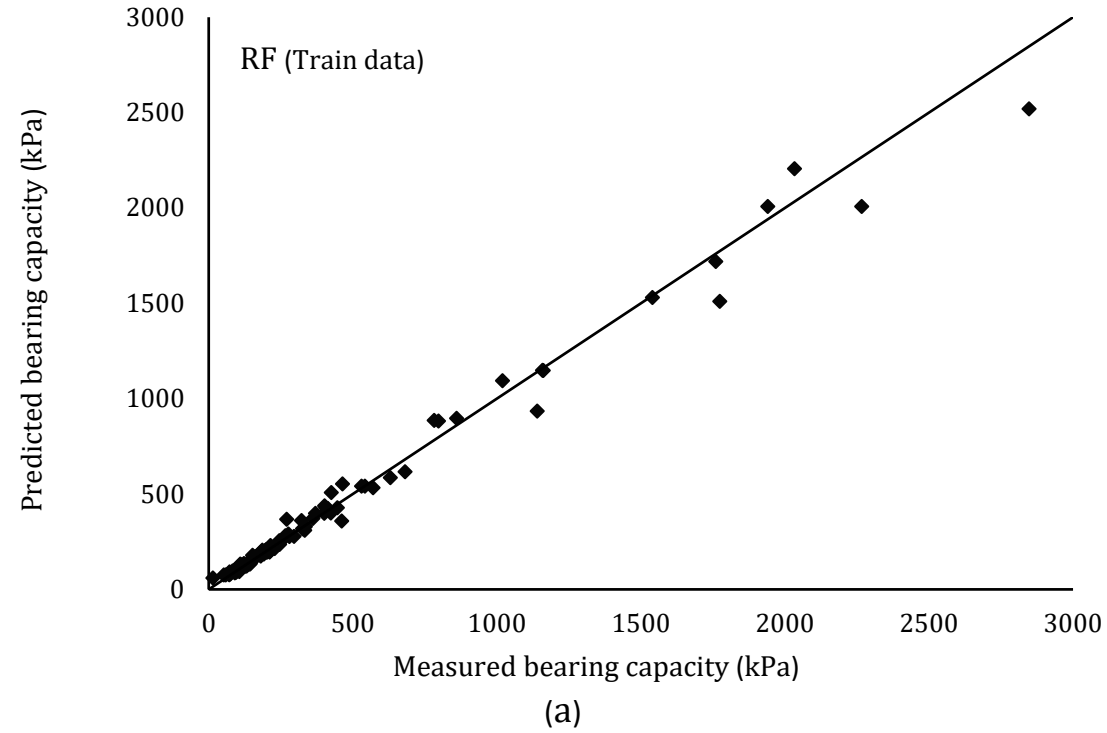
Table 5. Comparison between RF model with theoretical methods

Method	CC	RMSE (kPa)	MAE (kPa)
RF model	0.9926	65.57	33.88
Meyerhof (1963)	0.9389	178.96	96.12
Vesic (1974)	0.9444	221.53	108.83
Hansen (1970)	0.9408	277.81	140.80

Table 6. Compression between RF model with other soft computing methods

Model	Training			Testing		
	R	RMSE (kPa)	MBE (kPa)	R	RMSE (kPa)	MBE (kPa)
RF	0.9932	65.33	2.58	0.9871	66.88	11.5
ANN[a]	0.995	52.9	-1.78	0.992	77.2	-12.04
ANFIS[a]	0.9986	26.4	0	0.996	52.3	11.50
FIS[a]	0.990	71.1	0	0.9989	98	13.93
SVM[b]	0.996	46.59	-0.88	0.993	50.04	2.60
RVM[b]	0.998	28.19	1.93	0.996	29.93	5.73
ACO[c]	0.9989	26.4	0	0.9990	29.2	9.78
GP[d]	0.98	112	-	0.982	121	-

[a] (Padmini et al., 2008)
[b] (Samui, 2012)
[c] (Kalinli et al., 2011)
[d] (Shahnazari and Tutunchian, 2012)



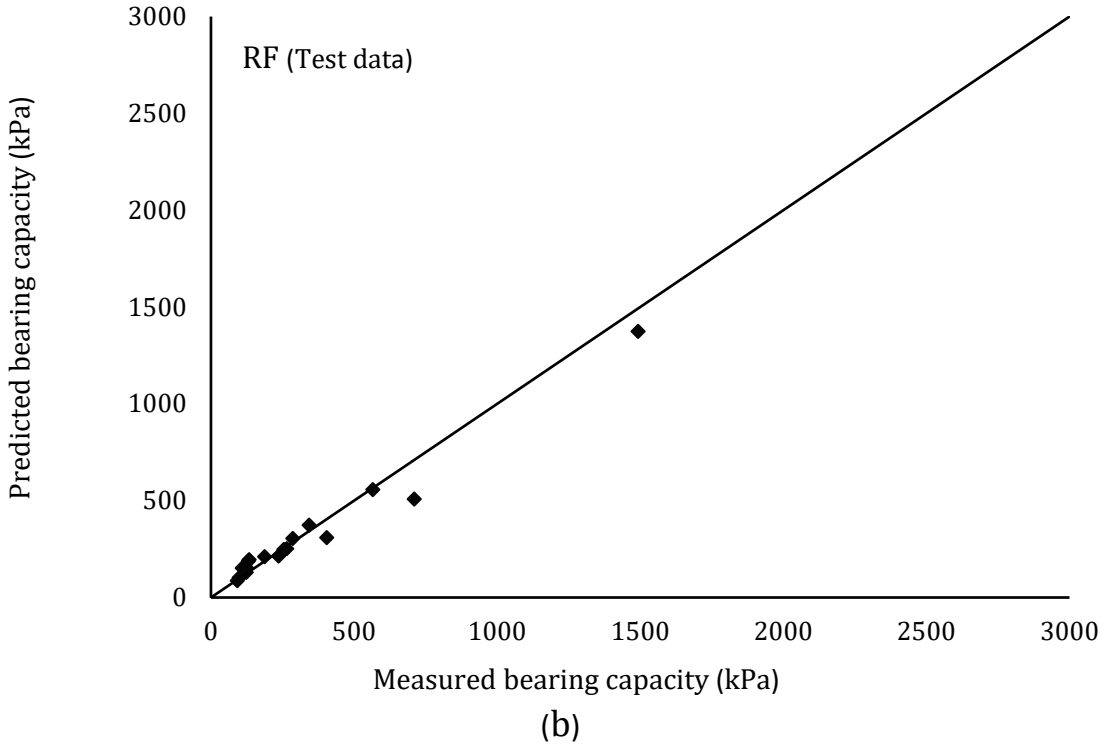
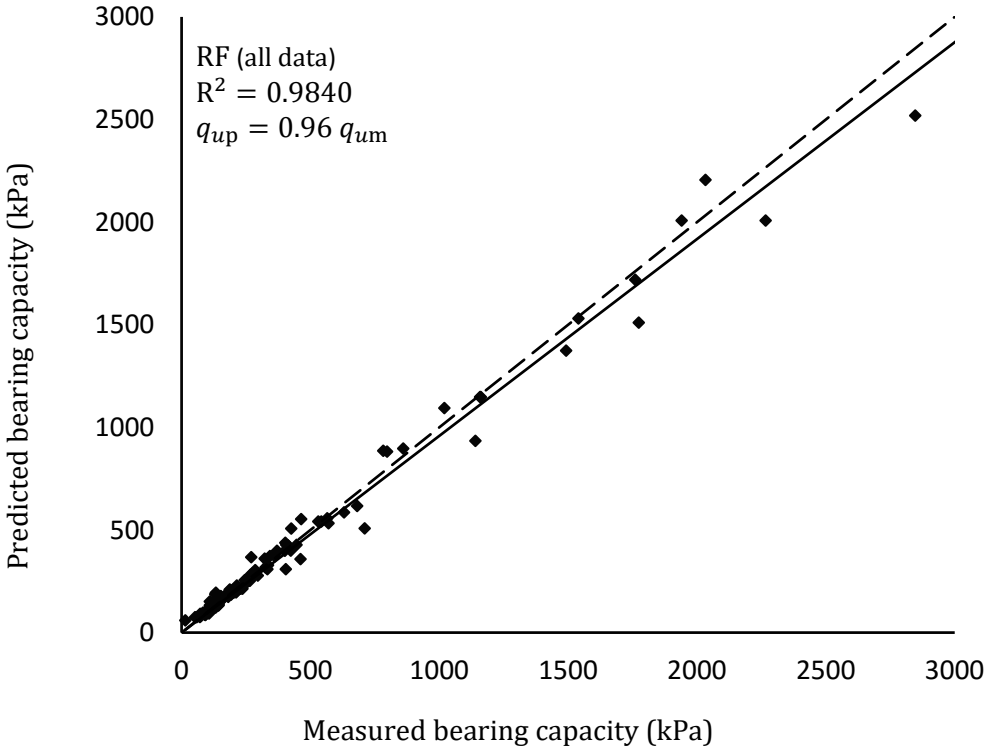
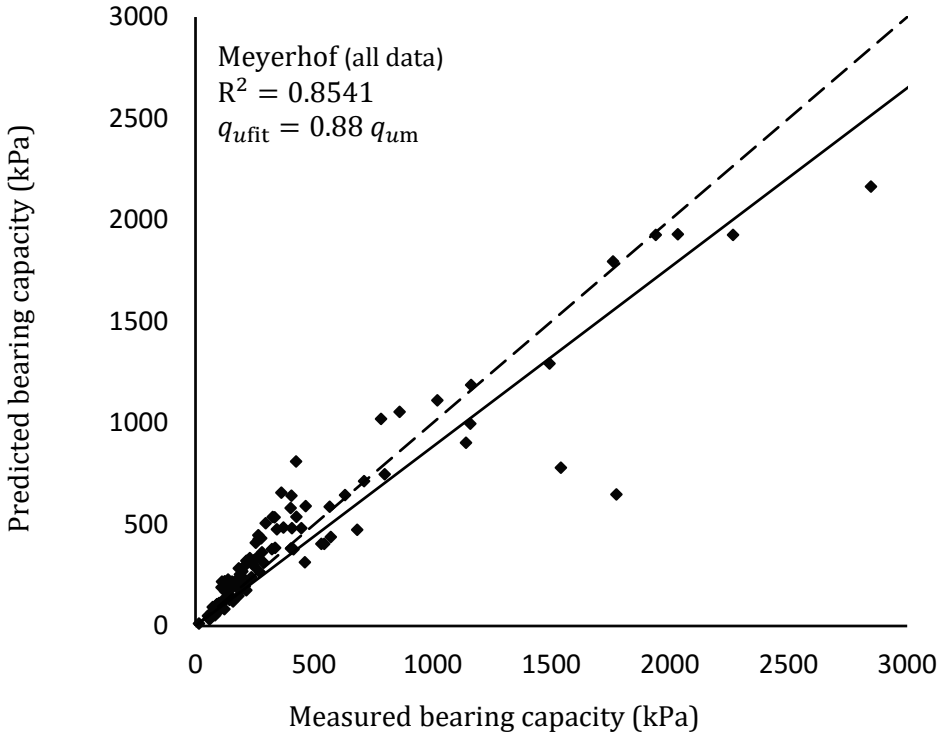


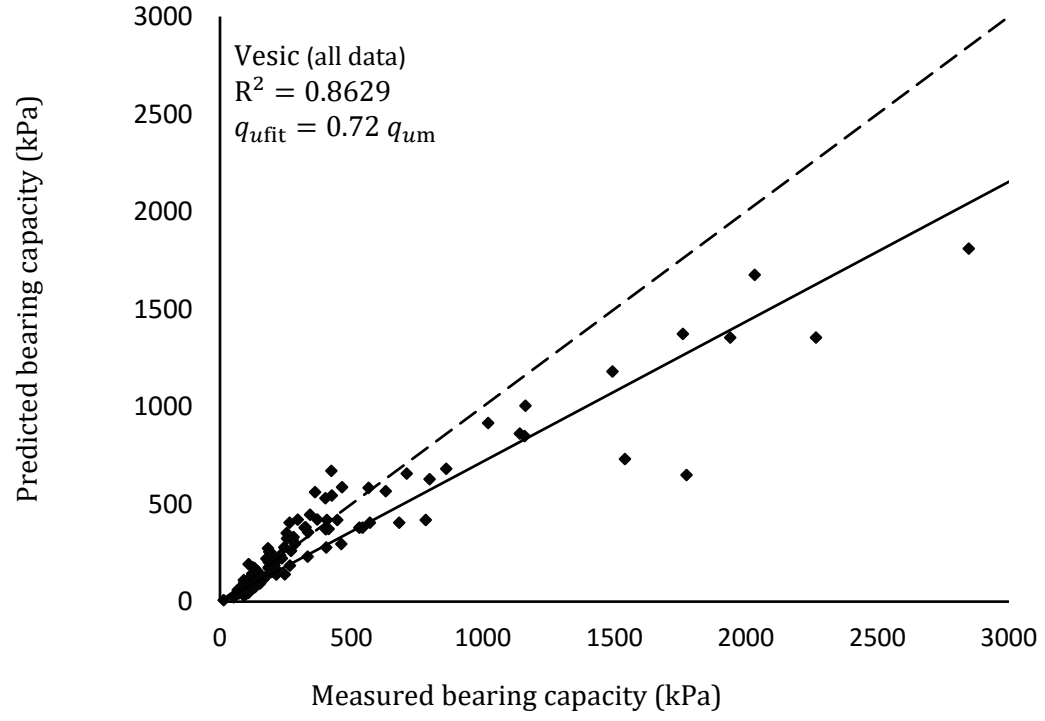
Fig. 4. Measured versus predicted ultimate bearing capacity for RF model, a) training set and b) testing set (solid line indicate best fit)


RF (all data)
$R^2 = 0.9840$
$q_{up} = 0.96\ q_{um}$
Predicted bearing capacity (kPa)
Measured bearing capacity (kPa)
0
500
1000
1500
2000
2500
3000

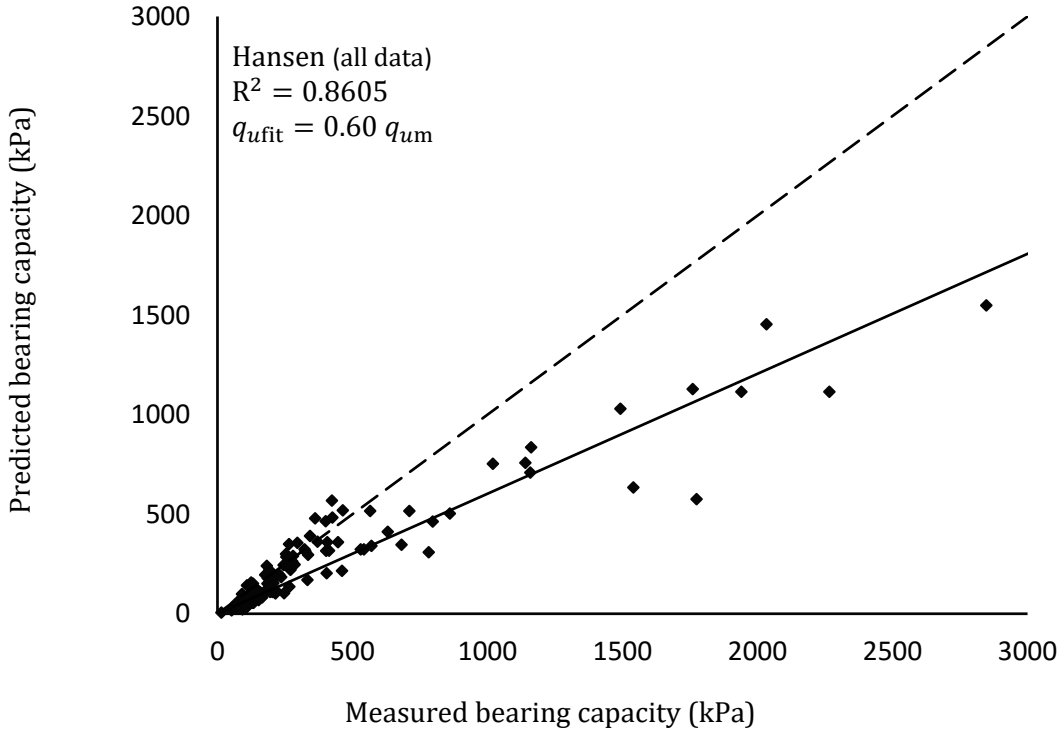
(a)


Meyerhof (all data)
$R^2 = 0.8541$
$q_{ufit} = 0.88\ q_{um}$
Predicted bearing capacity (kPa)
Measured bearing capacity (kPa)
0
500
1000
1500
2000
2500
3000

(b)



(c)



(d)

Fig. 5. Measured versus predicted capacity for (solid line and dashed line indicate best fit and perfect fit lines, respectively): a) RF model; b) Meyerhof; c) Vesic; d) Hansen

DISCUSSION

It is very worthy to be mentioned that there were not parameter tuning, data preprocessing or feature selection used for RF in the results reported here, illustrating that RF is a powerful tool for the regression problems. However, when applying ANN, there are needs of some data preprocessing with decorrelation and normalization to increase the convergence speed of network, or experts to determine a lot of parameters by applying their experiences and priori knowledge, which makes it difficult for the beginners to obtain ANN models with better performance when dealing with actual problems.

RF has some attracted advantages: 1) it is robust against overfitting; 2) it is very user-friendly that there are only two parameters needed to be considered, and RF is usually not very sensitive to their values; 3) it can offer the data internal structure measure, which suggests there is no need of extra feature selection procedure; 4) The internal OOB error rate of RF could be used for classification accuracy assessment when there are limited samples for independent accuracy assessments; 5) it is immune to irrelevant variables and outliers; 6) it is not sensitive to the differences of data units and magnitudes, which suggests it is not necessary to conduct data preprocessing, such as normalizing or centering and 7) it can cope with badly unbalanced data (Liaw and Wiener, 2002). It is worth noting that the RF is only applicable to the range of training data.

CONCLUSIONS

This study presents a new approach using the Random Forest (RF) for estimating the ultimate bearing capacity of shallow foundations based on the recorded experimental data. RF uses the “ensemble learning” strategy which generates many predictors and report the average of predictions. The model inputs consisted of footing geometry and soil parameters. The performance of the developed RF-based model was compared with the equations developed by Terzaghi (1943), Meyerhof (1963), Vesic (1974). The resulted statistical measures and provided scatter plots showed that the RF model is more reliable and has better performance than the traditional equations.

Advantage of the RF model is that this statistical learning modeling framework does not require assumptions of normality of model variables and can deal with non-linear relationships. The RF is fast since splitting the data set into learning and validation subsets for estimating the error which is generally require in some other soft computing techniques (such as SVM and ANN) is not necessary in RF and only two parameters to be tuned experimentally. Taking into account these advantages, it is suggested to use RF model in other fields of geotechnical engineering.

REFERENCES

Ardakani, A. and Kohestani, V. (2015). "Evaluation of liquefaction potential based on CPT results using C4. 5 decision tree", *Journal of AI and Data Mining*, 3(1), 85-92.

Bańczyk, K., Kempa, O., Lasota, T. and Trawiński, B. (2011). "Empirical comparison of bagging ensembles created using weak learners for a regression problem", *Proceedings of Asian Conference on Intelligent Information and Database Systems*, Springer, 312-322.

Barzegari, F., Yousefi, M. and Talebi, A. (2015). "Estimating suspended sediment by Artificial Neural Network (ANN), Decision Trees (DT) and Sediment Rating Curve (SRC) models (Case study: Lorestan province, Iran)." *Civil Engineering Infrastructures Journal*, 48(2), 373-380.

Breiman, L. (1996). "Bagging predictors", *Machine Learning*, 24(2), 123-140.

Breiman, L. (2001). "Random forests", *Machine Learning*, 45(1), 5-32.

Criminisi, A., Shotton, J. and Konukoglu, E. (2011). "Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning", Microsoft Research Cambridge, Technical Report, MSRTR-2011-114, 5(6), 12.

Criminisi, A., Shotton, J. and Konukoglu, E. (2012). "Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning", *Foundations and Trends® in Computer Graphics and Vision*, 7(2-3), 81-227.

De Beer, E. (1965). "The scale effect on the phenomenon of progressive rupture in cohesionless soils", *Proceedings of the 6th International Conference on Soil Mechanism and Foundation Engineering*, Canada, 7-13.

Foye, K., Salgado, R. and Scott, B. (2006). "Assessment of variable uncertainties for reliability-based design of foundations", *Journal of Geotechnical and Geoenvironmental Engineering*, 132(9), 1197-1207.

Hansen, J.B. (1970). *A revised and extended formula for bearing capacity*, Geoteknisk Institut.

Hassanlourad, M., Vosoughi, M., and Sarrafi, A. (2014). "Predicting the grouting ability of sandy soils by Artificial Neural Networks based on experimental tests", *Civil Engineering Infrastructures Journal*, 47(2), 239-253.

Ho, T.K. (1998). "The random subspace method for constructing decision forests", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(8), 832-844.

Kalinli, A., Acar, M. C. and Gündüz, Z. (2011). "New approaches to determine the ultimate bearing capacity of shallow foundations based on Artificial Neural Networks and Ant Colony optimization", *Engineering Geology*, 117(1), 29-38.

Kohestani, V., Hassanlourad, M. and Ardakani, A. (2015). "Evaluation of liquefaction potential based on CPT data using random forest", *Natural Hazards*, 79(2), 1079-1089.

Kohestani, V.R., Bazarganlari, M.R. and Asgari marnani, J. (2016). "Prediction of maximum surface settlement caused by earth pressure balance shield tunneling using random forest", *Journal of AI and Data Mining*, 5(1), 127-135.

Kohestani, V., Hassanlourad, M. and Bazargan-Lari, M.R. (2016). "Prediction the ultimate bearing capacity of shallow foundations on the cohesionless soils using M5P model tree", *Journal of Civil Engineering*, 27(2), 99-110, (in Persian).

Liaw, A. and Wiener, M. (2002). "Classification and Regression by randomForest", *R news*, 2(3), 18-22.

Meyerhof, G.G. (1963). "Some recent research on the bearing capacity of foundations", *Canadian Geotechnical Journal*, 1(1), 16-26.

Padmini, D., Ilamparuthi, K. and Sudheer, K. (2008). "Ultimate bearing capacity prediction of shallow foundations on cohesionless soils using neurofuzzy models", *Computers and Geotechnics*, 35(1), 33-46.

Samui, P. (2012). "Application of statistical learning algorithms to ultimate bearing capacity of shallow foundation on cohesionless soil", *International Journal for Numerical and Analytical Methods in Geomechanics*, 36(1), 100-110.

Shahin, M.A., Maier, H.R. and Jaksa, M.B. (2004). "Data division for developing neural networks applied to geotechnical engineering", *Journal of Computing in Civil Engineering*, 18(2), 105-114.

Shahnazari, H. and Tutunchian, M.A. (2012). "Prediction of ultimate bearing capacity of shallow foundations on cohesionless soils: An evolutionary approach", *KSCE Journal of Civil Engineering*, 16(6), 950-957.

Tatsuoka, F., Okahara, M., Tanaka, T., Tani, K., Morimoto, T. and Siddiquee, M. (1991). "Progressive failure and particle size effect in bearing capacity of a footing on sand", *Proceedings of Geotechnic Engrgineering Congress*, ASCE Geotechnical Special Publication, 788-802.

Terzaghi, K. (1943). *Theoretical soil mechanics*, John Wiley & Sons, New York.

Tsai, H.-C., Tyan, Y.-Y., Wu, Y.-W. and Lin, Y.-H. (2013). "Determining ultimate bearing capacity of shallow foundations using a genetic programming system", *Neural Computing and Applications*, 23(7-8), 2073-2084.

Vesic, A.S. (1974). "Analysis of ultimate loads of shallow foundations", *International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts*, 11(11), A230, Pergamon.

Witten, I.H. and Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*, Morgan Kaufmann.

Yamaguchi, H., Kimura, T. and Fujii, N. (1977). "On the scale effect of footings in dense sand", *Proceedings of the 9th International Conference on Soil Mechanism and Foundation Engineering*, 795-798.