Prediction of ultimate bearing capacity of shallow foundation on granular soils using Imperialist Competitive Algorithm based ANN

Reza Dinarvand\textsuperscript{a}, Mahdi Sadeghian\textsuperscript{a}, Reza Einolvand\textsuperscript{b}
\textsuperscript{a} Master of Civil Engineering, Imam Khomeini International University, Iran
\textsuperscript{b} Associate Degree of Civil Engineering, Islamic Azad University of Dezful, Iran

1. Introduction

Shallow foundations are the foundations with depth to width ratio of less than or equal to four (Das, 1999). The two basic performance criteria of foundations are the ultimate bearing capacity and the settlement (general or relative settlement). The ultimate bearing capacity (\(q_u\)) represents the maximum stress that the underlying soil can withstand without having a shear rupture. The ultimate bearing capacity depends on the mechanical properties of the soil and also the foundation properties and some researchers such as Terzaghi, Meyerhof, and Vesic have provided relationships for determining this quantity (Terzaghi, 1943; Meyerhof, 1963 and Vesic, 1973). Several laboratory experiments have been carried out to determine the bearing capacity of shallow foundations on granular soils, such as De Beer (1965), Yamaguchi et al. (1977), Kutter et al. (1988) and Aiban & Znidarcic (1995). The exact prediction of the ultimate bearing foundation capacity requires an accurate determination of the factors affecting the load bearing capacity. But the exact determination of these factors is sometimes difficult and impossible. Most researchers with hypotheses simplified the problem, whose research results often differ from laboratory results. The best way to study the ultimate bearing capacity of the foundations is using large-scale experiments. But due to the difficulty in using these methods, small scale laboratory models are the appropriate alternative used (Kohestani et al., 2015).

In recent years, the use of soft computing methods, such as artificial neural networks and support vector machines has been widespread in the modeling of complicated geotechnical engineering problems (Lee and Lee, 1996; Shahin, 2010; Kalinli et al., 2011). The advantage of these methods is that the pattern governing the phenomena is directly learned from laboratory and experimental results, and the errors in the lab and experimental data do not have much effect on the network.

So far, soft computing methods such as artificial neural networks, fuzzy logic networks and support vector machines have been used to estimate bearing capacity of shallow foundations (Padmini et al., 2008; Samui, 2012; Tsai et al., 2013). The results indicate that the soft computing methods are superior to experimental methods in predicting the load...

bearing capacity of the shallow foundations. Gupta et al. (2016) investigated the allowable bearing capacity of shallow foundations on granular soils using artificial neural networks. The results showed that artificial neural networks are able to estimate the bearing capacity of the shallow foundations with great accuracy. Sethy et al. (2013) investigated the application of artificial neural networks and fuzzy logic neural networks is estimating the load bearing capacity of rectangular foundations with the out of the central load. The results indicated that fuzzy logic neural networks were more suitable for predicting the load bearing capacity of the shallow foundations.

The quality of the neural network used to predict the behavior of phenomena is examined using network error and generalizability of the proposed model. A network with the lowest error and most generalizability will have the best performance. In recent years, optimization algorithms such as the Genetic Algorithm (GA), the Imperialist Competitive Algorithm (ICA), and the Ant Colony Algorithm has been used to minimize neural network errors. Momeni et al., (2014) by using genetic algorithm based artificial neural networks, investigated the final bearing capacity of the candles, whose results showed a higher accuracy of this method than conventional artificial neural networks. Kalinli et al., (2011) used artificial neural networks based on the ant colony algorithm to predict the ultimate bearing capacity of shallow foundations. The results showed that this model has high accuracy to predict the ultimate bearing capacity of shallow foundations. Marto et al., (2014) have shown that artificial neural networks based on the particle swarm optimization algorithm accurately predict the ultimate bearing capacity of shallow foundations.

The imperialist competitive algorithm, by inspiration from the colonial phenomenon and through the simulation of the socio-political process of colonialism, solves the problems of optimization. The purpose of the present study is to investigate the application of artificial neural networks based on imperialist competitive algorithm in predicting the ultimate bearing capacity of shallow foundations on granular soils.

2. Multi-Layer Perceptron Artificial Neural Networks

An artificial neural network is a collection of simple, interconnected computing elements called the neuron, whose learning ability allows the system to learn complex relationships. These computing units are related to a large number of interconnections, in which all of the knowledge gained from the environment is stored in it. Figure 1 shows the mathematical model of a neural neuron.

![Figure 1. Neural neuron mathematical model (Baranti et al., 2015)](image)

Neural neurons include a set of connections that are characterized by weight or resistance values. A signal, connected to the neuron, is multiplied by weight. A collector collects inputs into the neuron with a bias value. Also, an activity function is used to limit the output of a neuron in a desirable range.

One of the most commonly used neural networks is the multilayer perceptron network, or MLP network. This network has one input layer, one output layer and some hidden layers. In this network, each neuron in each layer is connected with all the neurons of the preceding and the next layer, which has no backward connections in the network. MLP networks with a hidden layer with differentiable transfer functions in the middle and exit layers can approximate all functions with any degree of approximation, provided that they have enough hidden layer of the neuron (Baranti et al., 2015). Figure 2 shows the multi-layer nerve network model.

![Figure 2. Architecture of a multi-layer neural network (Mahinroosta and Farrokh, 2010)](image)
The number of neurons in hidden layer has a great influence on the behavior of the network. If the number of neurons is low, the network cannot accurately reflect the nonlinear mapping between the input and the output. On the other hand, if the hidden layer of the Neuron is more than necessary, the network will maintain the training data by producing a nonlinear complicated data map. But in contrast to the new data, it does not perform well and, in fact, the network loses its generalization power. The number of hidden neurons is empirically obtained (Mnahj, 2004). Back-propagation algorithm (BP) is one of the most popular and most effective algorithms for training MLP networks (Hajihassani et al., 2014).

By using the BP algorithm, the output error of the problem is reduced by adjusting the weight of the connections. At the beginning, the network is trained with randomly selected weights. Then, using the feed-forward back propagation algorithm, all signals are displaced between the output and the middle layers. At the end, the output of the problem is calculated using the network and the difference between the target output and the predicted output is calculated. The difference between these two outputs is the network error. Then the network weights are corrected. This process continues to reach the lowest error (Simpson, 1990 and Kosko, 1994).

3. Imperialist Competitive Algorithm

In 2007, Atashpaz and Lucas presented the Imperialist Competition Algorithm inspired by the colonial phenomenon. The solution of the optimization problem in this way is through the simulation of the socio-political process of colonialism.

In this way, the initial population of probable answers is called the “country.” To get the least error (lowest cost), countries with the best colonial position are called imperialist countries. Imperialists pull some of the colonies based on their power and form the “empires”. After the formation of the empire, the colonies begin to move toward the corresponding imperialists. In an N-dimensional optimization problem with the help of the ICA, each country is defined as an array of 1 × N as follows: Country = (P1, P2, P3, ….PNvariable)

In which P is the parameter that needs to be optimized. Each parameter is defined in a country as a political-social characteristic. The ICA is trying to determine the countries with the best political-social combination. This process can lead to finding the best solution to the problem. To continue this process, the cost function is defined as follows:

Cost = f(country) = f(P1, P2, P3, ….PNvariable)

The process of optimization starts with the selection of the strongest nations as imperialist (Nimp). The rest of the nations are known as the colony (Ncol). Each of these colonies is a member of an empire. The normalized cost of each imperialist is determined as follows:

\[ C_n = c_n - \max_i\{c_i\} \]  

(1)

Where \( c_n \) is the nth colonial cost, and \( C_n \) is the normalized cost. Similarly, the normalized power of each imperialist is defined using the normalization of all imperialists:

\[ p_n = \left| \frac{C_n}{\sum_{i=1}^{N_{imp}} C_i} \right| \]  

(2)

Then the number of initial colony in each empire is as follows:

\[ NC_n = \text{round}\{p_n \cdot N_{col}\} \]  

(3)

Where \( NC_n \) is the initial number of colonies for the nth empire and \( N_{col} \) is the total number of colonies. The next stage is the integration phase in which imperialists are trying to attract colonies. Figure 3 illustrates the process of moving the colonies toward the imperialist. The parameter x is determined as follows:

\[ x \sim U(0, \beta \times d) \]  

(4)
Where $\beta$ is a number which is larger than one and $d$ is the distance between colony and imperialist. The colonial movement is not always a direct vector. To increase the search capability around imperialist, as shown in Fig. 4, a deviation value (parameter $\theta$) is added randomly to the direction of the movement.

$$\theta \sim U(-\gamma, \gamma)$$

Where $\gamma$ is a variation to adjust the deviation. $\beta$ and $\gamma$ are arbitrary values, but the values of 2 for $\beta$ and $\pi/4$ rad for $\gamma$ converge countries to a general minimum (Atashpaz and Lucas, 2007). In the next step, a revolution will occur, resulting in a sudden change in the characteristics of the colonies. This will be done to increase the ability to search and prevent falling into local minimum. Consequently, the colony with the lowest cost changes its position to the imperialist, and the algorithm continues with the imperialist in the new position. The total power of the empire is as follows:

$$TC_n = \text{Cost(imperialist}_n) + \xi \text{mean}\{\text{Cost(colonies of empire}_n)\}$$

Where $T_{C_n}$ is the total cost of the $n$th empire, and $\xi$ is a small positive number. When $\xi$ is a small number, the power of a whole empire is only affected by imperialist power. As the value of $\xi$ increases, the role of the colonies increases in the power of the whole empire. However, the number of 0.1 for $\xi$ in most cases has good results (Atashpaz and Lucas, 2007).

The final stage is the process of optimizing imperialist competition. At this stage, all imperialists try to control the colonies of other empires. Hence, the power of the strongest empire increases and the power of the weakest empire decreases gradually. Moreover, this process will help countries converge to a minimum, provided that it continues. The optimization process is then stopped using an endpoint algorithm. The ICA flowchart is shown in Figure 5.
4. Design of the model

4.1. Database

In order to achieve an efficient model for predicting piles bearing capacity, there is a quest for information about the pile and the surrounding soil.

In this study, the data published by Gandhi (2003) were used. These data include the foundation width (B), the depth of foundation (D), the length to width ratio (L/B), density (γ) and internal friction angle (φ) of soil. This data includes 53 loading tests on small-scale shallow foundations based on granular soil. All the foundations were built without slope. In this study, 80% of the data was allocated to the training and the rest of the data was devoted to the evaluating. Table 1 show the statistical characteristics of the input and output data of the model.

**TABLE 1. STATISTICAL DATA OF INPUT AND OUTPUT DATA MODEL GANDHI (2003)**

<table>
<thead>
<tr>
<th>Statistical characteristics</th>
<th>Depth (cm)</th>
<th>Width (cm)</th>
<th>Length to width ratio</th>
<th>Density (kPa)</th>
<th>Internal friction angle of soil (°)</th>
<th>Foundation bearing capacity (kPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>42.5</td>
<td>17.10</td>
<td>6</td>
<td>15</td>
<td>15.20</td>
<td>420.60</td>
</tr>
<tr>
<td>Min</td>
<td>32</td>
<td>13.20</td>
<td>1</td>
<td>0</td>
<td>5.85</td>
<td>14</td>
</tr>
<tr>
<td>Mean</td>
<td>38.89</td>
<td>16.33</td>
<td>3.81</td>
<td>73.7</td>
<td>10.73</td>
<td>181.41</td>
</tr>
<tr>
<td>STD</td>
<td>3.23</td>
<td>0.71</td>
<td>2.48</td>
<td>4.30</td>
<td>3.77</td>
<td>95.69</td>
</tr>
</tbody>
</table>

Data was transmitted to the range [-1, 1] due to the fact that input and output data of the problem consist of a wide range. This action has an effect on the convergence and proper performance of the network (Fausett, 1994). Hence, in the present research, the following linear transformation was used to transfer data to the desired range:

\[ NP = \frac{UB-LB}{MaxP-MinP} \times (SP - MinP) + LB \]  

In which UB and LB are the upper and lower bounds of the range, MinP and MaxP are the minimum and maximum value of data, and SP and NP are the values of real and normalized data.

4.2. Model Performance Evaluation Indicators

In order to evaluate the accuracy of the model obtained from the neural network, statistical indices such as coefficient of correlation (CC), mean square error (MSE), root mean square error (RMSE) and mean absolute error (MAE) were used. The values of the mentioned indices are obtained using the following relationships:

\[ CC = \frac{\sum_{i=1}^{n}(s_i - \bar{s})(c_i - \bar{c})}{\sqrt{\sum_{i=1}^{n}(s_i - \bar{s})^2(c_i - \bar{c})^2}} \]
\[ \text{MSE} = \frac{\sum_{i=1}^{n} E_i^2}{n} \]  \hspace{1cm} (9)

\[ \text{RMSE} = (\text{MSE})^{0.5} \]  \hspace{1cm} (10)

\[ \text{MAE} = \frac{\sum_{i=1}^{n}|E_i|}{n} \]  \hspace{1cm} (11)

\[ E_i = (s_i - c_i) \]  \hspace{1cm} (12)

In these equations, \( c_i \) is the mean of the observed values of the variable, \( s_i \) is the mean of the calculated model, \( c_i \) is the actual observational value of the variable, \( s_i \) is the calculated value of the variable by the model and the number of observational data.

### 4.3. Model used

Various methods have been used for training neural networks such as genetic algorithm, particle swarm algorithm, ant colony algorithm, and so on. The purpose of these algorithms is to obtain the weights and the constant value, in such a way that the network error is minimized. In this paper, an imperialist competitive algorithm was used for network training.

Achieving the best answer depends on the number of countries and the number of imperialist countries. Nine different states of the total number of countries and imperialist countries were studied. According to Table 2, the case with the total number of countries of 50 and the number of imperialist countries of 5 had the best performance.

<table>
<thead>
<tr>
<th>Model number</th>
<th>Number of countries</th>
<th>Number of imperialists</th>
<th>Best cost of imperialists</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>2</td>
<td>0.0089</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>3</td>
<td>0.0058</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>4</td>
<td>0.0110</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>5</td>
<td>0.0039</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>6</td>
<td>0.0106</td>
</tr>
<tr>
<td>6</td>
<td>70</td>
<td>7</td>
<td>0.0098</td>
</tr>
<tr>
<td>7</td>
<td>80</td>
<td>8</td>
<td>0.0086</td>
</tr>
<tr>
<td>8</td>
<td>90</td>
<td>9</td>
<td>0.0125</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>10</td>
<td>0.0108</td>
</tr>
</tbody>
</table>

To get an efficient model, network architecture has a great impact on network performance. The ICA is only able to adjust the weights and the constant value (bias) of the network and minimize the network error. A network with a hidden layer is able to estimate any continuous functions (Hornik et al., 1989). In this research, a network with two hidden layers was used. The number of neurons in each layer is the most fundamental characteristic of network architecture (Sonmez et al., 2006; Sonmez and Gokceoglu, 2008). There is no definite rule for determining the number of neurons in a layer. In the work of various researchers, a number is proposed between 1 and 18 neurons per layer. In this study, using a trial-error method, a number of networks with a hidden layer and with a different number of neurons were trained and tested. 80% of the data was used for training and the 20% of the remaining were used for testing. The analysis was done in two ways, ANN-BP and ANN-ICA, and these two methods were compared with each other. According to the Figures 6 and 7, a number of 16 neurons for layers were selected as the optimal number.

![Figure 6. Neural network error for different number of neurons (training data)](image)
5. Model results

In Figures 8 the results of training and testing data for ANN-ICA networks are presented. With regards to the correlation between the results of the experiments and the developed models, it can be concluded that the artificial neural networks based on the imperialist competitive algorithm can accurately estimate the final bearing capacity of shallow foundations on granular soils. Also, table 3 illustrates the performance of the ANN-ICA model. According to the obtained results, the model presented by the proposed model is highly accurate. Figures 9 and 10 also show that the predicted output is closely related to the output observed from the experiment. According to the results, the correlation coefficient for training data is 0.9908 and for the testing data is 0.9882.
Figure 10. The results of the model for the testing data

Table 3. Performance results of the ICA-based neural network model

<table>
<thead>
<tr>
<th>Method</th>
<th>CC</th>
<th>MAE</th>
<th>MSE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP-based on ICA</td>
<td>0.9908</td>
<td>0.0462</td>
<td>0.0034</td>
<td>0.0581</td>
</tr>
<tr>
<td>Training data</td>
<td>0.9882</td>
<td>0.0942</td>
<td>0.0158</td>
<td>0.1259</td>
</tr>
</tbody>
</table>

Table 4 shows the comparison between artificial neural networks based on imperialist competitive algorithm with back-propagation artificial neural networks, Meyerhof method, Vesic method and Hansen method. As it is known, artificial neural networks based on the imperialist competitive algorithm, due to the fact that they are not trapped in the local extremes, have a better performance than other methods. The results of the Meyerhof, Vesic and Hansen methods are taken from Gandhi (2003).

Table 4. Comparison of the performance of various methods

<table>
<thead>
<tr>
<th>Method</th>
<th>CC</th>
<th>RMSE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP-based on ICA</td>
<td>0.9903</td>
<td>15.01</td>
</tr>
<tr>
<td>MLP-based on BP</td>
<td>0.9790</td>
<td>21.42</td>
</tr>
<tr>
<td>Meyerhof method</td>
<td>0.9247</td>
<td>95.49</td>
</tr>
<tr>
<td>Vesic method</td>
<td>0.9272</td>
<td>60.68</td>
</tr>
<tr>
<td>Hansen method</td>
<td>0.9325</td>
<td>44.48</td>
</tr>
</tbody>
</table>

For further investigation, the ratio of the predicted lateral bearing capacity to the measured load bearing capacity (Qp/Qm) against the cumulative potential (p) was also plotted. This chart is one of the probabilistic graphs that provide valuable information. In this way, the predicted lateral load bearing capacity ratio was sorted ascending and cumulative probability was obtained using equation 13.

\[ p = \frac{i}{r+1} \]  

(13)

In which, i is the case where the cumulative probability factor is calculated, and r is the total number of reviews. This method offers valuable information in analyzing the dispersion of predictive methods:

1. The ratio of the predicted to the measured data at p = 50% indicates the tendency of the method to be low or high. The closer the number to one is the better.
2. The slope of the line shows the dispersion and standard deviation. If the slope of the line is less indicates better performance of this model.

Figure 11 shows the ratio of the predicted bearing capacity to the measured load bearing capacity against the cumulative potential. As can be seen, in the p = 50% the ratio of the load bearing capacity is very close to value of one, indicating the proper performance of the model. Also, the slope of the ANN-ICA line in relation to the ANN-BP network suggests a more accurate prediction of the model based on the imperialist competitive algorithm. In addition, at p = 90%, the relative bearing capacity of the ANN-ICA network is close to 1.00, indicating that the model is close to reality for the
most of the data. This value for the ANN-BP network is close to 1.22.

![Graph showing the ratio of predicted to measured data against cumulative potential.](image)

**Figure 11.** The ratio of the predicted to the measured data against the cumulative potential

6. **Sensitivity analysis**

Artificial neural networks can only learn the pattern of the phenomenon through training, but they are not able to determine how different parameters influence the output of the model. By the sensitivity analyzing, we can see the effect of each input on the output of the problem. Table 3 shows the results of the sensitivity analysis of input data and their effect on the output of the model. In the method used, with the removal of an input, the network was re-trained and the correlation coefficient and model error were recorded. According to the results of the sensitivity analysis, the ultimate bearing capacity of the foundations is affected by the depth of the foundation, the width of the foundation, the internal friction angle of the soil, the soil density and the length to width ratio, respectively.

<table>
<thead>
<tr>
<th>Input</th>
<th>CC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>B+D+L/B+γ</td>
<td>0.982</td>
<td>0.1075</td>
</tr>
<tr>
<td>B+D+L/B+ϕ</td>
<td>0.971</td>
<td>0.04598</td>
</tr>
<tr>
<td>B+D+ϕ+γ</td>
<td>0.9905</td>
<td>0.0158</td>
</tr>
<tr>
<td>B+ϕ+L/B+γ</td>
<td>0.89558</td>
<td>0.18166</td>
</tr>
<tr>
<td>ϕ+D+L/B+γ</td>
<td>0.97194</td>
<td>0.10776</td>
</tr>
<tr>
<td>B+D+L/B+ϕ+γ+ϕ</td>
<td>0.9903</td>
<td>0.0717</td>
</tr>
</tbody>
</table>

7. **Conclusion**

In the present paper, multi-layer perceptron (MLP) neural networks based on the Imperialist Competitive Algorithm (ICA) were used to predict the ultimate bearing capacity of shallow foundations on granular soils. Data from 53 small scale tests were used to train and test the model. By comparing the results obtained from the neural network and the laboratory results, it was observed that the training data with statistical indices CC = 0.9908 and RMSE = 0.0581, as well as test data with CC = 0.9882 and RMSE = 0.1259 is able to predict experiment results. Also, the results of the model showed the superiority of ICA-based artificial neural networks compared to back-propagation based artificial neural networks as well as the Meyerhof, Vesic and Hansen methods. The results of the sensitivity analysis of the model showed that the depth of the foundation had the greatest effect and the length to width ratio had the least effect on the pile ultimate bearing capacity.

The scope of application of the results obtained in the present study is limited to the data used in constructing the model. Therefore the neural network model is always able to learn new data, and it can be re-train the network by entering a wider range of data.

8. **References**


Sethy, B. P., Patra, C. R., Sivakugan, N., and Das, B. M. (2017). Application of ANN and ANFIS for Predicting the


