**Geomechanics and Engineering**, *Vol. 43, No. 2 (2025) 89-98*
https://doi.org/10.12989/gae.2025.43.2.089


# Optimization of ANN and determination of optimal network parameter range to predict suction caisson foundation capacity

Jaehyeok Han[1], Uichan Lee[1], Jaehun Ahn[2],
Jongmuk Won[3], Nhat-Duc Hoang[4,5] and Jongwon Jung*[1]

[1]*School of Civil Engineering, Chungbuk National University, Republic of Korea*
[2]*Department of Civil and Environmental Engineering, Pusan National University, Republic of Korea*
[3]*Department of Civil Urban, Earth, and Environmental Engineering,*
*Ulsan National Institute of Science and Technology, Republic of Korea*
[4]*Institute of Research and Development, Duy Tan University, Da Nang, 550000, Vietnam*
[5]*Faculty of Civil Engineering, Duy Tan University, Da Nang, 550000, Vietnam*

(*Received August 4, 2025, Revised October 4, 2025, Accepted October 10, 2025*)

**Abstract.** Suction caisson foundations are frequently used to moor offshore structures in the oil drilling and wind power generation industries. Though artificial neural network (ANN) models have been successfully applied to predict pile foundation capacity, the considerable differences between the characteristics of pile and suction caisson foundations imply that an ANN model trained using data from the former cannot be applied to predict the capacity of the latter. This study accordingly employed suction caisson foundation data to develop an ANN capable of accurately predicting the capacity of such foundations. The early stopping and model checkpoint techniques were applied to prevent overfitting by saving the immediately prior optimal weight. To obtain the optimal hyperparameter conditions efficiently, a Bayesian optimization algorithm was employed, which significantly reduced the optimization time. This algorithm produced four hyperparameter combinations that exhibited excellent performance; these were each used to train the ANN 500 times, thereby accounting for the uncertainty owing to randomly assigned initial weights. The proposed ANN was subsequently developed using two approaches: parameter analysis and optimization. The parameter analysis determined that the optimal number of network parameters for the selected hyperparameter combinations was 7,638, which was within the 500-650,000 range determined by a general analysis. The verification root mean square error(RMSE) of the ANN model developed using the optimization process was 8.88 with a coefficient of determination of 0.9998. Notably, because suction caisson foundation data have characteristics consistent with general geotechnical engineering practices, the optimal network parameter range and optimization method employed in this study to develop the ANN can be used with other data obtained in the geotechnical field.

**Keywords:** artificial neural network; network parameter; optimization; overfitting; suction caisson foundation

## 1. Introduction

The suction caisson, first proposed by Goodman *et al.* (1961), is a basic foundation system with a sealed cylindrical upper part and a lower part inserted into the ground using its own weight. Water is subsequently drained from the inside of the caisson and the foundation resistance is secured by the pressure difference between the inside and outside of the caisson. Suction caissons have been widely used as foundations for marine structures in the oil drilling industry and are increasingly applied with floating offshore wind power generation structures to meet social demands for green energy.

Many studies on the behavior of suction caisson foundations have been conducted using centrifugal model experiments (Brown and Nacci 1971, Fuglsang and Steensen-Bach 1991, Andersen *et al*. 1993, Watson and Randolph 1997, Watson *et al*. 2000, House and Randolph 2001, Clukey and Morrison 1993, Kelly *et al*. 2006, Murillo *et al*. 2009). The horizontal and vertical loads and moments resisted by suction caisson foundations have also been evaluated through indoor and field experiments (Bryne and Houlsby 2004, Houlsby *et al*. 2005). The results of such experiments have been applied to evaluate the holdin g capacity of the suction caisson foundation in a marginal equilibrium state. However, the difficulties associated with physical experimentation have led to the recommendation that evaluations of support and subsidence be conducted using three-dimensional finite element analyses instead (DNV 2005). Such analyses may need to consider nonlinear effects depending on the environmental and design conditions of the target application. A variety of studies have been conducted to predict the capacity and pullout resistance of suction caissons based on numerical analyses (Deng 2001, Cao *et al*. 2002, House and Randolph 2001, Aubeny *et al*. 2003, Zhan and Liu 2010, Ahn *et al*. 2014, Kim *et al*. 2017).

Recently, research has been actively conducted to predict pile capacity using machine learning or artificial neural network (ANN) techniques (Al-Swaidani *et al*. 2024, Savvides and Papadopoulos 2024, Tariq *et al*. 2024). Such efforts have employed data from pile loading tests (Abu Kiefa 1998, Alzo'ubi and Ibrahim 2019, Bajaj *et al*. 2019, Chow *et al*.

*Corresponding author, Professor
E-mail: jjung@chungbuk.ac.kr

**Copyright © 2025 Techno-Press, Ltd.**

ISSN: 2005-307X (Print), 2092-6219 (Online)

1995, Harandizadeh *et al*. 2021, Ismail *et al*. 2013;, Jebur *et al*. 2021, Shahin 2014a, Shaik *et al*. 2019, Borthakur and Dey 2020, Samui 2011, Samui 2019), cone penetration tests (Ardalan *et al*. 2009, Ghorbani *et al*. 2018, Harandizadeh 2020, Shahin 2010, Shahin 2014b, Alkroosh *et al*. 2015), or dynamic loading tests (Park and Cho 2010, Armaghani *et al*. 2020, Momeni *et al*. 2014, Momeni *et al*. 2015, Shatnawi *et al*. 2019, Tarawneh and Imam 2014, Teh *et al*. 1997, Yong *et al*. 2021). Studies have also been conducted to predict the ultimate pile capacity using various ANN algorithms and numerical analysis data (Benali 2021, Das and Dey 2018, Kardani *et al.* 2020, Liu *et al*. 2010, Moayedi and Armaghani 2018, Prayogo and Susanto 2018, Samui 2012, Shaikh *et al.* 2019, Singh and Walia 2017, Singh *et al.* 2018, Sun *et al*. 2020, Pham *et al.* 2020).

Though research using machine learning or ANN techniques to estimate pile capacity is ongoing, ANNs trained on pile data cannot be directly applied to suction caisson foundations as the latter relies upon the difference between internal and external pressures induced during construction and are primarily used in deep seas where they are affected by waves and currents. As a result, the development of such modeling methods for the prediction of suction caisson capacity remains lacking. This study accordingly developed a technique for optimizing an ANN to predict the capacity of suction caisson foundations using the appropriate data.

## 2. ANN analysis and optimization

Three aspects must be considered when developing analysis methods based on ANNs: first, overfitting must be avoided by careful selection of the learning process; second, detailed descriptions of the trial-and-error method applied for optimization and of the final hyperparameter determination method should be provided; third, the variability in the learning results should be considered even under the same conditions owing to the randomness inherent to the initial model weights. This study developed an ANN model considering all three of these aspects. Overfitting was prevented by saving the optimal weights immediately prior to overfitting using the early stopping and model checkpoint techniques. The optimization time was shortened using a Bayesian algorithm and the basis for the optimal hyperparameter conditions was established. Finally, four hyperparameter combinations that exhibited high performance were selected from the Bayesian optimization results and used to train the ANN 500 times each, thereby accounting for the uncertainty owing to the randomly assigned initial weights.

### *2.1 Data describing the suction caisson foundation*

The holding capacity of the suction caisson foundation is provided by the tip resistance and skin friction after installation via gravity. The schematic in Fig. 1 defines the parameters used to calculate holding capacity, in which $L$ denotes the length of the suction caisson, $D$ denotes the diameter of the suction caisson, $\Psi$ denotes the angle of the load at yield, $S_{u0}$ denotes the undrained shear strength of the caisson–soil interface, and $S_{u1}$ denotes the undrained shear strength per unit depth.

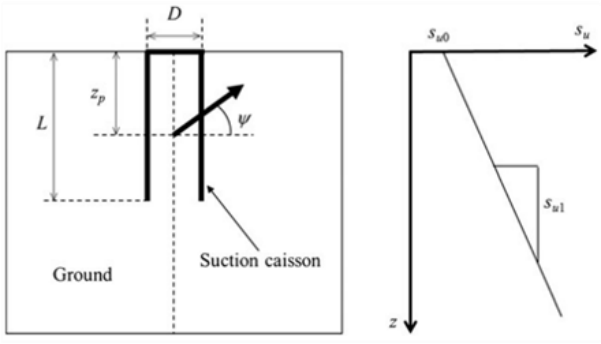



Fig. 1 Suction caisson parameters (Kim *et al*. 2017)

In this study, the variables necessary to predict the holding capacity of a suction caisson foundation were normalized to consider only three input variables in the model: the aspect ratio of the suction caisson, defined as $L/D$ ; the normalized undrained shear strength characteristics, defined as $S_{u1}D/S_{u0}$; and the load angle $\Psi$ at the time of yield. The normalized suction caisson holding capacity was selected as the output value. Fig. 2 depicts the input variable values from Kim *et al*. (2017) used in this study; of these 127 data points, 92 were used for training and 35 for validation. And the data were obtained through Abaqus, a finite element analysis program. Data normalization was performed during preprocessing to ensure that the model consistently recognized the input data and considered all the input variables equally. The Z-score normalization method was applied to do so as follows

$$\text{Data Normalization} = \frac{(x-\mu)}{s} \tag{1}$$

where $x$ is the data point value, the $\mu$ is the average value of all data points, and $s$ is the standard deviation of the data.

This normalization approach yields a set of dimensionless characteristic input variables of the same size. The holding capacity of asuction caisson foundation defined as $P_{fm}$. The foundation capacity was calculated as $P_{fm}/S_{u0}D^2$ considering the undrained shear strength of the caisson–soil interface and the caisson diameter, and the normalized suction caisson holding capacity was selected as the characteristic output value.

Table 1 shows the ranges, averages, and covariances of the normalized characteristic input variables and output values used to train and validate the model. All input and output values of the validation data were within the range of the training data. In addition, the normalized input variables $L/D$ and $S_{u1}D/S_{u0}$ indicated a positive covariance with respect to the output value $P_{fm}/S_{u0}D^2$ whereas the input variable $\Psi$ indicated a negative covariance. Table 1 also reports the Pearson coefficients expressing the correlation between the characteristic input and output values. An analysis of the correlation between input values indicated a Pearson coefficient of 0.012 between $L/D$ and $S_{u1}D/S_{u0}$, -0.18 between $L/D$ and $\Psi$, and -0.03 between $S_{u1}D/S_{u0}$ and $\Psi$. In addition, the Pearson coefficients between the input variables $L/D$, $S_{u1}D/S_{u0}$, and $\Psi$ and output value $P_{fm}/S_{u0}D^2$ were 0.71, 0.44, and -0.37, respectively. Thus, the

Table 1 Statistics describing the training and validation data

| Characteristic Training data | $L/D$ | $S_{u1}D/S_{u0}$ | $\Psi$ | $P_{fm}/S_{u0}D^2$ |
|---|---|---|---|---|
| Min | 2 | 0 | 0 | 13.3 |
| Max | 10 | 5 | 90 | 3371.6 |
| Average | 6.04 | 2.46 | 30.97 | 677.78 |
| Covariance | 1838.66 | 754.17 | 8506.63 | - |
| Pearson coefficient | 0.66 | 0.51 | -0.33 | - |
| Characteristic Training data | $L/D$ | $S_{u1}D/S_{u0}$ | $\Psi$ | $P_{fm}/S_{u0}D^2$ |
| Min | 2 | 0 | 0 | 19.5 |
| Max | 10 | 5 | 90 | 3119.90 |
| Average | 5.89 | 2.6 | 26.29 | 711.97 |
| Covariance | 1961.75 | 622.63 | 7251.28 | - |
| Pearson coefficient | 0.71 | 0.44 | -0.37 | - |

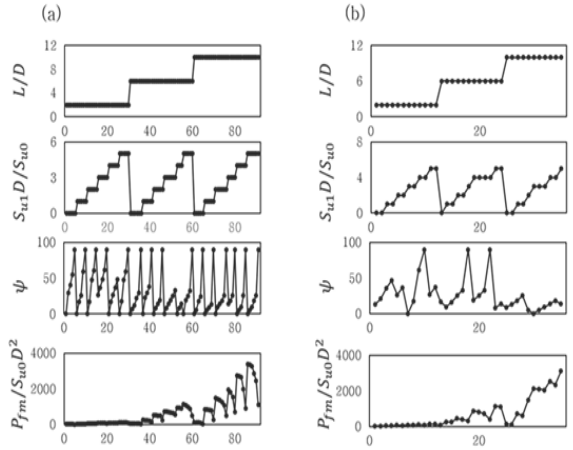



Fig. 2 Normalized data used for: (a) Training and (b) Validation

Pearson coefficients between input variables were all close to 0, and the Pearson coefficients between the input variables and output value remained less than 0.8, indicating that the dataset did not exhibit any multicollinearity problem (Bae 2002).

### *2.2 Parametric analysis*

A parametric analysis was conducted using the condition search method to determine the influences of each hyperparameter (the activation function, weight initialization method, optimizer, number of hidden layers, number of hidden layer nodes, and learning rate) on the number of network parameters and model performance. The rectified linear unit (ReLU), sigmoid, and tanh functions were evaluated as activation functions. The He normal or Glorot normal weight initialization methods were applied. The Adam, Adagrad, AdaDelta, Nadam, and RMSProp optimizers were considered. Ranges of 1–5 hidden layers and 2–4096 hidden layer nodes were evaluated along with learning rates from 0.001 to 0.1. All combinations of all hyperparameters were included in the considered cases. When performing the analysis for each hyperparameter, four different combinations of the remaining hyperparameters were evaluated for comparison. In contrast to the hyperparameters, which were varied according to the test case, the parameters that change according to the learning process within the model were considered network parameters. The number of network parameters was determined by the number of input variables, hidden layers, and hidden nodes as follows

$$N_1(\mathrm{k}+1)+\sum_{i=2}^{L} N_i(N_{i-1}+1)+N_O(N_L+1) \qquad (2)$$

where $N_1$ is the number of nodes in the first hidden layer, $k$ is the number of input variables, $L$ is the number of hidden layers, and $N_0$ is the number of nodes in the output layer. Based on Eq. (2), a consistent number of network parameters was generally applied; however, as stated above, the number of hidden layers (a hyperparameter) was changed to analyze the effects of an increase or decrease in the number of network parameters.

### *2.3 Optimization*

Three approaches were applied under the same conditions to optimize the performance of the model: (1) model checkpoint and early stopping techniques, (2) shortening the optimization time and securing optimal hyperparameter conditions using a Bayesian algorithm, and (3) consideration of uncertainty through iterative learning. A detailed description of each method is provided below.

#### *2.3.1 Model checkpoint and early stopping methods*

The model checkpoint method stores the weights of the model as long as the root mean square error (RMSE) used for validation is smaller in the current learning epoch than in the previous training epoch. Thus, overfitting can be prevented by obtaining the weight corresponding to the number of training iterations just before overfitting is achieved.

The early stopping method observes the RMSE of the validation data during the model training process and terminates training when the RMSE for the current epoch is greater than that of the previous epochs. Thus, the training time can be reduced and overfitting prevented.

#### *2.3.2 Bayesian optimization*

Bayesian optimization reduces the need for researcher intervention and shortens the optimization process. In this

study, Bayesian optimization algorithms were applied to control the number of hidden nodes in the first layer, the number of hidden nodes in the second layer, and the learning rate, and explore the hyperparameter combinations in the suction caisson foundation data. The remaining hyperparameters, determined to be optimal through the parametric analysis, were set as the Adam optimizer, ReLU activation function, He normal weight initialization, and two hidden layers were set, which satisfied the minimum conditions for deep learning and provided differentiation from the low-layer structure of the first layer, as has been typically used in previous ANN studies. The Bayesian optimization algorithm was applied to search randomly for the initial five searches, then searching 1000 iterations. The search range was set to select from 1–4096 hidden nodes in each layer and from learning rates of 0.001–0.1.

#### *2.3.3 Iterative learning*

Because the initial weights of the model were assigned random values, they changed each time training was attempted, which affected the results. To consider the change in training results according to the assignment of random values, four hyperparameter combinations with low validation RMSE values were selected from among the results explored through Bayesian optimization, and training was repeated 500 times for each. A model was selected once the initial weights were distributed using the random values that were most suitable for training, thereby accounting for the variability owing to the initial weights.

## 3. Results

The model training results are discussed in terms of the parametric analysis and optimization processes in this section.

### *3.1 Parametric analysis*

The parametric analysis was conducted by varying one hyperparameter at a time within a selected set of conditions for the remaining hyperparameters. Only the RMSEs of the validation data were used to determine the effects of each hyperparameter. The early stopping method was applied to stop the iteration of epochs and terminate the training process when the validation RMSE increased.

For clarity, each "Condition" label shown in Figs. 3-5 represents a specific combination of the number of hidden layers, hidden nodes, and learning rate used during training.

In particular, Conditions 1 and 2 correspond to relatively shallow networks (one hidden layer) with high and low learning rates, respectively, whereas Conditions 3 and 4 represent deeper networks (two hidden layers) trained under similar learning rate variations.

These distinctions allow the influence of each hyperparameter to be independently examined and facilitate consistent comparison across figures.

#### *3.1.1 Optimizer*

First, we examined the use of five different loss function optimizers during training. The validation RMSE value for each optimizer is compared in Fig. 3 under the four hyperparameter conditions listed in Table 2. The Adam optimizer exhibited the lowest validation RMSE and thus the best optimization performance under all four conditions. Therefore, the Adam optimizer was applied when performing the subsequent parametric analyses.

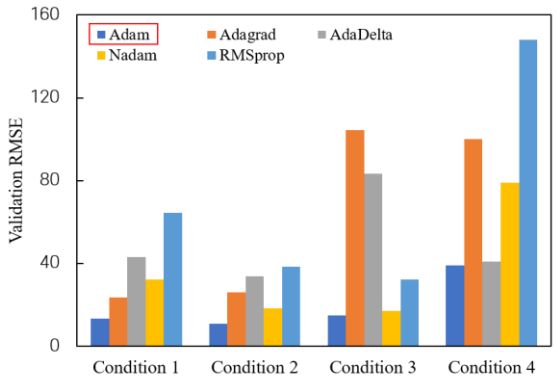



Fig. 3 Comparison of validation RMSEs according to the optimizer

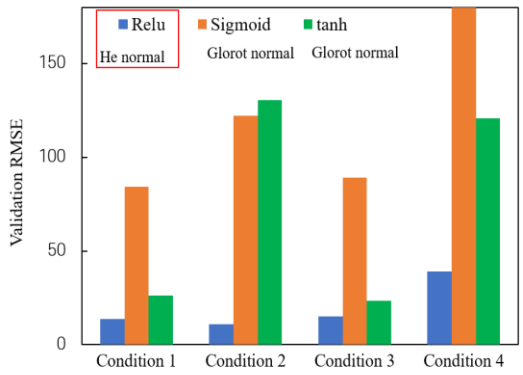



Fig. 4 Comparison of validation RMSEs according to activation function and weight initialization method

Table 2 Conditions for parametric analysis of the optimizer hyperparameter

| Condition | Number of hidden layers | Number of hidden nodes | Learning rate | Activation function |
|---|---|---|---|---|
| Condition 1 | 1 | 512 | 0.05 | ReLU |
| Condition 2 | 2 | 256 | 0.005 | |
| Condition 3 | 1 | 1024 | 0.005 | |
| Condition 4 | 2 | 1024 | 0.005 | |

#### *3.1.2 Activation function*

Fig. 4 shows the change in validation RMSE according to the activation function employed under the four conditions listed in Table 3. Because there exists an initial weight distribution suitable for training according to the activation function, the He normal weight initialization was used with the ReLU activation function, and the Glorot normal weight initialization was used with the Simoid and tanh activation functions (He *et al.* 2015). The results

Table 3 Conditions for parametric analysis of activation function and weight initialization method

| Condition | Number of hidden layers | Number of hidden nodes | Learning rate | Optimizer |
|---|---|---|---|---|
| Condition 1 | 1 | 512 | 0.05 | Adam |
| Condition 2 | 2 | 256 | 0.005 | |
| Condition 3 | 1 | 1024 | 0.005 | |
| Condition 4 | 2 | 1024 | 0.005 | |

Table 4 Conditions for parametric analysis of the number of hidden nodes

| Condition | Number of hidden layers | Learning rate | Optimizer | Activation function |
|---|---|---|---|---|
| Condition 1 | 1 | 0.05 | Adam | ReLU |
| Condition 2 | 1 | 0.005 | | |
| Condition 3 | 2 | 0.05 | | |
| Condition 4 | 2 | 0.005 | | |

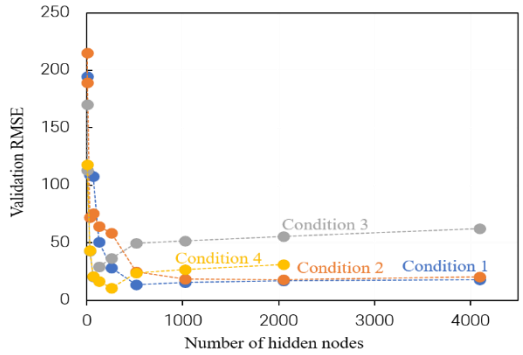
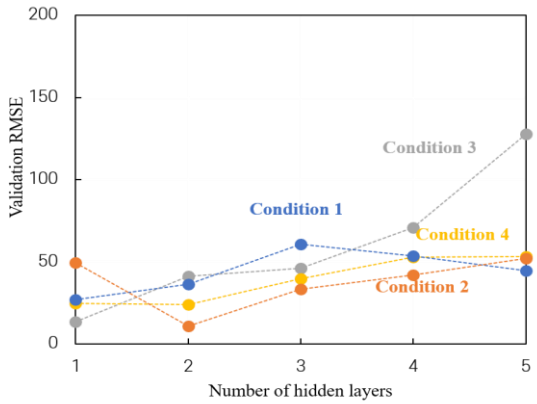
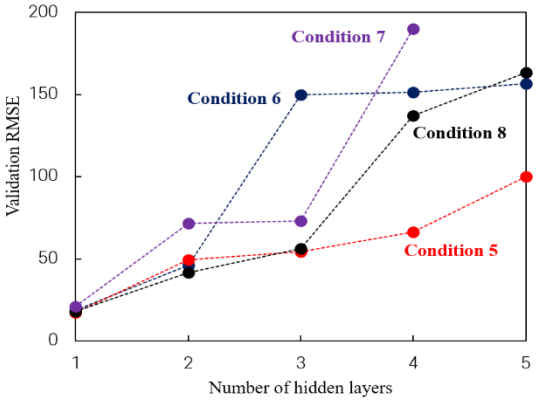



Fig. 5 Comparison of validation RMSEs according to the number of hidden nodes




(a) Conditions 1–4




(b) Conditions 5–8

Fig. 6 Comparison of validation RMSEs according to the number of hidden layers

indicated that the ReLU function provided the smallest validation RMSE; this is consistent with the results of previous studies showing that the ReLU function improves model performance by solving the vanishing gradient problem (He *et al*. 2015). Therefore, the ReLU activation function and He normal weight initialization method were used in the subsequent parametric analyses.

### *3.1.3 Numbers of hidden nodes and hidden layers*

Next, the change in the validation RMSE was evaluated according to the number of hidden nodes and hidden layers. When training an ANN using data from suction caisson foundations, which exhibit relatively independent and simple characteristics compared to the complex data used in other recent ANN applications, a quantitative analysis of overfitting should be attempted. Fig. 5 shows the change in the validation RMSE according to the number of hidden nodes under the conditions listed in Table 4. Under all four conditions, when the number of hidden nodes was small, the validation RMSE was greater than 100; as the number of hidden nodes increased, the validation RMSE decreased suddenly before slowly increasing again. Conditions 3 and 4, which included two hidden layers, exhibited their minimum validation RMSE values at smaller numbers of hidden nodes than Conditions 1 and 2, which included one hidden layer.

Fig. 6 shows the change in validation RMSE according to the number of hidden layers under the conditions listed in Table 5. The validation RMSEs tended to increase with the number of hidden layers in seven of the eight conditions (all except Condition 1). Condition 1 did not fall into a local minimum as it had a higher learning rate than Condition 2, and the number of network parameters for the corresponding model did not become excessively large as training was conducted using the smallest number of hidden nodes among the eight conditions. Therefore, even if three or more hidden layers are used, the validation RMSE will decrease. Notably, the learning rate comprises the only difference between Conditions 1 and 2; thus, as the number of hidden layers increases at a low learning rate, the probability of falling into a local minimum in the loss function increases.

### *3.1.4 Learning rate*

Fig. 7 shows the change in the validation RMSE according to the learning rate under the conditions listed in Table 6. The validation RMSEs for Conditions 1 and 2, in which each possessed only one hidden layer, changed little with the learning rate. However, the validation RMSEs for Conditions 3 and 4, in which each possessed two hidden

Table 5 Conditions for parametric analysis of the number of hidden layers

| Condition | Number of hidden layers | Learning rate | Optimizer | Activation function |
|---|---|---|---|---|
| Condition 1 | 256 | 0.05 | Adam | ReLU |
| Condition 2 | 256 | 0.005 | | |
| Condition 3 | 512 | 0.05 | | |
| Condition 4 | 512 | 0.005 | | |
| Condition 5 | 1024 | 0.05 | | |
| Condition 6 | 1024 | 0.005 | | |
| Condition 7 | 2048 | 0.05 | | |
| Condition 8 | 2048 | 0.005 | | |

Table 6 Conditions for parametric analysis of learning rate

| Condition | Number of hidden layers | Number of hidden nodes | Optimizer | Activation function |
|---|---|---|---|---|
| Condition 1 | 1 | 1024 | Adam | ReLU |
| Condition 2 | 1 | 2048 | | |
| Condition 3 | 2 | 1024 | | |
| Condition 4 | 2 | 2048 | | |

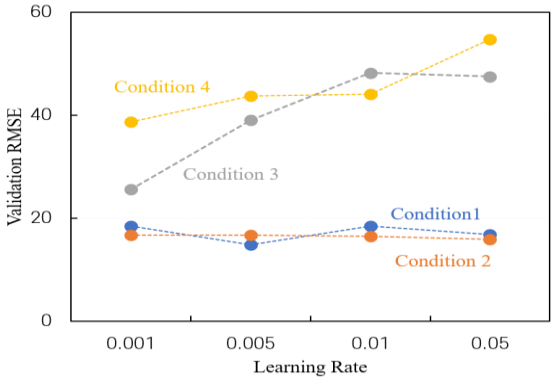



Fig. 7 Comparison of validation RMSEs according to the learning rate

layers, increased with the learning rate owing to the influence of the network parameters, which will be described later. In fact, the validation RMSE likely increased with the learning rate as the number of network parameters in Conditions 3 and 4 exceeded the optimal number of network parameters.

#### 3.1.5 Number of network parameters

An examination of the change in the validation RMSEs according to the number of nodes in each hidden layer and the number of hidden layers demonstrated that even when the former was increased significantly, the RMSE changed relatively less than when the latter was increased. In addition, the numbers of hidden nodes and hidden layers determine the number of network parameters in the model, hindering attempts to clearly observe overfitting by analyzing the two hyperparameters separately. Therefore, we examined the validation RMSE when varying the

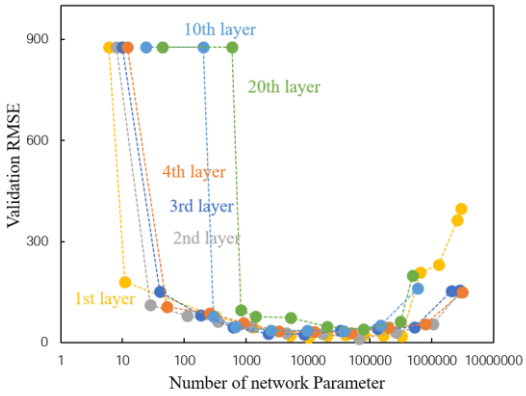



Fig. 8 Comparison of validation RMSEs according to the number of network parameters

number of network parameters rather than the number of hidden nodes and hidden layers. By inserting the number of hidden layers into Eq. (2), the number of hidden nodes can be adjusted to create a model with the desired number of network parameters.

Fig. 8 shows the change in the validation RMSE according to the number of network parameters in models with 1, 2, 3, 4, 10, and 20 hidden layers. Under all six conditions, the maximum validation RMSE was 878.39 with the smallest number of network parameters when there was one hidden node in each layer. In addition, as the number of network parameters increased, the values of the validation RMSEs decreased rapidly before relatively stabilizing, then eventually began increasing once the number of network parameters was excessive. When the number of hidden layers ranged from 1 to 4, the validation RMSE appeared to maintain a value of less than 50 given at least 500 network parameters. When the number of hidden layers was 10 or 20, the validation RMSE remained at 878.39 for up to 500 or 1000 network parameters, respectively, indicating that learning did not occur at all. Thus, when the hidden layers of an ANN are deeply stacked, the model becomes sensitive to underfitting given a small number of network parameters. When the number of network parameters was between 500 and 650,000, the difference in the validation RMSE values according to the number of hidden layers was negligible.

### 3.2 Optimization results

The optimization process is discussed in terms of the results of Bayesian optimization and the results of iterative learning the four selected hyperparameter conditions, evaluated according to the values of their double-validation RMSEs.

#### 3.2.1 Bayesian optimization result

Fig. 9 shows the validation RMSE values according to the number of network parameters and learning rate. Bayesian optimization has the advantage of being able to efficiently search for optimal values even with a limited number of iterations. In contrast, Grid Search or Random Search often rely on the researcher's intuition or luck, take a

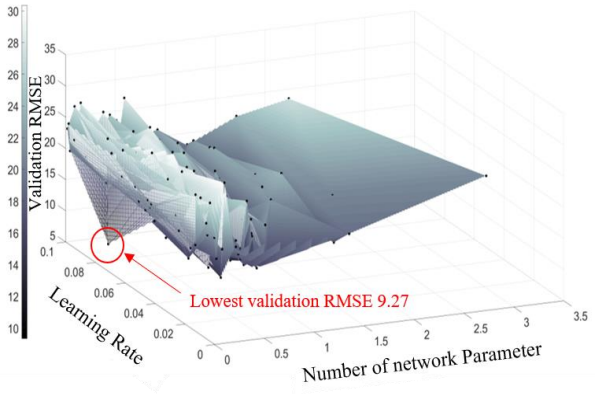



Fig. 9 Network parameter-learning rate-RMSE graph through Bayesian optimization

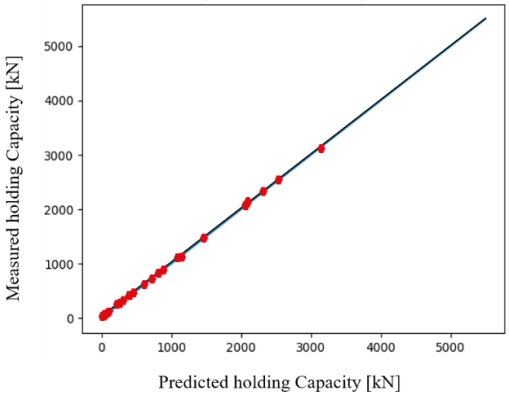



Fig. 10 Comparison of predicted and measured holding capacity

long time, and allow for the potential intervention of human bias. Considering the size of the dataset in this study, Bayesian optimization allowed us to find the optimal conditions (57/121 hidden nodes, learning rate 0.07191) much faster and more stably, confirming that this is an effective method in terms of both computational efficiency and accuracy.When using 57 hidden nodes in the first layer, 121 hidden nodes in the second layer, and a learning rate of 0.07191, the validation RMSE exhibited a minimum value of 9.27. As the number of network parameters increased, the number of searches decreased and the validation RMSE values increased, presumably because the probabilistic estimation characteristics of the Bayesian optimization technique determined that there was a low probability of obtaining optimal performance given a large number of network parameters. Furthermore, the lowest validation RMSE value was achieved with 7,638 network parameters, which is within the optimal range of network parameters obtained in the parameter analysis presented in Section 3.1. In this study, an initial random sample of 5 was used, followed by 1000 iterations for exploration. Even when the number of exploration iterations was varied, similar optimal conditions were derived, and the combination of 57/121 nodes and a learning rate of 0.07191 was obtained through repeated training. However, it is fundamentally difficult to achieve perfectly identical results because the initial weights are randomly assigned when training the model, and this random assignment is known to ensure better model performance. Nevertheless, by restricting the hyperparameter range within the optimal network parameter range suggested in this study, it is possible to derive the optimal value faster than before.

According to the results in Fig. 9, training was not effective when the number of parameters was less than 500, and overfitting occurred when it exceeded 650,000. The RMSE was stably minimized at approximately 7,638 parameters, which is an appropriate result for the size (127 samples) and characteristics of this dataset. Therefore, this value is optimized for a specific dataset, and the significance of this study lies in its quantitative presentation of a reasonable network complexity relative to the data scale.

Table 7 Result of iterative learning

| Number of hidden nodes in the first layer | Number of hidden nodes in the second layer | Learning rate | RMSE | |
|---|---|---|---|---|
| | | | Bayesian optimization | Iterative learning |
| 45 | 125 | 0.004333 | 12.48 | 10.83 |
| 59 | 77 | 0.03165 | 13.67 | 11.57 |
| 331 | 328 | 0.03237 | 11.96 | 10.91 |
| 57 | 121 | 0.07191 | 9.27 | 8.88 |

Table 8 Optimal hyperparameter conditions

| Hyperparameter | Optimal value |
|---|---|
| Number of hidden layers | Two |
| Number of hidden nodes | 57 in the first layer, 121 in the second layer |
| Learning rate | 0.07191 |
| Activation function | ReLU |
| Weight initialization method | He normal |
| Optimizer | Adam |
| Epochs | Model checkpoint, Early stopping |

### *3.2.2 Iterative learning result*

Table 7 lists the four conditions with the lowest Bayesian optimization RMSEs along with their RMSEs after 500 rounds of iterative learning. The condition employing 57 hidden nodes in the first layer, 121 hidden nodes in the second layer, and a learning rate of 0.07191 exhibited the lowest validation RMSE value regardless of whether learning was accomplished using Bayesian optimization only or with iterative learning. Comparing the remaining conditions, even when the RMSE value for Bayesian optimization was low, it was consistently higher than after iterative learning; therefore, iterative learning was necessary to clearly identify the optimal performance under each condition.

Fig. 10 compares the predicted and actual suction caisson holding capacities obtained from the validation data using the model with 57 first-layer hidden nodes, 121 second-layer hidden nodes, a learning rate of 0.07191, and

both Bayesian optimization and iterative learning. The RMSE was 8.88 with a coefficient of determination of 0.9998, indicating that the holding capacity was predicted with a small error.

Fig. 10 compares the predicted and actual suction caisson holding capacities obtained from the validation data using the model with 57 first-layer hidden nodes, 121 second-layer hidden nodes, a learning rate of 0.07191, and both Bayesian optimization and iterative learning. The RMSE was 8.88 with a coefficient of determination of 0.9998, indicating that the holding capacity was predicted with a small error.

Table 8 presents the optimal hyperparameter combinations for the suction caisson foundation capacity prediction model developed in this study. The activation function, weight initialization, and optimizer hyperparameters were determined through parametric analysis. Bayesian optimization techniques were employed to derive the remaining optimal hyperparameters: 57 hidden nodes in the first layer, 121 hidden nodes in the second layer, and a learning rate of 0.07191.

## 5. Conclusions

Suction caisson foundations are quite large and typically applied in the deep sea, hindering experimental evaluations of their performance. As a result, the holding capacity of a suction caisson is typically evaluated using a three-dimensional finite element analysis. Although suction caisson foundation data are acquired consistently, the unique characteristics of these geotechnical data must be considered independently for each case. Therefore, this study developed and optimized an ANN model to predict the holding capacity of suction caissons using 127 suction caisson data points by conducting a parametric analysis considering overfitting, applying an optimization process, and accounting for random initial weights. By using the model checkpoint and early stopping techniques together, the weights were obtained immediately prior to overfitting, confirming that effective overfitting prevention is possible by simultaneously utilizing the two techniques. The use of the Adam optimizer, ReLU activation function, and He normal weight initialization method exhibited the lowest validation RMSE; the optimal number of network parameters corresponding to these optimal hyperparameters was 7,638, which was within the range of 500–650,000 anticipated by the parametric analysis. Finally, applying iterative learning under various conditions, despite one condition indicating a lower RMSE than the others, the RMSE obtained for all conditions remained lower than for those obtained by Bayesian optimization alone, indicating that the optimal model can be obtained by accounting for the random distribution of initial weights. The validation RMSE of the optimized ANN model was 8.88 with a coefficient of determination of 0.9998.

The proposed modeling method was applied to determine the optimal hyperparameters for predicting the holding capacity of a suction caisson based on training and validation with 127 data points. Notably, a change in performance owing to overfitting was observed depending on the number of network parameters used in the trained ANN model. Based on this observation, the optimal number of network parameters required to suppress overfitting was quantitatively determined. But, the internal ANN mechanism has not been analyzed; therefore, future research could apply SHAP, LIME, or similar explainable artificial intelligence (XAI) techniques to enhance the model's interpretability. These results can be potentially applied using data from a wide variety of applications in the geotechnical engineering field. However, the generalized use of the proposed ANN model requires extensive validation; therefore, its effectiveness remains to be demonstrated through further research using additional data.

Since this ANN model was trained based on the finite element analysis data from Kim *et al*. (2017), it is currently appropriate to view it as an auxiliary tool that quickly and accurately predicts the results of that numerical analysis. for integration into design standards such as DNV, the following are required: (1) additional verification across diverse numerical analysis conditions, (2) comparison with field measurement results, (3) standardized reporting of model performance, and (4) verification procedures by professional societies or certification bodies. This study represents the initial phase, demonstrating the potential for the ANN optimization technique to be incorporated into design guidelines.Although separate calibration reflecting design standards and field conditions is necessary for actual design application, the ANN optimization technique presented in this paper can significantly enhance efficiency during the design review process utilizing various analysis results.

## Acknowledgments

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (2020R1A2C1012352 / 2022R1A4A3029737), and by the Glocal30 project (2024) of Chungbuk National University.

## References

Abu Kiefa, M.A. (1998), “General regression neural networks for driven piles in cohesionless soils”, *J. Geotech. Environ. Eng.*, **124**(12), 1177-1185. https://doi.org/10.1061/(ASCE)1090-0241(1998)124:12(1177).

Ahn, J., Lee, H. and Kim, Y.T. (2014), “Holding capacity of suction caisson anchors embedded in cohesive soils based on finite element analysis”, *Int. J. Numer. Anal. Method. Geomech.*, **38**(15), 1541-1555. https://doi.org/10.1002/nag.2268.

Al-Swaidani, A.M., Meziab, A., Khwies, W.T., Al-Bali, M. and Lala, T. (2024), “Building MLR, ANN and FL models to predict the strength of problematic clayey soil stabilized with a combination of nano lime and nano pozzolan of natural sources for pavement construction”, *Int. J. Geo-Eng.*, **15**(1), 2. https://doi.org/10.1186/s40703-023-00201-1.

Alkroosh, I.S., Bahadori, M., Nikraz, H. and Bahadori, A. (2015), “Regressive approach for predicting bearing capacity of bored piles from cone penetration test data”, *J. Rock Mech. Geotech. Eng.*, **7**(5), 584-592. https://doi.org/10.1016/j.jrmge.2015.06.011.

Alzo'ubi, A.K. and Ibrahim, F. (2019), "Predicting loading–unloading pile static load test curves by using artificial neural networks", *Geotech. Geol. Eng.*, **37**, 1311-1330. https://doi.org/10.1007/s10706-018-0687-4.

Andersen, K.H., Dyvik, R., Schrøder, K., Hansteen, O.E. and Bysveen, S. (1993), "Field tests of anchors in clay II: Predictions and interpretation", *J. Geotech. Eng.*, **119**(10), 1532-1549. https://doi.org/10.1061/(ASCE)0733-9410(1993)119:10(1532).

Ardalan, H., Eslami, A. and Nariman-Zadeh, N. (2009), "Piles shaft capacity from CPT and CPTu data by polynomial neural networks and genetic algorithms", *Comput. Geotech.*, **36**(4), 616-625. https://doi.org/10.1016/j.compgeo.2008.09.004.

Armaghani, D.J., Asteris, P.G., Fatemi, S.A., Hasanipanah, M., Tarinejad, R., Rashid, A.S.A. and Huynh, V.V. (2020), "On the use of neuro-swarm system to forecast the pile settlement", *Appl. Sci.*, **10**(6), 1904. https://doi.org/10.3390/app10061904.

Aubeny, C.P., Han, S.W. and Murff, J.D. (2003), "Inclined load capacity of suction caissons". *Int. J. Numer. Anal. Method. Geomech.*, **27**(14), 1235-1254. https://doi.org/10.1002/nag.319.

Bae, B.R. (2002), "Understanding and utilizing structural equation mModels", Daegyeong

Bajaj, P., Yadu, L. and Chouksey, S.K. (2019), "Study on vertical and batter piles subjected to lateral loads in different non-cohesive sub-soil conditions", *Int. J. Geotech. Eng.*, **14**(6), 603-613. https://doi.org/10.1080/19386362.2018.1564181.

Benali, A., Hachama, M., Bounif, A., Nechnech, A. and Karray, M. (2021), "A TLBO-optimized artificial neural network for modeling axial capacity of pile foundations", *Eng. with Comput.*, **37**(1), 675-684. https://doi.org/10.1007/s00366-019-00847-5.

Borthakur, N. and Dey, A.K. (2020), "Evaluation of group capacity of micropile in soft clayey soil from experimental analysis using SVM-based prediction model", *Int. J. Geomech.*, **20**(3), 04020008. https://doi.org/10.1061/(ASCE)GM.1943-5622.0001606.

Brown, G.A. and Nacci, V.A. (1971), "Performance of hydrostatic anchors in granular soils", Offshore Technology Conference, OnePetro. https://doi.org/10.4043/1472-MS.

Byrne, B.W. and Houlsby, G.T. (2004), "Foundations for offshore wind turbines", *Philos. T. R. Soc. A*, **361**(1813), 2909-2930. https://doi.org/10.1098/rsta.2003.1286.

Cao, J., Audibert, J.M.E., Ak-Khafaji, Z., Phillips, R. and Popescu, R. (2002), "Numerical analysis of the behavior of suction caissons in clay", *Proceedings of the 12th International Offshore and Polar Engineering Conference*, Kitakyushu, Japan.

Chow, Y.K., Chan, W.T., Liu, L.F. and Lee, S.L. (1995), "Prediction of pile capacity from stress-wave measurements, A neural network approach", *Int. J. Numer. Anal. Method. Geomech.*, **19**(2), 107-126. https://doi.org/10.1002/nag.1610190204.

Clukey, E.C. and Morrison, M.J. (1993), "A centrifuge and analytical study to evaluate suction caissons for TLP applications in the Gulf of Mexico", Design and Performance of Deep Foundations: Piles and Piers in Soil and Soft Rock, ASCE, 141-156.

Das, M. and Dey, A.K. (2018), "Prediction of bearing capacity of stone columns placed in soft clay using ANN model", *Geotech. Geol. Eng.*, **36**, 1845-1861. https://doi.org/10.1007/s10706-017-0436-0.

Deng, W. (2001), "Study on the inclined uplift capacity of suction caissons", Ph.D. Thesis, University of Sydney, Darlington, Australia.

Det Norske Veritas (DNV) (2005), "Geotechnical design and installation of suction anchors in Cclay", DNV Recommended Practice RP-E303, Høvik.

Fuglsang, L.D. and Steensen-Bach, J.O. (1991), "Breakout resistance of suction piles in clay", *Proceedings of the International Conference: Centrifuge*, **91**, 153-159.

Ghorbani, B., Sadrossadat, E., Bazaz, J.B. and Oskooei, P.R. (2018), "Numerical ANFIS-based formulation for prediction of the ultimate axial load bearing capacity of piles through CPT data", *Geotech. Geol. Eng.*, **36**, 2057-2076. https://doi.org/10.1007/s10706-018-0445-7.

Goodman, L.J., Lee, C.N. and Walker, F.J. (1961), "The feasibility of vacuum anchorage in soil", *Geotechnique*, 1**1**(4), 356-359. https://doi.org/10.1680/geot.1961.11.4.356.

Harandizadeh, H. (2020), "Developing a new hybrid soft computing technique in predicting ultimate pile bearing capacity using cone penetration test data", *Artif. Intel. Eng. Des. Anal. Manufact.*, **34**(1), 114-126. https://doi.org/10.1017/S0890060420000025.

Harandizadeh, H., Jahed Armaghani, D. and Khari, M. (2021), "A new development of ANFIS–GMDH optimized by PSO to predict pile bearing capacity based on experimental datasets", *Eng. with Comput.*, **37**, 685-700. https://doi.org/10.1007/s00366-019-00849-3.

He, K., Zhang, X., Ren, S. and Sun, J. (2015), "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification", *Proceedings of the IEEE International Conference on Computer Vision*, 1026-1034. https://doi.org/10.1109/ICCV.2015.123.

Houlsby, G.T., Ibsen, L.B. and Byrne, B.W. (2005), "Suction caissons for wind turbines", *Proceedings of the 1st International Symposium on Frontiers in Offshore Geotechnics (ISFOG)*, Perth, Australia.

House, A.R. and Randolph, M.F. (2001), "Installation and pull-out capacity of stiffened suction caissons in cohesive sediments", *Proceedings of the 11th International Offshore and Engineering Conference*, Stavanger, Norway, ISOPE-I-01-212.

Ismail, A., Jeng, D.S. and Zhang, L.L. (2013), "An optimised product-unit neural network with a novel PSO-BP hybrid training algorithm: Applications to load-deformation analysis of axially loaded piles", *Eng. Appl. Artif. Intel.*, **26**(10), 2305-2314. https://doi.org/10.1016/j.engappai.2013.04.007.

Jebur, A.A., Atherton, W., Al Khaddar, R. and Loffill, E. (2021), "Artificial neural network (ANN) approach for modelling of pile settlement of open-ended steel piles subjected to compression load", *Eur. J. Environ Civil Eng.*, **25**(3), 429-451. https://doi.org/10.1080/19648189.2018.1531269.

Kardani, N., Zhou, A., Nazem, M. and Shen, S.L. (2020), "Estimation of bearing capacity of piles in cohesionless soil using optimised machine learning approaches", *Geotech. Geol. Eng.*, **38**(9), 2271-2291. https://doi.org/10.1007/s10706-019-01085-8.

Kelly, R.B., Houlsby, G.T. and Byrne, B.W. (2006), "A comparison of field and laboratory tests of caisson foundations in sand and clay.", *Geotechnique*, **56**(9), 617-626. https://doi.org/10.1680/geot.2006.56.9.617.

Kim, Y., Ahn, J. and Jung, J. (2017), "Fuzzy modeling of holding capacity of offshore suction caisson anchors", *Int. J. Numer. Anal. Method. Geomech.*, **41**(7), 1038-1054. https://doi.org/10.1002/nag.2664.

Liu, Y.J., Liang, S.H., Wu, J.W. and Fu, N. (2010), "Prediction method of vertical ultimate bearing capacity of single pile based on support vector machine", *Adv. Mater. Res.*, **168**, 2278-2282. https://doi.org/10.4028/www.scientific.net/AMR.168-170.2278.

Moayedi, H. and Armaghani, D.J. (2018), "Optimizing an ANN model with ICA for estimating bearing capacity of driven pile in cohesionless soil", *Eng. with Comput.*, **34**, 347-356. https://doi.org/10.1007/s00366-017-0545-7.

Momeni, E., Nazir, R., Armaghani, D.J. and Maizir, H. (2014), "Prediction of pile bearing capacity using a hybrid genetic

algorithm-based ANN", *Measurement*, **57**, 122-131. https://doi.org/10.1016/j.measurement.2014.08.007.

Momeni, E., Nazir, R., Armaghani, D.J. and Maizir, H. (2015), "Application of artificial neural network for predicting shaft and tip resistances of concrete piles", *Earth Sci. Res. J.*, **19**(1), 85-93. https://doi.org/10.15446/esrj.v19n1.38712.

Murillo, C.A., Thorel, L. and Caicedo, B. (2009), "Spectral analysis of surface waves method to assess shear wave velocity within centrifuge models", *J. Appl. Geophys.*, **68**(2), 135-145. https://doi.org/10.1016/j.jappgeo.2008.10.007.

Park, H.I. and Cho, C.W. (2010), "Neural network model for predicting the resistance of driven piles", *Mar. Georesour. Geotechnol.*, **28**(4), 324-344. https://doi.org/10.1080/1064119X.2010.514232.

Pham, T.A., Ly, H.B., Tran, V.Q., Giap, L.V., Vu, H.L.T. and Duong, H.A.T. (2020), "Prediction of pile axial bearing capacity using artificial neural network and random forest", *Appl. Sci.*, **10**(5), 1871. https://doi.org/10.3390/app10051871.

Prayogo, D. and Susanto, Y.T.T. (2018), "Optimizing the prediction accuracy of friction capacity of driven piles in cohesive soil using a novel self-tuning least squares support vector machine", *Adv. Civil Eng.*, 6490169. https://doi.org/10.1155/2018/6490169.

Samui, P. (2011), "Prediction of pile bearing capacity using support vector machine", *Int. J. Geotech. Eng.*, **5**(1), 95-102. https://doi.org/10.3328/IJGE.2011.05.01.95-102.

Samui, P. (2012), "Application of relevance vector machine for prediction of ultimate capacity of driven piles in cohesionless soils", *Geotech. Geol. Eng.*, **30**(5), 1261-1270. https://doi.org/10.1007/s10706-012-9539-9.

Samui, P. (2019), "Determination of friction capacity of driven pile in clay using Gaussian Process Regression (GPR) and Minimax Probability Machine Regression (MPMR)", *Geotech. Geol. Eng.*, **37**(3), 4643-4647. https://doi.org/10.1007/s10706-019-00928-8.

Savvides, A.A. and Papadopoulos, L. (2024), "A neural network approach for the reliability analysis on failure of shallow foundations on cohesive soils", *Int. J. Geo-Eng.*, **15**(1), 15. https://doi.org/10.1186/s40703-024-00217-1.

Shahin, M.A. (2010), "Intelligent computing for modeling axial capacity of pile foundations", *Can. Geotech. J.*, **47**(2), 230-243. https://doi.org/10.1139/T09-094.

Shahin, M.A. (2014a), "Load-settlement modeling of axially loaded steel driven piles using CPT-based recurrent neural networks", *Soils Found.*, **54**(3), 515–522. https://doi.org/10.1016/j.sandf.2014.04.015.

Shahin, M.A. (2014b), "Use of evolutionary computing for modelling some complex problems in geotechnical engineering", *Geomech. Geoeng.*, **10**(2), 109-125. https://doi.org/10.1080/17486025.2014.921333.

Shaik, S., Krishna, K.S.R., Abbas, M., Ahmed, M. and Malvaluru, D. (2019), "Applying several soft computing techniques for prediction of bearing capacity of driven piles", *Eng. with Comput.*, **35**(4), 1463-1474. https://doi.org/10.1007/s00366-018-0674-7.

Shatnawi, A., Bodour, W.A.L., Abdel-Jaber, M. and Tarawneh, B. (2019), "Empirical formulas to predict the axial capacity of driven piles using in-situ dynamic load testing data", *Int. J. Mach. Learn. Comput.*, **9**(2), 129-134. https://doi.org/10.18178/ijmlc.2019.9.2.776.

Singh, G. and Walia, B.S. (2017), "Performance evaluation of nature-inspired algorithms for the design of bored pile foundation by artificial neural networks", *Neural Comput. Appl.*, **28**, 289-298. https://doi.org/10.1007/s00521-016-2345-1.

Singh, T., Pal, M. and Arora, V.K. (2019), "Modeling oblique load carrying capacity of batter pile groups using neural network, random forest regression and M5 model tree", *Front. Structu. Civil Eng.*, **13**, 674-685. https://doi.org/10.1007/s11709-018-0505-3.

Sun, G., Hasanipanah, M., Amnieh, H.B. and Foong, L.K. (2020), "Feasibility of indirect measurement of bearing capacity of driven piles based on a computational intelligence technique", *Measurement*, **156**(1), 107577. https://doi.org/10.1016/j.measurement.2020.107577.

Tarawneh, B. and Imam, R. (2014), "Regression versus artificial neural networks: Predicting pile setup from empirical data", *KSCE J. Civil Eng.*, **18**, 1018-1027. https://doi.org/10.1007/s12205-014-0072-7.

Tariq, A., Abualshar, B., Deliktas, B., Song, C.R., Al-Nimri, B., Barret, B., Silvey, A. and Glennie N. (2024), "ANN-based evaluation system for erosion resistant highway shoulder rocks", *Int. J. Geo-Eng.*, **15**(1), 17. https://doi.org/10.1186/s40703-024-00216-2.

The, C.I., Wong, K.S., Goh, A.T.C. and Jaritngam, S. (1997), "Prediction of pile capacity using neural networks", *J. Comput. Civil Eng.*, **11**(2), 129-138. https://doi.org/10.1061/(ASCE)0887-3801(1997)11:2(129).

Watson, P.G. and Randolph, M.F. (1997), "Vertical capacity of caisson foundations in calcareous sediments", *Proceedings of the 7th International Offshore and Polar Engineering Conference*, Honolulu, HI, ISOPE-I-97-121.

Watson, P.G., Randolph, M.F. and Bransby, M.F. (2000), "Combined lateral and vertical loading of caisson foundations", *Proceedings of the Offshore Technology Conference*, Houston, TX, OTC-12195-MS.

Yong, W., Zhou, J., Armaghani, D.J., Tahir, M.M., Tarinejad, R., Pham, B.T. and Van, H.V. (2021), "A new hybrid simulated annealing-based genetic programming technique to predict the ultimate bearing capacity of piles", *Eng. with Comput.*, **37**(4), 2111-2127. https://doi.org/10.1007/s00366-019-00932-9.

Zhan, Y.G. and Liu, F.C. (2010), "Numerical analysis of bearing capacity of suction bucket foundation for offshore wind turbines", *Electron. J. Geotech. Eng.*, **15**(10), 633-644.

*IC*