Research Article

Application of Ensemble Learning Using Weight Voting Protocol in the Prediction of Pile Bearing Capacity

Tuan Anh Pham and Huong-Lan Thi Vu

University of Transport Technology, Hanoi 100000, Vietnam

Correspondence should be addressed to Tuan Anh Pham; anhpt@utt.edu.vn

Received 25 February 2021; Revised 30 May 2021; Accepted 19 June 2021; Published 8 July 2021

Copyright © 2021 Tuan Anh Pham and Huong-Lan Thi Vu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate prediction of pile bearing capacity is an important part of foundation engineering. Notably, the determination of pile bearing capacity through an in situ load test is costly and time-consuming. Therefore, this study focused on developing a machine learning algorithm, namely, Ensemble Learning (EL), using weight voting protocol of three base machine learning algorithms, gradient boosting (GB), random forest (RF), and classic linear regression (LR), to predict the bearing capacity of the pile. Data includes 108 pile load tests under different conditions used for model training and testing. Performance evaluation indicators such as $R^2$, root mean square error (RMSE), and MAE (mean absolute error) were used to evaluate the performance of models showing the efficiency of predicting pile bearing capacity with outstanding performance compared to other models. The results also showed that the EL model with a weight combination of $w_1 = 0.482$, $w_2 = 0.338$, and $w_3 = 0.18$ corresponding to the models GB, RF, and LR gave the best performance and achieved the best balance on all datasets. In addition, the global sensitivity analysis technique was used to detect the most important input features in determining the bearing capacity of the pile. This study provides an effective tool to predict pile load capacity with expert performance.

1. Introduction

The cost for the foundation of the work usually accounts for a very large proportion of the total construction cost [1]. Therefore, choosing the right foundation structure solution, as well as determining the load capacity of the foundation, also plays an important role in reducing construction costs [2]. Among the commonly used foundation solutions today, the pile foundation is one of the most popular deep foundations for its advantages. In particular, the pile foundation is capable of transmitting into the deep soil layers [3, 4]. Determining the exact bearing capacity of the pile will help determine the correct foundation size and depth of the pile, thereby choosing the most appropriate foundation solution. To determine the bearing capacity of the pile, people can use different methods [5–10], for example, dynamic load testing (PDA) and static pile load test. In addition, many traditional formulas had been proposed, which were mainly based on the in situ testing result of soil such as cone penetration test (CPT) and standard penetration test (SPT) [5–9]. In some other studies, the authors also used the finite element method to analyze the pile load and pile displacement relationship of [10–12].

In the specific conditions, the above methods all prove to have many advantages, but, besides that, there are still many problems that need to be carefully considered before being widely applied in construction practice. Specifically, in their study, Jesswein et al. [13] showed that the calculation of pile load capacity according to SPT is not completely reliable because it still lacks precision, although it is cheap and easy to perform. Similarly, the analytical method is considered impractical, since the theoretical approach is based on too many assumptions and simplifications [3]. In another argument by Abu-Farsakh and Titi [14], it is assumed that the methods of piles empirical analysis or piles static analysis are expensive and have low accuracy due to the large selection of safety factors. On the other hand, for the pile load test method, although the percentage of reliability is high, it is
time-consuming and costly and the equipment is often very cumbersome [15]. The dynamic method is highly dependent on the characteristics of the pile, the hammer, and the position of the pile to predict the bearing capacity of the pile without considering soil effects [11, 16]. Last but not least, the characteristics of the numerical simulation method are finite-element-based; basically still approximate method and the results depend a lot on the modeling process [17].

In recent studies, many scientists have used a new approach to the problems related to the foundation of the building. That method is known as artificial intelligence (AI). Based on the development of computer science, AI has gradually proven its remarkable effectiveness in numerous different fields such as construction [18], transportation [19], security [20], and medicine [21]. The essence of AI algorithms is a combination of math, algorithm, and creativity. AI allows solving complex problems and has many unknowns, so it is suitable for nonsnow problems in the field of geotechnical engineering [3]. Scientists have used a lot of different AI algorithms to solve the problem of predicting the pile bearing capacity, which can be named as artificial neural network (ANN) [3, 22–25], deep neural network (DNN) [26–28], adaptive neurofuzzy inference system (ANFIS) [29, 30], and random forest (RF) [23]. The above studies all showed good predictive efficiency and were expected to become a highly generalizable tool in the prediction of pile load capacity. However, the published literature mainly focuses on a certain algorithm, making it difficult to confirm the optimum of the model for the data. Consequently, the usage model can lead to imbalances in bias and variance or overfitting and underfitting problem which is difficult to be fully solved due to the model’s algorithm [31, 32]. Therefore, in this study, the Ensemble Learning model [33–35] combining several base models according to a certain rule is expected to balance bias and variance, bringing outstanding efficiency in predicting the bearing capacity of piles. In addition, published studies are only interested in input parameters related to soil characteristics and pile geometry parameters but not other important factors such as pile type, method of pile testing, pile construction method, and pile tip type. These additional parameters which greatly influence the prediction performance were carefully considered during the modeling process in this study.

In this study, three different machine learning models were developed to independently predict the pile load capacity on the same data set, namely, gradient boosting (GB), random forest (RF), and classic linear regression (LR). Then, an Ensemble Learning (EL) algorithm was developed, using the voting protocol of base models to make the final prediction. A total of 108 data sets from the published literature were used to train and test the model; and since the data set is aggregated from projects around the world, the resulting model can expect the ability to predict the bearing capacity of different types of piles in different regions and can be widely beneficial internationally. The reminder of the paper is structured as follows: Section 2 presents the materials needed for the research, including an introduction to base machine learning methods such as RF, GB, and LR, as well as an algorithm to combine the underlying algorithms into a voting EL model. Section 3 presents the database used to develop the machine learning models, as well as the statistical information of the input and output variables. Section 4 shows the results analyzed by the machine learning models and compares those machine learning models with each other to see the outstanding performance of the EL model. Finally, Section 5 offers some conclusions and opinions.

2. Methods

2.1. Random Forest. The scientist Breiman [36] was the first to mention the random forest algorithm, also known as the group learning method. Later, this algorithm was exploited and developed by many other scientists; for example, in the research of Kuhn and Johnson [37], the authors used decision trees of the same distribution to establish forests and train and predict sample data. Decision trees are capable of solving classification and regression problems based on a supervised learning approach [23, 38]. The bootstraps technique (also known as the bagging method) is used for the RF algorithm, which has the effect of bringing together different base learners of different backgrounds and random construction. The biggest advantage of RF is that it is possible to build a set of basic learners while controlling variations. The RF algorithm is developed through the following steps [39]:

Step one: create random vectors. The random sample is selected from the input data set.

Step two: for each sample, a decision tree is built, and from there, choose 1 prediction result.

Step three: vote for each prediction results, and then choose the most voted prediction. The result is the final prediction.

The schematic diagram of the random forest model is shown in Figure 1.

2.2. Gradient Boosting. Boosting algorithm was first introduced in 1990 by the machine learning community. By repeating some simple patterns, weak learners will be improved to become good learners [40]. In the first phase, boosting is designed as a machine learning process to improve the prediction of binary results, which is done through base learners (using sets of decision tree weights) [41]. Boosting is then considered to reduce the slope in functional space and can be used to accommodate statistical regression models [42].

Gradient boosting (GB) has advantages over other statistical learning methods in that it can interpret results and simplify complex interactions, and it also requires preprocessing of data and adjustment of parameters number [43]. In addition, GB also helps increase the slope, improves model accuracy, and makes selection [44]. In this study, gradient boosting used the decision trees model of a fixed size as base learners and used boosting algorithm developed by Friedman [45, 46].
2.3. Linear Regression. Linear regression (LR) models have been widely used to solve problems in statistics [47–49]. In essence, linear regression tries to model the relationship between input and output variables by adjusting a linear equation. In this study, since there are 9 input variables and one output variable, the linear regression equation has the following form:

\[ y_i = f(x_{i1}, x_{i2}, \ldots, x_{i9}) = \sum_{j=1}^{9} \beta_j \cdot x_{ji} + \beta_0, \quad (1) \]

where \( \beta_j \) is the weight of \( j \)th input variables and \( \beta_0 \) is the bias (or intercept).

The least-squares method [50, 51] was used to find the values of the optimal weights by minimizing the sum \( S \) of squared residuals:

\[ S = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2 \longrightarrow \text{min}. \quad (2) \]

2.4. Ensemble Learning. Ensemble Learning (EL) is a machine learning technique that combined several base models to create one optimal predictive model. The combining models were expected to provide better predictive performance than member models and used to solve out any problem in technique [52, 53]. The random forest and gradient boosting models were bagging Ensemble Learning themselves based on multiple decision trees to make final predictions. Therefore, a combined model of random forest, gradient boosting, and classical linear regression based on average voting protocol was expected to give the best prediction of pile load capacity. There are two ways to use the voting protocol in the EL method, namely, average voting (AV) and weight voting (WV). In AV-EL, the base model’s weights are the same and are equal to 1, while in WV-EL the weights are in the \([0, 1]\) range and the sum of all weights is equal to 1. The flow chart of the EL model used in this study is presented in Figure 2, in which weight voting of three base models was used to make the final prediction.

The final prediction equation has the following form:

\[ y_i = \sum_{j=1}^{n} w_j \cdot P_{ji}, \quad (3) \]

where \( w_j \) is the weight of the \( j \)th learner; \( P_{ji} \) is the prediction of the \( j \)th learner; and \( n \) is the number of learners in the EL model.

In this study, the optimum weight of WV-EL was determined by random search (RS) technique. In this technique, the weights are randomly selected within the range \([0, 1]\) which allows the sum of all weights to be 1. The performances of the WV-EL models using different combinations are compared, and then the model that gives the best performance is selected as the final EL model.

The pseudocode of the RS technique used in this study is written as follows:
FOR each j learner in the EL model

INITIALIZE weight vector \( w[j] = 1 \)

END FOR

FIT the EL model with weight vector \( w \)

CALCULATE the initialize fitness value \( F_b \) of EL model with 5-fold CV technique

Iteration \( k = 1 \)

WHILE \( k < \) maximum Iteration

FOR each learner \( j \) in the EL model

GENERATE \( s[j] \) in random \([0, 1]\) range

END FOR

FOR each learner \( j \) in the EL model

CALCULATE weight vector \( w \) by normalized \( s \) vector as follow: \( w[j] = (s[j]/\sum_{j=1}^{n}s[j]) \)

FIT EL model with weight vector \( w \)

CALCULATE the fitness value of the EL model with 5-fold CV technique.

IF fitness value > \( F_b \) THEN

\( F_b = \) fitness value

END IF

\( k = k + 1 \)

END WHILE

2.5. Performance Evaluation. To evaluate the consistency between the predicted results and the actual results, three performance indicators were used: RMSE (root mean square error) [54, 55], \( R^2 \) (coefficient of determination) [56, 57], and MAE (mean absolute error) [58, 59]. \( R^2 \) allows determining the statistical relationship between the predicted values with the experimental results. \( R^2 \) ranged from \([-\infty, 1]\), and the model will be said to be more accurate when \( R^2 \) approaches 1. Conversely, the lower the values of RMSE and MAE are, the more accurate the calculation result is. The values of \( R^2 \), RMSE, and MAE were calculated through the following formulas:

\[
RMSE = \sqrt{\frac{1}{k} \sum_{i=1}^{k} (y_i - \bar{y}_i)^2},
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{k} (y_i - \bar{y}_i)^2}{\sum_{i=1}^{k} (y_i - \bar{y})^2},
\]

(4)

\[
MAE = \frac{1}{k} \sum_{i=1}^{k} (y_i - \bar{y}_i),
\]

where \( k \) denotes the number of samples, \( y_i \) and \( \bar{y}_i \) are the actual and predicted outputs, respectively, and \( \bar{y} \) is the average value of \( y_i \).

3. Data Used

In this study, a database of 108 pile load test reports was collected from published literature [30]. All input parameters that might affect the determination of the pile load were considered, including the type of test \( (T) \), the type of pile \( (P) \), the installation method (denoted as \( I \)), end of pile type \( (EP) \), the pile tip cross-sectional area \( (A_t) \), and the shaft area \( (A_f) \). The soil properties were shown through parameters obtained from cone penetration test (CPT) results, including the average cone tip resistance along the embedded length of the pile \( (q_{tc}) \), the average cone tip resistance over influence zone \( (q_{d}) \), and the average sleeve friction along the embedded length of the pile \( (f_{sa}) \). The considered output is the ultimate bearing capacity of the pile (denoted as \( P_u \)).

The summary of the database statistics is presented in Table 1, which includes the min, mean, max, median, and standard deviation (SD) values of all parameters used in this study. Besides, the histogram charts of all the input and output variables were illustrated in Figure 3.

4. Result and Discussion

4.1. Effect of Training Data Size. Since the size of the training set affected the accuracy of the model, the size of the training data set will be carefully considered [60]. In this section, the
suitable training part size was found by changing it from 20% to 80%; meanwhile the size of the testing set decreased from 80% to 20% of the original data set. To prevent data leakage, the testing set was hidden from the training process and the performance of the model was based on the 5-fold CV technique [61, 62]. The random forest model was selected for training and verification during this period. It is important to note that 300 samplings were performed in shuffling the database indexes to make the results more objective. The initialization of the RF model is shown in Table 2.

![Histogram chart of the features used in this study.](image)

Figure 3: Histogram chart of the features used in this study.

| Unit | T (cm²) | P (m²) | I (MPa) | Ap (kN) | qca (kN) | fua (MPa) | qu (kN) | Pu (kN) |
|------|---------|--------|---------|---------|---------|-----------|---------|--------|
| Min  | 1       | 1      | 1       | 1       | 20      | 5.45      | 0.833   | 9.39   |
| Median| 1.98    | 1.69   | 1.84    | 1.15    | 1736.02 | 26.46     | 5.84    | 101.89 |
| Max  | 2       | 3      | 2       | 2       | 7854    | 194.65    | 24.707  | 349.64 |
| SD   | 0.14    | 0.94   | 0.37    | 0.36    | 1674.09 | 26.35     | 4.23    | 66.29  |

SD: standard deviation; T = 1, constant rate of penetration; 2, maintained load; P = 1, concrete pile; 2, composite pile; 3, steel pile; I = 1, driven pile; 2, bored pile; Ap = 1, closed pile; 2, ppen pile.

Table 1: Summary of the statistical information of the database used in the present study.
Table 2: Initialization parameters of the RF model.

| Parameters          | Random forest |
|---------------------|---------------|
| Number of estimators | 15            |
| Learning rate       | —             |
| Min samples leaf    | 1             |
| Min samples split   | 2             |
| Max depth           | 12            |
| Cost function       | MSE           |
| Data used           | Training/5-fold CV |
| Number of runs      | 300           |

The comparison of RE for the testing set is shown in Table 4. It can be seen that all three base models give a quite good predictive performance. However, no model outperformed the EL model for the entire testing data set. The predictive performance of one model may be satisfactory for some data samples but poor for the rest. On the other hand, the performance of the EL model seems to balance the predicted performance among the base models, leading to the best overall prediction result for the entire test data set.

Figure 5 presented a visual comparison of the ultimate bearing capacity of pile-based experimental and predicted results from ML models. The performance indicator results were summarized in Table 5. It can be seen that the GB and RF models seem to suffer from overfitting as the performance indicator for the training set is significantly better than that for the testing set. In contrast, the LR model suffered from underfitting when the performance for the testing set was better than that for the training set. At the same time, the combination model seems to have solved all of the overfitting and underfitting problems of the three base models and gave the best balance in performance on training and testing sets. To be more specific, the performance index values of the combined model were $R^2 = 0.928$, $MAE = 292.897$, and $RMSE = 407.041$ on the training set and $R^2 = 0.955$, $MAE = 263.49$, and $RMSE = 372.338$ on the testing set.

4.3. Improved Performance of EL Model Using Weight Voting Optimization. In the previous section, the performances of the base models were proved to be dissimilar. Models such as GB and RF have good predictive performance, while LR has lower predictive performance. Moreover, the AV-EL model seems to exhibit a slight imbalance between bias and variance, based on the fact that the performance on the testing set was still better than that on the training set. Therefore, to create even a better EL model, member models should play a different role in the final prediction. To do that, the RS technique was used to optimize the weights of the base models. In this technique, 1000 simulations were performed to find out the best weight combination for the WV-EL model. It is important to notice that, to avoid data leak, 5-fold CV technique was used to evaluate the WV-EL model instead of the testing set. Besides, performance indicator $R^2$ was selected to be the fitness value of optimization.

The results of the optimization process are illustrated in Figure 6. It can be seen that, out of three base models, GB played the most important role for the EL model in predicting the pile bearing capacity. To be more specific, the higher the first weight was, the better the WV-EL model performed. Besides, the second weight should be in the $[0.2–0.7]$ range and the third weight should be less than 0.4. The best fitness value of optimization was $R^2 = 0.932$, and the optimum weights are summarized in Table 6.

The regression graph for the WV-EL model is shown in Figure 7. The visual comparison between $P_u$ results of the WV-EL model and experimental data is illustrated in Figure 8 (for the training data set) and in Figure 9 (for the testing data set). It can be seen that the WV-EL model gave a
Table 3: Initialization parameters of base models.

| Parameters                        | Gradient boosting | Random forest | Linear regression |
|-----------------------------------|-------------------|---------------|-------------------|
| Number of estimators              | 50                | 50            | Solver = Newton methods |
| Learning rate                     | 0.1               | —             | Max iteration = 100 |
| Min samples leaf                  | 2                 | 2             | —                 |
| Min samples split                 | 4                 | 4             | —                 |
| Max depth = 3                     | 8                 | 8             | —                 |
| Cost function                     | MSE               | MSE           | —                 |
| Data used                         | Training/testing set | Training/testing set | Training/testing set |

Table 4: Comparison of percentage relative error in model predicted values for the testing set.

| No. | Actual data (kN) | GB  | RF  | LR  | AV-EL |
|-----|------------------|-----|-----|-----|-------|
| 1   | 1845             | 7.2 | 9.8 | 63.8 | 26.9  |
| 2   | 1240             | −10.1| −13.1| 13.1 | −3.4  |
| 3   | 1140             | −12.2| −2.0 | 9.0  | −1.7  |
| 4   | 1500             | 21.4 | 2.3  | 9.3  | 11.0  |
| 5   | 1210             | −15.7| −4.0 | −40.3| −20.0 |
| 6   | 1350             | −21.5| −11.5| 11.3 | −7.2  |
| 7   | 1390             | −29.5| −35.5| −39.7| −34.9 |
| 8   | 1260             | −6.8 | −2.5 | 0.2  | −3.0  |
| 9   | 1170             | 27.1 | 28.5 | 7.4  | 21.0  |
| 10  | 725              | 19.4 | 21.2 | −40.9| −0.1  |
| 11  | 1485             | −16.5| −13.7| 2.0  | −9.4  |
| 12  | 7500             | −4.0 | −28.8| 43.4 | 3.5   |
| 13  | 1710             | −17.4| −5.6  | −24.0| −15.6 |
| 14  | 5435             | 32.8 | 8.2  | 22.8 | 21.3  |
| 15  | 2257             | 7.5  | 13.4 | −5.9 | 5.0   |
| 16  | 470              | 101.5| 125.1| 81.9 | 102.8 |
| 17  | 800              | −16.7| 3.0  | −68.3| −27.3 |
| 18  | 1070             | −13.8| −18.2| 13.2 | −6.3  |
| 19  | 2900             | −22.2| −23.9| −21.5| −22.5 |
| 20  | 645              | 13.2 | 12.8 | −56.9| −10.3 |
| 21  | 2100             | −9.6 | −20.1| 11.8 | −6.0  |
| 22  | 1020             | 19.8 | 20.7 | 79.5 | 40.0  |

Figure 5: Continued.
Figure 5: Regression graph for the models: gradient boosting for (a) training and (b) testing; random forest for (c) training and (d) testing; logistic regression for (e) training and (f) testing; AV-Ensemble Learning for (g) training and (h) testing.
better performance in the training set, with performance indicators of $R^2 = 0.951$, MAE = 242.406, and RMSE = 339.735 compared to the AV-EL model. In addition, the WV-EL model also gave the same performance as that of the AV-EL model in the testing set, with performance indicators of $R^2 = 0.95$, MAE = 269.87, and RMSE = 375.612. In other words, the WV-EL model exhibited a more balanced performance and was a better choice than the AV-EL model.

4.4. Comparison with the Previous Study. Table 7 presents some results on ML applications in pile foundation. The results of the present study as well as those of the previous studies showed the expected ML technique in predicting pile bearing capacity with good prediction. The comparison between study results does not make much sense because of the different data sets and major inputs. However, among those models, the WV-EL model showed high predictive performance and provides a balanced prediction across the entire data set. Besides, the data set size in the study is limited, so further studies on other data sets are needed to provide the most general model for the prediction of pile bearing capacity.

SPT is standard penetration test value, CPT is cone penetration test value, $G$ denotes pile geometrical properties, $\phi$ denotes internal friction angle of soil, $\sigma'$ denotes effective stress of the soil, $HW$ is hammer weight, $PS$ is pile set, $DH$ is

| Model                          | Data set | $R^2$ | MAE (kN) | RMSE (kN) | Training rank | Testing rank |
|-------------------------------|----------|-------|----------|-----------|---------------|--------------|
| Gradient boosting             | Training set | 0.976 | 191.139  | 249.088   | 1             |              |
|                              | Testing set | 0.929 | 310.801  | 466.769   | 2             |              |
| Random forest                 | Training set | 0.91  | 235.532  | 437.26    | 3             |              |
|                              | Testing set | 0.827 | 319.848  | 546.329   | 5             |              |
| Linear regression             | Training set | 0.781 | 549.559  | 727.217   | 5             |              |
|                              | Testing set | 0.861 | 515.661  | 856.316   | 4             |              |
| AV-Ensemble Learning          | Training set | 0.928 | 292.897  | 407.041   | 2             |              |
|                              | Testing set | 0.955 | 263.49   | 372.338   | 1             |              |

Table 5: Summary of prediction capability of models.
Figure 7: Regression graph for the WV-EL model for (a) training set and (b) testing set.

Figure 8: Prediction results of WV-EL for the training data set.

Figure 9: Prediction results of WV-EL for the testing data set.
drop height of hammer, \( T \) is type of test, \( P \) is type of pile, \( I \) denotes the installation method, and \( E_p \) denotes end of pile type.

4.5. Sensitivity Analysis. In this section, a global sensitivity technique is conducted to evaluate the importance of input parameters for the model using the Sobol method [64]. The survey data set is taken from the permitted range by Saltelli’s sampling scheme [65]. To make the results more objective, 300 simulations taking into account the shuffling in the order of the training data were performed to construct the EL model.

The analysis result is shown in Figure 10. It can be seen that, among the 9 input variables used to detect the pile bearing capacity, the shaft area (\( A_f \)) was the most important feature, which scored an average sensitivity index of 0.537. From a soil mechanic point of view, \( A_f \) affects the contact area between the soil and the pile; in other words, it affects the total positive friction force on the pile and so plays an important role in predicting the bearing capacity of the friction pile. \( A_t \) pile tip cross-sectional area (\( A_t \)) was the next important variable, which refers to pile tip resistance, with a score sensitivity index of 0.970. The variables \( g_{crt} \), \( T \), and \( f_{oa} \) were ranked as the third to the fifth important predictors, with average sensitivity index ranging from 0.039 to 0.094.

### Table 7: Summary of comparison with previous studies.

| References                  | Model  | Data set    | Number of samples | Foundation type | Major inputs          | \( R^2 \) | \( R \)  |
|-----------------------------|-------|-------------|-------------------|-----------------|------------------------|----------|--------|
| Momeni et al. [3]           | GA-ANN| Training set| 50                | Pile            | G, PS, HW, DH          | 0.96     | 0.980  |
| Kiefa [63]                  | GRNN  | Training set| 59                | Pile            | G, \( \phi \), \( \sigma' \) | 0.963    | 0.981  |
| Shahin and Jaksa [24]       | ANN   | Training set| 94                | Drilled shaft   | G, CPT                 | 0.9409   | 0.970  |
| Shahin [25]                 | ANN   | Training set| 80                | Driven pile     | G, CPT                 | 0.9409   | 0.970  |
| Shahin [25]                 | ANN   | Training set| 94                | Drilled shaft   | G, CPT                 | 0.9409   | 0.970  |
| Ghorbani et al. [30]        | ANFIS | Training set| 108               | Pile            | G, CPT                 | 0.921    | 0.960  |
| Pham et al. [28]            | GA-DNN| Training set| 472               | Driven pile     | G, SPT                 | 0.944    | 0.972  |
| AV-EL (the present study)   | EL    | Training set| 108               | Pile            | G, T, P, I, \( E_p \), CPT | 0.928    | 0.928  |
| WV-EL (the present study)   | EL    | Training set| 108               | Pile            | G, T, P, I, \( E_p \), CPT | 0.951    | 0.975  |

The analysis result is shown in Figure 10. It can be seen that, among the 9 input variables used to detect the pile bearing capacity, the shaft area (\( A_f \)) was the most important feature, which scored an average sensitivity index of 0.537. From a soil mechanic point of view, \( A_f \) affects the contact area between the soil and the pile; in other words, it affects the total positive friction force on the pile and so plays an important role in predicting the bearing capacity of the friction pile. \( A_t \) pile tip cross-sectional area (\( A_t \)) was the next important variable, which refers to pile tip resistance, with a score sensitivity index of 0.970. The variables \( g_{crt} \), \( T \), and \( f_{oa} \) were ranked as the third to the fifth important predictors, with average sensitivity index ranging from 0.039 to 0.094.

### Figure 10: Sensitivity analysis index.
0.031. The remaining variables including $E_p$, $q_{cat}$, $I$, and $P$ got a sensitivity index less than 0.012, indicating that they did not affect the result much.

5. Conclusions

In this study, an EL model was developed to predict the pile bearing capacity. The combination model was made up of three base models: GB, RF, and LR. A database containing 108 samples was used to develop and evaluate ML models. An optimization technique, namely, the RS, was used to choose the optimal weights for the WV-EL model.

The results showed that the two EL models used in this study, namely, AV-EL and WV-EL, exhibited superior performance compared to the base models. Besides, the EL models seem to solve the overfitting and underfitting problems of the base models by demonstrating excellent performance on both the training and testing data sets. In particular, the WV-EL model which has weights of $w_1 = 0.482, w_2 = 0.338,$ and $w_3 = 0.18$ for the GB, RF, and LR model, respectively, showed the best balance in the ability to predict pile bearing capacity on both the training and testing data sets.

In addition, the global sensitivity analysis technique was used to detect the most important input variables. The results showed that, out of all the input parameters, the shaft area ($A_s$) and the pile tip cross-sectional area ($A_t$) were considered to be the most important features for modeling the pile bearing capacity.

The research results have provided an effective tool for predicting the pile capacity according to the CPT test results and also showed potential in the use of the weight voting protocol Ensemble Learning model in solving foundation engineering problems.

Data Availability

The processed data are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors extend their appreciation to the University of Transport Technology for funding this research under grant number DTTD2021-24.

References

[1] T. Chapman and A. Marcetteau, “Achieving economy and reliability in piled foundation design for a building project,” The Structural Engineer, vol. 2, pp. 32–37, 2004.
[2] J. B. Burland, B. B. Broms, and V. F. De Mello, Behaviour of Foundations and Structures, BREF, Berkshire, UK, 1978.
[3] E. Momeni, R. Nazir, D. Jahan Armaghani, and H. Maizir, “Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN,” Measurement, vol. 57, pp. 122–131, 2014.
[4] W. Chen, P. Sarir, X.-N. Bui, H. Nguyen, M. M. Tahir, and D. J. Armaghan, “Neuro-genetic, neuro-empirical and genetic programming models in predicting ultimate bearing capacity of pile,” Engineering with Computers, vol. 36, pp. 1101–1115, 2019.
[5] G. G. Meyerhof, “Bearing capacity and settlement of pile foundations,” Journal of the Geotechnical Engineering Division, vol. 102, no. 3, 1976.
[6] J. H. Schmertmann, Guidelines for Cone Penetration Test (Performance and Design), Federal Highway Administration, Washington, DC, USA, 1978.
[7] J. de Kuiter and F. L. Beringen, “Pile foundations for large north sea structures,” Marine Geotechnology, vol. 3, no. 3, pp. 267–314, 1979.
[8] N. Shariatmadari, A. Eslami, and M. Karimpour-Fard, “Bearing capacity of driven piles in sands from SPT-applied to 60 case histories,” Iranian Journal of Science and Technology Transaction B-Engineering, vol. 32, p. 17, 2008.
[9] A. R. Bazaar and M. M. Kurkur, “N-values used to predict settlements of piles in Egypt, 1986,” 2020, https://cedb.asce.org/cedbsearch/record.jsp?dockey=0048882.
[10] C. Zhang, G. D. Nguyen, and I. Einav, “The end-bearing capacity of piles penetrating into crushable soils,” Geotextile Geomembrane, vol. 63, no. 5, pp. 341–354, 2013.
[11] I. Shooshpasha, A. Hasanzadeh, and A. Taghavi, “Prediction of the axial bearing capacity of piles by SPT-based and numerical design methods,” International Journal of Geomat, vol. 4, no. 2, pp. 560–564, 2013.
[12] Z. H. Elsherbiny and M. H. El Naggar, “Axial compressive capacity of helical piles from field tests and numerical study,” Canadian Geotechnical Journal, vol. 50, no. 12, pp. 1191–1203, 2013.
[13] M. Jesswein, J. Liu, and M. Kwak, “Predicting the side resistance of piles using a genetic algorithm and SPT N-values,” in Proceedings of the 2018 GeoEdmonton Conference, Edmonton, Canada, October 2018.
[14] M. Y. Abu-Farsakh and H. H. Titi, “Assessment of direct cone penetration test methods for predicting the ultimate capacity of friction driven piles,” Journal of Geotechnical and Geoenvironmental Engineering, vol. 130, no. 9, pp. 935–944, 2004.
[15] A. A. Ozok, “Survey design and implementation in HCI,” Human-Computer Interaction: Development Process, vol. 253, 2009.
[16] F. Rausche, F. Moses, and G. G. Goble, “Soil resistance predictions from pile dynamics,” in Current Practices and Future Trends in Deep Foundations, pp. 418–440, ASCE, Reston, VA, USA, 2004.
[17] P. Liu, Q. Xing, Y. Dong, D. Wang, M. Oeser, and S. Yuan, “Application of finite layer method in pavement structural analysis,” Applied Sciences, vol. 7, no. 6, p. 611, 2017.
[18] A. I. Dounis and C. Caraiscos, “Advanced control systems engineering for energy and comfort management in a building environment—a review,” Renewable and Sustainable Energy Reviews, vol. 13, no. 6-7, pp. 1246–1261, 2009.
[19] F.-Y. Wang, “Parallel control and management for intelligent transportation systems: concepts, architectures, and applications,” IEEE Transactions on Intelligent Transportation Systems, vol. 11, no. 3, pp. 630–638, 2010.
[20] G. Allen and T. Chan, Artificial Intelligence and National Security, Belfer Center for Science and International Affairs, Cambridge, MA, USA, 2017.
[21] X. Yang, Y. Wang, R. Byrne, G. Schneider, and S. Yang, “Concepts of artificial intelligence for computer-assisted drug
discovery,” Chemical Reviews, vol. 119, no. 18, pp. 10520–10594, 2019.
[22] I.-M. Lee and J.-H. Lee, “Prediction of pile bearing capacity using artificial neural networks,” Computers and Geotechnics, vol. 18, no. 3, pp. 189–200, 1996.
[23] T. A. Pham, H.-B. Ly, V. Q. Tran, L. V. Giap, H.-L. T. Vu, and H.-A. T. Duong, “Prediction of pile axial bearing capacity using artificial neural network and random forest,” Applied Sciences, vol. 10, no. 5, p. 1871, 2020.
[24] M. A. Shahin and M. B. Jaksa, “Intelligent computing for predicting axial capacity of drilled shafts,” in Contemporary Topics in In Situ Testing, Analysis, and Reliability of Foundations, pp. 26–33, ASCE, Orlando, FL, USA, 2009.
[25] M. A. Shahin, “Intelligent computing for modeling axial capacity of pile foundations,” Canadian Geotechnical Journal, vol. 47, no. 2, pp. 230–243, 2010.
[26] T. Bong, S.-R. Kim, and B.-I. Kim, “Prediction of ultimate bearing capacity of aggregate pier reinforced clay using multiple regression analysis and deep learning,” Applied Sciences, vol. 10, no. 13, p. 4580, 2020.
[27] S. Fong, S. Deb, and X.-s. Yang, “How meta-heuristic algorithms contribute to deep learning in the hype of big data analytics,” in Progress in Intelligent Computing Techniques: Theory, Practice, and Applications, vol. 518, pp. 3–25, Springer, Singapore, 2018.
[28] T. A. Pham, V. Q. Tran, H.-L. T. Vu, and H.-B. Ly, “Design deep neural network architecture using a genetic algorithm for estimation of pile bearing capacity,” PLoS One, vol. 15, no. 12, Article ID e0243030, 2020.
[29] H. Moayedi and S. Hayati, “Artificial intelligence design charts for predicting friction capacity of driven pile in clay,” Neural Computing and Applications, vol. 31, no. 11, pp. 7429–7445, 2019.
[30] B. Ghorbani, E. Sadrossadat, J. Bolouri Bazaz, and P. Rahimzadeh Oskooei, “Numerical ANFIS-based formulation for prediction of the ultimate axial load bearing capacity of piles through CPT data,” Geotechnical and Geological Engineering, vol. 36, no. 4, pp. 2057–2076, 2018.
[31] X. Ying, “An overview of overfitting and its solutions,” Journal of Physics: Conference Series, vol. 1168, Article ID 022022, 2019.
[32] A. Ghasemian, H. Hosseinimardi, and A. Clauset, “Evaluating overfit and underfit in models of network community structure,” IEEE Transactions on Knowledge and Data Engineering, vol. 32, pp. 1722–1735, 2019.
[33] A. Mosavi, M. Golshan, S. Janizadeh, B. Choubin, A. M. Melesse, and A. A. Dineva, “Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: a priority assessment of sub-basins,” Geocarto International, vol. 35, pp. 1–20, 2020.
[34] A. Mosavi, A. Shirzadi, B. Choubin et al., “Towards an ensemble machine learning model of random subspace based functional tree classifier for snow avalanche susceptibility mapping,” IEEE Access, vol. 8, pp. 145968–145983, 2020.
[35] A. Mosavi, F. Sajedi Hosseini, B. Choubin, M. Goodarzi, A. A. Dineva, and E. Rafiei Sardooi, “Ensemble boosting and bagging based machine learning models for groundwater potential prediction,” Water Resources Management, vol. 35, no. 1, pp. 23–37, 2021.
[36] L. Breiman, Random Forests, Statistics Department University of California, Berkeley, CA, USA, 1999.
[37] M. Kuhn and K. Johnson, Applied Predictive Modeling, Vol. 26, Springer, Berlin, Germany, 2013.
[38] W. Zhang, C. Wu, Y. Li, L. Wang, and P. Samui, “Assessment of pile drivability using random forest regression and multivariate adaptive regression splines,” Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, vol. 15, pp. 27–40, 2019.
[39] N. Kardani, A. Zhou, M. Nazem, and S.-L. Shen, “Estimation of bearing capacity of piles in cohesionless soil using optimised machine learning approaches,” Geotechnical and Geological Engineering, vol. 38, no. 2, pp. 2271–2291, 2020.
[40] S. Tourzani, J. Granderson, and S. Fernandes, “Gradient boosting machine for modeling the energy consumption of commercial buildings,” Energy and Buildings, vol. 158, pp. 1533–1543, 2018.
[41] R. E. Schapire, “Using output codes to boost multiclass learning problems,” in Proceedings of the 14th International Conference on Machine Learning, vol. 97, pp. 313–321, San Francisco, CA, USA, 1997.
[42] L. Breiman, “Arcing classifier (with discussion and a rejoinder by the author),” The Annals of Statistics, vol. 26, no. 3, pp. 801–849, 1998.
[43] L. Guelman, “Gradient boosting trees for auto insurance loss cost modeling and prediction,” Expert Systems with Applications, vol. 39, no. 3, pp. 3659–3667, 2012.
[44] R. Tibshirani, “Regression shrinkage and selection via the lasso,” Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, no. 1, pp. 267–288, 1996.
[45] J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
[46] J. H. Friedman, “Stochastic gradient boosting,” Computational Statistics & Data Analysis, vol. 38, no. 4, pp. 367–378, 2002.
[47] S. H. G. Silva, A. F. d. S. Teixeira, M. D. d. Menezes, L. R. G. Guilherme, F. M. d. S. Moreira, and N. Curí, “Multiple linear regression and random forest to predict and map soil properties using data from portable X-ray fluorescence spectrometer (pXRF),” Ciência e Agrotecnologia, vol. 41, no. 6, pp. 648–664, 2017.
[48] I. Sharma and S. Kakchapati, “Linear regression model to identify the factors associated with carbon stock in Chure forest of Nepal,” Scientifica, vol. 2018, Article ID 1383482, 8 pages, 2018.
[49] W. J. Boscardin, “The use and interpretation of linear regression analysis in ophthalmology research,” American Journal of Ophthalmology, vol. 150, no. 1, pp. 1–2, 2010.
[50] A. Bjöörk, Numerical Methods for Least Squares Problems, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1996.
[51] C. R. Rao and H. Toutenburg, Linear Models: Least Squares and Alternatives, Springer, Berlin, Germany, 2nd edition, 1995.
[52] W. Sun and B. Trevor, “A stacking ensemble learning framework for annual river ice breakup dates,” Journal of Hydrology, vol. 561, pp. 636–650, 2018.
[53] Z. Ma, P. Wang, Z. Gao, R. Wang, and K. Khalighi, “Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose,” PLoS One, vol. 13, no. 10, Article ID e0205872, 2018.
[54] A. G. Barnston, “Correspondence among the correlation, RMSE, and Heidke forecast verification measures: refinement of the Heidke score,” Weather and Forecasting, vol. 7, no. 4, pp. 699–709, 1992.
[55] F. Fortin, M. Abaza, F. Ancil, and R. Turcotte, “Why should ensemble spread match the RMSE of the ensemble mean?”
[56] F. Moksony, “Small is beautiful. The use and interpretation of R² in social research,” Szociológiai Szemle, Special Issue, pp. 130–138, 1990.

[57] H. Tang and M. Mayersohn, “Utility of the coefficient of determination (r²) in assessing the accuracy of interspecies allometric predictions: illumination or illusion?” Drug Metabolism and Disposition, vol. 35, no. 12, pp. 2139–2142, 2007.

[58] C. Willmott and K. Matsuura, “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,” Climate Research, vol. 30, no. 1, pp. 79–82, 2005.

[59] W. Wang and Y. Lu, “Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model,” IOP Conference Series: Materials Science and Engineering, vol. 324, no. 1, Article ID 012049, 2018.

[60] H. Sug, “The effect of training set size for the performance of neural networks of classification,” WSEAS Transactions on Computers, vol. 9, no. 11, pp. 1297–1306, 2010.

[61] D. Anguita, L. Ghelardoni, A. Ghio, L. Oneto, and S. Ridella, “The “K” in K-fold cross validation,” in Proceedings of the 2012 European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), Bruges, Belgium, 2012.

[62] S. Yadav and S. Shukla, “Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification,” in Proceedings of the 2016 IEEE 6th International Conference on Advanced Computing (IACC), pp. 78–83, Bhimavaram, India, February 2016.

[63] M. A. A. Kiefa, “General regression neural networks for driven piles in cohesionless soils,” Journal of Geotechnical and Geoenvironmental Engineering, vol. 124, no. 12, pp. 1177–1185, 1998.

[64] I. M. Sobol’, “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates,” Mathematics and Computers in Simulation, vol. 55, no. 1–3, pp. 271–280, 2001.

[65] A. Saltelli, “Making best use of model evaluations to compute sensitivity indices,” Computer Physics Communications, vol. 145, no. 2, pp. 280–297, 2002.