Data-driven modeling of ultimate load capacity of closed- and open-ended piles using machine learning

M.B.C. Ülker, E. Altinok and G. Taşkin

*Earthquake Engineering Program, Istanbul Technical University Institute of Disaster Management, Istanbul, Turkey; ‡Department of Civil Engineering, Istanbul University-Cerrahpaşa Faculty of Engineering, Istanbul, Turkey

ABSTRACT
Field pile load tests are fairly expensive experiments that can be applied to certain pile types required to be installed in full scale. Hence, it is neither practical nor efficient to perform a load test for every installed pile. While there exist many empirical relations for predicting pile capacities, such methods typically suffer from accuracy and generality. Therefore, current geotechnical practice still looks for methods to accommodate full-scale pile load testing to serve as accurate and practical tools. In this study, load bearing capacities of closed- and open-ended piles in cohesive and cohesionless soils are predicted using machine learning. Nine such methods are utilized in the analyses where Cone Penetration Test (CPT) and pile data are considered as the learning features necessary to teach those methods the database gathered via a comprehensive search. Then, machine learning models are developed, and the databases are separated into five-folds according to the cross-validation-principle, which are used for both training and testing of the machine learning methods. Model predictions are validated with classical CPT-based equations. Results indicate that Relevance Vector Regression and the Random Forest methods typically generate considerably better predictions than the other methods and empirical equations. Thus, machine learning methods are found as reliable tools to predict the pile load capacities of both open-ended and closed-ended pile provided that there is a large enough database and that an appropriate method is used.

1. Introduction
With the extent of advanced technology, machine learning has become one of the most commonly used topics in computer science that is increasingly being applied to solving civil engineering problems in the last decades (Reich 1997). Machine learning methods (MLM) can be used in a wide range of application areas from analysing statistical distributions of design parameters used in the analysis of engineering systems to predicting material behaviour which can be used in developing theoretical procedures for estimating load capacities. For that matter, MLM have been developed over the years. Lately, MLM are found useful in geotechnical engineering essentially to predict bearing capacities of soils from laboratory and/or field experiments (Shahin 2016).

Predicting bearing capacity of piles from Cone Penetration Test (CPT) is a compelling task. Abu-Farsakh and Titi (2004) evaluate the reliability of traditional CPT-based pile capacity methods on driven piles. Laboratoire Central des Ponts et Chaussées (LCPC) methods by Bustamante and Gianselli (1982), De Ruiter and Beringen (1979) show better performance than others (Abu-Farsakh and Titi 2004). Schmertmann-Nottingham method (Schmertmann 1978) provides reliable capacity results. Esami and Fellenius (1997) propose a direct method to estimate pile capacity from CPT with pore pressure measurements. Thereafter, an enhanced procedure is proposed to improve this method (Niazi and Mayne 2016). Lehane et al. (2007) develop the UWA-05 method for capacity of piles in siliceous sand from cone tip resistance.

In the last three decades, artificial intelligence (AI) is applied to many geotechnical engineering problems such as predicting the index parameters of soils, modelling the pile load tests, optimizing the cost of retaining structures, etc. (Alzouabi and Ibrahim 2019; Banimad, Yasrobi, and Woodward 2005; Dağdeviren and Kaymak 2018; Jafari, Habibaghah, and Ghahramani 2006; Kumar and Samui 2019; Lamorski et al. 2008; Mohanty and Das 2018; Yurtcu and Özocak 2016; Zhu, Zaman, and Anderson 1998). Predicting load capacity of foundations by MLM needs more comprehensive studies due to irregularity among the results and requirement of large datasets. Artificial Neural Networks (ANN) is the earliest of such methods applied to capacity predictions (Lee and Lee 1996; Teh et al. 1997). Shahin et al. (2002) use Neural Networks to estimate the settlement rate of shallow foundations. Samui (2008) applies the Support Vector Machine (SVM) for prediction of shaft capacity of piles by using the undrained shear strength and CPT friction resistance parameters as the input features. Pal and Deswal (2010) predict axial capacity of piles including tip and shaft resistances using the Gaussian Process Regression (GPR). Their results show that SVM, Polynomial Neural Networks (PNN), Genetic Algorithms (GA) and GPR demonstrate better performances.
than the traditional methods (Bustamante and Gianselli 1982; De Ruiter and Beringen 1979; Schmertmann 1978). Subsequently, studies are conducted on predicting the load-settlement behaviour of piles by applying optimized NN (Ismail and Jeng 2011; Ismail, Jeng, and Zhang 2013). Debnath and Dey (2018) and Moayedi and Hayati (2018) contribute to predicting bearing capacity and settlement of piles through AI. Pham et al. (2020) use Standard Penetration Test (SPT) to predict the capacity of driven concrete piles using Deep Learning. Gomes et al. (2021) compare the performance of six MLM. Jesswein and Liu (2022) introduce a new GA-based pile design method using SPT-N. Some others are noteworthy (Cao, Nguyen, and Wang 2022; Hoang, Tran, and Huynh 2022; Nguyen et al. 2022; Pham, Tran, and Mumtaz 2022). Nguyen et al. (2022b) predict the capacity of Precast High-Strength Concrete (PHC) piles from SPT using a feed-forward NN.

While SPT are widely available, they do not necessarily provide reliable information for soils. Therefore, predictions made by using databases composed solely of SPT may not always be that accurate. Thus, in this study we make use of CPT in our ML analyses.

Estimating pile capacity from CPT through ML regression has been used in the last decades. Ardalan et al. (2009) estimate pile shaft capacity from CPT results utilizing the PNN and GA. Shahin (2010) examines the performance of an ANN algorithm for 80 driven piles and 94 drilled shafts using full-scale tests and CPT. Performance of their ANN model is analysed in terms of two statistical metrics namely R-square (R²) and Root Mean Squared Error (RMSE). Gene Expression Programming (GEP) is used to build a new correlation between pile capacity and CPT data in cohesive soils by Akkroosh and Nikraz (2011, 2012). SVM algorithm is implemented by Kordjazi et al. (2014). Akkroosh et al. (2015) develop a new model for bored piles with the Least Square Support Vector Machines (LSSVM). Ebrahimian and Movahed (2017) propose two formulations for predicting the axial capacity of piles applying evolutionary based approaches. Kardani et al. (2020) use six different MLM to predict pile capacity with a relatively small database containing 59 data samples. While there has been considerable research on MLM that estimates the pile capacity, most of these studies focus on predictive abilities of a small group of methods. As computer technology advances, new MLM are developed leading to better approximations. In this study, nine of those most common MLM are used to predict the axial capacities of closed- and open-ended piles. Static tests and CPT data are used in the training and test datasets.

2. Machine learning methods

In machine learning (ML) linear regression provides an explicit model making it more interpretable than nonlinear methods allowing for parameters’ effect on the output. Nonlinear methods, however, provide an implicit function (i.e. black-box) achieving higher performance with less interpretability. Here focus is on understanding the regression methods rather than rigorous theoretical details.

In ML, a hyper-parameter should be set before training a model. Hyper-parameters are used to control the model to get the best performance on a given dataset. Selection of proper hyper-parameters often depends upon the complexity of the algorithm and can have a significant impact on the performance of a model. In this study, the hyper-parameters are tuned by the cross-validation principle in each method.

**Ridge Linear Regression (RLR):**

RLR is a variant of linear regression, where we have the following model:

\[ f(x_1, \cdots, x_n; w_1, \cdots, w_n) = w_0 + w_1 x_1 + \cdots + w_n x_n \]  

(1)

where \( w_i \) refers to the weight associated with the parameter, \( x_i \).

The loss function of the ridge regression is defined by penalizing residual sum of squares as follows:

\[ \min_{w_1, \cdots, w_n} \sum_{i=1}^{m} ||f'(x'_i, \cdots, x'_n; w_1, \cdots, w_n) - y'_i||^2 + \lambda \sum_{i=1}^{n} w_i^2 \]  

(2)

where \( \lambda \) is the regularization parameter controlling the smoothness of the model, while the last term is called \( l_2 \) norm regularizer. The loss function is minimized by taking the partial derivatives with respect to model parameters, yielding the following system of equations:

\[ w = (X^T X + \lambda I)^{-1} X^T y \]  

(3)

where \( I \) is an \( n \times n \) identity matrix, and \( y \) is a vector that holds all the outputs associated with given observations.

**Multivariate Adaptive Regression Splines (MARS):**

MARS is a multivariate non-parametric regression method, which is a generalization of stepwise linear regression (Friedman 1991). The learning model is expressed as:

\[ f(x) = w_0 + \sum_{j=1}^{q} w_j P_j(x) \]  

(4)

where \( q \) is the number of basis functions included in the MARS, and \( w_j \) is the coefficient of the \( j \)-th basis function \( P_j \). MARS has two steps: i) Creating a collection of basis functions based on a dataset, ii) Conducting a least square regression to find the weights of those basis functions.

**k-Nearest Neighbors Regression (kNNR):**

kNNR is a non-parametric regression based on the similarity between samples in the input space. To predict the output value \( (y_i) \) for a given sample \( (x_i) \), a neighbourhood set is determined using the \( k \)-nearest neighbours of the corresponding sample according to Euclidean distance. The output value is then calculated as the average of the outputs with \( k \)-nearest neighbours of the sample.

**Decision Trees (TREE):**

TREE is a non-parametric method based on a set of hierarchical nodes each of which performs a decision tree algorithm with a testing function, \( f_m(.) \) using individual input features. Here, \( m \) is the number of nodes. Main advantage of the decision tree is its interpretability because each path from root to leaf can be converted into a set of understandable ‘if/else’ structures.

**Random Forests (RF):**

RF consists of a collection of randomized base tree classifiers, which are constructed by some partitioning rule such as the random split selection (Breiman 2001). RF is an ensemble-
based highly robust method used for classification and regression. Prediction of the output for a given sample is determined by averaging the predictions estimated from each tree.

Extreme Learning Machines (ELM):

ELM is a learning algorithm that allows performing non-linear, multivariate regression for generalized single-hidden-layer feedforward NN, which has one hidden layer and one linear output layer (Huang, Zhu, and Siew 2006). Its learning model is defined for a given input, \( x \), as follows:

\[
f_i(x) = \sum_{l=1}^{L} w_l h_l(x)
\]

where \( L \) is the number of nodes at the hidden layer, and \( w_l \) is the output weight connecting the \( l \)-th hidden node to the output nodes. \( h_l \) represents a activation function at each node.

Relevance Vector Regression (RVR):

RVR is a probabilistic sparse kernel model identical in functional form to the support vector machines (SVM) (Tipping 2000). The probabilistic learning model of RVR is given as follows:

\[
p(y|\alpha, \sigma^2) = (2\pi)^{-\frac{3}{2}}|B^{-1} + \Phi A^{-1} + \Phi^T|^{-\frac{3}{2}} 
\exp\left\{-\frac{1}{2}y^T(B^{-1} + \Phi A^{-1} + \Phi^T)^{-1}y\right\}
\]

where \( A = diag(a_0, \ldots, a_n) \) and \( B = \sigma^{-2}I_n \). \( \sigma \) is determined using the training set, and \( \alpha \) is calculated based on solving an optimization problem. The matrix \( \Phi \) is called kernel functions \( \Phi = K(x, x_i) \).

Gaussian Process Regression (GPR):

GPR is a non-parametric Bayesian multivariate regression method defined as a collection of random variables with a joint Gaussian distribution. The Gaussian Process (GP) is defined as:

\[
f(x) \sim GP(m(x), k(x, x_i))
\]

where \( x \) and \( k(x, x_i) \) correspond to mean and covariance functions, respectively. Given a sample \( x \), desired learning model, \( f_x \), is obtained as follows:

\[
p(f_x | x, f) = N(f_x | \mu_x, \sigma_x^2)
\]

where

\[
\mu_x = k^T_x [K(x, x) + \sigma_x^2 I]^{-1} y
\]

\[
\sigma_x^2 = k(x, x) - k^T_x [K(x, x) + \sigma_x^2 I]^{-1} k_x
\]

Here, \( k_x \) is a vector of covariance values of training samples \( x \) and test sample \( (x) \), \( \sigma_x^2 \) is the noise variance. The \( \sigma_x^2 \) expresses the confidence measure associated with the model output.

Variational Heteroscedastic Gaussian Process Regression (VHGPR):

In the standard GPR, the variance for all observations is constant throughout the input space, which can be unrealistic in some problems. A model considering input-dependent variance is known as heteroscedastic GPR (HGPR) (Lázaro-Gredilla and Titsias 2011), which is a special case of homoscedastic GPR. However, inference in HGPR is very challenging since the predictive density and marginal likelihood are no longer analytically tractable. To deal with this, a variational inference method is used for HGPR based on variational Bayes and Gaussian approximation.

3. Database details

3.1 Collection of the database

Two different databases are developed for closed-ended (CEP) and open-ended piles (OEP). Firstly, a wide database is gathered for CEP with 219 case data including full scale tests, pile geometry, and CPT. Driven bored piles and drilled shafts (concrete and steel piles with various cross sections) are considered. Soil conditions are determined from CPT in terms of cone resistance (\( q_c \)) and sleeve friction (\( f_s \)). 145 pile data are taken directly from Alsanman (1995) and Eslami (1996). Other pile load test data are collected from various studies. The database is publicly available at https://github.com/emirhanaltinok/database-for-paper. Ultimate pile capacities are determined as failure loads as long as the load-displacement relationship indicated clear failure. For those that did not, their capacity is estimated using the Chin – Kondner, the Decourt, and Brinch – Hansen 80% methods (Fellenius 2001).

The second database constitutes 60 OEP load tests and CPT data with steel pipe piles embedded into non-cohesive soils. Piles are subjected to static tensile loads. Ultimate bearings are determined based upon friction capacities.

3.2 Model input and output

CEP Database:

Unit tip resistance and the unit shaft friction over the embedded depth of pile as well as the pile length and cross-sectional area are the necessary main features. Five input features: i) Embedded length of the pile, \( L \); ii) Cross-sectional area, \( A \); iii) Average cone tip resistance around the pile tip, \( (q_c)_{end} \); iv) Average sleeve friction along the embedded length, \( (f_s)_{ave} \); v) Average cone tip resistance along the pile length, \( (q_c)_{ave} \). The first two features are associated with pile geometry, while others are taken from CPT. \( (q_c)_{end} \) is taken as the geometric average of the \( q_c \) along the influence zone defined by Eslami and Fellenius (1997). Geometric mean yields more accurate and repeatable values to analyse CPT as opposed to arithmetic mean since it is less affected by the outliers. The influence zone is defined as the distance measured from \( 4D \) below the pile tip to \( 8D \) above it when the soils above the pile tip are weaker than they are below. \( D \) is the pile diameter. If the soil above the pile tip is dense, then the influence zone starts from \( 2D \) above the pile tip decreasing the effect of denser soils lying above. Figures 1(a, b) shows the interpretation of \( (q_c)_{end} \) when the soil above the pile tip is weaker or denser, respectively. \( (q_c)_{ave} \) and \( (f_s)_{ave} \) are determined by taking the geometric average of measured cone end resistance and sleeve friction. The first 2D distance is not considered due to soil disturbance. These are shown in Figures 1(c, d). As for output, ultimate bearing capacity (\( Q_u \)) is the only variable, which is directly taken from the static tests. The last feature is determined from only 120 data because of the missing sleeve friction values of the other 99. Description of the database is in Table 1.
Once all the input are prepared, feature scaling is applied to eliminate the dimensions. All the feature variables are reduced to suitable forms by applying a zero-mean standardization technique:

\[ x' = \frac{x - \bar{x}}{\sigma} \]  

(11)

where \( x \) is the initial variable; \( \bar{x} \) is the mean of the feature vector and \( \sigma \) is standard deviation.

**OEP Database:**

Similar to CEP, geometric measures and soil conditions are considered for the input. Another geometrical feature, *incremental filling ratio* (IFR), is also specified to mimic the pile-soil interaction. IFR shows the amount of soil plugged in used in predicting the compression/tension capacities of OEP (Paik and Salgado 2003).

Another novel feature is the number of *setup days* representing the days passed after the pile driving until the test. This is important in identifying the piles with the same properties tested in different periods (Chow et al. 1998; Schneider 2007). Although it is generally not considered in design, we investigate the time effect in this study. \( q_{ave} \) is taken as the fifth feature. Description is given in Table 2.

### 4. Training and validation

The \( k \)-fold cross-validation technique is used to train the database and validate its predictions. The training dataset is randomly divided into \( k' \) number of mutually exclusive subsets (folds) of approximately equal size. MLM are trained on \( k-1 \) folds, and the remaining is used for testing the model performance. This process is repeated \( k \) times in a way that each fold is used for testing only once. Expected performance is estimated by averaging the measure, which could be either in terms of the predictive error or accuracy of the learning method over the \( k \) number of trials. The predictions made using the test folds are then compared with the remaining single fold in terms of \( R^2 \) and RE. The flowchart is shown in Figure 2.

### 4.1 Performance measurement

Statistically \( R^2 \) and RE are calculated to determine how close the predicted pile capacities are to the actual values. \( R^2 \) is used to measure the goodness-of-fit indicating if the fit is appropriate for a regression as:

![Figure 1. Input features from CPT; a) \( q_{lax} \) for weak soil above the pile tip, b) \( q_{lax} \) for strong soil above the pile tip, c) \( q_{ave} \), d) \( f_{ave} \).](image)

| Input Features | Output |
|----------------|--------|
| **Length** (m) | **Area** (m²) | \( q_{lax} \) (MPa) | \( q_{ave} \) (MPa) | \( f_{ave} \) (MPa) | \( Q_o \) (kN) |
| Max | 45 | 2.54 | 65.95 | 24.64 | .225 | 14590 |
| Min | 5.6 | 0.06 | 17 | 39 | .005 | 210 |
| Range | 39.4 | 2.48 | 65.79 | 24.26 | .221 | 14380 |
| Std. Dev. | 8.7 | 0.35 | 7.61 | 4.42 | .039 | 2224 |
| Mean | 14.6 | 0.25 | 9.42 | 5.64 | .059 | 2203 |
| Number of data | 219 | 219 | 219 | 219 | 120 | 219 |

*Table 1. Statistical parameters of input/output for CEP database.*
Table 2. Statistical parameters of input/output for OEP database.

| Input Features | Output |
|----------------|--------|
|                | Length | Diameter | IFR | Setup Days | \( Q_{\text{ave}} \) | \( Q_I \) |
|                | (m)    | (mm)     | (%) | (Day)      | (Mpa)      | (kN)   |
| Max            | 47.0   | 1220     | 95  | 1950       | 42.24      | 19482  |
| Min            | 2.5    | 34       | 45  | 2          | 2.65       | 6      |
| Range          | 44.5   | 1186     | 50  | 1948       | 39.59      | 19476  |
| Std. Dev.      | 12.5   | 304      | 15  | 283        | 8.53       | 3875   |
| Mean           | 15.4   | 432      | 72  | 163        | 11.12      | 2173   |
| Number of Data | 60     | 60       | 60  | 60         | 60         | 60     |

Figure 2. Flowchart of the proposed approach.


\[ R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \]  

(12)

where \( y_i \) and \( \hat{y}_i \) correspond to the actual and the predicted value of the \( i \)th sample, respectively. \( \bar{y} \) is the mean. There is a strong correlation if the correlation coefficient is greater than 0.8. Thus, \( R^2 \) gets larger than 0.64 (Sadrossadat et al. 2013) calculated as:

\[ RE = \frac{\hat{y}_i - y_i}{y_i} \]  

(13)

\[ R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \]  

(14)

4.2 Classical methods for Axial Pile capacity

We revisit some empirical relationships to calculate the pile capacity. Among those that exhibit sufficient accuracy, three methods stand out. Results are then used in validation of ML predictions (Table 3).

5. Results

5.1 CEP database

A database of 219 pile tests with 120 having \((f_s)\) and \((q_s)\), while 99 missing \(f_s\), is generated. The input features are \(L, N, q_c, converse_f, \alpha, \) and \(q_s\). The output is \(Q_s\). Therefore, the first set of analyses are performed on a database of 120 tests having both \(q_s\) and \(f_s\). The accuracy is estimated based on 5-fold cross validation. We follow a naming convention to avoid confusion and call the first analysis set 120S_5F_9ML referring to 120 sample, 5 features and 9 MLM.

Based on preliminary studies (Shahin 2010), in the second set, \(f_s\) is removed from the feature set, and the analyses are repeated with 120 samples. In addition, 4 of the 9 MLM resulting in the highest correlation with the database are used in the subsequent analyses. This second set is then called 120S_4F_4ML. It is observed that removing \(f_s\) enhances the performance. Thus, the current dataset is expanded with additional 99 tests without \(f_s\). This enables us to gather a comprehensive database with 219 tests. In the rest (219S_4F_4ML) this database is used.

MLM are first applied to initial database including the \(f_s\) whose results are in Table 4. The highest \(R^2\) and the lowest RE of each method are highlighted. Then performances are assessed with boxplots in Figure 3.

Table 3. Classical method equations.

| Method                  | End Capacity | Shaft Capacity |
|-------------------------|--------------|----------------|
| Schmertmann             | \(q_p = a_0 + a_1 r_s\) | For clay: \(Q_s = \frac{\sigma f_s A_t}{\sqrt{3}}\) | For Sand: \(Q_s = K \left( \sum \frac{z_i f_i A_i}{d_i^{1.5}} + \sum \frac{z_i f_i A_i}{d_i^{1.5}} \right)\) |
| (Schmertmann 1978)      |              |                |
| De Ruiter and Beringen  | \(q_p = \sqrt{N s_o}; S_o = \frac{N}{k}\) | For Clay: \(q_i = \alpha S_o\) | For Sand: \(q_i = \min\left( f_s; \frac{q_i}{\alpha S_o} \right)\) |
| (De Ruiter and Beringen 1979) |              |                |
| LCPC                    | \(q_p = k_q q_c\) |                |
| (Bustamante and Gianeselli 1982) |              |                |

Open Ended Piles

| Method                  | End Capacity | Shaft Capacity |
|-------------------------|--------------|----------------|
| UWA                     | \(\tau_{rs} = \frac{\sigma_{rs} K}{\sigma_{rs} + \Delta \sigma_{rs}} \tan \delta_o\) | \(\sigma_{rs} = 0.03 q_s (A_s)^{0.3} \left( \frac{\sqrt{21}}{2} \right)^{0.5}, A_s = 1 - \frac{\sigma_{rs}}{\sigma_{rs0}}\) |
| (Lehane, Schneider, and Xu 2007) |              |                |
| NGI                     | \(\Delta \sigma_{rs0} = 4G_0 S_o \tanh 185 \frac{\left( \frac{q_s}{\sigma_{rs0/\sigma_{rs0}}} \right)^{0.25}}{\left( \frac{q_s}{\sigma_{rs0/\sigma_{rs0}}} \right)^{0.25}}\) | \(\tau_{rs} = 0.45 \frac{\sigma_{rs0}}{\sigma_{rs0}} \left( \frac{\sigma_{rs0}}{\sigma_{rs0}} \right)^{0.15}\left( \frac{\sigma_{rs0}}{\sigma_{rs0}} \right)^{0.15}\) |
| (Clausen et al. 2005)   |              |                |
| Fugro                   | \(R^* = \left( R_{critical} - R_{cmax} \right)^{0.5}\) |                |
| (Yang et al. 2015)      |              |                |

Table 4. \(R^2\) and RE calculated for each fold in 120S_5F_9ML.

| Fold | RLR | MARS | KNNR | TREE | RF | ELM | RVR | GPR | VHGP | Schmertmann | DeRuiter | LCPC |
|------|-----|------|------|------|----|-----|-----|-----|------|-------------|----------|------|
| 1    | 0.83 | 0.85 | 0.74 | 0.61 | 0.86 | 0.45 | 0.87 | 0.85 | 0.87 | 0.79        | 0.74     | 0.81 |
| 2    | 0.39 | 0.31 | 0.53 | 0.40 | 0.27 | 0.56 | 0.27 | 0.25 | 0.34 | 0.28        | 0.32     | 0.29 |
| 3    | 0.71 | 0.81 | 0.66 | 0.60 | 0.85 | 0.76 | 0.80 | 0.86 | 0.84 | 0.80        | 0.52     | 0.87 |
| 4    | 0.37 | 0.32 | 0.52 | 0.42 | 0.30 | 0.38 | 0.29 | 0.27 | 0.30 | 0.40        | 0.59     | 0.34 |
| 5    | 0.78 | 0.75 | 0.76 | 0.80 | 0.75 | 0.77 | 0.81 | 0.76 | 0.48 | 0.62        | 0.58     |      |
| Mean | 0.88 | 0.79 | 0.70 | 0.89 | 0.88 | 0.91 | 0.89 | 0.89 | 0.72 | 0.74        | 0.84     |      |

5.2 CEP database

A database of 219 pile tests with 120 having \((f_s)\) and \((q_s)\), while 99 missing \(f_s\), is generated. The input features are \(L, N, q_c, converse_f, \alpha, \) and \(q_s\). The output is \(Q_s\). Therefore, the first set of analyses are performed on a database of 120 tests having both \(q_s\) and \(f_s\). The accuracy is estimated based on 5-fold cross validation. We follow a naming convention to avoid confusion and call the first analysis set 120S_5F_9ML referring to 120 sample, 5 features and 9 MLM.

Based on preliminary studies (Shahin 2010), in the second set, \(f_s\) is removed from the feature set, and the analyses are repeated with 120 samples. In addition, 4 of the 9 MLM resulting in the highest correlation with the database are used in the subsequent analyses. This second set is then called 120S_4F_4ML. It is observed that removing \(f_s\) enhances the performance. Thus, the current dataset is expanded with additional 99 tests without \(f_s\). This enables us to gather a comprehensive database with 219 tests. In the rest (219S_4F_4ML) this database is used.

MLM are first applied to initial database including the \(f_s\) whose results are in Table 4. The highest \(R^2\) and the lowest RE of each method are highlighted. Then performances are assessed with boxplots in Figure 3.
Figure 3. Distribution of statistical measures among the MLM and traditional methods in 120S_5F_9ML, (a) $R^2$, (b) RE.

Figure 4. Distribution of statistical parameters among the MLM and traditional methods in 120S_4F_4ML, (a) $R^2$, (b) RE.
Figure 5. Boxplot of the 219S_4F_4ML set, (a) $R^2$, (b) RE.

Table 5. $R^2$ and RE values of the 60S_SF_9ML set for OEP.

| Fold | RLR | MARS | KNNR | TREE | RF | ELM | RVR | GPR | VHGP | Schmertmann | DeRui \( \text{MLM} \) | UWA | NGI | Fugro |
|------|-----|------|------|------|----|-----|-----|-----|------|----------|-----------|-----|-----|-------|
| 1    | R2  | 0.72 | 0.90 | 0.75 | 0.72 | 0.91 | 0.90 | 0.96 | 0.94 | 0.79   | 0.96      | 0.96 | 0.89 |
|      | RE  | 0.55 | 0.64 | 0.48 | 0.86 | 0.45 | 0.53 | 0.57 | 0.61 | 0.63   | 0.18      | 0.21 | 0.21 |
| 2    | R2  | 0.54 | 0.25 | 0.85 | 0.02 | 0.76 | 0.68 | 0.98 | 0.40 | 0.86   | 0.98      | 0.99 | 0.84 |
|      | RE  | 0.64 | 2.05 | 0.64 | 1.43 | 0.49 | 0.54 | 0.30 | 0.69 | 0.43   | 0.37      | 0.24 | 0.24 |
| 3    | R2  | 0.95 | 0.94 | 0.99 | 0.48 | 0.88 | 0.95 | 0.93 | 0.96 | 0.50   | 0.89      | 0.92 | 0.60 |
|      | RE  | 0.26 | 0.21 | 0.20 | 0.74 | 0.39 | 0.26 | 0.29 | 0.21 | 0.73   | 0.44      | 0.30 | 0.30 |
| 4    | R2  | 0.91 | 1.00 | 0.94 | 0.93 | 0.99 | 0.94 | 0.95 | 0.98 | 0.94   | 0.92      | 0.95 | 0.78 |
|      | RE  | 0.51 | 0.35 | 0.61 | 0.58 | 0.37 | 0.45 | 0.42 | 0.38 | 0.67   | 0.50      | 0.38 | 0.38 |
| 5    | R2  | 0.60 | 0.24 | 0.66 | 0.09 | 0.23 | 0.61 | 0.98 | 0.85 | 0.71   | 0.89      | 0.81 | 0.83 |
|      | RE  | 0.58 | 0.78 | 0.94 | 1.31 | 0.92 | 0.57 | 0.15 | 0.32 | 0.28   | 0.28      | 0.38 | 0.38 |
| Mean | R2  | 0.74 | 0.67 | 0.84 | 0.45 | 0.76 | 0.81 | 0.96 | 0.83 | 0.76   | 0.93      | 0.92 | 0.79 |
|      | RE  | 0.51 | 0.81 | 0.57 | 0.98 | 0.52 | 0.47 | 0.34 | 0.44 | 0.61   | 0.35      | 0.30 | 0.30 |

According to Table 4, the RF, RVR, GPR, and VHGP result in high $R^2$ and low RE. Boxplots are found useful for identifying the interquartile range and standard deviation. Height of each box represents the interquartile of the represented data. As in Figure 3a, average $R^2$ are higher than the traditional methods. Also standard deviations of the $R^2$ are considerably less than the traditional. RE distributions (Figure 3b) indicate that these methods predict the capacity with less RE, while the classical methods yield a higher RE.

Predicted capacities with Schmertmann and de Ruiter's methods yield less accurate results compared to the tests with lower $R^2$. LCPC gives more reliable results, while RVR shows the highest prediction with mean $R^2$ of 0.842; GPR and VHGP give fairly high $R^2$ of 0.84. RF provides acceptable prediction. Therefore, we do not consider the remaining MLM in the rest of the analyses due to their poor performances.

The analyses are subsequently repeated with the same database with only 4 features ($L$, $A$, $q_{c,end}$, $q_{c,ave}$). No change is made in 5-folds or in the order of datasets. Results of four MLM are given. Performances of the selected MLM increase compared to the previous analysis. The most apparent change in the $R^2$ is observed in the RVR, whose mean $R^2$ becomes 0.87. We conclude that sleeve friction does not significantly affect the results through RF, GPR, and VHGP except for RVR.

From the foundation engineering standpoint, we find this result interesting, in that the sleeve friction, although an important parameter in predicting pile load capacity, is not highly correlated with the $Q_s$ based on the nonlinear methods used in this study, which seems to be supporting an earlier claim on that matter (Shahin 2010).

As given in Figure 4, standard deviations of the 4 MLM are less than the ones by the classical methods. Hence, we have
Figure 6. Boxplot of the 60S_5F_9ML, (a) $R^2$, (b) RE.

Figure 7. Boxplot of the 49S_5F_9ML, (a) $R^2$, (b) RE.
extended our database with 99 more test data resulting in a total of 219 with the same four features. Firstly, a 5-fold analysis on the new database is made for trial. Secondly, a 10-fold cross-validation is applied. Accuracy does not increase as much.

For the third analysis (219S_4F_4ML), results (Figure 5) show that performances decrease to some extent. While there is more scatter of data obtained from various sources that may have created some inconsistency, this may also be attributed to the absence of $f_s$. That is, although poorly correlated with $Q_u$, such a mechanics-wise important parameter in geotechnical engineering practice is not easy to discard in the prediction of $Q_u$ of CEP irrespective of the MLM used. The RF method still demonstrates its robustness with a mean $R^2$ of 0.84. Given the relatively large database, RF provides the best capability compared to the traditional methods and nonlinear MLM. However, one can argue that as the database is extended with more consistent data, RVR, GPR and VHGPR might ultimately improve.

5.2. OEP database

The second database contains 60 tensile tests applied on open-ended steel pipe piles embedded in cohesionless soils (Table 2). The analysis of the OEP database is performed on two sets of data. Firstly, the shaft capacity of all piles in the OEP database are predicted, where the performance of MLM and conventional methods is compared. Then the outliers with a relatively larger capacity are eliminated to provide consistency among the data, whose count is now 49. First analysis is called 60S_5F_9ML, and the second 49S_5F_9ML. The results of the first analysis are given in Table 5 and its boxplot in Figure 6. We limit the tabular presentation for this set only for the sake of brevity.

Results show that the KNNR, ELM, RVR and GPR methods provide more consistent predictions than the other MLM. RVR gives the best prediction in terms of mean $R^2$ and RE. Comparing the MLM with the empiricals, it is clear that the UWA and NGI are better than all the MLM except RVR with

Figure 8. Regression graphs of 120S_4F_4ML, (a) RF, (b) GPR, (c) RVR, (d) VHGPR.
$R^2$ of 0.96. This shows a very good correlation between the predicted friction capacities and the measured ones in the load tests. There is a significant difference between RVR and the Fugro method, while it is closer in $R^2$ to the UWA and NGI
Figure 10. Regression graph of RVR and traditional methods, ($R^2 = 0.91$-RVR, 0.72-Schmertmann, 0.74-DeRuiter, 0.84-LCPC).

Figure 11. Regression graphs of 495_SF_9ML, (a) MARS, (b) RF, (c) GPR, (d) VHGPR, (all folds are used as testing data).
methods. Besides, per Figure 6, standard deviation of the $R^2$ values is less than the other MLM and the conventional ones. Thus, our results indicate that RVR provides the best prediction for 60S_5F_9ML analysis.

The second set of data used in 49S_5F_9ML analysis is prepared by eliminating 11 pile test data with long and large diameter piles since these piles have considerably large shaft capacities causing an outlier behaviour. The results are given in Figure 7. Eliminating the outliers, the predictions of UWA, NGI and Fugro dramatically decrease, which concludes that these methods work better in predicting the large diameter piles with higher shaft capacities. The mean $R^2$ for MARS, RF, GPR and VHGPR methods are higher than the other MLM and empirical methods with lower RE. VHGPR method provides the best prediction for the 49S_5F_9ML analysis with a higher mean $R^2$ of 0.89.

5.3. Model verification

The scatter of actual load test results versus the predicted results along with empirical data is compared. For CEP database, although 219S_4F_4ML set gives the highest $R^2$ of 0.96, since the mean values are not increased significantly in the expanded database, we focus our attention on 49S_5F_5ML. MARS, RF, GPR and VHGPR stand out due to their relatively high mean $R^2$. Figure 11 shows the actual shaft capacities of piles with respect to the predicted values. VHGPR method makes the best predictions with 0.89 $R^2$ and 0.24 RE. The predictions are too close to actual capacities when folds 1, 2, 3, and 5 are used for training again. Regression plots are given in Figure 12.

6. Conclusions

In this study, static load capacities of closed- and open-ended piles in both cohesive and cohesionless soils are predicted using machine learning. Nine of such methods are comparatively used. A comprehensive database is generated for the CEP, and a moderate database is gathered for the OEP based on static load test and CPT data. Databases are then used to teach these algorithms using five features associated with pile geometry and CPT. Generally nonlinear methods perform better than the linear ones particularly the RF, RVR, GPR, and VHGPR. Among these four, the RVR performs well for the CEP and VHGPR predicts well for the OEP compared to empirical and nonlinear. RF is considered as the one with the most capability of accurately predicting the friction capacity of both types of piles. In more details;

- For the CEP database, although 219S_4F_4ML set gives the highest $R^2$ of 0.96, 120S_4F_4ML presents a more consistent dataset.
- RVR gives the highest $R^2$ value of 0.91 and the lowest RE of 0.27 compared to the classical methods.
- For the OEP database, the UWA and NGI are better than all the other MLM except RVR. This shows a very good correlation between the predicted friction capacities and the measured ones.
- For the database of 49S_5F_5ML, MLM provide a stronger correlation than the empirical methods.
Thus, given a large enough database, it is possible to predict the static pile load capacity of both CEP and OEP with sufficient accuracy, particularly when a specific set of MLM is used. While this study demonstrates the efficacy of ML methods in predicting pile load capacity, it is essential to acknowledge the limitations inherent in our approach. One of the primary limitations is the quality of the published data. Although we made significant efforts to compile a comprehensive database considering data quality, the inclusion of diverse data sources and the possibility of errors or missing parameters in the collected data could introduce variability and hence, impact the accuracy of the predictions.

Furthermore, as with any modelling technique, bias can exist in ML algorithms. Despite our efforts to mitigate bias through data collection from multiple sources and the careful selection of input parameters, there may still be inherent bias within the models. Therefore, it is important to recognize the potential impact of bias and exercise caution when applying our findings to different contexts or datasets. Moreover, generalizing a single ML algorithm across various geotechnical engineering problems is challenging due to site-specific conditions, geological variability, and other factors unique to each scenario. While these complexities can affect the performance and applicability of the models, utilizing robust ML methods for predicting pile load-bearing capacity with a comprehensive database containing a variety of pile types and different soil conditions helps us reduce such variability.

In conclusion, by addressing these limitations and considering the potential risks associated with bias and generalizability, the application of machine learning can be refined and enhanced in predicting pile load capacity. Continued efforts to improve data quality, mitigate bias, and develop specialized models for specific site conditions will contribute to the advancement of geotechnical engineering practice.

Disclosure statement
No potential conflict of interest was reported by the author(s).

ORCID
M.B.C. Ulker (http://orcid.org/0000-0001-7632-2303

References
Abu-Farsakh, M. Y., and H. H. Titii. 2004. “Assessment of Direct Cone Penetration Test Methods for Predicting the Ultimate Capacity of Friction Driven Piles.” Journal of Geotechnical and Geoenvironmental Engineering 130 (9): 935–944. https://doi.org/10.1061/(ASCE)1090-0241.
Alkroosh, I. S., M. Bahadori, H. Nikraz, and A. Bahadori. 2015. “Regressive Approach for Predicting Bearing Capacity of Bored Piles from Cone Penetration Test Data.” Journal of Rock Mechanics & Geotechnical Engineering 7 (5): 584–592. https://doi.org/10.1016/j.jrmge.2015.06.011.
Alkroosh, I., and H. Nikraz. 2011. “Correlation of Pile Axial Capacity and CPT Data Using Gene Expression Programming.” Geotechnical and Geological Engineering 29 (5): 725–748. https://doi.org/10.1007/s10706-011-9413-1.
Alkroosh, I., and H. Nikraz. 2012. “Predicting Axial Capacity of Driven Piles in Cohesive Soils Using Intelligent Computing.” Engineering Applications of Artificial Intelligence 25 (3): 618–627. https://doi.org/10.1016/j.engappai.2011.08.009.
Alsammam, O. M. 1995. "The Use of CPT for Calculating Axial Capacity of Drilled Shafts." University of Illinois at Urbana-Champaign. ProQuest Dissertations & Theses Global (304199126). https://www.proquest.com/dissertations-theses/use-cpt-calculating-axial-capacity-drilled-shafts/docview/304199126/se-2.
Alzoubi, A. K., and F. Ibrahim. 2019. “Predicting Loading–Unloading Pile Static Load Test Curves by Using Artificial Neural Networks.” Geotechnical and Geological Engineering 37 (3): 1311–1330. https://doi.org/10.1007/s10706-018-0687-4.
Ardalan, H., A. Eslamipour, and N. Nariman-Zadeh. 2009. “Piles Shaft Capacity from CPT and CPTu Data by Polynomial Neural Networks and Genetic Algorithms.” Computers and Geotechnics 36 (4): 616–625. https://doi.org/10.1016/j.compgeo.2008.09.003.
Banimahd, M., S. S. Yasrobi, and P. K. Woodward. 2005. "Artificial Neural Network for Stress–Strain Behavior of Sandy Soils: Knowledge Based Verification." Computers and Geotechnics 32 (5): 377–386. https://doi.org/10.1016/j.compgeo.2005.06.002.
Breiman, L. 2001. “Random Forests.” Machine Learning 45 (1): 5–32. https://doi.org/10.1023/a:1010933404324.
Bustamante, M., and L. Gianselli. 1982. Pile Bearing Capacity Prediction by Means of Static Penetrometer CPT, in: Proceedings of the Second European Symposium on Penetration Testing ESOPT-2. Amsterdam, May 24–27, pp. 493–500. https://www.cpt-robertson.com/PublicationsPDF/PileBearingCap%2C2CPT-LCPMethod-Bustamante-ESOPT2-198.pdf.
Cao, M.-T., N.-M. Nguyen, and W.-C. Wang. 2022. “Using an Evolutionary Heterogeneous Ensemble of Artificial Neural Network and Multivariate Adaptive Regression Splines to Predict Bearing Capacity in Axial Piles.” Engineering Structures 268:114769. https://doi.org/10.1016/j.engstruct.2022.114769.
Chow, F. C., R. J. Jardine, F. Brucy, and J. C. Nauroy. 1998. “Effects of Time on Capacity of Pipe Piles in Dense Marine Sand.” Journal of Geotechnical and Geoenvironmental Engineering 124 (3): 254–264. https://doi.org/10.1061/(ASCE)1090-0241(1998)124:3(254).
Clausen, C. J. F., P. M. Aas, and K. Karlsrud. 2005. Bearing capacity of driven piles in sand, the NGi approach. In Proceedings of International Symposium on Frontiers in Offshore Geotechnics, Perth, 574–580. Retrieved from https://www.researchgate.net/publication/290599390_Bearing_capacity_of_driven_piles_in_sand_the_NGi_approach.
Dağdeviren, Ü., and B. Kaymak. 2018. “Yapay arı koloni algoritması kullanılarak betonarme istiğim duvarlarının optimum maliyeti tasarımı etkileyen parametrelerin incelenmesi.” Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi 33.” Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi 33 (1). https://doi.org/10.17341/gazimfd.406796.
Debnath, P., and A. K. Dey. 2018. “Prediction of Bearing Capacity of Geogrid-Reinforced Stone Columns Using Support Vector Regression.” International Journal of Geomechanics 18 (2): 1–15. https://doi.org/10.1061/(ASCE)GM.1943-5622.0001067.
De Ruiter, J., and F. L. Beringen. 1979. “Pile Foundations for Large North Sea Structures.” Marine Geotechnology 3 (3): 267–314. https://doi.org/10.1080/0641197990379805.
Ebrahimian, B., and V. Movahed. 2017. “Application of an Evolutionary-Based Approach in Evaluating Pile Bearing Capacity Using CPT Results.” Ships and Offshore Structures 12 (7): 937–953. https://doi.org/10.1744532015.1116243.
Eslami, A. 1996. “Bearing Capacity of Piles from Cone Penetration Test Data.” University of Ottawa.” University of Ottawa Theses Collection. https://ruor.uottawa.ca/handle/10393/9922.
Eslami, A., and B. H. Fellenius. 1997. “Pile Capacity by Direct CPT and CPTu Methods Applied to 102 Case Histories.” Canadian Geotechnical Journal 34 (6): 886–904. https://doi.org/10.1139/cgj-34-6-886.
Fellenius, B. H. 2001. “What Capacity Value to Choose from the Results a Static Loading Test.” Fulcrum - the Newsletter of the Deep Foundations Institute 19–22. https://www.fellenius.net/papers/230%20&%20240%20Analysis%20of%20Pipe%20Capacity-DIF.pdf.
Friedman, J. H. 1991. “Multivariate Adaptive Regression Splines. The Annals of Statistics.” The Annals of Statistics 19 (1). https://doi.org/10.1214/aos/1176347963.
Using 2022 Regression Selection. Cohesionless 841–848. Conference Water 55:91–102. Vector 2271–2291. Learning https://doi.org/10.1061/(ASCE)1090-0241(2006)132:5(661) Engineering Applications 489–501. Pollination-Optimized Bearing I. M., B. A. Prediction H., N.-D. and G. K. Characteristic K., and G. M. and S. Deswal. 2010. “Modelling Pile Capacity Using Gaussian Process Regression.” Computers and Geotechnics 37 (7–8): 942–947. https://doi.org/10.1016/j.compgeo.2010.07.012

Pham, T. A., V. Q. Tran, and W. Muntau. 2022. “Developing Random Forest Hybridization Models for Estimating the Axial Bearing Capacity of Pile.” PLOS ONE 17 (3): e0265747. https://doi.org/10.1371/journal.pone.0265747.

Pham, T. A., V. Q. Tran, H.-L. T. Vu, and H.-B. Ly. 2020. “Design Deep Neural Network Architecture Using a Genetic Algorithm for Estimation of Pile Bearing Capacity.” PLOS ONE 15 (12): e0243030. https://doi.org/10.1371/journal.pone.0243030.

Reich, Y. 1997. “Machine Learning Techniques for Civil Engineering Problems.” Computer-Aided Civil and Infrastructure Engineering 12 (4): 295–310. https://doi.org/10.1111/1088-9507.00065.

Sadrossadat, E., F. Soltani, S. M. Mousavi, S. M. Marandi, and A. H. Alavi. 2013. “A New Design Equation for Prediction of Ultimate Bearing Capacity of Shallow Foundation on Granular Soils.” Journal of Civil Engineering and Management 19 (Supplement_1): S78–S90. https://doi.org/10.3843/13923730.2013.801902.

Samui, P. 2008. “Prediction of Friction Capacity of Driven Piles in Clay Using the Support Vector Machine.” Canadian Geotechnical Journal 45 (2): 288–295. https://doi.org/10.1139/T07-072.

Schmertmann, J. H. 1978. Guidelines for Cone Penetration Test (Performance and Design), FHWA-TS-78-209. Washington: United States Federal Highway Administration. https://rosap.nhtl.bts.gov/view/dot/958.

Schneider, J. A. 2007. “Analysis of Piezocene Data for Displacement Pile Design.” The University of Western Australia. https://research-repository.uwa.edu.au/en/publications/analysis-of-piezocene-data-for-displacement-pile-design.

Shahin, M. A. 2010. “Intelligent Computing for Modeling Axial Capacity of Pile Foundations.” Canadian Geotechnical Journal 47 (2): 230–243. https://doi.org/10.1139/T10-048.

Shahin, M. A. 2016. “State-Of-The-Art Review of Some Artificial Intelligence Applications in Pile Foundations.” Geoscience Frontiers 7 (1): 33–44. https://doi.org/10.1016/j.gsf.2014.10.002.

Shahin, M. A., H. R. Maier, and M. B. Jaksa. 2002. “Predicting Settlement of Shallow Foundations Using Neural Networks.” Journal of Geotechnical and Geoenvironmental Engineering 128 (9): 785–793. https://doi.org/10.1061/(ASCE)1090-0241.

Teh, C. I., K. S. Wong, A. T. C. Goh, and S. Jariingam. 1997. “Prediction of Pile Capacity Using Neural Networks.” Journal of Computing in Civil Engineering 11 (2): 129–138. https://doi.org/10.1061/(ASCE)0887-3801(1997)11:2(129).

Tipping, M. E. 2000. “The Relevance Vector Machine.” Advances in Neural Information Processing Systems 12 (NIPS 1999), Cambridge, MA, 652–658. MIT Press. https://papers.nips.cc/paper/1999/hash/3f144cefe89ad6a6a1afaf7859c5076b-Abstract.html.

Yang, Z., R. Jardine, W. Guo, and F. Chow. 2015. A Comprehensive Database of Tests on Axially Loaded Piles Driven in Sand, a Comprehensive Database of Tests on Axially Loaded Piles Driven in Sand. London: Academic Press.

Yurtçu, Ş., and A. Özçak. 2016. “İnçe daniel zeminlerde sıkışma indisi in istatistiksel ve yapay zeka yöntemleri ile tahmin edilmesi.” Gazette Universitatis Muhendislik-Mimarlik Fakultesi Dergisi 31 (3). https://doi.org/10.17341/gumfmd.95986.

Zhu, J. H., M. M. Zaman, and S. A. Anderson. 1998. “Modeling of Soil Behavior with a Recurrent Neural Network.” Canadian Geotechnical Journal 35 (5): 858–872. https://doi.org/10.1139/98-042.