
# Use of ML techniques for predicting the bearing capacity of piles and its relative errors

Yago F. Gomes[1], Dimas B. Ribeiro[1]

[1]*Dept. of Civil Engineering, Instituto Tecnológico de Aeronáutica*
*Praça Marechal Eduardo Gomes, 50, Vila das Acácias, 12.228-900, São José dos Campos, São Paulo, Brasil*
*yago@ita.br, dimasbetioli@gmail.com*

**Abstract.**

This work presents an application of machine learning techniques for estimating the bearing capacity of pile foundations and the relative error from these techniques. It uses as raw data 165 load tests results associated with SPT soundings, taken from several Brazilian regions. A dataset based on the inputs from Decourt-Quaresma and Meyerhof semi-empirical methods was created and applied to several machine learning techniques with a leave-one-out cross validation approach for training and testing the algorithms. Using the results obtained from each model, the metrics RMSE and $R^2$ were calculated through a stacking strategy. The Random Forest technique presented the best performance for both bearing capacity (RMSE = 640,26) and relative error ($R^2 = 0.77$) prediction problems. The other five ML techniques performance overcame the semi-empirical methods, which obtained an RMSE close to 900, indicating the potential of these tools. Then, the errors obtained from the predictions were used to propose a new machine learning problem, aiming to predict the error of new examples. Although the preliminary results were not accurate, the authors believe that the study justifies further investigations.

**Keywords:** Standard penetration test, Pile capacity, Machine learning

## 1 Introduction

Geotechnical engineers need to predict the bearing capacity of piles to design safe constructions, considering the expected loads and resistances that will take place at the real construction. The most accurate method is measuring the capacity of one or more piles in the construction site, using standards as ABNT [1]. Although these experiments are mandatory in most cases, they are not useful in the first phases of the designing process, when only basic information like SPT results are available for design. In these situations, geotechnical engineers seek SPT based methods like Aoki and Velloso [2], Décourt and Quaresma [3] and Meyerhof [4] to obtain reasonable predictions.

The number of studies that use machine learning (ML) techniques in geotechnical problems has increased in the last years. One can find researches in areas like soil classification (Neto et al. [5], Carvalho and Ribeiro [6]), slope stability (Bui et al. [7], Maxwell et al. [8]) and soil liquefaction (Livingston et al. [9], Kohestani et al. [10]). Although one can find works dedicated to predicting pile capacity using ML kike Jesswein and Liu [11] and Pham et al. [12], many gaps still remain in this research area. The main limitations of the available studies are the use of only one or two ML techniques and restricting the dataset to very specific soil types, most cases all taken from the same site.

This work presets an approach for predicting the bearing capacity of precast concrete piles and the expected errors using machine learning techniques. The used dataset includes information from 165 static load tests with the corresponding SPT tests, taken from different regions of Brazil. The inputs of Meyerhof and Décourt Quaresma semi-empirical methods are the reference for assembling the dataset, which combines the inputs of these semi-empirical methods. The study starts training several machine learning techniques to predict the bearing capacity, computing the errors of each prediction. Second, it uses the found errors as inputs to train the techniques again, aiming to predict these errors. The machine learning techniques presented reasonable results for the first part, surpassing linear regression and the semi-empirical techniques, used as baselines. Nonetheless, only preliminary

results were achieved for the second part. In spite of that, the authors believe that further investigations are justified.

## 2 Dataset and ML procedure

The proposed dataset uses the inputs of the semi-empirical methods of Meyerhof [4] and Décourt and Quaresma [3]. Both methods use pile diameter $D$ and length $L$, a SPT average for the pile length and a SPT average for the pile tip. The difference between them is how they calculate these average SPT values. Based on preliminary tests, this work uses a combination of these two sets, totalizing 6 inputs: $D$, $L$, both average SPT values for pile length and both average SPT values for pile tip. The output of each example is the pile capacity obtained from a static load test, performed according to ABNT [1] and using the Van der Veen method (Van Der Veen [13]). The information used to calculate all values was taken from Lobo [14], Vianna [15] and Santos Jr. [16], totalizing 165 examples.

First step is organizing the dataset as a matrix, where each line represents an example and columns represent the inputs and the output. In other words, it becomes a $165 \times 7$ matrix. Next, the procedure divides the dataset in two parts, composing the training and test datasets. It trains the ML techniques using the training dataset and test them later with the separated test dataset. This approach seeks to evaluate what would be the performance of the techniques when subjected to completely new data. In this study the leave-one-out cross validation approach is employed. This means that the full dataset is used for training, while one example is separated for test. After testing all examples, the average accuracy gives the overall performance of the model [17].

One disadvantage of this approach is that the complete dataset can be underrepresented by the test dataset. One way of minimizing this problem is to divide the dataset into folds, separating one of them for test and using the remaining ones for training. The procedure repeats the calculation of performance using each fold as test dataset and the mean performance considering all folds gives the final performance of the model. This is the so called cross validation method. Increasing the number of folds tends to improve accuracy and increase computational cost. This work employs the limit case in which the number of folds is the number of examples, which leads to the so called leave-one-out technique. This approach is possible only with relatively small datasets, which is the case of this work.

## 3 Predicting models

This work uses two semi-empirical methods, six machine learning techniques and a multiple linear regression to predict the bearing load capacity of piles. The next sections present a brief description of each one of these techniques.

### 3.1 Semi-empirical methods

Décourt and Quaresma [3] and Meyerhof [4] describe the semi-empirical methods selected for this study. They are both based on the concept that the total resistance of a pile $R_t$ is the sum of its lateral resistance and its tip resistance. The Décourt-Quaresma method was chosen for its popularity in Brazil, and the Meyerhof method for being well known in engineering practice around the world.

For Décourt and Quaresma [3]:

$$R_t = \alpha.K.SPT_p.A_p + U.\beta.10\left(\frac{SPT_l}{3} + 1\right)L\ . \tag{1}$$

where $K$ depends on soil type, $\alpha$ and $\beta$ are parameters to be calibrated, $A_p$ is the cross sectional area of the pile, $U$ is its perimeter, $SPT_p$ is the average SPT index at the pile tip and $SPT_l$ is the average SPT index at the pile length.

For Meyerhof [4]:

$$R_t = A_p.q_p + U.L.q_s\ . \tag{2}$$

where

$$q_p = 40.L\frac{SPT_p}{D} \leq 400.SPT_p \ . \tag{3}$$

and

$$q_s = 2.SPT_l \ . \tag{4}$$

One should notice that these methods do not use the same definition for $SPT_l$ and $SPT_p$. Further detail about this difference can be consulted in Décourt and Quaresma [3] and Meyerhof [4].

### 3.2 K nearest neighbor (KNN) and kernel k nearest neighbor (KKNN)

The first step to use the KNN technique is creating a space where each coordinate represents an input of the dataset and a point represents each example. One should observe that this approach is sensitive to data scale, which requires normalizing all input data to a standard range. KNN uses the concept that points that are close in this input space should be similar. In this context, one way of estimating the output of an unknown example within a cloud of known examples is to equal it to the output of its closest neighbor.

There are several ways of improving the accuracy of this technique. The first is not using only one neighbor to predict the new example, but the average of two or more nearest neighbors. In most cases, weighting the output of these neighbors with respect to its distance of the new example also leads to better accuracy (Dudani [18]).

One disadvantage of KNN is that it usually presents poor performance for complex problems. Kuo et al. [19] proposes a solution to this problem mapping the examples of the input space to a higher dimensional space, so that they become easier to predict. The KKNN technique uses functions called kernels for this mapping.

### 3.3 Decision tree (DT) and random forest (RF)

A DT works as a one-directional graph that starts with a root node, that splits the examples of the dataset to other nodes using a rule. This rule is usually based on inequalities applied to one of the inputs. If a node receives part of the examples and splits it to others using a new rule, it is called decision node. If it receives examples and finishes the procedure assigning outputs to them, it is a leaf node.

One problem of decision trees is that they tend to become overspecialized in the training dataset. This means that, although the training stage presents high performance, poor performance occurs if the tree is applied to new data. This is the so called overfit. One alternative to minimize this problem is using RF, which creates many trees selecting different inputs for training each one of them. Ho [20] presents a detailed description of RF.

### 3.4 Artificial neural networks (ANN)

ANN is a mathematical model inspired by the human brain. It is usually composed by layers, which are composed by units called neurons. A neuron receives one or more signals, multiplies each one of them by a weight and sums the results. The neuron discounts a threshold and applies an activation function to the obtained value. The output signal can be the final prediction of the ANN or an input to other neurons (Moselhi et al. [21]).

The basic architecture of an ANN has one input layer, one or more hidden layers and one output layer. The input layer contains one neuron for each input and, in regression problems, the output layer contains a single neuron that gives the prediction of the ANN. One way of obtaining a good architecture is testing different numbers of hidden layers and neurons in each one of them, subjected to computational cost limitations.

### 3.5 Support vector machines (SVM)

An SVM model uses a space where coordinates are inputs of the dataset and points represent examples, similarly to the KNN. The basic idea is using statistical learning principles to propose a hyperplane that represents predictions and is optimized with respect to its distance to the known examples. The model tries to define a margin around the hyperplane to contain all known examples, with smaller margins meaning smaller errors.

Given that no solution is possible in most cases, usually it is necessary to smooth the margin so that some points are allowed outside. Another limitation of this approach is that the resulting model is linear. One way of

solving this problem is using kernels to map points of the input space into a higher dimensional space, similarly to the strategy used by KKNN.

### 3.6 Multiple linear regression (LR)

This work uses LR as one of the baselines for the performance of the ML techniques. It uses a linear function to predict the output from the inputs, as presented in eq. (5):

$$\hat{y} = \beta_0 + x_1.\beta_1 + x_2.\beta_2 + ... + x_n.\beta_n \ . \tag{5}$$

where $\hat{y}$ is the prediction, $x_i$ are inputs and $\beta_i$ are coefficients to be calibrated.

## 4 Results and discussions

As presented in previous sections, this work uses a dataset that includes 6 inputs to predict the bearing capacity of precast concrete piles. Table 1 presents a sample of this dataset. The average value of the pile capacity for the dataset is approximately 1360 KN.

Table 1. Sample of the used dataset

| N | $SPT_{lD}$ | $SPT_{lM}$ | $SPT_{pD}$ | $SPT_{pM}$ | $L$ | $D$ | $Q_u$ |
|---|---|---|---|---|---|---|---|
| 1 | 6.46 | 4.16 | 27.33 | 27.33 | 18.90 | 26.00 | 1115.00 |
| 2 | 7.94 | 6.04 | 27.33 | 27.33 | 21.12 | 26.00 | 1005.00 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 165 | 22.71 | 12.80 | 42.25 | 51.67 | 7.00 | 40.00 | 1800.00 |

The study used this dataset to train and test all ML techniques. Results are compared to LR and to the semi-empirical methods of Décourt and Quaresma [3] and Meyerhof [4]. Table 2 presents the results.

Table 2. RMSE and $R^2$ obtained when calculating the bearing capacity of piles

| Technique | $R^2$ | RMSE |
|---|---|---|
| RF | 0.770 | 640 |
| KNN | 0.761 | 662 |
| Décourt and Quaresma [3] | 0.748 | 910 |
| ANN | 0.746 | 668 |
| SVM | 0.743 | 688 |
| KKNN | 0.730 | 692 |
| LR | 0.724 | 696 |
| DT | 0.719 | 701 |
| Meyerhof [4] | 0.662 | 896 |

Table 2 orders the values with respect to $R^2$. RF was the technique that showed best performance for both RMSE and $R^2$, followed by KNN. One can observe that, theoretically, KNN was expected to have lower performance than KKNN, but this was not observed in practice. This shows that the mapping proposed in KKNN to improve accuracy does not suit this type of problem. Considering RMSE alone, the semi-empirical techniques presented the worst performance. Nonetheless, Décourt and Quaresma [3] presented the third best performance with respect to $R^2$. Notice that DT presented a relatively poor performance, worse than LR which is one of the baselines of the study. The tendency of DT to overfit can explain this behavior, which is remedied by the random input selection used by RF. It is also interesting to observe that the semi-empirical methods failed to surpass even

LR with respect to RMSE. In the end, Table 2 shows that ML techniques are capable of modeling this type of problem.

The second part of this study presents an evaluation of the expected errors for RF, which presented the best performance in Table 2. The procedure uses the relative error $RE$ of each prediction $P$ of the RF algorithm, with respect to the measured load bearing capacity of the pile $Q_u$, as presented in eq. (6):

$$RE = \frac{|Pr - Q_u|}{Q_u}. \tag{6}$$

Next, the procedure updates the dataset, substituting the output $Q_u$ by the relative errors calculated using eq. (6). Table 3 presents a sample of this new dataset.

Table 3. Sample of the new dataset

| N | $SPT_{lD}$ | $SPT_{lM}$ | $SPT_{pD}$ | $SPT_{pM}$ | $L$ | $D$ | $RE$ |
|---|---|---|---|---|---|---|---|
| 1 | 6.46 | 4.16 | 27.33 | 27.33 | 18.90 | 26.00 | 0.131 |
| 2 | 7.94 | 6.04 | 27.33 | 27.33 | 21.12 | 26.00 | 0.602 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 165 | 22.71 | 12.80 | 42.25 | 51.67 | 7.00 | 40.00 | 0.060 |

The idea is to propose a new machine learning regression using the same inputs, but the errors as outputs. The study tested some combinations of inputs, but achieved only preliminary results with $R^2$ close to zero in all cases. Nevertheless, the authors believe that some alternatives can still be investigated, like testing other machine learning techniques and other combinations of inputs. One explanation for this result is the low precision of the SPT, as reported in the literature [22]. Thus, high variability is expected for the large dataset used. Lower errors would be expected if the techniques were calibrated with load tests taken from the same region.

## 5 Conclusions

This work proposed a methodology for predicting the bearing capacity of piles using machine learning techniques. It used inputs from classical methods from the literature and results from static load tests as outputs. The used techniques presented reasonable results when compared to baselines like linear regression and the original semi-empirical methods. The latter were surpassed by all other techniques, presenting RMSE close to 900 while RF presented RMSE$= 640$. The second part of the study aimed at predicting the error of the machine learning techniques and achieved only preliminary results. Although none of the tests achieved reasonable accuracy, the authors believe that further investigation is justified, including other input combinations and techniques. In the end, the obtained results show that the use of machine learning techniques is promising for predicting the bearing capacity of piles. Thus, the authors recommend future research investigating the prediction of errors using machine learning. One suggestion is correcting SPT with energy calibration specially for the Meyerhof method, once Brazilian equipment was used for the SPT.

**Acknowledgements.** To the Coordination of Improvement of Higher Education Personnel (CAPES) for granting the first author's scholarship.

**Authorship statement.** The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

## References

[1] ABNT. Projeto e execução de fundações. In *Assiciação Brasileira de Normas Técnicas*. ABNT, Rio de Janeiro, Brazil, 2010.

[2] N. Aoki and D. Velloso. An approximate method to estimate the bearing capacity of piles. In *Proc., 5th Pan-American Conf. of Soil Mechanics and Foundation Engineering*, pp. 367–376, Buenos Aires, Argentina. International Society of Soil Mechanics and Geotechnical Engineering, 1975.
[3] L. Décourt and A. Quaresma. Capacidade de carga de estacas a partir de valores de spt. In *Congresso Brasileiro de Mecânica dos Solos e Engenharia de Fundações*, pp. 45–53. Associação Brasileira de Mecânica dos Solos e Engenharia Geotécnica, 1978.
[4] G. Meyerhof. Bearing capacity and settlements of pile foundations. *Journal of Geotechnical Engineering*, vol. 102, pp. 197–228, 1976.
[5] A. B. Neto, C. d. S. B. Bonini, B. S. Bisi, dos A. R. Reis, and L. F. S. Coletta. Artificial neural network for classification and analysis of degraded soils. *IEEE Latin America Transactions*, vol. 15, n. 3, pp. 503–509, 2017.
[6] L. O. Carvalho and D. B. Ribeiro. Soil classification system from cone penetration test data applying distance-based machine learning algorithms. *Soils and Rocks*, vol. 42, n. 2, pp. 167–178, 2019.
[7] D. T. Bui, H. Moayedi, M. Gör, A. Jaafari, and L. K. Foong. Predicting slope stability failure through machine learning paradigms. *ISPRS International Journal of Geo-Information*, vol. 8, n. 9, pp. 395, 2019.
[8] A. E. Maxwell, M. Sharma, J. S. Kite, K. A. Donaldson, J. A. Thompson, M. L. Bell, and S. M. Maynard. Slope failure prediction using random forest machine learning and lidar in an eroded folded mountain belt. *Remote Sensing*, vol. 12, n. 3, pp. 486, 2020.
[9] G. Livingston, M. Piantedosi, P. Kurup, and T. G. Sitharam. An approximate method to estimate the bearing capacity of piles. In *Geotechnical Earthquake Engineering and Soil Dynamics IV Congress*, pp. 1–10. ASCE, 2008.
[10] V. R. Kohestani, M. Hassanlourad, and A. Ardakani. Evaluation of liquefaction potential based on cpt data using random forest. *Natural Hazards*, vol. 79, n. 2, pp. 1079–1089, 2015.
[11] M. Jesswein and J. Liu. Predicting the side resistance of piles using a genetic algorithm and spt n-values. In *Proceedings of the $71^{st}$ Canadian Geotechnical Conference and the $13^{th}$ Joint CGS/IAH-CNC Groundwater Conference*, pp. 1–8, Edmonton, Alberta, Canada. GeoEdmonton 2018, 2018.
[12] T. Pham, H. Ly, V. Tran, L. Giap, H. Vu, and H. Duong. Prediction of pile axial bearing capacity using artificial neural network and random forest. *Applied Sciences*, vol. 10, n. 5, pp. 1871, 2020.
[13] C. Van Der Veen. The bearing capacity of pile. In *International Conference on Soil Mechanics and Foundation Engineering*, pp. 367–376, 1953.
[14] B. Lobo. *Método de previsão de capacidade de carga de estacas: aplicação dos conceitos de energia do ensaio SPT*. PhD thesis, Universidade Federal do Rio Grande do Sul, Escola de Engenharia, Porto Alegre, Brasil, 2005.
[15] A. Vianna. Análise de provas de carga estática em estacas pré-moldadas cravadas na cidade de curitiba e região metripolitana. M.S. thesis, Universidade de São Paulo, São Carlos, Brasil, 2000.
[16] O. Santos Jr. Previsão do comportamento carga-recalque de estacas pré-moldadas de concreto. M.S. thesis, Universidade de São Paulo, São Carlos, Brasil, 1988.
[17] T. Wong. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, vol. 48, n. 9, pp. 2839–2846, 2015.
[18] S. A. Dudani. The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-6, n. 4, pp. 325–327, 1976.
[19] B. Kuo, J. Yang, T. Sheu, and S. Yang. Kernel-based knn and gaussian classifiers for hyperspectral image classification. In *2008 International Geoscience and Remote Sensing Symposium*, pp. 1006–1008. IEEE, 2008.
[20] T. K. Ho. Random decision forests. In *Proceedings of $3^r d$ International Conference on Document Analysis and Recognition*, volume 1, pp. 278–282, 1995.
[21] O. Moselhi, T. Hegazy, and P. Fazio. Potential applications of neural networks in construction. *Canadian Journal of Civil Engineering*, vol. 19, n. 3, pp. 521–529, 1992.
[22] G. B. Baecher and J. T. Christian. *Reliability and statistics in geotechnical engineering*. John Wiley & Sons, 2005.