*Article*

# Factors Influencing Pile Friction Bearing Capacity: Proposing a Novel Procedure Based on Gradient Boosted Tree Technique

Chia Yu Huat [1], Seyed Mohammad Hossein Moosavi [1], Ahmed Salih Mohammed [2],
Danial Jahed Armaghani [3,*], Dmitrii Vladimirovich Ulrikh [3], Masoud Monjezi [4] and Sai Hin Lai [1,*]

1   Department of Civil Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia;
    17107717@siswa.um.edu.my (C.Y.H.); mh.moosavi@um.edu.my (S.M.H.M.)
2   Civil Engineering Department, College of Engineering, University of Sulaimani, Sulaymaniyah 46001, Iraq;
    ahmed.mohammed@univsul.edu.iq
3   Department of Urban Planning, Engineering Networks and Systems, Institute of Architecture and
    Construction, South Ural State University, 454080 Chelyabinsk, Russia; ulrikhdv@susu.ru
4   Department of Mining, Faculty of Engineering, Tarbiat Modares University, Tehran 14115-143, Iran;
    monjezi@modares.ac.ir
*   Correspondence: danialarmaghani@susu.ru (D.J.A.); laish@um.edu.my (S.H.L.)

**Abstract:** In geotechnical engineering, there is a need to propose a practical, reliable and accurate way for the estimation of pile bearing capacity. A direct measure of this parameter is difficult and expensive to achieve on-site, and needs a series of machine settings. This study aims to introduce a process for selecting the most important parameters in the area of pile capacity and to propose several tree-based techniques for forecasting the pile bearing capacity, all of which are fully intelligent. In terms of the first objective, pile length, hammer drop height, pile diameter, hammer weight, and N values of the standard penetration test were selected as the most important factors for estimating pile capacity. These were then used as model inputs in different tree-based techniques, i.e., decision tree (DT), random forest (RF), and gradient boosted tree (GBT) in order to predict pile friction bearing capacity. This was implemented with the help of 130 High Strain Dynamic Load tests which were conducted in the Kepong area, Malaysia. The developed tree-based models were assessed using various statistical indices and the best performance with the lowest system error was obtained by the GBT technique. The coefficient of determination ($R^2$) values of 0.901 and 0.816 for the train and test parts of the GBT model, respectively, showed the power and capability of this tree-based model in estimating pile friction bearing capacity. The GBT model and the input selection process proposed in this research can be introduced as a new, powerful, and practical methodology to predict pile capacity in real projects.

**Keywords:** tree-based techniques; feature selection; pile bearing capacity; gradient boosted tree; random forest

## 1. Introduction

There are several types of deep foundations, for instance, piles and caissons, which are required in situations where the soil is not able to support structural loads at a shallow depth. The main objective of the pile foundation is to transmit the structural load to deeper bearing strata in order to withstand the axial, lateral, and uplift load and to minimize the settlement. The load applied at the top of the pile head is transferred to the soil where the load is partially taken by normal stress at the pile base and the remaining load is taken by the lateral pile-soil interface via shear stress [1]. Thus, the piles can be classified into two types, which are end bearing pile and friction pile. The end bearing pile is a pile that transmits the structural load to a hard and incompressible stratum where the required bearing capacity is derived from end bearing at the pile base [2]. As for the friction pile, the pile-bearing capacity is derived from skin friction and cohesion between the pile surface and

the shaft that is encompassed by soil or rock along a pile [3]. Hence, the base and friction capacity of piles are crucial for carrying the axial loading. In the event of no stiff stratum at a reasonable depth, the loads are required to transfer by friction through the pile shafts [4].

The pile-bearing capacity can be governed by soil and pile properties [5,6]. The contribution of soil typically consists of cohesion and friction between the pile and the shaft of a pile at a depth. The pile friction capacity is calculated by a combination of the interface shear strength ($\tau_m$) along the pile length and the pile surface area to compute the shaft resistance ($Q_{su}$) [7]. In addition, during the installation of driven piles, which are usually prefabricated, the loose deposit soil that encompasses the pile can be locally densified due to soil displacement and, thus, the pile capacity can be increased [7]. As such, it can be stated that the installation method of piles can be one of the factors that contributes to the pile capacity [8].

The pile-bearing capacity can be determined using several techniques, such as empirical, semi-empirical and finite element (FE). The extent of the FE model and computation time is limited with model boundaries in contexts where this can be done by redoing the model boundaries by taking the boundaries to be further away from the modeling object and comparing the results. This process can be more time consuming [9]. In industry practice, the Standard Penetration Test (SPT)-N is widely-used to determine the pile capacity [8,10–12]. There are many empirical formulas of pile friction bearing correlated with SPT-N in a general form of equation, as shown below:

$$q_s = n_s \, N \tag{1}$$

where, $q_s$ is the limit skin friction stress at a given depth, which is proportional to the $N$ value at the particular depth, and $n_s$ is the skin friction factor proposed by researchers as presented in Table 1. However, according to previous studies, these empirical equations are not reliable in terms of accuracy [13,14]. This is because some of the pile's empirical analysis relationships are made by simplification in contexts where a large factor of safety is applied. This factor reduces the accuracy of the predictions and the deprivation of resources [15]. Other than this, there is a simple correlation between the pile bearing capacity and in-situ tests, for instance, the Cone Penetration Test (CPT) or SPT. However, this correlation method overestimates the pile bearing capacity [16].

**Table 1.** Some of empirical equations for determining the pile friction bearing capacity using $n_s$ results.

| No | References | Equation | $n_s$ (kPa) | Type of Installation Pile |
|---|---|---|---|---|
| 1 | Bazaraa and Kurkur [17] | $q_s = 0.67\,N$ if $D \leq 0.5$ m. else $q_s = 1.34\,N$ | 0.67 if $D \leq 0.5$ m, 1.34 for the other $D$ values (where $D$ is pile diameter in m) | Bored |
| 2 | Decourt [18] | $q_s = 10\,(N/3+1)$ | - | Bored |
| 3 | Lopes and Laprovitera [19] | $q_s = 1.62\,N$ $q_s = 1.94\,N$ | 1.62 in sand 1.94 in silty sand | Bored |
| 4 | Meyerhof [20] | $q_s = 1\,N$ $q_s = 2\,N$ | 1 2 | Bored Driven |
| 5 | Shioi and Fukui [21] | $q_s = 1\,N$ $q_s = 2\,N$ | 1 2 | Bored Driven |
| 6 | Aoki and Veloso [22] | $q_s = 2\,N$ $q_s = 2.28\,N$ | 2.00 in sand 2.28 in silty sand | Bored |
| 7 | Reese and O'Neill [23] | $q_s = 3.3\,N$ | 3.3 | Bored |
| 8 | Robert [24] | $q_s = 1.9\,N$ | 1.90 1.90 | Bored Driven |

Pile tests are required during the construction process to reassure the design calculation, because the estimation of axial pile capacity at various soil types will never be more accurate than approximately 30% [25]. There are a few methods of pile tests used to calculate the axial capacity of the piles. The typical methods are Static Load Test (SLT) and High Strain Dynamic Testing (HSDT) [26]. SLT is considered as the most reliable predictor of long term pile performance. However, this testing is expensive and time consuming [27–29]. Other than the SLT test, HSDT is one of the methods used to determine the pile bearing capacity. In comparison with SLT, HSDT is quick and economical [26]. This test is carried out based on the theory of one-dimensional wave propagation and is given by a Pile Driving Analyzer (PDA). PDA has proven that the predicted bearing capacity values are closely related to SLT results [30]. Nevertheless, all pile tests are, basically, expensive and time consuming to set up at the site [31,32]. Due to the aforementioned situation, it is important to predict pile bearing capacity using new and effective calculation approaches, such as Machine Learning (ML) and Artificial Intelligence (AI).

AI, ML and data mining techniques have been used widely to solve many civil engineering and more specifically geotechnical problems [33–45]. In terms of piling related issues, such as pile bearing capacity, there are several studies that have applied and proposed AI and ML techniques [12,46–50]. One of the most-used models in this regard is the Artificial Neural Networks (ANN). These approaches demonstrated a number of successful predictions [29,50]. As discussed before, pile driving formulae were used to provide an approximate estimation of the driven pile capacity. This formula is derived from impulse-momentum principles. However, the accuracy of neural network predictions are significantly higher compared to the conventional pile driving formulae [51]. In another study, Pal [52] stated that the General Regression Neural Network (GRNN) model has shown higher accuracy of the pile load bearing capacity prediction in comparison to empirical approaches, but slightly lower accuracy than the ANN technique. In addition, the Gene Expression Programming (GEP) model was in good agreement with the results of the experiment, indicating that pile capacity has a good relationship with some inputs, such as pile geometry [53]. Alavi et al. [54] concluded that Linear Genetic Programming (LGP) model is the best behavior in modelling uplift capacity of suction caissons, followed by the GEP and tree-based genetic programming models in comparison with regression and FE models. A Gaussian Process Regression (GPR) approach was suggested by Momeni et al. [55] in the area of pile capacity after comparison with other ANN-based models. Another group of scholars applied and proposed a combination of at least two AI techniques for prediction of pile bearing capacity [26–28,31,47]. Actually, these combined techniques enjoyed the advantages of all the used AI models for prediction purposes and, due to that, they achieved higher performance compared to the single AI models. Table 2 presents the most important AI and ML studies for predicting the pile capacity, together with their soil types, number of data, model performance, and input parameters.

According to Table 2, many studies used ANN, ANN-based and genetic-based models for estimating pile capacity. In addition, several studies applied and proposed neuro-fuzzy and Support Vector Machine (SVM) models for the same purpose. However, there are very few approaches using tree-based techniques, like Random Forest (RF), to predict the pile capacity as far as the authors know. Due to this, this paper is aimed at applying and proposing the full applications of tree-based models only, i.e., Decision Tree (DT), RF and Gradient Boosted Tree (GBT) for the prediction of pile friction bearing capacity. To do this, a feature selection (i.e., input selection) will be conducted to select the most crucial input variables for pile friction bearing capacity. The mentioned models will then be constructed and the model with the highest accuracy will be selected and introduced for estimating the pile friction bearing capacity.

**Table 2.** Summary of important AI and ML studies to predict pile capacity.

| No | Reference | Input | Model | Soil Type | Datasets | Performance |
|---|---|---|---|---|---|---|
| 1 | Goh [56] and Goh [51] | Pile length and diameter, effective overburden stress and undrained shear strength | ANN | Clay | 65 driven piles (timber and steel) | R = 0.956 |
| 2 | Lee and Lee [29] | Model pile load test: penetration depth ratio, mean normal stress of calibration chamber, number of hammer blows. In-situ pile load test: penetration depth ratio, average SPT-N value with the pile shaft depth and close to the end of pile, pile set, final penetration depth/blow, hammer energy | ANN | Sand | Model pile load test: 28 steel tube. In-situ pile load test: 24 piles | Maximum error of prediction < 25% |
| 3 | Goh [57] | Pile elastic modulus, pile length and cross sectional area, weight of pile, hammer drop height and weight, pile set and type of hammer | ANN | Sand | 116 piles (timber precast concrete and steel) | R = 0.965 |
| 4 | Kiefa [58] | Soil shear resistance encompassing the pile shaft, soil shear resistance at the tip of pile, effective overburden stress, pile length and pile area | GRNN | Sand | 59 load tests on different type of driven pile | $R^2 = 0.912$ |
| 5 | Das and Basudhar [59] | Pile diameter, pile length, eccentricity of load and undrained shear strength | MFPNN | Clay | 38 short and rigid piles | R = 0.947 |
| 6 | Pal and Deswal [60] | Force, velocity multiplied by impedance, hollow pipes piles diameter, wall thickness, and pile depth | SVM and GRNN | - | 105 pre-stressed precast spun piles | R SVM = 0.964 GRNN = 0.977 |
| 7 | Pal and Deswal [61] | Pile diameter, pile length, eccentricity of load and undrained shear strength | GPR | Non-cohesive soil | 94 load tests piles (timber, precast concrete and steel piles) | R = 0.950 |
| 8 | Alavi et al. [54] | Pile length/pile diameter, lateral force point of application distance/pile length, chain force angle with the horizontal, undrained shear strength at pile tip, and soil permeability | LGP | All soil types | 62 suction caissons | $R^2 = 0.994$ |
| 9 | Gandomi et al. [62] | Pile diameter and length, eccentricity of load, and pile tip undrained shear strength | Multi-Gene GP | Clay | 38 short and rigid piles | R = 0.985 |
| 10 | Alkroosh and Nikraz [53] | Pile diameter and length, CPT pile tip resistance, cone sleeve friction along pile length, CPT resistance along pile shaft, elastic modulus of pile and type of pile | GEP | Non-cohesive soil | 25 driven piles (concrete and steel) | R = 0.94 |

**Table 2.** *Cont.*

| No | Reference | Input | Model | Soil Type | Datasets | Performance |
|---|---|---|---|---|---|---|
| 11 | Kordjazi et al. [63] | Cross section area of the tip, perimeter of pile, embedded pile length, average cone tip resistance, average cone sleeve friction, and average cone tip resistance below the pile tip | Different SVM kernels | Sands, clays and silts | 108 pile tests | R = range from 0.966–0.982 |
| 12 | Momeni et al. [26] | Pile set, pile length and cross-sectional area, hammer drop height and weight | GA-ANN | - | 50 pre-cast concrete piles | $R^2$ = 0.990 |
| 13 | Ghorbani et al. [64] | Area of the pile at tip, unit shaft resistance of the soil, average of cone tip resistance for shaft and tip, and average of sleeve friction value along the pile embedded length | ANFIS | All soil types | 108 concrete, steel and composite piles | R = 0.96 |
| 14 | Harandizadeh et al. [27] | Pile length, cross-section shape and material, cone tip resistance and sleeve friction of cone | ANFIS-GMDH-PSO | Sand, silty sand, clay, sandy clay, and silty clay | 41 Concrete and 31 Steel piles | R = 0.96 |
| 15 | Dehghanbanadaki et al. [65] | Pile area, pile length, flap number, average soil cohesion and friction angle, average soil specific weight and average pile-soil friction angle | ANFIS-GWO | - | 100 steel and concrete driven piles | R = 0.930 |
| 16 | Momeni et al. [55] | Pile length and diameter, pile set, ram weight, and hammer drop height | GPR | - | 296 precast driven piles | $R^2$ = 0.81 |
| 17 | Pham et al. [66] | Pile diameter, pile tip segment length, second pile segment length, pile top segment length, natural ground elevation, pile top elevation, guide pile segment sop driving elevation, pile tip elevation, and average SPT | RF and ANN | - | 2314 reinforced concrete piles | $R^2$ RF = 0.861 ANN = 0.811 |

Gray wolf optimization (GWO); Correlation coefficient (R); Determination coefficient ($R^2$); Genetic algorithm (GA); Adaptive neuro-fuzzy inference system (ANFIS); Group method of data handling (GMDH); Particle swarm optimization (PSO); Random forest (RF); Multilayer feedback propagation neural network (MFPNN); Support vector machine (SVM).

## 2. Methods and Material

### 2.1. Case Study and Established Database

In order to predict the pile friction bearing capacity values, there is a need to prepare a series of experiments on site. The case study in this research was located at Kepong, Malaysia and the area was about 4.4 acres in size. This site was located in the Limestone formation, as indicated in Figure 1. In Kuala Lumpur, there are many commercial centers that are built on heavily karstified limestone formations [67]. The study area was ex-mining land, a swampy area and a pond, as shown in Figure 2.

- Zone (1): Fresh water (Swamps)—The region is continuously or seasonally submerged by freshwater and commonly seen in the lower sections of rivers and near freshwater.
- Zone (2): Vegetation—The region is covered by plants.
- Zone (3): Mining Land—The region that used for the extraction of valuable minerals.
- Zone (4): Pond and Lake—The region that comprise of freshwater and living creatures.
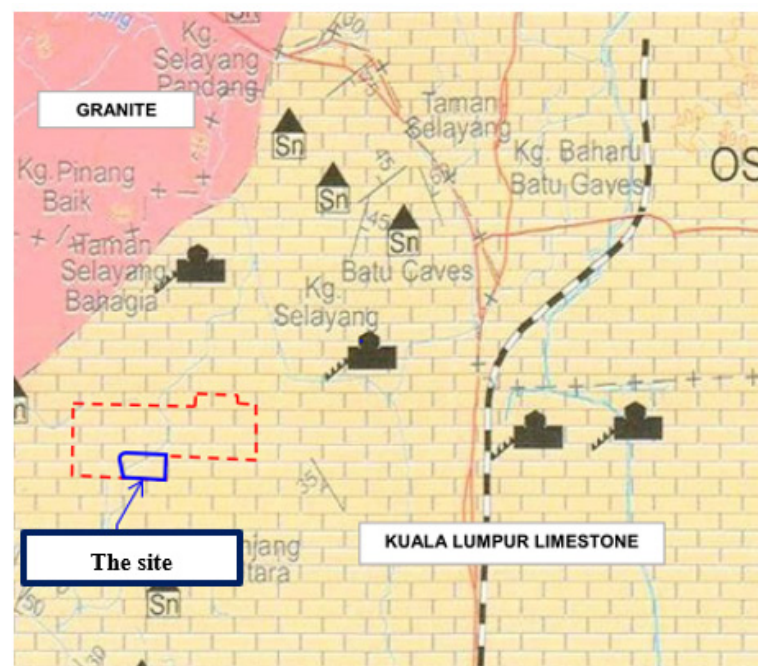- Zone (5): Building—The area covered by building.



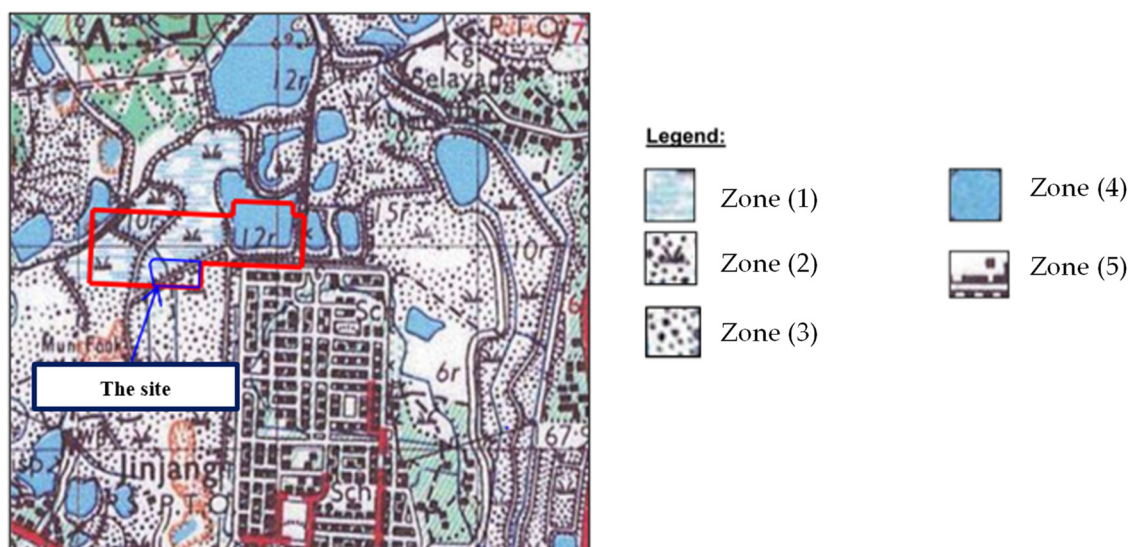**Figure 1.** Geological map of the site.



**Figure 2.** Topography map of Kuala Lumpur and the studied region.

The topography of the site is relatively flat ground, with the level of the ground ranging from about RL 55 m to RL 57.5 m. The site was proposed for the construction of high-rise buildings, with two towers of approximately 150 m in height and an eight-story car park. Therefore, deep foundations are necessary to withstand the structural load. The Jack-in installation method of pre-cast spun piles was the foundation of this development. Therefore, subsurface investigations have been carried out at the site to identify the ground condition. A total of 26 boreholes were investigated in order to determine the subsoil condition of the site. From the boreholes, the overburden of the site is mostly silt and clay material, with the SPT-N in the range of generally less than 10, as displayed in Figure 3.
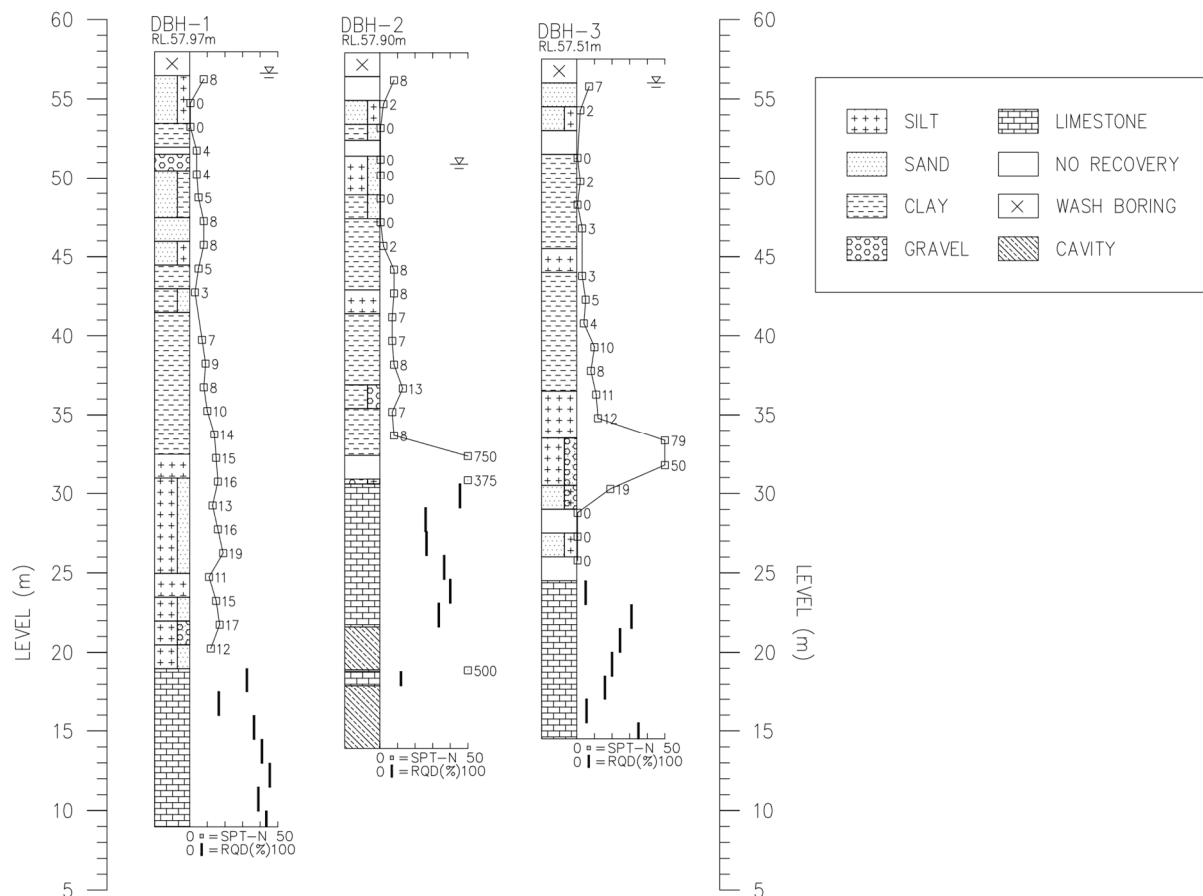


**Figure 3.** Example of three (3) boreholes soil profile of the ground with different geo-material and their SPT-N values.

SLT tests were carried out on the first installed piles to verify the pile design capacity. However, since SLT tests are expensive and time-consuming, the number of these tests is limited. Nevertheless, a total of 130 piles were selected to carry out HSDT. During the process of HSDT, a hydraulic hammer is dropped on the pile head with cushion and the force with velocity at the upper end of the pile are measured followed by a signal matching procedure. The HSDT is carried out on-site by testers who use a PDA system for data collection. Prior to the test, the soil around the test pile was excavated to ease the installation of transducers at about 1.5-3 times the pile's diameter from the pile head. A total of two strain and two accelerometer transducers were attached to opposite sides of each other close to the pile top, as shown in Figure 4. The applied load from the hammer drop was derived from the Strain transducers which act as strain measurements when the load is applied from a hammer drop on the pile head, whereas the movement of the pile head is measured by accelerometers during the impact.
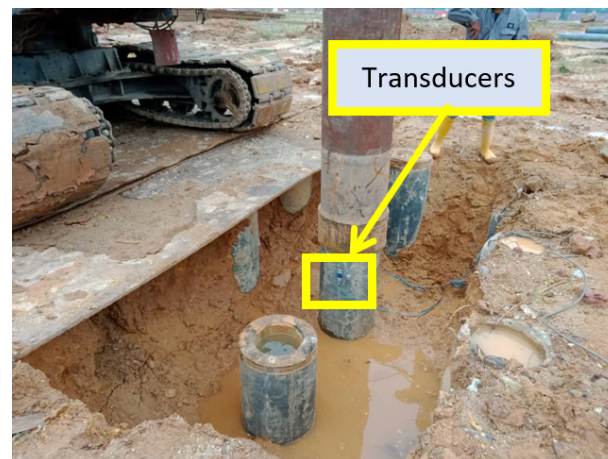
**Figure 4.** Indication of transducers on the piles.

When evaluating a model, the importance and contribution of parameters on the pile bearing capacity are significant. This can be extracted from the empirical equations or the statistical-based techniques. For example, as shown in Table 1, the SPT-N parameter is considered as an important factor in this regard. Apart from the empirical equations, several researchers have carried out sensitivity analyses on the influential parameters on the pile bearing capacity. Momeni et al. [26] stated that the weight of the hammer and pile geometrical properties, such as pile length and cross sectional area, have the highest impact on the pile bearing capacity. Ghorbani et al. [64] performed the sensitivity index of each parameter (pile shaft and tip area, the average cone tip resistance along the embedded length of the pile, the average cone tip resistance over the influence zone and the average sleeve friction along the embedded length of the pile which are obtained from CPT data) and found that pile soil surface area is the most contributing parameter. In another study, Pham et al. [66] found that average value of SPT-N number along the embedded pile length is the most crucial parameter in terms of pile capacity. Pile cross sectional area and length parameters were introduced as the most effective variables in another interesting study in the pile capacity estimation conducted by Momeni et al. [8].

According to the above discussion, and the available data for collection on the site, a total of six parameters, including pile diameter, hammer weight, pile length, hammer drop height, SPT-N average and pile friction bearing capacity (shaft friction) were measured on the site while conducting pile tests and from the borehole data. In order to calculate the SPT-N average for each pile, the zoning of the nearest SI to the pile test was considered. Subsequently, the SPT-N average of each pile was calculated based on the nearest SI work and varied according to the pile length. The mentioned parameters were collected for a total of 130 piles, which resulted in 130 data samples comprised of all six parameters. The next step is the identification and removal of outliers where outliers can be known as data that were significantly different from the observed data. This process is considered as a mandatory step when the data quantity is large, which is the case with the data in this study. With the presence of outliers, the database of variability will be increased, which can cause possible modeling errors. In this study, the statistical approach, which is based on the interquartile range rule, was used to identify outliers and these outliers were removed from the database. The interquartile range is the range between the first and third quartiles, namely Q1 and Q3. Any of the data that are smaller than Q1-1.5x interquartile range or higher than Q3 + 1.5 x interquartile range are considered outliers.

Eventually, five data samples were reduced from the whole 130 datasets and the used data samples reached 125 data samples. More information about these parameters can be found in Table 3. Among these factors, as highlighted in this study, the pile friction bearing capacity was considered as model output and the remaining factors were set as predictors.

However, the model predictors will be analyzed to select the most effective ones later in this paper.

**Table 3.** The parameters measured during conducting pile tests and from the boreholes data.

| Parameter | Unit | Range |
| --- | --- | --- |
| Pile Length | m | 18.5–47 |
| Hammer Drop Height | m | 0.55–1.1 |
| Pile Diameter | mm | 400–450 |
| Hammer Weight | Kg | 7000 |
| SPT-N Average | - | 4–17 |
| Pile Friction Bearing Capacity | kN | 2028–4844 |

### 2.2. Decision Tree (DT)

ML involves algorithms that use historical data with independent and target variables to learn and produce decisions by referring to a certain objective. One of the advantages of ML techniques, in comparison with conventional statistical approaches, such as regression, is that they are applicable for more than two-dimensional data. Many researchers have used tree-based techniques for data-driven prediction analysis for various geotechnical problems [68,69]. Thus, in this study, tree-based ML algorithms, such as DT, were applied to construct models and identify the crucial predictors of pile soil friction. DT can be represented graphically, displaying certain decision conditions with the complex branching that happens in a constructed decision. This approach is one of the top and most widely used supervised learning algorithms for predicting the accuracy of a model.

DT is able to carry out all tasks related to recognition, classification and prediction issues. DT is a "tree" shaped model that comprises of a series questions, with each question being described by various variables. A real tree consists of roots, branches and leaves. Similarly, the graph for DT is comprised of nodes, which are leaves, and branches that represent connections between nodes [70]. During the process of DT, a variable is selected as a root, which is known as the first node. The first node is separated into multiple internal nodes by referring to the appointed features. DT is a top-down tree, with the roots is located at the top. The final product of the branches consists of roots, branches and nodes [71]. Every node can be separated into two branches and each node has a relation to a certain characteristic and branches that have been described by a specific range of input. A flowchart related to DT technique is shown in Figure 5.

### 2.3. Random Forest (RF)

RF is a method that based on several DTs with boostrap aggregation and is one of the supervised ensembles ML approaches. It is also a part of ensemble learning that is based on a bagging algorithm. RF comprises of three (3) main attributes, which are presented as follows [72,73].

1. Automatically provide estimation of missing value.
2. Weight data to balance the errors found in imbalanced data.
3. Determine the crucial variables by estimation for classification.

In comparison with bagging, during the process of constructing each tree, RF utilizes random sample prediction before each node segmentation in order to reduce bias. Every DT is generated in parallel by RF and these trees can be classification or regression trees. At each constructed DT, each note is separated using the best features that can generate the most optimal solution among all of the attributes. The RF algorithm is a well-known method used to extract useful but hidden information within huge amounts of data. Figure 6 shows the process of the RF algorithm in modeling an output parameter.
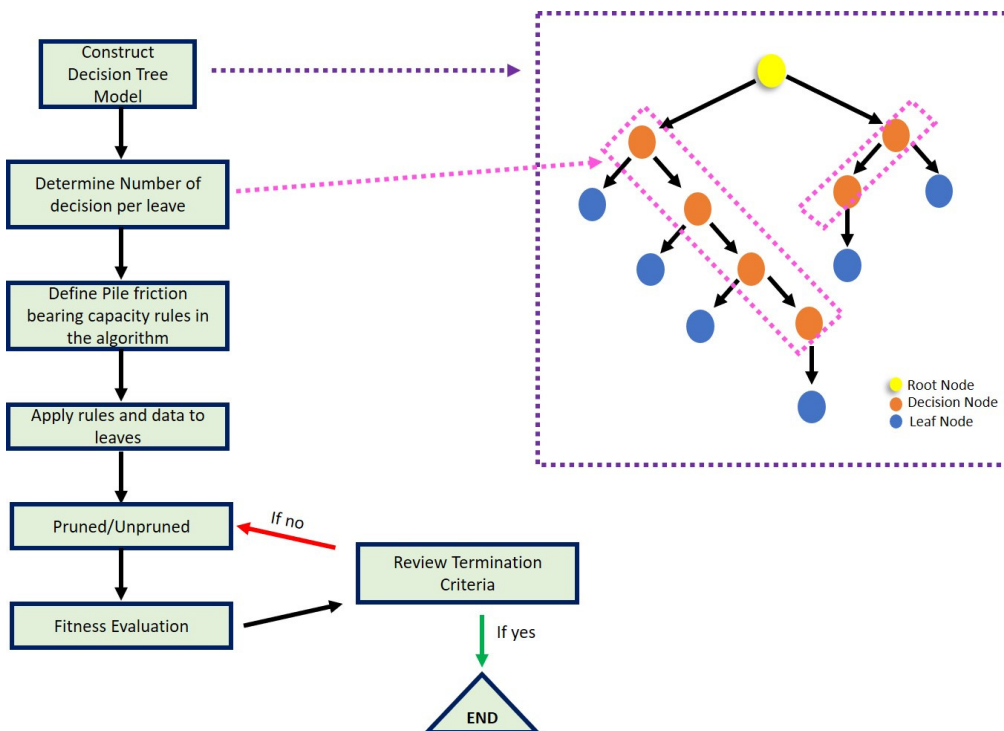
**Figure 5.** A Flowchart of DT technique for prediction purposes.
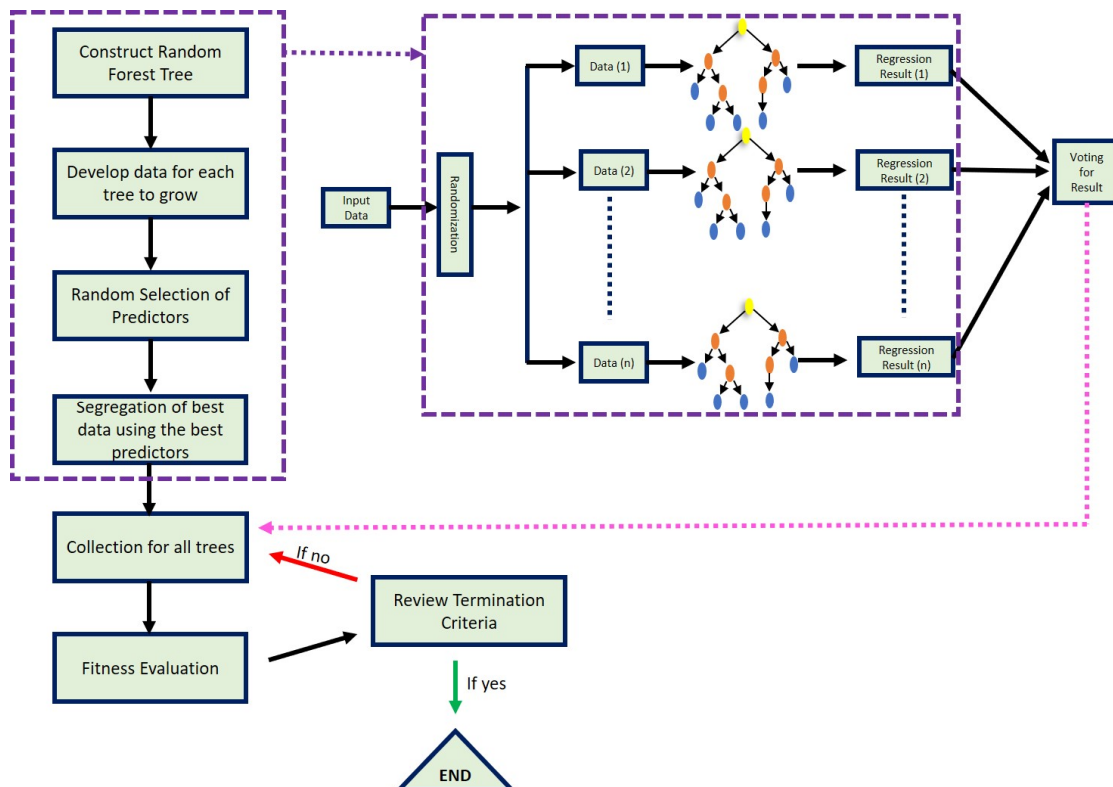


**Figure 6.** A Flowchart of RF technique for prediction purposes.

## 2.4. Gradient Boosted Tree (GBT)

A GBT is one of the tree-based methods that works on the principle of boosting. The models with low variance errors and high bias are combined with the purpose of lowering the bias and, at the same time, maintaining low variance [74]. Boosting is the process by

which it learns several classifiers by altering the sample weight for every process of training. All of the classifiers are combined linearly to enhance the classification performance. Unlike other tree-based methods, this approach uses similar training datasets in boosting. Similar datasets are trained and constructed as shallow trees in boosting trees but every tree has a different specific feature of the relationships between inputs and outputs. The objectives of the ($n$)th shallow tree are trained in series to reduce the prediction errors from the previous ($n-1$)th trees. The objective of GBT is to form a supplementary model that reduces the loss function. The process of the GBT method is described as follows:

1.  A constant value is begun in the model to lower down the loss function.
2.  During the iteration of training process, the residual value of the model is estimated from the negative gradient of the loss function.
3.  The current residual value is fit by newly trained regression tree.
4.  The combination of final regression with past models and residual is updated.
5.  When the maximum number of iterations set by the user is achieved, the iteration in the algorithm is ceased.

Overall, the GBT model is able to improve the previous poorly executed data by continuously using a regression tree to fit the residual. The applications of this technique have been highlighted in several problems related to geotechnical engineering [74–77]. A GBT flowchart in modeling a predictive technique is displayed in Figure 7.
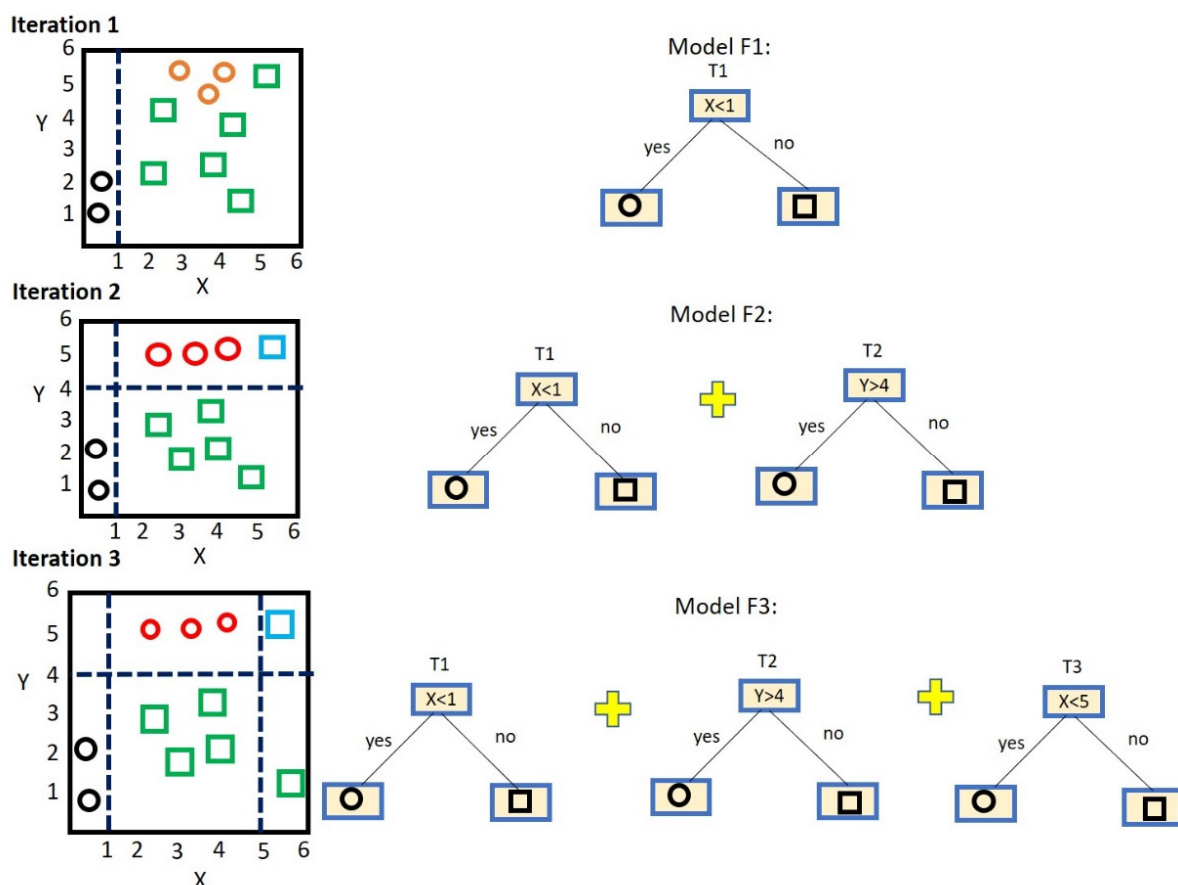


**Figure 7.** A GBT flowchart in modeling a predictive technique.

### 2.5. Performance Indices

Evaluation of the models in ML is a way of assessing the size effect in conventional statistics [78]. The ability of a model to predict for an unknown sample is a critical step in ML that increases trust in the model for use on other datasets. The accuracy, in terms of percentage, is the measurement for model evaluation. The authors of this study decided to

use several important performance indices, including root mean square error (RMSE), $R^2$, and absolute error. Willmott and Matsuura [79] stated that the total square error is affected by the larger error, rather than the minor error. When the variances associated with the frequency distribution of error magnitudes increase, the error values will be increased. The mentioned performance indices were utilized and computed in different studies and their formulas and process of calculation can be found in the literature [80–83]. In this study, the mentioned performance indices will be used to evaluate the model's prediction capacity.

*2.6. Study Steps*

  This study was planned to introduce a process of modeling and a superior tree-based model for solving a problem related to piling technology. The prediction of the pile capacity is always important for geotechnical engineers right after pile installation, because the measurement of this parameter needs specific equipment and its setting in the site, which is not easy to do. In addition, conducting such tests are costly and sometimes includes human and machine errors [26]. The modeling process of this study was started by identifying and removing the outliers. The next step is related to feature selection or input selection so that the most effective parameters will be selected. In this way, the supervised and unsupervised feature selection methods will be utilized. After selecting the model predictors, three model trees, i.e., DT, RF and GBT, will be conducted to predict friction pile bearing capacity. These models' trees and their performance capacities will be assessed and discussed. The best tree model will be selected and introduced based on both model development and model assessment. A schematic diagram of the study steps is presented in Figure 8.
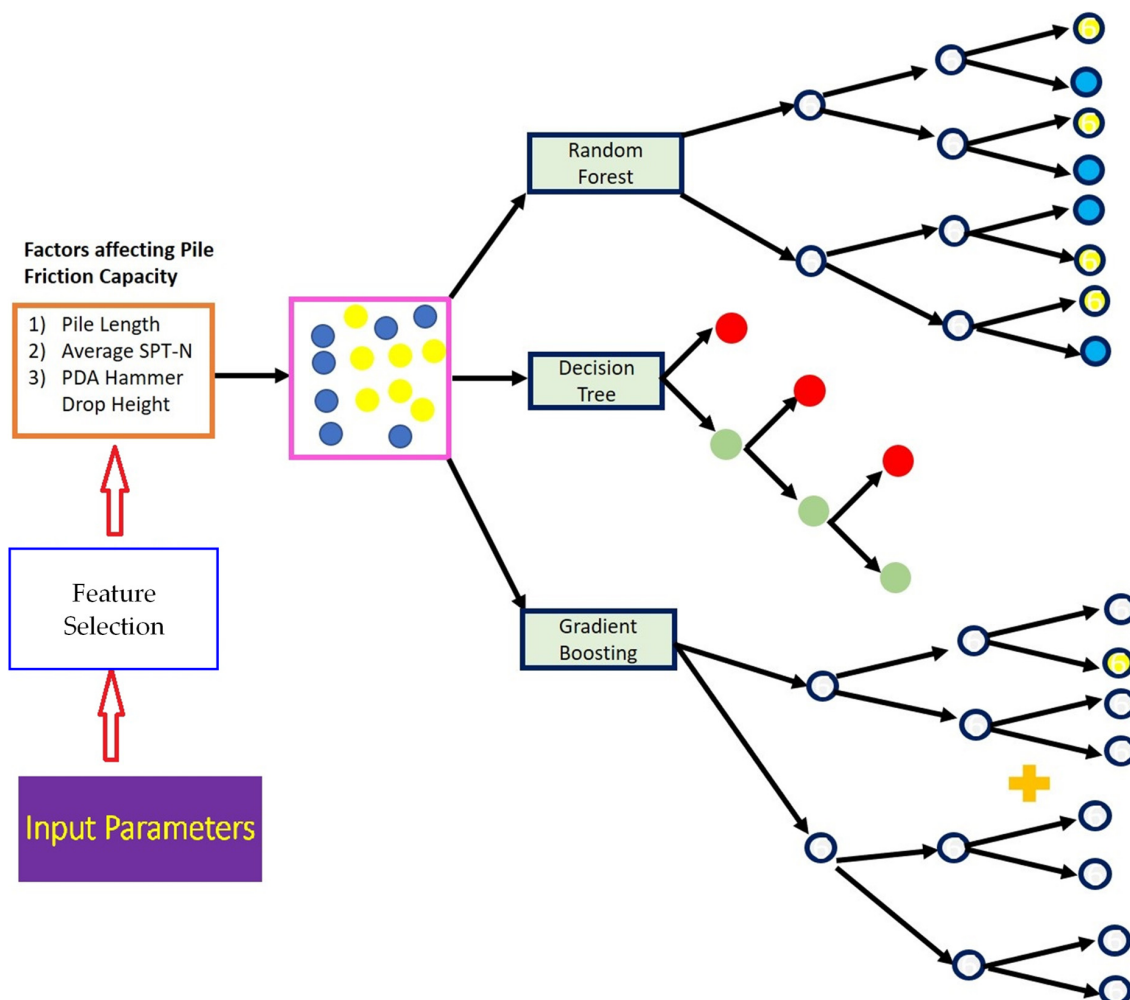


**Figure 8.** The process of this study to predict pile capacity.

## 3. Input Selection

Feature/input selection is an alternative to identifying significant factors in conventional statistics using measures of confidence interval and hypothesis testing. After performing model evaluation, the elements (independent variables) need to be explored further in terms of how they contribute to the accuracy measure. This technique removes variables that are insignificant or highly correlated with any other variable. The rank of variables based on importance score can be visualized to understand the prediction accuracy. The supervised and unsupervised feature selection methods differ in terms of the target variables. The supervised learning model needs a target variable to determine the important variables, while unsupervised learning ignores the target variable and selects important variables based on correlation. In the following sub-sections, supervised and unsupervised feature selection methods will be applied and their results will be discussed.

### 3.1. Correlation

Table 4 presents the correlations between the used input variables. Since the purpose of this study is to predict the shaft friction or pile friction bearing capacity, variables with strong positive correlation were considered for developing the final model. According to correlation analysis, pile capacity and pile length are highly positively correlated (0.794). The correlation of hammer weight with other variables is not reported in Table 4, which indicates no correlation between this variable and other variables. It should be noted that, in the database, there is only one value for the hammer weight and, because of this issue, this parameter was removed from the analysis of the correlation technique.

**Table 4.** Correlations between the independent variables.

| Attributes | Hammer Drop Height Test | Pile Diameter | Pile Length | Shaft Friction | SPT-N Average |
|---|---|---|---|---|---|
| Hammer Drop Height Test | 1 | 0.358 | 0.504 | 0.476 | −0.025 |
| Pile Diameter | 0.358 | 1 | 0.103 | 0.436 | 0.207 |
| Pile Length | 0.504 | 0.103 | 1 | 0.794 | −0.030 |
| Shaft Friction | 0.476 | 0.436 | 0.794 | 1 | 0.036 |
| SPT-N Average | −0.025 | 0.207 | −0.030 | 0.036 | 1 |

### 3.2. Supervised Feature Selection

This study adopts different feature selection methods to select only important variables and develop a prediction model based on the selected variables. The main reason behind reducing the number of variables (based on their level of importance and correlations) is to decrease the complexity and improve the applicability of the final model. Armaghani et al. [84] stated that a lower number of model inputs is considered as an advantage for the developed models, since the model complexity cannot be minimized. After conducting unsupervised clustering and understanding the correlation of the variables, it was compared the variables' importance based on three different tree-based supervised ML techniques. The importance of variables based on GBT, RF and DT results are shown and compared in Table 5. According to this table, pile length is indicated as the most important variable based on all three techniques. On the other hand, hammer weight does not have any impact on pile capacity. These results are in line with correlation analysis results, too. In addition, in order to make a better conclusion out of three feature selection methods, weight values of each variable were summed up and compared, as shown in Figure 9. The accumulated weight values were then sorted out from highest to lowest values. It can be concluded from Figure 9 that pile diameter and hammer weight is not significant enough to be considered as model inputs. Therefore, the authors decided to not consider these two attributes for developing the final predictive models in this research. However, it is

necessary to note that there are only two values of pile diameters in the provided database. Therefore, the impact of this parameter is not significant in our database (Figure 9). In general, pile geometry is considered as a significant predictor category for prediction of the pile capacity, as suggested in the literature [28,85] and it is suggested to use pile diameter as a model input with various values in future studies.

**Table 5.** Importance score (weight) of variables based on three supervised ML methods.

| GB | | RF | | DT | |
|---|---|---|---|---|---|
| **Attribute** | **Weight** | **Attribute** | **Weight** | **Attribute** | **Weight** |
| Pile Length | 0.75 | Pile Length | 0.80 | Pile Length | 0.35 |
| SPT-N Average | 0.125 | SPT-N Average | 0.17 | Pile Dimeter | 0.04 |
| Hammer drop height test | 0.031 | Hammer drop height test | 0.11 | Hammer drop height test | 0.037 |
| Pile Dimeter | 0.030 | Pile Dimeter | 0.025 | SPT-N Average | 0.018 |
| Hammer Weight | 0.00 | Hammer Weight | 0.00 | Hammer Weight | 0.00 |
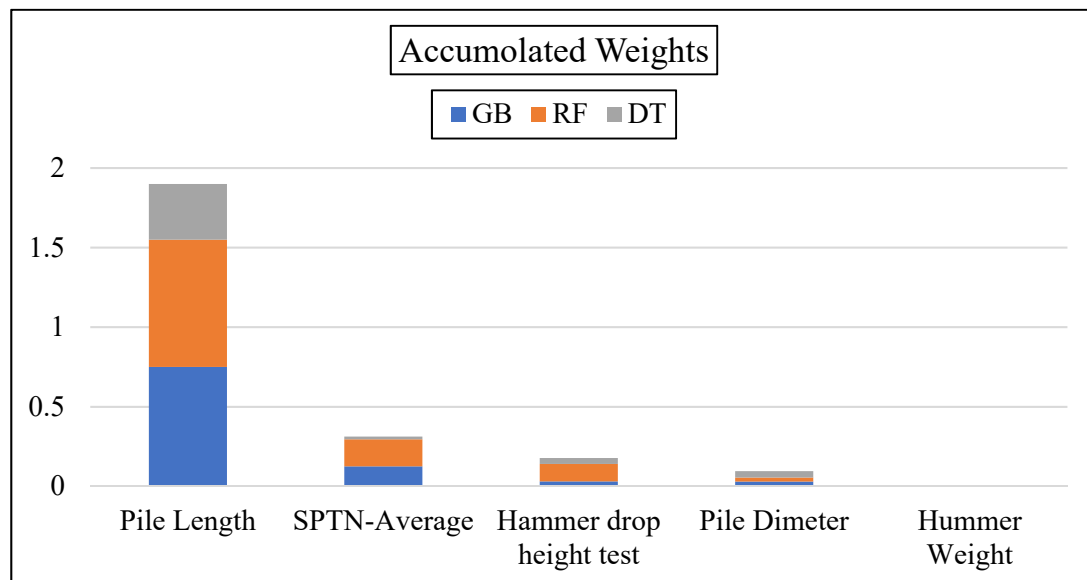


**Figure 9.** Accumulated weights of variables.

## 4. Modeling and Results

In order to have an accurate model proposal, both model development and model evaluation parts should be at an acceptable level. In the first stage of modeling all data samples were normalized using the following equation in the range of [0–1]:

$$Z = \frac{X - min(X)}{max(X) - min(X)} \quad (2)$$

where, *X* presents each parameter that needs to be normalized (i.e., each input and output), *min(X)* and *max(X)* are the minimum and maximum values of whole data of that specific parameter, respectively.

For this purpose, there is a need to divide the whole database into two groups: train and test. In this study, among all of the available suggestions in the literature, the authors decided to use a combination of 80–20% for train and test phases, respectively. Therefore, before starting the modeling, all 125 data samples were divided to 25 and 100 data samples for model evaluation and model development, respectively. As discussed

before, three tree-based ML techniques were employed to determine the most accurate model for predicting the pile friction bearing capacity. To do this, several parametric investigations were performed for different parameters of DT, RF and GBT techniques. In these analyses, three and five model inputs were utilized. Finally, the model performance results with three and five input variables are presented in Tables 6 and 7, respectively. In these tables, train and test results of $R^2$, RMSE and absolute error were presented. According to the results, GBT technique achieved the highest accuracy rate for both models with five and three variables, with $R^2$ equal to 0.911 and 0.901, respectively, for training datasets in predicting pile friction bearing capacity. In addition, the GBT model achieved the lowest RMSE and absolute error compared to the RF and DT models. The next best model after GBT is related to RF for three and five input variables, followed by the DT model. The R values of (0.813 and 0.761) and (0.773 and 0.712) were obtained for testing data samples of RF and DT techniques for three and five model inputs, respectively. It is obvious that the results obtained by the GBT technique are more accurate compared to the RF and DT models for both three and five input parameters.

**Table 6.** Models evaluation results using five input variables.

| Technique | Train | | | Test | | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| | $R^2$ | RMSE | Absolute Error | $R^2$ | RMSE | Absolute Error |
| GBT | 0.911 | 0.073 | 0.052 | 0.841 | 0.089 | 0.098 |
| RF | 0.892 | 0.082 | 0.073 | 0.813 | 0.071 | 0.100 |
| DT | 0.861 | 0.135 | 0.121 | 0.773 | 0.121 | 0.167 |

**Table 7.** Models evaluation results using three important input variables.

| Technique | Train | | | Test | | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| | $R^2$ | RMSE | Absolute Error | $R^2$ | RMSE | Absolute Error |
| GBT | 0.901 | 0.094 | 0.077 | 0.816 | 0.127 | 0.098 |
| RF | 0.881 | 0.095 | 0.081 | 0.761 | 0.081 | 0.100 |
| DT | 0.853 | 0.170 | 0.152 | 0.712 | 0.130 | 0.167 |

As was expected, the accuracy of the models decreased by decreasing the number of variables after feature selection. However, considering the train results of the GBT technique, the accuracy reduction is not significant (only 1%). For testing data, the GBT results based on $R^2$ are 0.841 and 0.816 for three and five models, respectively, which show a close model accuracy when three input parameters are used. Therefore, as discussed before, proposing a new predictive model with a lower number of model predictors is of importance in the area of piling and geotechnical engineering. The other researchers and designers can easily use a simpler model because they need a lower number of parameters to be measured. Therefore, in this study, the authors decided to propose and introduce a predictive model with lower model inputs, even though it has lower performance prediction results. Hence, the results presented in Table 7 will be considered in this study and, as such, the GBT model for three inputs will be discussed in more detail in the following paragraphs.

After reducing the number of variables through the feature selection process, the GBT model was conducted using three important selected variables. Table 8 presents the importance of input variables using the GBT technique with the final three variables. The importance values of 0.81, 0.21 and 0.075 were obtained, respectively, for pile length, SPT-N average, and hammer drop height. According to the results, pile length, with an importance of 0.81, plays the most important role in predicating pile friction bearing capacity using the GBT technique. On the other hand, hammer drop height has the lowest impact on the model output, which is the pile capacity.

**Table 8.** GBT importance of variables.

| Attribute | Weight |
|---|---|
| Pile Length | 0.81 |
| SPT-N Average | 0.21 |
| Hammer drop height | 0.075 |

In modelling GBT, there were many models constructed in order to see the difference between different parameters of GBT on the system performance. As presented in Table 9, 27 GBT models were built in this study with different properties in order to predict pile bearing capacity. In these 27 models, the authors considered different values for the number of trees, maximum depth and learning rate in the modeling. In addition, error results are presented in Table 9 for each GBT model. As a result, the optimal/best model is achieved when the number of trees is 90 with a maximum depth of two and 0.1 learning rate (i.e., GBT model number 20). The lowest error rate of 0.1889 and the highest accuracy (R values of 0.901 and 0.816 according to Table 7) were observed at the described point. Figure 10 shows the schematic tree generated by the proposed GBT model. More discussions regarding this technique will be given in the next section.

**Table 9.** Different GBT models to predict pile friction bearing capacity.

| GBT Model No. | No. of Trees | Maximal Depth | Learning Rate | Error Rate |
|---|---|---|---|---|
| 1 | 30 | 2.0 | 0.001 | 0.2891 |
| 2 | 90 | 2.0 | 0.001 | 0.2814 |
| 3 | 150 | 2.0 | 0.001 | 0.2745 |
| 4 | 30 | 4.0 | 0.001 | 0.2891 |
| 5 | 90 | 4.0 | 0.001 | 0.2814 |
| 6 | 150 | 4.0 | 0.001 | 0.2745 |
| 7 | 30 | 7.0 | 0.001 | 0.2891 |
| 8 | 90 | 7.0 | 0.001 | 0.2814 |
| 9 | 150 | 7.0 | 0.001 | 0.2745 |
| 10 | 30 | 2.0 | 0.01 | 0.2597 |
| 11 | 90 | 2.0 | 0.01 | 0.2366 |
| 12 | 150 | 2.0 | 0.01 | 0.2189 |
| 13 | 30 | 4.0 | 0.01 | 0.2597 |
| 14 | 90 | 4.0 | 0.01 | 0.2370 |
| 15 | 150 | 4.0 | 0.01 | 0.2209 |
| 16 | 30 | 7.0 | 0.01 | 0.2597 |
| 17 | 90 | 7.0 | 0.01 | 0.2370 |
| 18 | 150 | 7.0 | 0.01 | 0.2209 |
| 19 | 30 | 2.0 | 0.1 | 0.1986 |
| 20 | 90 | 2.0 | 0.1 | 0.1889 |
| 21 | 150 | 2.0 | 0.1 | 0.1953 |
| 22 | 30 | 4.0 | 0.1 | 0.1985 |
| 23 | 90 | 4.0 | 0.1 | 0.1912 |
| 24 | 150 | 4.0 | 0.1 | 0.1955 |
| 25 | 30 | 7.0 | 0.1 | 0.1985 |
| 26 | 90 | 7.0 | 0.1 | 0.1912 |
| 27 | 150 | 7.0 | 0.1 | 0.1955 |

**Figure 10.** GBT model tree using three variables.

## 5. Discussion

In this study, a series of experimental data were measured and recorded during SLT tests, and the capacity values of friction piles, together with some other important parameters on them, were collected. The idea was to propose a series of fully tree-based techniques, i.e., DT, RF and GBT, for estimation of the pile bearing capacity. Through feature selection, in order to propose a simpler model, the three most important parameters were identified as pile length, SPT-N average and hammer drop height. The mentioned tree-based models were then built to predict pile friction bearing capacity. In order to construct DT, RF and GBT models, many attempts have been made to achieve higher performance capacities based on the used statistical indices. These attempts were performed by setting different values for the most influential DT, RF and GBT parameters. As expected, the developed GBT model was able to provide a better performance capacity in estimating the actual results of pile friction bearing capacity. The training and testing results of the proposed GBT model are presented in Figures 11 and 12, respectively. It is important to note that the pile capacity values presented in these figures are normalized between [0–1], as described previously. The $R^2$ and other statistical indices are presented in these figures, which confirms that the GBT is a powerful tree-base technique in both phases of model development and model evaluation. RMSE results and absolute error results of (0.094, and 0.077) and (1.27, and 0.098) for train and test data samples, respectively, reveal that the GBT tree model is applicable in the field of piling and deep foundation. It is able to predict pile bearing capacity values with a low level of system error, which is of importance and advantage in the geotechnical engineering field.
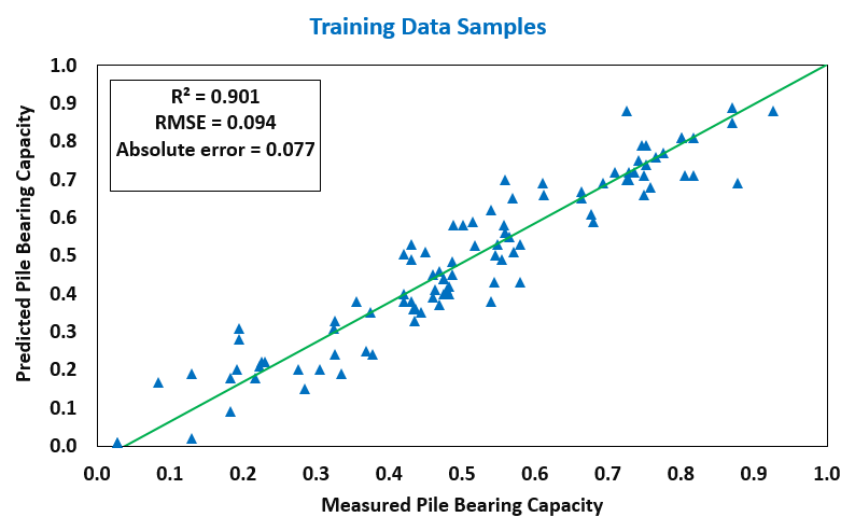


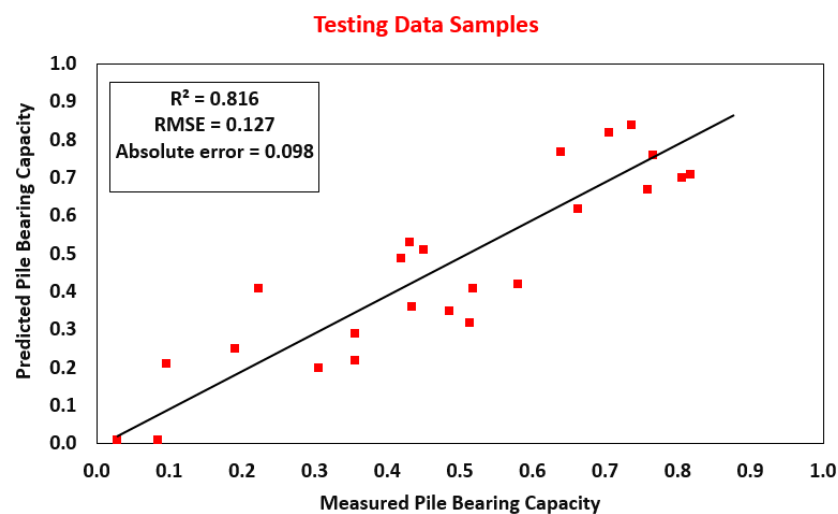**Figure 11.** Statistical indices results of GBT for model development part.

**Figure 12.** Statistical indices result of GBT for model evaluation part.

Compared to the previous ML related studies, this study focuses on only tree-based ML techniques. To date, only a few researchers have proposed similar techniques in this regard. The developed GBT model in this study was based on only three input parameters, and based on these three inputs, the GBT model provided $R^2$ values of 0.901 and 0.816 for the train and test phases, respectively. The results of the GBT model are not better than many of the relevant studies presented in Table 2. However, as presented in Table 2, most of the studies used five or more input parameters to predict the pile bearing capacity. This makes them complicated models for further use by other researchers. This is because they have to provide the related values for all inputs if they want to use the proposed models. Nevertheless, in this study, the presented results were constructed based on only three model inputs/predictors. In other words, the proposed GBT technique in this study is easier to implement by other researchers, designers, or engineers. Hence, the modelling process and the proposed GBT model in this study can be suggested as a reliable and applicable technique/process with a high level of accuracy in forecasting pile bearing capacity.

It is good to know that the GBT model can depict the promising accuracy of the prediction, provided that this study is carried out for different types of soils, piles, installation methods, and types of hammers. This study was carried out with limited data with one hammer weight, two types of pile dimension, an SPT-N value of generally about 10, and one installation method. Thus, it is highly recommended to further carry out this study with more variables in order to provide higher accuracy of the prediction.

## 6. Optimum Parameters Based on Simulation Model

In order to gain deeper insight into the factors affecting pile capacity, sensitivity analysis was conducted based on a desirable scenario: what is the optimal value of independent variables in order to gain the maximum pile friction bearing capacity? The simulation-based sensitivity analysis was conducted using RapidMiner Studio Educational Software version 9.8.001. The graphical results in this section are the outputs of the RapidMiner Software. The RapidMiner software conducts the tree-based models under the Python software environment. The optimization was run and determined the best input variables to meet our target under the specified constraints. In addition, the simulation-based sensitivity analysis is suitable for assessing and answering "what if" questions. Table 10 presents the optimal values of attributes based on the described scenario.

**Table 10.** Optimized value of attributes based on maximum scenario.

| Attribute | Optimum Value (Normalized) | Optimum Value (Actual) | Maximum Pile Capacity |
|---|---|---|---|
| Pile Length | 0.90 | 44 | |
| Hammer Drop Height | 0.419 | 1.1 | 4844 |
| SPT-N Average | 0.702 | 6 | |

According to the simulation-based optimization results, pile friction bearing capacity will be equal to 4844, when pile length, hammer drop height and SPT-N average are equal to 44, 1.1 and 6, respectively (Figure 13). These values are the optimum values (not maximum) of independent variables in order to achieve the maximum pile capacity values. For example, the optimum value for pile length is 44, which is almost equal to the maximum value (0.90 as normalized). Therefore, pile capacity will be decreased after this amount of pile length.
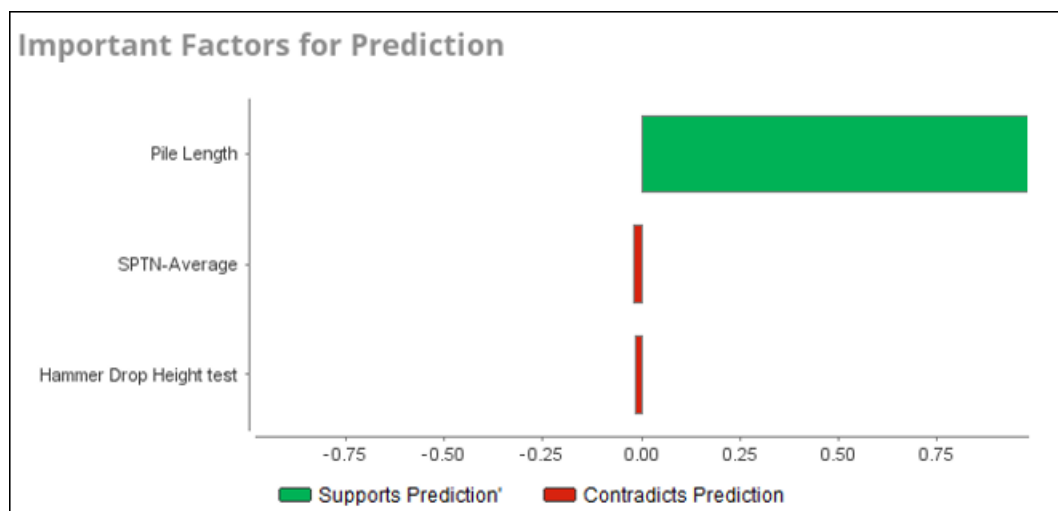


**Figure 13.** Optimization results and importance of variables.

### 7. Summary and Conclusions

The purpose of this research is to propose a more accurate and applicable model/approach for predicting pile friction bearing capacity, which is fully tree-based with a limited number of model inputs/predictors. To achieve this aim, first, among the initial five input variables, the three most effective ones, i.e., pile length, SPT-N average, and hammer drop height, were selected for the modelling part, based on a comprehensive feature selection process. Three tree-based techniques, i.e., DT, RF and GBT, were then built to estimate pile friction bearing capacity. In building these models, a series of parametric investigations based on their effective variables were planned and performed in order to obtain the best model in each category. In the next step, model assessment has been done using different performance prediction indices, and their results have been compared to each other. Overall, the findings demonstrated the successful application of tree-based techniques for the purpose of this paper. However, the best tree model was related to GBT with $R^2$ values of 0.901 and 0.816 for model development and model assessment parts, respectively. It should be noted that the other tree-based models received acceptable and applicable results for prediction of pile friction bearing capacity. The parametric study results showed that the optimum values of pile length, hammer drop height and SPT-N average are 44, 1.1, and 6, respectively, in order to get the maximal pile capacity values. The proposed tree-based techniques and their processes are easy to implement and can be

used by other researchers and designers to obtain very accurate pile capacity values for similar conditions. However, other researchers can prepare a larger database for the same problem and develop more comprehensive tree-based techniques, or even a combination of these techniques with new optimization techniques, such as the sparrow search algorithm, in order to achieve a higher accuracy level.

## References

1. Viggiani, C.; Mandolini, A.; Russo, G. *Piles and Pile Foundations*; Spon Press: London, UK, 2012; p. 278, ISBN 1498725538.
2. Torshizi, M.F.; Saitoh, M.; Álamo, G.M.; Goit, C.S.; Padrón, L.A. Influence of pile radius on the pile head kinematic bending strains of end-bearing pile groups. *Soil Dyn. Earthq. Eng.* **2018**, *105*, 184–203. [CrossRef]
3. Ma, Y.; Deng, N. Deep foundations. *Substruct. Des.* **2014**, *239*, 18.
4. Vesic, A.S. *Design of Pile Foundations. National Cooperative Highway Research Program Synthesis of Practice no. 42*; Transportation Research Board: Washington, DC, USA, 1977; p. 3248.
5. Helwany, S. *Applied Soil Mechanics with ABAQUS Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2007; ISBN 0471791075.
6. Liu, Y.J.; Liang, S.H.; Wu, J.W.; Fu, N. Prediction method of vertical ultimate bearing capacity of single pile based on support vector machine. *Adv. Mater. Res.* **2011**, *168–170*, 2278–2282. [CrossRef]
7. Shah, D.L.; Shroff, A.V. *Soil Mechanics and Geotechnical Engineering*; CRC Press: Boca Raton, FL, USA, 2003; ISBN 9058092356.
8. Momeni, E.; Nazir, R.; Armaghani, D.J.; Maizir, H. Application of artificial neural network for predicting shaft and tip resistances of concrete piles. *Earth Sci. Res. J.* **2015**, *19*, 85–93. [CrossRef]
9. Brinkgreve, R.B.J.; Engin, E. Validation of geotechnical finite element analysis. In Proceedings of the 18th International Conference on Soil Mechanics and Geotechnical Engineering, Paris, France, 2–6 September 2013; Volume 2, pp. 677–682.
10. Nawari, N.O.; Liang, R.; Nusairat, J. Artificial intelligence techniques for the design and analysis of deep foundations. *Electron. J. Geotech. Eng.* **1999**, *4*, 1–21.
11. Wardani, S.P.R.; Surjandari, N.S.; Jajaputra, A.A. Analysis of ultimate bearing capacity of single pile using the artificial neural networks approach: A case study. In Proceedings of the 18th International Conference on Soil Mechanics and Geotechnical Engineering, Paris, France, 2–6 September 2013; pp. 837–840.
12. Jianbin, Z.; Jiewen, T.; Yongqiang, S. An ANN model for predicting level ultimate bearing capacity of PHC pipe pile. In *Earth and Space 2010: Engineering, Science, Construction, and Operations in Challenging Environments*; ASCE: Reston VA, USA, 2010; pp. 3168–3176.
13. Suman, S. Prediction of Pile Capacity Parameters Using Functional Networks and Multivariate Adaptive Regression Splines. Doctoral Dissertation, Department of Civil Engineering National Institue of Technology Rourkela, Odisha, India, 2015.
14. Doherty, P.; Gavin, K. The shaft capacity of displacement piles in clay: A state of the art review. *Geotech. Geol. Eng.* **2011**, *29*, 389–410. [CrossRef]
15. Shahin, M.A. Artificial intelligence in geotechnical engineering: Applications, modeling aspects, and future directions. In *Metaheuristics in Water, Geotechnical and Transport Engineering*; Elsevier: Amsterdam, The Netherlands, 2013; p. 169204.
16. Momeni, E.; Maizir, H.; Gofar, N.; Nazir, R. Comparative study on prediction of axial bearing capacity of driven piles in granular materials. *J. Teknol.* **2013**, *61*, 15–20. [CrossRef]
17. Bazaraa, A.R.; Kurkur, M.M. N-values used to predict settlements of piles in Egypt. In Proceedings of the Use of In Situ Tests in Geotechnical Engineering, Virginia, VA, USA, 23–25 June 1986; pp. 462–474.

18. Décourt, L. Prediction of the bearing capacity of piles based exclusively on N values of the SPT. In *Penetration Testing*; Routledge: Oxfordshire, UK, 2021; pp. 29–34.

19. Lopes, F.R.; Laprovitera, H. On the prediction of the bearing capacity of bored piles from dynamic penetration tests. In Proceedings of the International Geotechnical Seminar on Deep Foundations on Bored and Auger Piles, Ghent, Belgium, 1–7 June 1988; pp. 537–540.

20. Meyerhof, G.G. Penetration tests and bearing capacity of cohesionless soils. *J. Soil Mech. Found. Div.* **1956**, *82*, 861–866. [CrossRef]

21. Shioi, Y.; Fukui, J. Application of N-value to design of foundations in Japan. In *Penetration Testing*; Routledge: Oxfordshire, UK, 2021; pp. 159–164.

22. Aoki, N.; Velloso, D.A. An approximate method to estimate the bearing capacity of piles. In Proceedings of the 5th Pan-American Conf. of Soil Mechanics and Foundation Engineering, Bueno Aires, Argentina, 17–22 November 1975; International Society of Soil Mechanics and Geotechnical Engineering Buenos. Volume 1, pp. 367–376.

23. Reese, L.C.; O'Neill, M.W. New design method for drilled shafts from common soil and rock tests. In *Foundation Engineering: Current Principles and Practices*; ASCE: Reston, VA, USA, 1989; pp. 1026–1039.

24. Robert, Y. A few comments on pile design. *Can. Geotech. J.* **1997**, *34*, 560–567. [CrossRef]

25. Randolph, M.F. Science and empiricism in pile foundation design. *Géotechnique* **2003**, *53*, 847–875. [CrossRef]

26. Momeni, E.; Nazir, R.; Armaghani, D.J.; Maizir, H. Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. *Measurement* **2014**, *57*, 122–131. [CrossRef]

27. Harandizadeh, H.; Armaghani, D.J.; Khari, M. A new development of ANFIS–GMDH optimized by PSO to predict pile bearing capacity based on experimental datasets. *Eng. Comput.* **2021**, *37*, 685–700. [CrossRef]

28. Chen, W.; Sarir, P.; Bui, X.-N.; Nguyen, H.; Tahir, M.M.; Armaghani, D.J. Neuro-genetic, neuro-imperialism and genetic programing models in predicting ultimate bearing capacity of pile. *Eng. Comput.* **2020**, *36*, 1101–1115. [CrossRef]

29. Lee, I.-M.; Lee, J.-H. Prediction of pile bearing capacity using artificial neural networks. *Comput. Geotech.* **1996**, *18*, 189–200. [CrossRef]

30. Likins, G.E.; Rausche, F. Correlation of CAPWAP with static load tests. In Proceedings of the Seventh International Conference on the Application of Stresswave Theory to Piles, Petaling Jaya, Malaysia, August 2004; pp. 153–165.

31. Moayedi, H.; Armaghani, D.J. Optimizing an ANN model with ICA for estimating bearing capacity of driven pile in cohesionless soil. *Eng. Comput.* **2018**, *34*, 347–356. [CrossRef]

32. Jahed Armaghani, D.; Shoib, R.S.N.S.B.R.; Faizi, K.; Rashid, A.S.A. Developing a hybrid PSO–ANN model for estimating the ultimate bearing capacity of rock-socketed piles. *Neural Comput. Appl.* **2017**, *28*, 391–405. [CrossRef]

33. Huang, J.; Asteris, P.G.; Pasha, S.M.K.; Mohammed, A.S.; Hasanipanah, M. A new auto-tuning model for predicting the rock fragmentation: A cat swarm optimization algorithm. *Eng. Comput.* **2020**. Available online: https://link.springer.com/article/10.1007/s00366-020-01207-4 (accessed on 24 October 2021). [CrossRef]

34. Chen, H.; Asteris, P.G.; Jahed Armaghani, D.; Gordan, B.; Pham, B.T. Assessing Dynamic Conditions of the Retaining Wall: Developing Two Hybrid Intelligent Models. *Appl. Sci.* **2019**, *9*, 1042. [CrossRef]

35. Parsajoo, M.; Armaghani, D.J.; Mohammed, A.S.; Khari, M.; Jahandari, S. Tensile strength prediction of rock material using non-destructive tests: A comparative intelligent study. *Transp. Geotech.* **2021**, *31*, 100652. [CrossRef]

36. Armaghani, D.J.; Hajihassani, M.; Mohamad, E.T.; Marto, A.; Noorani, S.A. Blasting-induced flyrock and ground vibration prediction through an expert artificial neural network based on particle swarm optimization. *Arab. J. Geosci.* **2014**, *7*, 5383–5396. [CrossRef]

37. Xie, C.; Nguyen, H.; Choi, Y.; Armaghani, D.J. Optimized functional linked neural network for predicting diaphragm wall deflection induced by braced excavations in clays. *Geosci. Front.* **2021**, 101313. [CrossRef]

38. Apostolopoulou, M.; Asteris, P.G.; Armaghani, D.J.; Douvika, M.G.; Lourenço, P.B.; Cavaleri, L.; Bakolas, A.; Moropoulou, A. Mapping and holistic design of natural hydraulic lime mortars. *Cem. Concr. Res.* **2020**, *136*, 106167. [CrossRef]

39. Asteris, P.G.; Kolovos, K.G. Self-compacting concrete strength prediction using surrogate models. *Neural Comput. Appl.* **2019**, *31*, 409–424. [CrossRef]

40. Khandelwal, M.; Mahdiyar, A.; Armaghani, D.J.; Singh, T.N.; Fahimifar, A.; Faradonbeh, R.S. An expert system based on hybrid ICA-ANN technique to estimate macerals contents of Indian coals. *Environ. Earth Sci.* **2017**, *76*, 399. [CrossRef]

41. Khandelwal, M.; Singh, T.N. Prediction of blast induced air overpressure in opencast mine. *Noise Vib. Worldw.* **2005**, *36*, 7–16. [CrossRef]

42. Armaghani, D.J.; Harandizadeh, H.; Momeni, E. Load carrying capacity assessment of thin-walled foundations: An ANFIS–PNN model optimized by genetic algorithm. *Eng. Comput.* **2021**. Available online: https://link.springer.com/article/10.1007/s00366-021-01380-0 (accessed on 24 October 2021).

43. Gajurel, A.; Chittoori, B.; Mukherjee, P.S.; Sadegh, M. Machine learning methods to map stabilizer effectiveness based on common soil properties. *Transp. Geotech.* **2021**, *27*, 100506. [CrossRef]

44. Jahed Armaghani, D.; Asteris, P.G.; Askarian, B.; Hasanipanah, M.; Tarinejad, R.; Huynh, V. Van Examining Hybrid and Single SVM Models with Different Kernels to Predict Rock Brittleness. *Sustainability* **2020**, *12*, 2229. [CrossRef]

45. Mohammed, A.S.; Asteris, P.G.; Koopialipoor, M.; Alexakis, D.E.; Lemonis, M.E.; Armaghani, D.J. Stacking Ensemble Tree Models to Predict Energy Performance in Residential Buildings. *Sustainability* **2021**, *13*, 8298. [CrossRef]

46. Hajihassani, M.; Jahed Armaghani, D.; Kalatehjari, R. Applications of Particle Swarm Optimization in Geotechnical Engineering: A Comprehensive Review. *Geotech. Geol. Eng.* **2018**, *36*, 705–722. [CrossRef]

47. Harandizadeh, H.; Toufigh, M.M.; Toufigh, V. Application of improved ANFIS approaches to estimate bearing capacity of piles. *Soft Comput.* **2019**, *23*, 9537–9549. [CrossRef]

48. Mayerhof, G.G. Bearing capacity and settlemtn of pile foundations. *J. Geotech. Geoenviron. Eng.* **1976**, *102*, 196–228.

49. Armaghani, D.J.; Faradonbeh, R.S.; Rezaei, H.; Rashid, A.S.A.; Amnieh, H.B. Settlement prediction of the rock-socketed piles through a new technique based on gene expression programming. *Neural Comput. Appl.* **2016**, *29*, 1115–1125. [CrossRef]

50. Shahin, M.A. Intelligent computing for modeling axial capacity of pile foundations. *Can. Geotech. J.* **2010**, *47*, 230–243. [CrossRef]

51. Goh, A.T.C. Back-propagation neural networks for modeling complex systems. *Artif. Intell. Eng.* **1995**, *9*, 143–151. [CrossRef]

52. Pal, M. Modelling pile capacity using generalised regression neural network. In Proceedings of the Indian Geotechnical Conference, Kochi, India, 15–17 December 2011.

53. Alkroosh, I.; Nikraz, H. Predicting axial capacity of driven piles in cohesive soils using intelligent computing. *Eng. Appl. Artif. Intell.* **2012**, *25*, 618–627. [CrossRef]

54. Alavi, A.H.; Aminian, P.; Gandomi, A.H.; Esmaeili, M.A. Genetic-based modeling of uplift capacity of suction caissons. *Expert Syst. Appl.* **2011**, *38*, 12608–12618. [CrossRef]

55. Momeni, E.; Dowlatshahi, M.B.; Omidinasab, F.; Maizir, H.; Armaghani, D.J. Gaussian Process Regression Technique to Estimate the Pile Bearing Capacity. *Arab. J. Sci. Eng.* **2020**, *45*, 8255–8267. [CrossRef]

56. Goh, A.T.C. Nonlinear modelling in geotechnical engineering using neural networks. *Trans. Inst. Eng. Aust. Civ. Eng.* **1994**, *36*, 293–297.

57. Goh, A.T.C. Pile driving records reanalyzed using neural networks. *J. Geotech. Eng.* **1996**, *122*, 492–495. [CrossRef]

58. Kiefa, M.A.A. General regression neural networks for driven piles in cohesionless soils. *J. Geotech. Geoenvironmental Eng.* **1998**, *124*, 1177–1185. [CrossRef]

59. Das, S.K.; Basudhar, P.K. Undrained lateral load capacity of piles in clay using artificial neural network. *Comput. Geotech.* **2006**, *33*, 454–459. [CrossRef]

60. Pal, M.; Deswal, S. Modeling pile capacity using support vector machines and generalized regression neural network. *J. Geotech. Geoenvironmental Eng.* **2008**, *134*, 1021–1024. [CrossRef]

61. Pal, M.; Deswal, S. Modelling pile capacity using Gaussian process regression. *Comput. Geotech.* **2010**, *37*, 942–947. [CrossRef]

62. Gandomi, A.H.; Alavi, A.H. A new multi-gene genetic programming approach to nonlinear system modeling. Part I: Materials and structural engineering problems. *Neural Comput. Appl.* **2012**, *21*, 171–187. [CrossRef]

63. Kordjazi, A.; Nejad, F.P.; Jaksa, M.B. Prediction of ultimate axial load-carrying capacity of piles using a support vector machine based on CPT data. *Comput. Geotech.* **2014**, *55*, 91–102. [CrossRef]

64. Ghorbani, B.; Sadrossadat, E.; Bazaz, J.B.; Oskooei, P.R. Numerical ANFIS-based formulation for prediction of the ultimate axial load bearing capacity of piles through CPT data. *Geotech. Geol. Eng.* **2018**, *36*, 2057–2076. [CrossRef]

65. Dehghanbanadaki, A.; Khari, M.; Amiri, S.T.; Armaghani, D.J. Estimation of ultimate bearing capacity of driven piles in c-φ soil using MLP-GWO and ANFIS-GWO models: A comparative study. *Soft Comput.* **2020**, *25*, 4103–4119. [CrossRef]

66. Pham, T.A.; Ly, H.-B.; Tran, V.Q.; Van Giap, L.; Vu, H.-L.T.; Duong, H.-A.T. Prediction of pile axial bearing capacity using artificial neural network and random forest. *Appl. Sci.* **2020**, *10*, 1871. [CrossRef]

67. Zabidi, H.; De Freitas, M.H. Re-evaluation of rock core logging for the prediction of preferred orientations of karst in the Kuala Lumpur Limestone Formation. *Eng. Geol.* **2011**, *117*, 159–169. [CrossRef]

68. Gandomi, A.H.; Fridline, M.M.; Roke, D.A. Decision tree approach for soil liquefaction assessment. *Sci. World J.* **2013**, *2013*, 346285. [CrossRef]

69. Ramesh Murlidhar, B.; Yazdani Bejarbaneh, B.; Jahed Armaghani, D.; Mohammed, A.S.; Mohamad, E.T. Application of Tree-Based Predictive Models to Forecast Air Overpressure Induced by Mine Blasting. *Nat. Resour. Res.* **2021**, *30*, 1865–1887. [CrossRef]

70. Tiryaki, B. Predicting intact rock strength for mechanical excavation using multivariate statistics, artificial neural networks, and regression trees. *Eng. Geol.* **2008**, *99*, 51–60. [CrossRef]

71. Hasanipanah, M.; Faradonbeh, R.S.; Amnieh, H.B.; Armaghani, D.J.; Monjezi, M. Forecasting blast-induced ground vibration developing a CART model. *Eng. Comput.* **2017**, *33*, 307–316. [CrossRef]

72. Khalilia, M.; Chakraborty, S.; Popescu, M. *Predicting Disease Risks from Highly Imbalanced Data Using Random Forest*; Springer Link: New York, NY, USA, 2011.

73. Aghaabbasi, M.; Shekari, Z.A.; Shah, M.Z.; Olakunle, O.; Armaghani, D.J.; Moeinaddini, M. Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques. *Transp. Res. Part A Policy Pract.* **2020**, *136*, 262–281. [CrossRef]

74. Kardani, N.; Zhou, A.; Nazem, M.; Shen, S.-L. Estimation of bearing capacity of piles in cohesionless soil using optimised machine learning approaches. *Geotech. Geol. Eng.* **2020**, *38*, 2271–2291. [CrossRef]

75. Zhou, J.; Qiu, Y.; Zhu, S.; Armaghani, D.J.; Khandelwal, M.; Mohamad, E.T. Estimation of the TBM advance rate under hard rock conditions using XGBoost and Bayesian optimization. *Undergr. Space* **2021**, *6*, 506–515. [CrossRef]

76. Nguyen, H.; Bui, X.N.; Choi, Y.; Lee, C.W.; Armaghani, D.J. A Novel Combination of Whale Optimization Algorithm and Support Vector Machine with Different Kernel Functions for Prediction of Blasting-Induced Fly-Rock in Quarry Mines. *Nat. Resour. Res.* **2021**, *30*, 191–207. [CrossRef]

77. Zhou, J.; Qiu, Y.; Armaghani, D.J.; Zhang, W.; Li, C.; Zhu, S.; Tarinejad, R. Predicting TBM penetration rate in hard rock condition: A comparative study among six XGB-based metaheuristic techniques. *Geosci. Front.* **2021**, *12*, 101091. [CrossRef]
78. Raschka, S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. *arXiv* **2018**, arXiv:1811.12808.
79. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]
80. Xu, H.; Zhou, J.; Asteris, G.P.; Jahed Armaghani, D.; Tahir, M.M. Supervised Machine Learning Techniques to the Prediction of Tunnel Boring Machine Penetration Rate. *Appl. Sci.* **2019**, *9*, 3715. [CrossRef]
81. Harandizadeh, H.; Armaghani, D.J.; Mohamad, E.T. Development of fuzzy-GMDH model optimized by GSA to predict rock tensile strength based on experimental datasets. *Neural Comput. Appl.* **2020**, *32*, 14047–14067. [CrossRef]
82. Kardani, N.; Bardhan, A.; Samui, P.; Nazem, M.; Zhou, A.; Armaghani, D.J. A novel technique based on the improved firefly algorithm coupled with extreme learning machine (ELM-IFF) for predicting the thermal conductivity of soil. *Eng. Comput.* **2021**. Available online: https://link.springer.com/article/10.1007/s00366-021-01329-3 (accessed on 24 October 2021). [CrossRef]
83. Armaghani, D.J.; Yagiz, S.; Mohamad, E.T.; Zhou, J. Prediction of TBM performance in fresh through weathered granite using empirical and statistical approaches. *Tunn. Undergr. Space Technol.* **2021**, *118*, 104183. [CrossRef]
84. Armaghani, D.J.; Mohamad, E.T.; Momeni, E.; Narayanasamy, M.S. An adaptive neuro-fuzzy inference system for predicting unconfined compressive strength and Young's modulus: A study on Main Range granite. *Bull. Eng. Geol. Environ.* **2015**, *74*, 1301–1319. [CrossRef]
85. Armaghani, D.J.; Harandizadeh, H.; Momeni, E.; Maizir, H.; Zhou, J. An optimized system of GMDH-ANFIS predictive model by ICA for estimating pile bearing capacity. *Artif. Intell. Rev.* **2021**. Available online: https://link.springer.com/article/10.1007/s10462-021-10065-5 (accessed on 24 October 2021). [CrossRef]