Proceedings of the 17th Pan-American Conference on Soil Mechanics and Geotechnical Engineering (XVII PCSMGE), and 2nd Latin-American Regional Conference of the International Association for Engineering Geology and the Environment (IAEG), La Serena Chile, 2024.

Machine learning applied to estimate the specific weight of mining tailings

Aprendizado de máquina aplicado para estimativa do peso específico de rejeitos de mineração

Helena Paula Nierwinski & Ricardo José Pfitscher
Full Professor, Federal University of Santa Catarina, Brazil, helena.paula@ufsc.br

Talita Menegaz
Ph.D student in Geotechnical Engineer, University of São Paulo, Brazil

ABSTRACT: The soil-specific weight is a property of fundamental importance for the determination of stresses in Geotechnical Engineering designs and for the interpretation of in situ test results. In many cases obtaining it directly through undisturbed samples is not feasible. Therefore, empirical equations are often used to obtain this property. Several correlations in the literature allow estimations of the specific weight of natural soils. However, concerns arise regarding materials whose soil-specific gravity (G) is outside the range of natural soils for which the correlations were developed, as is the case of mining tailings designs. Therefore, the present paper aims to evaluate machine learning techniques for estimating specific weights in mining tailings. To this end, this work relies on a database with results of field and laboratory tests carried out in different Brazilian mining tailings deposits to estimate the soil specific-weight with an increased precision. A multiple linear regression and an artificial neural network model were proposed, with the latter demonstrating greater similarity between the estimated and measured specific weight values (R^2=0.91 considering global dataset).

KEYWORDS: Soil specific weight; CPTu tests; Multiple linear regression; Artificial neural networks.

1 INTRODUCTION

Geotechnical investigation is a mandatory practice for the development of any geotechnical design. The results obtained from the geotechnical investigation allow the definition of the design parameters and the prediction of the soil's behavior in the most diverse conditions imposed by engineering works (Schnaid & Odebrecht 2012).

Lunne *et al*. (1997) state that although field and laboratory tests complement each other, the vast majority of geotechnical designs are supported by field tests. Field tests provide information regarding the entire soil stratigraphy more quickly and economically. Based on field test results, specific profile layers can be selected to collect samples for laboratory tests. Laboratory tests, in turn, are only carried out at specific depths, and the information they provide often becomes extrapolated to thick adjacent layers.

A field test widely used in geotechnical campaigns, mainly for the characterization of soil deposits of lower resistance and in saturated conditions, is the Cone Penetration Test (CPTu) (ASTM D5778:20). This test provides records of three soil properties simultaneously (cone tip resistance - q_t, sleeve friction - f_s and pore pressure - u), obtained every 2 cm, from electrical sensors installed in a metallic conical tip, which is driven into the ground at a constant speed. Despite being quite robust, the CPTu test does not allow the collection of soil samples and, therefore, some soil properties need to be estimated, based on the measured parameters, to interpret the test results, or there is dependence on test results laboratory tests carried out on samples collected promptly at the study site.

The natural specific weight of the soil (γ_t) is a fundamental parameter for accurately interpreting field tests (such as CPTu), comprehending soil behavior, and formulating geotechnical designs. For example, the geotechnical engineer must define the soil's natural specific weight value to assess the geostatic stresses and evaluate material strength parameters. In current practice, the specific weight is defined through empirical correlations based on field test results or obtained through the characterization of samples used in special laboratory tests (e.g., direct shear, triaxial tests, consolidation tests). For tests carried out in the laboratory to have reliable results, it is necessary to collect an undisturbed sample, that is, one that preserves the characteristics of the material in the field (Coile 1936, Stewart 1943). However, the procedures to obtain an undisturbed sample require time and costs for correct execution and provide specific parameters, which will be extrapolated to the entire area when the project is developed.

Additionally, in some types of soils, like mining tailings, that are cohesionless, the standard methods to collect undisturbed samples need to be revised, and it is crucial to possess modern and advanced technological equipment and a well-trained staff. However, in South America, particularly in Brazil, there is a scarcity of resources necessary for geotechnical research and a need for more emphasis on comprehensive geotechnical-geological studies. As a result, accurately measuring specific weight becomes challenging, leading many designs to rely on standard values found in the literature for these parameters.

Using the CPTu test results it is possible to apply empirical or statistical formulas to estimate the natural soil-specific weight (Mayne 2014). However, most of these techniques encompass empirical formulas derived from natural soil datasets, with a specific gravity of solids (G) ranging typically from 2.5 to 2.7 (Mayne 2007). However, particular characteristics in specific type of soils hinder the adoption of such formulation with reliability. For instance, organic soils (Lengkeek *et al*. 2018, Straz and Borowiec 2020) and mining tailings (Menegaz *et al*. 2022), have non-usual properties that the formulation cannot achieve.

In the last years, some researchers introduced statistical regressions and machine learning techniques to estimate the soil specific weight. Considering soil characteristics and employing statistical regressions, Mayne (2014) has developed an equation that incorporates variations in q_t, f_s and, m_q (cone resistance-depth ratio) to estimate the γ_t. The research demonstrates a satisfactory fit, with an R^2 value varying around 0.62. Similarly, Robertson and Cabal (2010) have employed cone dimensionless parameters of resistance (q_t/σ_{atm}), sleeve friction ($Rf = f_s/\sigma_{atm}$), and the average specific gravity (G). They have proposed two different equations; however, the original study did not provide an R^2 value for estimation. Nevertheless, Menegaz *et al.* (2022) applied the formulations to various soils and found R^2 values ranging from 0.2 to 0.79.

More recently, machine-learning techniques were introduced to develop models for soil weight estimation. Straz and Borowiec (2020) applied an artificial neural network (ANN) model to provide estimations of γ_t for organic soils based on laboratory-determined leading parameters and obtained an R^2 value of 0.94. Nierwinski *et al.* (2023) proposed a practical method for determining the soil specific weight based on CPTu test results and specific gravity (G) (including the G parameter allows the differentiation between soil types, representing an intrinsic soil characteristic). First, the authors proposed a multiple regression model. After uncovering the hidden relationships between CPTu parameters and γ_t in different soil types, they also proposed an ANN model, obtaining an R^2 of 0.82. The methods were available as an internet-based software tool allowing someone to load a dataset and automatically calculate the soil specific weight.

Acknowledging the necessity of accurately estimating soils' specific weight and considering the favorable outcomes of machine learning-based models highlighted in existing literature, this article endeavors to employ machine-learning techniques to develop a model for specific weight estimation in mining tailings. The development of the model improved the methodology used in Nierwinski *et al.* (2023) by including the soil type resulting from a cluster analysis as input for the estimation models. Distinctly from the general model proposed in the former work, this improvement makes the model more specific for mining tailings, increasing the precision of estimates in such soils. Also, the training phase relies on a database composed of CPTu and G test results solely from Brazilian mining tailings, encompassing bauxite, iron, zinc, and gold mining tailings.

2 ESTIMATIONS OF SOIL ESPECIFIC WEIGHT USING CPTU TEST RESULTS – STATE OF ART

A proposal for estimating the specific weight from data obtained by the CPTu test was initially presented by Robertson and Cabal (2010). These authors introduced two dimensionless equations aimed at determining γ_t. The first equation (Eq. 1) establishes γ_t as dependent on the friction ratio (Rf) and corrected tip resistance (q_t). The second one (Eq. 2) incorporates the specific gravity of solids (G) to estimate the specific weight of soils, particularly those where G falls outside the range of 2.5 to 2.7.

$$\gamma_t/\gamma_w = 0.27\left[\log R_f\right] + 0.36\left[\log(q_t\sigma_{atm})\right] + 1.236 \quad (1)$$

where q_t = corrected tip resistance and σ_{atm} = atmospheric pressure (kPa).

$$\gamma_t/\gamma_w = \frac{\left[0.27\left[\log R_f\right] + 0.36\left[\log\left(\frac{q_t}{\sigma_{atm}}\right)\right] + 1.236\right]\mathrm{G}}{2.65} \quad (2)$$

In subsequent study, Mayne and Peuchen (2013) explore the correlation between the soil specific weight and their plasticity index, leading to the development of Equations (3) and (4). Since cone tests lack the ability to directly measure soil plasticity, the authors utilized a relationship between tip resistance and cone depth ($m_q = q_t/z$), where z represents the cone depth.

$$\gamma_t = \gamma_w + \mathrm{m_q}/8 \quad (3)$$

$$\gamma_t = 0.636\,(q_t)^{0.072}(10 + \mathrm{m_q}/8) \quad (4)$$

Another parameter derived from CPTu tests and employed in soil specific weight estimation is the sleeve cone friction (f_s). Mayne (2014) introduces two statistical equations for estimating γ_t based on sleeve friction. While Equation (3) demonstrates a close fit with a nonlinear relationship, Equation (6) is derived through linear regression.

$$\gamma_t = 26 - \frac{14}{1 + [0.5\log(f_s+1)]^2} \quad (5)$$

$$\gamma_t = 12 + 1.5\ln(f_s + 1) \quad (6)$$

Taking into account the limitations of regression-based models and the promising outcomes demonstrated in estimations using machine learning, Nierwinski *et al.* (2023) have proposed an approach supported by machine learning, employing both regression and artificial neural network (ANN) models and utilizing data extracted from CPTu tests. The dataset analyzed by the authors comprises 1862 entries, containing geotechnical data from 10 distinct soil types (45% sourced from sites in Brazil). The regression model proposed by Nierwinski *et al.* (2023) is a multiple variable model, with the values of q_t and u being deemed significant for the model. Furthermore, the authors of this study found that transforming the variables to a logarithmic scale enhanced the accuracy of the estimations. Equation 7 represents the multilinear regression model proposed by Nierwinski *et al.* (2023) for estimating the specific weight of soils:

$$\gamma_t = -1.1795 + 3.33\mathrm{G} + 2.90\log_{10} q_t + 0.21\log_{10} u \quad (7)$$

The ANN model proposed by Nierwinski *et al.* (2023), on the other hand, incorporated all the variables obtained in the CPTu test (q_t, f_s, and u) as well as the G parameter. The authors illustrated that the developed ANN model was more suitable for estimating the specific weight, given the input data evaluated. Figure 1 depicts the ANN model proposed by the authors, applied to the global database under study.

Nierwinski *at al.* (2023) also compared the estimates generated by applying other proposals from the literature, using the same database employed in the development of their models. Table 1 illustrates the R^2 values obtained for each of the equations (previous presented), highlighting the achievement of higher R^2

values through the models developed in their study, utilizing machine learning tools.

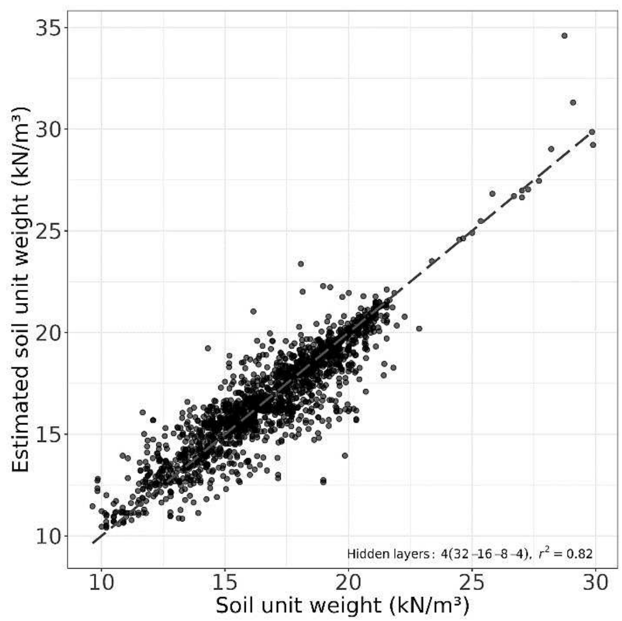


Figure 1. Soil specific weight estimation using artificial network model (source: Nierwinski *et al*. 2023)

Table 1. Reliability comparison between different models for soil specific weight estimative (source: adapted from Nierwinski *et al*. 2023)

Reference Model	*Soil type considered*	R^2
Equation 1	all soil types	0.39
Equation 2	all soil types	0.52
Equation 3	soft to firm normally consolidated clays	0.08
Equation 4	soft to firm normally consolidated clays	0.08
Equation 5	all soil types	0.34
Equation 6	all soil types	0.34
Equation 7	all soil types	0.61
ANN model	all soil types	0.82

3 DATASET PRESENTATION AND METHODOLOGY

This section introduces the dataset utilized to develop the model designed to estimate the specific weight of mining tailings, along with the methodological procedures employed in its development.

3.1 *Dataset*

This paper utilizes data obtained from experimental investigation campaigns conducted by various companies in Brazil. The dataset comprises 219 entries containing geotechnical data from four distinct types of tailings: bauxite, iron, zinc, and gold mining tailings, distributed according to the depicted in Figure 2. These campaigns provide data of corrected cone tip resistance (q_t), sleeve friction (f_s), and pore pressure (u_2) for all materials present in the database. The soil-specific gravity (G) value was available in 75% of the test reports, while for the remaining 25%, values were obtained from literature sources. Although this may bias the model, removing such values will reduce the model's generalization, as we rely on a small dataset compared to the current practice for machine learning purposes. To obtain the reference values for γ we analysed laboratory tests results, for which undisturbed samples were obtained (e.g. consolidation tests and triaxial tests). Subsequent items will elucidate the content of the database:

- Bauxite: The dataset comprises 159 entries (72.6%) of bauxite mining tailings. These entries are exclusively sourced from investigation campaigns conducted by a private company at sites located in the north and northeast regions of Brazil;
- Zinc: There are 35 samples of zinc mining tailings in the dataset, accounting for 15.98% of the total. This data originates from a field campaign conducted by a private company at a dam site in Minas Gerais, southeast Brazil;
- Iron: The dataset includes 16 entries (7.31%) of iron mining tailings collected during experimental campaigns conducted by a private company in Brazil;
- Gold: Data from a field trial conducted by a private company in Brazil contribute 9 entries (4.11%) to the dataset.

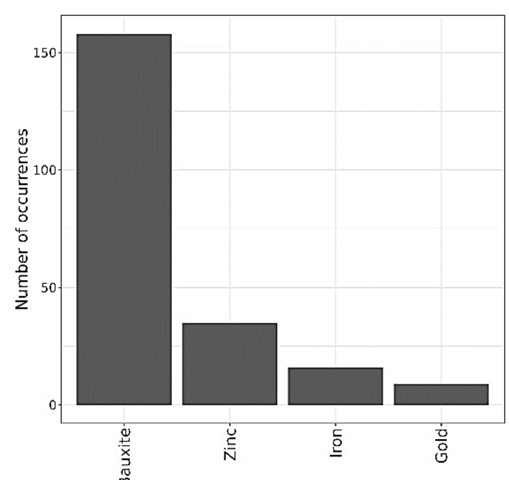


Figure 2. Distribution of soil types in the dataset

3.1 *Methodology*

The database initially underwent pre-processing to formulate models for predicting the specific weight of mining tailings. A clustering analysis was implemented following an investigation into whether distinct soil types exhibit analogous behaviours, potentially necessitating tailored models for each category. Given the prevalent use of statistical regressions in contemporary methodologies, a statistical analysis was conducted to estimate γ_t using a multiple linear regression model. An artificial neural network was developed to encapsulate the non-linear interrelations among geotechnical parameters.

The following items present details of the steps followed to develop the models.

3.1.1 *Cluster analysis*

A clustering analysis is an unsupervised machine learning method employed to categorize a dataset, aiming to group similar objects while distinguishing them from dissimilar ones (Jain *et al*. 1999). Following the establishment of the database, our focus shifted to identifying potential clusters within the data, enabling us to scrutinize specific correlations within each group, if present. Given its widespread popularity in the field, we opted for the K-means algorithm to conduct the clustering analysis. This algorithm endeavors to partition the dataset into K clusters by minimizing the sum of squared distances between each data point and the centroid of its corresponding cluster. We feed the K-means algorithm with G, q_t, f_s, and u values, removing the dependent variable γ from the dataset. We experimented with varying values of K, ranging from two (representing mining tailings and natural soils) to those indicated by silhouette and elbow techniques (Saputra *et al*. 2020). Subsequently, suppose the resulting clusters consist of mining tailings exhibiting similar characteristics (or behaviors). In that case, we consider the groupings generated by K-means valid, indicating the need to develop distinct models for each cluster.

3.1.2 *Multiple linear regression model*

Adhering to contemporary methodologies, the first model evaluated was linear regression. According to Chambers and Hastie (1992), linear regression models serve as statistical tools employed to examine the correlation between a dependent variable and one or more independent variables. When dealing with multiple regressions, the equation conforms to the formula outlined in Equation 8.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{8}$$

In Equation 8, y represents the forecasted value of the dependent variable, β_0 denotes the y-intercept, $\beta_1 X_1$ signifies the regression coefficient (β_1) associated with the first independent variable (X_1), and $\beta_n X_n$ denotes the regression coefficient about the final independent variable.

The initial step involves exploring the relationships between the dependent variable, γ_t, and the independent geotechnical parameters. In multiple linear regression, there could be intercorrelation among some independent variables, necessitating a thorough examination before model development. If the correlation coefficient (R^2) between two independent variables exceeds 0.6, only one of them should be included in the regression model (Bevans 2022). One approach to understanding such correlations entails analyzing scatter plots of each independent variable against the dependent variable, γ_t, and calculating Pearson's correlation coefficient. Following selecting appropriate independent variables, we employ the *lm* function within *R* software to determine the coefficients for the multiple linear regression models. The *lm* function uses the Ordinary Least Squares (OLS) algorithm and determines the set of values for the coefficients of independent variables that minimize the vector of errors or residuals (Chambers, 1992).

3.1.3 *Artificial neural network (ANN) model*

Artificial neural networks (ANNs) are computational models inspired by the architecture and operation of the human brain. ANNs are comprised of interconnected neurons organized into layers, with each neuron processing input data and transmitting it to subsequent layers until generating a final output. At its core, a neuron executes an activation function that aggregates multiple inputs and produces an output. The collective results of multiple neurons within a layer are then propagated to the next layer. For example, the threshold-based activation function introduced by McCulloch and Pitts (1943) calculates a weighted sum of input signals and yields an output of 1 if the sum surpasses a defined threshold, otherwise 0. The sigmoid-based activation function is widely adopted due to its smooth, continuous nature (Jain *et al*., 1996).

The initial layer receives inputs corresponding to the independent variables when employing Artificial Neural Networks (ANNs) for regression tasks. In contrast, the final layer generates the estimated value of the dependent variable. Throughout the training phase of the ANN model, the coefficients of each activation function within the neurons are adjusted to optimize the alignment between the output produced by the ANN and the independent variable. Given that activation functions can exhibit nonlinear behaviors, ANNs possess significant abstraction capabilities for conducting estimations. This study leverages the abstraction capabilities of ANNs by incorporating, as an independent variable (input for the initial ANN layer), the value generated by multiple linear regressions. This approach thereby encompasses potential nonlinearities that traditional linear regressions fail to capture.

. The structured architecture of the Artificial Neural Network (ANN) comprises six layers, four hidden layers. In an ANN, a hidden layer is a neuron layer between the input and output layers. The term "hidden" refers to the fact that this layer is not directly exposed to the input or output of the model. We conducted an iterative search to define such an architecture. With fewer layers, the ANN has the worst accuracy; with more layers, on the other hand, it does not improve the R^2 and increases the inference time. The initial layer serves to intake inputs for the model, encompassing parameters such as G, q_t, f_s, and u_2. Subsequently, the subsequent four layers remain concealed, with the first layer housing 32 neurons, the second containing 16, the third comprising 8, and the fourth encompassing 4 neurons. Ultimately, the computations within the fourth hidden layer yield the model's output: the estimated value of γ_t.

The activation function employed in every neuron within the Artificial Neural Network (ANN) is a sigmoid logistic function. As Anastasiadis *et al*. (2005) outlined, the algorithm utilized to calculate the network's weights is resilient back-propagation with weight backtracking. We randomly split the dataset regarding the training process, allocating 70% for training and 30% for testing purposes.

4 RESULTS AND DISCUSSION

This section covers the results obtained through the modeling carried out and discusses them.

4.1 *Cluster analysis*

The database underwent analysis to ascertain the ideal number of clusters, employing both the silhouette technique (Figure 3(a)) and the elbow technique (Figure 3(b)). Silhouette and elbow techniques are commonly used to evaluate and determine the optimal number of clusters in clustering algorithms such as K-means. The Silhouette method measures how similar an object is to its cluster compared to others. In interpreting the average silhouette plot (Figure 3(a)), the analyst selects the cluster number that results in the more significant average. The Elbow method is used to determine the optimal number of clusters by plotting the explained variation as a function of the number of clusters and looking for the "elbow point" where the rate of decrease sharply slows. In other words, the analyst should select the point stabilizing the WSS plot's decrease (Figure 3(b)). Both techniques indicate the presence of two distinct clusters. Figure 4 illustrates a distinct partition within the dataset concerning the behavior of variables G and γ_t, which directly correlate with soil type delineation. Given the presence of these two distinct groups within the dataset, the material type was a variable considered for constructing the regression model.

Table 2 shows the clustering results with the number of soils of each type classified for each group. As one can see, there is a clear distinction between the groups, with Cluster 1 comprising bauxite, zinc, and gold tailings and Cluster 2 with iron tailings.

Table 2. Soil type distribution among clusters

	Soil type			
	Bauxite	Zinc	Iron	Gold
Cluster 1	158	35	1	9
Cluster 2	0	0	15	0

4.2 *Multiple linear regression model*

This section outlines the procedures for developing a statistical model using linear regression to estimate soil specific weight based on the examined database. The model considers two clusters, as demonstrated in section 4.1. The next step in constructing this model involves assessing the relationships between the model's potential parameters and the data's behavior when juxtaposed.

The scatter plot presented in Figure 4 illustrates the correlations among G, q_t, f_s, u, and γ_t (designated as *gamma*). Correlation analysis reveals a robust association between q_t, f_s, and G (R^2 = 0.608). The variable G represents an intrinsic material characteristic facilitating its identification. Both q_t and f_s stem from the Cone Penetration Test (CPTu), providing similar explanatory power within the model. Therefore, considering q_t's extensive historical usage and its strong correlations with parameters in the literature, only q_t will be incorporated into the current model. Furthermore, examination of Figure 4 data demonstrates the limited contribution of the variable u in elucidating specific weight (*gamma*), warranting its exclusion from the model. Ultimately, Figure 4 indicates a nonlinear relationship between q_t (and f_s) and specific weight (*gamma*), suggesting the potential necessity for logarithmic transformation before regression analysis.

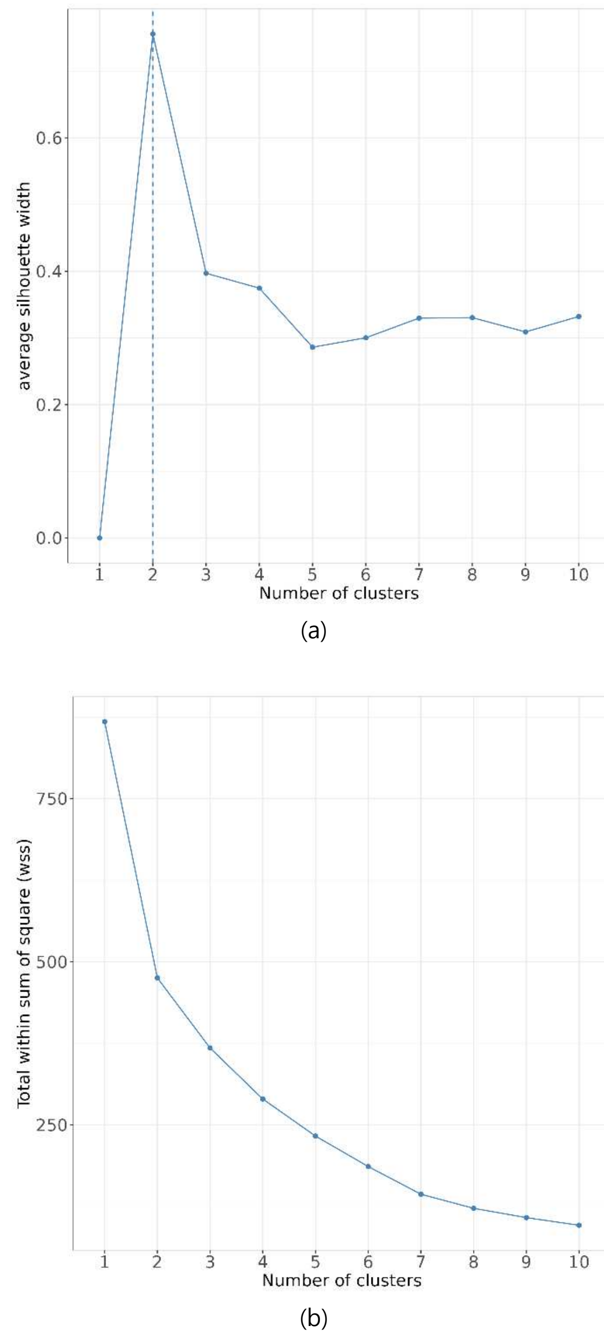


Figure 3. Silhouette (a) and elbow (b) techniques results for determining the number of clusters of dataset

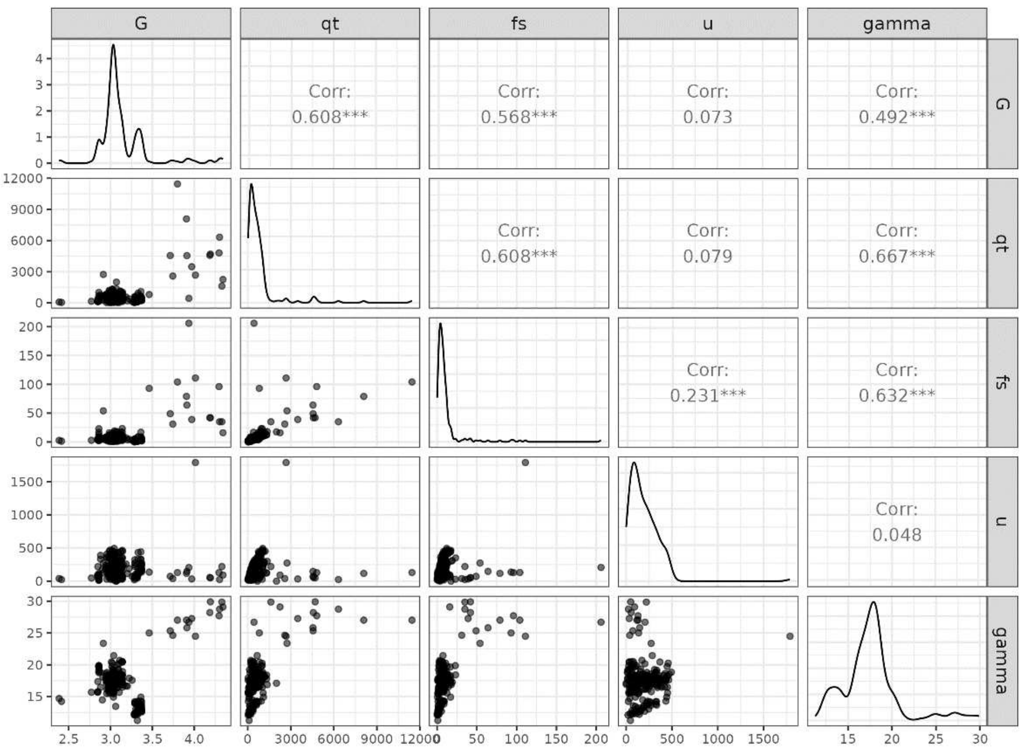


Figure 4. Scatter plot for the correlations between G, q_t, f_s, u and γ_t (*gamma*)

Thus, the variables G and q_t were utilized in constructing the multiple linear regression model, with the value of q_t varying logarithmically in correlation with specific weight. Additionally, the model considered the soil type from each of the two pre-defined clusters. Equation 9 presents the statistical model derived from a multiple linear regression encompassing the complete dataset, yielding an adjusted R^2 of 0.73 for estimations, as shown in Figure 5.

$$\gamma_t = 26.7 + 2.43 \log_{10} q_t - 5.24\mathrm{G} + 12.8 soilType \qquad (9)$$

where *soilType* refers to the cluster considered, assuming 0 for cluster 1 and 1 for cluster 2.

The data points on the scatter plot in Figure 5 appear to follow a linear trend, indicating a robust correlation between the estimated and measured values of soil specific weight within the model. Moreover, the proposed model yielded a higher R^2 value compared to other linear regression models discussed in section 2 of this study.

4.3 *Artificial Neural Network (ANN) model*

To try to improve specific weight estimative, we employed an Artificial Neural Network (ANN) model to predict the mining tailing specific weight utilizing the geotechnical parameters from the dataset as inputs. Given the enhanced abstraction capabilities of neural network models, we incorporated all variables (G, q_t, f_s, and u) without applying log transformation, as considered for multiple linear regression models.

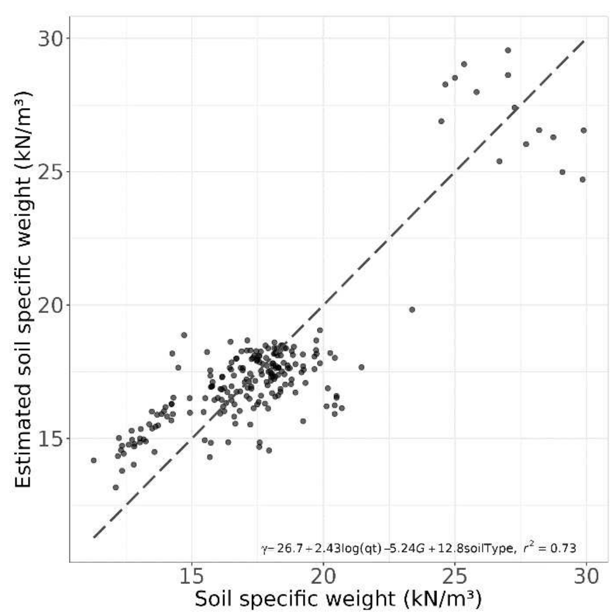
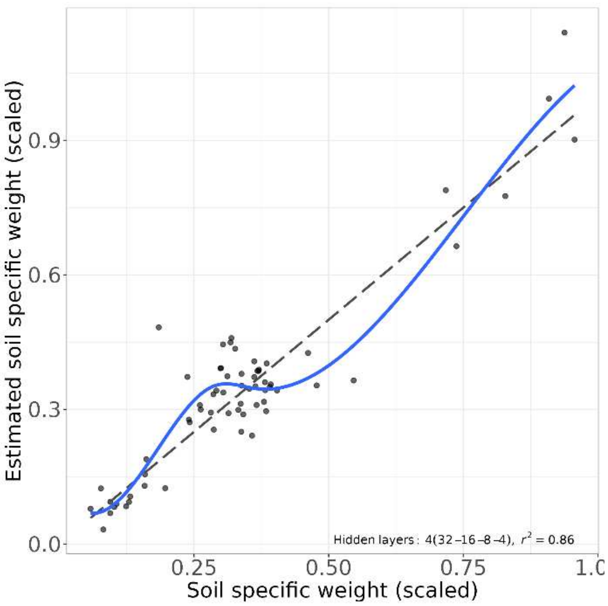
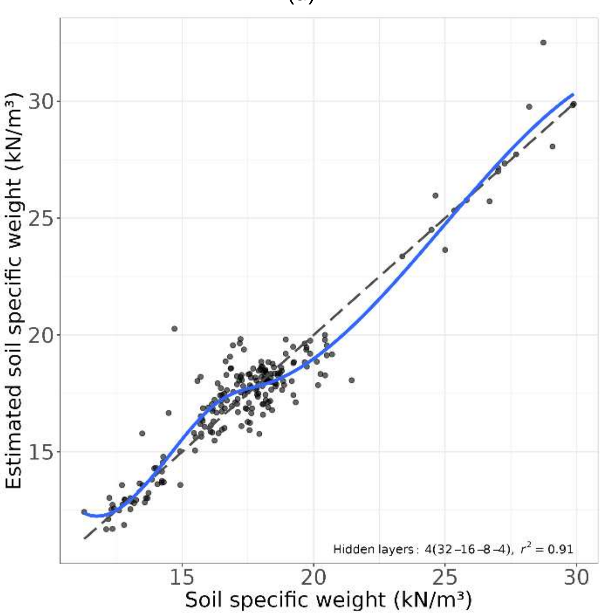


Figure 5. Scatter plot for the comparison between estimated values of γ_t using the linear model and the measured values

Figure 6(a) illustrates the scaled values of γ_t estimated by the ANN model for the test dataset (30%) compared to the actual values. The model tested with 30% of the data yielded an $R^2 = 0.86$, surpassing those estimated by Equation (9). However, upon examining the model tested with the entire dataset (Figure 6(b)), it becomes evident that it provides the best fit, with an $R^2 = 0.91$. This outcome underscores that an ANN model trained with in-situ parameters is the most suitable approach for estimating the specific weight of mining tailings.



(a)



(b)

Figure 6. Specific weight estimations using ANN models: predictions for the test data (a) and all data (b)

4.4 *Results summary*

To compare our research findings to the state-of-the-art, we summarize in Table 3 the R^2 values obtained through the models presented in this paper and the obtained with the literature equations discussed in Section 3. The results show an improvement in all models' precision when used to estimate the specific weight of mining tailings and that the proposed models that use the clustering identification offer the best values of R^2 (0.73 for the regression model and 0.91 for the ANN model).

Table 3. Comparison of the results with the state-of-art estimations applied to the mining tailings dataset

Reference Model	R^2
Equation 1	0.47
Equation 2	0.50
Equation 3	0.39
Equation 4	0.41
Equation 5	0.54
Equation 6	0.53
Regression model (this research)	0.73
ANN model (this research)	0.91

5 CONCLUSIONS

In contemporary practice, soil specific weight determination typically relies on empirical correlations derived from field test results or characterization of undisturbed samples via laboratory tests such as simple direct shear, triaxial tests, and consolidation tests. Recent advancements in the field have embraced statistical equations tailored to individual soil characteristics to overcome the limitations of laboratory tests. While these statistical-based methods yield accurate estimations for soils incorporated in the regressions, they struggle to estimate γ_t for distinct soil types. This study explores the application of machine learning as a pragmatic approach to estimating mining tailings specific weight (γ_t) based on parameters gathered from CPTu tests.

Considering the results and analysis of this research, the conclusions can be summarized as the following points:

- clustering analysis revealed that the dataset could be divided into two distinct groups, which was considered in the multiple regression model development;
- it is found that to estimate γt using a multiple linear regression model, G, q_t, and, the *soilType* (clusters) were considered significant variables and the q_t variable must be logarithmically transformed. The multiple linear model achieves an R^2 of 0.73;
- an Artificial Neural Network (ANN) model was determined, enabling estimates within the database evaluated with R^2 of 0.91, considering global data and

promising applications for specific weight estimates in mining tailings.

6 ACKNOWLEDGEMENTS

The authors would like to acknowledge Brazilian Companies for gently providing the reports to develop our database.

7 REFERENCES

Anastasiadis A.D., Magoulas G.D., Vrahatis M.N., 2005. New globally convergent training scheme based on the resilient propagation algorithm. *Neurocomputing*, 64, 253–270.

ASTM D5778. 2020. *Standard Test Method for Electronic Friction Cone and Piezocone Penetration Testing of Soils*, 2020 Edition, June 1, 2020.

Bevans R. 2022. *Multiple Linear Regression | A Quick Guide (Examples)*. URL https://www.scribbr.com/statistics/multiple-linear-regression/.

Chambers J., Hastie T. 1992. *Statistical Models in S*. Wadsworth & Brooks/Cole Advanced Books & Software.

Chambers J. M. 1992. Linear models. Chapter 4 of Statistical Models in S eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

Coile T. S. 1936. Soil samplers. *Soil Sci*. 42: 139-142.

Jain A.K., Murty M.N., Flynn, P.J., 1999. Data Clustering: A Review. Vol. 31, no. 3. *Association for Computing Machinery*, New York, NY, USA, pp. 264–323.

Jain A.K., Mao J., Mohiuddin K.M. 1996. Artificial neural networks: A tutorial. *Computer* 29 (3), 31–44.

Lengkeek H. J.; De Greef J.; Joosten S. 2018. CPT based unit weight estimation extended to soft organic soils and peat. *In: Cone Penetration Testing* . CRC Press, 2018. p. 389-394.

Lunne T., Robertson P. K. and Powell J.J.M. 1997. *Cone Penetration Testing*. Londres: Blackie Academic & Professional.

Mayne P. W. 2007. *Cone penetration testing*. Transportation Research Board.

Mayne P., 2014. Interpretation of geotechnical parameters from seismic piezocone tests. *In: Proceedings, 3rd International Symposium on Cone Penetration Testing*. pp. 47–73.

Mayne P., Peuchen, J., 2013. Unit weight trends with cone resistance in soft to firm clays. *In: Geotechnical and Geophysical Site Characterization: Proceedings of the 4th International Conference on Site Characterization ISC-4*. Vol. 1. Taylor & Francis Books Ltd, pp. 903–910.

McCulloch W.S., Pitts W. 1943. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys*. 5, 115–133.

Menegaz T., Odebrecht E., Nierwinski H.P., Schnaid F. 2022. Soil unit weight prediction from CPTs for soils and mining tailings. *In: Cone Penetration Testing 2022*. CRC Press, pp. 566–569.

Nierwinski H. P.; Pfitscher R. J.; Barra B. S.; Menegaz T.; Odebrecht E. 2023. A practical approach for soil unit weight estimation using artificial neural networks. *Journal of South American Earth Sciences*, V.131, 2023, 104648, ISSN 0895-9811.

Robertson P.K., Cabal K. 2010. Estimating soil unit weight from CPT. *In: 2nd International Symposium on Cone Penetration Testing*. pp. 2–40.

Saputra D.M., Saputra D., Oswari L.D. 2020. Effect of distance metrics in determining k-value in k-means clustering using elbow and silhouette method. *In: Sriwijaya International Conference on Information Technology and Its Applications*. SICONIAN 2019, Atlantis Press, pp. 341–346.

Schnaid F.; Odebrecht, E. 2012. *Ensaios de Campo e suas aplicações à Engenharia de Fundações*. (in portuguese). 2nd Edition. São Paulo: Oficina de Textos.

Stewart A. 1943. A soil tube for obtaining wet clay cores in an undisturbed structural condition. *Soil Sci*. 55 (3), 247–252.

Straż G., Borowiec A., 2020. Estimating the unit weight of local organic soils from laboratory tests using artificial neural networks. *Appl. Sci*. 10 (7), 2261.