**ORIGINAL ARTICLE**

# A novel approach based on soft computing techniques for unconfined compression strength prediction of soil cement mixtures

Joaquim Tinoco[1] · António Alberto[2] · Paulo da Venda[3] · António Gomes Correia[4] · Luís Lemos[3]

## Abstract

The prediction of the uniaxial compression strength ($q_u$) of soil cement mixtures is of up most importance for design purposes. This is done traditionally by extensive laboratory tests which is time and resources consuming. In this paper, it is presented a new approach to assess $q_u$ over time based on the high learning capabilities of data mining techniques. A database of 444 records, encompassing cohesionless to cohesive and organic soils, different binder types, mixture conditions and curing time, were used to train three models based on support vector machines (SVMs), artificial neural networks (ANNs) and multiple regression. The results show a promising performance in $q_u$ prediction of laboratory soil cement mixtures, being the best results achieved with the SVM model ($R^2 = 0.94$) and with an average of SVM and ANN model ($R^2 = 0.95$), well reproducing the major effects of the input variables water/cement ratio, cement content, organic matter content and curing time, which are known as preponderant in soil cement mixtures behaviour.

**Keywords** Soil–cement mixtures · Jet grouting · Deep soil mixing · Soft computing · Data mining · Sensitivity analysis

## 1 Introduction

Mechanical properties study of soil cement mixtures is a complex task due to high number of parameters involved. Over the last decades, several researches have been conducted, following different approaches but with the same purpose of a better understand of soil cement mixtures behaviour over time.

Concerning to uniaxial compression strength ($q_u$), this mechanical property is obtained through laboratory tests that involve time and resources consuming, which are generally very limited. Therefore, it is important to reduce the number of laboratory tests without compromising safety or confidence issues. A common practice is to prepare (before construction works) and test some laboratory samples aiming to simulate the field conditions. These samples, prepared with the same soil, cement and water used in the field, will give an important idea about the behaviour of the infield mixture. However, this laboratory samples also represent an important cost for the project and therefore should be minimized.

This scenario underlines the necessity, at least upon at a pre-design stage, to have available prediction tools to obtain the best design parameters. However, due to the high number of parameters affecting the behaviour of soil cement mechanical properties, in particularly the $q_u$, the traditional statistical analysis are unable to deal with.

Aiming to overcome this limitation, a first and successful attempt has been recently made, taking advantage of the high learning capabilities of data mining (DM) techniques [6, 27–29], which have been successfully applied in the pass in different knowledge

✉ Joaquim Tinoco
jtinoco@civil.uminho.pt

1 ISISE - Institute for Sustainability and Innovation in Structural Engineering/ALGORITMI Research Center, School of Engineering, University of Minho, Guimarães, Portugal

2 CIEPQPF - Research Centre on Chemical Processes and Forest Products Engineering, Department of Civil Engineering, University of Coimbra, Coimbra, Portugal

3 ISISE - Institute for Sustainability and Innovation in Structural Engineering, Department of Civil Engineering, University of Coimbra, Coimbra, Portugal

4 ISISE - Institute for Sustainability and Innovation in Structural Engineering, School of Engineering, University of Minho, Guimarães, Portugal

domains [11, 12, 18], including in civil engineering field [13, 21].

Although a good performance has been achieved in $q_u$ prediction of laboratory soil cement mixtures (Fig. 1), there are some limitations that still need to be overcome. In particular, the model dependence on the mixture proprieties, such as its porosity, is one of its main drawbacks. As can be observed in Fig. 2, which shows the relative importance of each input variables in $q_u$ prediction, the mixture porosity (only measured after mixture preparation) has a relative importance higher than 15%. Moreover, the models shown in Fig. 1 were developed based on a database regarding soil cement samples covering higher dosages of cement [6]. Hence, aiming to eliminate models

dependence on the final mixtures properties, namely its porosity, as well as increase its applicability domain, a new data-driven model is here proposed for $q_u$ prediction over time without considering any property of the final soil cement mixture and covering a wide range of cement contents.

The proposed model, based on advanced statistics analysis usually known as data mining techniques, allows estimating $q_u$ of laboratory soil cement mixtures over time based on ten input variables such as the cement content, soil grain size distribution or type of binder. A cross-validation approach under 5 runs was applied for model generalization assessment.

## 2 Methodology

### 2.1 Modelling

For $q_u$ modelling it was followed a data-driven approach where three different DM algorithms were fitted to a database previously compiled and prepared containing unconfined compression tests results related to laboratory soil cement mixtures, as well as a set of ten input variables related to the soil and cement characteristics used to prepare the mixture. In particular, two of the high flexible learning DM algorithms were trained, namely Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs). As a baseline comparison also a Multiple Regression (MR) algorithm was fitted to the database.

Bellow is presented a brief overview of the three DM algorithms applied in this study, highlighting the adopted parameters for each one.

Initially developed for classification tasks [8], SVMs were latter adapted to regression tasks thanks to the introduction of $\epsilon$-insensitive loss function [26]. The main purpose of the SVMs is to transform input data into a high-dimensional feature space using nonlinear mapping. The SVM then finds the best linear separating hyperplane, related to a set of support vector points, in the feature space. This transformation depends on a kernel function. In this work, the popular Gaussian kernel was adopted. In this context, its performance is affected by three parameters: $\gamma$, the parameter of the kernel; C, a penalty parameter; and $\epsilon$ (only for regression), the width of an $\epsilon$-insensitive zone [14]. The heuristics proposed by Cherkassky and Ma [3] were used to define the first two parameter values, $C = 3$ (for a standardized output) and $\epsilon = \hat{\sigma}/\sqrt{N}$, where $\hat{\sigma} = 1.5/N \cdot \sum_{i=1}^{N}(y_i - \hat{y_i})^2$, $y_i$ is the measured value, $\hat{y_i}$ is the value predicted by a 3-nearest neighbour algorithm and $N$ is the number of examples. A grid search of
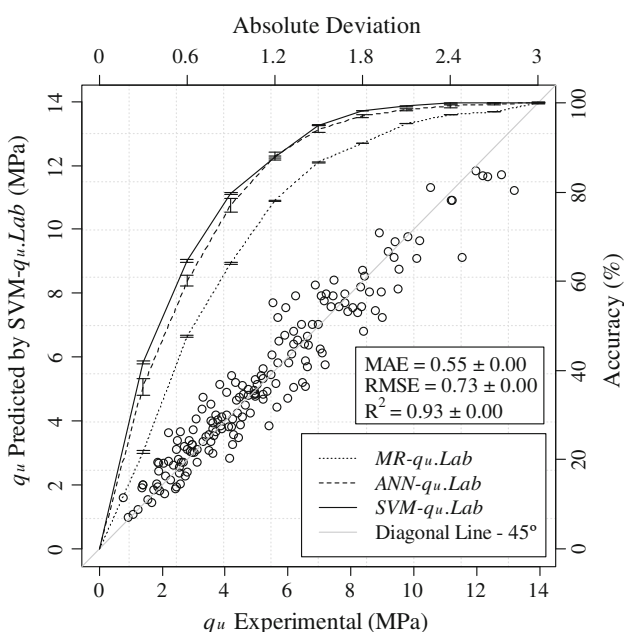


**Fig. 1** Data mining models performance in $q_u$ prediction (old model) of laboratory soil cement mixtures. (adapted from Correia et al. [6])
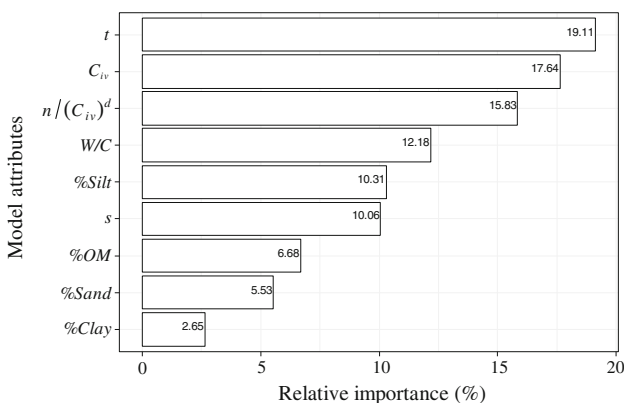


**Fig. 2** Relative importance of each input variable in $q_u$ prediction of laboratory soil cement mixtures according to SVM algorithm [28]

$2^{\{-1,-3,-7,-9\}}$ was adopted to optimize the kernel parameter $\gamma$, under an internal threefold cross-validation scheme.

Concerning to ANNs, they are a method of artificial intelligence, which seeks to simulate the biological structure of the human brain and nervous system through their architecture [19]. This concept was firstly introduced in 1943 by McCulloch and Pitts [23] although its use was expanded by Werbos [33] through the development of the backpropagation algorithm, which became a practical tool in the field of forecasting and prediction. ANNs are a technique capable of modelling complex nonlinear mappings and is robust in exploration of data with noise. In this study, the multilayer perceptron that contains only feed-forward connections, with one hidden layer containing $H$ processing units, was adopted. Because the network's performance is sensitive to $H$ (a trade-off between fitting accuracy and generalization capability), it was adopted a grid search (similar to the one used for SVM) of 0; 2; 4; 6; 8 during the learning phase to find the best $H$ value. Such grid search only considered training data, dividing it into fitting (70%) and validation data (30%), where the validation error was used to select the best $H$. After selecting the best $H$ value, the ANN is retrained with the whole training data. The neural function of the hidden nodes was set to the popular logistic function $1/(1 + e^{-x})$.

For a baseline comparison, also MR was implemented in this work. According to a MR algorithm several independent variables are linearly combined to predict the dependent (output) variable [15]. Due to its additive nature, this model is easy to interpret and is widely used in regression tasks. However, one of its main limitations is its inefficiency at modelling problems of a nonlinear nature.

R statistical environment [24], a free and open-source software, was used to conduct all experiments. The rminer package [9], which facilitates the implementation of several DM algorithms, namely ANNs and SVMs algorithms, as well as different validation approaches such as the cross-validation implemented in this study, was also adopted in this work.

## 2.2 Models evaluation

Models assessment is an important step after its training allowing to measure models performance, not only in terms of accuracy but also concerning to their interpretability.

For models comparison and accuracy measurement, three metrics currently used in regression problems were calculated [15]: mean absolute error (MAE), root mean square error (RMSE) and coefficient of correlation ($R^2$). A low value of MAE and RMSE and an $R^2$ close to the unit value means a higher predictive capacity. The main difference between MAE and RMSE is that the latter one is more sensitive to extreme values since it uses the square of the distance between the real and predicted values [27]. When compared with MAE, RMSE penalizes more heavily a model that in a few cases produces high errors. Thus, these two error measurements give different and complementary perspectives about the behaviour of the induced models, allowing its comparison. In addition to this three metrics, it was taken also advantage of regression error characteristic (REC) curve proposed by Bi and Bennett [1], which plots the error tolerance on the $x$-axis versus the percentage of points predicted within the tolerance on the $y$-axis, allowing a quick and easy comparison of different DM models.

Models generalization is another point when assessing a data-driven model. For that purpose, a cross-validation ($k$-fold = 10) approach [15] was applied and the entire process was repeated 5 times. A $k$-fold validation evaluates the data across the entire training set, but it does so by dividing the training set into $k$ folds (or subsections, where $k$ is a positive integer) and then training the model k times, each time leaving a different fold out of the training data and using it instead as a validation set. At the end, the performance metric (e.g. MAE, RMSE, etc.) is averaged across all $k$ tests. Lastly, as before, once the best parameter combination has been found, the model is retrained on the full data.

Understanding what was learned by the models is also a key point in any data-driven project. Since data-driven models, particularly SVM or ANN that rely on complex statistical analysis and are frequently referred to as "black boxes", are mathematically very complex it urges the necessity to "open" such models in order to facilitate its understanding. Aiming to overcome this drawback, Cortez and Embrechts [10] proposed a novel visualization approach based on sensitivity analysis (SA), which is used in this work. SA is a simple method that is applied after the training phase and measures the model responses when a given input is changed, allowing the quantification of the relative importance of each attribute as well as its average effect on the target variable. In particular, it was applied the Global Sensitivity Analysis (GSA) method [10], which is able to detect interactions among input variables. This is achieved by performing a simultaneous variation of $F$ inputs. Each input is varied through its range with $L$ levels and the remaining inputs fixed to a given baseline value. In this work, it was adopted the average input variable value as a baseline and set $L = 12$, which allows an interesting detail level under a reasonable amount of computational effort.

With the sensitivity response of the GSA, different visualization techniques can be computed. The input importance barplot shows the relative influence ($R_a$) of each input variable in the model (from 0 to 100%). The

rational of GSA is that the higher the changes produced in the output, the more important is the input. To measure this effect, first the gradient metric ($g_a$) for all inputs was calculated. After that, the relative influence was computed according to the following equation:

$$R_a = g_a \bigg/ \sum_{i=1}^{I} g_i \cdot 100(\%)$$

$$\text{where,} \; g_a = \sum_{j=2}^{L} |\hat{y}_{a,j} - \hat{y}_{a,j-1}|/(L-1) \tag{1}$$

where $a$ denotes the input variable under analysis and $\hat{y}_{a,j}$ is the sensitivity response for $x_{a,j}$.

## 2.3 Database

For models training and testing purposes, a database with 444 records was collected and compiled. These samples make part of different laboratory studies carried out on University of Minho and University of Coimbra with two main purposes. On one hand analyse the influence of several variables in mechanical behaviour of jet grouting and cutter soil mixing laboratory mixtures [27]. On the other hand define the binder mixture to obtain the best technical, economical and environmental soil stabilization [5, 7, 30–32].

The soils used in the preparation of the laboratory samples were collected from eight test sites. One of them is Coimbra area (located near Coimbra city, Portugal), ranging from cohesive to cohesionless soils, organic to nonorganic soils, presenting different geotechnical properties. Fourteen different binders were tested, including Portland cement, slag, fly ash, lime and silica fume, applied individually or combined. Concerning to the seven remaining sites, all of them are of a clayey nature, containing different percentages of sand, silt, clay and organic matter [27]. These samples were prepared with cement type CEM I 42.5R (Portland cement with 100% clinquer) and CEM II 42.5R (composed Portland cement with $\geq$ 65% clinquer). In addition, a couple of samples were also prepared with pozzolanic cement (CEM IV/A 35.5R with $\geq$ 20% clinquer).

As models input a set of 10 variables were selected. The definition of such variables took into account the empirical knowledge related to soil cement mixtures behaviour, particularly concerning to the $q_u$ evolution over time [2, 20, 22, 25]. Also the feedback obtained from the learning process was used in the input variables selection. Bellow are listed all 10 input variables considered in this study for $q_u$ prediction.

- Soil clay content (%)—%Clay
- Soil sand content (%)—%Sand
- Soil silt content (%)–%Silt
- Soil organic matter content (%)—%OM
- Water content (%)—$\omega_0$
- Cement content (%)—$a_w$
- Water/cement ratio—W/C
- Age of the mixture (days)—$t$
- Coefficient related with the binder type—$C_s$
- Coefficient related with a secondary binder—$L_2$

Table 1 summarizes the main statistics of all 10 inputs variables as well as of the output variable, showing the wide range of cement content as well as the $q_u$ values.

## 3 Results and discussion

This section summarizes the main achievements of this work that aims the development of a predictive model for $q_u$ of laboratory soil cement mixtures, through the application of advanced statistics analysis.

The average hyperparameters and fitting time values (and respective 95% level confidence intervals according to a t-student distribution) of the three DM algorithms trained for $q_u$ prediction of laboratory soil cement mixtures (i.e. MR, ANN and SVM) are shown in Table 2.

The achieved results shows a promising performance in $q_u$ prediction of laboratory soil cement mixtures based on the set of inputs selected that not include any information about the mixture properties. In fact, as shown in Table 3, both ANN and SVM algorithms were able to predict $q_u$ very accurately, haven achieved an $R^2 = 0.94$. Based on MAE or RMSE, it is possible to observe that SVM.Lab is able to predict $q_u$ with a slightly higher accuracy when compared with ANN.Lab. As expected, MR.Lab has achieved the lower performance with an $R^2 = 0.68$, which

**Table 1** Summary of the main statistics of the input and output variables used in $q_u$ prediction

| Variable | Minimum | Maximum | Mean | Standard deviation |
|---|---|---|---|---|
| %Clay | 0.00 | 45.00 | 19.84 | 14.31 |
| %Sand | 0.00 | 99.00 | 22.97 | 22.1 |
| %Silt | 1.00 | 79.00 | 57.17 | 18.15 |
| %OM | 0.00 | 19.40 | 5.87 | 4.64 |
| $\omega_0$ | 7.17 | 113.05 | 64.96 | 24.48 |
| $a_w$ | 3.00 | 284.32 | 55.42 | 69.21 |
| W/C | 0.63 | 10.91 | 3.30 | 2.05 |
| $t$ | 3.00 | 90.00 | 25.71 | 15.42 |
| $C_s$ | 0.20 | 0.38 | 0.22 | 0.06 |
| $L_2$ | 0.00 | 1.00 | 0.61 | 0.49 |
| $q_u$ | 0.10 | 13.19 | 2.77 | 2.72 |

**Table 2** Hyperparameters and computation time for each fitted model

| Model | Hyperparameter | Time (s) |
|---|---|---|
| MR.Lab | – | $0.41 \pm 0.03$ |
| ANN.Lab | $H = 7 \pm 1$ | $17.18 \pm 0.45$ |
| SVM.Lab | $\gamma = 0.21 \pm 0.04;\ C = 4.84 \pm 0.22;\ \epsilon = -5.57 \pm 0.50$ | $9.63 \pm 0.28$ |

**Table 3** Models performance comparison based on metrics MAE, RMSE and $R^2$

| Model | MAE | RMSE | $R^2$ |
|---|---|---|---|
| MR.Lab | $1.08 \pm 0.01$ | $1.53 \pm 0.01$ | $0.68 \pm 0.00$ |
| ANN.Lab | $0.46 \pm 0.02$ | $0.69 \pm 0.05$ | $0.94 \pm 0.01$ |
| SVM.Lab | $0.43 \pm 0.01$ | $0.67 \pm 0.03$ | $0.94 \pm 0.01$ |
| SVM and ANN.Lab (average) | $0.41 \pm 0.01$ | $0.61 \pm 0.02$ | $0.95 \pm 0.00$ |

represent a low performance when compared with SVM.Lab or ANN.Lab.

Although ANN.Lab and SVM.Lab models present a very high performance, it was observed that $q_u$ prediction accuracy can be improved by averaging ANN.Lab and SVM.Lab predictions. With this trick, an $R^2 = 0.95$ is achieved as well as an RMSE very close to 0.61 MPa (Table 3). Figure 3, that plots the REC curves of each model, illustrates this slightly better performance in $q_u$ prediction by averaging ANN.Lab and SVM.Lab prediction. Moreover, Fig. 3 also underlines the huge difference between MR.Lab and ANN.Lab or SVM.Lab performance in $q_u$ prediction of laboratory soil cement mixtures.

Figure 4 illustrates clearly the very promising performance of both ANN.Lab and SVM.Lab models in $q_u$ prediction by plotting the relation between observed and predicted values. As shown, all points are very close to the diagonal line, which represents a perfect model. Figure 5 plots the same representation but considering the average of ANN.Lab and SVM.Lab predictions, illustrating once again the very high performance achieved.

From an engineering point of view, in addition to the model accuracy it is also important to understand what have been learned by it, particularly when dealing with ANN and SVM algorithms that are mathematically very complex. With this in mind, we have run a GSA [10] methodology over the models in order to measure the influence of each model attribute in $q_u$ prediction. Figure 6 plots the relative importance of each input variable, showing that W/C is the most relevant variable in $q_u$ prediction according to both ANN.Lab and SVM.Lab models, with a relative importance higher than 20%. The three next key variables are, according to SVM.Lab model, $a_w$, %OM and t. Based on ANN.Lab, the ranking is slightly different, being $\omega_0$, %Silt and %Sand the next three most influent variables after W/C. Comparing both ANN.Lab and SVM.Lab models, the last one seems to be more realistic. In fact, among the four most relevant variables, SVM.Lab model includes the influence of the water and cement contents (W/C and $a_w$), soil organic matter content (%OM) and age of the mixture (t), which are known as preponderant in soil cement mixtures behaviour [4, 16, 17, 22]. According to ANN.Lab model, the effect of the cement content is less representative (only present on W/C) and the effect of the cure time only takes the sixth position in the ranking (less than 10%). As well known, the age of the mixture is one of the most influent variables in soil cement mixtures behaviour. Thus, considering models accuracy as well as the relative importance of each variable, SVM.Lab seems to be a better choice to estimate $q_u$ development over time of laboratory soil cement mixtures. Concerning to MR.Lab, beyond its lower performance, the high influence of the soil properties (close to 80%) and the lower effect of the curing time and cement content (around 4%) is not rational.
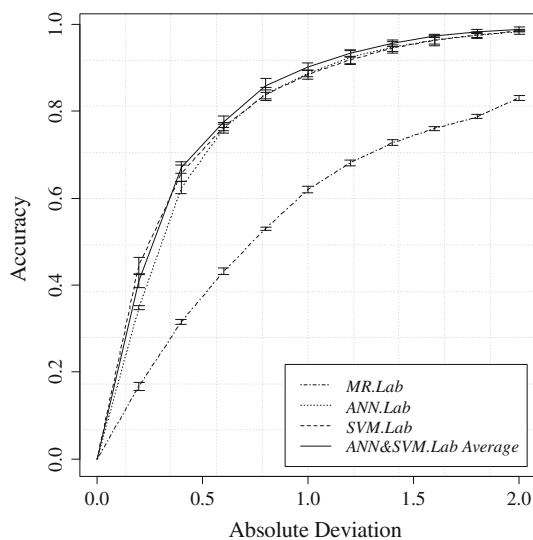


**Fig. 3** Comparison of SVM, ANN and MR models performance in $q_u$ prediction of laboratory soil cement mixtures based on REC curves
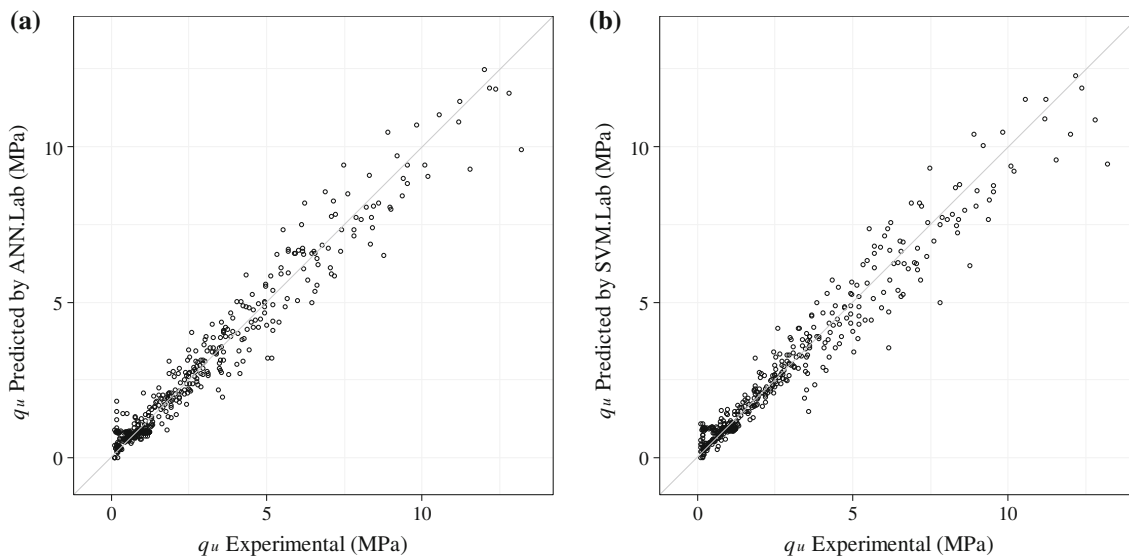
**(a)**



**(b)**

**Fig. 4** Relationship between $q_u$ experimental versus predicted values according to: **a** ANN.Lab model; **b** SVM.Lab model
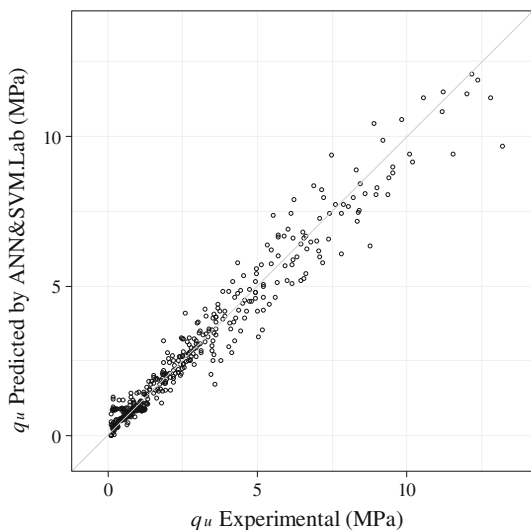


**Fig. 5** Scatterplot of the average of SVM.Lab and ANN.Lab predictions
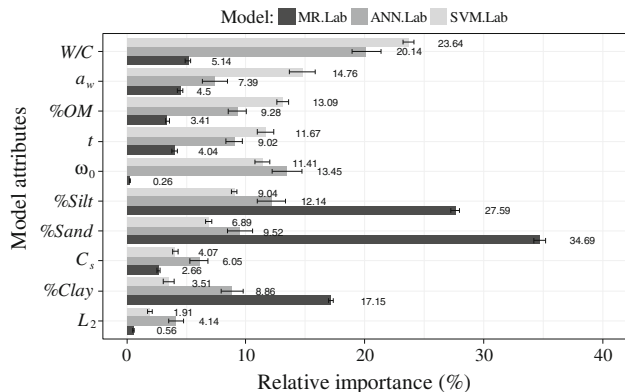


**Fig. 6** Comparison of the relative importance of each input variable based on a GSA

## 4 Conclusions

A data-driven approach is proposed for uniaxial compressive strength ($q_u$) prediction of laboratory soil cement mixtures. The proposed models, supported on a representative database comprising 444 records, are able to predict $q_u$ over time with a very promising accuracy ($R^2 = 0.95$) and only by taken as model inputs information that is available during the project stage, such as soil properties, binder and water content, etc. This means that the project design can calculate the expected $q_u$ for different scenarios (formulations) taken into account the available material without the need to prepare/test any sample. As a result, a better optimization of the available resources can be done and consequently important economic benefits can be achieved.

Through the application of a global sensitive analysis (GSA), it was possible to identify the most influent variables in $q_u$ prediction over time. This GSA allowed a better understanding of the proposed models that due to its natures are mathematically very complex and allows to conclude that the water/cement ratio ($W/C$) is the most relevant variable followed by cement content, soil organic matter content and age of the mixture.

As a final observation, it should be stressed out the important contribution of data mining techniques to solve and better understanding of complex problems, namely support vector machines and artificial neural networks algorithms.

# References

1. Bi J, Bennett K (2003) Regression error characteristic curves. In: Proceedings of the twentieth international conference on machine learning. AAAI Press, Washington, DC, USA, pp 43–50

2. Chen H, Wang Q (2006) The behaviour of organic matter in the process of soft soil stabilization using cement. Bull Eng Geol Environ 65(4):445–448

3. Cherkassky V, Ma Y (2004) Practical selection of SVM parameters and noise estimation for SVM regression. Neural Netw 17(1):113–126

4. Consoli NC, Rosa DA, Cruz RC, Dalla Rosa A (2011) Water content, porosity and cement content as parameters controlling strength of artificially cemented silty soil. Eng Geol 122(3–4):328–333

5. Correia A (2011) Applicability of deep mixing technique to the soft soil of Baixo Mondego. Ph.D. thesis, School of Engineering, University of Coimbra, Coimbra, Portugal

6. Correia AG, Tinoco J, Cortez P (2014) Use of data mining in design of soil improvement by jet grouting. In: Toll DG (ed) Second international conference on information technology in geo-engineering (ICITG 2014). IOS Press, Durham, pp 43–63

7. Correia AA, Oliveira PJV, Custódio DG (2015) Effect of polypropylene fibres on the compressive and tensile strength of a soft soil, artificially stabilised with binders. Geotext Geomembr 43(2):97–106

8. Cortes C, Vapnik V (1995) Support vector networks. Mach Learn 20(3):273–297

9. Cortez P (2010) Data mining with neural networks and support vector machines using the R/rminer tool. In: Perner P (ed) Advances in data mining: applications and theoretical aspects, 10th industrial conference on data mining. LNAI 6171. Springer, Berlin, pp 572–583

10. Cortez P, Embrechts M (2013) Using sensitivity analysis and visualization techniques to open black box data mining models. Inf Sci 225:1–17

11. Deepa N, Ganesan K (2016) Multi-class classification using hybrid soft decision model for agriculture crop selection. Neural Comput Appl 30(4):1025–1038

12. Domingos P (2012) A few useful things to know about machine learning. Commun ACM 55(10):78–87

13. Emamgholizadeh S, Bahman K, Bateni SM, Ghorbani H, Marofpoor I, Nielson JR (2017) Estimation of soil dispersivity using soft computing approaches. Neural Comput Appl 28(1):207–216

14. Gilan S, Bahrami Jovein H, Ramezanianpour A (2012) Hybrid support vector regression–particle swarm optimization for prediction of compressive strength and RCPT of concretes containing metakaolin. Constr Build Mater 34:321–329

15. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, New York

16. Horpibulsk S, Rachan R, Suddeepong A, Chinkulkijniwat A (2011) Strength development in cement admixed bangkok clay: laboratory and field investigations. Soils Found 51(2):239–251

17. Horpibulsuk S, Rachan R, Suddeepong A (2011) Assessment of strength development in blended cement admixed Bangkok clay. Constr Build Mater 25(4):1521–1531

18. Inbarani HH, Kumar SU, Azar AT, Hassanien AE (2018) Hybrid rough-bijective soft set classification system. Neural Comput Appl 29(8):67–78

19. Kenig S, Ben-David A, Omer M, Sadeh A (2001) Control of properties in injection molding by neural networks. Eng Appl Artif Intell 14(6):819–823

20. Lee FH, Lee Y, Chew SH, Yong KY (2005) Strength and modulus of marine clay–cement mixes. J Geotech Geoenviron Eng 131(2):178–186

21. Liao S, Chu P, Hsiao P (2012) Data mining techniques and applications. A decade review from 2000 to 2011. Expert Syst Appl 39(12):11303–11311

22. Lorenzo G, Bergado D (2004) Fundamental parameters of cement-admixed clay—new approach. J Geotech Geoenviron Eng 130(10):1042–1050

23. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133

24. R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

25. Sariosseiri F, Muhunthan B (2009) Effect of cement treatment on geotechnical properties of some Washington state soils. Eng Geol 104(1–2):119–125

26. Smola A, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14(3):199–222

27. Tinoco J, Gomes Correia A, Cortez P (2011) Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time. Constr Build Mater 25(3):1257–1262

28. Tinoco J, Correia AG, Cortez P (2014) A novel approach to predicting Young's modulus of jet grouting laboratory formulations over time using data mining techniques. Eng Geol 169:50–60. https://doi.org/10.1016/j.enggeo.2013.11.015

29. Tinoco J, Alberto A, da Venda P, Correia AG, Lemos L (2016) A data-driven approach for qu prediction of laboratory soil-cement mixtures. Procedia Eng 143:566–573. https://doi.org/10.1016/j.proeng.2016.06.073

30. Venda Oliveira PJ, Correia AA, Garcia MR (2012) Effect of organic matter content and curing conditions on the creep behavior of an artificially stabilized soil. J Mater Civ Eng 24(7):868–875

31. Venda Oliveira PJ, Correia AA, Garcia MR (2013) Effect of stress level and binder composition on secondary compression of an artificially stabilized soil. J Geotech Geoenviron Eng 139(5):810–820

32. Venda Oliveira PJ, Correia AA, Lopes TJ (2014) Effect of organic matter content and binder quantity on the uniaxial creep behavior of an artificially stabilized soil. J Geotech Geoenviron Eng 140(9):04014053

33. Werbos P (1974) Regression: new tools for prediction and analysis in the behavioral sciences. Ph.D. thesis, Harvard University, Harvard, USA