Computers and Concrete, *Vol. 19, No. 3 (2017) 233-241*
DOI: https://doi.org/10.12989/cac.2017.19.3.233

Compressive strength prediction of CFRP confined concrete using data mining techniques

Aires Camões[*1] and Francisco F. Martins[2a]

[1]CTAC, Department of Civil Engineering, University of Minho, Guimarães, Portugal
[2]ISISE, Department of Civil Engineering, University of Minho, Guimarães, Portugal

(Received July 8, 2016, Revised December 3, 2016, Accepted December 15, 2016)

Abstract. During the last two decades, CFRP have been extensively used for repair and rehabilitation of existing structures as well as in new construction applications. For rehabilitation purposes CFRP are currently used to increase the load and the energy absorption capacities and also the shear strength of concrete columns. Thus, the effect of CFRP confinement on the strength and deformation capacity of concrete columns has been extensively studied. However, the majority of such studies consider empirical relationships based on correlation analysis due to the fact that until today there is no general law describing such a hugely complex phenomenon. Moreover, these studies have been focused on the performance of circular cross section columns and the data available for square or rectangular cross sections are still scarce. Therefore, the existing relationships may not be sufficiently accurate to provide satisfactory results. That is why intelligent models with the ability to learn from examples can and must be tested, trying to evaluate their accuracy for composite compressive strength prediction. In this study the forecasting of wrapped CFRP confined concrete strength was carried out using different Data Mining techniques to predict CFRP confined concrete compressive strength taking into account the specimens' cross section: circular or rectangular.

Based on the results obtained, CFRP confined concrete compressive strength can be accurately predicted for circular cross sections using SVM with five and six input parameters without spending too much time. The results for rectangular sections were not as good as those obtained for circular sections. It seems that the prediction can only be obtained with reasonable accuracy for certain values of the lateral confinement coefficient due to less efficiency of lateral confinement for rectangular cross sections.

Keywords: CFRP confined concrete; data mining; artificial neural networks; support vector machines

1. Introduction

During the last two decades, fiber reinforced polymers (FRP) have been accepted by civil engineers and used in different construction applications such as the repair and rehabilitation of existing structures as well as in new construction applications.

The characteristics of FRP materials, including highly specific strength and stiffness, low thickness and weight, and high resistance to corrosion, are favorable properties justifying the increased use of these composites in structural rehabilitation and strengthening, namely to increase load carrying and energy absorption capacities.

Many types of FRP composites are available for external strengthening and repair such as: glass fiber reinforced polymers, carbon fiber reinforced polymers (CFRP), and aramid fiber reinforced polymers.

The strengthening of concrete columns can be executed by a FRP jacket, which provides lateral confinement to the column. FRP confinement can be of different types, such as spiral, wrapped and tube. The FRP wrapping technique is widely used to increase load carrying capacity, ductility and the shear strength of concrete columns (Berthet *et al.* 2005, Harajli *et al.* 2006, Mirmiran and Shahawy 1997, Pessiki *et al.* 2001, Toutanji 1999, Toutanji and Deng 2001).

In this application, the FRP sheets are generally wrapped around the columns with fibers mainly oriented in the circumferential (hoop) direction. When a FRP confined concrete column is axially loaded in compression, the concrete tends to expand laterally and this expansion is restrained by the FRP. Therefore, the fibers confine the concrete and increase its compressive strength by creating a triaxial stress state. The FRP wraps also increase the shear resistance of columns and prevent premature failures when columns are subjected to lateral loadings typical of those observed during earthquakes (Green *et al.* 2006).

One of the successful and most popular structural applications of FRP composites is the external strengthening, repair and ductility enhancement of reinforced concrete columns in both seismic and corrosive environments using CFRP (Hollaway 2004, Leung *et al.* 2006).

In this context, the effect of FRP, namely CFRP, confinement on the strength and deformation capacity of concrete columns has been extensively studied and is now sufficiently understood and well documented. Therefore, several empirical and theoretical models have been

∗Corresponding author, Professor
E-mail: aires@civil.uminho.pt
[a]Professor
E-mail: ffm@civil.uminho.pt

http://www.techno-press.org/?journal=cac&subpage=8 ISSN: 1598-8198 (Print), 1598-818X (Online)

developed and proposed by different authors (Deniaud and Neale 2006, Lam *et al.* 2006, Lee and Hegemier 2009, Matthys *et al.* 2005, Teng *et al.* 2007).

However, the majority of such studies have focused on the performance of circular cross section columns. The data available for columns of square or rectangular cross sections have increased over recent years but are still limited (Benzaid and Mesbah 2013). Moreover, these relationships are usually empirical ones or based on correlation analysis, as so far there is no general law describing the phenomena and able to explain such a hugely complex system. In this context, the traditional methods based on generalization of previous experience may not be sufficiently accurate to provide satisfactory relationships. That is why intelligent models with the ability to learn from examples have been successfully applied in the prediction of concrete compressive strength (Gupta 2007, Kim and Kim 2002, Lai and Serra 1997, Martins and Camões 2013, Saridemir 2009 and Topçu and Saridemir 2008) and also in the prediction of CFRP confined concrete compressive strength (Cevik 2011, Cevik and Guzelbey 2008, Cevik *et al.* 2010, Doran *et al.* 2015 and Jalal *et al.* 2013).

In this study the forecasting of wrapped CFRP confined concrete strength was carried out using Data Mining (DM) techniques taking into account the compressive strength prediction of the composite. These techniques are powerful, intelligent tools that learn from examples and experiences and were applied successfully to predict concrete strength by other authors (Gupta 2007, Kim and Kim 2002, Lai and Serra 1997, Saridemir 2009 and Topçu and Saridemir 2008).

This research compared the predictive capacity of several DM techniques to forecast CFRP confined concrete compressive strength taking into account the specimens' cross section: circular or rectangular.

This paper begins by presenting a brief description of CFRP confined concrete and its compressive strength prediction. This is followed by a definition of the data mining techniques and their application in predicting CFRP confined concrete behavior. How to evaluate the different algorithms of DM is also explained. Finally the results, discussion and conclusions are presented.

2. Compressive strength prediction of CFRP confined concrete

The use of CFRP in columns to increase deformation capacities and compressive strengths is a relatively new technique and has been extensively used for more than two decades. In this context, many experimental and analytical investigations have been conducted in recent decades to evaluate the axial load capacity and stress-strain behavior of concrete confined with CFRP polymers (Doran *et al.* 2015).

Reinforced concrete columns need to be laterally confined in order to ensure major deformation under applied loads before failure and to provide an adequate flexural behavior. In the case of a seismic event, energy dissipation allowed by a well-confined concrete column can often save lives. On the contrary, a poorly confined concrete column behaves in a brittle manner leading to sudden and catastrophic failures (Dong *et al.* 2015 and Kumutha *et al.* 2007).

When a CFRP-wrapped concrete core is subjected to an axial compressive load, the concrete core expands laterally. This expansion is restricted by the CFRP wrap, and therefore the concrete core is changed from an uniaxial to a three-dimensional compressive stress state. In this state, the concrete core performance is significantly influenced by the confinement pressure (Parvin and Jamwal 2005). Several parameters influence the confinement effectiveness of the CFRP wrap, which include concrete strength, wrap thickness or the number of CFRP layers, and wrap angle orientation (Sadeghian *et al.* 2008).

Although many investigations have been carried out on the behavior of CFRP-wrapped concrete, the effects of concrete strength, CFRP confining stiffness and CFRP rupture strain on the strength and ductility of CFRP confined concrete are quite complex and up to now, there are still no generally accepted guidelines for the design of CFRP confinement.

2.1 Modelling FRP-confined concrete

According to Hollaway (2004), Lee and Hegemier (2009) and Teng *et al.* (2007), various models for confinement of concrete with FRP have been developed. Most of the existing strength models for FRP-confined concrete adopted the concept of Richart *et al.* (1929), in which the compressive strength for confined concrete takes the following form

$$f_{cc}^{'} = f_{c0}^{'} + k_1 \ f_l \tag{1}$$

Or

$$K_s = \frac{f_{cc}^{'}}{f_{c0}^{'}} = 1 + \frac{k_1 \ f_l}{f_{c0}^{'}} \tag{2}$$

Where $f_{cc}^{'}$ and $f_{c0}^{'}$ are the compressive strength of confined and unconfined concrete respectively, f_l is the lateral confinement pressure, k_1 is the confinement effectiveness coefficient and K_s could be defined as the lateral confinement coefficient.

Eq. (1) assumes a linear relation between $f_{cc}^{'}$ and f_l. However, other nonlinear empirical relations can be found in the literature (Jalal *et al*. 2013).

3. Data mining

3.1 Definition and applications to CFRP confined concrete

The Data Mining (DM) process is one step of so-called Knowledge Discovery in Databases (KDD). According to Fayyad *et al.* (1996), KDD begins with the selection of the data base, which constitutes the support to proceed with the study. After the selection, the target data is pre-processed and DM is applied. The KDD process ends with presentation of the knowledge discovery after interpretation

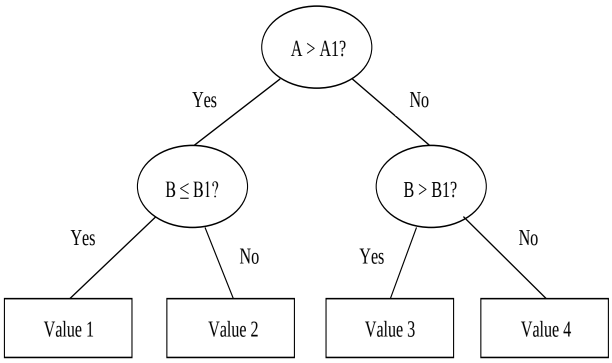


Fig. 1 Example of a regression tree

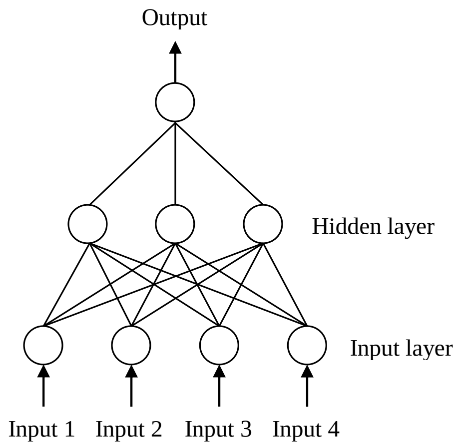


Fig. 2 Example of a multilayer perceptron

of the results. DM can be applied to classification and regression tasks. The regression task consists of mapping several input variables to a numeric output. In the DM process it is usual to divide the dataset into two subsets. One, called the training set, is used in a learning process of the algorithms, and the other, called the testing set, is used to test the algorithms. During the learning process the various parameters of the algorithms are adjusted to optimize the results. The accuracy of the algorithms is assessed through metrics based on errors and the correlation coefficient. The validated algorithms are used as models to predict the value of the output variables.

Many authors have been successful in applying intelligent tools. Cevik and Guzelbey (2008) applied neural networks to model the strength enhancement of CFRP confined concrete cylinders. Cevik and Cabalar (2008), proposed a genetic programming approach to the formulation of strength enhancement of FRP confined concrete cylinders. Gandomi *et al.* (2010) presented a new approach to the formulation of compressive strength of CFRP confined concrete cylinders using linear genetic programming. Cevik (2011) applied genetic programming and step-wise regression, neuro fuzzy and neural network to model strength enhancement of FRP confined concrete cylinders. Jalal *et al.* (2013) carried out modelling of strength enhancement of concrete cylinders with retrofitted CFRP composites using the adaptive neuro fuzzy inference system (ANFIS) and genetic programming. Doran *et al.* (2015), implemented a new artificial intelligence-based algorithm to model the strength enhancement of CFRP confined reinforced concrete columns using fuzzy logic methodology.

3.2 DM algorithms

There are several DM algorithms such as Regression Trees (RT), Multiple Regressions (MR), Artificial Neural Networks (ANN), Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN). These algorithms were already explained by the authors in previous papers (Martins and Camões 2013, Martins and Miranda 2012). Therefore, a brief description of them will be provided below. Further details can be found in many publications. Breiman *et al.* (1984), Berk (2008), Coimbra *et al.* (2014) and Czajkowski and Kretowski (2016) for RT; Aleksander and Morton (1990), Ilonen *et al.* (2003), Downing (2015) and Souza and Soares (2016) for ANN; Vapnik (1998), Cristianini and Shawe-Taylor (2000), Dibiki *et al.* (2001), Ben-Hur and Weston (2010) and Liang *et al.* (2011) for SVM; Cover (1968), Cover and Hart (1967), Nguyen *et al.* (2016) and Yang *et al.* (2016) for k-NN.

The Decision Trees (Quinlan 1986) have an inverted tree structure composed of nodes and descendent branches. The result of a test performed at each node indicates the branch to continue the process. This process is repeated until the final decision can be made and a class is attributed to the register. Regression trees are a particular case of decision trees where classes are replaced by values (Fig. 1).

Multiple regressions are similar to simple regressions but with several independent variables instead of one independent variable.

ANN is a technique that attempts to mimic the human brain. The artificial neurons communicate with each other sending signals through liaisons that define the ANN architecture. A weight, w_i, j (i and j are neurons or nodes) is associated with each link and each neuron has an activation function that introduces a non-linear component (Cortez 2010 and Haykin 1999). This study used a logistic activation function f given by 1/(1+*ex*) and the following general equation

$$\hat{y} = w_{o,0} + \sum_{j=I+1}^{o-1} f\left(\sum_{i=1}^{I} x_i w_{j,i} + w_{j,0}\right) w_{o,i} \tag{3}$$

Where x_i are the input parameters or nodes, I is the number of input parameters and o is the output parameter.

This study adopted the multilayer perceptron (feed forward network) architecture (Haykin 1999) with one hidden layer that contains HN processing units (Fig. 2). Since the network's performance is sensitive to HN, a grid search of (0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20) was adopted during the learning phase to find the best HN value. The neural network can learn their weights and bias using the gradient descent algorithm, known as backpropagation, presented by Nielsen (2016).

The SVM technique was initially developed for classification problems by Cortes and Vapnik (1995). This method uses nonlinear mapping to transform the input data into a multidimensional feature space (Fig. 3). After this

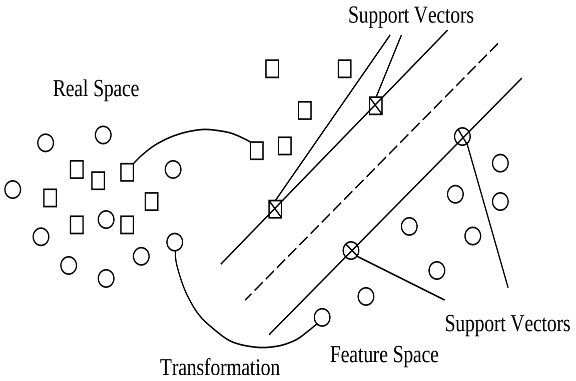


Fig. 3 Example of a multilayer perceptron

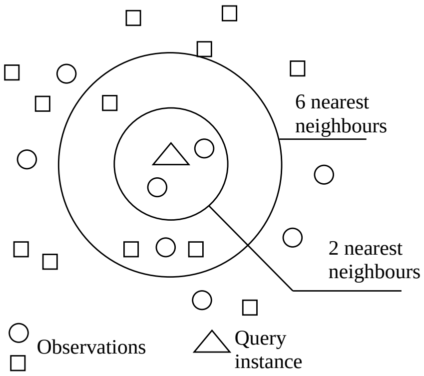


Fig. 4 Example of k-nearest neighbors

transformation the SVM finds the best hyperplane of linear separation within the feature space. The nonlinear mapping depends on a kernel function $k(x, x')$. This work uses the Gaussian kernel function which is the most popular one

$$k(x,x') = e^{\left(-\gamma \cdot \|x-x'\|^2\right)}, \quad \gamma > 0 \tag{4}$$

After introduction of the ε-insensitive loss function it was possible to apply SVM to regression problems (Smola and Scholkopf 2004). Both the width of the ε-insensitive zone and the kernel parameter, γ, affected the performance of the regression. In addition to these two parameters there is a penalty parameter C, which also affects the performance of the regression. In order to limit searching space, C and ε were set using heuristics proposed by Cherkassy and Ma (2004): C=3 and $\varepsilon = \hat{\sigma}/\sqrt{N}$, where $\hat{\sigma} = 1.5 \times \sum_{i=1}^{N} (y_i - \hat{y}_i)$, $\hat{y}_i$ is the value predicted by a 3-nearest neighbor algorithm and N the number of examples. Under this setup, the search space was limited to the input values of γ.

The k-Nearest Neighbor (Hechenbichler and Schliep 2004) is a simple supervised learning algorithm that can be used in classification and regression problems. In classification problems an instance query is classified according to its neighbors' classes (Fig. 4). The dominant class among the nearest neighbors is attributed to the query instance. In regression problems the property value for the instance query is obtained as the average of the weighted values of the k nearest neighbors. This implies calculation of the distance between the target and its neighbors in the multidimensional space. Generally, weights are attributed according to distance. The closest neighbors are given more weight than more distant ones.

In this work the whole dataset was divided in two subsets: the training set with two thirds of the whole dataset and the testing set with the remaining data.

Using the training dataset the parameters involved in the different techniques (H, γ and k) were optimized through a grid search of H(0, 2, 4, 6, ..., 20), $\gamma(2^{-15}, 2^{-13}, ..., 2^{3})$ and k(2, 3, 4, ..., 12). The predictive performance of the models was evaluated by adopting a 10-fold cross-validation (Effron 1993) corresponding to the division of the training dataset, formed of two thirds of the whole dataset, in 10 equal subsets. One subset is tested each time and the remaining data are used to fit the model. This means that all training data are used for training and testing. The process is repeated sequentially until all subsets have been tested. After selection of the best parameters, the model is retrained with all training data. Finally, the future performance of the model is verified using the testing dataset (one third of the whole dataset). It must be stressed that this testing dataset was not used to fit the model.

There are several metrics to assess the performance of the regression models. This study uses the mean absolute deviation (MAD), the root mean squared error (RMSE) and the coefficient of correlation (R). The three metrics are defined as

$$MAD = \frac{1}{N} \times \sum_{i=1}^{N} \left| y_i - \hat{y}_i \right| \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \tag{6}$$

$$R = \frac{\sum_{i=1}^{N} (y_i - \bar{y}) \times (\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N} (y_i - \bar{y})^2} \times \sqrt{\sum_{i=1}^{N} (\hat{y}_i - \bar{\hat{y}})^2}} \tag{7}$$

Where N denotes the number of examples, y_i the real value, $\hat{y}_i$ the value estimated by the model, $\bar{y}$ the mean of the real values and $\bar{\hat{y}}$ the mean of the estimated values.

The results obtained should be interpreted with great care, and the relative importance of the input parameters in the model can be very useful to support this interpretation. Hence the need to perform a sensitivity analysis (Kewley *et al.* 2000) aiming to evaluate the model's response to changes in the input variables. Thus, each input parameter is ranged from its lowest value to its highest value while the remaining input parameters keep their mean values. For a given input parameter, the higher the variance induced by it in the model output, the greater its importance.

4. Data used in data mining

As already mentioned, this study includes two analyses corresponding to circular and rectangular cross sections. This required two databases. The database for circular cross

Table 1 Statistics of the input and the output parameters for circular sections

Parameters		Min.	Mean	Max.	Standard Deviation	Coefficient of variation (%)
Inputs	d(mm)	51	131.20	200	34.99	26.67
	h(mm)	102	294.81	610	120.87	41.00
	t(mm)	0.089	0.416	2.0	0.368	88.62
	E_{CFRP}(MPa)	19900	611600	211221.6	108110.5	51.18
	ε_{rup}(mm)	0.0017	0.0091	0.0207	0.0033	36.11
	f'_{c0}(MPa)	17.39	39.69	171	25.61	64.53
Output	K_s	1.05	2.15	5.23	0.91	42.37

Table 2 Statistics of the input and the output parameters for rectangular sections

Parameters		Min.	Mean	Max.	Standard Deviation	Coefficient of variation (%)
Inputs	b(mm)	20	181.14	457	73.26	40.44
	l(mm)	108	185.19	457	46.98	25.37
	t(mm)	0.056	0.528	3.0	0.413	78.15
	E_{CFRP}(MPa)	10500	193886.5	439000	77994.87	40.23
	f'_{c0}(MPa)	10.83	26.89	55.36	11.60	43.14
Output	K_s	0.94	1.77	4.79	0.72	40.75

sections was taken from Jalal *et al.* (2013). This database was built from several existing tests on CFRP-confined concrete cylinders corresponding to twenty published papers. It is composed of 128 records, each record having the following parameters: diameter (d) and height (h) of the cylindrical specimen, total thickness of CFRP layer used (t), elastic modulus of CFRP (E_{CFRP}), ultimate circumferential strain in the CFRP jacket (ε_{rup}), unconfined concrete strength (f'_{c0}) and confined compressive strength (f'_{cc}). Nevertheless, in this study f'_{cc} was substituted by K_s which was calculated using the Eq. (2). The database for rectangular sections was taken from Doren *et al.* (2015). This database was collected from fifteen references and is composed of 163 records, each record having the following parameters: column width (b), column length (l), radius of the corner (r), total thickness of CFRP (t), elastic modulus of CFRP (E_{CFRP}), unconfined concrete strength (f'_{cc}) and lateral confinement coefficient (K_s).

Tables 1 and 2 present some statistical data of the parameters used in the analyses. In relation to rectangular cross sections the coefficients of variation of four parameters (b, E_{CFRP}, f'_{c0} and K_s) are quite similar, which means they have analogous variability. The variability is more pronounced among the parameters corresponding to circular cross sections. Total thickness of CFRP (t) for both circular and rectangular cross sections presented the highest variability. It must be stressed that K_s presented similar variability for both circular and rectangular cross sections.

5. Results and discussion

Table 3 Mean values of the metrics obtained in the training phase for circular sections using six input parameters

	RT	MR	ANN	SVM	k-NN
MAD	0.574	0.687	0.509	0.359	0.442
RMSE	0.767	0.883	0.797	0.518	0.618
R	0.603	0.414	0.729	0.839	0.760

Table 4 Mean values of the metrics obtained in the training phase for circular sections using five input parameters

	RT	MR	ANN	SVM	k-NN
MAD	0.578	0.690	0.469	0.355	0.394
RMSE	0.761	0.882	0.691	0.501	0.549
R	0.611	0.413	0.757	0.850	0.814

Table 5 Mean values of the metrics obtained in the training phase for circular sections using four input parameters

	RT	MR	ANN	SVM	k-NN
MAD	0.551	0.697	0.407	0.366	0.386
RMSE	0.757	0.873	0.660	0.506	0.533
R	0.615	0.414	0.806	0.845	0.825

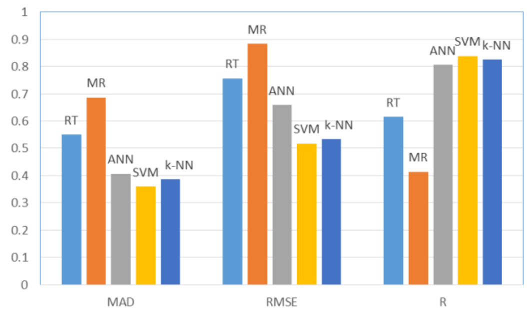


Fig. 5 Metrics to evaluate the performance of the models for circular cross sections

With the databases built, the predictive models were trained to forecast the lateral confinement coefficient (K_s).

5.1 Circular cross sections

For this geometry three analyses were performed using a different number of input parameters: six parameters (d, h, t, E_{CFRP}, ε_{rup} and f'_{c0}), five parameters (d, t, E_{CFRP}, ε_{rup} and f'_{c0}) and four parameters (d, t, E_{CFRP} and f'_{c0}).

Tables 3, 4 and 5 show the errors and the coefficient of correlation obtained in the training phase. Analysis of these tables reveals that the best RT, ANN and k-NN models are obtained with four parameters while the best MR and SVM models are obtained with six and five parameters, respectively. To allow a visual comparison of the different methods used, with the metrics close to each other, the bar graph in Fig. 5 was plotted. Fig. 5 is composed of the weakest SVM model and by the best models of the remaining techniques (RT, MR, ANN and k-NN). The use of the weakest SVM model is justified because, as can be seen in Tables 3 to 5, all SVM models are better than all the

Table 6 Importance of the input variables in the evaluation of K_s (%) for circular sections using six input parameters

	RT	MR	ANN	SVM	k-NN
d	21.68	12.00	3.58	11.87	8.11
h	0	2.60	6.01	4.36	10.95
t	1.68	7.38	52.90	24.48	6.98
E_{FRP}	23.95	15.92	13.94	11.02	32.01
f'_{c0}	52.69	52.44	15.41	37.98	24.14
ε_{rup}	0	9.66	8.16	10.29	17.81

Table 7 Importance of the input variables in the evaluation of K_s (%) for circular sections using five input parameters

	RT	MR	ANN	SVM	k-NN
d	21.68	20.45	5.75	21.17	7.99
t	1.68	4.68	22.41	16.79	10.45
E_{FRP}	23.95	15.70	49.46	8.96	25.48
f'_{c0}	52.69	47.18	14.52	45.66	31.62
ε_{rup}	0	11.99	7.86	7.42	24.46

Table 8 Importance of the input variables in the evaluation of K_s (%) for circular sections using four input parameters

	RT	MR	ANN	SVM	k-NN
d	18.51	23.20	12.92	6.45	5.30
t	1.44	6.19	26.43	37.61	20.68
f'_{c0}	35.04	5.05	26.40	10.48	23.46
ε_{rup}	45.01	65.56	34.25	45.46	50.56

models obtained with the other techniques. A visual analysis of Fig. 5 confirms that the SVM model has the smallest errors and the highest coefficient of correlation, whereas the MR model has the highest errors and the lowest coefficient of correlation.

The importance of the input variables in the models is presented in Tables 6, 7 and 8. As explained earlier, the models are based on different algorithms. Therefore, the importance attributed to the input parameters differs from model to model.

When only four parameters are used, all the models attributed the greatest importance to f'_{c0}. The RT, MR and SVM models maintained f'_{c0} as the most important parameter when using five and six input parameters. When the six input parameters are used all the models attributed low importance to h.

The difference in performance verified in the training phase of the three SVM models is small. Therefore, at first sight, the three SVM models could be used to predict Ks. However, in the model with 4 parameters f'_{c0} presents about four times less importance than ε_{rup} and t. As this relationship does not seem reasonable, the SVM model with four parameters should be discarded.

The accuracy of the SVM models with six and five input parameters can be verified in Figs. 6-7, which present the relationships between the measured and calculated values of K_s, using the training and testing data. It is recalled that the testing data were not used in generating the model. These figures confirm the good predictive capacity of both models. However, to perform a mathematical analysis of these figures it is necessary to calculate the associated errors and the coefficient of correlation. These metrics are presented in Table 9. In relation to the metrics obtained in the training phase (see Tables 3 to 5) it can be seen that the errors are lower and the coefficients of correlation are higher. Table 9 shows that the performance of the SVM model is better with six parameters using the testing dataset.

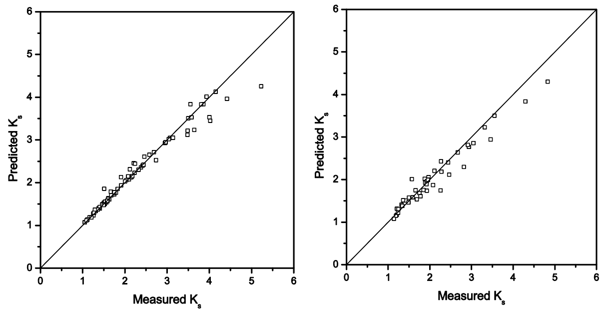


Fig. 6 Predicted versus measured K_s using SVM model with six input parameters for circular cross sections obtained with: (a) training data set, (b) testing data set

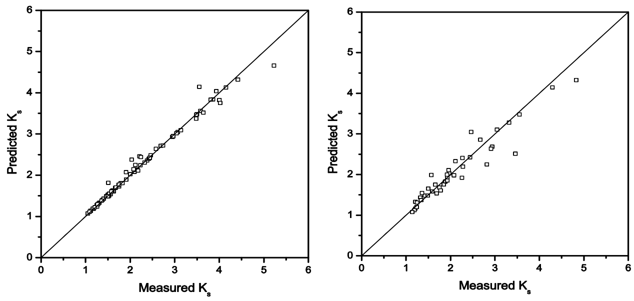


Fig. 7 Predicted versus measured K_s using SVM model with five input parameters for circular cross sections obtained with: (a) training data set, (b) testing data set

Table 9 Metrics corresponding to SVM models of Figs. 6 and 7

Parameters		MAD	RMSE	R
5	training	0.060	0.120	0.992
	testing	0.162	0.247	0.958
6	training	0.086	0.176	0.986
	testing	0.151	0.216	0.977

5.2 Rectangular cross sections

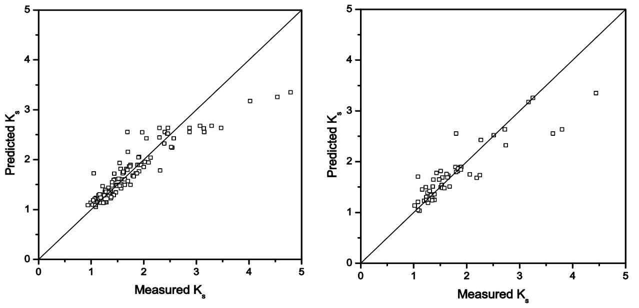
Table 10 shows the errors and the coefficient of correlation obtained in the training phase. The SVM model has the best forecasting capacity and the RT model the poorest forecasting capacity. In relation to the importance attributed to the input parameters by the models, only RT

Table 10 Mean values of the metrics obtained in the training phase for rectangular sections

	RT	MR	ANN	SVM	k-NN
MAD	0.405	0.375	0.325	0.279	0.302
RMSE	0.586	0.541	0.684	0.445	0.498
R	0.595	0.658	0.629	0.794	0.718

Table 11 Importance of the input variables in the evaluation of K_s (%) for rectangular sections

	RT	MR	ANN	SVM	k-NN
b	0	2.05	3.73	0.96	6.18
h	9.59	9.99	15.49	3.39	5.46
t	6.43	37.90	27.25	36.71	16.85
E_{CFRP}	0	35.62	46.57	29.87	67.60
f'_{c0}	83.98	14.44	6.96	29.07	3.91



(a) (b)

Fig. 8 Predicted versus measured K_s using SVM model for rectangular cross sections obtained with: (a) training data set; (b) testing data set

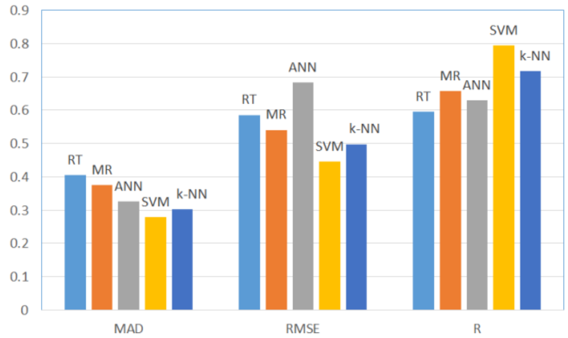


Fig. 9 Metrics to evaluate the models' performance for rectangular cross sections

attributes the greatest importance to f'_{c0} (Table 11). k-NN and ANN attribute marginal importance to f'_{c0}. SVM attributed the highest importance to t. However, it attributed almost 30% of importance to f'_{c0} and E_{CFRP}.

Fig. 8 shows the comparison between the measured and the predicted lateral confinement coefficient for rectangular cross sections and the SVM model using the training and testing set, respectively. The results are reasonable until K_s is more or less equal to 3. For higher values of K_s the measured values tend to be lower than the predicted values.

Fig. 9 allows visual comparison of the models' performance, confirming the best performance of the SVM model followed by the k-NN model. Furthermore, the RT model has the highest MAD value and the lowest R value. However, it is the ANN model that gives the highest value of RMSE, which can be explained by the high variability of these errors obtained in the cross-validation scheme.

6. Conclusions

Data mining techniques have the capacity to learn from examples. In this study several data mining techniques were used to predict the lateral confinement coefficient. The training phase indicates the SVM as the model with the best predictive capacity for all analyzed situations. Furthermore, the SVM model demonstrates sensitivity to parameters known to affect the lateral confinement coefficient. In the case of circular cross sections, the height (h) of the specimens has no significant influence on the models' performance. This was confirmed by the lesser importance of h and the better performance of the models when using five parameters, without h, mainly ANN, SVM and k-NN models. However, the validation of the SVM model with six input parameters using the unseen testing dataset reveals that the influence of h is enough to slightly improve the performance of the SVM model with five parameters. Moreover, based on the analyzed data, K_s can be accurately predicted for circular cross sections using SVM with five input parameters without spending too much time.

The results for rectangular sections were not as good as those obtained for circular sections. It seems that K_s can only be obtained with a reasonable accuracy for values less than 3.

Despite the good performance presented by SVM, it should be stressed that this model is only valid for conditions found in the dataset used in this study. Therefore, it is only valid for circular and rectangular sections and for CFRP confined concrete. Its application to other materials of confinement, such as steel, was not studied in this research and should be verified using an enlarged dataset with data for those materials.

Acknowledgements

This work was partly financed by FEDER funds through the Competitivity Factors Operational Programme-COMPETE and by national funds through FCT-Foundation for Science and Technology within the scope of the project POCI-01-0145-FEDER-007633.

References

Aleksander, I. and Morton, H. (1990), *An Introduction to Neural Computing*, Chapman & Hall.

Ben-Hur, A. and Weston, J. (2010), *A User's Guide to Support Vector Machines*, Humana Press, New York, U.S.A.

Benzaid, R. and Mesbah, H.A. (2013), "Circular and square concrete columns externally confined by CFRP composite: Experimental investigation and effective strength models", *InTech*, 167-201.

Berk, R.A. (2008), *Statistical Learning from a Regression Perspective*, Springer-Verlag, New York, U.S.A.

Berthet, J.F., Ferrier, E. and Hamelin, P. (2005), "Compressive behaviour of concrete externally confined by composite jackets", *Constr. Build. Mater.*, **19**(3), 223-232.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984), *Classification and Regression Trees*, Chapman & Hall/CRC.

Cevik, A. (2011), "Modeling strength enhancement of FRP confined concrete cylinders using soft computing", *Exp. Syst. Appl.* **38**(5), 5662-5673.

Cevik, A. and Cabalar, A.F. (2008), "A genetic-programming-based formulation for the strength of fiber-reinforced-polymer-confined concrete cylinders", *J. Appl. Poly. Sci.*, **110**(5), 3087-3095.

Cevik, A. and Guzelbey, I.H. (2008), "Neural network modeling of strength enhancement for CFRP confined concrete cylinders", *Build. Environ.*, **43**(5), 751-763.

Cevik, A., Gögüs, M.T., Güzelbey, I.H. and Filiz, H. (2010), "Soft computing based formulation for strength enhancement of CFRP confined concrete cylinders", *Adv. Eng. Soft.*, **41**(4), 527-536.

Cherkassy, V. and Ma, Y. (2004), "Practical selection of SVM parameters and noise estimation for SVM regression", *Neur. Net.*, **17**(1), 113-126.

Coimbra, R., Rodriguez-Galiano, V., Olóriz, F. and Chica-Olmo, M. (2014), "Regression trees for modelling geomechanical data-an application to late jurassic carbonates (ammonitico rosso)", *Comput. Geosci.*, **73**, 198-207.

Cortes, C. and Vapnik, V. (1995), "Support vector networks", *Mach. Learn.*, **20**(3), 273-297.

Cortez, P. (2010), "Data mining with neural networks and support vector machines using the r/rminer tool", *Proceedings of the 10th Industrial Conference on Data Mining, Advances in Data Mining, Applications and theoretical aspects*, Berlin, Germany.

Cover, T.M. (1968), "Estimation by the nearest neighbor rule", *IEEE Trans. Informat. Theor.*, **14**(1), 50-55.

Cover, T.M. and Hart, P.E. (1967), "Nearest neighbor pattern classification", *IEEE Trans. Informat. Theor.*, **13**(1), 21-27.

Cristianini, N. and Shawe-Taylor, J. (2000), *An Introduction to Support Vector Machine*, University Press, London, U.K.

Czajkowski, M. and Kretowski, M. (2016), "The role of decision trees representation in regression problems-an evolutionary perspective", *Appl. Soft Comput.*, **48**, 458-475.

Deniaud, C. and Neale, K.W. (2006), "An assessment of constitutive models for concrete columns confined with fiber composite sheets", *Compos. Struct.*, **73**(3), 318-330.

Dibike, Y.B., Velickov, S., Solomatine, D.P. and Abbott, M.B. (2001), "Model introduction with support vector machines; introduction and applications", *J. Comput. Civil Eng.*, **15**(3), 208-216.

Dong, C.X., Kwana, A.K.H. and Hob, J.C.M. (2015), "Effects of confining stiffness and rupture strain on performance of FRP confined concrete", *Eng. Struct.*, **97**, 1-14.

Doran, B., Yetilmezsoy, K. and Murtazaoglu, S. (2015), "Application of fuzzy logic approach in predicting the lateral confinement coefficient for RC columns wrapped with CFRP", *Eng. Struct.*, **88**, 74-91.

Downing, K.L. (2015), *Intelligence Emerging: Adaptative and Search in Evolving Neural System*, MIT Press, U.S.A.

Efron, B. and Tibshirani, R. (1993), *An Introduction to the Bootstrap*, Chapman & Hall.

Fayyad, U., Piatesky-Shapiro, G. and Smyth, P. (1996), *From Data Mining to Knowledge Discovery: An Overview*, IAAAI Press/The MIT Press, Cambridge MA, 471-493.

Gandomi, M., Alavi, A.H. and Sahab, M.G. (2010), "New formulation for compressive strength of CFRP confined concrete cylinders using linear genetic programming", *Mater. Struct.*, **43**(7), 963-983.

Green, M.F., Bisby, L.A., Fam, A.Z. and Kodur, V.K.R. (2006), "FRP confined concrete columns: Behaviour under extreme conditions", *Cement Concrete Compos.*, **28**(10), 928-937.

Gupta, S.M. (2007), "Support vector machines based modelling of concrete strength", *World Acad. Sci. Eng. Technol.*, **36**, 305-311.

Harajli, M.H., Hantouche, E. and Soudki, K. (2006), "Stress-strain model for fiber-reinforced polymer jacketed concrete columns", *ACI Struct. J.*, **105**(5), 672-682.

Haykin, S. (1999), *Neural Networks-A Comprehensive Foundation*, 2nd Edition, Prentice-Hall, New Jersey, U.S.A.

Hechenbichler, K. and Schliep, K. (2004), "Weighted k-nearest-neighbor techniques and ordinal classification", Ph.D. Dissertation, Ludwig-Maximilians University Munich, Germany.

Hollaway, L.C. (2004), *Advanced Polymer Composites for Structural Applications in Construction: ACIC*, Woodhead Publishing, U.K.

Ilonen, J., Kamarainen, J.K. and Lampinen, J. (2003), "Differential evolution training algorithm for feed-forward neural network", *Neur. Proc. Lett.*, **17**(1), 93-105.

Jalal, M., Ramezanianpour, A.A., Pouladkhan, A.R. and Tedro, P. (2013), "Application of genetic programming (GP) and ANFIS for strength enhancement modeling of CFRP-retrofitted concrete cylinders", *Neur. Comput. Appl.*, **23**(2), 455-470.

Kewley, R., Embrechts, M. and Brenemam, C. (2000), "Data strip mining for the virtual design of pharmaceuticals with neural networks", *IEEE Trans. Neur. Net.*, **11**(3), 668-679.

Kim, J.I. and Kim, D.K. (2002), "Application of neural networks for estimation of concrete strength", *KSCE J. Civil Eng.*, **6**(4), 429-438.

Kumutha, R., Vaidyanathan, R. and Palanichamy, M.S. (2007), "Behaviour of reinforced concrete rectangular columns strengthened using GFRP", *Cement Concrete Compos.*, **29**(8), 609-615.

Lai, S. and Serra, M. (1997), "Concrete strength prediction by means of neural network", *Constr. Build. Mater.*, **11**(2), 93-98.

Lam, L., Teng, J.G., Cheng, C.H. and Xiao, Y. (2006), "FRP-confined concrete under axial cyclic compression", *Cement Concrete Res.*, **28**(10), 949-958.

Lee, C. and Hegemier, G.A. (2009), "Model of FRP-confined concrete cylinders in axial compression", *J. Compos. Constr.*, **13**(5), 442-454.

Leung, C.K.Y., Ng, M.Y.M. and Luk, H.C.Y. (2006), "Empirical approach for determining ultimate FRP strain in FRP-strengthened concrete beams", *J. Compos. Constr.*, **10**(2), 125-138.

Liang, Y., Xu, Q.S., Li, H.D. and Cao, D.S. (2011), *Support Vector Machines and Their Application in Chemistry and Biotechnology*, Taylor & Francis CRC Press.

Martins, F.F. and Camões, A. (2013), "Prediction of compressive strength of concrete containing fly ash using data mining techniques", *Cement WapnoBeton*, **XVIII/LXXX**(1), 39-51.

Martins, F.F. and Miranda, T.F.S. (2012), "Estimation of the rock deformation modulus and RMR based on data mining techniques", *Geotech. Geol. Eng.*, **30**(4), 787-801.

Matthys, S., Toutanji, H., Audenaert, K. and Taerwe, L. (2005), "Axial load behavior of largescale columns confined with fiber-reinforced polymer composites", *ACI Struct. J.*, **102**(2), 258-267.

Mirmiran, A. and Shahawy, M. (1997), "Behavior of concrete columns confined by fiber composites", *J. Struct. Eng.*, **123**(5), 583-590.

Nguyen, B., Morell, C. and Baets, B.D. (2016), "Large scale distance metric learning for k-nearest neighbors regression", *Neurocomput*., **214**, 805-814.

Nielsen, M. (2016), *Neural Networks and Deep Learning*.

Parvin, A. and Jamwal, A.S. (2005), "Effects of wrap thickness and ply configuration on composite-confined concrete cylinders", *Compos*. *Struct*., **67**(4), 437-442.

Pessiki, S., Harries, K.A., Kestner, J.T., Sause, R. and Ricles, J.M. (2001), "Axial behaviour of reinforced concrete columns confined with FRP jackets", *Compos*. *Constr*., **5**(4), 237-245.

Quinlan, J. (1986), "Induction of decision trees", *Mach*. *Learn*., **1**(1), 81-106.

Richart, F.E., Brandtzaeg, A. and Brown, R.L. (1929), "The failure of plain and spirally reinforced concrete in compression", Ph.D. Dissertation, University of Illinois, Urbana, U.S.A.

Sadeghian, P., Rahai, A.R. and Ehsani, M.R. (2008), "Numerical modeling of concrete cylinders confined with CFRP composites", *J*. *Reinfor*. *Plast*. *Compos*., **27**(12), 1309-1321.

Saridemir, M. (2009), "Prediction of compressive strength of concretes containing metakaolin and silica fume by artificial neural networks", *Adv*. *Eng*. *Soft*., **40**(5), 350-355.

Smola, A. and Scholkopf, B. (2004), "A tutorial on support vector regression", *Stat*. *Comput*., **14**(3), 199-222.

Souza, A.M.F. and Soares, F.M. (2016), *Neural Network Programming with Java*, Packt Publishing Ltd, Birmingham, U.K.

Teng, J.G., Yu, T., Wong, Y.L. and Dong, S.L. (2007), "Hybrid FRP-concrete-steel tubular columns: Concept and behavior", *Constr*. *Build*. *Mater*., **21**(4), 846-854.

Topçu, I.B. and Saridemir, M. (2008), "Prediction of compressive strength of concrete containing fly ash using artificial neural networks and fuzzy logic", *Comput*. *Mater*. *Sci*., **41**(3), 305-311.

Toutanji, H.A. (1999), "Stress-strain characteristics of concrete columns externally confined with advanced fiber composites sheets", *ACI Mater*. *J*., **96**(3), 397-404.

Toutanji, H.A. and Deng, Y. (2001), "Strength and durability performance of concrete axially loaded members confined with AFRP composites sheets", *Compos*. *Part B: Eng*., **33**(4), 255-261.

Vapnik, V.N. (1998), *Statistical Learning Theory*, Wiley, New York, U.S.A.

Yang, L., Dong, L. and Bi, X. (2016), "An improved location difference of multiple distances based nearest neighbors searching algorithm", *Optik-J*. *Light Electr*. *Opt*., **127**(22), 10838-10843.

CC