THE APPLICATION OF BOOSTING ALGORITHMS IN THE PREDICTION OF BOND STRENGTH BETWEEN THIN STEEL BARS AND CONCRETE

Vanderci F. Arruda[1], Gray F. Moita[1], Eliene P. Carvalho[1], Priscila F. S. Silva[1], Marco A. A. Grossi[1]

[1] *Centro Federal de Educação Tecnológica de Minas Gerais*

Av. Amazonas, 5.253, Nova Suíça, CEP: 30.421-169, Belo Horizonte, MG, Brasil.

vanderci-engcivil@hotmail.com; gray@cefetmg.br; eliene@cefetmg.br; 201422800040@aluno.cefetmg.br; marcoaraujo.grossi@gmail.com

Abstract. The current study discusses the application of intelligent algorithms and machine learning techniques to predict the bond strength between steel and concrete. The paper focuses on three boosting algorithms employed for this prediction task. The research exploited a database derived from pull-out tests conducted on thin steel bars to assess the bond between steel and concrete. The experimental program involved the use of three different classes of concrete and two types of steel bars. The goal was to analyze the steel-concrete bond strength, which is influenced by various factors. For the computational simulations, the input variables considered in this study were the bar surface, bar diameter (ϕ), concrete compressive strength (fc), and anchorage length (Ld). The output was the pull-out strength at the steel-concrete interface. It is important to highlight that most previous studies in this field have mainly focused on bars with diameters greater than 10.0 mm, while there is limited research available to evaluate the performance of bars with diameters smaller than 10.0 mm. The paper describes the computational experiments conducted using different boosting algorithms: Adaptive Boosting (AdaBoost), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB). These machine learning-based models achieved highly accurate predictions, applying specific hyperparameters. The following metrics were used to compare the performance of the different methods: Root Mean Squared Error (RMSE), the coefficient of variation (CV), and the error. These metrics were used to evaluate the reliability of each algorithm in predicting the bond strength in the samples. The results indicate the accuracy and goodness of fit of the model's predictions. Based on them, it can be concluded that the presented model can satisfactorily predict the bond strength of samples between thin steel bars and concrete.

Keywords: Intelligent Systems, Machine Learning, Adaptive Boosting, Gradient Boosting, Extreme Gradient Boosting.

1 Introduction

In reinforced concrete structures, the joint operation of the steel-concrete combination is essential. This joint functioning prevents the relative displacement between the steel and the concrete in its surroundings Belo [1]. This behavior causes the stresses, whether compressive, tensile, shear or bending, to be supported by the materials and ensures the transfer of loads between them. According to Leohard and Mong [2] and Pinheiro et al [3], adhesion can be decomposed into three parts: bonding, friction, and mechanical adhesion. These parts can be divided into several stages: Stage I refers to chemical adhesion resulting from the micromechanical interaction of materials. In stage II, as the stresses increase, the chemical adhesion portion breaks and the first cracks appear at the ribs tips; however, the steel-concrete sliding is limited because there is a fitting of the bar in the concrete. In stage III, when the stresses increase, the cracks disperse radially, but are still contained by the effect of the wedges of the bar protrusions with the concrete. Finally, in stage IV, concrete cracking or steel bar pullout occurs.

Usually, bonding strength is obtained by pull-out and beam tests, as established in E. N. BS [4], involving expensive and lengthy experimental tests. As shown in Figure 1, the purpose of the pull-out test is to pull out a steel bar positioned in the center of a concrete cube. This test can be performed on steel bars of any diameter;

however, the literature presents, mainly, studies regarding thick bars (diameter greater than 10mm), such as that of Almeida et al [5].

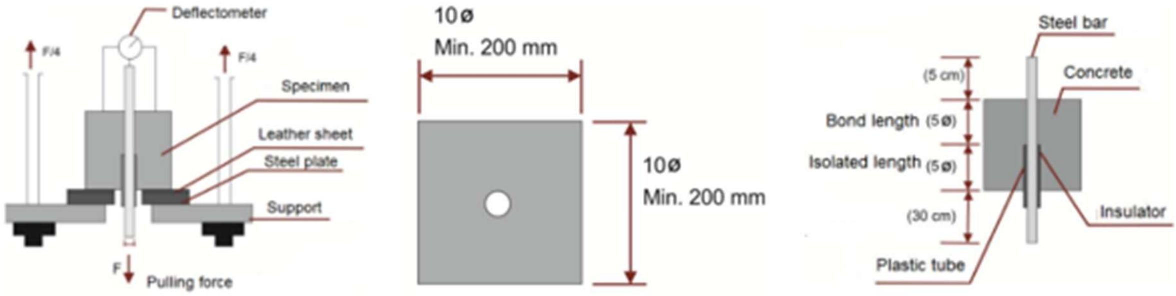

Figure 1 - Pull-out test set-up [11].

Analytical techniques and numerical methods have been developed to determine the steel-concrete bonding stress, but they have limitations. This is because the number of variables involved in this problem is large and also because there is a non-linear relationship between these parameters. Many authors have proposed empirical equations, derived from laboratory tests, but they face the same problem. As Makni [6] describes, machine learning algorithms are alternative techniques to overcome these limitations. Furthermore, the difficulty related with destructive testing makes the search for unconventional methods relevant. Hence, this paper proposes the use of computational intelligence methods in the prediction of the steel-concrete bonding behavior, corroborating with other studies on the steel-concrete bonding, especially with thin bars.

Recent studies of the authors´ research team have developed machine learning algorithms for the prediction of adhesion stress in thin bars (diameter less than 10mm) [7,8,9,10,11]. The samples used in this and other analyses come from the study of Carvalho et al [12].

This work uses machine learning techniques, called boosting algorithms, to predict the maximum steel-concrete bonding strength. The following metrics will be calculated for each sample in the database: the root mean square error (RMSE), the coefficient of variation (CV) and the error. This study introduces a new approach to determine the steel-concrete bonding strength by establishing the maximum applied load using boosting algorithms.

2 Methodology

2.1 Data Description and Preprocessing

The current study proposes the use of intelligent algorithms to predict the steel-concrete bonding strength using a database from the academic literature - the studies by Carvalho et al. [12]. The work used as reference deals with the pull-out test that evaluated the steel-concrete bonding behavior in thin bars. The experimental program employed concretes of class C25, C35, and C40; CA-50 ribbed bars (with diameters of 6.3, 8.0, and 10 mm); and CA-60 notched bars (with diameters of 5.0, 6.0, 8.0, and 9.5 mm).

This work required the acquisition of reliable experimental data to determine the steel-concrete bonding by means of computational intelligence. The data from Carvalho's study [12] were processed using a statistical technique to remove outliers and is presented by Arruda [10].

Four input parameters and one output parameter were used. The input parameters are as follows: concrete compressive strength (fc), bar diameter (ϕ), anchorage length (Ld) and the surface geometry of steel bars. The output parameter is the maximum applied load (in N). Table 1 shows the maximum and minimum values of the input and output components. As stated above, the types of steel bars used in this study are ribbed and notched.

Table 1. Database attributes

Database attributes	Values	
	Maximum	Minimum
Diameter (mm)	10.0	5.0
Compressive strength of concrete (MPa)	47.0	22.8
Anchorage length (mm)	50.0	25.0
Maximum Pull-out Force (N)	19578.0	1129.3

Machine Learning algorithms are divided according to their learning type, i.e., the way they learn. There are three main categories: supervised, unsupervised and reinforcement learning. The models proposed in this work use reinforcement learning, with the division of data for training (85%), and testing (15%). The procedure adopted for development is described in the flowchart shown in Figure 2.

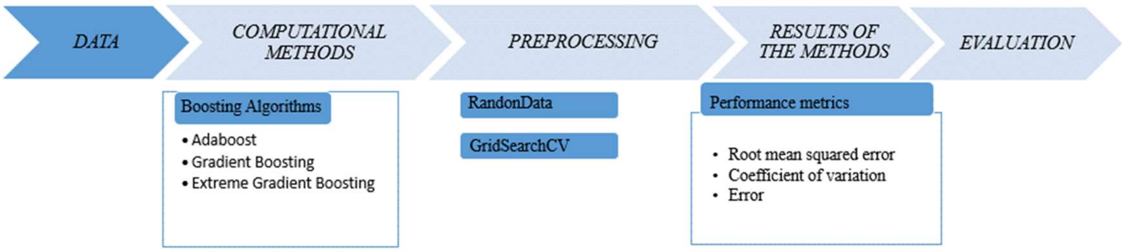

Figure 2- Boosting Algorithms Sequence

2.2 The boosting algorithms

Boosting algorithms are computational modeling techniques first presented by Freund and Schapire [13]. These algorithms improve predictive power by converting several weak learners into strong learners. They share the central idea of combining several weak models (usually shallow decision trees) to create a strong model that is better able to make accurate predictions. The procedure goes as follows: (1) create one model on the training dataset, and then (2) create a second model to rectify the errors present in the first model. This procedure continues until, and unless, the errors are minimized, and the dataset is predicted correctly. The final model is a weighted combination of the weak predictions. The main reinforcement algorithms include:

Adaboost: The so-called Adaptive Boosting is a popular machine learning technique. The most common estimator used with AdaBoost is decision trees, called Decision Stumps. The method assigns higher weights to instances that were poorly ranked in the previous iteration and trains on the modified dataset. The final model is the measured sum of the predictions of these weak learners. According to Zhang et al [14], the AdaBoost method is sensitive to noisy data and outliers. This means that if there is a lot of noisy data, the time taken by the AdaBoost algorithm will increase and its efficiency will decrease.

Gradient Boosting: Like the above technique, Gradient Boosting creates a robust model by combining several weak learners, commonly decision trees. The system adjusts the new models considering the residual error (difference between predicted and actual values) of the previous iteration. This minimizes the residual errors of subsequent iterations, gradually improving the overall performance of the method.

Extreme Gradient Boosting: This is an advanced and optimized version of Gradient Boosting which features an optimized decision tree that is built in such a way as to minimize the loss function, which is defined according to the type of problem (classification or regression). Its formulation includes regularization techniques (Lasso and Ridge) which are added to the cost function, thus avoiding overfitting. Another positive aspect is that the method offers the possibility of cross-validation during the training process, helping to adjust the model's hyperparameters more precisely.

In short, the goal of these algorithms is common, which is to improve accuracy by combining weak models. However, they differ in the way they adjust subsequent models, how they handle model weights, and how they apply regularization and optimization techniques.

Several authors have used boosting algorithms as a computational method for various purposes. In Cai [15], the Gradient boosting regression algorithm is used to predict the Net ecosystem carbon Exchange, which is affected by a set of meteorological variables to different degrees, helping to assess the balance of the carbon cycle between biological organisms and the atmosphere. In Wang et al [16], two boosting algorithms, adaptive boosting (Adaboost) and light gradient boosting machine (LightGBM), were developed among the models tested, with the aim of evaluating the efficiency of an explainable ensemble learning framework in accurately predicting the bond strength between steel sections with different surface treatments and various types of concrete.

Another authors that presented reinforcement algorithms in their work were Li and Song [17], whose work aimed to obtain the compressive strength and tensile strength of high-performance concrete (HPC) from data obtained from experimental tests. Predictive models were created, including adaboost, gradient boosted decision tree and XGB. The models were evaluated and it was concluded that, among those tested, the gradient boosted decision tree had the best performance. In Kim [18], the main objective of was to predict the strength of the interfacial bonding between fiber-reinforced polymer and concrete using boosting algorithms, including Catboost, histogram gradient boosting algorithm, extreme gradient boosting algorithm. CatBoost outperformed the other methods.

2.3 Performance Metrics

As mentioned, $RMSE_i$, given by Equation (1), $ERROR_i$, as per Equation (2), and CV_i, given by Equation (3), are used as performance metrics. In the equations, y_i is the observed value of sample, $\bar{y}$ is the mean of the observations of sample, $\hat{y}$ is the predicted value of y_i of sample, σ_i sample standard deviation of sample, i sample number and N_i is number of sample members.

$$RMSE_i = \sqrt{\frac{1}{N_i}\sum_{i=1}^{N_i}(y_i - \hat{y})^2} \tag{1}$$

$$ERROR_i = \frac{RMSE_i}{\hat{y}_i} \tag{2}$$

$$CV_i = \frac{\sigma_i}{\bar{y}_i} \tag{3}$$

3 Presentation and analysis of the results

3.1 The dataset results

For this work, it was necessary to acquire the treated database available in Arruda et al [10]. Once the database was determined, the database used was divided into 17 samples according to the similarity of the input parameters, which yielded the mean maximum pull-out forces and standard deviations shown in Figure 3.

The coefficient of variation (CV) is a statistical measure used when one wants to compare the variation of sets of observations that differ in mean or are measured at different magnitudes. The CV is widely used when conducting quality assurance studies and evaluating repeatability and reproducibility. It is a pure dimensionless number, which, when considered low, indicates a reasonably homogeneous data set. It can be difficult to classify a low, medium, high or very high CV, but it can be very useful in comparing two variables that are not initially comparable.

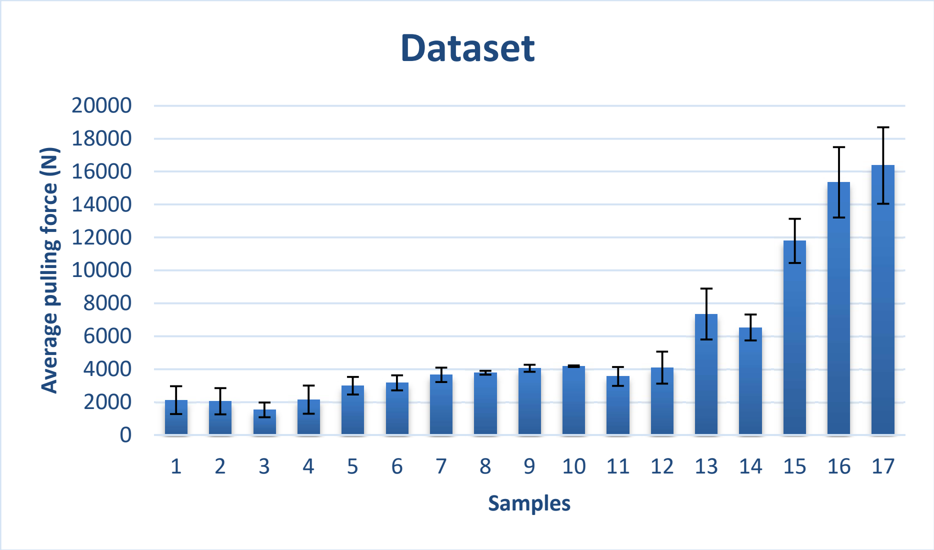

Figure 3 - Average pulling force x samples

Table 2 shows the CV of each sample set. For notched bars, the CV values shows averages of 37%, 15%, 4% and 9% for diameters of 5 mm, 6 mm, 8 mm and 9.5 mm, respectively. The 5 mm bar presented the worst CV in this data set, which can be explained by the difficulties inherent in the test, i.e., to concrete a 5 mm bar in a 200 mm specimen and ensuring the correct anchoring length. For ribbed bars, the coefficient of variation averaged 22.5%, 11.5% and 14% for diameters of 6.3 mm, 8 mm and 10 mm respectively.

Table 2. Metrics of the dataset

INPUT DATA								
i	Bar type	Ø (mm)	fck (Mpa)	Lb (mm)	$\bar{y}_i$	σ_i	CV_i	N_i
1	Notched CA-60	5	29.05	25	2127.88	848.60	0.40	6
2	Notched CA-60	5	27.9	25	2052.68	794.81	0.39	5
3	Notched CA-60	5	41.55	25	1533.32	451.38	0.29	5
4	Notched CA-60	5	45.3	25	2161.13	849.23	0.39	4
5	Notched CA-60	6	28.25	30	2997.20	530.31	0.18	4
6	Notched CA-60	6	28.9	30	3177.98	461.89	0.15	5
7	Notched CA-60	6	45.1	30	3666.93	433.69	0.12	4
8	Notched CA-60	8	31.6	40	3790.75	114.34	0.03	2
9	Notched CA-60	8	40.1	40	4056.10	209.85	0.05	4
10	Notched CA-60	9.5	30.2	47.5	4179.93	47.06	0.01	4
11	Notched CA-60	9.5	41.4	47.5	3570.23	575.93	0.16	4
12	Ribbed CA-50	6.3	22.8	31.5	4102.08	977.30	0.24	6
13	Ribbed CA-50	6.3	46.95	31.5	7353.02	1546.31	0.21	5
14	Ribbed CA-50	8	29.5	40	6530.18	786.86	0.12	5
15	Ribbed CA-50	8	46.6	40	11797.00	1341.82	0.11	3
16	Ribbed CA-50	10	27.9	50	15353.89	2139.94	0.14	9
17	Ribbed CA-50	10	45.3	50	16370.44	2326.07	0.14	9

3.2 The boosting algorithms results

When using machine learning techniques, such as boosting algorithms, it is paramount to carefully choose hyperparameters to be used in the implementation. In the present work, the GridSearchCV method was employed to help in this choice. GridSearchCV is used to choose the best parameters for the model. The GridSearchCV output of the GridSearchCV defines the "optimal" parameters for each method.

The use of various metrics to evaluate the performance of an intelligent system is commonplace in the academic environment. In general, when calculating metrics for an intelligent algorithm, the target and predicted values of all the test data are used. However, the use of a general metric is not applicable in the case of predicting the adhesion strength of thin bars, due to the great variability of the samples used in this work (Figure 3). Once the database was set, as well asand the methods to be used, the input parameters in each method were chosen (GridSearchCV) and the methods were executed, the methods were implemented, thus determining the predicted value for each sample. With these values, the RMSE and ERROR values were calculated for each sample. The metrics of the dataset are shown in Table 3.

Table 3. Performance metrics of the boosting algorithms

SAMPLES	Ø (mm)	ADABOOST		GB		XGB	
i		$RMSE_i$	$ERROR_i$	$RMSE_i$	$ERROR_i$	$RMSE_i$	$ERROR_i$
1	5	779.01	0.35	775.67	0.36	776.95	0.38
2	5	729.08	0.33	723.94	0.33	710.90	0.35
3	5	772.59	0.35	575.75	0.30	413.01	0.25
4	5	735.46	0.34	748.45	0.33	803.31	0.32
5	6	599.61	0.18	510.91	0.16	460.68	0.15
6	6	461.07	0.14	498.55	0.14	414.41	0.13
7	6	414.81	0.12	397.95	0.10	410.09	0.11
8	8	114.58	0.03	217.49	0.05	88.48	0.02
9	8	394.17	0.11	192.49	0.05	185.93	0.05
10	9,5	472.02	0.13	122.19	0.03	89.50	0.02
11	9,5	513.78	0.14	547.32	0.14	498.82	0.14
12	6,3	1133.24	0.23	867.55	0.19	803.84	0.20
13	6,3	1399.41	0.18	1404.80	0.18	1383.06	0.19
14	8	868.69	0.14	703.83	0.11	704.18	0.11
15	8	1095.74	0.09	1127.39	0.10	1095.95	0.09
16	10	2088.54	0.14	2047.71	0.14	2019.58	0.13
17	10	2201.50	0.13	2369.84	0.15	2193.04	0.13

In Adaboost, for notched bars, the error value yielded averages of 34%, 14%, 7% and 13%, for diameters of 5 mm, 6 mm, 8 mm, and 9.5 mm, respectively. For ribbed bars, the error averages were 21%, 12% and 14% for diameters of 6.3 mm, 8 mm, and 10 mm, respectively. In Gradient Boosting, for notched bars, the error value showed averages of 33%, 10%, 5%, and 9%, for diameters of 5 mm, 6 mm, 8 mm and, 9.5 mm, respectively. For ribbed bars, the error averages were 21%, 10% and 14% for diameters of 6.3 mm, 8 mm, and 10 mm, respectively. In Extreme Gradient Boosting, for notched bars, the error value showed averages of 33%, 13%, 3% and 8%, for diameters of 5 mm, 6 mm, 8 mm, and 9.5 mm, respectively. For ribbed bars, the error averages were 19%, 10% and 13% for diameters of 6.3 mm, 8 mm, and 10 mm, respectively. The worst performance for the methods was obtained by the 5 mm notched bar, which can be explained by the difficulties inherent to the test, as mentioned above. It is therefore possible to conclude that a trend was observed: the better the test's CV, the better the predicted value and, consequently, the error obtained by the computational method. This shows that the computational method responds to the characteristics of the adopted database.

4 Conclusions

This work had as general objective to present a study of the application of machine learning techniques in determining the maximum pull-out force in pull-out tests. In the original database [11], statistical techniques were developed aiming at the removal of spurious data [10]. Three reinforcement algorithms were implemented: Adaboost, Gradient Boosting, and Extreme Gradient Boosting. Once the input and output parameters were defined, RandonData and GridSearchCV techniques were used to place the data in random order and determine the hyperparameters for each method. After setting the methods and hyperparameters, the reinforcement algorithms were implemented, thus determining the maximum pull-out force of each sample. With these data, the performance metrics were calculated for each sample. The results obtained in the test and in the computational method were compared, verifying that there is a trend between the coefficient of variation of the test and the error obtained in the implementations. In the end, the values predicted by the methods, and consequently the error, and the coefficients of variation obtained are highly consistent. Based on them, one can conclude that the presented model can satisfactorily predict the steel-concrete bonding strength of samples, even with thin bars.

Acknowledgements. CAPES and CEFET-MG supported the work described in this paper.

References

[1] G. F. Belo, *Bond resistence between steel and fresh concrete in concreting conditions subject to vibration.* MSc Dissertation - Federal University os Paraíba (UFPB), 101 p. João Pessoa:2017

[2] F. Leohard and E. Mong, Construções de concreto-princípios básicos do dimensionamento de estruturas de concreto armado. *Rio de Janeiro: Interciência*, v. 1, 1977.

[3] L. M. Pinheiro et al. *Estruturas de concreto–capítulo 11*. 2016

[4] BS, E. N. Steel for the Reinforcement of Concrete: Weldable Reinforcing Steel: General. *Central Secretariat: Brussels, Belgium*, 2005.

[5] F. M. Almeida et al., "Numerical approach of the steel-concrete bond behavior using pull-out models". *Matéria (Rio de Janeiro)*, v. 24, 2019.

[6] M. Makni et al. Artificial neural network for the prediction of the steel–concrete bond behaviour. *European journal of environmental and civil engineering*, v. 18, n. 8, p. 862-881, 2014.

[7] P. F. S. Silva, et al. A Computational Approach to Predict the Bond Strength of Thin Steel Rebars in Concrete by Means of Artificial Neural Networks. *Proceedings of the joint XLIII Ibero-Latin-American Congress on Computational Methods in Engineering*, ABMEC, 2022.

[8] P. F. S Silva, et al. A Computational Approach to Predict the Bond Strength of Thin Steel Rebars in Concrete by Means of Support Vector Machine . *Proceedings of the joint XLII Ibero-Latin-American Congress on Computational Methods in Engineering and III Pan-American Congress on Computational Mechanics, ABMEC-IACM*, 2021.

[9] P. F. S. Silva, et l. A Computational Method to Predict the Concrete Compression Strength Using Decision Trees and Random Forest. *Proceedings of the XLI Ibero-Latin-American Congress on Computational Methods in Engineering*, 2020.

[10] V. F. Arruda, et al. Técnicas estatísticas para o pré-processamento de dados experimentais para o uso em algoritmos inteligentes. *Anais da XXXIX Jornadas Sulamericanas de Engenharia Estrutural. Porto Alegre: ASAEE, 2022.* v. 1. p. 1-8.

[11] V. F. Arruda, et al. The use of intelligent algorithms in the prediction of bonding strength in steel-concrete interfaces. *Proceedings of the joint XLIII Ibero-Latin-American Congress on Computational Methods in Engineering, ABMEC*, 2022.

[12] E. P. Carvalho, et al., "Experimental investigation of steel-concrete bond for thin reinforcing bars". *Latin American Journal of Solids and Structures*, v. 14, p. 1932-1951, 2017.

[13] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.

[14] Y. Zhang et al. Research and application of AdaBoost algorithm based on SVM. *IEEE 8th joint international information technology and artificial intelligence conference (ITAIC). IEEE*, 2019. p. 662-666.

[15] J. Cai et al. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. *Applied energy*, v. 262, p. 114566, 2020.

[16] X. Wang et al. Explainable ensemble learning model for predicting steel section-concrete bond strength. *Construction and Building Materials*, v. 356, p. 129239, 2022.

[17] Q. Li and Z. Song. High-performance concrete strength prediction based on ensemble learning. *Construction and Building Materials*, v. 324, p. 126694, 2022.

[18] B. Kim et al. Ensemble machine learning-based approach for predicting of FRP–concrete interfacial bonding. *Mathematics*, v. 10, n. 2, p. 231, 2022.