



## ORIGINAL ARTICLE

# Comparison of machine learning techniques to predict the compressive strength of concrete and considerations on model generalization

*Comparação de técnicas de aprendizado de máquina para prever a resistência à compressão do concreto e considerações sobre a generalização de modelos*

Rafael Christian Fonseca da Paixão<sup>a</sup>

Rúben El-Katib Penido<sup>b</sup>

Alexandre Abrahão Cury<sup>c</sup>

Júlia Castro Mendes<sup>b</sup>



<sup>a</sup>Universidade Federal de Ouro Preto - UFOP, Departamento de Engenharia de Controle e Automação, Ouro Preto, MG, Brasil

<sup>b</sup>Universidade Federal de Ouro Preto – UFOP, Departamento de Engenharia Civil, Ouro Preto, MG, Brasil

<sup>c</sup>Universidade Federal de Juiz de Fora – UFJF, Departamento de Mecânica Aplicada e Computacional, Juiz de Fora, MG, Brasil

Received 20 October 2021

Accepted 31 January 2022

**Abstract:** The compressive strength of concrete is an essential property to ensure the safety of a concrete structure. However, estimating this value is usually a laborious and uncertain process since the mix design is based on empirical methods and its confirmation in the laboratory demands time and resources. In this context, this work aims to evaluate Machine Learning (ML) models to predict the compressive strength of concrete from its constituents. For this purpose, a dataset from the literature was used as input to four ML models: Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), Artificial Neural Networks (ANN) and Gaussian Process Regression (GPR). The accuracy of the models was evaluated through 10-fold cross-validation, and quantified by  $R^2$ , Mean Absolute Error (MAE), and Root-Mean-Square Error (RMSE) metrics. Subsequently, a new dataset was put together with mixtures from the literature and used to validate the previous models. In the model creation step, all algorithms obtained similar and positive results, with MAE between 1.96-2.26 MPa and  $R^2$  varying from 0.79 to 0.83. However, in the validation step, the accuracy of the models dropped sharply, with MAE growing to 3.04-4.04 MPa and  $R^2$  decreasing to 0.37-0.59. ANN and GPR showed the best results, while SVR had the worst predictions. This work showed that ML tools are promising techniques to predict the compressive strength of concrete. However, care must be taken with the input data to guarantee that models are not overfitted to a given region, set of materials, or type of concrete.

**Keywords:** machine learning, concrete mix design, generalization ability, compressive strength, concrete database.

**Resumo:** A resistência à compressão do concreto é uma propriedade essencial para garantir a segurança de uma estrutura. No entanto, estimar este valor é atualmente um processo trabalhoso e impreciso, uma vez que o a dosagem é baseada em métodos empíricos e sua confirmação em laboratório demanda tempo e recursos. Nesse contexto, este trabalho tem como objetivo avaliar modelos de Aprendizado de Máquina (ML) para prever a resistência à compressão do concreto a partir de seus componentes. Para tanto, um banco de dados da literatura foi utilizado como entrada para quatro modelos de ML: *Extreme Gradient Boosting* (XGBoost), Regressão de Vetor de Suporte (SVR), Redes Neurais Artificiais (ANN) e Processo Gaussiano de Regressão (GPR). A precisão dos modelos foi avaliada por meio de validação cruzada (10-*fold*) e medida com as métricas de  $R^2$ , Erro Médio Absoluto (MAE) e a Raiz do Erro Quadrático Médio (RMSE). Posteriormente, um novo banco de dados foi montado com traços da literatura e utilizado para validar os modelos anteriores. Na etapa

**Corresponding author:** Júlia Castro Mendes. E-mail: [julia.mendes@ufop.edu.br](mailto:julia.mendes@ufop.edu.br)

**Financial support:** FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais, project PPM-00001-18 for A.A.C. and APQ-01838-21 for J.C.M.), PROPPI/UFOP (UFOP Research and Innovation Dean's Office, for Undergraduate Research Scholarship for R.C.F.P. and Research Assistance Grant for J.C.M.), CNPq (National Council for Scientific and Technological Development, project 304329/2019-3 for A.A.C.) and CAPES (Coordination of Superior Level Staff Improvement, for master's scholarship to R.E.P.).

**Conflict of interest:** Nothing to declare.

**Data Availability:** The data that support the findings of this study are available from the corresponding author, JM, upon reasonable request.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

de criação do modelo, todos os algoritmos obtiveram resultados semelhantes e satisfatórios, com MAE entre 1,96-2,26 MPa e  $R^2$  variando de 0,79 a 0,83. No entanto, na etapa de validação, a precisão dos modelos caiu drasticamente, com o MAE crescendo para 3,04-4,04 MPa e o  $R^2$  diminuindo para 0,37-0,59. As ANN e o GPR mostraram os melhores resultados, enquanto a SVR teve as piores previsões. Este trabalho mostrou que as ferramentas de ML são técnicas promissoras para prever a resistência à compressão do concreto, porém, deve-se ter cuidado com os dados de entrada para garantir que os modelos não sejam sobreajustados (*overfitted*) a uma determinada região, conjunto de materiais ou tipo de concreto.

**Palavras-chave:** aprendizado de máquina, dosagem de concreto, habilidade de generalização, resistência a compressão, banco de dados de concreto.

---

**How to cite:** R. C. F. Paixão, R. E. K. Penido, A. C. Cury, and J. C. Mendes, "Comparison of machine learning techniques to predict the compressive strength of concrete and considerations on model generalization" *Rev. IBRACON Estrut. Mater.*, vol. 15, no. 5, e15503, 2022, <https://doi.org/10.1590/S1983-41952022000500003>.

## 1 INTRODUCTION

The compressive strength of concrete is one of its most important properties. This feature directly impacts the structural design and is related to the cost, safety, and stability of a concrete structure. This strength is usually expressed in MPa, and is traditionally obtained from the rupture of cylindrical or cubic specimens in a hydraulic press, a procedure standardized worldwide [1], [2]. Due to the evolution of cement hydration over time, engineers stipulate that this target strength is reached after 28 days of cure in most conventional projects.

As structural projects demand a given compressive strength, the engineers responsible for the construction sites need to establish an optimized proportion among the constituents of the concrete to guarantee the safety of the building. This is done using mix design methods, such as the ones developed by the American Concrete Institute (ACI), the Brazilian Association of Portland Cement (ABCP) and the Brazilian Technological Research Institute (IPT). These methods seek to achieve an average target value (above the minimum) so that the minimum value is met with a safety margin [3]. This average value is obtained statistically, as it is possible for a concrete specimen to obtain a lower strength than specified, given the heterogeneous nature of its components and mixing procedure [1]. Therefore, in practice, the economically viable target strength is defined as the value to be exceeded by a certain proportion of all results (usually 95% when a single test is considered, or 99% when an average of 3 or 4 tests is taken) [1].

These well-established methods are nowadays still performed through charts and empirical formulae [1], [4]. Additionally, they are only valid for conventional concrete. For other types of concrete, such as high-strength, self-compacting, lightweight, and recycled concretes, the scenario is even more uncertain, with scarce and divergent mix design techniques [5], [6].

Like strength evaluation, other concrete-related areas deal with empirical processes and time-consuming tests. To improve these processes, or at least reduce the need for experimental tests, several studies of Machine Learning (ML) techniques applied to civil engineering problems have been published in recent years. ML techniques consist of computational models capable of autonomously acquire knowledge. These models make decisions and can predict new results based on patterns acquired from previous data. As examples, we can cite Yaseen et al. [7], who applied ML techniques to measure the shear strength of reinforced concrete beams and concluded that these algorithms can be useful tools for professionals. Pettres and de Lacerda [8] obtained positive results in the recognition of defect patterns in concrete with the use of Artificial Neural Networks (ANN). ML-based algorithms are also being successfully used in the field of Structural Health Monitoring, especially in applications involving damage detection in large-scale concrete structures, such as bridges, dams, and buildings [9]–[10].

Some authors have also tried to predict the compressive strength of concrete using ML techniques. For example, Hoang et al. [11] applied the Gaussian Process Regression (GPR) to predict concrete strength using a dataset of 246 mixtures, defined according to the Vietnamese standard. The authors achieved an  $R^2$  (coefficient of determination) of 0.90, concluding that these models are a promising alternative to assist engineers in construction sites. In turn, Dao et al. [12] tested the accuracy of ANN and GPR to the dataset assembled by Yeh [13], currently one of the most used worldwide, using a Monte Carlo simulation. The dataset was simply split into 70% of the observations for training and 30% for testing. The authors obtained an  $R^2$  of 0.89 with the GPR and indicated that these algorithms may contribute to the mix design process. Likewise, Mustapha and Mohamed [14], also using Yeh's [13] dataset without cross-validation, obtained an  $R^2$  of 0.93 by applying the Support Vector Regression (SVR). Finally, Cui et al. [15] used a decision tree model for this same purpose, obtained an  $R^2$  above 0.80, and concluded that these models are suitable to assist in the mix design of concretes.

Thus, ML techniques are promising tools to predict the compressive strength of concrete. However, no article was found comparing the Extreme Gradient Boosting Decision Tree (XGBoost), GPR, SVR, and ANN to this purpose within

the same dataset and boundary conditions. Furthermore, to the author’s best knowledge, no article validated the models trained from the traditional Yeh dataset [13] with a different dataset to test the generalization ability of the models.

In this sense, the present work compares the accuracy of these four ML techniques in predicting the compressive strength of conventional concrete specimens and evaluates the resulting models in terms of their generalizing capabilities to a different dataset. The authors seek, therefore, to find the most suitable technique to use in future predictions and to reflect on the limitations of applying these models to concretes in diverse contexts.

## 2 METHODOLOGY

### 2.1 Methodology Overview

Figure 1 shows an overview of the present work. Initially, four supervised ML models were developed to relate the input features (concrete components and proportions) to the target variable (compressive strength). These models were built using a classic dataset available in the literature, gathered by Yeh [13]. We subsequently evaluated the quality of the prediction through cross-validation and three statistical metrics: coefficient of determination ( $R^2$ ), the mean absolute error (MAE) and the root mean square error (RMSE). The significance of each input variable (concrete component) in the prediction of the final compressive strength was also investigated. In a second stage, for the validation of these models, the authors assembled a second dataset from 11 articles in the literature, with 22 new observations (in the Appendix). This dataset was then used as test values for the previously created models. The accuracy of this prediction was again assessed using the 3 metrics described above.

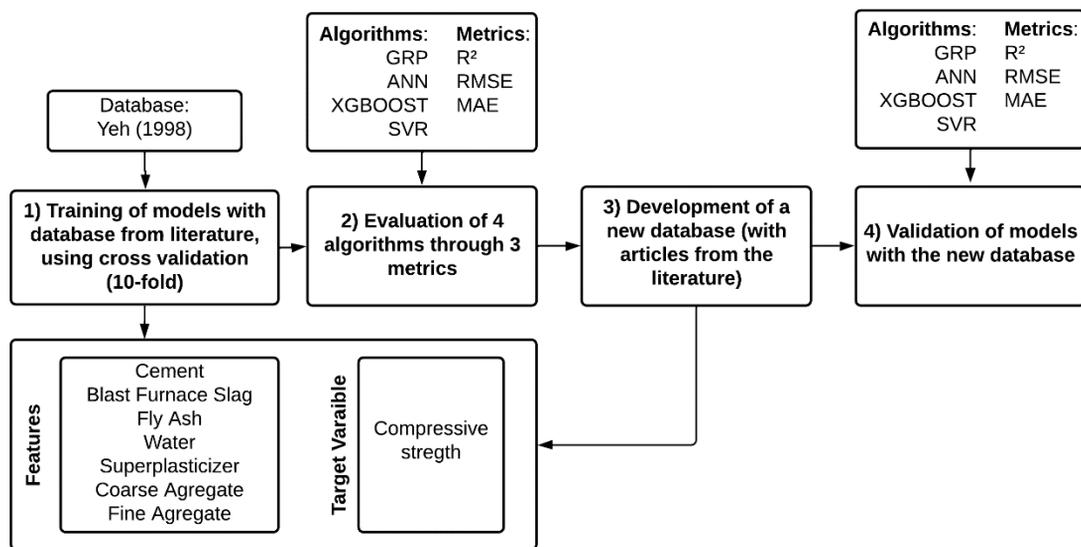


Figure 1. Methodology Overview

### 2.2 Machine Learning Techniques

As there is no single general model perfectly adaptable to all engineering problems, four supervised models were chosen to be applied to the present study: XGBoost, SVR, ANN and GPR. They were selected based on a preliminary literature analysis, in which we gathered the techniques that had different learning-based backgrounds. Among them, XGBoost, SVR, ANN and GPR were the ones with the most promising performance to deal with similar complex problems.

The authors opted to manually adjust the hyperparameters of the techniques without focusing on specific optimization methods for each one, so that there would be no distinction in the creation processes of these models. The experiments were carried out on a computer with an Intel Core i5-10210U processor and 8GB of RAM. The algorithms were implemented in Python (version 3.8.6) using the Pandas library to analyze and manipulate datasets, and the scikit-learn, TensorFlow and XGBoost libraries to apply the ML models.

The following sections will provide a summarized description of these methods. For more detailed explanations, the reader may consult the references given at the end of each part.

### 2.2.1 Extreme Gradient Boosting (XGBoost)

XGBoost has been increasingly used in several research fields because it presents suitable predictions and a short execution time to solve classification and regression problems [16]. This algorithm is based on the classical decision tree technique.

The structure of a decision tree can be described as follows: the tree starts with a major node called “root” that splits into several other nodes. Each of these nodes carries a condition to separate the dataset into subsets that have similar characteristics [17]. Generally, using only one decision tree leads to poor predictions; therefore, ensemble techniques are usually adopted to improve the performance of these models [18]. Ensemble methods consist of combining several trees to achieve more reliable results.

An example of ensemble is the boosting technique, which uses “n” weak trees sequentially to create a more robust predictor at the end of training [19]. The focus of the boosting method is to reduce bias and variance with each new model created, based on the difficulties faced by the previous model [20]. XGBoost uses gradient boosting, an extension of the previous method, in which a descending gradient is applied to improve the trees, according to the error of the previous models.

The XGBoost can be briefly described as follows: for a given dataset  $D = \{(x_i, y_i)\} (|D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R})$ , with  $x_i$  and  $y_i$  variables (inputs and outputs, respectively),  $m$  features, and  $n$  observations, the model uses  $K$  additive functions to predict outputs:

$$\hat{y}_i = \phi(X_i) = \sum_{k=1}^K f_k(x_i), f_k \in F \tag{1}$$

with  $\hat{y}_i$  being the model output and  $F$  the space of the regression tree, defined as:

$$F = \{f(x) = w_{q(x)}\} (q : \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T) \tag{2}$$

The structure of each tree is represented by  $q$ , while the number of leaves and their weights are represented by  $T$  and  $w$ , respectively. Also, the term  $f_k$  represents an independent tree structure  $q$  with  $w$  leaf weights.

In the regression tree optimization process, the following objective function must be minimized:

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{3}$$

There is also a convex loss function  $l$  that measures the difference between  $\hat{y}_i$  and  $y_i$  which are, respectively, the prediction given by the model and the real value. The term  $\Omega$  penalizes the complexity of the regression trees and is given by:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \epsilon \|w\|^2 \tag{4}$$

However, models that use gradient boosting are trained in an additive way. In these cases, the following objective function is minimized:

$$L(\phi) = \sum_i l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{5}$$

$f_t$  is added in the objective function, with  $t$  being the number of iterations [16].

Regarding the implementation of the algorithm, this model does not need many adjustments. Hence, the authors carried out some preliminary tests to define its optimal hyperparameters. For a more detailed explanation about this method, the authors recommend the references Chen & Guestrin [16] and Suen *et al* [20].

### 2.2.2 Support Vector Regression (SVR)

Support Vector Machine (SVM) is a supervised learning model that creates a hyperplane capable of separating data into distinct classes [21]. There are infinite hyperplanes able to perform this task. However, this algorithm seeks to find the one that yields the greatest distance between the classes. To this purpose, the SVM finds the points located on the margins (the support vectors) and maximizes the margin [22]. In other words, the algorithm initially defines a hyperplane that separates the data to later determine the points of each class that are closest to this separator. Finally, it seeks the hyperplane that leads to the greatest distance between the two classes, called the “optimum” hyperplane [23].

In addition to linear problems, these algorithms can be used to solve non-linear problems, by using kernels. Applying the kernel to the model increases the number of dimensions of the input space, thus transforming the initially non-separable data into data that is separable by the algorithm [23].

Given that the prediction of concrete strength is a regression problem, the authors used the Support Vector Regression (SVR) variant in this work. It has the same principle as SVM but focuses on solving regression problems.

The SVR can be briefly described as follows: for a dataset  $\{(X_1, y_1), \dots, (X_l, y_l) \in X \times \mathbb{R}\}$ , where  $X_i$  represents the space of the input variables, the purpose of the regression is to find a function  $f(x)$  that has at most one deviation  $\varepsilon$  from the real values  $y_i$ . For the linear function:

$$f(x) = \langle w, x \rangle + b \text{ with } w \in X, b \in \mathbb{R} \tag{6}$$

the SVR will transform this problem into a constrained optimization problem:

$$\min \frac{1}{2} \|w\|^2 \tag{7}$$

subject to the following restrictions:

$$\begin{cases} y_i - \langle w, X_i \rangle - b \leq \hat{a} \\ \langle w, X_i \rangle + b - y_i \leq \hat{a} \end{cases} \tag{8}$$

The error of the model's predictions is dealt with within the constraints. The SVR model adopts an  $\varepsilon$ -insensitive loss function, which penalizes predictions that are farther than  $\varepsilon$  from the desired output [24].

To perform the hyperparameter tuning for the SVR model, the authors varied the kernel coefficient (a.k.a. gamma) and the ‘C’ regularization parameter randomly from  $10^{-2}$  to  $10^3$ . The best results were achieved with gamma and C as 0.6 and 33, respectively. For a more detailed explanation about this method, the authors recommend the references Smola and Schölkopf [24] and Noble [23].

### 2.2.3 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) were developed based on studies of the human brain [25]. These algorithms have been widely applied to solve problems in various fields around the world, due to their robustness to deal with complex tasks [26]–[27]. ANNs consist of several processing elements, called neurons, connected to each other. Figure 2 represents the single neuron model, also known as perceptron.

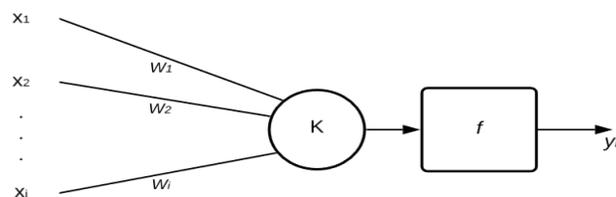


Figure 2. Representation of one perceptron

The neuron will receive the input values  $X_i$ ; these entries are multiplied by the synaptic weights  $w_i$ . Each neuron also has a bias  $b$ . This bias has no input data associated with it, allowing the neuron to change the output independently of the input values. Neuron  $k$  performs the weighted sum of the received signals. Finally, this sum passes through the activation function  $f$  to produce the output  $y_k$ :

$$y_k = f \left( \sum_{i=1}^n w_i X_i + b \right) \tag{9}$$

One of the architectures most used by ANN models is the Multilayer Perceptron (MLP). In MLPs, neurons are divided into the input layer, hidden layers, and output layers, as shown in Figure 3 [28].

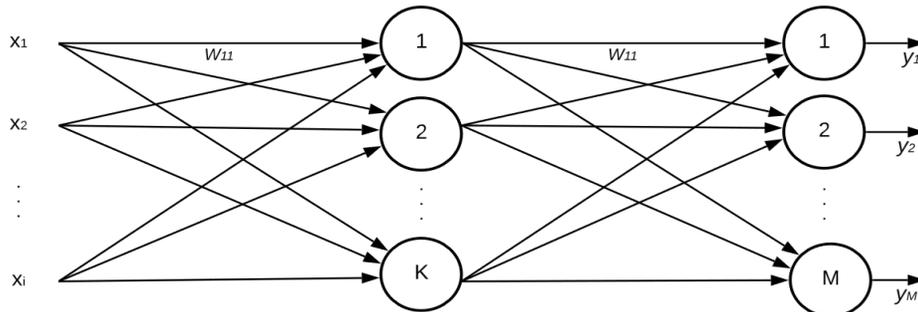


Figure 3. Representation of an MLP with a hidden layer

For an MLP like the one depicted in Figure 3, the mechanism of only one neuron is used for each of the layers:

$$y_j = f \left( \sum_{i=1}^n w_{ij}^{(l-1)} X_i^{(l-1)} + b^{(l-1)} \right), \text{ para } j = 1, \dots, k^l \tag{10}$$

Therefore,  $y_j$  will provide the output of each neuron in its respective layer  $l$  [22]. For a more detailed explanation about this method, the authors recommend the references Garcia [25] and Barreto [28].

To define the number of hidden layers and the number of neurons per ANN layer for the present work, the authors conducted a sensitivity analysis. The model was trained several times with Yeh’s dataset [13], varying the number of layers from 1 to 7, and the number of neurons from 4 to 512, per layer. From the analysis of the evaluation metrics (section 2.3.5), the final model with 5 hidden layers and 256 neurons was implemented.

### 2.2.4 Gaussian Process Regression (GPR)

The Gaussian Process Regression (GPR) is a non-parametric regression technique that uses the probability distribution to predict the outcome. Through the provided training data, this technique uses the Bayes’ rule to update the probabilities of each function representing the model [29]. The main advantage of the GPR is that it provides an approximation of the uncertainty of each forecast [29].

The GPR can be defined as follows:

$$f(x) \sim GP \left( m(x), k(x_i, x_j) \right) \tag{11}$$

where  $m(x)$  is an average function and  $k(x_i, x_j)$  a covariance (or kernel) function of the  $GP$  Gaussian distribution [30] for samples  $x_i \in x_j$ . Choosing the kernel function is one of the most important steps in implementing this model. As in the SVR models, these functions are responsible for smoothing the function being modelled, which will affect the quality of the prediction [30].

This work adopted the Radial Basis Function kernel (RBF). RBF is a stationary kernel function that uses the squared Euclidean distance between two vectors, as follows [31]:

$$k(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2l^2}\right) \quad (12)$$

with  $d(x_i, x_j)$  being the Euclidean distance and  $l$  the kernel function length scale [30]. Based on previous validations, considering several different simulations, this function proved to be the most suitable for the present study.

As in the previous models, a hyperparameter optimization for the GPR was also performed. To this purpose, the parameter “alpha” of the model was randomly varied from  $10^{-3}$  to  $10^2$ . This hyperparameter is the value added to the diagonal of the kernel matrix during the process. The value 0.2 was set. For a more detailed explanation about this method, the authors recommend the references Rasmussen [29] and Williams and Rasmussen [30].

## 2.3 Data analysis

Choosing the right technique, as well as defining a proper dataset, are essential steps in the framework of machine learning. For instance, using a tool that performs well on several problems, but training it with unrepresentative data, will result in poor predictions [32], [33].

### 2.3.1 Training dataset

In the present work, the dataset of concrete compositions was gathered from data available in the literature. For the first part of the construction of the models, the authors used the “Concrete Compressive Strength Data Set” from the studies carried out by Yeh [13]. This dataset has eight input features: Cement, Blast Furnace Slag, Fly Ash, Water, Superplasticizer, Coarse Aggregate, Fine Aggregate, and Age. The set also has the output feature Compressive Strength of Concrete, ranging from 2 to 82 MPa. The complete dataset has 1030 distinct observations (entries). The dataset comprises mixtures from 17 different sources, most of them originated from research carried out between 1987 and 1997, in Taiwan. These mixtures comprised specimens of different shapes and sizes. Thus, the original author performed a standardization, through correlation indices from the literature, so that all the compressive strength results corresponded to 15-cm cylindrical specimens. In addition, the author specified that the coarse aggregate of all the mixtures had dimensions below 20mm and that the superplasticizers were originated from several manufacturers [13].

Pre-processing steps include data preparation prior to making predictions. In general, this part consists in solving scaling problems, analyzing the outliers, and missing values that directly impact the performance of the models [32]. In the present work, the feature responsible for informing the concrete curing time (age) was not used. Due to the design convention of 28 days to achieve the target concrete strength for conventional purposes, only instances that had this age were used. All data referring to other curing times were removed from the set to avoid biases. This step reduced the number of observations from 1030 to 419.

A second adjustment included filtering the values of the output variable. This work aims to evaluate normal-strength concretes, whose values vary between 15-50 MPa [34]. As the outliers are largely responsible for hindering the modelling of the phenomenon, the authors decided to remove all data above 50 and below 15 MPa, seeking to obtain a more robust model for the defined resistance range. Thus, in total, 329 observations (32% of the initial dataset) were adopted for the creation of the models. Table 1 shows the characteristics of the final dataset used.

**Table 1.** Overview of Yeh’s dataset (1998) after pre-processing.

Parameter	Unit	Min	Max	Mean	Std. Dev.
Portland cement	kg/m <sup>3</sup>	102.00	516.00	242.66	87.72
Blast furnace slag	kg/m <sup>3</sup>	0.00	359.40	89.68	89.50
Fly ash	kg/m <sup>3</sup>	0.00	200.10	64.73	65.82
Water	kg/m <sup>3</sup>	121.80	247.00	186.72	17.69
Superplasticizer	kg/m <sup>3</sup>	0.00	22.10	6.18	5.03
Coarse aggregate	kg/m <sup>3</sup>	801.00	1145.00	958.13	80.68
Fine aggregate	kg/m <sup>3</sup>	594.00	945.00	763.81	70.91
Compressive strength	MPa	15.09	49.90	32.36	8.52

### 2.3.2 Validation dataset

To test the generalization ability of the implemented models, the authors assembled a new dataset with concrete mixtures available in the literature. Mixtures were taken from 11 articles (listed in the Appendix), which originated from 8 different countries, including Brazil. For the validation set to be compatible with the model, the compressive strength was standardized to correspond to 150×300mm cylindrical specimens (the same ones used in Yeh's dataset), using the correlations from Yi et al. [35]. In addition, the same data filtering was performed to consider only strengths between 15 and 50MPa. Thus, the final validation set had 22 observations, described in Table 2. The complete dataset can be provided by request to the corresponding author.

**Table 2.** Overview of the authors' validation dataset after pre-processing

Parameter	Unit	Min	Max	Mean	Std. Dev.
Portland cement	kg/m <sup>3</sup>	220.0	568.8	379.8	90.4
Blast furnace slag	kg/m <sup>3</sup>	0.0	410.5	81.0	112.7
Fly ash	kg/m <sup>3</sup>	0.0	25.0	4.5	9.9
Water	kg/m <sup>3</sup>	138.0	250.3	204.9	28.8
Superplasticizer	kg/m <sup>3</sup>	0.0	11.3	2.0	3.2
Coarse aggregate	kg/m <sup>3</sup>	656.3	1029.0	879.2	126.9
Fine aggregate	kg/m <sup>3</sup>	477.7	1029.3	802.4	155.7
Compressive strength	MPa	19.6	40.3	33.5	6.0

### 2.3.3 Data Rescaling

When working with ML, another important factor is the scale of the data. Some models do not perform well with inputs that have different scales, which can lead the model to prioritize a given input simply because it has a bigger scale [36]. Regarding the present work, Table 1 shows that the data referring to the superplasticizer range from 0 to 22 Kg/m<sup>3</sup>, while the coarse aggregate values range from 801 to 1145 Kg/m<sup>3</sup>, evidencing that the data from different features are not in the same magnitude. Thus, the authors rescaled the input data, as follows:

$$X_i^{new} = \frac{X_i^{old} - \mu}{\sigma} \quad (13)$$

where  $X_i^{old}$  is the original input value,  $\mu$  is the average,  $\sigma$  is the standard deviation, and  $X_i^{new}$  the modified input value. After the rescaling step, all values are centered at zero with a standard deviation equal to 1.

### 2.3.4 Cross-validation (k-fold)

Cross-validation is a technique widely used to assist in the evaluation of ML models [37]. It consists of randomly dividing the data into “k” sets, each of which is used to validate the model once [38], [39]. This strategy provides a less biased assessment compared to common techniques such as just splitting data once into training and testing.

This study adopted k=10, which is widely used in the literature for similar problems [5], [11], [40]. Initially, the complete dataset is randomly divided into 10 subsets or folds. In the first iteration, the first subset is used to test the model, after all the others have been used to train it. In the next iteration, the algorithm uses the second segmentation to test the model after it has used everything else for training. This procedure is repeated until all 10 sets have been used to test the model, as illustrated in Figure 4. The results shown in this work correspond to the mean of the 10 iterations.

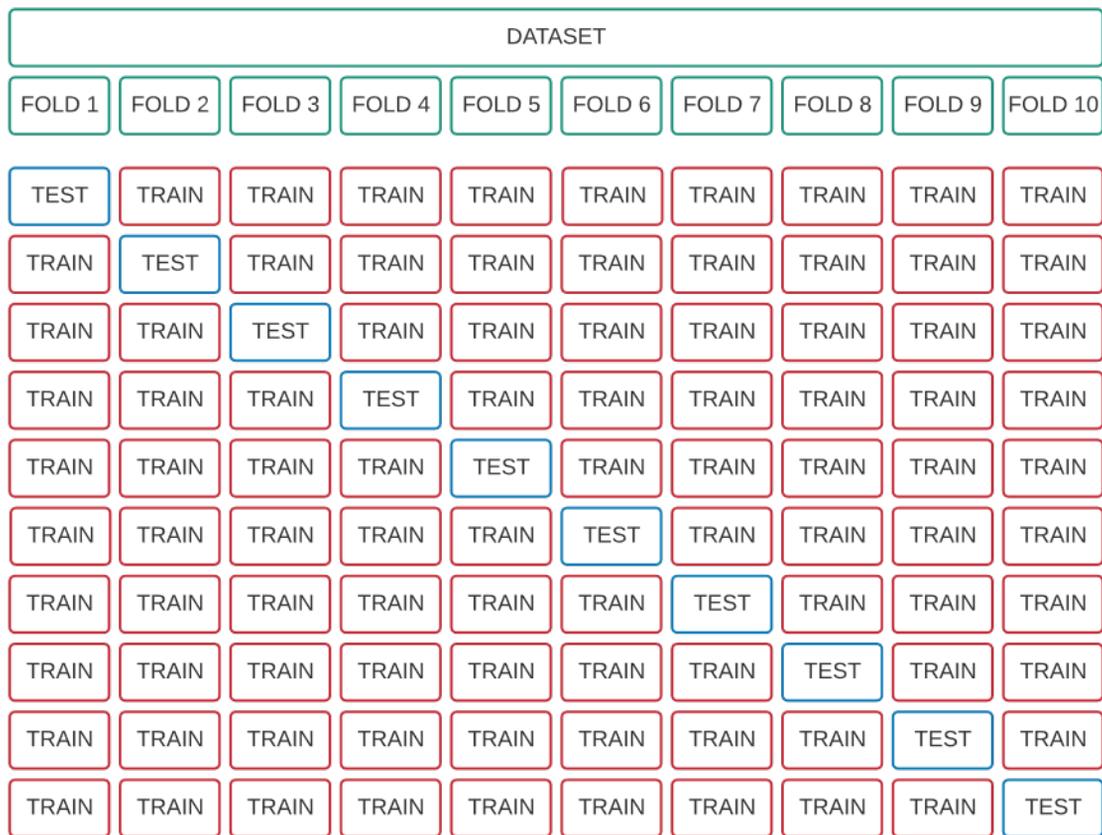


Figure 4. Scheme of the cross-validation process for k=10

### 2.3.5 Assessment Metrics

Three quantitative metrics were used to assess the performance of each model, aiming, together, to provide a global analysis of its accuracy. They are the coefficient of determination ( $R^2$ ), the mean absolute error (MAE) and the root mean square error (RMSE). They are vastly used to assess regression models for this type of problem [5] [41] [42].

The  $R^2$  is calculated using (14) [43], where  $\hat{y}$  is the value predicted by the model and  $y$  is the observed value.  $R^2$  results in a number between minus infinity and 1. When the analyzed model fits perfectly to the data, the  $R^2$  will assume the value 1, indicating that the predictors are able to explain all the variability of the data [44]. As the  $R^2$  compares the performance of the tested model with a flat line (a baseline model in which all predictions will be the mean value of the outputs), if the assessed model presents a worse fit than the line that represents the mean value, the  $R^2$  will be negative.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{14}$$

The MAE measures the average magnitude of the errors (the difference between observed and predicted values), regardless of their direction. It can be determined using (15) [45]. In MAE, large errors caused by outliers are not so important, because this metric is absolute and not quadratic [44].

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{15}$$

Finally, the RMSE ((16 [45]) is a vastly used metric when the researcher wants to measure the average magnitude of the errors [44]. Unlike MAE, in RMSE, as the error of each prediction increases, the RMSE increases considerably.

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{16}$$

Both MAE and RMSE range from zero to positive infinity. The lower these metrics, the better the model.

### 2.4 Significance of the input features

Finally, the authors sought to understand the impact that each feature had on the predictions. For this evaluation, the decision tree technique (XGBoost) was used. In these models, each node has a condition to split the values so that similar instances end up in the same set. The condition is based on the Gini impurity for classification problems and in the variance for the regression problems [46]. Thus, when a decision tree-based model is trained, it intrinsically calculates how much each variable contributes to reducing the variance and, consequently, it can estimate how useful each variable is to the construction of the model. For a  $L$  dataset with  $j$  classes, (17) calculates the Gini impurity, with  $p_i$  being the class probability [47]. The Gini impurity ranges from 0 to 1, with 0 relating to an impure node. The smaller the Gini, the more important that variable is for the tree.

$$GINI(L) = 1 - \sum_{i=1}^j p_i^2 \tag{17}$$

## 3 RESULTS

### 3.1 Creation and evaluation of the models

Table 3 summarizes the evaluation metrics ( $R^2$ , MAE and RMSE) of the four models created to predict the compressive strength of conventional concrete specimens. In this initial stage, the models were trained and cross-validated with the Yeh [13] dataset. The XGBoost achieved the best correlation between predicted and observed values, reaching an  $R^2$  of 0.83. On the other hand, SVR had the worst performance ( $R^2 = 0.79$ ), although it was very close to the other models ( $R^2 = 0.82$ ).

As part of the assessment of the best models to develop future studies, the authors have also recorded the time required to process each algorithm. Due to the small amount of data available for training, the running time ranged from 0.15 (SVR) to 69.73 seconds (ANN). Despite both being relatively short periods, the processing time for the ANN model was approximately 465 times that of the SVR, 50 times the XGBoost and 19 times the GPR. This result means that the application of ANN to larger datasets may be impractical depending on the situation.

**Table 3.** Evaluation of the models developed from 4 ML algorithms: Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), Artificial Neural Networks (ANN), Gaussian Process Regression (GPR).

Model	Running time (s)	$R^2$	RMSE (MPa)	MAE (MPa)	Maximum absolute error (MPa)	Minimum absolute error (MPa)
XGBoost	1.40	0.83	3.41	2.24	18.78	0.00
SVR	0.15	0.79	3.73	2.26	19.90	0.00
ANN	69.73	0.82	3.40	2.26	23.79	0.01
GPR	3.67	0.82	3.43	1.96	21.12	0.00

The best model in this article obtained a lower  $R^2$  than that of other authors who used the same dataset put together by Yeh [13]. For example, Dao et al. [12] used GPR and ANN to obtain the compressive strength of concrete and reached  $R^2$  of 0.89 (against our 0.82 shown in Table 3). However, as opposed to the current work, these authors used the curing time as one of the features and evaluated all strength ranges. It means that they had access to a bigger dataset and their metrics were boosted by “easier” predictions (since the variability of the concrete strength at 3 and 7 days is usually much lower than that at 28 days). For comparison purposes, applying the complete dataset to our models would result in  $R^2$  ranging from 0.87 to 0.93.

Similarly, Mustapha and Mohamed [14] applied SVR to the Yeh [13] dataset, obtaining  $R^2$  up to 0.93 (versus 0.79 in this work). However, Mustapha and Mohamed [14] not only used the complete dataset (all ages and strengths) but also did not perform cross-validation to remove possible bias when splitting the data for training and testing.

It is also possible to compare the accuracy of our models with works in which the authors produced their own concrete specimens. For example, Lam et al. [48] produced 75 specimens to obtain the data used in their algorithms. They built an ANN-based model that obtained  $R^2 = 0.92$ , (versus  $R^2 = 0.82$  in the current work). However, this type of approach can limit the generalization ability of the model, as the algorithms learned from only one homogeneous source of concrete.

Regarding the other metrics, the XGBoost and ANN models obtained very similar RMSE and MAE results, around 3.40 MPa and 2.24 MPa, respectively. GPR obtained a lower MAE, 1.96 MPa, and a slightly higher RMSE, 3.43 MPa. As with the  $R^2$  results, the SVR presented the worst results, MAE of 2.26 MPa and RMSE of 3.73 MPa. It is noteworthy that the models proposed in the current work resulted in relatively close MAE and RMSE values. At a first glance, these results indicate a good performance of the models.

Comparatively, Dao et al. [12], mentioned above, obtained a RMSE of 5.46 MPa and a MAE of 3.86 MPa – while using the complete dataset, including compressive strengths higher than 50 MPa. In the same conditions, Mustapha and Mohamed [14] reached a MAE of 5.89 MPa. We can also mention Hoang et al. [11], who achieved a RMSE of 4.04 MPa, even though they created their own dataset of 246 specimens (ranging from 13.5 – 85.2 MPa).

It is important to remember that the RMSE is influenced by the square of the individual errors [44]. Thus, large errors are weighted more heavily than small ones. Therefore, this metric is recommended to evaluate models when large errors are particularly undesirable (such as in the prediction of concrete strength). However, Willmott and Matsuura [45] argue that the RMSE should not be used to compare two or more models, as this value varies according to the scale of the errors. The authors claim that the MAE is a metric that represents the magnitude of the error more naturally and, therefore, comparisons between different models should be based on the MAE.

Given the heterogeneous nature of cement-based composites and the infrastructure of construction sites, the calculation of the target mean strength of concrete is usually influenced by the quality control of its preparation. In Brazil, these parameters are set by NBR 12655 [49]. The smallest standard deviation value for the calculation of this strength, considering the best preparation conditions, normal-strength concrete, and no prior experiments, is 4.0 MPa [49]. Thus, both the RMSE and MAE values for all models were below the standard deviation indicated by NBR 12655. Important note: this comparison is not a measure of the safety of this mix design methodology, but it shows that the weighted average of errors obtained through the ML algorithms is smaller than the typical variability considered among specimens at a construction site.

Regarding individual errors, Figure 5 shows the frequency distribution of absolute errors (the difference between predicted and observed values) for all the mixtures in the dataset, regardless of direction. For all the models, at least 84% (275 instances) of the errors fell below 5MPa (for SVR), reaching 91% (300 instances) (for ANN). Conversely, for any algorithm, less than 3% of the predictions (10 instances) deviated more than 10 MPa from the real values. However, the maximum absolute error reached 18.78 – 21.12 MPa, which is a significant value.

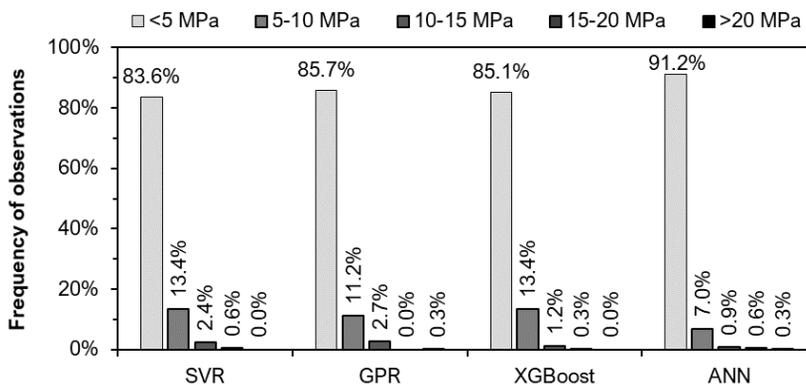


Figure 5. Frequency of errors (difference between predicted and observed values) for the 329 evaluated mixtures.

Seeking to understand the factors that led to these high singular errors, the authors assembled the 10 concrete mixtures that led the models to the biggest deviations, shown in Table 4. This table reveals that 3 observations are repeated in all models (being the top 3 errors of the XGBoost, ANN, and GPR); and another 3 are repeated in 3 models.

When analyzing the observations that presented the highest errors, one notices that they refer to concretes with unconventional proportions of materials. For example, mixture #1 of XGBoost (that was also #1 in SVR, ANN, and GPR), has only 200 kg/m<sup>3</sup> of Portland cement (and another 200 kg/m<sup>3</sup> of blast furnace slag), an unusual w/c ratio of 0.95, and still reached 49.25 MPa (versus an average of 27.7 MPa, predicted by the algorithms). Conversely, mixture #2 in XGBoost (that was #3 in GPR and ANN, and #7 in SVR) has a cement consumption of 436 kg/m<sup>3</sup>, w/c ratio of 0.5 and only reached 23.85 MPa (while the algorithms predicted approximately 38.9 MPa). The other observations that were repeated in the top 10 errors also showed mix proportions that are not commonly found in conventional concretes (e.g., over 30% of mineral admixtures in relation to cement mass).

**Table 4.** The 10 worst observations of each model regarding absolute error

Algori- thm	#	Portland Cement	Blast Furnace Slag	Fly Ash	Water	Super- plasti- cizer	Coarse Aggreg.	Fine Aggreg.	Observed Strength	Predicted Strength	Error
XGBoost	1 <sup>1</sup>	200.00	200.00	0.00	190.00	0.00	1145.00	660.00	49.25	30.47	18.78
	2 <sup>1</sup>	436.00	0.00	0.00	218.00	0.00	838.40	719.70	23.85	38.46	14.61
	3 <sup>1</sup>	277.10	0.00	97.40	160.60	11.80	973.90	875.60	48.28	35.19	13.09
	4	165.00	0.00	150.00	182.00	12.00	1023.00	729.00	18.03	29.09	11.06
	5 <sup>3</sup>	194.70	0.00	100.50	170.20	7.50	998.00	901.80	37.27	26.87	10.40
	6	212.00	0.00	124.80	159.00	7.80	1085.40	799.50	38.50	29.18	9.32
	7	165.00	128.50	132.10	175.10	8.10	1005.80	746.60	46.39	37.36	9.03
	8 <sup>2</sup>	249.10	0.00	98.80	158.10	12.80	987.80	889.00	30.85	39.85	9.00
	9 <sup>3</sup>	297.20	0.00	117.50	174.80	9.50	1022.80	753.50	47.40	38.52	8.88
	10	164.60	0.00	150.40	181.60	11.70	1023.30	728.90	18.03	26.75	8.72
SVR	1 <sup>1</sup>	200.00	200.00	0.00	190.00	0.00	1145.00	660.00	49.25	29.35	19.90
	2 <sup>1</sup>	277.10	0.00	97.40	160.60	11.80	973.90	875.60	48.28	31.46	16.82
	3 <sup>2</sup>	249.10	0.00	98.80	158.10	12.80	987.80	889.00	30.85	44.99	14.14
	4 <sup>2</sup>	381.40	0.00	0.00	185.70	0.00	1104.60	784.30	22.49	36.14	13.65
	5 <sup>3</sup>	144.00	15.00	195.00	176.00	6.00	1021.00	709.00	15.34	27.84	12.50
	6 <sup>2</sup>	385.00	0.00	0.00	186.00	0.00	966.00	763.00	31.35	43.57	12.22
	7 <sup>1</sup>	436.00	0.00	0.00	218.00	0.00	838.40	719.70	23.85	35.97	12.12
	8	446.00	24.00	79.00	162.00	11.60	967.00	712.00	44.42	33.46	10.96
	9	142.00	167.00	130.00	174.00	11.00	883.00	785.00	44.61	34.10	10.51
	10	141.90	166.60	129.70	173.50	10.90	882.60	785.30	44.61	34.20	10.41
ANN	1 <sup>1</sup>	200.00	200.00	0.00	190.00	0.00	1145.00	660.00	49.25	25.46	23.79
	2 <sup>1</sup>	277.10	0.00	97.40	160.60	11.80	973.90	875.60	48.28	29.52	18.76
	3 <sup>1</sup>	436.00	0.00	0.00	218.00	0.00	838.40	719.70	23.85	42.52	18.67
	4 <sup>3</sup>	168.00	42.10	163.80	121.80	5.70	1058.70	780.10	24.24	36.01	11.77
	5 <sup>2</sup>	385.00	0.00	0.00	186.00	0.00	966.00	763.00	31.35	41.89	10.54
	6	142.00	167.00	130.00	174.00	11.00	883.00	785.00	44.61	34.59	10.02
	7 <sup>2</sup>	381.40	0.00	0.00	185.70	0.00	1104.60	784.30	22.49	32.34	9.85
	8 <sup>3</sup>	297.20	0.00	117.50	174.80	9.50	1022.80	753.50	47.40	37.80	9.60
	9	151.00	0.00	184.00	167.00	12.00	991.00	772.00	15.57	24.89	9.32
	10	400.00	0.00	0.00	187.00	0.00	1025.00	745.00	43.70	34.50	9.20
GPR	1 <sup>1</sup>	200.00	200.00	0.00	190.00	0.00	1145.00	660.00	49.25	28.13	21.12
	2 <sup>1</sup>	277.10	0.00	97.40	160.60	11.80	973.90	875.60	48.28	33.34	14.94
	3 <sup>1</sup>	436.00	0.00	0.00	218.00	0.00	838.40	719.70	23.85	38.45	14.60
	4 <sup>2</sup>	249.10	0.00	98.80	158.10	12.80	987.80	889.00	30.85	42.14	11.29
	5 <sup>2</sup>	381.40	0.00	0.00	185.70	0.00	1104.60	784.30	22.49	33.73	11.24
	6 <sup>3</sup>	144.00	15.00	195.00	176.00	6.00	1021.00	709.00	15.34	26.24	10.90
	7 <sup>3</sup>	168.00	42.10	163.80	121.80	5.70	1058.70	780.10	24.24	34.51	10.27
	8 <sup>2</sup>	385.00	0.00	0.00	186.00	0.00	966.00	763.00	31.35	41.60	10.25
	9	162.00	207.00	172.00	216.00	10.00	822.00	638.00	39.84	29.69	10.15
	10 <sup>3</sup>	194.70	0.00	100.50	170.20	7.50	998.00	901.80	37.27	27.23	10.04

<sup>1</sup> Observations that appear in the top 10 worst predictions of all techniques. <sup>2</sup> Observations that appear in the top 10 worst predictions of 3 techniques. <sup>3</sup> Observations that appear in the top 10 worst predictions of 2 techniques.

Assuming that these results are not due to typing mistakes or experimental issues, they indicate:

- the relevance of the input data for the construction of models with good quality predictions, bearing in mind that the data must be like the problem studied.
- the necessity to collect multiple observations of concrete mixes of all types if one wants to create mix design tools that are as generalizable as possible.

### 3.2 Significance of each feature

Figure 6 shows the importance of each feature to the construction of the boosted decision trees within the model, obtained with the XGBoost technique. The more a feature is used to make key decisions during the construction of the model, the higher will be its relative significance. As expected, the cement has the greatest relative impact among the input features, while the aggregates had the smallest. Following the cement, we observed a significant influence of supplementary cementitious materials (mineral admixtures, such as blast furnace slag and fly ash). This is explained because all these binders have characteristics that significantly increase the strength of concrete [1] [50] [51]. These results indicate that the model had a good interpretation of the data.

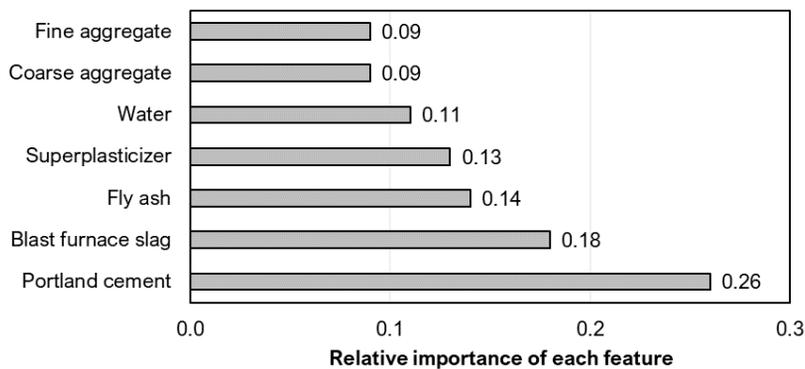


Figure 6. Relative importance of each input feature in predicting the compressive strength, according to the Extreme Gradient Boosting Decision Tree (XGBoost) model.

### 3.3 Considerations on model generalization

Table 5 presents the performance of the models trained with Yeh’s dataset [13], when validated with the new dataset with 22 instances elaborated by the authors. No model performed a good prediction, with the  $R^2$  falling from 0.79-0.83 (Table 3) to 0.37-0.59. MAE rose from 1.96-2.26 MPa to 3.04-4.04 MPa, and RMSE rose from 3.40-3.73 MPa to 3.75-4.67 MPa. This result demonstrates the low generalization ability of the models for the evaluation of new concrete mixes.

The characteristics of Yeh’s dataset [13] may explain this scenario. First, this dataset was built from relatively old studies (between 1987 and 1997), which is probably a major source of inaccuracies given the technological advancements of construction materials, especially Portland cement and chemical admixtures. Additionally, most of these works were carried out in Taiwan, using relatively homogeneous local materials, and coarse aggregates with a maximum size of 20mm. Thus, the dataset is incapable of representing the variability of concretes on a global scale. And this low generalization ability is even more worrisome because a significant portion of articles on the application of artificial intelligence for concrete mix design uses this dataset.

Table 5. Results of the model validation step carried out with the new dataset developed by the authors.

Model	$R^2$	RMSE	MAE	Maximum absolute error (MPa)	Minimum absolute error (MPa)
XGBoost	0.42	4.46	3.57	9.75	0.26
SVR	0.37	4.67	4.04	9.22	0.67
ANN	0.51	4.09	3.23	9.75	0.13
GPR	0.59	3.75	3.04	8.40	0.07

The regional peculiarities of concrete components are well known by professionals in this field. For example, even within Brazil, cements and concretes from the South region tend to adopt pozzolanic admixtures, while cements and concretes from the Southeast region commonly incorporate blast furnace slag [52]. However, despite this heterogeneity being empirically known, studies are still lacking to measure its impact on algorithms for concrete mix design.

To allow the development of safe, efficient, and economical mix design tools, the authors see two possibilities: 1) each country or region must work with its own dataset to generate models that are adapted to the local reality or 2) the creation of databases with more input features, such as country of origin, maximum aggregate size, type of cement, etc., thus allowing the creation of fewer mix design tools, but highly adaptable to different types of concrete.

## 4 CONCLUSION

This article compared four machine learning techniques to predict the compressive strength of conventional concrete specimens from their components. A well-known database, elaborated by Yeh [13], was used to train four models: Gaussian Process Regression (GPR), Extreme Gradient Boosting Decision Tree (XGBoost), Artificial Neural Networks (ANN), and Support Vector Regression (SVR). After evaluating these models, a new database was put together by the authors to validate them. This test sought to analyze the models' generalization ability to new concrete mixes.

In the first stage, the GPR, XGBoost, and ANN models obtained  $R^2 > 0.82$ , while SVR had the worst performance,  $R^2 = 0.79$ . For all algorithms, the MAE was below 2.26 MPa and the RMSE, below 3.73 MPa, which the authors consider relatively positive results compared to the minimal standard deviation prescribed in real mix design procedures. Although better correlations have been found in the literature, our work adopted a more conservative approach, looking only for resistance at 28 days.

To identify the causes of the inaccuracies in the proposed models, we ranked the top 10 mix proportions with the greatest deviations between the predicted and observed results. Most of them appeared in at least 3 algorithms, indicating that the issue was probably related to these particular mix proportions rather than with the proposed models. Indeed, the authors observed that these entries had unconventional percentages of admixtures front what is normally observed in conventional concretes. This result highlights the importance of the input data for the development of high-quality prediction models.

The relatively small number of observations meant that the running time was not significant to select the best algorithm. In this sense, a study analyzing how the dataset size would affect the processing time of the models should be carried out in the future.

In the validation step, the quality of the models dropped sharply, with the best  $R^2$  being only 0.59 (for the GPR model). The probable main contribution to this result was the difference between the characteristics of the dataset used for validation and the one used for model training. The models were created from the classic Yeh's dataset [13], which can be considered relatively homogeneous in terms of the origin of concrete observations and aggregate sizes.

This result shows that the regionalization and homogeneity of some datasets can lead to false-positive results in the search for universal concrete mix design strategies. In a future study, the authors intend to quantitatively assess this ability to generalize models. Furthermore, joint initiatives are needed to build a more comprehensive and varied database of concrete properties. Until that happens, the authors recommend that ML models for concrete mix design should be limited to predicting the strength of specimens from the same laboratories that trained them.

In summary, this article showed that ML techniques are potentially viable to predict the compressive strength of concrete. For now, more studies regarding the creation and validation of bigger and more varied databases are needed. However, soon, this approach may reduce the time and resources currently spent on the mix design processes.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the agencies CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico - Finance Code 304329/2019-3 for Alexandre Cury), FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais, project PPM-00001-18 for Alexandre Cury and APQ-01838-21 for Julia Mendes), PROPPI/UFOP (Research and Innovation Dean's Office, for Undergraduate Research Scholarship for Rafael Paixão and Research Assistance for Júlia Mendes), and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Finance Code 001), for master's scholarship to Rúben Penido) for providing financial support. We are also grateful for the collaboration of the Research Group on Data Science in Engineering (CIDENG - CNPq).

## 5 REFERENCES

- [1] A. M. Neville, *Properties of Concrete*, 5th ed., Bookman Editora, 2015.
- [2] BSI, *Testing Hardened Concrete Compressive Strength of Test Specimens*, BS EN 12390-3:2019, 2019.
- [3] ACI Committee 211, *Standard Practice for Selecting Proportions for Normal, Heavyweight, and Mass Concrete*, ACI PRC-211.1-91, 2002.
- [4] K. W. Day, J. Aldred, and B. Hudson, *Concrete Mix Design, Quality Control and Specification*, Boca Raton: CRC press, 2013.
- [5] A. L. Bonifácio, J. C. Mendes, M. C. Farage, F. S. Barbosa, C. B. Barbosa and A. L. Beaucour, "Application of Support Vector Machine and Finite Element Method to predict the mechanical properties of concrete," *Latin American Journal of Solids and Structures*, vol. 16, 2019.
- [6] B. F. Tutikian and M. Pacheco, "Self-compacting concretes (SCC): comparison of methods of dosage," *Rev. IBRACON Estrut. Mater.*, vol. 5, no. 4, pp. 500–529, 2012.
- [7] Z. M. Yaseen, M. T. Tran, S. Kim, B. T. and R. C. Deo, "Shear strength prediction of steel fiber reinforced concrete beam using hybrid intelligence models: a new approach," *Eng. Struct.*, vol. 177, pp. 244–255, 2018.
- [8] R. Pettres and L. A. de Lacerda, "Reconhecimento de padrões de defeitos em concreto a partir de imagens térmicas estacionárias e redes neurais artificiais," *Ágora: Revista de Divulgação Científica*, pp. 1–12, 2010.
- [9] V. Alves and A. Cury, "A fast and efficient feature extraction methodology for structural damage localization based on raw acceleration measurements," *Struct. Contr. Health Monit.*, vol. 28, no. 7, pp. e2748, 2021.
- [10] R. Almeida Cardoso, A. Cury, and F. Barbosa, "A clustering-based strategy for automated structural modal identification," *Struct. Health Monit.*, vol. 17, no. 2, pp. 201–217, 2018.
- [11] N. D. Hoang, A. D. Pham, Q. L. Nguyen and Q. N. Pham, "Estimating compressive strength of high performance concrete with Gaussian process regression model," *Advances in Civil Engineering*, vol. 2016, 2016.
- [12] D. V. Dao et al., "A sensitivity and robustness analysis of GPR and ANN for high-performance concrete compressive strength prediction using a Monte Carlo simulation," *Sustainability*, vol. 12, no. 3, pp. 830, 2020.
- [13] I. C. Yeh, "Modeling of strength of high-performance concrete using artificial neural networks," *Cement Concr. Res.*, vol. 28, no. 12, pp. 1797–1808, 1998.
- [14] R. Mustapha and E. A. Mohamed, "High-performance concrete compressive strength prediction based weighted support vector machines," *Int. J. Eng. Res. Appl.*, vol. 7, no. 1, pp. 68–75, 2017.
- [15] L. Cui, P. Chen, L. L. J. Wang, and H. Ling, "Application of extreme gradient boosting based on grey relation analysis for prediction of compressive strength of concrete," *Adv. Civ. Eng.*, 2021.
- [16] T. Chen and C. Guestrin "Xgboost: A scalable tree boosting system," in *Proc. 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, 2016.
- [17] S. O. Rezende, *Sistemas Inteligentes: Fundamentos e Aplicações*, Manole Ltda., 2003.
- [18] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants," *Mach. Learn.*, vol. 36, pp. 105–139, 1999.
- [19] M. Woźniak, M. Grana, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Inf. Fusion*, vol. 16, pp. 3–17, 2014.
- [20] Y. L. Suen, P. Melville, and R. J. Mooney "Combining bias and variance reduction techniques for regression trees," in *European Conference on Machine Learning*, pp. 741-749, October 2005.
- [21] D. Meyer and F. T. Wien "Support Vector Machines, The Interface to libsvm in Package," 2015, pp. 28.
- [22] R. P. Finotti, A. A. Cury, and F. D. S. Barbosa, "An SHM approach using machine learning and statistical indicators extracted from raw dynamic measurements," *Lat. Am. J. Solids Struct.*, 2019.
- [23] W. S. Noble, "What is a support vector machine," *Nat. Biotechnol.*, vol. 24, pp. 1565–1567, 2006.
- [24] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, pp. 199–222, 2004.
- [25] J. L. Garcia, *Fundamentos da Inteligência Artificial*, Rio de Janeiro: LTC, 2011.
- [26] R. M. Sadek et al., "Parkinson's Disease Prediction Using Artificial Neural Network," *Int. J. Acad. Health Med. Res.*, vol. 3, no. 1, pp. 1–8, 2019.
- [27] R. Mohammad, T. L. McCluskey, and F. A. Thabtah, "Predicting phishing websites using neural network trained with back-propagation," in *World Congress in Computer Science, Computer Engineering, and Applied Computing*, 2013.
- [28] J. M. Barreto "Introdução as redes neurais artificiais," UFSC, Florianópolis, 2002.
- [29] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*, pp. 63-71, February 2003.
- [30] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*, Cambridge: MIT Press, 2006.

- [31] I. H. D. Steinwart and C. Scovel, "An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4635–4643, 2006.
- [32] G. E. D. A. P. Batista, "Pré-processamento de dados em aprendizado de máquina supervisionado," PhD Thesis, Instituto de Ciências Matemáticas e de Computação - ICMC/USP, São Carlos, 2003.
- [33] T. Borovicka, et al., "Selecting representative data sets," in A. Karahoca, Ed., *Advances in data mining knowledge discovery and applications*, London: IntechOpen, pp. 43–70, 2012.
- [34] Associação Brasileira de Normas Técnicas, *Concrete for Structural Use - Density, Strength and Consistence Classification*, ABNT NBR 8953, 2015.
- [35] S. T. Yi, E. I. Yang, and J. C. Choi, "Effect of specimen sizes, specimen shapes, and placement directions on compressive strength of concrete," *Nucl. Eng. Des.*, vol. 236, no. 2, pp. 115–127, 2006.
- [36] P. L. D. C. Ferreira and N. Zincir-Heywood "Exploring feature normalization and temporal information for machine learning based insider threat detection.," in *15th Int. Conf. Network and Service Management*, October 2019, pp. 1–7.
- [37] M. S. Santos, J. P. Soares, P. H. Abreu, H. Araujo and J. Santos, "Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches," *IEEE Comput. Intel. Magaz.*, vol. 13, no. 4, pp. 59–76, 2018.
- [38] J. Shao, "Linear model selection by cross-validation," *J. Am. Stat. Assoc.*, vol. 88, no. 422, pp. 486–494, 1993.
- [39] H. Blockeel and J. Struyf, "Efficient algorithms for decision tree cross-validation," *J. Mach. Learn. Res.*, vol. 3, pp. 621–650, 2002.
- [40] D. C. Feng et al., "Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach," *Constr. Build. Mater.*, vol. 230, pp. 230, 2020.
- [41] C. Deepa, K. Sathiyakumari and V. P. Sudha, "Prediction of the compressive strength of high performance concrete mix using tree based modeling," *Int. J. Comput. Appl.*, pp. 18–24, 2010.
- [42] B. A. Young et al., "Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods," *Cement Concr. Res.*, vol. 115, no. 12, pp. 379–388, 2019.
- [43] A. C. Cameron and F. A. Windmeijer, "An R-squared measure of goodness of fit for some common nonlinear regression models," *J. Econom.*, vol. 77, no. 2, pp. 329–342, 1997.
- [44] M. A. DeRousseau, E. Laftchiev, J. R. Kasprzyk, B. Rajagopalan and W. V. Srubar III, "A comparison of machine learning methods for predicting the compressive strength of field-placed concrete," *Constr. Build. Mater.*, vol. 228, no. 6, pp. 116661, 2019.
- [45] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Clim. Res.*, vol. 30, no. 1, pp. 79–82, 2005.
- [46] Z. Zhou and G. Hooker, "Unbiased measurement of feature importance in tree-based methods," *ACM Transactions on Knowledge Discovery from Data*, vol. arXiv:1903.05179v2, pp. 1–21, 2021.
- [47] S. Tangirala, "Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, pp. 612–619, 2020.
- [48] N. T. M. Lam, D.-L. Nguyen and D.-H. Le, "Predicting compressive strength of roller-compacted concrete pavement containing steel slag aggregate and fly ash," *Int. J. Pavement Eng.*, pp. 1–14, 2020.
- [49] Associação Brasileira de Normas Técnicas, *Portland Cement Concrete - Preparation, Control, Receipt and Acceptance - Procedure*, ABNT NBR 12655, 2015.
- [50] P. Mehta and P. J. M. Monteiro, *Concrete: Microstructure, Properties, and Materials*, 2nd ed. São Paulo: IBRACON, 2014.
- [51] G. C. Cordeiro and J. M. Désir, "Potencial de argila caulínica de Campos dos Goytacazes, RJ, na produção de pozolana para concreto de alta resistência," *Cerâmica*, vol. 56, no. 337, pp. 71–76, 2010.
- [52] J. Natalli, E. C. S. Thomaz, J. C. Mendes and R. A. F. Peixoto, "A review on the evolution of Portland cement and chemical admixtures in Brazil," *Rev. IBRACON Estrut. Mater.*, vol. 14, no. 6, e14603, 2021.
- [53] ASTM International, *Standard Test Method for Compressive Strength of Cylindrical Concrete Specimens*, C39M-14, 2014.
- [54] China GB-National Standards, *Standard for Test Method of Mechanical Properties on Ordinary Concrete*, GB/T 50081-2002, 2002.
- [55] L. Andrade Nunes, R. Piazzaroli Finotti Amaral, F. D. Souza Barbosa, and A. Abrahão Cury, "A hybrid learning strategy for structural damage detection," *Struct. Health Monit.*, vol. 20, no. 4, pp. 2143–2160, 2021.
- [56] U. Atici, "Prediction of the strength of mineral admixture concrete using multivariable regression analysis and an artificial neural network," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 9609–9618, 2011.

---

**Author contributions:** RCFP: data curation, simulations, formal analysis, methodology, writing; REP: data curation, writing; AAC: methodology, writing, supervision; JCM: study conception; methodology, writing, supervision. All authors discussed the results and reviewed the manuscript.

**Editors:** Rebecca Gravina, Guilherme Aris Parsekian

## APPENDIX

The articles listed below were consulted for the creation of the dataset used to validate the models. Although they refer to steelmaking slag concrete, we only collect the data related to conventional aggregates (used as reference) for this work.

- Anastasiou, E., Filikas, K. G., & Stefanidou, M. (2014). Utilization of fine recycled aggregates in concrete with fly ash and steel slag. *Construction and Building Materials*, 50, 154-161.
- Andrade, H. D. (2018). Carbonatação em concreto de escória de aciaria. Universidade Federal de Ouro Preto (Dissertação de Mestrado).
- Lee, J. Y., Choi, J. S., Yuan, T. F., Yoon, Y. S., & Mitchell, D. (2019). Comparing properties of concrete containing electric arc furnace slag and granulated blast furnace slag. *Materials*, 12(9), 1371.
- Liu, S., Wang, Z., & Li, X. (2014). Long-term properties of concrete containing ground granulated blast furnace slag and steel slag. *Magazine of concrete Research*, 66(21), 1095-1103.
- Mengxiao, S., Qiang, W., & Zhikai, Z. (2015). Comparison of the properties between high-volume fly ash concrete and high-volume steel slag concrete under temperature matching curing condition. *Construction and Building Materials*, 98, 649-655.
- Miñano, I., Benito, F. J., Valcuende, M., Rodríguez, C., & Parra, C. J. (2019). Improvements in aggregate-paste interface by the hydration of steelmaking waste in concretes and mortars. *Materials*, 12(7), 1147.
- Pang, B., Zhou, Z., & Xu, H. (2015). Utilization of carbonated and granulated steel slag aggregate in concrete. *Construction and building materials*, 84, 454-467.
- Paula Stief, J. N., da Silva Maia, N., & Peixoto, R. A. F. (2011). Determinação experimental do módulo de elasticidade do concreto convencional e com agregados de escória de aciaria. *Educação & Tecnologia*, 14(2).
- Qasrawi, H., Shalabi, F., & Asi, I. (2009). Use of low CaO unprocessed steel slag in concrete as fine aggregate. *Construction and Building Materials*, 23(2), 1118-1125.
- Roslan, N. H., Ismail, M., Abdul-Majid, Z., Ghoreishiamiri, S., & Muhammad, B. (2016). Performance of steel slag and steel sludge in concrete. *Construction and building materials*, 104, 16-24.
- Souza, B. P., da Costa, E. C. P., de Carvalho, J. M. F., Peixoto, R. A. F., de Resende Mol, R. M., & Fontes, W. C. Caracterização físico-química de agregados de escória de aciaria LD pós-processada para concretos sustentáveis, 1-388.