XLIII CILAMCE
21-25, November, 2022
FOZ DO IGUAÇU - PR - BRAZIL

Comparison among four techniques to predict the compressive strength of concrete: Extreme Gradient Boosting, Support Vector Regression, Artificial Neural Networks, and Gaussian Process Regression

Rafael C. F. da Paixão[1], Rúben E. Penido[2], Vítor F. Mendes[2], Alexandre A. Cury[3], Júlia C. Mendes[2]

[1]*Dept. of Automation Control Engineering, Federal University of Ouro Preto*
Campus Morro do Cruzeiro, 35400-000, Ouro Preto, Minas Gerais, Brazil
rafael.paixao@aluno.ufop.edu.br
[2]*Graduate Program in Civil Engineering, Federal University of Ouro Preto*
Campus Morro do Cruzeiro, 35400-000, Ouro Preto, Minas Gerais, Brazil
ruben.penido@aluno.ufop.edu.br, vitor.mendes@aluno.ufop.edu.br, julia.mendes@ufop.edu.br
[3]*Dept. of Applied and Computational Mechanics, Federal University of Juiz de Fora*
Rua José de Lourenço Kelmer, 36036-900, São Pedro, Juiz de Fora, Minas Gerais, Brazil
alexandre.cury@engenharia.ufjf.br

Abstract. The compressive strength (Rc) of concrete is an important feature that influences the safety, durability, and cost of a structure. To achieve the desired Rc, professionals generally use mix design methods based on empirical tables. Then, the Rc must be confirmed in laboratory with tests that cost time and resources. To mitigate this issue, this study proposes and compares the use of four Machine Learning (ML) techniques to predict the Rc of concretes from their components. The techniques are: Extreme Gradient Boosting, Support Vector Regression, Artificial Neural Networks, and Gaussian Process Regression. Initially, a dataset vastly used in the literature for this purpose was used as input. Secondly, a dataset built by the authors was used to validate the models' generalization ability. All models were cross-validated (10-fold) and their accuracies were measured by R^2, MAE, and RMSE. XGBoost and GPR presented the best performance, while SVR presented the worst. Despite the positive performances measured in all models with the first dataset, the metrics dropped sharply in the validation step involving the second dataset. Thus, the ML techniques are promising tools for the mix design of concretes, but attention must be taken to guarantee that models are not overfitted because of the homogeneity of the input data.

Keywords: machine learning, concrete mix design, compressive strength.

1 Introduction

The bearing capacity, durability, and cost of any concrete structure depend on the quality of the concrete used. In this sense, the compressive strength of concrete, usually denoted by Rc, is one of its most important properties, defined with safety margins since the design process. In this sense, the engineer, having the required Rc value, portions the ingredients of the concrete in order to guarantee the desired strength. The traditional mix design methods (e.g., ACI, ABCO, IPT etc.) are based on empirical tables and formulations that must be confirmed in laboratory, with specimens at 28 days of curing (normally), in a procedure that follows local standards. This iterative process demands a high expenditure of human labour, natural resources, and time.

In this scenario, the Machine Learning (ML) techniques are a promising solution to efficiently predict the behaviour of concrete. ML models can predict future results based on patterns that are learned autonomously from a database of previous results. As an example, Hoang et al. [1] employed the Gaussian Process Regression (GPR) to predict the Rc of concrete using a dataset of 246 mixtures in Vietnam, obtaining a coefficient of determination (R^2) of 0,90. Various other studies [2, 3, 4] have used different ML models to predict the Rc of concretes using a well-known concrete mixture dataset assembled by Yeh [5]. Dao et al. [2] tested the accuracy of Artificial Neural Networks (ANN) and GPR to Yeh's dataset, in a 70/30 train/test simple separation, obtaining an R^2 of 0.89 with the GPR. Mustapha & Mohamed [4], also using Yeh's [5] dataset without cross-validation, obtained an R^2 of 0.93 by applying the Support Vector Regression (SVR). Cui et al. [3] used a decision tree model for this same dataset, obtaining R^2 above 0.80. All these authors agree on the potential of ML techniques to predict the Rc of concrete. However, most of the entries in Yeh's dataset originated from research carried out between 1987 and 1997 in Taiwan.

In this sense, the present work compares four extensively used ML techniques to predict the Rc of concrete: Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), Artificial Neural Networks (ANN), and Gaussian Process Regression (GPR). We also built a new dataset, with mixtures from all over the world, to validate the models trained with the traditional Yeh's dataset, seeking to interpret their generalization ability. We seek to find the most appropriate technique, among these four ones, to use in future predictions, and we aim to show the limitations of employing ML models to concrete mix design in diverse contexts.

2 Methodology

After a preliminary literature analysis, XGBoost, SVR, ANN, and GPR were found as the supervised ML techniques that showed the most promising performance when predicting concrete properties, while maintaining different learning-based backgrounds. After selecting these techniques, this study was split in two stages. In the first stage, the models were developed and cross-validated with Yeh's dataset [5] to extract the relationships from the concrete ingredients and predict their Rc (the target variable), expressed in MPa. The input features for each model were the proportions (expressed in kg/m³) of cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregates, and fine aggregates.

Yeh's dataset is called "Concrete Compressive Strength Data Set" and contains 1030 entries ranging from 2 to 82 MPa [5]. It is publicly available in the literature and is vastly used for prediction purposes. Beyond the previously cited features, this dataset has the age of the hydration of the concretes, which was not used in the present work. We only selected the observations with 28 days and between 15 - 50 MPa since these age and strength range are conventionally used for normal-strength concretes in buildings [6]. After these pre-processing steps, in total, 329 observations were used as input in the models (68% of the original size).

In the second step, these models were trained with the reduced Yeh's dataset and validated with another dataset built by the authors with 22 entries from 11 sources. This new database comprises mixtures between 15 - 50 MPa from 8 different countries, collected from articles published from 2009 to 2019. Since not all specimens had not the same geometry, they were converted to 150×300mm cylindrical specimens according to the correlations described by Yi et al. [8]. The overview of both datasets after pre-processing is shown in Table 1.

Table 1. Overview of the two pre-processed datasets with the Rc expressed in MPa and the other parameters in kg/m³

	Yeh's dataset				Validation dataset			
# Instances	329				22			
Parameter	Min	Max	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.
Portland cement	102.00	516.00	242.66	87.72	220.0	568.8	379.8	90.4
Blast furnace slag	0.00	359.40	89.68	89.50	0.0	410.5	81.0	112.7
Flay ash	0.00	200.10	64.73	65.82	0.0	25.0	4.5	9.9
Water	121.80	247.00	186.72	17.69	138.0	250.3	204.9	28.8
Superplasticizer	0.00	22.10	6.18	5.03	0.0	11.3	2.0	3.2
Coarse aggregate	801.00	1145.00	958.13	80.68	656.3	1029.0	879.2	126.9
Fine aggregate	594.00	945.00	763.81	70.91	477.7	1029.3	802.4	155.7
Rc	15.09	49.90	32.36	8.52	19.6	40.3	33.5	6.0

As shown in Table 1, the features have different scales, which can lead some models to prioritize a given input because of its higher value. Thus, we rescaled the input features, assuming the data is normally distributed within each feature and scaling the values so that the distribution centered around 0, with a deviation equal to 1. In the two stages, we used the k-fold cross-validation, with k=10, according to other studies with the similar problem [1].

The ML models were implemented in Python, using TensorFlow, Scikit-learn, and XGBoost libraries. We also used the Pandas library to assess and manipulate the database. The authors manually adjusted the hyperparameters of the techniques, avoiding focusing on specific optimization methods for each one. More information in these adjustments can be found in [7].

To measure these models' ability of prediction, the coefficient of determination (R^2), the mean absolute error (MAE), and the root mean square error (RMSE) metrics were used. These metrics is commonly used in regression models for this type of problem [9, 10, 11].

The R^2 ranges from $-\infty$ to 1. When the model has good accuracy, the R^2 tends to 1. Regarding MAE, it measures the average of the absolute deviations between the predicted and the observed results, i.e., the magnitude of the errors. Lastly, RMSE is similar to MAE, but in a quadratic approach. Therefore, when the error of a given prediction increases, the RMSE increases considerably. Both MAE and RMSE range from 0 to $+\infty$ and lower values indicate better predictions.

3 Results

Table 2 presents the results of the models in the two stages of this project.

Table 2. Metrics of the four implemented models: Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), Artificial Neural Networks (ANN), and Gaussian Process Regression (GPR).

	Stage 1: models trained and tested with Yeh's dataset					Stage 2: models trained with Yeh's dataset and validated with the new global dataset				
Model	R^2	RMSE (MPa)	MAE (MPa)	Max. abs. error (MPa)	Min. abs. error (MPa)	R^2	RMSE (MPa)	MAE (MPa)	Max. abs. error (MPa)	Min. abs. error (MPa)
XGBoost	0.83	3.41	2.24	18.78	0.00	0.42	4.46	3.57	9.75	0.26
SVR	0.79	3.73	2.26	19.90	0.00	0.37	4.67	4.04	9.22	0.67
ANN	0.82	3.40	2.26	23.79	0.01	0.51	4.09	3.23	9.75	0.13
GPR	0.82	3.43	1.96	21.12	0.00	0.59	3.75	3.04	8.40	0.07

3.1 Discussion of Stage 1: models trained and tested with Yeh's dataset

In the first stage, the R^2 of all models ranged around 0.80. This indicates a good correlation between the predicted and observed values, given the relatively small dataset and the knowingly complicated relationship among the concrete components. The XGBoost was the best technique, with $R^2 = 0.83$, with SVR presenting the worst performance $R^2 = 0.79$, although still close to the other models. Regarding MAE and RMSE metrics, the XGBoost and ANN models obtained very similar results, around 2.24 MPa and 3.40 MPa, respectively. As observed for the R^2, the SVR presented the worst results, MAE of 2.26 MPa and RMSE of 3.73 MPa.

In comparison with the literature, the best R^2 obtained in this study was lower than that of other authors that also used the Yeh's dataset. Using GPR and ANN tools, Dao et al. [2] reached $R^2 = 0.89$, for instance. Part of this reason relies on the fact that these authors did not pre-process the dataset. They used the entire dataset as inputs, with all strength values and curing ages, meaning that they had many more instances to train the algorithms. Additionally, it is known that the scale of the variation of Rc at early ages is usually much lower than that at 28 days. This can indicate that some ML models that test all ages have their metrics artificially boosted by lower deviations. Mustapha and Mohamed [4] obtained a higher R^2 (0,93) using SVR, but they also used all ages and did not cross-validate the model. It is well-known that simple separations into train/test groups can lead to bias in prediction problems.

In turn, Lam et al. [12] developed a model from their own experimentally built specimens. Using ANN, they achieved $R^2 = 0.92$. However, this approach can limit the generalization ability of the model since the algorithms only learn from a homogeneous source of concrete. Hoang et al. [1] also created their own dataset with 246 instances and achieved a RMSE of 4.04 MPa (higher than the values obtained by us).

Regarding the minimum absolute error obtained in the models, their values are equal to 0 or very close to it. This is expected to a good prediction model. However, the values of the maximum absolute results ranged close to 20 MPa, even with the dataset limited to 50 MPa, so care must be taken when using ML for mix design purposes. Analyzing the frequency of errors for the 329 instances, at least 83.6% of the those fell below 5 MPa (value regarding SVR), reaching up to 91.2% for ANN (Table 3). In all models, less than 3% of the predictions deviated more than 10 MPa from the real values.

Table 3. Frequency of errors deviation values for the Yeh's dataset

Error deviation	SVR	GRP	XGBoost	ANN
<5 MPa	83.6%	85.7%	85.1%	91.2%
5-10 MPa	13.4%	11.2%	13.4%	7.0%
>10 MPa	3.0%	3.0%	1.5%	1.8%

In a comparative analysis among the instances of the Yeh's dataset pre-processed, it was noted that the instances that lead to a high absolute error refers to the concrete with unconventional proportions of materials. For instance, there were some observations with only 200 kg/m³ of Portland cement reaching 49.25MPa and others with more than 30% of mineral admixtures in relation to cement mass. This fact indicates the need to have a vast observation set of concrete mixes of all type to create tools that are as generalizable as possible.

3.2 Stage 2: models trained with Yeh's dataset and validated with the new global dataset

The generalization ability of the models was tested using the Yeh's dataset pre-processed as a training input and the dataset built by us the validation one. Their accuracy obtained in the models is dropped sharply, as show in Table 2. The R^2 felt from about 0.80 in stage 1 to 0.36 - 0.59 in this new validation. Regarding RMSE and MAE, the values rose as well. These results indicate that the models were not able to extrapolate the correlations learned from the Yeh's dataset to new concrete mixes.

One of the reasons behind this poor generalization ability is that the Yeh's dataset [5] comprises relatively old studies (1987 – 1997), while our new dataset started from 2009. Over the last few years, the technology of construction materials has greatly improved, especially for Portland cement and chemical admixtures. Furthermore, the Yeh's datset mostly comprises studies carried out in Taiwan and with maximum coarse aggregate size of 20mm, which makes the dataset homogeneous in concern with the materials' properties. The regional specificities of concrete ingredients are significant to the Rc value. For instance, in Brazil, Portland cement made in the Southern region generally incorporates pozzolanic admixtures, meanwhile, in the Southeast region, blast furnace slag is more adopted [13].

In this scenario, the excellent results found in the literature with the application of ML for concrete mix design using Yeh's dataset or other homogeneous datasets can be misleading. Further studies are necessary to measure the impact of regional specificities on the ML models.

4 Conclusion

The present work developed and compared models with four different ML tools (XGBoost, SVR, ANN, and GPR) to predict the Rc of concretes. Initially, we trained the models and assessed their fitness through cross-validation (k-fold) using the well-known Yeh's dataset, available in the literature [5]. Next, we used the Yeh's dataset as training input and validated the models with a new set of concrete mixtures collected all over the world.

In the first stage, the models obtained promising results, with R^2 around 0.80. The XGBoost model presented the best performance, while the SVR presented the worst. Better correlations have been found by other authors, but they did not limit their datasets to 28 days nor used cross-validation.

In the second stage, the GPR was the best predictor of the Rc, while SVR was again the worst. The quality of the prediction from all models dropped significantly when they were validated with the new global dataset. This is due to the differences between the datasets. The training dataset (Yeh's) comprised concretes developed at least 10 years prior to the testing dataset, originated from the same country (in majority), and with small aggregate sizes.

The testing dataset had none of these limitations. As result, the correlations developed by the ML models from Yeh's dataset were not representative when applied to a much more heterogeneous set of mixtures.

We can conclude that the regionality and period of the dataset strongly influence ML models aimed at predicting the Rc. This fact must be factored in the search for a universal concrete mix design tool. Thus, we propose that more studies are carried out to understand and quantify this phenomenon.

Among the different ML models evaluated, XGBoost and GPR presented the best performance, while SVR presented the worst. This is not an ultimate result – the authors recommend that these techniques are applied to a larger and more varied dataset in order to verify if the dataset's particularities influence the prediction quality of each technique. In closing, this work showed that ML is a promising tool to predict the Rc of concrete, although, at the moment, it should be limited to concrete specimens with similar characteristics.

Acknowledgements. The present work was partially financially supported by FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais – Grant PPM-00001-18 for Alexandre Cury and Grant APQ-01838-21 for Júlia Mendes), PROPPI/UFOP (Research and Innovation Dean's Office for Undergraduate Research Scholarship for Rafael Paixão), and CAPES (Coordination of Superior Level Staff Improvement, for master's scholarship to Rúben Penido and Vítor Mendes). We gratefully acknowledge the collaboration of the Research Group on Data Science in Engineering (CIDENG – CNPq).

The datasets from this work are available at the repository https://github.com/cidengcnpq . The authors request that, should the dataset be used, the article [8] be cited.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors or has the permission of the owners to be included here.

References

[1] R. Pettres e L. A. de Lacerda, "Reconhecimento de padrões de defeitos em concreto a partir de imagens térmicas estacioárias e redes enurais artificiais". *Ágora: revista de divulgação científica*, pp. 1-12, 2010.
[2] 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.
[3] D. V. Dao, H. Adeli, H. B. Ly, L. M. Le, V. M. Le, T. T. Le, and B. T. Pham, "A sensitivity and robustness analysis of GPR and ANN for high-performance concrete compressive strength prediction using a Monte Carlo simulation". *Sustainability*, p. 830, 2020.
[4] 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". *Advances in Civil Engineering*, 2021.
[5] R. Mustapha and E. A. Mohamed, "High-performance concrete compressive strength prediction based weighted support vector machines". *International Journal of Engineering Research and Applications*, vol. 7, n. 1, pp. 68-75, 2017.
[6] I. C. Yeh, "Modeling of strength of high-performance concrete using artificial neural networks". *Cement and Concrete research*, pp. 1797-1808, 1998.
[7] ABNT, "NBR 12655 - Portland cement concrete - Preparation, control, receipt and acceptance - Procedure". Associação Brasileira de Normas Técnicas, Rio de Janeiro, 2015.
[8] S. T. Yi, E. I. Yang, and J. C. Choi, "Effect of specimen sizes, specimen shapes, and placement directions on compressive strength of concrete". *Nuclear Engineering and Design*, vol. 236, n. 2, pp. 115-127, 2006.
[9] C. Deepa, K. SathiyaKumari, and V. P. Sudha, "Prediction of the compressive strength of high performance concrete mix using tree based modeling". *International Journal of Computer Applications*, pp. 18-24, 2010.
[10] 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.
[11] B. A. Young, A. Hall, L. Pilon, P. Gupta, and G. Sant, "Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods". *Cement and Concrete Research*, pp. 379-388, 2019.
[12] 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". *International Journal of Pavement Engineering*, pp. 1-14, 2020.
[13] J. Natalli, E. Thomaz, J. Mendes, and R. Peixoto, "A review on the evolution of Portland cement and chemical admixtures in Brazil". *IBRACON Structures and Materials Journal*, pp. (accepted, in press), 2021.