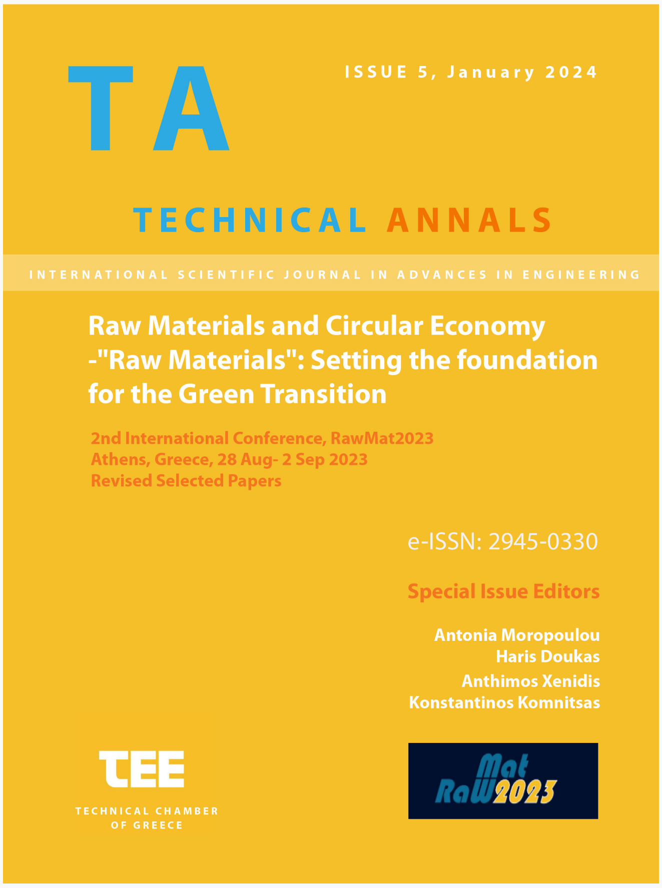
Technical Annals

Vol 1, No 5 (2024)

Technical Annals



The application of data science and machine learning techniques in predicting the compressive strength of confined concrete

Christos Papakonstantinou, Maria Valasaki, Filippos Sofos, Theodoros Karakasidis

doi: [10.12681/ta.37190](https://doi.org/10.12681/ta.37190)

The application of data science and machine learning techniques in predicting the compressive strength of confined concrete

C. G. Papakonstantinou [1[0000-1111-2222-3333]], M. Valasaki [1[0000-0002-2597-8027]]
F. Sofos [2[0000-0001-5036-2120]], T. Karakasidis [2[0000-0001-9580-0702]]

[1]Department of Civil Engineering, University of Thessaly, Volos, Greece
[2]Department of Physics, University of Thessaly, Volos, Greece
cpapak@uth.gr

Abstract. The integration of machine learning (ML) techniques into industrial and manufacturing applications has seen great growth in recent years. Various numerical and analytical models have been proposed, based either on experimental results or simulation results, and have helped to understand phenomena that take place during the life cycle of a material. In this direction, a large experimental data set to determine the compressive strength of fiber reinforced polymer (FRP) confined concrete specimens has been used as a basis in this work. The obtained measurements are correlated with the mechanical and structural properties of the material and fed into a ML model. The model is trained on the experimental values and can provide predictions for conditions within or outside the value range of the input data. Various ML algorithms are implemented and studied for their prediction accuracy, and the results show that ML can be an important computational tool, which can act as a complement to expensive experiments or time-consuming simulations in engineering sciences.

Keywords: Machine Learning, Data Science, Confined Concrete, Compressive Strength, FRP.

1 General

Property extraction in materials science has seen a significant shift in the last decade towards the adoption of techniques based on data science. The large amount of experimental data generated and stored in various databases has enabled scientists and engineers to incorporate this information into innovative statistical techniques and methods. This integration aims to propose new methodologies and materials with enhanced properties, which is especially crucial for technological and scientific advancement. The majority of data-driven methods are based on concepts from artificial intelligence (AI) and machine learning (ML).

The term ML refers to the implementation of a computational model of complex non-linear data-driven relationships and AI is the framework for decision-making and actions based on ML [1]. ML uses statistical approaches to analyze data with the help of appropriate algorithms and lead to predictions. The two main categories of ML are supervised and unsupervised learning. In supervised ML, the characteristics of the input data are known in advance, while in unsupervised ML, no information is available and the characteristics of the input data need to be searched [2].

Particularly in the field of engineering and construction, ML techniques have been used to predict concrete properties that affect quality measures such as the compressive strength [3,4]. The compressive strength of concrete can indeed be influenced by either the proportions of the constituent materials in the mix or by their mechanical properties. Both factors play a crucial role in determining the overall strength and durability of the concrete structure [5–7]. In this study, we investigate a database where the concrete compressive strength (f_{cc}) is modeled as a function of the mechanical properties of concrete using ML (machine learning) methods. The workflow, depicted in Figure 1, involves normalizing the experimental data and training an ML algorithm. By employing both linear and non-linear approximations, the model determines how input parameters map to the output variable, enabling predictions even when input information is scarce.

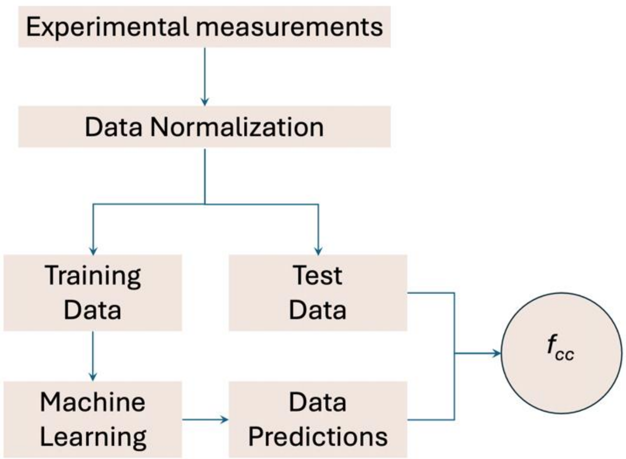


Fig. 1. Property prediction model with ML methods.

2 Data Handling

2.1 Database

The dataset comprises experimental records from various concrete specimens, which were confined using different FRP materials—including carbon, glass, and aramid. These specimens span a range of concrete strengths, from low to high. The practice of reinforcing concrete with composite materials has been shown to significantly enhance compressive strength [8] and is widely adopted in construction.

There are eight crucial input parameters related to the mechanical properties of the concrete specimens and the physical and mechanical characteristics of the composites. These parameters directly impact the confined compressive strength. For a detailed breakdown of each parameter, refer to Table 1.

2.2 Data Pre-processing

A total of 1476 experimental measurements are split into two subsets: one for training the ML model and the other for comparison. The division ratio is 80% for training and 20% for comparison. Before feeding the input data to the ML algorithm, it undergoes preprocessing. This preprocessing includes checking for empty or problematic records and applying a normalization step. The normalization aims to restrict the range of input values and is expressed by the following relation:

$$\bar{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

In summary, the data is carefully prepared before training the ML model, ensuring its quality and suitability for the algorithm.

Table 1. Description of 8 input parameters and output f_{cc}.

Parameter	Description
D	Concrete specimen diameter (mm)
H	Concrete specimen height (mm)
E_f	Fiber modulus of elasticity (commercial value)
e_{fu}	FRP ultimate axial strain at failure
f'_{co}	Unconfined concrete compressive strength (MPa)
f_{frp}	Ultimate axial FRP stress $f_{frp} = E_{frp}\,\varepsilon_{frp}$ (MPa)
t	Total FRP thickness (mm)
L	FRP number of layers
f_{cc}	Confined compressive concrete strength (MPa)

3 Machine Learning Algorithms

Linear regression, decision trees, and artificial neural networks stand out as some of the most extensively employed machine learning algorithms in the realms of science and technology. This study will scrutinize these three cases by inputting our data, evaluating their performance based on the coefficient R2, mean absolute error (MAE), and mean square error (MSE). The implementation leverages the Python library, sci-kit learn [9].

3.1 Multiple Linear Regression

Linear regression pertains to a method of analyzing the relationship between a dependent variable and a single independent variable. When dealing with multiple input variables, the model is denoted as multiple linear regression (MLR) [9] . In MLR we consider n independent input variables, linearly combined to derive the dependent variable Y as:

$$Y = \sum_{i=1}^{n} w_i X_i \tag{2}$$

Where $w_1, w_2, \ldots, w_n$ are the weight for each independent input $X_1, X_2, \ldots, X_n$. The method is shown graphically in Figure 2(a).

3.2 Decision Trees

The Decision Trees (DT) algorithm operates in the structure of a tree diagram, encompassing nodes, branches, and leaves (Figure 2(b)). Each node signifies a test conducted on a feature, and each branch represents the outcome of that test. The DT model's response follows the decisions made from the initial node to the terminal node (leaf). The feature space undergoes recursive partitioning based on the partition feature. A value is assigned to each terminal region for estimating the target output. Despite its ease of implementation, it's worth noting that input from other statistical methods may be required to prevent overfitting in the DT algorithm [11]. The use of decision trees can be advantageous compared to other algorithms, since decision trees are relatively robust to outliers, as the splitting criteria focus on dividing the data into homogenous sets. They are quite simple but powerful tools that model decisions and their potential outcomes. Techniques like pruning and ensemble methods can be used to address their limitations and increase their efficiency. Decision trees are adaptable, easy to understand, and user-friendly machine learning algorithms, that are effective for both classification and regression tasks.

3.3 Perceptron Algorithm

When arranged in multiple layers, such as input, output, and various internal hidden layers (Figure 2(c)), the conventional perceptron evolves into a neural network commonly known as a multi-layer perceptron (MLP). The determination of the number of hidden layers typically involves an iterative trial-and-error process. The data flow between neurons is contingent upon activation functions applied to each internal node and a weight function imposed on each input. These weights undergo adjustments to ensure that the predicted output closely resembles the expected output with minimal error. The training of the multi-layer perceptron (MLP) occurs iteratively, employing a backward calculation process. Despite being the favored choice in concrete property prediction applications, MLPs remain widely popular due to their effectiveness in minimizing error prediction through iterative adjustments [12,13], their multiparameter implementation can create difficulties in model convergence, trapping in local minima and overmodelling [14]. In this study, a neural network comprising three hidden layers, each consisting of 20 nodes (20,20,20), was employed.

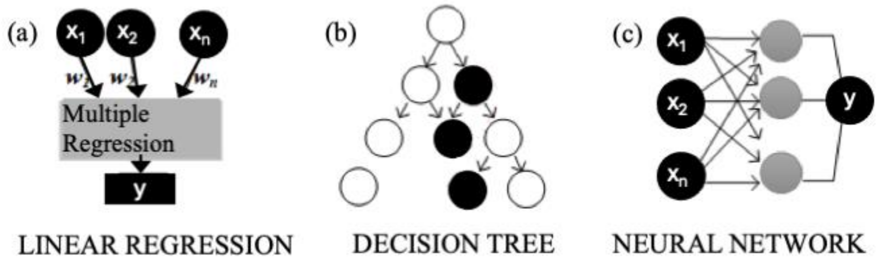
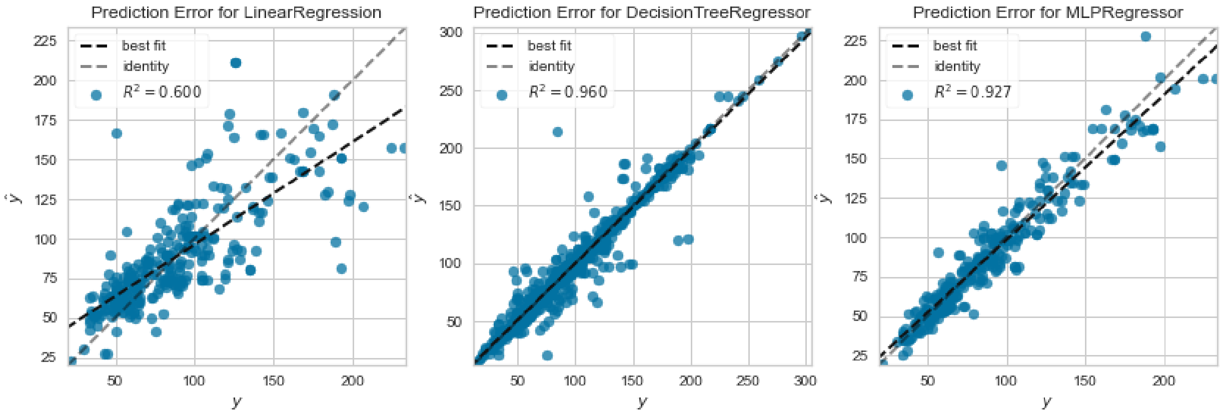


Fig. 2. ML algorithms, a) multiple linear regression, b) decision tree, c) neural network

4 Results and Discussion

Following the application of the three representative machine learning algorithms to predict the compressive strength behavior based on eight distinct mechanical and structural material characteristics, the regression results are presented in Figure 3. The accuracy of the predictions is contingent on the distance of the data points from the central 450 line.

The linear regression model (Figure 3a) demonstrates moderate performance, whereas the neural networks (Figure 3c) exhibit notably accurate results, achieving an R2 value of 0.927. The highest level of accuracy, however, is observed in the decision tree model, attaining an R2 value of 0.960. (Figure 3b). Additional measures of accuracy (mean absolute error and mean squared error) are given in Table 2. Based on these criteria, decision trees emerge as the optimal choice for handling this particular type of input data.[15].



(a) (b) (c)

Fig. 3. Regression results for the ML algorithms, a) multiple linear regression, b) decision tree, c) neural network, showing the fit of the experimental value (y) to the predicted value ($\hat{y}$).

Table 2. Accuracy measures for each ML method

Model	MAE	MSE
MLR	18.05	667.46
DT	9.33	233.48
MLP	10.27	236.63

5 Conclusions

Machine Learning Algorithms presented in this paper have shown that they are able to provide accurate predictions of compressive strength based on an eight-variable parameter space, even in the presence of significant noise in the observations, due to the experimental measurements. Additionally, we have demonstrated that machine learning algorithms based on decision trees exhibit strong performance when compared to conventional linear models. Due to their ease of implementation in contrast to intricate neural network architectures, decision tree-based models can be seamlessly integrated into comparable applications.

In conclusion, it becomes clear that statistical and machine learning techniques could potentially be applied to many real-life applications, overcoming the traditional route, which, most of the time, requires either expensive experiments or time-consuming, computationally intensive techniques. Along with knowledge of the physical problem, ML could contribute to the development of existing techniques and thus could be integrated into traditional scientific and technological methods.

References

1. Dimiduk DM, Holm EA, Niezgoda SR. Perspectives on the Impact of Machine Learning, Deep Learning, and Artificial Intelligence on Materials, Processes, and Structures Engineering. Integrating Materials and Manufacturing Innovation 2018; 7(3): 157–172. DOI: 10.1007/s40192-018-0117-8.
2. Wang T, Zhang C, Snoussi H, Zhang G. Machine Learning Approaches for Thermoelectric Materials Research. Advanced Functional Materials 2020; 30(5): 1906041. DOI: https://doi.org/10.1002/adfm.201906041.
3. Cook R, Lapeyre J, Ma H, Kumar A. Prediction of Compressive Strength of Concrete: Critical Comparison of Performance of a Hybrid Machine Learning Model with Standalone Models. Journal of Materials in Civil Engineering 2019; 31(11): 04019255. DOI: 10.1061/(ASCE)MT.1943-5533.0002902.
4. Sofos F, Papakonstantinou CG, Valasaki M, Karakasidis TE. Fiber-Reinforced Polymer Confined Concrete: Data-Driven Predictions of Compressive Strength Utilizing Machine Learning Techniques. Applied Sciences. 2023; 13(1):567. https://doi.org/10.3390/app13010567.
5. Young BA, Hall A, Pilon L, Gupta P, Sant G. 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 2019; 115: 379–388. DOI: 10.1016/j.cemconres.2018.09.006.
6. Asteris PG, Skentou AD, Bardhan A, Samui P, Pilakoutas K. Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. Cement and Concrete Research 2021; 145: 106449. DOI: 10.1016/j.cemconres.2021.106449.
7. Słoński M. A comparison of model selection methods for compressive strength prediction of high-performance concrete using neural networks. Computers & Structures 2010; 88(21): 1248–1253. DOI: 10.1016/j.compstruc.2010.07.003.
8. Papakonstantinou, C.G. Fiber Reinforced Polymer (FRP) Confined Circular Columns: Compressive Strength Assessment. Journal of Engineering Science & Technology Review 13.6, 2020.DOI: 10.25103/jestr.136.01
9. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011; 12(85): 2825–2830.
10. Sofos F, Karakasidis TE. Machine learning techniques for fluid flows at the nanoscale. Fluids 2021; 6(3). DOI: 10.3390/fluids6030096.
11. Schmidt J, Marques MRG, Botti S, Marques MAL. Recent advances and applications of machine learning in solid-state materials science. Npj Computational Materials 2019; 5(1): 1–36. DOI: 10.1038/s41524-019-0221-0.
12. Ikumi T, Galeote E, Pujadas P, de la Fuente A, López-Carreño RD. Neural network-aided prediction of post-cracking tensile strength of fibre-reinforced concrete. Computers & Structures 2021; 256: 106640. DOI: 10.1016/j.compstruc.2021.106640.
13. Roberson MM, Inman KM, Carey AS, Howard IL, Shannon J. Probabilistic neural networks that predict compressive strength of high strength concrete in mass placements using thermal history. Computers & Structures 2022; 259: 106707. DOI: 10.1016/j.compstruc.2021.106707.
14. Nguyen H, Vu T, Vo TP, Thai HT. Efficient machine learning models for prediction of concrete strengths. Construction and Building Materials 2021; 266: 120950. DOI: 10.1016/j.conbuildmat.2020.120950.

15. Sofos F, Stavrogiannis C, Exarchou-Kouveli KK, Akabua D, Charilas G, Karakasidis TE. Current Trends in Fluid Research in the Era of Artificial Intelligence: A Review. Fluids 2022; 7(3): 116. DOI: 10.3390/fluids7030116.