# Comparative study of soft computing techniques for the prediction of concrete strength containing waste material

# Comparative study of soft computing techniques for the prediction of concrete strength containing waste material

Ayaz Ahmad [1,2], William Finnegan[1,2], Yadong Jiang[1,2], Jamie Goggins[1,2,3]

[1] Civil Engineering, School of Engineering, College of Science & Engineering, National University of Ireland Galway, University Road, Galway, Ireland.
[2] MaREI Centre, Ryan Institute, National University of Ireland Galway, University Road, Galway, Ireland.
[3] ERBE Centre for Doctoral Training, National University of Ireland Galway, University Road, Galway, Ireland.

Email : a.ahmad8@nuigalway.ie ; william.finnegan@nuigalway.ie; yadong.jiang@nuigalway.ie; jamie.goggins@nuigalway.ie;

ABSTRACT: The application of artificial intelligence algorithms to anticipate the strength properties of various types of concrete is increasing in prominence. This study describes the use of two artificial intelligence algorithms, such as decision tree (DT) and bagging algorithm (BA), to anticipate the compressive strength of concrete containing fly ash. Python instructions were executed on the appropriate models using the anaconda navigator software. The models were conducted with seven input variables (cement, water, fly ash, superplasticizers, coarse aggregate, fine aggregate, and age) and one output parameter (i.e. compressive strength). Results show that the precision level of the BA towards the prediction of concrete's strength is high compared to the DT model. The said accuracy is indicated by the coefficient of determination value, which equals 0.93 for the BA and 0.86 for the DT model. The statistical checks also verified the accuracy level of the employed algorithms. The low values of the mean absolute error and root mean square error also confirm high accuracy for BA compared to DT.

KEYWORDS: concrete, fly ash, decision tree, bagging, modelling.

## 1. INTRODUCTION

Globally, demand for construction materials is expanding tremendously as infrastructure develops at a breakneck pace [1]. However, in the current context of sustainable development, the durability of building materials is not the only factor to consider. Additionally, economic and environmental considerations are becoming increasingly relevant. Portland cement, a commonly used construction material, is well-known for its high energy consumption [2], [3]. Numerous international initiatives and conferences aimed at reducing greenhouse gas emissions have spurred the adoption of supplementary cement-based materials (SCMs) as binders in place of ordinary cement [4], [5]. The application of fly ash in concrete as a partial replacement for cement is gaining popularity since it not only reduces the emission of carbon dioxide from the production and use of cement but also satisfies the strength requirements of concrete [6], [7]. Mehta et al. [8] used two types of waste materials (ground granulated blast furnace slag and fly ash) in concrete to investigate the numerous properties. A 52% increase in compressive strength was reported during the experimental program. The study of Stolz et al. [9] was also based on the several properties of activated fly ash (cellular alkali) concrete. Significant changes have been reported with various mixes.

The use of machine learning (ML) approaches in the civil engineering field is one of the popular trends in investigating and predicting the performance of materials [10], [11], [12]. This is especially true when predicting the mechanical properties of concrete since it takes time for concrete to reach the desired strength [13], [14], [15], [16]. The number of ML algorithms to resolve the regression and classification problems that are commonly used to anticipate the strength of concrete [17], [18]. Shahmansouri et al. [19] introduced the GEP model to investigate the CS of the concrete material, which shows an appreciable result in terms of prediction. Akande et al. [20] employed the neuron-based model (ANN) and SVM to anticipate and compare the compressive strength (CS) of selected concrete. The result shows that the SVM performed better than the ANN technique in the prediction of the required result.

This research explains the application of two ML algorithms, decision tree (DT) and bagging regressor (BR), to foretell the same property of the concrete having waste material (fly ash). A comparative study of the individual (DT) and ensemble (BR) ML approaches was conducted to investigate their performance. The accuracy level of the BR was better than the DT, as illustrated by the coefficient of determination ($R^2$) value. The $R^2$ value for BR was reported as 0.92, while DT gave the value of 0.86. This research aims to apply and compare the various ML approaches to forecast the strength of concrete. The study's main objective is to compare and recommend the highly precise ML algorithm that can be successfully employed to predict the mechanical properties of any type of concrete.

## 2. DATA DESCRIPTION

The Anaconda navigator software utilises Python codding to run the appropriate models. The database having input parameters and required output was used in the software for predicting the outcome. A total of 570 data points were retrieved from the literature with seven inputs (cement, water, superplasticizers, fly ash, fine aggregate, coarse aggregate, and a number of days) and one output (strength) parameter. The statistical descriptive explanation (analysis) for all variables is shown in table 1.

Table 1: Input variables used in the present study that were retrieved from the literature (cement, water, superplasticizers, fly ash, fine aggregate, coarse aggregate, and number of days)

| Statistics | Cement ($kg/m^3$) | Fly ash ($kg/m^3$) | Water ($kg/m^3$) | SP* | CA* | FA* | Age (days) |
|---|---|---|---|---|---|---|---|
| Mean values | 288.31 | 73.14 | 181.33 | 5.24 | 1003.73 | 793.41 | 45.42 |
| Standard Error result | 4.06 | 2.66 | 0.76 | 0.23 | 3.09 | 2.93 | 2.56 |
| Median data | 275.00 | 98.80 | 185.70 | 5.50 | 1006 | 792.50 | 28 |
| Mode values | 213.50 | 0.00 | 192 | 0.00 | 968 | 613.00 | 28 |
| Standard Deviation | 96.86 | 63.54 | 18.06 | 5.42 | 73.69 | 69.85 | 61.04 |
| Sample Variance | 9381.53 | 4037.84 | 326.29 | 29.39 | 5430.84 | 4878.38 | 3726.0 |
| Variable's range | 405.30 | 200.10 | 88.00 | 28.20 | 324 | 351.00 | 364.0 |
| Low values | 134.70 | 0.00 | 140.00 | 0.00 | 801 | 594.00 | 1.0 |
| High values | 540.0 | 200.1 | 228.0 | 28.2 | 1125 | 945 | 365 |
| Total result | 164047.4 | 41618.3 | 103177.2 | 2982.7 | 571123.9 | 451449.4 | 25845.0 |
| Count | 569.00 | 569.00 | 569.00 | 569.00 | 569.00 | 569.00 | 569.00 |

Superplasticizers ($kg/m^3$); *CA = Coarse aggregate ($kg/m^3$); *FA = Fine aggregate ($kg/m^3$).

## 3. METHODOLOGY

Both the ensemble and individual model strategies are introduced to anticipate the properties of materials in a short period of time. The $R^2$ value (which varies from 0–0.99) is commonly used to determine the amount of accuracy between the actual and predicted levels. A high $R^2$ value suggests that the chosen technique produces satisfactory results. The CS of concrete containing waste material is predicted by employing two types of ML approaches in the study.

### 3.1. DECISION TREE ALGORITHM

This approach is a supervised technique of learning that is normally introduced for classification plus problems related to regression [21]. This method uses a classifier with a tree structure, whose inner nodes describe the database properties. The conclusion rules are described by the branches, while the outcome is shown by each leaf node. The essential decision-making nodes are a decision node and a leaf node. Decision nodes have a number of branches and can make any tentative decision, whereas leaf nodes operate without branching and are regarded as the output of the decisions. It's known as a DT because it looks like a tree whose root starts from a node and grows into a bunch of branches [22]. At each location, the program identifies the variation between the actual and anticipated value. At each splatted point, the errors are evaluated, and the parameter with the lowest fitness function result is selected as the split point. The process is then repeated.

### 3.2. BAGGING ALGORITHM

The ensemble technique is an artificial intelligence (AI) paradigm that uses a similar learning algorithm to train several models [23]. Many algorithms, called multi-classifiers, are used in the ensemble. Thousands of learners are brought together with a single goal in mind to solve the problem. Bagging is an alternative ensemble approach that generates additional data during the training phase to explain the prediction model's variance. This is a result of sampling with irregularity. This includes data substitutions from the genuine set. Some of the values can be repeated in an additional training data set by making samples with replacement. All the components have the same opportunity for representation in the other dataset when bagging. Increasing the size of the training set will not increase the predictive power. With the fine-tuning of the prediction to an expected outcome, the variation can be minimized even further. All these sources of data are typically utilized to train many models. The approximate average of all the forecasts from the number of models is used in this ensemble of models. In regression, the prediction can be the mean or average of the forecasts from several models. [24].

## 4. K-FOLD CROSS-VALIDATION (C-V) AND STATISTICAL MEASURES

K-fold C-V is employed to check the model's performance in terms of bias and variance. The data is separated into ten stratified groups, each of which is randomly assigned to a training and test set. As shown in Figure 5, this procedure divides the total data into two halves, one for the test data and the other for the training data. The precision and efficiency of the selected model are then verified via C-V by averaging 10 rounds of varying mistakes. Likewise, statistical indicators are employed to assess the model [25]. In our current study, we use two different sorts of indicators, which are given below, mean absolute error (MAE) and root mean square error (RMSE). (Equation 1-2)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |x_i - x| \qquad 1$$

$$\mathrm{RMSE} = \sqrt{\sum \frac{\left(y_{pred} - y_{ref}\right)^2}{n}} \qquad 2$$

## 5. RESULT AND DISCUSSION

### 5.1. DECISION TREE

The DT model gives a strong relationship between the experimental result and the result obtained from the model. The better result can be observed from the coefficient of determination ($R^2$) values equalling 0.86 as depicted in Figure 1. However, the dispersals of the error results for the DT model

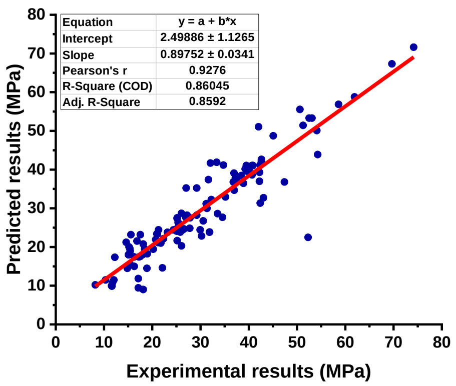



Figure 1. A comparison of experimental and anticipated DT model output.

are shown in Figure 2. The variation for DT models gives a maximum value of 29.81 MPa, and the average value is equal to 3.17 MPa. It was also noted that the 34.21 % of the error's data were lie between 0 to 1 MPa, and 42.10 % of the data were lie between 1 and 5 MPa. However, 23.68% of the error data was located above 5 MPa.

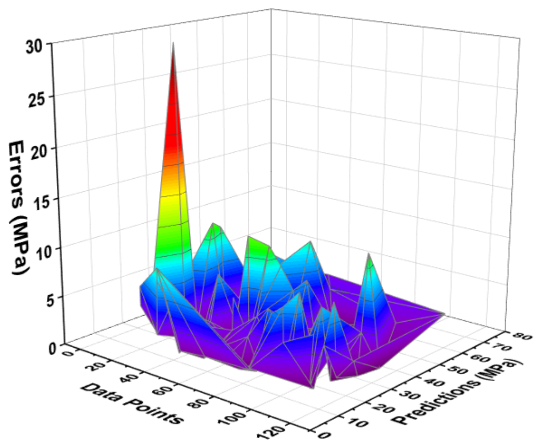



Figure 2. Representation of error's distribution for DT model

### 5.2. BAGGING ALGORITHM

The bagging model was reported to be more accurate as opposed to the DT model while investigating the CS of the selected concrete. The better performance of the said model was indicated by the better relationship between the actual and forecasted output, as depicted in Figure 3, while the dispersion of its errors is shown in Figure 4. The distribution gives the high, lower, and average values of 14.93 MPa, 0.0015 MPa, and 2.60 MPa, respectively. 32.45% of the error data lies between 0 MPa and 1 MPa, and 56.14% of the data lies between 1 MPa and 5 MPa. However, only 11.40% of the error's data were lying above 5 MPa 1 MPa, and 5 MPa. However, only 11.40% of the error's data were lie above 5 MPa.

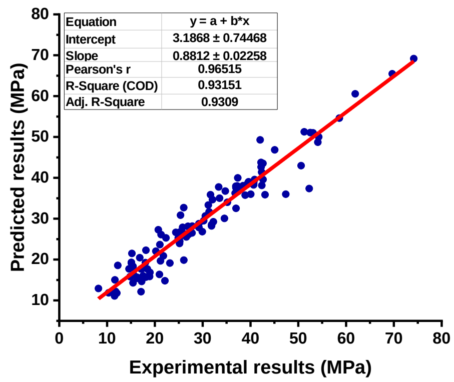
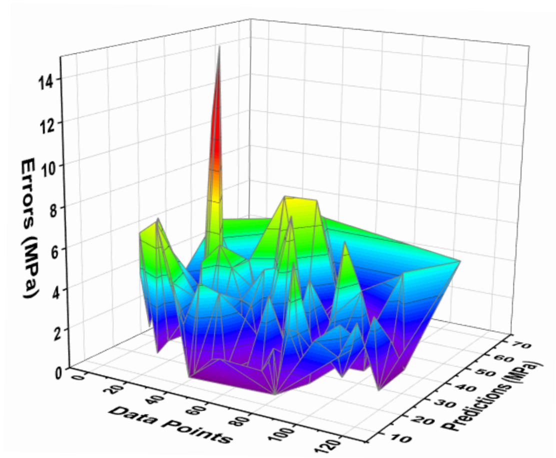



Figure 3. A comparison of experimental and anticipated bagging model output




Representation of error's distribution for bagging model

## 6. CONCLUSIONS

The implementation of machine learning algorithms to predict the CS of concrete, including waste material, is described in this research. The two ML techniques (DT and bagging algorithm) were introduced for forecasting the CS of flay ash-based concrete. The performance comparison was carried out on the precision level of both the employed models. The following conclusion can be drawn from the study:

The bagging model shows a better relationship between the actual and predicted outcome and shows an impressive result as indicated by the coefficient of determination value equal to 0.92.The DT model predicted result for CS of fly ash-based concrete was also in the acceptable range by giving an $R^2$ value of 0.86.

To summarize, both the employed show a significant effect on the predicted result. To achieve more accuracy, it is necessary to enhance the data set through experiments. However, other ensemble ML approaches like boosting, Adaboost, and random forest can also give better results in terms of prediction. Moreover, it is clear that the ML approaches can be successfully employed for predicting different types of strength properties such as compressive strength, flexural strength, and splitting tensile strength. This successful adoption will not only reduces the physical effort on the experimental work in the laboratory but also minimize the cost and time of the construction project.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the support of Science Foundation Ireland (SFI) through the MaREI Research Centre for Energy, Climate and Marine (Grant no. 12/RC/2302_2) and the Sustainable Energy Authority of Ireland (SEAI) through the WindLEDeRR Project (Grant no. 21/RDD/601).

## REFERENCES

[1] G.J. Parra-Montesinos, High-performance fiber-reinforced cement composites: an alternative for seismic design of structures, ACI Structural Journal 102(5) (2005) 668.
[2] C. Shi, A.F. Jiménez, A. Palomo, c. research, New cements for the 21st century: The pursuit of an alternative to Portland cement, Cement 41(7) (2011) 750-763.
[3] C. Dow, F.P. Glasser, Calcium carbonate efflorescence on Portland cement and building materials, Cement Concrete Research 33(1) (2003) 147-154.
[4] U. Environment, K.L. Scrivener, V.M. John, E.M. Gartner, Eco-efficient cements: Potential economically viable solutions for a low-CO2 cement-based materials industry, Cement Concrete Research 114 (2018) 2-26.
[5] Á. Fernández, J.G. Calvo, M.C. Alonso, Ordinary Portland Cement composition for the optimization of the synergies of supplementary cementitious materials of ternary binders in hydration processes, Cement Concrete Composites 89 (2018) 238-250.
[6] G.N.K. Reddy, G.H. Vardhan, S.V.B. Reddy, Partial replacement of cement in concrete with sugarcane bagasse ash and its behaviour in aggressive environments, IOSR Journal of Mechanical Civil Engineering e-ISSN: 1684 (2016).
[7] F.A. Olutoge, O.M. Okeyinka, O.S. Olaniyan, Assessment of the suitability of periwinkle shell ash (PSA) as partial replacement for ordinary Portland cement (OPC) in concrete, International Journal of Research Reviews in Applied Sciences 10(3) (2012) 428-434.
[8] A. Mehta, R. Siddique, T. Ozbakkaloglu, F.U.A. Shaikh, R. Belarbi, Fly ash and ground granulated blast furnace slag-based alkali-activated concrete: Mechanical, transport and microstructural properties, Construction Building Materials 257 (2020) 119548.
[9] J. Stolz, Y. Boluk, V. Bindiganavile, Mechanical, thermal and acoustic properties of cellular alkali activated fly ash concrete, Cement Concrete Composites 94 (2018) 24-32.
[10] S.-C. Lee, Prediction of concrete strength using artificial neural networks, Engineering structures 25(7) (2003) 849-857.
[11] A. Ahmad, K. Chaiyasarn, F. Farooq, W. Ahmad, S. Suparp, F. Aslam, Compressive strength prediction via gene expression programming (GEP) and artificial neural network (ANN) for concrete containing RCA, Buildings 11(8) (2021) 324.
[12] M. Shariati, M.S. Mafipour, P. Mehrabi, M. Ahmadi, K. Wakil, N.T. Trung, A. Toghroli, Prediction of concrete strength in presence of furnace slag and fly ash using Hybrid ANN-GA (Artificial Neural Network-Genetic Algorithm), Smart Structures Systems 25(2) (2020) 183-195.
[13] Z. Keshavarz, H. Torkian, Application of ANN and ANFIS models in determining compressive strength of concrete, Journal of Soft Computing in Civil Engineering 2(1) (2018) 62-70.
[14] H. Naderpour, A.H. Rafiean, P. Fakharian, Compressive strength prediction of environmentally friendly concrete using artificial neural networks, Journal of Building Engineering 16 (2018) 213-219.
[15] J.-S. Chou, C.-F. Tsai, A.-D. Pham, Y.-H. Lu, Machine learning in concrete strength simulations: Multi-nation data analytics, Construction Building Materials 73 (2014) 771-780.
[16] W.Z. Taffese, E. Sistonen, Machine learning for durability and service-life assessment of reinforced concrete structures: Recent advances and future directions, Automation in Construction 77 (2017) 1-14.
[17] P. Chopra, R.K. Sharma, M. Kumar, T. Chopra, Comparison of machine learning techniques for the prediction of compressive strength of concrete, Advances in Civil Engineering 2018 (2018).
[18] F. Farooq, M. Nasir Amin, K. Khan, M. Rehan Sadiq, M. Faisal Javed, F. Aslam, R. Alyousef, A comparative study of random forest and genetic engineering programming for the prediction of compressive strength of high strength concrete (HSC), Applied Sciences 10(20) (2020) 7330.
[19] A.A. Shahmansouri, H.A. Bengar, E. Jahani, Predicting compressive strength and electrical resistivity of eco-friendly concrete containing natural zeolite via GEP algorithm, Construction Building Materials 229 (2019) 116883.
[20] K.O. Akande, T.O. Owolabi, S. Twaha, S.O. Olatunji, Performance comparison of SVM and ANN in predicting compressive strength of concrete, IOSR Journal of Computer Engineering 16(5) (2014) 88-94.
[21] A. Karbassi, B. Mohebi, S. Rezaee, P. Lestuzzi, Damage prediction for regular reinforced concrete buildings using the decision tree algorithm, Computers Structures 130 (2014) 46-56.
[22] W.B. Chaabene, M. Flah, M.L. Nehdi, Machine learning prediction of mechanical properties of concrete: Critical review, Construction Building Materials 260 (2020) 119889.
[23] M.H.D.M. Ribeiro, L. dos Santos Coelho, Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series, Applied Soft Computing 86 (2020) 105837.
[24] J. Dou, A.P. Yunus, D.T. Bui, A. Merghadi, M. Sahana, Z. Zhu, C.-W. Chen, Z. Han, B.T. Pham, Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan, Landslides 17(3) (2020) 641-658.
[25] F. Aslam, F. Farooq, M.N. Amin, K. Khan, A. Waheed, A. Akbar, M.F. Javed, R. Alyousef, H. Alabdulijabbar, Applications of gene expression programming for estimating compressive strength of high-strength concrete, Advances in Civil Engineering 2020 (2020).