Experimental and Data-Driven Investigation of the Influence of Coarse Aggregate Type on the Compressive Strength of Concrete

Muhammad Sufian[1,a,*], Xin Wang[1,b,*], Mohamed F.M Fahmy[1,c], Zhishen Wu[1,d], Muhammad Rahman[1,e], Amr M.A Moussa[1,f]

[1]Key Laboratory of C & PC Structures Ministry of Education, National and Local Unified Engineering Research Center for Basalt Fiber Production and Application Technology, Southeast University, Nanjing, Jiangsu, China

[a]drsufian@seu.edu.cn, [b]xinwang@seu.edu.cn, [c]m.fahmy@aun.edu.eg, [d]zswu@seu.edu.cn, [e]rahman.m@seu.edu.cn, [f]amr.moussa@eng.svu.edu.eg

*Corresponding author

Keywords: Coarse Aggregate, Basalt, Mix Design Code, Compressive Strength, Machine Learning

Abstract: This study investigates the influence of coarse aggregate (CA) type and shape on the compressive strength of concrete. Two types of coarse aggregates were used, i.e., basalt crushed and naturally rounded of 15 mm size. Eight concrete mixtures were designed according to four distinct concrete mix design (CMD) codes. A total of 64 concrete samples were prepared and tested under compression. The results show that concrete made with basalt CA achieved significantly higher CS (by 7% to 39%) compared to concrete made with natural CA, across different CMD codes. The incorporation of basalt CA enhanced the toughness and ductility of concrete, making it a better option for normal and medium-strength concrete structures. In addition to the experimental program, two ensembled machine learning (ML) models, i.e., extreme gradient boosting (XGB), and random forest (RF) were employed to forecast the CS of concrete. RF and XGB showed remarkable accuracy as evident by R^2= 0.93 and R^2= 0.92, respectively, and low error matrices. Moreover, the feature importance analysis identified cement content and CA type as the primary determinants of strength, while the water-cement ratio served as a crucial regulator.

1. Introduction

The impact of CA characteristics on the compression behavior of concrete is critical and needs extensive investigation. In the construction industry, the diverse utilization of CA is essential for enhancing the load-bearing capability and extending structural longevity. Coarse aggregates constitute 60%-80% of the concrete volume, making their role in concrete mix fundamental to mechanical properties [1]. Research has indicated that while CA size slightly affects the CS of concrete [2], the type and shape of CA significantly influence the concrete CS and energy dissipation [3–5]. Some studies have revealed that concrete with basalt CA provides superior CS and fracture energy compared with granite, limestone, and river gravel [6–8]. Basalt CA has been recognised to enhance the strength, permeability, and thermal conductivity of concrete [9]. These findings emphasize the importance of CA characteristics in optimizing concrete mechanical performance, providing a baseline for further research in this area.

Modern infrastructure development relies heavily on concrete. In addition to guaranteeing structural integrity, precise predictions of its mechanical characteristics also help to optimize material use, decline environmental impact, and increase the effectiveness of CMD [10]. ML models, especially ensemble approaches such as random forest (RF) and gradient boosting models have become more popular for their enhanced generalizability and accuracy, making them ideal for handling complicated data and producing reliable predictions [11]. Some common ML techniques employed to forecast the mechanical properties of concrete include RF, support vector machine, gradient boosting, and artificial neural networks [12-15]. These techniques have shown significant effectiveness in enhancing predictive accuracy in the prediction of concrete mechanical properties.

DOI: 10.25236/iwmecs.2025.055

The advancement of ML techniques has significantly improved our understanding of concrete's mechanical behavior, enabling more accurate and efficient predictions for modern construction, including optimizing CMD, enhancing environmental sustainability, and project efficiency.

In summary, the current study conducted a comparative experimental assessment using two types of CA (basalt and natural) of 15 mm size following four distinct CMD codes to evaluate their impacts on the compressive strength of concrete. Expanding on the experimental results, ML models (XGB and RF) were developed and trained on the experimental and literature dataset to (i) accurately predict 28-day CS and (ii) clarify mechanistic drivers through feature-importance analysis. This integrated experimental–machine learning system offers useful insights on CA selection and mix proportioning, serving as a practical tool for data-driven decision-making in civil engineering.

2. Experimental Program

Ordinary Portland cement of 42.5 MPa grade strength was utilized as a binder. The fine aggregate was river sand with a specific gravity of 2.4, water absorption of 0.89%, fineness modulus of 2.22, and moisture content of 1.07%. Basalt CA with a specific gravity of 2.7, water absorption of 1.1%, and natural CA with a specific gravity of 2.54 and water absorption of 0.56% were used. Tap water was used for concrete batching and curing. Four CMD codes were evaluated: the American Concrete Institute (ACI) code [16], the British (BS) code [17], the Indian (IS) code [18], and the Chinese (JGJ) code [19]. A total of eight concrete mixtures with a strength of 25 MPa were proportioned as shown in Figure 1.

This program encompassed a total of 64 concrete samples, i.e., 32 cubes (150x150x150 mm) and 32 cylinders of size (150x300 mm). The specimens were cured for 28 days and tested under axial compression. A universal compression test machine with a capacity of 1000 MPa was used to measure the CS of cubes at a loading rate of 0.3 MPa/sec. The cylinders were loaded at a steady rate of 0.3 mm/min in displacement control mode, using a 1000 kN servo-hydraulic machine with a load drop of 100 kN/m^2. Figure 2(a, b & c) illustrates the testing setup.

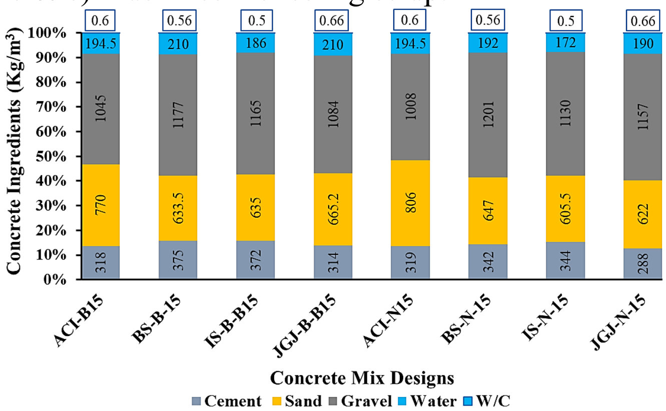

Figure 1 Concrete mix proportions.

(a). Specimen presentation (b). Testing of cubes (c). Testing of cylinders

Figure 2 Testing setup.

3. Machine Learning Applications

This research investigated the use of two robust ML models, i.e., XGB and RF, to validate the experimental program and predict the concrete strength properties. These models were chosen because of their strong applicability over a range of datasets, high-dimensional dataset management, and handling of nonlinear interactions. These models provide a complete toolset for solving a variety of regression problems in civil engineering applications. XGB is an ensembled ML technique often used for classification and regression problems due to its high efficiency and scalability. This technique was designed to combine several trained models, enhancing predictive accuracy beyond what individual models can achieve. XGB employs decision trees as base learners, and repeatedly improves its predictions by using gradient descent to optimize a loss function. The objective function and regularization terms are presented as in Equations 1 and 2.

$$\mathcal{L}(\emptyset) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (1)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|\omega_k\|^2 \quad (2)$$

Where, $l(y_i, \hat{y}_i)$ is loss function that measures the error between the true value y_i and the predicted value $\hat{y}_i$. $\Omega(f_k)$ is the regularization term that penalizes the complexity of the individual tree f_k. n is the number of samples. K is the number of trees. The XGB model for concrete strength estimation can be presented as in Equation 3.

$$F_{XGB} = \sum_{k=1}^{K} f_k(x) \quad (3)$$

RF is an ensemble learning technique that builds a multitude of decision trees during the training phase and produces the class label (for classification) or the mean prediction (for regression) from the individual trees. RF enhances decision trees by minimizing variance, thereby increasing the model's resilience to overfitting. RF generates a collection of decision trees, each trained on a random subset of features and a random portion of training data, i.e., with bootstrapping. The Equation 4 shows how the predictions generated by each tree are averaged to get the final outcome.

$$\hat{Y} = \frac{1}{T}\sum_{t=1}^{T} \hat{Y}_i \quad (4)$$

Where, $\hat{Y}$ is the predicted result; T is the number of trees in the forest; $\hat{Y}i$ is the prediction from an individual tree (i=1, 2, ... n), n is the total number of trees.

For ML models, the CS values of 28 days were computed. The preliminary dataset was established containing experimental data and relevant literature data of normal strength concrete. The specimen geometry did not lead to any bias in the ML model. A dataset containing 344 CS results was employed for the development of the ML models. The dataset was divided in 75% (training) and 25% (testing). The output parameter was CS, while the input variables were cement (kg/m^3), water (kg/m^3), sand (kg/m^3), CA content (kg/m^3), W/C ratio, and CA type.

Statistical parameters, i.e., coefficient of determination (R^2), mean absolute error (MAE), and mean squared error (MSE) were employed. These parameters are explained by their purpose, benefits and formulas, as given in Table 1. These statistical metrics assess model fitting and facilitate sensitivity analysis, thereby highlighting the significance of each input in the prediction process and offer a framework for comparing the performance of different algorithms.

Table 1 Statistical parameters to assess ML models.

Parameters	Formula	Purpose and Benefits
R^2	$R^2 = 1 - \frac{\sum_{i=1}^{n}(X_{Pi}-X_{Ei})^2}{\sum_{i=1}^{n}(X_{Ei}-\overline{X_{Ei}})^2}$	R^2 is a measure of the model's ability to explain the data variability and ranges from 0 to 1. R^2>80% indicate a good model fit.
MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{X_{Ei}-X_{Pi}}{n}\right\rvert$	MAE provides a reliable and understandable measure of average error that is less susceptible to outliers than MSE and clarifies forecast accuracy.
MSE	$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_{Ei} - X_{Pi})^2$	MSE measures average squared deviation between the actual and predicted results. It is sensitive to large errors.

4. Results and Discussions

4.1. Compressive Strength

Figure 3 presents the CS of cube specimens made with Basalt and natural CA. The first letter of the specimen ID indicates the CMD code, the second letter indicates the specimen type, and third letter represents CA type and size 15 mm. The ACI-C-B15, BS-C-B15, IS-C-B15, and JGJ-C-B15 specimens achieved 23.4%, 14.6%, 30.5%, and 8% higher CS than the designed CS, respectively. The natural CA cubes, ACI-C-N15, BS-C-N15, IS-C-N15, and JGJ-C-N15 exhibited 3.8%, 4.9%, 17.7%, and 3% higher CS than the designed CS, respectively, confirming the accuracy of the CMD calculations. The CS results for cylinders incorporated with basalt and natural CA are presented in Figure 4. The basalt CA cylinders, i.e., ACI-L-B15, BS-L-B15, IS-L-B15, and JGJ-L-B15, offered notably higher CS, i.e., 30%, 25%, 39%, and 37% than the natural CA cylinders across the four CMD codes, respectively. Concrete with basalt CA offered significantly higher CS than natural CA concrete due to basalt CA's higher crushing strength, shape, and surface texture. All these properties are involved in providing a strong bond between the cement matrix and CA, which results in superior compression characteristics.

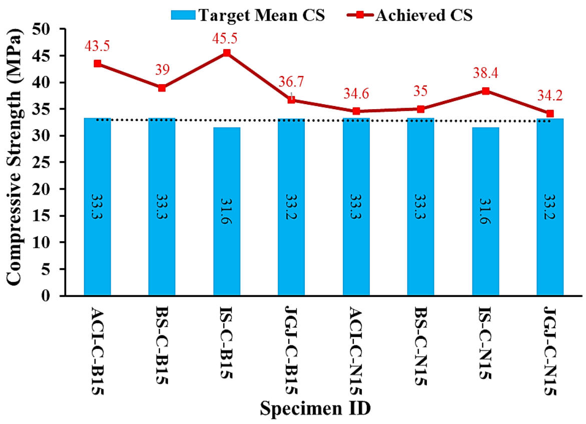

Figure 3 Compressive strength results of cube.

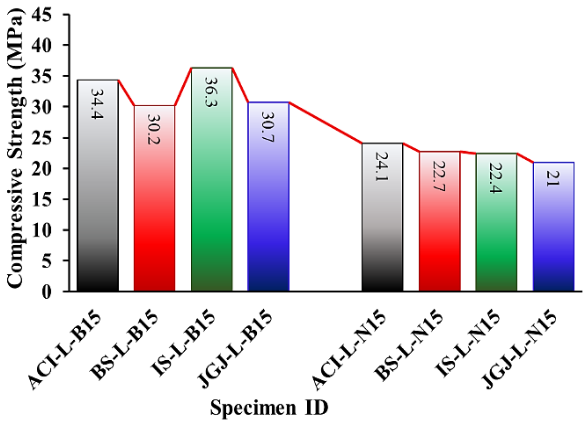

Figure 4 Compressive strength results of cylinders.

4.2. Performance Evaluation of ML Algorithms

The ML models (XGB and RF) were developed using the Python Scikit-learn package. To guarantee consistent scaling, the Standard Scaler module from Scikit-learn was first used to standardize the dataset's features. The statistical parameters such as R^2, MAE, and MSE, were employed to quantitatively assess the performance of these ML models. The graphical comparison of experimental and ML models' predicted CS results is illustrated in Figures 5 and 6. The statistical analysis demonstrated that RF and XGB models consistently showed excellent prediction abilities. RF marginally surpassed the XGB with the highest R^2 of 0.93, signifying enhanced fit and generalization capability. The XGB model obtained R^2 of 0.92, showing that the model accounts for more than 92% of the variation in the experimental CS data. Furthermore, in the case of average prediction error compared to the actual CS values, RF and XGB both demonstrated excellent accuracy and resilience, with MAE error of 0.59. Furthermore, RF attained lower MSE= 0.94 as compared to XGB's MSE= 1.07, underscoring its exceptional capacity to reduce significant discrepancies between predicted and actual CS results. This superior performance makes RF and XGB models as the most dependable ones for predicting the CS of concrete. Additionally, XGB's impressive predictive accuracy positions it as a viable alternative, especially in situations where boosting methods are preferred. The feature importance analysis identified cement content and CA type as the most influential factors with importance scores of 0.36 and 0.30, respectively, while the water-cement ratio served as a crucial regulator with an importance score of 0.09. The amount of sand and gravel showed minor contributions in determining the CS of concrete. The feature importance analysis results are illustrated in Figure 7.

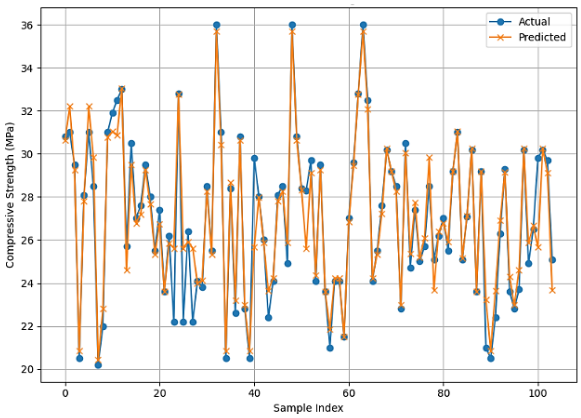

Figure 5 Actual and XGB model predicted CS

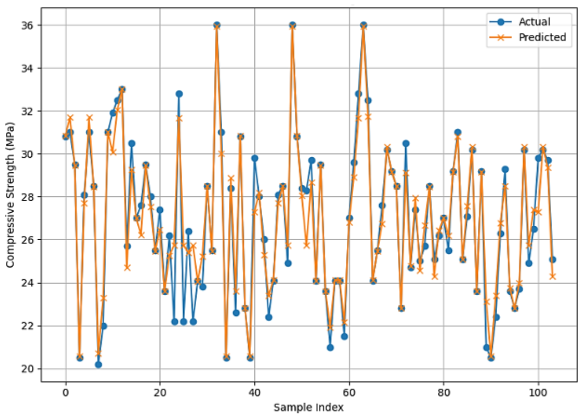

Figure 6 Actual and RF model predicted CS

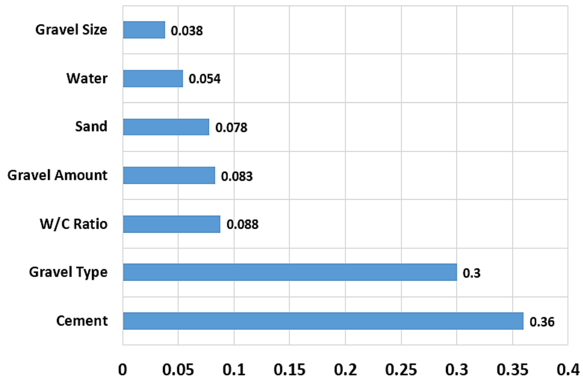

Figure 7 Feature importance analysis: RF model.

5. Conclusion

1) Basalt CA concrete exhibited 7% – 39% higher compressive strength than concrete made with natural CA across four CMD codes, demonstrating the importance of CA type, mineralogy, and surface texture in concrete strength.

2) Ensemble ML models reliably enhanced the experimental comprehension by demonstrating superior effectiveness in capturing nonlinear concrete mix-structure property relationships for 28-day CS prediction. This is evident by statistical analysis results as RF and XGB models attained R^2 of 0.93 and 0.92, respectively. Furthermore, RF and XGB both exhibited outstanding error matrices, i.e., MAE= 0.59 for both, and in the case of MSE, RF exhibited lower MSE= 0.94, as compared to XGB's MSE= 1.07.

3) The analysis of feature importance identified cement content as the primary positive factor, while the w/c ratio served as the key regulator. Additionally, the type of CA demonstrated a significant and consistent secondary impact. These clear rankings support the understanding of the roles played by hydration capacity, porosity control, and ITZ quality.

Acknowledgement

The authors gratefully acknowledge the financial support provided by National Key Research and Development Program of China (2024YFE0198400), the Fundamental Research Funds for the Central Universities (No. 2242024k30039, 2242024k30050), National Natural Science Foundation of China (52208233) and the Taishan Industrial Experts Program.

References

[1] Mehta, P.K., Monteiro, P.J.M. (2014) Concrete: Microstructure, Properties, and Materials, 4th

edition. McGraw-Hill Education, New York, USA.

[2] Ezeldin, A., Aitcin, P.C. (1991) Effect of Coarse Aggregate on the Behavior of Normal and High-Strength Concretes. Cement, Concrete, and Aggregates 13:121–124.

[3] Deng, P., Xu, K., Guo, S. (2023) Effects of coarse aggregate morphology on concrete mechanical properties. Journal of Building Engineering 63:105408.

[4] Góra, J., Piasta, W. (2020) Impact of mechanical resistance of aggregate on properties of concrete. Case Studies in Construction Materials 13: e00438.

[5] Konitufe, C., Abubakar, A., Sabo, B.A. (2023) Influence of Aggregate Size and Shape on the Compressive Strength of Concrete. Construction 3:15–22.

[6] Sharmin, R., Ahmed, M., Lubna, F.A. (2006) Comparison of strength performance of concrete with uncrushed or crushed coarse aggregates. Journal of Engineering and Applied Sciences 1:1–4.

[7] Golewski, G.L. (2024) Effect of Coarse Aggregate Type on the Fracture Toughness of Ordinary Concrete. Infrastructures 9:185.

[8] Al-Rawashdeh, M., Ashraf, S. (2015) Effect of basalt aggregates and plasticizer on the compressive strength of concrete. International Journal of Engineering & Technology 4:520.

[9] Al-Baijat, H.M. (2008) The use of basalt aggregates in concrete mixes in Jordan. Jordan Journal of Civil Engineering 2:63–70.

[10] Editorials. (2021) Concrete needs to lose its colossal carbon footprint. The international journal of science, Nature 597(7878):593–594.

[11] Li, Q.F., Song, Z.M. (2022) High-performance concrete strength prediction based on ensemble learning. Construction Building Material, 324:126694.

[12] Ben Chaabene, W., Flah, M., Nehdi, M.L. (2020) Machine learning prediction of mechanical properties of concrete: Critical review. Construction Building Materials 260, 119889.

[13] Miao, X., Zhu, J.X., Zhu, W.B., Wang, Y., Peng, L., Dong, H.L., Xu, L.Y. (2024) Intelligent prediction of comprehensive mechanical properties of recycled aggregate concrete with supplementary cementitious materials using hybrid machine learning algorithms, Case Stud. Constr. Mater. 21: e03708.

[14] Hosseinzadeh, M., Dehestani, M., Hosseinzadeh, A. (2023) Prediction of mechanical properties of recycled aggregate fly ash concrete employing machine learning algorithms, Journal of Building Engineering 76:107006.

[15] Golafshani, E., Khodadadi, N., Ngo, T., Nanni, A., Behnood, A. (2024) Modelling the compressive strength of geopolymer recycled aggregate concrete using ensemble machine learning, Advances in Engineering Software, 191:103611.

[16] ACI 211. (2002) Standard Practice for selecting proportions for normal, heavyweight, and mass concrete (reapproved 2009). American Concrete Institute, Farmington Hills, MI, USA.

[17] BS EN 206-1. (2001) Concrete Specifications, Performance, Production, and Conformity, Department of Environment BSI. Chiswick High Road London, England.

[18] IS 10262. (2009) Guidelines for concrete mix design proportioning. Bureau of Indian Standards, New Delhi, India.

[19] JGJ55. (2011) Specification for mix proportion design of ordinary concrete. China Building Industry Press, Beijing, China.