CANCOM2024 – CANADIAN INTERNATIONAL CONFERENCE ON COMPOSITE MATERIALS

EMPIRICAL AND BOOSTING MACHINE LEARNING BASED PREDICTION MODELS FOR THE STRENGTH OF RICE HUSK ASH-CONCRETE

Ashraf, Muhammad Waqas [1], Khan, Adnan [2], Tu, Yongming[1, 3, *], Ullah Safi[1], Wang, Chao[3, *]

[1] School of Civil Engineering, Southeast University, 211189 Nanjing, PR China

[2] School of Transportation, Southeast University, 211189 Nanjing, PR China.

[3] Division of Structural and Fire Engineering, Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, SE-97187 Luleå, Sweden

* Corresponding author (tuyongming@seu.edu.cn; yongmingtu@ltu.se)

Keywords: *Sustainable concrete, Predictive models, Model interpretation*

Abstract: The concrete construction industry is accountable for fifty per cent of all worldwide greenhouse gas emissions. Recycling and utilizing waste materials in concrete production is a viable solution to decrease its environmental impact. The utilization of rice husk ash (RHA) as a replacement for cement in concrete production has several potential benefits, including reducing cement consumption and decreasing its environmental impact. This research aims to review the potential uses of RHA in concrete and its implications for strength. Furthermore, the study aims to use various machine learning (ML) techniques, such as Genetic Programming (GP), Gradient Boosting Regression (GBR), and eXtreme Gradient Boosting Regression (XGBR), to predict the strength of RHA concrete (RHAC). The study utilized the ShaPley Adaptive exPlanations (SHAP) analysis on XGBR. A comprehensive parametric analysis was executed for GP to understand input trends linked to strength. This study facilitates the development and refinement of RHAC compositions and illustrates the capabilities of ML and statistical techniques in anticipation of improving structural material functionality.

1. INTRODUCTION

Currently, the construction industry is witnessing an increase in the incorporation of supplementary cementitious materials (SMCs) for manufacturing cement-based materials. This improvement can be attributed to the practice's favorable environmental, economic, and technical consequences. Some advantages that can be obtained are reducing CO_2 emissions into the atmosphere, minimizing energy use, cost-effectiveness, and sustainable management of solid waste. Reactive silica or aluminosilicate-containing industrial byproducts are the main source of SCMs. These substances can react with the byproducts of cement hydration to produce secondary cementitious compounds that enhance the strength and durability of cementitious materials. [1] Commonly used SCM include slag, fly ash, metakaolin, and rice husk ash.

Rice husk ash (RHA) is a pozzolanic substance produced through rice husk combustion at temperatures ranging from 600 to 850 °C. It contains silica as the main oxide and can replace up to 30% cement, depending on its pozzolanic activity [2]. RHA offers numerous technical advantages along with sustainability and economic benefits. According to [3], concrete mixtures containing 5-30% RHA can exhibit higher strength in early and late stages than control concrete mixtures. In addition, including 10-20% RHA in concrete mixtures significantly reduces chloride penetration by 81-89% [4]. This is due to the formation of a denser microstructure caused by calcium silicate hydrate (CSH) [5]. Concrete specimens in cubes and cylinders are usually made to evaluate the strength of mortar and concrete. However, evaluating the required strength in laboratories and fields is more expensive and time-consuming. Thus, the efficacy of concrete is assessed through the utilization of empirical and machine-learning regression methodologies [6].

To forecast the desired outcomes, numerous computational and numerical methods have been used in various fields [7], [8]. Similarly, Tie et al. [9] used artificial neural networks (ANN) analysis to generate heat and entropy in non-Newtonian fluid flow between rotating disks. Khan et al. [10] predicted geopolymer concrete mechanical properties using ensemble and individual methods. Ensemble learning approaches have become increasingly popular due to their greater prediction capabilities [11]. Some researchers have proposed bagging and boosting series algorithms for concrete strength prediction [12]. To predict concrete strength, Lyngdoh et al. used significant ML algorithms. According to their outcomes, the XGBoost model is the most effective [13]. In another study, XGBoost and CatBoost were employed to predict the concrete strength, and it was discovered that XGBoost and CatBoost had considerably fewer mean errors between predicted and actual values [14]. GEP algorithm, simple and practical mathematical equations predicted ground granulated blast-furnace slag (GGBS)-based Geopolymer concrete (GPC) mortar CS [15]. Although ML techniques have been widely accepted as effective modelling methodologies in numerous engineering applications, there remains a limited exploration of these techniques for sustainable concrete's tensile strength (TS) and Flexural Strength (FS). Therefore, this study has four main objectives: (I) To collect a comprehensive dataset of TS and FS of RHAC from existing literature. (II) Propose three novel ML approaches (Boosting and Empirical), namely gradient boosting (GBR), extreme gradient boosting regression (XGBR), and genetic Programming (GP) models, to predict the TS and FS of RHAC. (III) To compare the performance of the models proposed in this study. (IV) To determine the significance of different features by conducting SHAP and parametric analysis.

2. Model building and evaluation criteria:

This study chooses GBR, XGBR, and GP models to predict the TS and FS of RHAC. Before developing the prediction model, it is important to determine the best data split so that the model has higher generalization capability; the dataset was randomly divided into training and testing sets, i.e., 70% and 30%, respectively. The hyperparameters significantly influence the precision of the model. Optimization can improve the ML model's performance by identifying the best hyperparameters on the dataset. The grid search technique was utilized to optimize the hyperparameters of the ML model. The grid search method is a systematic strategy that extensively explores a range of hyperparameter variations and trains the model numerous times. The configuration that yields the best outcomes throughout several training sessions is referred to as the ideal combination of hyperparameters. Figure 1 shows the methodology flowchart adopted for the current study.

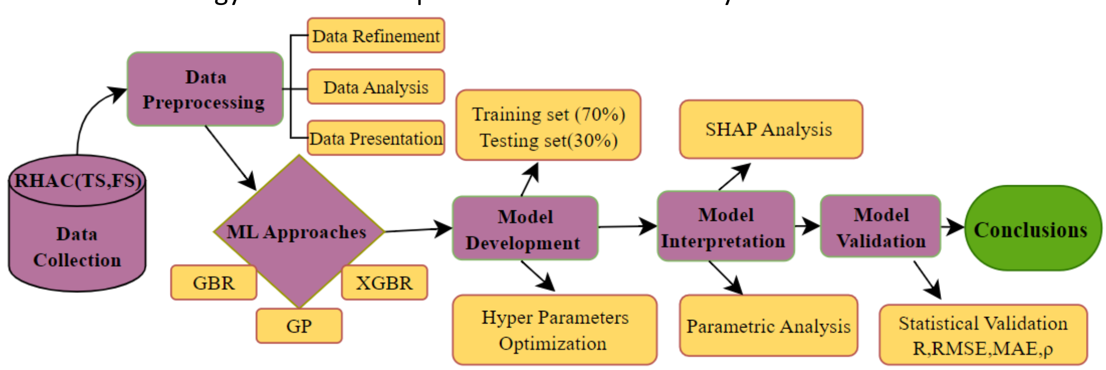

Figure 1 Methodology

The effectiveness of the constructed models was evaluated using various statistical measures, such as root mean square error (RMSE), relative root mean square error (RRMSE), mean absolute error (MAE), and correlation coefficient (R) [11]. The values for R are limited to the range of 0 to 1, with a greater R-value indicating a better model. Conversely, smaller RMSE, MAE, and RRMSE values indicate the model's superior performance. In addition, the researchers calculated a performance index (ρ) [16], which was a combined measure of the model's accuracy, considering both RRMSE and R. The variable ρ takes on values ranging from zero to positive infinity, where a lower value indicates a higher level of model performance.

3. Results and Discussion

3.1 GBR, XGBR, and GP outcomes for Split tensile strength (TS) and flexural strength (FS):

In Figure 2, the experimental and predicted values for TS for all models are shown juxtaposed with the absolute error and statistical metrics such as RMSE, MAE, etc. Figure 2 illustrates the efficacy of the models in precisely predicting the TS of RHAC. The robustness of the GBR, XGBR, and GP models was determined by analyzing their prediction outcomes and associated errors. During the training phase, the GBR, XGBR, and GP models obtained R values of 0.993, 0.992, and 0.895, respectively. During the testing phase, the recorded values were 0.990, 0.992, and 0.889, respectively. These results indicate a significant relationship between the experimental and predicted outcomes. The RMSE values for the models during the training phase were 0.19, 0.20, and 0.73, respectively. Similarly, during the testing phase, the RMSE values were 0.29, 0.29, and 0.86. During the training phase, the MAE values were recorded as 0.15, 0.16, and 0.60. In the subsequent testing phase, the MAE values were observed to be 0.24, 0.22, and 0.74. The statistics consistently exhibited little variance throughout all datasets. In addition, the explicit equation derived from the GP model tree for TS and FS is presented in Eq (1) and (2), respectively.

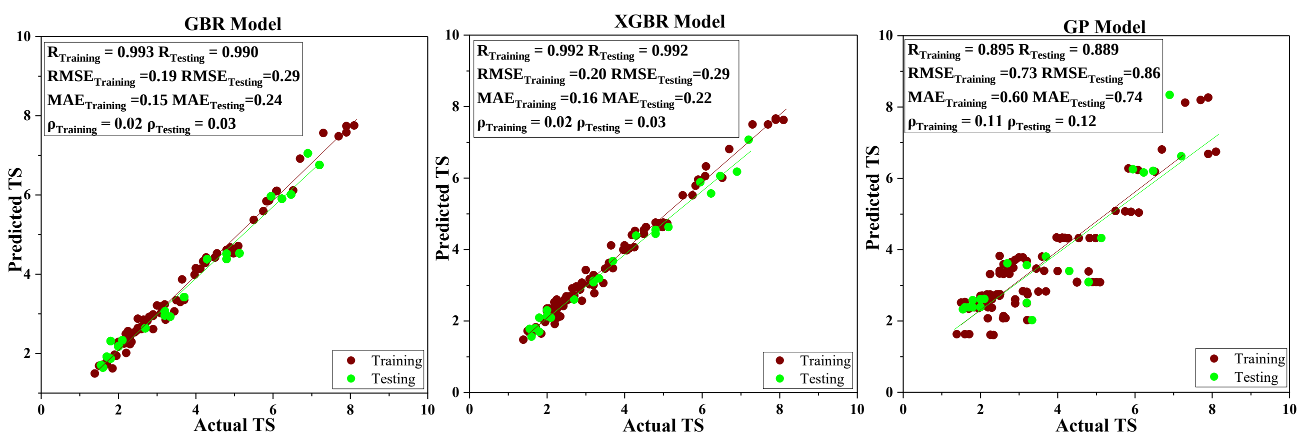

Figure 2 GBR, XGBR, and GP outcomes for TS

$$TS = \left(\left(\left(\left(c_0(W) + c_1(FA) + \frac{(c_2(A)+c_3(SP))\left((c_4(W)+\left((c_5(C)(c_6(RHA)+(c_7(SP))^2\right)*c_8\right)}{W*W*c_9}\right)\right)^2 * c_{10} + c_{11}\right) \quad (1)$$

Where, c_0=-0.05113, c_1=0.01015, c_2=1.0184, c_3=3.1485, c_4=2.6168, c_5=0.71398, c_6=0.6942, c_7=4.4812, c_8=-0.0190, c_9=0.5383, c_{10}=0.04919, c_{11}=1.5026

$$FS = \left(\left(c_0(W) + c_1(SP) + \frac{c_2(RHA)+c_3(C)}{c_4(W)+c_5(A)}\right)\left(c_6(WC) + c_7(W) + c_8(RHA)\right) * c_9 + c_{10}\right) \quad (2)$$

Where, c_0=-0.09899, c_1=-0.6848, c_2=1.0649, c_3=0.2885, c_4=-0.0428, c_5=0.0814, c_6=2.1539, c_7=1.2801, c_8=-0.9041, c_9=-0.0002, c_{10}=-4.3752.

The predicted and experimental results for FS can be seen in Figure 3. The performance of each ML model is gauged by the values of statistical indicators, as previously discussed. The GBR, XGBR, and GP models exhibited high R values for both the training and testing phases. During the training phase, the R-values were 0.987, 0.975, and 0.957, respectively, while in the testing phase, they were 0.978, 0.973, and 0.951. This demonstrates a strong correlation between the actual and predicted values. In the training phase, the RMSE values for these models were 0. 50, 0.68, and 0.85, respectively, and the MAE values were 0.36, 0.54, and 0.65. For the testing phase, the RMSE values were 0.68, 0.68, and 0.95, the MAE values were 0.54, 0.58, and 0.75. All these details are visually represented in Figure 3.

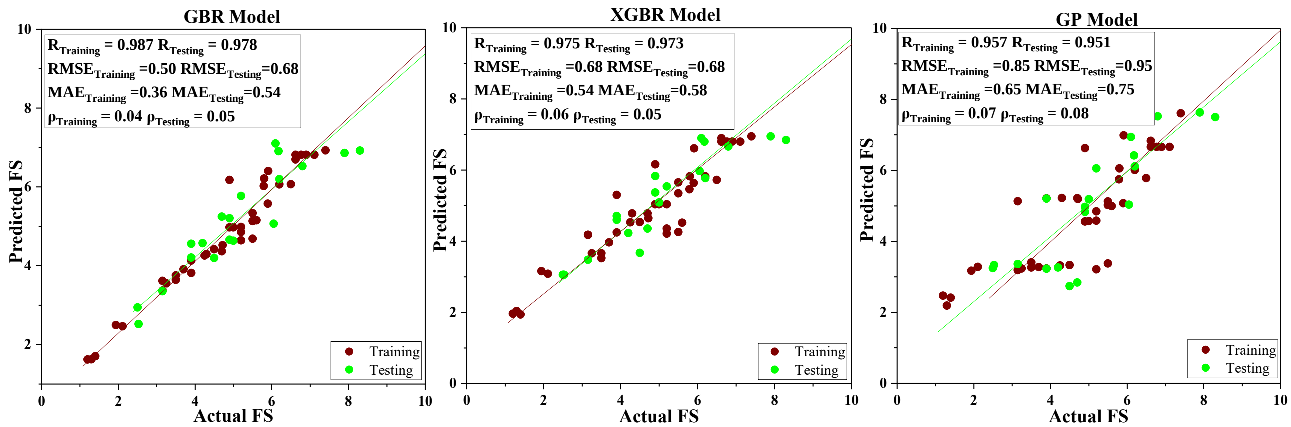

Figure 3 GBR, XGBR, and GP outcomes for FS

4. Model interpretation

4.1 SHAP analysis

The SHAP values for the model are illustrated in Figure 4 (a,b). All input parameters tend to affect the TS and FS of RHAC consistently. The beeswarm plot in SHAP analysis illustrates the significance of feature values using color-coding on the right side.

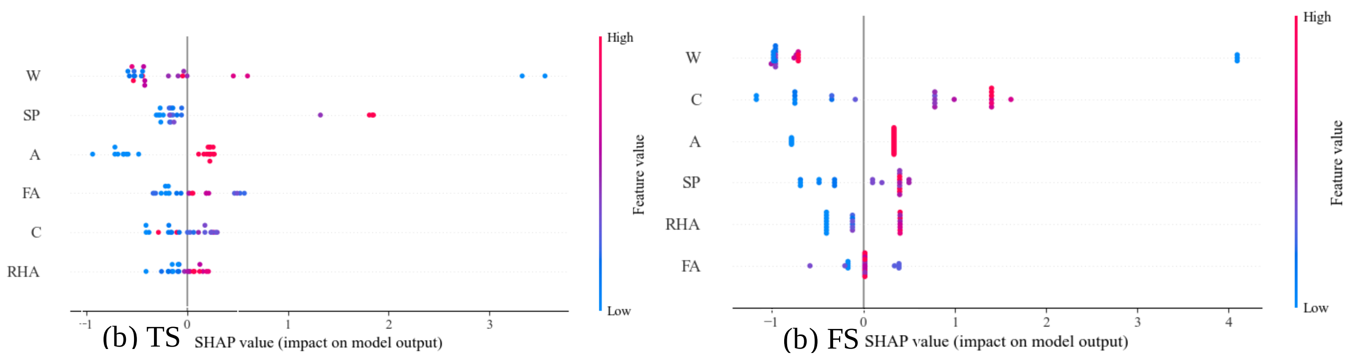

Figure 4 Beeswarm plot for (a) TS and (b) FS

The SHAP analysis, illustrated in a beeswarm plot, confirms that the outcomes derived from the XGBR model correlate with this comprehension. The TS and FS model's output is significantly impacted by the W, SP, and W, C, respectively, as shown in Figure 4 (a,b), where lower values of A have a negative influence and higher values have a favourable impact.

4.2 Parametric analysis:

To obtain a deeper understanding of the mechanical properties of RHAC, this study not only depends on historical trends but also does a parametric analysis. This investigation aims to validate the GP model and evaluate the impact of specific input features on the mechanical properties while considering other factors constant [11]. A single input variable was varied from its minimum to its highest value to assess the effects on the TS and FS. Figure 5 presents the outcomes of a parametric study conducted using the generated GP models. An increase in TS and FS of RHAC was positively correlated with increases in C, RHA, SP, and A. Furthermore, it was noted that an increase in W leads to a significant reduction in strength. According to.[17] the parametric analysis results validate the interpretations of earlier studies and are following the experimental investigation.

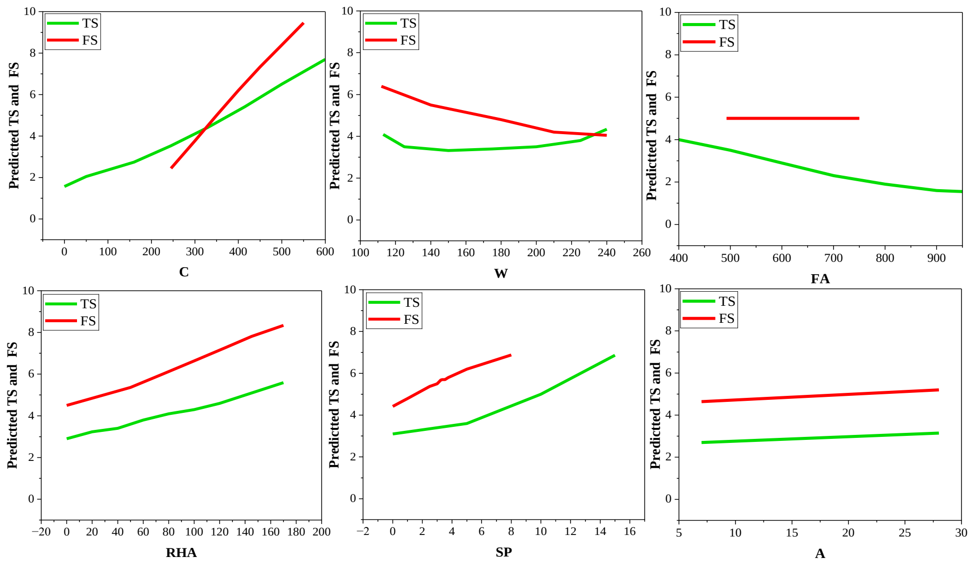

Figure 5 Parametric analysis for TS and FS

5. Conclusion:

This work involved the development of three machine learning prediction models, namely GBR, XGBR, and GP, to predict the TS and FS for RHAC. The models were trained using a comprehensive dataset of 110 and 67 data points, respectively.

- The GBR model demonstrated higher validation accuracy than the XGBR and GP models during cross-validation for TS and FS of RHAC. This was further supported by its stronger correlation coefficient and lower error rate, as demonstrated by statistical validation of the model. In addition, the simplified GP expression could be used by practitioners and investigators to predict the TS and FS of RHAC.
- The SHAP analysis for XGBR shows that W, SP, and W, C have the greatest impact on RHAC's TS and FS, respectively.
- The parametric analysis shows the suggested GP model's TS and FS prediction accuracy. The main contributing factors for each strength attribute were C, RHA, SP, and A, which have a positive correlation, and W, which has a negative correlation for each strength attribute.
- This proposed study has potential application for determining the TS and FS strength of RHAC. Further investigations could be implemented to fine-tune the proposed models on a larger dataset and comparison with more advanced ML approaches.

References:

[1] M. Ahmaruzzaman, "A review on the utilization of fly ash," *Prog Energy Combust Sci*, vol. 36, no. 3, pp. 327–363, Jun. 2010, doi: 10.1016/j.pecs.2009.11.003.

[2] R.-S. Bie, X.-F. Song, Q.-Q. Liu, X.-Y. Ji, and P. Chen, "Studies on effects of burning conditions and rice husk ash (RHA) blending amount on the mechanical behavior of cement," *Cem Concr Compos*, vol. 55, pp. 162–168, Jan. 2015, doi: 10.1016/j.cemconcomp.2014.09.008.

[3] J. M. Paris, J. G. Roessler, C. C. Ferraro, H. D. DeFord, and T. G. Townsend, "A review of waste products utilized as supplements to Portland cement in concrete," *J Clean Prod*, vol. 121, pp. 1–18, May 2016, doi: 10.1016/j.jclepro.2016.02.013.

[4] J. Sousa Coutinho, "The combined benefits of CPF and RHA in improving the durability of concrete structures," *Cem Concr Compos*, vol. 25, no. 1, pp. 51–59, Jan. 2003, doi: 10.1016/S0958-9465(01)00055-5.

[5] G. H. M. J. Subashi De Silva, S. Vishvalingam, and T. Etampawala, "Effect of waste rice husk ash from rice husk fuelled brick kilns on strength, durability and thermal performances of mortar," *Constr Build Mater*, vol. 268, p. 121794, Jan. 2021, doi: 10.1016/j.conbuildmat.2020.121794.

[6] W. Ben Chaabene, M. Flah, and M. L. Nehdi, "Machine learning prediction of mechanical properties of concrete: Critical review," *Constr Build Mater*, vol. 260, p. 119889, Nov. 2020, doi: 10.1016/j.conbuildmat.2020.119889.

[7] B. Liu, J. Shi, F. Zhou, S. Shen, Y. Ding, and J. Qin, "Effects of steam curing regimes on the capillary water absorption of concrete: Prediction using multivariable regression models," *Constr Build Mater*, vol. 256, p. 119426, Sep. 2020, doi: 10.1016/j.conbuildmat.2020.119426.

[8] Z. He, H. Zhu, J. Shi, J. Li, Q. Yuan, and C. Ma, "Multi-scale characteristics of magnesium potassium phosphate cement modified by metakaolin," *Ceram Int*, vol. 48, no. 9, pp. 12467–12475, May 2022, doi: 10.1016/j.ceramint.2022.01.112.

[9] T. Zhao, M. I. Khan, and Y. Chu, "Artificial neural networking (ANN) analysis for heat and entropy generation in flow of non-Newtonian fluid between two rotating disks," *Math Methods Appl Sci*, vol. 46, no. 3, pp. 3012–3030, Feb. 2023, doi: 10.1002/mma.7310.

[10] M. A. Khan *et al.*, "Geopolymer Concrete Compressive Strength via Artificial Neural Network, Adaptive Neuro Fuzzy Interface System, and Gene Expression Programming With K-Fold Cross Validation," *Front Mater*, vol. 8, May 2021, doi: 10.3389/fmats.2021.621163.

[11] A. Khan *et al.*, "An ensemble tree-based prediction of Marshall mix design parameters and resilient modulus in stabilized base materials," *Constr Build Mater*, vol. 401, p. 132833, Oct. 2023, doi: 10.1016/j.conbuildmat.2023.132833.

[12] A. G. Oyeyi, A. Khan, J. Huyan, W. Zhang, F. M.-W. Ni, and S. L. Tighe, "Ensemble and evolutionary prediction of layers temperature in conventional and lightweight cellular concrete subbase pavements," *International Journal of Pavement Engineering*, vol. 25, no. 1, Dec. 2024, doi: 10.1080/10298436.2024.2322525.

[13] G. A. Lyngdoh, M. Zaki, N. M. A. Krishnan, and S. Das, "Prediction of concrete strengths enabled by missing data imputation and interpretable machine learning," *Cem Concr Compos*, vol. 128, p. 104414, 2022, doi: https://doi.org/10.1016/j.cemconcomp.2022.104414.

[14] S. Pranav, M. Lahoti, and M. Gopalarathnam, "Concrete Compressive Strength Prediction Using Boosting Algorithms," 2023, pp. 307–315. doi: 10.1007/978-981-19-8979-7_26.

[15] A. A. Shahmansouri, H. Akbarzadeh Bengar, and S. Ghanbari, "Compressive strength prediction of eco-efficient GGBS-based geopolymer concrete using GEP method," *Journal of Building Engineering*, vol. 31, p. 101326, 2020, doi: https://doi.org/10.1016/j.jobe.2020.101326.

[16] A. H. Gandomi and D. A. Roke, "Assessment of artificial neural network and genetic programming as predictive tools," *Advances in Engineering Software*, vol. 88, pp. 63–72, Oct. 2015, doi: 10.1016/j.advengsoft.2015.05.007.

[17] A. A. K. Al-Alwan *et al.*, "The impact of using rice husk ash as a replacement material in concrete: An experimental study," *Journal of King Saud University - Engineering Sciences*, Apr. 2022, doi: 10.1016/j.jksues.2022.03.002.