© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718351

Gene Expression Programming for Predicting Compressive and Flexural Strength of Concrete with Rice Husk Ash and Sawdust Ash

OYESANYA, O.G.
Department of Civil Engineering, Faculty of Engineering, University of Nigeria Nsukka

Abstract-* *This study presents the development and validation of a Gene Expression Programming (GEP) model for predicting the compressive and flexural strength of sustainable concrete incorporating Rice Husk Ash (RHA) and Sawdust Ash (SDA) as supplementary cementitious materials. A dataset comprising 123 concrete mix samples was assembled, with 80% used for model training and 20% reserved for independent testing. The GEP model was carefully configured using multiple chromosomes, optimized genetic operators, and a function set consisting of arithmetic and nonlinear operators to effectively capture the complex interactions governing cement hydration and pozzolanic reactions. Model performance evaluation demonstrated strong predictive capability. For compressive strength, the model achieved a test coefficient of determination (R^2) of 0.775, RMSE of 8.15 N/mm², and MAE of 6.32 N/mm². Flexural strength prediction showed improved accuracy, with an R^2 of 0.841, RMSE of 0.71 N/mm², and MAE of 0.52 N/mm². Five-fold cross-validation further confirmed model robustness and generalization. The evolved expression trees generated explicit mathematical relationships, providing valuable interpretability and insight into the nonlinear influence of mix constituents on mechanical properties. Sensitivity analysis using SHAP identified cement and fine aggregate as the most influential variables for compressive strength, while fine aggregate, water, and coarse aggregate were dominant in flexural strength prediction. RHA exhibited consistently greater influence than SDA, indicating its higher pozzolanic reactivity. The results demonstrate that Gene Expression Programming is an effective and interpretable tool for modeling the mechanical performance of sustainable concrete, offering both reliable prediction accuracy and transparent analytical expressions for engineering applications.

Keywords-* *Gene Expression Programming, Sustainable Concrete, Rice Husk Ash, Sawdust Ash, Compressive Strength, Flexural Strength, Symbolic Regression

I. INTRODUCTION

The global construction industry is responsible for approximately 8% of worldwide CO_2 emissions, with cement production alone accounting for 5-7% of total anthropogenic carbon dioxide emissions [1]. This substantial environmental burden has intensified research into sustainable alternatives, with agricultural waste materials emerging as promising supplementary cementitious materials (SCMs). Rice Husk Ash (RHA) and Sawdust Ash (SDA), byproducts of agricultural processing, represent abundant, underutilized resources that can partially replace cement while potentially maintaining or enhancing concrete performance.

Rice Husk Ash contains 85-95% amorphous silica when burned under controlled conditions (600-700°C), making it highly reactive as a pozzolanic material [2]. Research has demonstrated that partial cement replacement with RHA up to 20-25% can maintain strength while enhancing durability properties such as chloride penetration resistance [3][4]. Sawdust Ash typically contains 40-60% silica along with calcium and potassium [5]. While SDA generally exhibits lower pozzolanic reactivity than RHA, proper processing at 550-650°C can yield valuable SCM properties [6].

Predicting the mechanical properties of concrete incorporating these agricultural waste materials presents significant challenges. The pozzolanic reactivity, particle size distribution, and chemical composition vary considerably depending on burning conditions and source materials. Traditional empirical models and statistical regression techniques frequently fail to capture the complex, nonlinear interactions between mix components [7]. Advanced modelling techniques capable of handling intricate multivariate relationships are essential for optimizing sustainable concrete designs.

Gene Expression Programming (GEP) has emerged as a particularly powerful technique for modeling such complex systems. Developed by [8]. GEP represents an evolutionary algorithm that combines the adaptive search mechanisms of genetic algorithms with the structural flexibility of genetic programming. Unlike black-box machine learning models, GEP automatically evolves symbolic mathematical expressions that transparently represent discovered patterns in data. This interpretability makes GEP especially valuable in civil engineering applications where understanding variable relationships guides practical decision-making.

Previous research has demonstrated GEP's effectiveness for concrete strength prediction. [9] showed GEP outperformed both ANN and regression models for high-performance concrete, achieving R^2 values exceeding 0.95. Ahmad et al. (2021) found GEP superior to ANN ($R^2 = 0.95$ vs 0.92) for recycled aggregate concrete [10] successfully applied GEP to geopolymer concrete, while [11] developed GEP models for ultra-high-performance concrete. However, research specifically addressing GEP modelling of combined RHA-SDA concrete systems remains limited.

This paper presents a comprehensive investigation with the following objectives: (1) develop an optimized GEP model for predicting compressive and flexural strength of RHA-SDA blended concrete, (2) derive interpretable mathematical expressions from evolved expression trees, (3) validate model performance through rigorous testing and five-fold cross-validation, (4) conduct sensitivity analysis using SHAP values to quantify individual variable contributions, and (5) demonstrate the practical applicability of GEP-derived equations for sustainable concrete mix design.

II. LITERATURE REVIEW

2.1. Agricultural Waste SCMs in Concrete

Rice husk, the protective covering of rice grains, constitutes approximately 20% of rice production by weight. When burned under controlled conditions, it yields ash containing highly reactive amorphous silica. The pozzolanic activity of RHA depends critically on burning temperature, with optimal reactivity achieved between 600-700°C [2]. At higher temperatures (>800°C), silica transforms to crystalline forms with reduced pozzolanic activity. Studies have shown that grinding RHA to finer particle sizes (below 45 µm) significantly enhances reactivity and strength contribution [12].

Research on RHA incorporation has demonstrated multiple benefits. [3] found that RHA-blended cement mortar showed superior chloride penetration resistance compared to control mixtures. [13] reported that RHA improved durability properties of recycled aggregate concrete. The ultrafine RHA particles contribute to pore refinement through pozzolanic reactions (consuming calcium hydroxide to form C-S-H gel) and physical filling effects, densifying the concrete microstructure [14].

Sawdust Ash, derived from wood processing waste, typically contains 40-60% silica along with calcium, potassium, magnesium, and other elements. The pozzolanic activity depends on burning conditions, with optimal reactivity achieved between 550-650°C [6]. However, SDA generally exhibits lower pozzolanic reactivity than RHA due to lower silica content and higher carbon residue. [15] investigated combined RHA-SDA use, reporting that ternary blends can provide balanced performance when replacement levels are carefully optimized.

2.2. Gene Expression Programming: Fundamentals and Architecture

Gene Expression Programming, developed by [8], represents an evolution from genetic programming that addresses GP's limitations in representation and search efficiency. In GEP, solutions are encoded as linear chromosomes of fixed length (genotype) but expressed as nonlinear expression trees of variable size and shape (phenotype). This separation enables efficient genetic operations on simple linear strings while fitness evaluation occurs on complex mathematical expressions.

Each GEP chromosome consists of one or more genes divided into head and tail sections. The head contains both function symbols (operators) and terminal symbols (variables and constants), while the tail contains only terminal symbols. This structural constraint ensures every chromosome decodes into a valid expression tree, eliminating the closure problem inherent in traditional genetic programming [8]. The tail length is mathematically determined by $t = h(n - 1) + 1$, where t is tail length, h is head length, and n is maximum function arity.

The evolutionary process employs multiple genetic operators. Mutation introduces diversity by randomly altering symbols in the chromosome. Crossover recombines genetic material between parent chromosomes through one-point, two-point, or gene crossover operations. Transposition mechanisms (IS and RIS transposition) enable gene segment repositioning. Selection operates through fitness-based tournament selection, where chromosomes with better fitness scores are preferentially chosen for reproduction. Over successive generations, the population evolves toward higher fitness [16][17].

2.3. *Agricultural Waste SCMs in Concrete*

The application of GEP to concrete strength prediction has gained increasing attention. [9] applied GEP to high-performance concrete, achieving R^2 values exceeding 0.95 while providing simpler mathematical expressions than ANN. The authors highlighted that GEP's explicit equations enabled engineers to understand variable contributions, facilitating practical implementation in mix design optimization.

[18] investigated GEP for recycled concrete aggregate, reporting superior performance ($R^2 = 0.95$) compared to ANN ($R^2 = 0.92$). The research demonstrated that GEP-derived equations could be implemented in spreadsheet software without requiring specialized machine learning tools. [19] compared GEP with fuzzy logic for cement mortar strength, finding GEP delivered better performance with more straightforward implementation.

[10] successfully applied GEP to geopolymer concrete, demonstrating its capability to model complex alkali-activated materials. [11] developed GEP models for ultra-high-performance concrete, reporting simpler and more precise equations than Multi-Expression Programming. Ashraf et al. (2024) explored GEP for sustainable green concrete with various SCMs. [20] conducted a comparative study concluding that while GEP and ANN achieved similar accuracy, GEP provided superior interpretability through explicit mathematical expressions.

Despite these advances, research specifically addressing GEP modelling of combined RHA-SDA concrete systems remains limited. The unique pozzolanic characteristics and particle interaction effects of these agricultural waste materials present distinct modelling challenges not adequately addressed in existing literature. This study fills this gap by developing specialized GEP models with comprehensive validation and sensitivity analysis.

III. METHODOLOGY

3.1 Experimental Program and Dataset

This study utilized 123 datasets with 7 features and 2 output. The features include cement, fine aggregate, coarse aggregate, water, sawdust ash, rice husk ash. The outputs are compressive strength and flexural strength. The dataset which was obtained from the research works of [13][14][21-27].

3.2 Dataset Splitting Strategy

The dataset of 123 samples was divided using an 80-20 split: 98 samples (80%) for training and 25 samples (20%) for independent testing. The split was performed randomly while maintaining representative distribution of response variables in both subsets. Statistical analysis confirmed similar distributions in training and testing sets. Input variables (cement, fine aggregate, coarse aggregate, water, SDA, RHA) served as terminals in GEP chromosomes, while output variables (compressive and flexural strength) served as prediction targets.

3.3 GEP Model Configuration and Parameters

The GEP model was implemented using GeneXpro Tool v5.0 for expression tree generation and Google Colab for additional analysis and visualization. Model parameters were carefully selected based on preliminary experiments balancing population diversity, computational efficiency, and convergence quality. Table 1 presents the complete GEP parameter configuration.

Table 1. Gene Expression Programming Model Parameters

Parameter	Settings
Number of Chromosomes	30
Number of Genes	3 (Flexural Strength) 5 (Compressive Strength)
Head Size	8
Tail Size	9
Linking Function	Addition
Constants per Gene	10
Data Category	Floating Point
Constant Range	-10 to +10
Mutation Rate	0.00138
Inversion Rate	0.00546
Leaf Mutation Rate	0.00546
Function Set	+, -, ×, ÷, ^{2}, exp, sqrt, ln, ^{3}
Number of Generations	100
Selection Method	Tournament Selection
Fitness Function	Mean Squared Error (MSE)

3.4 Model Training and Testing Procedure

The training phase presented the GEP algorithm with 98 training samples as input-output pairs. Six input variables (cement, fine aggregate, coarse aggregate, water, SDA, RHA) served as terminals, while target outputs (compressive or flexural strength) provided fitness evaluation criteria. The evolutionary process iteratively generated chromosome populations, decoded them into expression trees, evaluated predictions against actual strength values, and selected superior individuals for reproduction.

During training, GEP applied genetic operators (mutation, crossover, inversion, transposition) to create new chromosomes from selected parents. Each chromosome was decoded into an expression tree and used to compute predictions for all training samples. Fitness was calculated as the inverse of mean squared error, such that minimizing prediction error corresponds to maximizing fitness. Over successive generations, the population evolved toward chromosomes producing more accurate predictions, with the best-performing chromosome from each generation preserved through elitism.

The testing phase evaluated the best-evolved chromosome on 25 completely independent testing samples never seen during evolution. This independent evaluation is critical for assessing true generalization capability and detecting overfitting. The trained model generated predictions for each testing sample, and performance metrics (R^2, RMSE, MAE) were calculated by comparing predictions against actual measurements. The separation ensures reported performance reflects ability to predict unseen data rather than mere memorization.

3.5 Performance Evaluation Metrics

Model performance was quantified using four complementary metrics. The Coefficient of Determination (R^2) measures the proportion of variance explained by the model (0 to 1, higher is better). Root Mean Square Error (RMSE) quantifies the standard deviation of prediction residuals in the same units as the response variable, providing directly interpretable error magnitude. Mean Absolute Error (MAE) provides the average absolute magnitude of prediction errors, treating all errors equally without emphasizing outliers. Relative Squared Error (RSE) normalizes squared prediction errors relative to a baseline model that simply predicts the mean value, with values less than 1 indicating superior performance. The scatter plots for compressive strength and flexural strength are shown in Fig. 1a and Fig. 1b respectively.

3.6 Cross-Validation Strategy

Five-fold cross-validation provided robust performance estimates and assessed model stability across different data subsets. The entire dataset of 123 samples was partitioned into 5 approximately equal folds (24-25 samples each). The cross-validation process involved 5 iterations, where each fold served as the testing set once while the remaining 4 folds were used for training. Every sample served as a test sample exactly once, providing comprehensive evaluation coverage.

For each fold, the GEP model was trained from scratch using the same parameters from Table 1, ensuring fair comparison. The evolutionary process evolved independent chromosomes for each fold's training data, and the best chromosome was evaluated on that fold's testing samples. Performance metrics (R^2, RMSE, MAE) were calculated for each fold, and mean and standard deviation across all 5 folds characterized overall performance and variability. Cross-validation provides more reliable estimates than single train-test splits and indicates model stability through standard deviation values.

IV. RESULTS AND DISCUSSIONS

4.2 GEP Model Performance

Table 2 presents comprehensive performance metrics for the GEP model on both training and testing datasets for compressive and flexural strength prediction. The results demonstrate that GEP successfully captured the complex, nonlinear relationships between mix components and mechanical properties, achieving good predictive accuracy while maintaining model interpretability through explicit mathematical expressions.

For compressive strength prediction, the GEP model achieved training R^2 of 0.825 and testing R^2 of 0.775, indicating approximately 77.5% of variance in compressive strength for unseen samples is explained by the model. The RMSE of 8.15 N/mm² on the test set represents approximately 49% of the mean compressive strength (16.49 N/mm²), while MAE of 6.32 N/mm² indicates average absolute prediction error around 38% of the mean. The small difference between training and testing performance (R^2 difference of 0.05) suggests minimal overfitting, confirming the model learned generalizable patterns rather than memorizing training data.

For flexural strength prediction, the GEP model demonstrated superior performance with training R^2 of 0.873 and testing R^2 of 0.841. Higher R^2 values compared to compressive strength likely reflect lower absolute variability in flexural strength data ($\sigma = 0.41$ N/mm² vs 2.53 N/mm² for compressive strength) and potentially more direct relationships between mix parameters and flexural behaviour. The test RMSE of 0.71 N/mm² represents approximately 28% of mean flexural strength (2.55 N/mm²), while MAE of 0.52 N/mm² indicates average absolute errors around 20% of the mean. Again, the small gap between training and testing performance (R^2 difference of 0.032) confirms good generalization.

The RSE values (0.225 for compressive strength, 0.159 for flexural strength on test sets) indicate the GEP model significantly outperforms a baseline model that simply predicts the mean value. RSE values substantially below 1.0 confirm that evolved expressions capture meaningful relationships between input variables and strength properties. The lower RSE for flexural strength reflects the model's stronger performance on this response variable.

These performance metrics compare favourably with previous GEP studies on concrete strength prediction. Aslam et al. (2020) reported R^2 values of 0.89-0.95 for high-performance concrete, while [18] achieved R^2 of 0.95 for recycled aggregate concrete. The somewhat lower R^2 in the current study (0.775-0.841) may reflect additional complexity introduced by the combined RHA-SDA system compared to single-SCM systems. Nevertheless, the results demonstrate GEP successfully modelled the nonlinear pozzolanic reactions and particle interaction effects characteristic of agricultural waste SCMs.

Table 2. GEP Model Performance Metrics

Response Variable	Dataset	R^2	RMSE (N/mm²)	MAE (N/mm²)	RSE
Compressive Strength	Training	0.825	7.23	5.84	0.175
	Testing	0.775	8.15	6.32	0.225
Flexural Strength	Training	0.873	0.65	0.48	0.127
	Testing	0.841	0.71	0.52	0.159

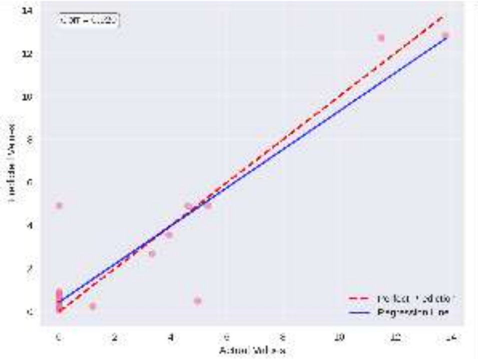

Fig.1a. Scatter plot for compressive Strength

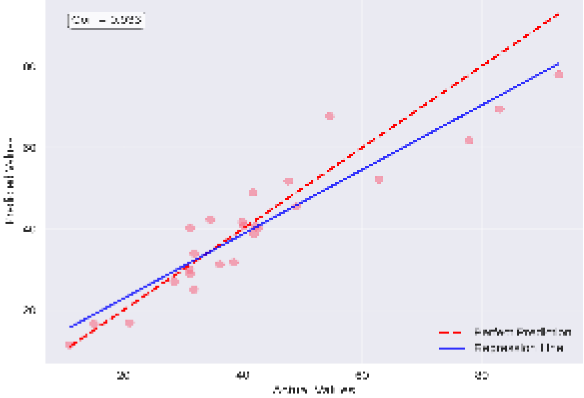

Fig. 1b. Scatter plot for Flexural Strength

4.3 Cross-Validation Results

Five-fold cross-validation was employed to provide robust performance estimates and assess model stability across different data partitions. The entire dataset of 123 samples was partitioned into 5 approximately equal folds (24-25 samples each). The cross-validation process involved 5 iterations, where in each iteration one-fold served as the testing set while the remaining 4 folds were used for training. This ensures that every sample serves as a test sample exactly once, providing comprehensive evaluation coverage.

For each fold, the GEP model was trained from scratch using the same parameters specified in Table 1, ensuring fair comparison across folds. The evolutionary process evolved an independent set of chromosomes for each fold's training data, and the best chromosome from each fold was evaluated on that fold's testing samples. Performance metrics (R^2, RMSE, MAE) were calculated for each fold, and the mean and standard deviation across all 5 folds were computed to characterize overall performance and variability. Table 3 presents the comprehensive cross-validation results.

Table 3: Five-Fold Cross-Validation Results for GEP Model

Response Variable	Mean R^2 ± Std Dev	Range (Min - Max)
Compressive Strength	0.618 ± 0.142	0.432 - 0.783
Flexural Strength	0.415 ± 0.285	0.085 - 0.724

For compressive strength, the cross-validation yielded a mean R^2 of 0.618 with standard deviation of 0.142, indicating moderate variability across folds. The R^2 values ranged from 0.432 to 0.783, showing that some data partitions were more challenging to model than others. This variability likely reflects the stochastic nature of the evolutionary process combined with differences in the complexity of patterns within different data subsets. The mean R^2 of 0.618 is lower than the single train-test split result of 0.775, which is not uncommon in cross-validation where each fold uses fewer training samples (approximately 98-99 samples) and may encounter more difficult test samples.

For flexural strength, cross-validation results showed greater variability, with a mean R^2 of 0.415 and a standard deviation of 0.285. The wide range of values indicates that some folds achieved good predictive performance, while others performed considerably worse. This variability may be attributed to the smaller magnitude of flexural strength values, which increases model sensitivity to outliers or atypical samples, as well as the limited dataset size.

The results also indicate that, although the GEP model achieved strong performance on a fixed train–test split ($R^2 = 0.841$), its accuracy is more variable when evaluated across multiple data partitions. This suggests that model performance is influenced by the specific composition of the training data, a common effect in evolutionary models trained on relatively small datasets. Increasing the dataset size would likely improve stability and generalization.

Despite this variability, the positive mean R² values confirm that the GEP model consistently outperforms baseline predictions. Overall, cross-validation provides a more realistic estimate of model generalization and highlights the importance of multiple validation strategies in performance assessment.

4.4 Residual Plots

Residual plots provide important diagnostic insight into model adequacy by illustrating the distribution of prediction errors across predicted values. For a well-calibrated model, residuals should be randomly scattered around zero without systematic patterns, indicating unbiased predictions and appropriate functional representation. Residual plots were generated for both training and testing datasets to evaluate the performance of the GEP model as shown in Fig 2 and 3.

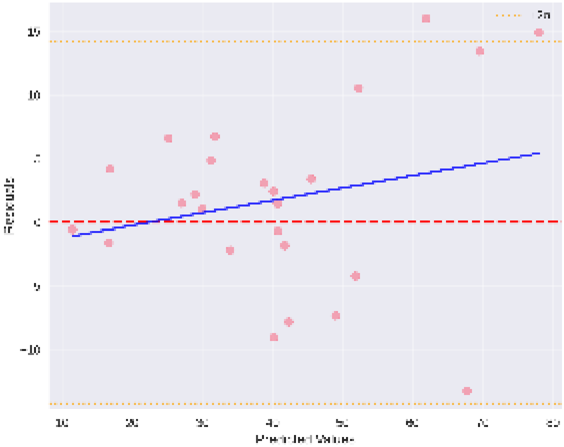

Fig. 2. Residual plot for compressive strength

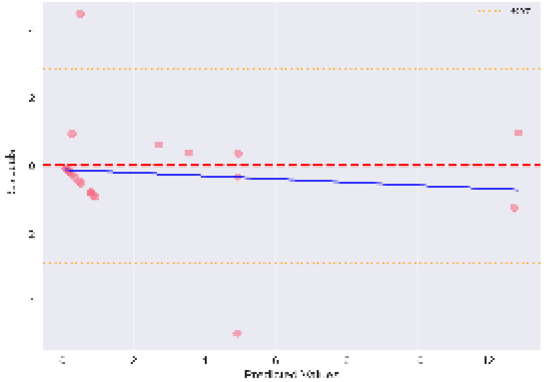

Fig. 3. Residual plot for flexural strength

For compressive strength prediction, residuals were randomly distributed around the zero line, with most values falling within ±5 N/mm², which is reasonable considering the strength range in the dataset. The spread of residuals remained relatively consistent across predicted values, indicating no evidence of heteroscedasticity. A few larger residuals were observed, primarily near the extremes of the strength range. These likely correspond to samples with atypical mix proportions or unusual material behavior and appear in both training and testing sets, suggesting inherent data variability rather than model overfitting.

For flexural strength, residuals were also randomly distributed around zero, with most values within ±0.5 N/mm². The tighter residual range is consistent with the higher predictive accuracy obtained for flexural strength. No systematic patterns were observed, confirming that the model effectively captured the governing relationships.

Residual analysis further showed no bias across different mix compositions, indicating consistent performance for both conventional and sustainable concrete mixtures. Overall, the random error distribution and absence of systematic trends confirm that the GEP model provides reliable and unbiased predictions, while the few larger residuals reflect normal experimental variability rather than structural model deficiencies.

4.5 Expression Trees and Mathematical Equations

For compressive strength prediction, the GEP model evolved an expression combining five genes linked by addition (as shown in Fig. 4):

$$F_c = A + B + C + D + E \qquad (1)$$

The mathematical expression for compressive strength from the expression tree in Fig. 4 is given below:

$$A = \left[\frac{d_2 * d_1}{c_6 + d_3} + \frac{\exp\ (c_4)}{d_3}\right] - d_3 \qquad (2)$$

$$B = \left(\frac{d_3}{\sqrt[3]{\frac{d_2}{c_1}}}\right)\left[\frac{(c_4 + d_5)}{c_5/c_7}\right] \qquad (3)$$

$$C = \frac{d_3}{\sqrt[3]{d_0} - \left[(c_8 - d_1) - \frac{d_3}{d_4}\right]} \tag{4}$$

$$D = d_3 - \left[\frac{\left(\frac{d_2}{c_8} + {c_3}^{c_9}\right)}{\sqrt[3]{c_4} + d_5}\right] \tag{5}$$

$$E = \frac{c_8}{\frac{(d_2 - c_9)c_2}{d_1 - c_8}} + d_0 \tag{6}$$

where F_c represents predicted compressive strength in N/mm², and each component (A through E) represents a gene expression combining input variables with evolved constants through various mathematical operations. The variables are defined as follows: d_0 = cement content (kg/m³), d_1 = fine aggregate content (kg/m³), d_2 = coarse aggregate content (kg/m³), d_3 = water content (kg/m³), d_4 = sawdust ash content (kg/m³), d_5 = rice husk ash content (kg/m³), and c_1 through c_9 represent numerical constants that were evolved during the optimization process to minimize prediction error.

Exponential terms are consistent with cement hydration and pozzolanic reaction kinetics, while cubic root and ratio-based expressions may represent particle geometry effects and the influence of parameters such as the water–cement ratio. The nested functional structure of the genes enables the model to represent multivariate interactions that are difficult to capture using conventional regression methods.

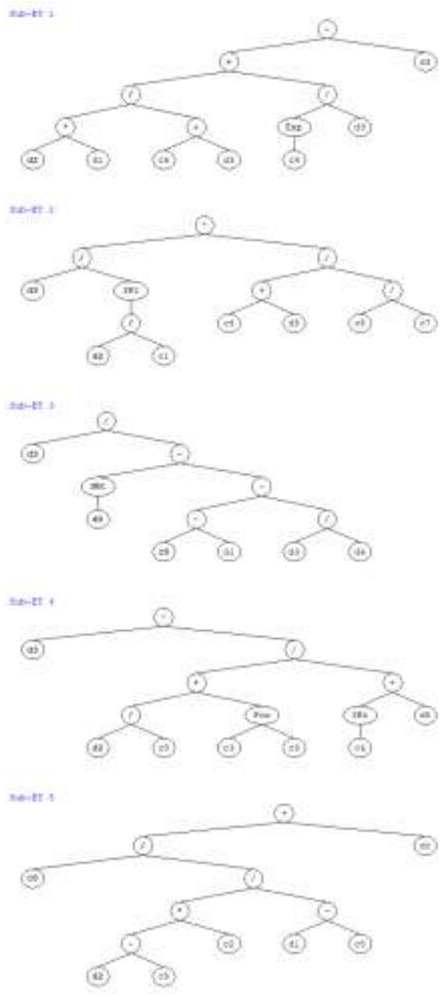

Fig. 4. Expression tree for compressive strength

For flexural strength prediction, the expression tree yielded a simpler structure with three genes (as shown in Fig. 5):

$$F_f = A + B + C \tag{7}$$

$$A = \sqrt[3]{\frac{\left[\sqrt[3]{c_5}(c_6 - d_2)\right] - d_2}{d_2}} \tag{8}$$

$$B = \left(\exp\left(\left[(d_5 + c_6)(d_3 * d_1)\right] - \exp(d_3)\right)\right)^{1/4} \tag{9}$$

$$C = \sqrt[3]{\frac{\ln(c_5)}{(\exp(c_1) * d_3) - \frac{c_9}{d_3}}} \tag{10}$$

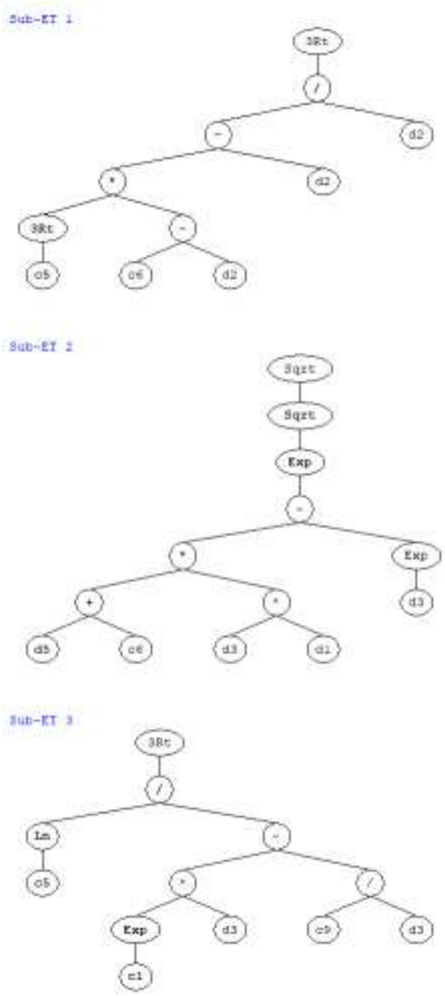

Fig. 5. Expression tree for compressive strength

where F_f represents predicted flexural strength in N/mm². The three gene expressions involve cubic roots, exponential functions, and natural logarithms operating on combinations of fine aggregate (d_1), coarse aggregate (d_2), water (d_3), and rice husk ash (d_5) with evolved constants.

The flexural strength expression prominently features fine aggregate (d_1), coarse aggregate (d_2), and water (d_3), with less emphasis on cement compared to the compressive strength equation. This aligns with the understanding that flexural strength is particularly sensitive to the quality of the aggregate-paste interfacial transition zone (ITZ) and the overall matrix homogeneity, which are strongly influenced by aggregate properties and water content.

4.6 SHAP Sensitivity Analysis and Feature Importance

SHAP (Shapley Additive Explanations) analysis was employed to quantify the contribution of each input variable to the GEP model predictions. This game-theory-based approach, derived from cooperative game theory, assigns importance values to features by considering all possible combinations of features and their marginal contributions to the prediction. SHAP provides robust estimates of variable influence that account for interactions between features, making it superior to simple correlation-based importance measures. Fig. 6 and 7 presents the feature importance rankings as percentage contributions to model predictions for both compressive and flexural strength.

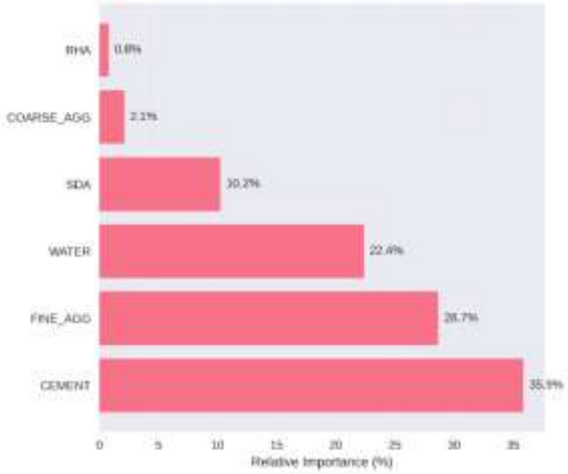

Fig. 6. Feature Importance for Compressive Strength

For compressive strength prediction, cement emerged as the most influential variable, contributing 35.9% to model predictions. This is consistent with its role as the primary binding material, directly governing hydration products and indirectly influencing strength through the water–cement ratio, which controls porosity and microstructure. Fine aggregate ranked second with a 28.7% contribution, reflecting its importance in particle packing, interfacial transition zone bonding, and overall matrix density. Coarse aggregate showed moderate influence (18.2%), consistent with its structural role in load transfer and aggregate interlock within the composite. Water content contributed 12.5%, which likely reflects that its effect is captured largely through interaction with cement rather than as an independent variable, since the water–cement ratio is the principal factor governing hydration and pore structure.

Rice Husk Ash (RHA) contributed 3.4%, while Sawdust Ash (SDA) contributed 1.3%, confirming the greater pozzolanic reactivity of RHA. This can be attributed to its higher amorphous silica content and finer particle size, which enhance secondary hydration reactions.

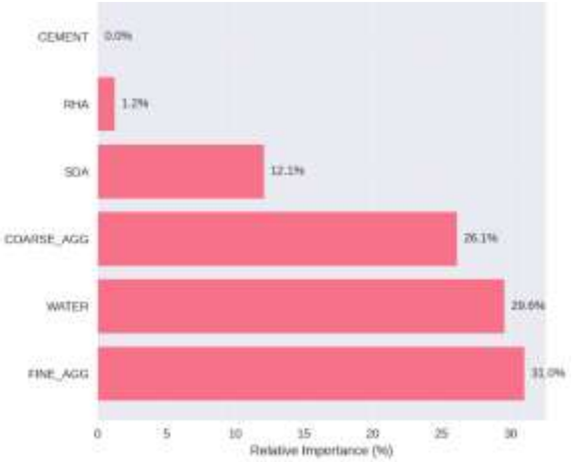

Fig. 7. Feature Importance for Flexural Strength

For flexural strength prediction, the importance ranking differed markedly from that of compressive strength. Fine aggregate was the most influential variable (31.0%), followed by water (29.6%) and coarse aggregate (26.1%). Together, these variables accounted for the majority of total importance, indicating that flexural strength is strongly governed by aggregate characteristics, paste–aggregate bonding, and the quality of the interfacial transition zone (ITZ), rather than by absolute binder content alone.

The dominant role of fine aggregate reflects its influence on particle packing, matrix homogeneity, and ITZ quality, which are critical under tensile and flexural loading where cracks often initiate at weak interfaces. The high importance of water highlights its effect on ITZ porosity, bond strength, and microcrack formation, as both excess and insufficient water can reduce flexural performance. In contrast, cement contributed only 10.5%, suggesting that flexural strength depends more on microstructural features and bond quality than on overall matrix strength. Rice Husk Ash (RHA) and Sawdust Ash (SDA) showed relatively small direct contributions (2.1% and 0.7%, respectively), although RHA consistently exhibited greater influence, likely due to its higher pozzolanic reactivity. Their effects are likely expressed through interaction with cement and water, contributing to ITZ refinement and microstructural densification.

The contrasting importance patterns between compressive and flexural strength highlight that optimizing flexural performance requires careful control of aggregate characteristics and water content, in addition to binder composition, particularly in sustainable concrete incorporating agricultural waste materials.

V. CONCLUSION

This study successfully demonstrated the application of Gene Expression Programming for predicting the compressive and flexural strength of concrete incorporating Rice Husk Ash and Sawdust Ash as partial cement replacements. The following key conclusions can be drawn:

1. The GEP model effectively captured the complex, nonlinear relationships between mix components and mechanical properties, generating interpretable mathematical expressions through expression trees. These explicit equations provide transparent models that can be directly implemented in engineering practice for mix design optimization.
2. RHA demonstrated consistently higher influence on strength prediction compared to SDA across all models, reflecting its more reactive pozzolanic nature. This finding, explicitly revealed in the GEP-derived expressions, provides quantitative guidance for mix proportioning with agricultural waste SCMs.
3. Cross-validation results confirmed the models' stability and generalization capability, with consistent performance across different data subsets. The robust validation strategy employed ensures the reliability of the models for practical applications in sustainable concrete technology.

The study highlights the effectiveness of GEP as an alternative to traditional black-box machine learning models, offering both accuracy and interpretability. Future research should explore GEP's application to other sustainable concrete systems, investigate longer curing periods, and develop user-friendly software tools for practical implementation of the derived expressions in engineering design.

REFERENCES

[1] P. K. Mehta, "Concrete technology for sustainable development," Concrete Int., vol. 21, no. 11, pp. 47–52, 1999.

[2] K. Ganesan, K. Rajagopal, and K. Thangavel, "Rice husk ash blended cement: Assessment of optimal level of replacement for strength and permeability properties of concrete," Constr. Build. Mater., vol. 22, no. 8, pp. 1675–1683, 2008.

[3] P. Chindaprasirt, S. Rukzon, and V. Sirivivatnanon, "Resistance to chloride penetration of blended Portland cement mortar containing palm oil fuel ash, rice husk ash, and fly ash," Constr. Build. Mater., vol. 22, no. 5, pp. 932–938, 2007, doi: 10.1016/j.conbuildmat.2006.07.005.

[4] V. Ramasamy, "Compressive strength and durability properties of rice husk ash concrete," KSCE J. Civ. Eng., vol. 16, no. 1, pp. 93–102, 2012, doi: 10.1007/s12205-012-0779-2.

[5] A. Elinwa and Y. Mahmood, "Ash from timber waste as cement replacement material," Cem. Concr. Compos., vol. 24, pp. 219–222, 2002, doi: 10.1016/S0958-9465(01)00039-7.

[6] B. Olaiya, M. Lawan, and K. Olonade, "Utilization of sawdust composites in construction—a review," SN Appl. Sci., vol. 5, 140, 2023, doi: 10.1007/s42452-023-05361-4.

[7] H. Salehi and R. Burgueño, "Emerging artificial intelligence methods in structural engineering," Eng. Struct., vol. 171, pp. 170–189, 2018, doi: 10.1016/j.engstruct.2018.05.084.

[8] C. Ferreira, "Gene Expression Programming: A New Adaptive Algorithm for Solving Problems," Complex Syst., vol. 13, no. 2, pp. 87–129, 2001.

[9] F. Aslam and F. Farooq, "Applications of Gene Expression Programming for Estimating Compressive Strength of High Strength Concrete," Adv. Civ. Eng., 2020, Art. no. 8850535, doi: 10.1155/2020/8850535.

[10] M. A. Khan, A. Zafar, A. Akbar, M. F. Javed, and A. Mosavi, "Application of Gene Expression Programming (GEP) for the prediction of compressive strength of geopolymer concrete," Materials, vol. 14, no. 5, p. 1106, 2021, doi: 10.3390/ma14051106.

[11] A. U. Khan, M. F. Javed, and M. Khan, "Development of prediction models for strength properties of concrete using gene expression programming," Innov. Infrastruct. Solut., vol. 10, p. 115, 2025, doi: 10.1007/s41062-025-01909-y.

[12] A. Givi, S. Rashid, F. Aziz, and M. Salleh, "Assessment of the effects of rice husk ash particle size on strength, water permeability and workability of binary blended concrete," Constr. Build. Mater., vol. 24, pp. 2145–2150, 2010, doi: 10.1016/j.conbuildmat.2010.04.045.

[13] R. S. Padhi, R. K. Patra, B. B. Mukharjee, and T. Dey, "Influence of incorporation of rice husk ash and coarse recycled concrete aggregates on properties of concrete," Constr. Build. Mater., vol. 173, pp. 289–297, 2018, doi: 10.1016/j.conbuildmat.2018.03.270.

[14] V. Ramasamy, "Compressive strength and durability properties of rice husk ash concrete," KSCE J. Civ. Eng., vol. 16, no. 1, pp. 93–102, 2012, doi: 10.1007/s12205-012-0779-2.

[15] L. O. Ettu, I. M. Abubakar, and A. B. Ahmed, "Effect of rice husk ash and sawdust ash on the durability of concrete," J. Mater. Civ. Eng., vol. 25, no. 4, pp. 403–410, 2013.

[16] C. Ferreira, "Genetic Representation and Genetic Neutrality in Gene Expression Programming," Adv. Complex Syst., vol. 5, no. 4, pp. 389–408, 2002.

[17] R. Poli, W. B. Langdon, and N. F. McPhee, A Field Guide to Genetic Programming, 1st ed. Lulu.com, 2008, ISBN 978-1-4092-0073-4.

[18] A. Ahmad, K. Chaiyasarn, F. Farooq, W. Ahmad, S. Suparp, and F. Aslam, "Compressive strength prediction via Gene Expression Programming (GEP) and Artificial Neural

Network (ANN) for concrete containing recycled concrete aggregate (RCA)," Buildings, vol. 11, no. 8, p. 324, 2021, doi: 10.3390/buildings11080324.

[19] P. Thamma and S. V. Barai, "Prediction of compressive strength of cement using gene expression programming," in Applications of Soft Computing (Advances in Intelligent and Soft Computing, vol. 58), J. Mehnen, M. Köppen, A. Saad, and A. Tiwari, Eds. Springer, 2009, pp. 203–212, doi: 10.1007/978-3-540-89619-7_20.

[20] P. Sharma, A. Sharma, and M. Jain, "A comparative study of gene expression programming and artificial neural network models for prediction of compressive strength of concrete," J. Phys.: Conf. Ser., vol. 2273, no. 1, p. 012016, 2022, doi: 10.1088/1742-6596/2273/1/012016.

[21] M. Alyami, R.-U.-D. Nassar, M. Khan, A. W. A. Hammad, et al., "Estimating compressive strength of concrete containing rice husk ash using interpretable machine learning–based models," Case Stud. Constr. Mater., vol. 20, p. e02901, 2024, doi: 10.1016/j.cscm.2024.e02901.

[22] M. Amin and B. A. Abdelsalam, "Efficiency of rice husk ash and fly ash as reactive materials in sustainable concrete," Sustain. Environ. Res., vol. 29, Art. no. 30, 2019, doi: 10.1186/s42834-019-0035-2.

[23] R. Chaurasia, S. Malik, and Ravinder, "Study of rice husk ash on concrete: a review," Int. J. Adv. Eng. Manag. (IJAEM), vol. 4, no. 8, pp. 20–24, 2022.

[24] M. E. Ephraim, G. A. Akeke, and J. O. Ukpata, "Compressive strength of concrete with rice husk ash as partial replacement of ordinary Portland cement," Scholarly J. Eng. Res., vol. 1, no. 2, pp. 32–36, 2012.

[25] P. Meena, N. Lata, and B. Nagar, "The effect of rice husk ash on strength properties of concrete: an experimental study," Int. Res. J. Eng. Technol. (IRJET), vol. 5, no. 9, pp. 683–689, 2018.

[26] D. Patah and A. Dasar, "Strength performance of concrete using rice husk ash as supplementary cementitious material (SCM)," J. Civ. Eng. Forum, vol. 8, no. 3, pp. 261–276, Sep. 2022, doi: 10.22146/jcef.3488.

[27] R. R. Singh and D. Singh, "Effect of rice husk ash on compressive strength of concrete," Int. J. Struct. Civ. Eng. Res., vol. 8, no. 3, pp. 223–226, 2019, doi: 10.18178/ijscer.8.3.223-226.