

The Nigerian Association of Mathematical Physics

Journal homepage: <https://nampjournals.org.ng>



Modelling the Compressive Strength of Costus afer-Modified Fly Ash Geopolymer Concrete as a Function of Curing Temperature, Activator Ratio, and NaOH Concentration

^{1,*}Ogigbo, O. R., ²Iyeke, S. D. and ³Kennedy C.

^{1,2,3}Department of Civil Engineering, Faculty of Engineering, University of Benin, Benin City, Edo State, Nigeria.

ARTICLE INFO

Article history:

Received xxxxx
Revised xxxxx
Accepted xxxxx
Available online
xxxxxx

Keywords:

Water saturation,
ensemble model,
neural network,
well log,
MATLAB,
core data,
learning, reservoir

ABSTRACT

The use of geopolymer concrete as a sustainable alternative to ordinary Portland cement concrete has increased, but predicting its compressive strength remains challenging due to complex interactions among mix and curing parameters. This study developed Multilinear Regression (MLR) and Random Forest (RF) models to predict the compressive strength of Costus afer-modified fly ash geopolymer concrete using curing temperature, activator ratio, and sodium hydroxide concentration. Model performance was evaluated using R^2 , MAE, MSE, and MAPE. The RF model outperformed MLR, achieving R^2 values of 0.9898 and 0.9882 for training and testing, respectively, with testing MAE, MSE, and MAPE of 1.3396 MPa, 2.2104 MPa², and 6.11%. Feature importance analysis identified curing temperature as the most influential parameter, followed by activator ratio, while sodium hydroxide concentration had the least effect. The RF model offers a reliable tool for strength prediction and mixture optimization.

1. Introduction

Concrete is the most widely used construction material due to its versatility, durability, and low cost. However, ordinary Portland cement (OPC) production is highly carbon-intensive, contributing 5–8% of global anthropogenic CO₂ emissions [1–3]. To reduce this footprint, supplementary cementitious materials partially replace OPC, but they do not eliminate dependence on it. This has driven interest in alkali-activated materials and geopolymer binders, which use aluminosilicate-rich industrial by-products to produce low-carbon construction materials [4].

*Corresponding author: **Ogigbo, O. R.**

E-mail address: okiemute.ogirigbo@uniben.edu, +234(0)8136962171

DOI: <https://doi.org/10.60787/jnamp.vol73no.755>

1118-4388 © 2024 JNAMP. All rights reserved

Unlike OPC, geopolymers form through alkali activation, where silica and alumina dissolve in alkaline environments and polymerize into a three-dimensional aluminosilicate network (N-A-S-H gel), eliminating the need for clinker [5, 6]. Low-calcium (Class F) fly ash is a particularly suitable precursor due to its high reactive silica and alumina content, favourable particle morphology, and availability. Its use also reduces landfill disposal. Life-cycle assessments confirm geopolymer concrete generally has a lower global warming potential than OPC concrete, though benefits depend on activator production and curing conditions [7, 8]. The engineering performance of fly ash-based geopolymer concrete is governed by precursor characteristics, alkaline activator composition, and curing conditions. The sodium silicate-to-sodium hydroxide ratio (activator ratio), NaOH concentration, and curing temperature are primary variables controlling compressive strength [9, 10]. The activator ratio regulates soluble silica for polymer chain development; insufficient silica limits polymerization, while excess increases viscosity and reduces workability. NaOH concentration controls dissolution rates; high concentrations accelerate early strength but can cause workability loss and microstructural defects. Heat curing accelerates geopolymerization and early-age strength, but excessive temperatures may induce moisture loss and microcracking [11, 12]. These factors interact nonlinearly, making optimization challenging.

To improve sustainability, agricultural residues such as rice husk ash, sugarcane bagasse ash, and palm oil fuel ash are being investigated as supplementary precursors [13, 14]. Among these, *Costus afer* ash—abundant in tropical regions—has gained interest. Controlled calcination produces reactive silica and alumina that can enhance particle packing and strengthen the geopolymer matrix [15]. However, its effectiveness depends on composition, fineness, calcination conditions, replacement level, and activator parameters. Given the complexity of these interactions, predictive modelling is essential. Multiple Linear Regression (MLR) offers simplicity and interpretability but assumes linear relationships, limiting its accuracy for geopolymer systems [16, 17]. Machine learning techniques like Random Forest (RF) capture nonlinear interactions and variable importance, making them well-suited for modelling complex cementitious materials [18, 19].

Despite advances, knowledge gaps remain: few studies examine combined effects of curing temperature, activator ratio, and NaOH concentration, and limited research exists on *Costus afer* ash in fly ash-based geopolymers. Existing models may not apply to mixtures with this additive, and systematic comparisons between MLR and RF are scarce. This study addresses these gaps by investigating the combined influence of curing temperature, activator ratio, and NaOH concentration on the compressive strength of *Costus afer*-modified fly ash geopolymer concrete. Using a controlled experimental dataset, MLR and RF models are developed and compared to evaluate predictive performance and characterize variable interactions. The findings aim to support the optimization of sustainable geopolymer mixtures and provide reliable predictive tools for environmentally responsible concrete design.

2. Methodology

2.1 Materials

2.1.1 Fly ash (FA)

The fly ash used for the study was Class F, in accordance with ASTM C618-19 [20]. It was obtained locally from a vendor in Port Harcourt, Rivers State, Nigeria. The raw fly ash was ground to finer particles and sieved with a 75 microns sieve, to ensure uniform particle size distribution. The particle size distribution of the fly ash used in the study is shown in Figure 1, while the chemical composition as obtained from X-ray fluorescence (XRF) is shown in Table 1.

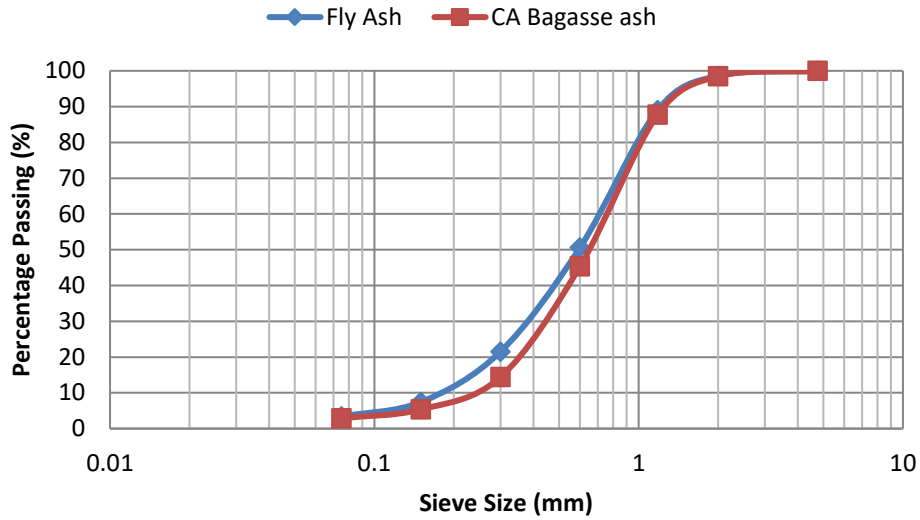


Figure 1: Particle size distribution of fly ash and *Costus afer* bagasse ash

2.1.2 *Costus afer* (CA) Bagasse ash

The *Costus afer* plant used in this study was collected from forests in Odiokwu village, which is located in the Ubie District of Ekpeye land, Ahoada West Local Government Area, Rivers State, Nigeria. After harvesting, the plant was carefully taken to the Civil Engineering lab at Ken Saro-Wiwa Polytechnic in Bori, Rivers State, where it underwent processing into ash. The process involved peeling of the fibers from the stalks. Thereafter, the stems were crushed into smaller bits to break down the fibrous material. Next, these crushed pieces were burned to ash in an electric furnace set to 600°C for about two hours, based on the method described by Palanisamy [21]. The complete process is shown in Figure 2. After all these, the ash was passed through a 75 μm sieve to make sure all particles were the same size. The particle size distribution of the *Costus afer* bagasse ash used in the study is shown in Figure 1, while the chemical composition as obtained from X-ray fluorescence (XRF) is shown in Table 1.

Table 1: Chemical composition of Fly ash and *Costus afer* bagasse ash

Chemical property	Percentage (%)	
	FA	CA
Silicate oxide (SiO_2)	35.61	69.51
Aluminium oxide (Al_2O_3)	14.89	1.93
Ferric oxide (Fe_2O_3)	21.81	0.55
Calcium oxide (CaO)	7.85	1.83
Sulphur oxide (SO_3)	2.81	0.42
Others	14.94	25.76
Loss on ignition (LOI)	1.92	9.49
(SiO_2) + (Al_2O_3) + (Fe_2O_3)	72.31	71.97



Figure 2: (a) Harvested *Costus afer* plant (b) Peeled and cut stems of *Costus afer* plant (c) Pulverized *Costus afer* fibre (d) *Costus afer* ash

2.1.3 Aggregates

The aggregates used in the study, which was comprised of fine and coarse aggregates, was obtained from local quarry sites in Port Harcourt and environs. The aggregates were treated by removal of impurities, and drying under air to ensure that the moisture content conforms with the specifications given in BS EN 12620:2013 [22]. The particle size distribution of the aggregates and their physical properties are shown in Figure 3 and Table 2 respectively.

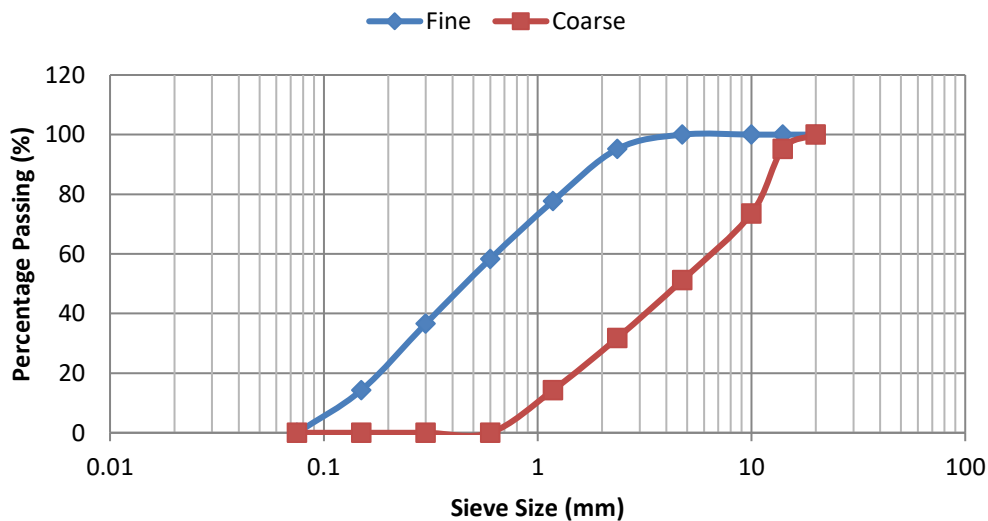


Figure 3: Particle size distribution for aggregates

Table 2: Physical properties of aggregates

Property	Coarse Aggregate	Fine Aggregate
Specific Gravity	2.65	2.55
Bulk Density (kg/m ³)	1600	1450
Water Absorption (%)	1.5	2.0
Fineness Modulus	6.5	2.8
Maximum Aggregate Size (mm)	20	4.75

2.1.4 Alkaline activators

In this study, two types of alkaline activators were used: sodium hydroxide (NaOH) and sodium silicate (Na₂SiO₃). Both were bought from local chemical suppliers in Port Harcourt, Rivers State, Nigeria. The sodium hydroxide came as pellets with a purity level of 97%, meeting the requirements of BS EN 15743:2010+A1:2015 [23]. These pellets were mixed with distilled water to prepare sodium hydroxide solutions at different concentrations – 8M, 10M, 12M, 14M, and 16M. The sodium silicate solution was prepared using the same approach. Once both solutions were ready, they were combined in various mixing ratios, referred to in this study as the activator ratio. The ratios used in the mixing were 0.5, 1.0, 1.5, 2.0, and 2.5. The final blended alkaline solutions were then used to make the geopolymer concretes.

2.2 Mix Design and Proportions

The mix design adopted in this study was based on a modified conventional concrete proportioning approach, tailored to accommodate the unique characteristics of geopolymer binder systems. A base mix ratio of 1:1.38:2.82 (binder: fine aggregate: coarse aggregate) by weight was adopted for this study. When adjusted for differences in specific gravity and packing behaviour between geopolymer materials and conventional cement-based systems, this proportion closely aligns with the commonly used 1:2:4 mix [24]. The adjustment was necessary to ensure that the mix maintained suitable workability, good particle distribution, and satisfactory mechanical properties within the geopolymer system. *Costus afer* bagasse ash was used to replace fly ash, which is the main precursor for the geopolymer, at various percentages of 5, 10, 15 and 20% by weight of fly ash.

2.3 Concrete Batching and Curing

For the preparation of the geopolymer concrete, the dry binder constituents – fly ash and *Costus afer* bagasse ash – were first combined in a mechanical mixer to achieve a homogeneous distribution of the bagasse ash throughout the mixture. Subsequently, the alkaline activator solution was introduced gradually, with continuous mixing, until a uniform and workable paste was obtained. Following the formation of a consistent paste, the fine and coarse aggregates were incorporated into the mix. Mixing was maintained to ensure complete coating of all aggregate particles and to preserve uniformity across the entire batch. The fresh concrete was then placed into moulds and compacted to eliminate entrapped air and secure adequate consolidation.

Thereafter, the geopolymer concrete specimens were subjected to a two-step curing regime, as described by Nath [25] and Ahmed [26]. Immediately after casting, the specimens were placed in a preheated oven for a duration of 24 hours. A range of curing temperatures was employed in the study to evaluate their effect on compressive strength development. These temperatures included ambient temperature (27°C), as well as 60°C, 70°C, 80°C, 90°C, and 100°C. Upon completion of the initial high-temperature curing period, the specimens were removed from the oven, demoulded, and wrapped in a waterproof covering to minimise moisture loss. They were subsequently transferred to a fog room with elevated humidity, which provided a consistently moist environment to support ongoing reactions within the material. The specimens remained under these conditions until testing at the age of 28 days.

2.4 Test Methods

2.4.1 Compressive Strength

Compressive strength testing of the geopolymer concrete specimens was conducted following the guidelines set out in BS EN 12390-3:2019 [27]. On the designated test day, the necessary samples were taken out of the fog room, dried off with a clean cloth, and positioned for testing. Each cube was then placed under a steadily rising compressive load until it failed, and the highest load it withstood was noted. A compression testing machine with a range of 0 to 2000 kN was used for all tests.

2.4.2 Modelling

Data Set Analysis

The modelling process considered five independent variables, namely fly ash content, *Costus afer* (CA) bagasse ash content, curing temperature, sodium hydroxide (NaOH) molarity, and activator ratio. The 28-day compressive strength was adopted as the response (dependent) variable since it is the standard indicator used to evaluate the structural performance of concrete.

The descriptive statistics of all the variables employed in the modelling are presented in Table 3. The fly ash content varied between 334.16 and 417.70 kg/m³, while the *Costus afer* bagasse ash content ranged from 0 to 62.66 kg/m³, reflecting the different levels of cementitious material replacement considered in the experimental programme. The curing temperature ranged from 27°C, representing ambient curing conditions, to 100°C, thereby capturing the influence of both ambient and elevated-temperature curing. Similarly, the concentration of the sodium hydroxide solution varied from 8 M to 16 M, whereas the activator ratio was adjusted within the range of 0.5 to 2.5 to investigate its effect on the geopolymerisation process.

To examine the degree of association between the response variable and the selected predictors, a heat map correlation matrix was generated, as presented in Figure 4. The analysis revealed that curing temperature exhibited the strongest positive relationship with the 28-day compressive strength, with a correlation coefficient of 0.81, indicating that curing temperature played a dominant role in determining the strength development of the geopolymer concrete mixtures. In contrast, fly ash content, *Costus afer* bagasse ash content, sodium hydroxide molarity, and activator ratio showed relatively weak linear correlations with compressive strength. This observation suggests that, although these variables contribute to the overall behaviour of the material, their individual effects on compressive strength are less pronounced when assessed using simple pairwise correlation. Their influence may instead arise through nonlinear relationships or interactions with other variables, thereby justifying the application of advanced machine learning techniques capable of capturing such complex relationships.

Table 3: Descriptive statistics of the input parameters

Variable	Count	Min.	Max.	Average	Std. Dev.
Fly ash (kg/m ³)	70	334.16	417.70	375.93	29.749
CA bagasse ash (kg/m ³)	70	0	83.54	41.772	29.749
Curing temperature (°C)	70	27	100	76.214	16.170
NaOH solution (M)	70	8	16	10.714	1.957
Activator ratio	70	0.5	2.5	1.821	0.489
Compressive strength (MPa)	70	11.72	52.71	38.575	9.167

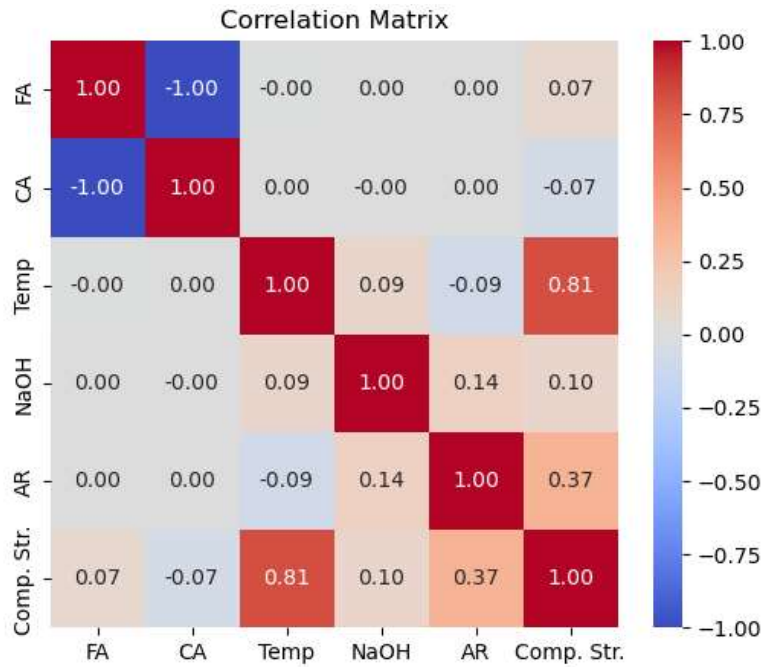


Figure 4: Heat map showing the correlation coefficients between all the variables

Train-Test Split

Before model development, the complete dataset, comprising 70 experimental observations, was randomly partitioned into training and testing subsets to enable both model development and independent performance evaluation. A total of 63 observations (90% of the dataset) were assigned to the training set, while the remaining 7 observations (10%) constituted the testing set. The training dataset was used to fit the models by estimating the regression parameters for the Multiple Linear Regression (MLR) model and training the Random Forest (RF) algorithm. In contrast, the testing dataset, which was not exposed to the models during the training phase, was reserved exclusively for assessing their predictive performance on previously unseen data. This approach helps to provide an unbiased evaluation of the models' generalization capability and reduces the likelihood of overestimating prediction accuracy.

The predictive performance of the developed models was assessed by comparing the predicted 28-day compressive strengths with the corresponding experimentally measured values. Four widely accepted statistical performance metrics were employed for this purpose: the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Square Error (MSE), and Mean Absolute Percentage Error (MAPE). A model was considered to exhibit satisfactory predictive performance if it produced a high coefficient of determination together with relatively low values of MAE, MSE, and MAPE. Collectively, these evaluation metrics provide complementary information on the goodness-of-fit, prediction accuracy, and robustness of the developed models and have been widely adopted in studies involving the prediction of mechanical properties of concrete and other cementitious materials.

Model development

The present study employed Multiple Linear Regression (MLR) and Random Forest (RF) to model the compressive strength of *Costus afer*-modified fly ash geopolymer concrete. The selection of these two modelling approaches was motivated by their complementary strengths and their widespread application in predictive modelling of engineering materials.

MLR was selected as the baseline model because it provides a straightforward and interpretable mathematical relationship between the response variable and the predictor variables [16, 17]. The model

quantifies the individual contribution of each explanatory variable through regression coefficients, making it possible to assess the direction and magnitude of the influence of fly ash content, *Costus afer* bagasse ash content, curing temperature, sodium hydroxide concentration, and activator ratio on compressive strength. In addition, MLR requires relatively low computational effort and serves as an established benchmark against which more advanced predictive algorithms can be evaluated [28, 29]. Although the method assumes a linear relationship between the predictors and the response, its simplicity and ease of interpretation make it an appropriate starting point for predictive modelling.

On the other hand, random forest was selected because the behaviour of geopolymer concrete is governed by complex physicochemical processes involving nonlinear interactions among constituent materials and curing conditions [30]. Such interactions are often difficult to capture adequately using conventional linear regression models. Random Forest (RF) is an ensemble learning algorithm that constructs multiple decision trees using bootstrap samples of the training data and combines their predictions through aggregation, thereby improving predictive accuracy and reducing the risk of overfitting [31, 32]. The algorithm is capable of modelling nonlinear relationships, handling interactions among variables without prior specification, accommodating correlated predictors, and providing measures of variable importance. These characteristics make it particularly suitable for modelling the behaviour of cementitious materials, where the effects of mixture proportions and curing parameters are rarely purely linear [33–35].

The combined use of MLR and RF enabled a comprehensive evaluation of the predictive problem from both statistical and machine learning perspectives. While MLR provides a transparent and easily interpretable analytical model, RF offers greater flexibility in capturing complex relationships embedded in the experimental data.

The general form of the MLR model as used in this study is expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon$$

where Y denotes the predicted compressive strength, β_0 is the intercept, $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are the regression coefficients associated with the predictor variables (X_1, X_2, X_3, X_4, X_5), and ϵ represents the random error term.

The RF regression model is expressed as:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B T(x; \theta_b)$$

where B represents the number of trees grown, and $T(x; \theta_b)$ denotes the prediction from the b^{th} tree grown on bootstrap sample θ_b . The RF regression was implemented in Python using the RandomForestRegressor function available in the scikit-learn library. Given the relatively small size of the dataset, the number of decision trees in the ensemble was set to 10. All other hyperparameters were retained at their default values as provided by the library.

3. Results and Discussion

3.1 Multilinear Regression (MLR)

The performance statistics of the MLR model (see Table 4) indicate that the model achieved a training R^2 of 0.8248 and a testing R^2 of 0.9331. These values suggest that the linear model was able to explain

approximately 82.5% of the variation in the training dataset and 93.3% of the variation in the testing dataset. The corresponding MAE values of 2.7784 MPa and 2.8478 MPa for the training and testing datasets, respectively, indicate moderate prediction errors, while the MSE values of 11.7742 and 12.5568 confirm that some deviations between predicted and measured compressive strengths were relatively large. Likewise, the MAPE increased from 7.62% in the training phase to 15.52% in the testing phase, implying a reduction in prediction accuracy when the model was applied to unseen data.

Table 4: Performance of MLR model

Performance metric	Training dataset	Test dataset
MAE	2.7784	2.8478
R ²	0.8248	0.9331
MSE	11.7742	12.5568
MAPE	0.0762	0.1552

The regression equation obtained for the model was:

Compressive Strength

$$= 166586546693425.72 - 398818641832.4994(FA) - 398818641832.5258(CA) + 0.4595(Temp) - 01267(NaOH) + 8.5532(AR)$$

Among the process variables, the activator ratio (AR) had the largest positive coefficient (+8.5532), indicating that an increase in the sodium silicate-to-sodium hydroxide ratio contributed most strongly to the increase in compressive strength within the linear framework of the model. This observation agrees with the established understanding that higher activator ratios promote greater dissolution of aluminosilicate species and enhance geopolymer gel formation [36–39]. The curing temperature also had a positive coefficient (+0.4595), showing that elevated temperatures accelerated geopolymerization and improved strength development, although its linear effect was less pronounced than that of the activator ratio.

The negative coefficient for NaOH concentration (−0.1267) suggests that increasing the molarity beyond the optimum range reduced compressive strength. Similar behaviour has been reported in geopolymer systems where excessive alkali concentration leads to rapid precipitation of reaction products, incomplete polycondensation, and increased microstructural defects [38, 39]. The extremely large negative coefficients associated with FA and CA are not physically meaningful because both variables remained essentially constant throughout the experimental programme; the coefficients therefore reflect multicollinearity rather than actual material effects.

The predicted-versus-actual plots for both the training and testing datasets (see Figure 5 and 6) show that the MLR model was able to capture the general trend of the experimental data. However, noticeable deviations from the line of perfect agreement occurred at higher compressive strength values, indicating that the linear assumption was insufficient to fully describe the behaviour of the geopolymer system. The residual plots (see Figure 7 and 8) also exhibited a wider dispersion around zero, with several residuals extending beyond ± 5 MPa, confirming the presence of larger prediction errors.

These findings are consistent with recent studies on geopolymer concrete modelling. Parhi [40] observed that conventional regression models produced lower predictive accuracy than nonlinear ensemble algorithms when modelling geopolymer concrete strength because the relationships among activator concentration, curing conditions, and precursor composition are inherently nonlinear. Similarly, Rathnayaka [41] concluded that linear regression techniques generally underperform compared with ensemble learning approaches when predicting the compressive strength of fly ash-based geopolymer concrete due to their inability to capture complex variable interactions.

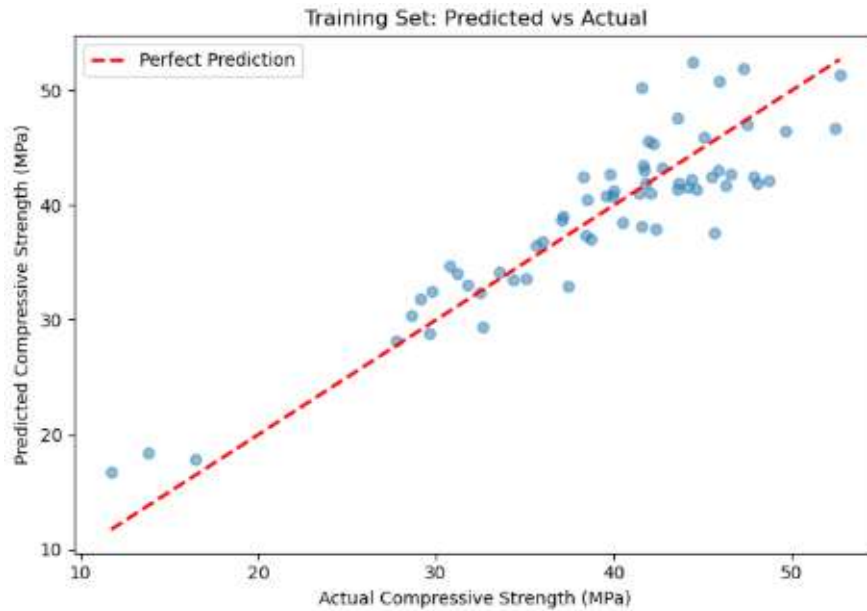


Figure 5: Plot showing comparison between actual compressive strength and predicted compressive strength for the training dataset of the MLR model

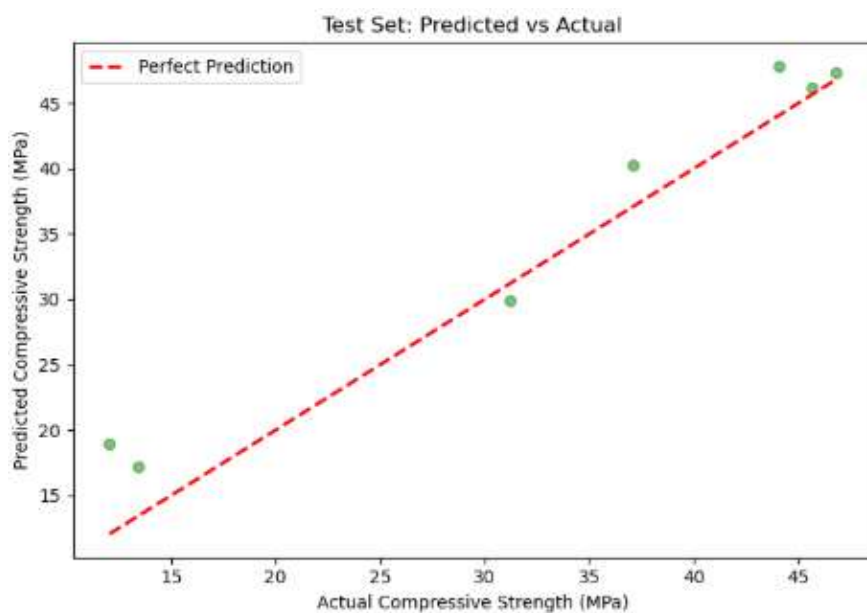


Figure 6: Plot showing comparison between actual compressive strength and predicted compressive strength for the test dataset of the MLR model



Figure 7: Plot showing the residuals i.e. the difference between the actual compressive strength and the predicted compressive strength for the training dataset of the MLR model

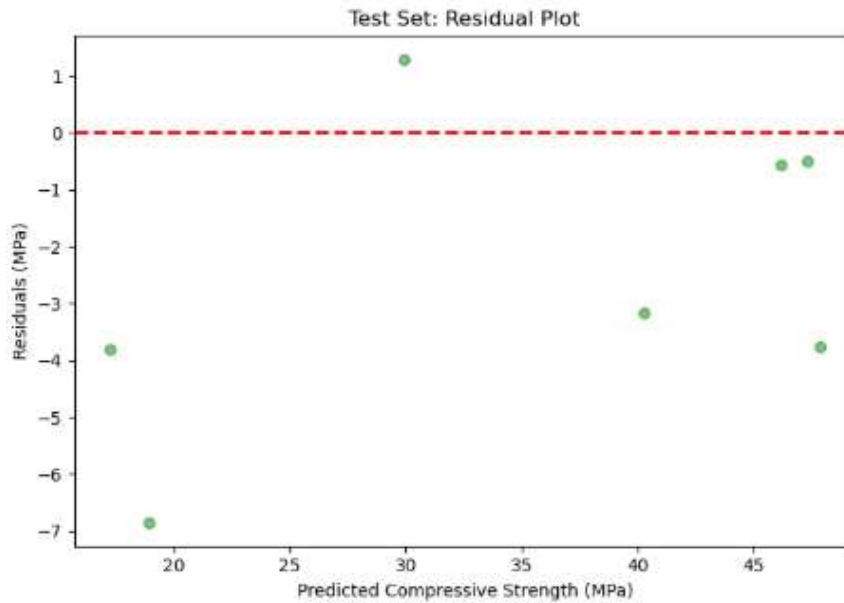


Figure 8: Plot showing the residuals i.e. the difference between the actual compressive strength and the predicted compressive strength for the test dataset of the MLR model

3.2 Random Forest Regression

The Random Forest (RF) model demonstrated substantially superior predictive performance (see Table 5). The model achieved R^2 values of 0.9898 for the training dataset and 0.9882 for the testing dataset, indicating that approximately 99% of the variability in compressive strength was explained by the model. The MAE values decreased to 0.6165 MPa and 1.3396 MPa for the training and testing datasets, respectively, while the MSE values were reduced to 0.6856 and 2.2104. Furthermore, the MAPE values of 1.79% and 6.11% indicate highly accurate predictions with very small percentage errors.

Table 5: Performance of RF regression model

Performance metric	Training dataset	Test dataset
MAE	0.6165	1.3396
R^2	0.9898	0.9882
MSE	0.6856	2.2104
MAPE	0.0179	0.0611

The close agreement between the training and testing statistics indicates that the RF model generalized well and did not suffer from significant overfitting. The predicted-versus-actual plots (shown in Figure 9 and 10) showed that the data points clustered tightly around the 45° line for both datasets, demonstrating excellent agreement between measured and predicted strengths. The residual plots (shown in Figure 11 and 12) exhibited a narrow distribution around zero, with most residuals confined within approximately ± 2 MPa, indicating a stable and unbiased predictive behaviour.

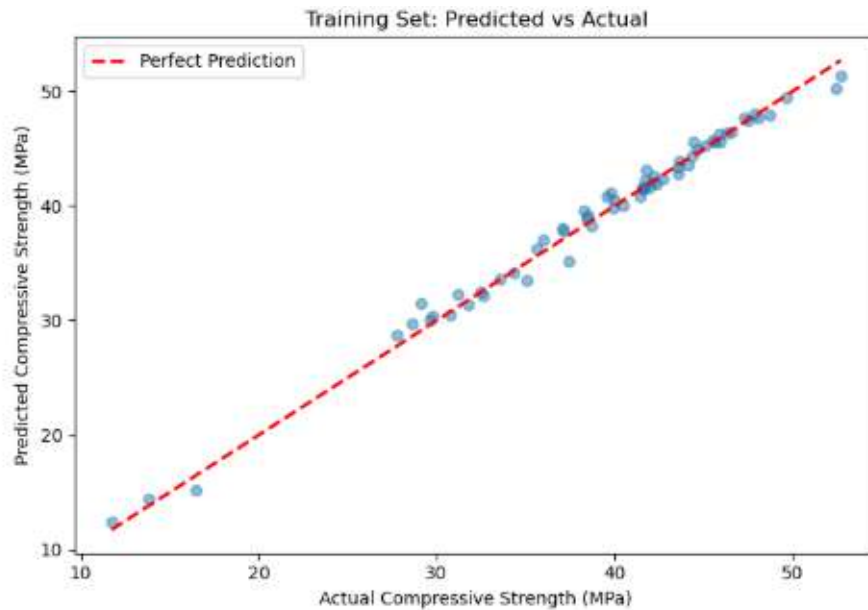


Figure 9: Plot showing comparison between actual compressive strength and predicted compressive strength for the training dataset of the RF regression model

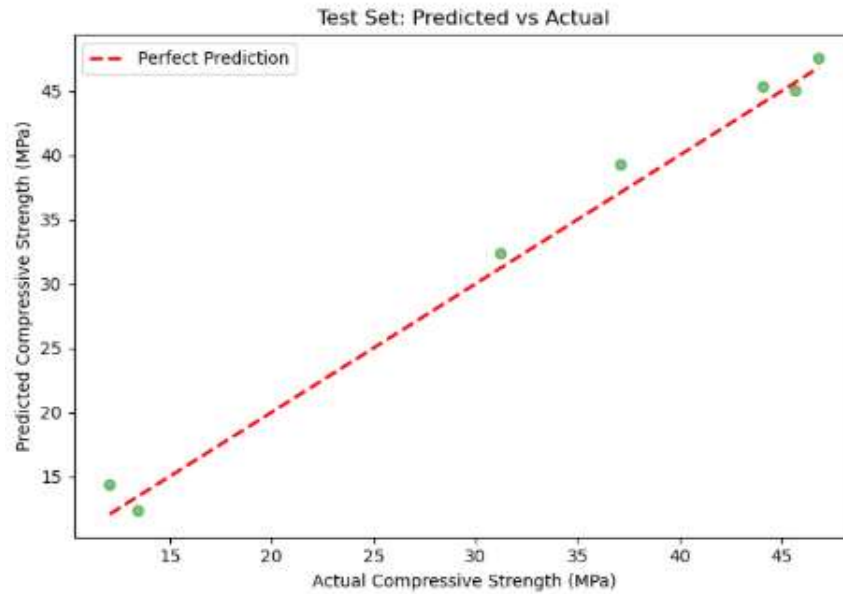


Figure 10: Plot showing comparison between actual compressive strength and predicted compressive strength for the test dataset of the RF regression model



Figure 11: Plot showing the residuals i.e. the difference between the actual compressive strength and the predicted compressive strength for the training dataset of the RF regression model

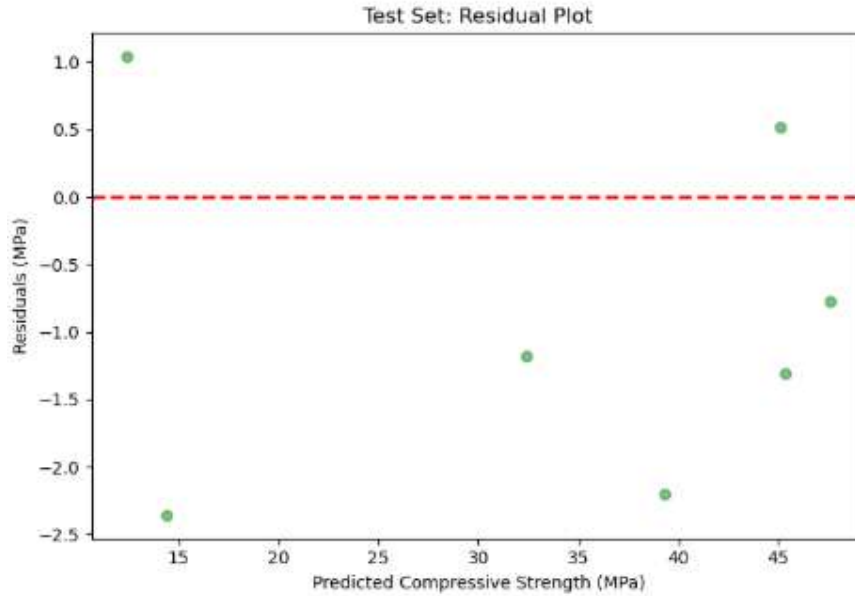


Figure 12: Plot showing the residuals i.e. the difference between the actual compressive strength and the predicted compressive strength for the test dataset of the RF regression model

The feature importance (see Figure 13) analysis revealed that curing temperature was the most influential variable, accounting for approximately 60% of the total model importance. The activator ratio contributed about 28%, while fly ash and *Costus afer* contents contributed roughly 7% and 4%, respectively. NaOH concentration had the smallest contribution, approximately 2%. This ranking indicates that curing temperature governs the nonlinear strength development process more strongly than any other variable.

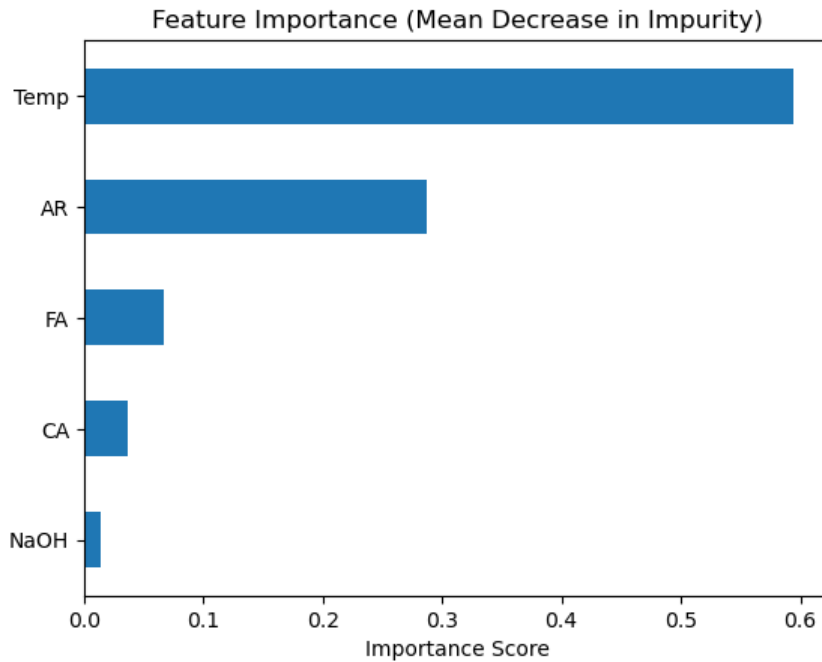


Figure 13: Plot showing the importance scores of the features of the RF regression model

The dominance of curing temperature agrees with recent investigations. Hu [42] identified curing temperature as one of the key factors controlling the compressive strength of fly ash-based geopolymer

concrete, together with activator modulus and Si/Al ratio. Their machine learning analysis showed that temperature had a stronger influence than several mix composition variables. Similarly, Abdellatief [43] reported that Random Forest models outperformed traditional regression approaches when predicting ultra-high-performance geopolymer concrete strength because ensemble algorithms effectively captured nonlinear interactions among curing conditions and binder chemistry.

The excellent performance of the RF model in the present study also agrees with the findings of Bin [44], who demonstrated that optimized Random Forest models significantly improved the prediction accuracy of geopolymer composite compressive strength compared with simpler statistical models.

3.3 Comparison between MLR and RF

A direct comparison of the two models (shown in Table 6) clearly demonstrates the superiority of the Random Forest approach.

Table 6: Comparison of the performance of the MLR and RF regression models

Performance metric (test data)	MLR	RF
MAE	2.8478	1.3396
R ²	0.9331	0.9882
MSE	12.5568	2.2104
MAPE	0.1552	0.0611

The RF model increased the coefficient of determination by approximately 5.9 percentage points relative to the MLR model. More importantly, the prediction errors were reduced substantially: the MAE decreased by about 53%, the MSE decreased by about 82%, and the MAPE decreased by about 61%. These improvements indicate that the RF model not only explained a greater proportion of the strength variability but also produced much more accurate individual predictions.

The difference in variable importance between the two models is particularly noteworthy. The MLR model identified the activator ratio as the dominant variable because it had the largest positive regression coefficient. In contrast, the RF model identified curing temperature as the most influential predictor. This discrepancy arises because regression coefficients represent only linear effects, whereas Random Forest feature importance reflects both nonlinear effects and interactions among variables. In geopolymer systems, curing temperature interacts strongly with activator chemistry and precursor dissolution kinetics, making its overall influence much larger than that suggested by a purely linear model.

The present findings are consistent with the broader literature. Rathnayaka [41] concluded that ensemble learning techniques generally provide superior prediction accuracy for fly ash-based geopolymer concrete because they capture nonlinear material behaviour more effectively than classical regression models. Likewise, Cihan [45] reported that Random Forest and other ensemble algorithms achieved very high predictive accuracy for geopolymer concrete strength and were less prone to overfitting when properly tuned. From an engineering perspective, the results suggest that curing temperature and activator ratio are the two critical parameters requiring careful optimization in *Costus afer*-modified fly ash geopolymer concrete. While the activator ratio directly influences the availability of soluble silicate species for geopolymer gel formation, curing temperature governs the rate and extent of geopolymerization. The relatively small contribution of NaOH concentration in the RF model indicates that, within the investigated range, further increases in molarity provide limited additional strength benefits and may even be detrimental when the optimum level is exceeded.

Overall, although the MLR model provides a simple mathematical expression that may be useful for preliminary engineering calculations, the Random Forest model offers substantially better predictive capability and a more realistic representation of the complex physicochemical interactions governing the compressive strength of *Costus afer*-modified fly ash geopolymer concrete. Consequently, the RF model is recommended as the preferred predictive tool for mixture optimization and performance forecasting in future studies.

4. Conclusions

This study developed and evaluated two predictive models, namely Multilinear Regression (MLR) and Random Forest (RF), for estimating the compressive strength of *Costus afer*-modified fly ash geopolymer concrete using curing temperature, activator ratio, and sodium hydroxide concentration as the principal process variables. The findings demonstrate that both modelling approaches were capable of predicting compressive strength with reasonable accuracy; however, their predictive capabilities differed considerably owing to the complexity of the relationships among the input variables.

The MLR model provided a satisfactory representation of the experimental data, achieving coefficients of determination (R^2) of 0.8248 and 0.9331 for the training and testing datasets, respectively. The regression analysis identified the activator ratio as the variable exerting the strongest positive linear influence on compressive strength, while curing temperature also contributed positively to strength development. In contrast, increasing the sodium hydroxide concentration beyond the investigated optimum range was associated with a slight reduction in compressive strength. Although the MLR model offers the advantage of a simple mathematical expression that can be readily applied for preliminary engineering estimations, the residual distribution and prediction errors indicate that it is limited in its ability to represent the nonlinear behaviour characteristic of geopolymer concrete.

The Random Forest model exhibited markedly superior predictive performance, attaining R^2 values of 0.9898 and 0.9882 for the training and testing datasets, respectively, together with substantially lower MAE, MSE and MAPE values. The close agreement between measured and predicted compressive strengths, coupled with the narrow distribution of residuals, demonstrates that the RF model generalized well and produced highly reliable predictions without evidence of significant overfitting. Feature importance analysis further revealed that curing temperature was the dominant predictor of compressive strength, followed by the activator ratio, whereas the contribution of sodium hydroxide concentration was comparatively small within the investigated range. These findings highlight the importance of accounting for nonlinear interactions among processing variables during the geopolymerization process.

A comparison of the two modelling approaches clearly established the superiority of the Random Forest algorithm. While the MLR model explained over 93% of the variability in the testing dataset, the RF model explained nearly 99%, with prediction errors reduced substantially across all evaluation metrics. The improved performance of the RF model reflects its ability to capture the complex interactions among curing conditions and activator chemistry that cannot be adequately represented by a conventional linear regression model. Consequently, the Random Forest algorithm provides a more realistic and dependable framework for predicting the compressive strength of *Costus afer*-modified fly ash geopolymer concrete.

Beyond model evaluation, the outcomes of this study have practical significance for the design and optimization of sustainable geopolymer concrete mixtures. The results indicate that careful control of curing temperature and activator ratio is essential for maximizing compressive strength, while unnecessary increases in sodium hydroxide concentration are unlikely to produce proportional improvements in performance. The developed Random Forest model therefore offers a practical decision-support tool for

reducing the number of laboratory trials required during mixture optimization, thereby saving time, materials, and experimental costs.

Future studies should expand the modelling framework by incorporating additional variables such as curing duration, liquid-to-binder ratio, aggregate characteristics, and supplementary precursor materials. Increasing the size and diversity of the experimental dataset and evaluating other advanced machine learning algorithms, including Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), CatBoost, and deep neural networks, would further enhance prediction accuracy and improve the robustness and general applicability of strength prediction models for geopolymer concrete.

Acknowledgement

This research was funded by TETFUND Institution Based Research (IBR) Grant, years 2024/2025 (merged) Batch 20.

References

1. UN Environment, Scrivener, K.L., John, V.M. and Gartner, E.M. (2018). Eco-efficient cements: Potential economically viable solutions for a low-CO₂ cement-based materials industry. *Cement and concrete Research*, 114, pp.2-26.
2. Abiodun, Y.O., Olanrewaju, O.A., Gbenedor, O.P., Ochulor, E.F., Obasa, D.V. and Adeosun, S.O. (2022). Cutting cement industry CO₂ emissions through metakaolin use in construction. *Atmosphere*, 13(9), p.1494.
3. Marandi, N. and Shirzad, S. (2025). Sustainable cement and concrete technologies: A review of materials and processes for carbon reduction. *Innovative Infrastructure Solutions*, 10(9), p.408.
4. Horan, C., Genedy, M., Juenger, M. and van Oort, E. (2022). Fly Ash-based Geopolymers as lower carbon footprint alternatives to Portland cement for well cementing applications. *Energies*, 15(23), p.8819.
5. Castillo, H., Collado, H., Droguett, T., Vesely, M., Garrido, P. and Palma, S. (2022). State of the art of geopolymers: A review. *e-Polymers*, 22(1), pp.108-124.
6. Kumar, S.G.K.M., Kinuthia, J.M., Oti, J. and Adeleke, B.O. (2025). Geopolymer Chemistry and Composition: A Comprehensive Review of Synthesis, Reaction Mechanisms, and Material Properties—Oriented with Sustainable Construction. *Materials*, 18(16), p.3823.
7. Shobeiri, V., Bennett, B., Xie, T. and Visintin, P. (2021). A comprehensive assessment of the global warming potential of geopolymer concrete. *Journal of Cleaner Production*, 297, p.126669.
8. Sharma, B., Rajput, A.S. and Barbhuiya, S. (2026). Life cycle assessment of energy and CO₂ emission in clinker, OPC, and PPC production: case studies and pathways to emission reduction. *Innovative Infrastructure Solutions*, 11(6), p.337.
9. Qaidi, S., Najm, H.M., Abed, S.M., Ahmed, H.U., Al Dughaisi, H., Al Lawati, J., Sabri, M.M., Alkhatib, F. and Milad, A. (2022). Fly ash-based geopolymer composites: A review of the compressive strength and microstructure analysis. *Materials*, 15(20), p.7098.
10. Khan, Q.S., Mumtaz, T., Qazi, A.U. and Pham, T.M. (2024). Influence of alkaline activators on mechanical properties of environmentally friendly geopolymer concrete under different curing regimes. *Environmental Science and Pollution Research*, 31(50), pp.60619-60639.
11. Mohamed, O.A. (2023). Effects of the curing regime, acid exposure, alkaline activator dosage, and precursor content on the strength development of mortar with alkali-activated slag and fly ash binder: a critical review. *Polymers*, 15(5), p.1248.
12. Mohammed, T.O., Aman, U.H., Bin, H.M.Z. and Fanijo, E.O. (2025). Recent Advances in Fly Ash-and Slag-Based Geopolymer Cements. *Sustainability*, 17(24), p.11167.

13. Ahmed, M.M., Sadoon, A., Bassuoni, M.T. and Ghazy, A. (2024). Utilizing agricultural residues from hot and cold climates as sustainable SCMs for low-carbon concrete. *Sustainability*, 16(23), p.10715.
14. Onsongo, S.K., Olukuru, J., Munyao, O.M. and Mwabonje, O. (2025). The role of agricultural ashes (rice husk ash, coffee husk ash, sugarcane bagasse ash, palm oil fuel ash) in cement production for sustainable development in Africa. *Discover Sustainability*, 6(1), p.62.
15. Devadharshini, M. and Vasugi, K. (2025). A Comprehensive Review and Future Directions on Utilization of Agricultural Waste Ash in Geopolymer Concrete. *International Journal of Engineering Trends and Technology*, 73(9), pp.125-140.
16. Harmaji, A., Kirana, M.C. and Jafari, R. (2024). Machine learning to predict workability and compressive strength of low-and high-calcium fly ash-based geopolymers. *Crystals*, 14(10), p.830.
17. Mushthofa, M., Thedy, J., Teguh, M., Purwanto, Pratama, A.G. and Han, A.L. (2026). Artificial Intelligence in Geopolymer Concrete Mix Design: A Comprehensive Review of Techniques and Applications. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, 50(1), pp.49-81.
18. Li, F., Rana, M.S. and Qurashi, M.A. (2025). Advanced machine learning techniques for predicting concrete mechanical properties: A comprehensive review of models and methodologies. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, 8(1), p.110.
19. Al-Shamasneh, A.A.R., Mahmoodzadeh, A., Karim, F.K., Saidani, T., Alghamdi, A., Alnahas, J. and Sulaiman, M. (2025). Application of machine learning techniques to predict the compressive strength of steel fiber reinforced concrete. *Scientific Reports*, 15(1), p.30674.
20. ASTM C618-19 (2022). Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete. ASTM International, West Conshohocken, USA.
21. Palanisamy, E. and Ramasamy, M. (2022). Dependency of sisal and banana fiber on mechanical and durability properties of polypropylene hybrid fiber reinforced concrete. *Journal of Natural Fibers*, 19(8), pp.3147-3157.
22. BS EN 12620:2013. Aggregates for concrete. BSI, Brussels.
23. BS EN 15743:2010+A1:2015. Supersulfated cement. Composition, specifications and conformity criteria. BSI, Brussels.
24. Albidah, A., Alghannam, M., Abbas, H., Almusallam, T. and Al-Salloum, Y. (2021). Characteristics of metakaolin-based geopolymer concrete for different mix design parameters. *Journal of Materials Research and Technology*, 10, pp.84-98.
25. Nath, P., and Sarker, P.K. (2017). Flexural Strength and Elastic Modulus of Ambient-Cured Blended Low-Calcium Fly Ash Geopolymer Concrete, *Construction and Building Materials*, 130, 22–31.
26. Ahmed, H.U., Mohammed, A.A., Rafiq, S., Mohammed, A.S., Mosavi, A., Sor, N.H., and Qaidi, S.M.A. (2021). Compressive Strength of Sustainable Geopolymer Concrete Composites: A State-of-the-Art Review, *Sustainability*, 13, 1-38.
27. BS EN 12390-3:2019. Testing hardened concrete – compressive strength of test specimens. BSI, Brussels.
28. Weisberg, S. (2005). Applied Linear Regression. 3rd Ed. John Wiley & Sons, NewJersey.
29. Pardoe, I. (2020). Multiple Linear Regression. In: Applied Regression Modeling, I. Pardoe (Ed.). <https://doi.org/10.1002/9781119615941.ch3>
30. Matsimbe, J., Dinka, M., Olukanni, D. and Musonda, I. (2024). Fundamental machine learning algorithms and statistical models applied in strength prediction of geopolymers: a systematic review. *Discover Applied Sciences*, 6(10), p.538.

31. Mienye, I.D. and Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, 10, pp. 99129-99149.
32. Halabaku, E. and Bytyçi, E. (2024). Overfitting in Machine Learning: A Comparative Analysis of Decision Trees and Random Forests. *Intelligent Automation & Soft Computing*, 39(6).
33. Hilloulin, B. and Umunnakwe, R. (2024). Machine learning-aided prediction of shrinkage in modern concrete: Focus on mix proportions and SCMs. *Journal of Building Engineering*, 98, p.111410.
34. Zhang, M., Zhang, C., Zhang, J., Wang, L. and Wang, F. (2023). Effect of composition and curing on alkali activated fly ash-slag binders: Machine learning prediction with a random forest-genetic algorithm hybrid model. *Construction and Building Materials*, 366, p.129940.
35. Velusamy, P. and Thangavel, R. (2026). Predictive modeling of age-dependent strength in concrete incorporating GGBS, silica fume, and phosphogypsum using XGBoost, Random Forest, and Weibull analysis. *Mechanics of Time-Dependent Materials*, 30(1), p.2.
36. Azevedo, A.G.S., Strecker, K., Barros, L.A., Tonholo, L.F. and Lombardi, C.T. (2019). Effect of curing temperature, activator solution composition and particle size in Brazilian fly-ash based geopolymer production. *Materials Research*, 22, p.e20180842.
37. Cheng, Y., Cong, P., Zhao, Q., Hao, H., Mei, L., Zhang, A., Han, Z. and Hu, M. (2022). Study on the effectiveness of silica fume-derived activator as a substitute for water glass in fly ash-based geopolymer. *Journal of Building Engineering*, 51, p.104228.
38. Nodehi, M. and Taghvaaee, V.M. (2022). Alkali-activated materials and geopolymer: a review of common precursors and activators addressing circular economy. *Circular Economy and Sustainability*, 2(1), pp.165-196.
39. Guo, L., Liu, J., Zhou, M. and An, S. (2024). Effect of an alkali activators on the compressive strength and reaction mechanism of coal gangue-slag-fly ash geopolymer grouting materials. *Construction and Building Materials*, 426, p.136012.
40. Parhi, S.K. and Patro, S.K. (2023). Prediction of compressive strength of geopolymer concrete using a hybrid ensemble of grey wolf optimized machine learning estimators. *Journal of Building Engineering*, 71, p.106521.
41. Rathnayaka, M., Karunasinghe, D., Gunasekara, C., Wijesundara, K., Lokuge, W. and Law, D.W. (2024). Machine learning approaches to predict compressive strength of fly ash-based geopolymer concrete: A comprehensive review. *Construction and Building Materials*, 419, p.135519.
42. Hu, H., Jiang, M., Tang, M., Liang, H., Cui, H., Liu, C., Ji, C., Wang, Y., Jian, S., Wei, C. and Song, S. (2025). Prediction of compressive strength of fly ash-based geopolymers concrete based on machine learning. *Results in Engineering*, p.106492.
43. Abdellatief, M., Hassan, Y.M., Elnabwy, M.T., Wong, L.S., Chin, R.J. and Mo, K.H. (2024). Investigation of machine learning models in predicting compressive strength for ultra-high-performance geopolymer concrete: A comparative study. *Construction and Building Materials*, 436, p.136884.
44. Bin, F., Hosseini, S., Chen, J., Samui, P., Fattahi, H. and Jahed Armaghani, D. (2024). Proposing optimized random forest models for predicting compressive strength of geopolymer composites. *Infrastructures*, 9(10), p.181.
45. Cihan, M.T. and Cihan, P. (2025). Bayesian-Optimized ensemble models for geopolymer concrete compressive strength prediction with interpretability analysis. *Buildings*, 15(20), p.3667.