Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.083654

ARTICLE

Accurate Compressive Strength Prediction of Fly Ash Geopolymers Using Advanced Ensemble Models and Morris Analysis

Arslan Qayyum Khan[1], **Muhammad Dawood Rasheed**[2], **Muhammad Huzaifa Naveed**[2] and **Amorn Pimanmas**[3,*]

[1]Department of Civil and Environmental Engineering, Florida International University, Miami, FL, USA
[2]Department of Civil Engineering, The University of Lahore, Lahore, Pakistan
[3]Department of Civil Engineering, Kasetsart University, Bangkok, Thailand

*Corresponding Author: Amorn Pimanmas. Email: amorn.pi@ku.th

Received: 08 April 2026; Accepted: 06 May 2026; Published: 27 May 2026

ABSTRACT: The construction industry's substantial carbon footprint, primarily attributed to the production of Ordinary Portland Cement, necessitates a transition toward more sustainable alternatives. Geopolymer concrete (GPC), an innovative binder synthesized from industrial by-products like fly ash (FA), offers a promising low-carbon solution but is hindered by performance variability and a lack of standardized design protocols. This research addresses this critical barrier by developing robust predictive models for the compressive strength of FA-based GPC. Six machine learning algorithms, including Bagging, Categorical Boosting (CatBoost), K-Nearest Neighbors (KNN), LightGBM, Random Forest Regressor (RFR), and eXtreme Gradient Boosting (XGBoost), were developed and evaluated. The results demonstrate that the XGBoost model achieved superior predictive accuracy, with the lowest average errors (Mean squared error of 4.66, Mean absolute error of 1.42) and the highest average coefficient of determination (0.962). To enhance model interpretability, a Morris sensitivity analysis was conducted. The analysis quantitatively identified the coarse aggregate as the most influential parameter governing compressive strength, followed by key chemical precursors such as silicon dioxide (SiO_2) and aluminum oxide (Al_2O_3). These findings not only align with established material science principles but also validate the physical realism of the machine learning model. This study provides a reliable computational framework for predicting the performance of FA-based GPC, facilitating mix design optimization and accelerating the adoption of this sustainable material in modern construction.

KEYWORDS: Geopolymer concrete; compressive strength; ensemble machine learning; XGBoost; Morris analysis

1 Introduction

Concrete is widely used, and it is vital as a primary construction material globally. The production of Ordinary Portland Cement (OPC) is fundamentally unsustainable as it involves energy-intensive calcination of limestone that contributes an estimated 7%–8% of global anthropogenic carbon dioxide (CO_2) emissions [1]. This ecological impact has catalyzed a critical search for viable alternative binders within the construction sector. Geopolymer concrete (GPC) is a prominent, technologically mature alternative engineered to avoid the environmental drawbacks of OPC. Instead of relying on cementitious hydration, GPC is synthesized through the alkaline activation of aluminosilicate-rich precursors—often industrial by-products such as fly ash (FA), ground granulated blast furnace slag (GGBS), or metakaolin. This process initiates a polycondensation reaction, forming a robust, three-dimensional aluminosilicate polymer network that serves as the binding matrix [2]. The primary advantage of this approach is a drastic reduction in

carbon footprint, potentially achieving an 80%–90% decrease compared to conventional concrete systems by eliminating the calcination process and promoting the valorization of industrial waste streams [3].

Beyond its environmental credentials, GPC also exhibits several superior engineering properties. These include high early compressive strength, low creep and shrinkage, and remarkable durability in aggressive chemical environments, particularly against acid and sulfate attacks. Such characteristics render GPC an ideal material for infrastructure applications demanding long-term resilience, including marine structures, wastewater systems, and industrial flooring [4]. The utilization of precursors like FA further aligns GPC with the principles of a circular economy, transforming a potential environmental liability into a high-value construction resource [3].

Despite these advantages, the widespread adoption of GPC is impeded by several critical challenges. The inherent heterogeneity of its aluminosilicate precursors, a lack of standardized mix design protocols, and uncertainty about the predictability of long-term performance collectively hinder its integration into mainstream construction practice. The reliable prediction of mechanical properties, most notably compressive strength, is therefore a fundamental prerequisite for structural design and engineering confidence.

This present study directly addresses this critical barrier. This study employed advanced machine learning (ML) methodologies to develop a high-fidelity predictive model for the compressive strength of FA-based GPC. By clarifying the complex, non-linear relationships between material composition and mechanical performance, this study aims to enhance design accuracy and reduce the risk of GPC application, thereby accelerating its adoption as a sustainable cornerstone of modern construction.

1.1 Limitations of Traditional Prediction Methods

Traditional empirical and regression-based methods for estimating concrete compressive strength are poorly suited to FA-based GPC. These models were developed for OPC systems and rely on assumptions of material uniformity and linear behavior that do not apply to geopolymer chemistry. As a result, they fail to capture the complex, non-linear interactions among key parameters such as the chemical composition of FA, alkaline activator concentration and ratio, and curing regime. Their lack of generalizability is further compounded by the inherent variability of industrial by-product materials; a model calibrated for one FA source often performs poorly for another, necessitating repeated experimental recalibration.

Moreover, traditional prediction workflows depend heavily on extensive laboratory testing, multiple trial mixes, and long curing periods, making them time-consuming, resource-intensive, and restrictive for early-stage design. This rigidity limits innovation and hinders optimization efforts to improve performance and sustainability. These limitations underscore the need for adaptive, data-driven techniques capable of modeling GPC's high-dimensional behavior. ML approaches provide a promising alternative, offering rapid, accurate, and transferable predictions that can support more efficient mix design and broader adoption of geopolymer technologies.

1.2 Research Significance

The central objective of this study is to develop accurate, data-driven machine learning (ML) models for predicting the compressive strength of FA-based GPC. While GPC offers a sustainable alternative to ordinary Portland cement by utilizing industrial by-products and significantly reducing carbon emissions, its complex and nonlinear behavior—governed by interdependent mix design and curing parameters—limits the effectiveness of traditional empirical modeling approaches. Although previous studies have applied ML techniques for GPC strength prediction, many are primarily focused on algorithm benchmarking using relatively homogeneous or single-source datasets, with limited emphasis on model interpretability and

generalizability. In addition, the combined use of predictive modeling and systematic sensitivity analysis to understand governing mechanisms remains underexplored.

To address these gaps, this study develops a unified framework that integrates (i) a heterogeneous dataset compiled from multiple independent studies, capturing a broad range of material compositions and curing conditions; (ii) advanced ensemble learning algorithms for robust and accurate strength prediction; and (iii) Morris global sensitivity analysis to quantify the relative importance and interaction effects of key input variables.

Unlike conventional studies, this work not only evaluates predictive accuracy but also provides physically interpretable insights into the factors controlling compressive strength. Furthermore, the performance of the developed models, particularly XGBoost, is benchmarked against state-of-the-art studies reported in the literature, ensuring a transparent and quantitative comparison of predictive capability.

From a practical perspective, the proposed framework enables rapid and reliable prediction of compressive strength, reducing reliance on extensive laboratory testing and supporting efficient mix design optimization. By bridging data-driven modeling with materials science understanding, this study contributes to the advancement of intelligent and sustainable construction practices.

2 Literature Review

The prediction of compressive strength in FA-based GPC has become a focal area of research, driven by the increasing demand for low-carbon construction materials. When activated by alkaline solutions, FA forms an inorganic polymer binder capable of replacing traditional OPC in concrete. Although environmentally beneficial, the geopolymerization process involves complex chemical and physical interactions that challenge conventional modeling approaches. In response, ML and deep learning (DL) techniques have emerged as promising tools for capturing the nonlinear relationships among multiple influencing variables.

A range of conventional ML models has been explored in previous studies. Techniques such as DT (Decision tree), SVR, and ANNs (Artificial Neural Networks) have demonstrated reasonable predictive performance in modeling the compressive strength of GPC [5]. For instance, a study employed SVM, linear regression, ensemble learning, and Gaussian process regression (GPR) to predict the compressive strength of FA-based geopolymer-modified sustainable concrete [6]. Their findings indicated that the optimized GPR model outperformed others, achieving an R^2 of 0.9590 and an RMSE of 1.7132 MPa, highlighting its superior predictive capability. In another comparative study, SVR outperformed both DT and multilayer perceptron (MLP), achieving an R^2 value above 0.90, particularly when key variables such as FA content, activator concentration, and curing temperature were incorporated into the training data [7]. Another study by [8], used ANNs to model compressive strength across various curing conditions and activator ratios, yielding a MAE of <2 MPa. However, both studies emphasized the limitations of small or homogeneous datasets, which may lead to overfitting and compromised model generalizability [6].

To enhance predictive accuracy and robustness, ensemble learning techniques have gained considerable attention. These methods, which combine predictions from multiple base learners, have shown superior performance over individual models. For instance, a study comparing XGBoost, GBM (Gradient Boosting Machines), and RFR found that XGBoost achieved the highest R^2 (0.95), attributed to its ability to manage multicollinearity and capture nonlinear feature interactions [9]. Another investigation evaluated bagging and boosting methods on a dataset of 500 GPC mix designs and concluded that boosting models such as CatBoost and LightGBM consistently outperformed bagging models like RF, particularly when hyperparameters were optimized through grid search and Bayesian optimization [9,10].

Ensemble learning methods have been widely used, combining multiple learning algorithms to achieve better predictive performance. For instance, a study by [8] applied boosting and AdaBoost algorithms to predict the compressive strength of high-calcium FA-based GPC. The boosting technique achieved an R^2 value of 0.96, outperforming the ANN and AdaBoost models, which had R^2 values of 0.87 and 0.93, respectively.

Another research by [11] utilized ensemble ML algorithms, including RFR and GBR, to forecast the compressive strength of high-performance concrete containing FA. The ensemble models demonstrated superior performance compared to individual learners, with RFR and GBR achieving strong correlations and lower error metrics.

Computationally intensive DL methods have also demonstrated promising results. Deep feedforward neural networks (DFNNs) with multiple hidden layers have been shown to outperform traditional ML models by effectively modeling high-order interactions among input features. One study reported an R^2 value of 0.96 using a DFNN architecture with five hidden layers [12]. Proper data preprocessing, including normalization and handling of missing values, was essential in achieving such results. Furthermore, unconventional architectures such as convolutional neural networks (CNNs) have been tested on structured GPC datasets, yielding competitive performance despite their more common use in image data. However, their high computational cost presents a limitation for deployment in real-time or resource-constrained environments [11,13].

Recent advancements have also explored hybrid models that combine ML or DL approaches with optimization algorithms. These models aim to enhance accuracy by fine-tuning model parameters using metaheuristic search techniques. For example, a hybrid model integrating genetic algorithms with ANN successfully improved prediction accuracy by optimizing network weights and biases [14]. Similarly, the use of the Whale Optimization Algorithm in conjunction with SVR led to improved generalization and reduced prediction error, highlighting the effectiveness of bio-inspired optimization methods [15].

Despite the promising performance of these models, a critical challenge remains their limited interpretability. Black-box behavior of complex ML models poses a barrier to adoption in practical settings where transparency and explainability are essential. To address this, researchers have increasingly employed explainable AI (X-AI) tools such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME). For example, a study applied SHAP analysis to XGBoost model has identified sodium hydroxide molarity, FA content, and curing time as key predictors of compressive strength [8]. These findings were consistent with prior domain knowledge, enhancing the credibility and usability of the models. Additionally, the necessity for X-AI techniques in regulatory contexts has been emphasized, where interpretability is as critical as accuracy [6,16].

In addition, it is vital to understand the influence of individual input variables for practical model deployment. Sensitivity analyses using techniques such as the Morris method and permutation importance have revealed that parameters like curing temperature, activator-to-FA ratio, and water-to-solid ratio significantly impact the strength development in GPC [7,9,17]. Moreover, permutation-based assessments of RF models showed notable performance deterioration when these variables were perturbed, reinforcing their central role in accurate prediction.

However, a barrier in the literature is the scarcity of diverse, high-quality datasets. Many studies are limited to region-specific or small-scale experimental data, which hinders the generalizability of the models. Researches by [8,12] stress the importance of compiling large, standardized datasets that encompass a broader range of materials, mixing proportions, and curing environments. Additionally, Ref. [18] highlights that

standard k-fold cross-validation techniques may be inadequate when applied to datasets lacking variability, as they fail to reflect real-world heterogeneity.

Despite the significant progress in applying ML techniques for predicting the compressive strength of geopolymer concrete, several research gaps remain. First, many existing studies rely on relatively small or single-source datasets, limiting the generalizability of the developed models. Second, although high predictive accuracy has been reported, limited emphasis has been placed on model interpretability and understanding the relative influence of input parameters. Third, consistent benchmarking across multiple algorithms and comparison with state-of-the-art models are often lacking. Finally, the interaction effects among key variables are not systematically quantified in most studies.

This study addresses these gaps through a structured approach. A heterogeneous dataset compiled from multiple independent studies is used to improve model robustness and representativeness. Advanced ensemble learning models are employed and systematically evaluated to ensure reliable prediction performance. In addition, Morris global sensitivity analysis is integrated to quantify both the influence and interaction of input variables, enhancing interpretability. Furthermore, the developed models are benchmarked against existing studies to provide a transparent assessment of their predictive capability. Through this framework, the study advances beyond conventional ML workflows by combining accuracy, interpretability, and practical applicability in a unified manner.

In conclusion, the application of ML and DL to predict the compressive strength of FA-based GPC has demonstrated substantial promise. Models such as SVR, XGBoost, and DFNN have achieved high predictive accuracy when combined with effective data preprocessing and optimization techniques. Nevertheless, limitations related to data availability, interpretability, and generalizability continue to restrict the practical deployment of these models. The broader adoption of X-AI tools and the establishment of standardized, diverse datasets are essential for ensuring the reliability and applicability of ML-based solutions in real-world construction scenarios.

3 ML Models and Evaluation Metrics

3.1 ML Models

In this study, a varied selection of ML algorithms was employed to model the complex, nonlinear relationships between mix design parameters and compressive strength in FA-based GPC, as shown in Fig. 1. The selected models are known for their effectiveness in regression tasks and their ability to handle multidimensional data.

eXtreme Gradient Boosting (XGBoost) is a robust and scalable ensemble method that builds multiple regression trees sequentially, where each new tree corrects the errors of the previous trees. It uses a combination of model regularization (L1 and L2), shrinkage, and column subsampling to enhance model generalization and prevent overfitting [19]. XGBoost is known for its superior performance on structured data tasks and is among the most widely used algorithms. The XGBoost objective function integrates a differentiable loss function, $l(y_i, \overline{y}_i)$, which measures prediction error, with a regularization term, $\Omega(f)$, to control model complexity and prevent overfitting. The regularization component penalizes both the number of leaves (T) and the magnitude of their weights (ω) as shown in Eqs. (1) and (2).

$$\mathcal{L}_k \cong \sum_{i=1}^{n} \left[g_i f_k(x_i) + \frac{1}{2} h_i {f_k}^2(x_i) \right] + \Omega(f_k) \tag{1}$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \tag{2}$$

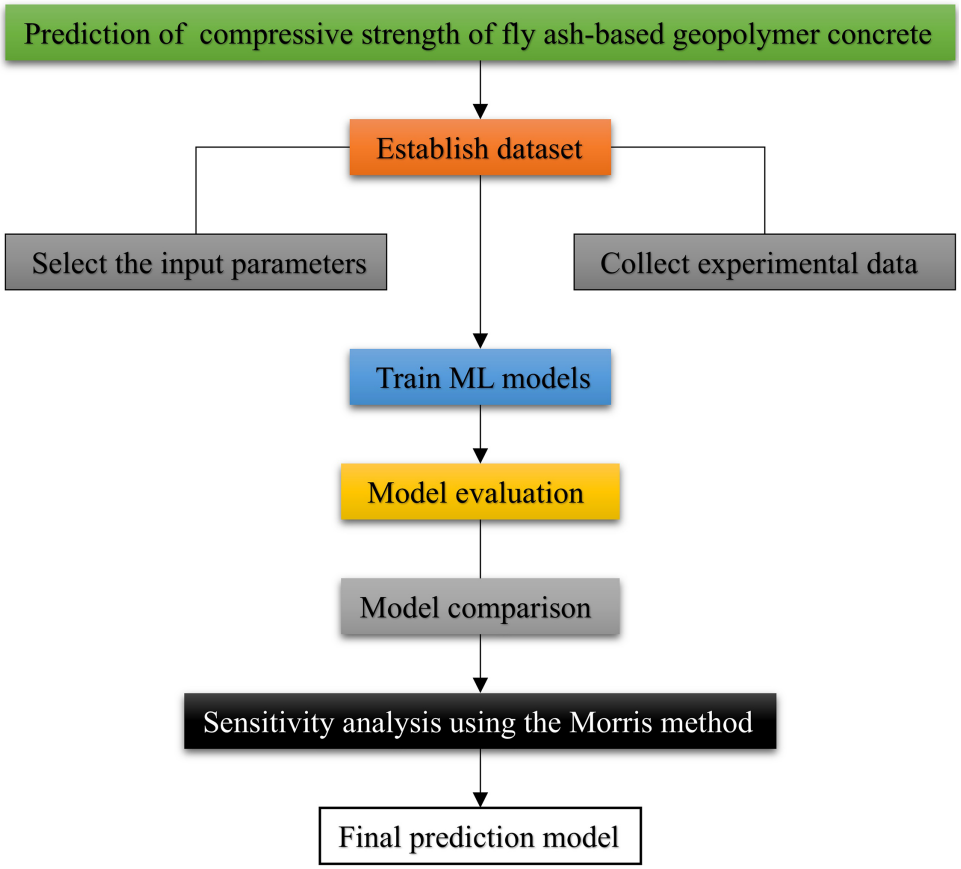

Figure 1: Workflow of data collection, preprocessing, model development, and sensitivity analysis.

To efficiently optimize this objective, XGBoost uses a second-order Taylor expansion, as shown in Eq. (3), reformulating the problem using the first (g_i) and second-order (h_i) gradients of the loss function.

$$\mathcal{L}_k \cong \sum_{j=1}^{T}\left[\left(\sum i \in Ijg_i\right) w_j + \frac{1}{2}\left(\sum i \in I_j h_i + \lambda\right) {w_j}^2\right] + \gamma\left(T\right) \tag{3}$$

This approach enables rapid optimization of the tree structure, enhancing both computational performance and the model's ability to generalize.

LightGBM is a fast, efficient, and high-performance gradient boosting framework developed by Microsoft. It utilizes histogram-based algorithms to bucket continuous feature values, significantly accelerating computation and reducing memory usage. LightGBM supports leaf-wise tree growth, which typically results in deeper trees and improved accuracy [20]. It is particularly well-suited for large datasets and high-dimensional data.

Random Forest Regressor (RFR) is a widely used ensemble learning algorithm that constructs a large number of decision trees during training and outputs the average prediction of all trees. It introduces randomness through both bootstrapping the data and randomly selecting subsets of features for node splitting [21]. This dual-randomization reduces overfitting and improves generalization, making it one of the most robust ML methods.

Categorical Boosting (CatBoost) is a gradient boosting algorithm optimized for datasets with categorical variables, though it also works effectively with continuous features. It uses an ordered boosting process that helps reduce overfitting and handles categorical variables internally without needing one-hot encoding [22,23]. This results in faster training times and improved accuracy with minimal preprocessing.

K-Nearest Neighbour (KNN) is a non-parametric, lazy learning algorithm that makes predictions by averaging the target values of the k closest training examples in the feature space [24,25]. Proximity is usually measured using the Euclidean distance. The method does not assume any underlying data distribution, making it versatile, though it can become computationally expensive and sensitive to the scale and distribution of the input features.

The Bagging Regressor is an ensemble method that builds multiple instances of a base estimator, typically a decision tree, on random subsets of the training data using bootstrapping. The predictions from each model are then averaged to form the final prediction [26,27]. This approach reduces model variance, enhances stability, and improves overall performance, particularly for algorithms prone to high variance.

3.2 Evaluation Metrics

To evaluate the performance of the ML models, three commonly used regression metrics were employed: the coefficient of determination (R^2), mean absolute error (MAE), and mean squared error (MSE), as shown in Eqs. (4)–(6). These metrics provide a comprehensive understanding of how well the models predict compressive strength relative to the experimental results.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i^{'} - y_i\right)^2}{\sum_{i=1}^{n}\left(y_i^{'} - \overline{y}\right)^2} \tag{4}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i^{'} - y_i)^2 \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i^{'} - y_i\right| \tag{6}$$

where n represents the number of observations, y_i denotes the actual value of the target variable, $\hat{y}_i$ is the predicted value, and $\overline{y}$ is the mean of the observed values. Lower values of MSE and MAE indicate better predictive performance, while higher R^2 values signify improved model fit.

By incorporating these performance metrics, the evaluation framework ensures a well-rounded assessment of model accuracy, robustness, and reliability in predicting compressive strength from the given mix design parameters.

4 Methodology

4.1 Data Collection

The predictive accuracy and generalizability of ML models are closely linked to the quality, heterogeneity, and completeness of the dataset employed for training and evaluation. In this study, a comprehensive dataset comprising 145 unique mix designs of FA-based geopolymer concrete (GPC) was systematically compiled from 11 peer-reviewed scientific publications [28–38]. These studies were selected based on the availability of detailed compositional data and on compressive strength outcomes reported under clearly defined experimental conditions. Although the dataset size is moderate, it incorporates data from multiple independent studies, introducing a high level of variability in mix design and curing conditions.

This heterogeneity enhances the robustness of the machine learning models by enabling them to learn from a broader range of practical scenarios. Furthermore, the use of normalization, cross-validation, and independent validation/testing sets helps ensure reliable model performance and reduces the risk of overfitting. It is important to note that the selection of input variables in this study is guided by established physical and chemical mechanisms governing geopolymer concrete behavior. The chosen features represent key factors influencing geopolymerization, including precursor composition (SiO_2 and Al_2O_3), alkaline activation parameters (NaOH and Na_2SiO_3), mix proportions, and curing conditions. The machine learning models are therefore used to capture the complex nonlinear interactions among these mechanism-informed variables rather than relying on purely data-driven feature selection.

By sourcing data from a diverse range of experimental setups, the resulting dataset captures a wide variety of mix proportions, activator formulations, aggregate combinations, and curing protocols. This level of heterogeneity is essential for developing ML models that can capture the highly nonlinear behavior of FA-based GPC and ensure their applicability across a broad range of practical construction scenarios.

Each of the mix design used in this study is defined by 13 input features known to affect the compressive strength concrete significantly. These parameters encompass material composition, chemical reactivity, mix ratios, and thermal curing conditions. The inclusion of these features reflects a holistic understanding of the geopolymerization process, enabling a more nuanced model training. A representative mix design illustrating the influence of FA, aggregates, alkaline activators (NaOH and Na_2SiO_3), and water content is presented in Table 1. The range of FA compositions reported across various studies is summarized in Table 2. These variations are critical, as the SiO_2/Al_2O_3 ratio plays a central role in strength development, with higher ratios favoring the formation of -Si-O-Si- bonds, which contribute to a denser and more stable geopolymer network. This is due to the superior bond strength of -Si-O-Si- linkages compared to -Si-O-Al- or -Al-O-Al- bonds [39]. The inclusion of Na_2SiO_3 alongside NaOH further enhances strength by promoting gel formation and improving matrix density, attributed to its high viscosity and reactivity. Additionally, the activation method used significantly influences the mechanical performance of FA-based geopolymers [40]. Beyond binders, coarse and fine aggregates were also considered as input variables. The type and size range of these aggregates are detailed in Table 3. However, parameters such as fineness modulus and precise grading of fine aggregates were excluded due to challenges in standardizing their quantification. Table 4 presents a comprehensive summary of these variables, including their units, ranges, and statistical metrics.

FA content varies from 254.5 to 600 kg/m^3. Its chemical composition is characterized by silicon dioxide (SiO_2) ranging from 36.2% to 75.66% and aluminum oxide (Al_2O_3) ranging from 9.2% to 31.25%. These chemical oxides directly influence the geopolymeric gel structure and the development of strength.

The coarse aggregate ranged from 554 to 1684 kg/m^3 and fine aggregate from 500–706 kg/m^3, both of which contribute to the concrete's packing density and load-bearing capacity. The activator system includes sodium hydroxide (NaOH) content (11.78–198 kg/m^3), its molarity (8–20 M), and sodium silicate (Na_2SiO_3) content (29.51–342 kg/m^3). These parameters collectively dictate the alkalinity and silica availability required for effective geopolymerization.

Table 1: Sample data set of FA-based geopolymer concrete.

FA	SiO_2	Al_2O_3	Coarse Aggregate	Fine Aggregate	NaOH	NaOH (M)	Na_2SiO_3	Na_2SiO_3 /NaOH	AA/FA	Water	Temperature (°C)	Duration (h)	Compressive Strength
550	38.7	20.8	838	600	95	12	239	2.515789	0.607	188.02	0	0	40.5

Table 2: Fly ash composition.

References		SiO_2	Al_2O_3	CaO	SO_3	Fe_2O_3	MgO	LOI
[28]	Pt. Augusta	49.37	31.25	4.8	0.24	4.47	1.28	0.51
	Collie	53.82	29.95	1.03	0.34	9.24	0.58	0.63
	Tarong	75.66	19	0.3	0.03	1.38	0	1.16
[29]		53.71	27.2	1.9	0.3	11.17	–	0.68
[30]		71.5	9.2	6.72	2.4	2.37	0.6	3.67
[31]		53.71	27.2	1.9	0.3	11.17	–	0.68
[32]		51.7	29.1	8.84	1.5	4.76	–	–
[33]		50.5	26.57	2.13	0.41	13.77	1.54	0.6
[34]		38.7	20.8	26.6	2.1	5.3	1.5	0.1
[35]		49	31	5	–	3	3	–
[36]	Type-I	47.87	28	3.81	0.27	14.09	0.93	0.43
	Type-II	49.37	31.25	4.8	0.24	4.47	1.28	0.51
	Type-III	53.82	29.95	1.03	0.34	9.24	0.58	0.63
[37]		45.23	19.95	15.51	–	13.15	–	–
[38]		59.7	28.36	2.1	0.4	4.57	0.83	1.06

Note: LOI: Loss of ignition.

Table 3: Aggregate characteristics based on particle size.

References	Coarse Aggregate		Fineness Modulus		
	Raw Material	Size Range (mm)	Raw Material	Size Range (mm)	Fineness Modulus
[28]	Crushed basalt	7.0–10.0	River sand	–	3
[29]	Crushed granite	7.0–20.0	Natural river sand	–	2.67
[30]	–	9.5–12.5	River sand	–	2.35
[31]	Crushed granite	7.0–20.0	Natural sand	–	2.64
[33]	Crushed stone	7.0–10.0	River sand	–	–
[34]	Crushed granite	10	River sand	–	2.8
[35]	Crushed bluestone gravel	7	Graded sand	<0.4	–
[36]	Crushed basalt aggregate	7.0–10.0	Uncrushed river sand	–	3
[37]	Limestone	20	River sand	–	2.9
[38]	Crushed granite rock	20	Natural river sand	–	2.64

Table 4: Statistical summary of mix design variables.

Input Parameter	Unit	Mean	Standard Deviation	Categories/Range		
				Minimum	Maximum	Range
Fly ash	kg/m^3	444.7324	92.19544	254.5	600	345.5
Silicon dioxide (SiO_2)	%	51.10414	14.5592	36.2	75.66	39.46
Aluminum oxide (Al_2O_3)	%	18.24241	6.635397	9.2	31.25	22.05

(Continued)

Table 4 (continued)

Input Parameter	Unit	Mean	Standard Deviation	Categories/Range		
				Minimum	Maximum	Range
Coarse aggregate	kg/m^3	1102.821	233.5686	554	1684	1130
Fine aggregate	kg/m^3	593.9938	45.55399	500	706	206
Sodium hydroxide (NaOH)	kg/m^3	75.27293	29.536	11.781	198	186.219
Sodium hydroxide molarity	M	11.98621	2.98372	8	20	12
Sodium silicate (Na_2SiO_3)	kg/m^3	166.0145	67.51812	29.512	342	312.488
Na_2SiO_3/NaOH ratio	–	2.287949	0.789268	1	8.769231	7.769231
AA/FA ratio	–	0.534371	0.137368	0.09	0.924757	0.834757
Water content	kg/m^3	143.831	44.35007	37.445	206.778	169.333
Curing temperature	°C	28.58276	32.20222	0	100	100
Curing duration	hours	13.57241	17.39282	0	72	72
Compressive strength	MPa	28.53048	11.51438	7.6	54.4	46.8

In addition to individual component quantities, the dataset incorporates derived ratios that provide deeper insight into chemical balance. These include the Na_2SiO_3 to NaOH ratio (1.00–8.77) and the alkaline activator to FA ratio ranging from 0.09–0.92. Both ratios are essential for optimizing gel formation, ensuring mix stability, and reducing the risk of defects such as efflorescence. Water content, ranging from 37.45 to 206.78 kg/m^3 facilitating precursor dissolution and influencing workability. Curing temperature (0°C–100°C) and duration (0–72 h) define the thermal history of each mix, significantly impacting the kinetics of the reaction and the development of mechanical strength.

The target output variable is the compressive strength of the GPC mixes, expressed in MPa. This metric is central to structural performance assessment and is widely used in both design validation and quality control. The compressive strength data were collected from the original studies, where they were measured using standardized test procedures, predominantly at 7-day and 28-day intervals.

The compressive strength values span a broad range, reflecting the influence of various mix formulations, FA reactivities, and curing strategies. Notably, around 70 out of the 145 mixes underwent elevated-temperature curing up to 100°C, capturing the significant role of thermal activation in strength development. This diversity further enhances the dataset's representativeness and supports robust model learning across different operational regimes.

It is important to mention that, to ensure data quality and consistency, several filtering and verification steps were applied during dataset compilation. Only studies reporting complete mix design parameters and corresponding compressive strength values under clearly defined testing conditions were included. The majority of compressive strength results were reported at commonly adopted curing ages, primarily 7 and 28 days, which are standard in geopolymer concrete research. This selection helps maintain reasonable comparability across different data sources. Furthermore, all collected data were screened for completeness and physical plausibility. Records with missing parameters were excluded, eliminating the need for data imputation. A statistical inspection of the dataset distributions was performed to identify potential outliers. No extreme values outside realistic material behavior ranges were observed; therefore, no data points were removed to preserve the inherent variability of the dataset. It is important to note that the dataset intentionally incorporates data from multiple independent studies with varying materials, mix proportions, and curing conditions. This heterogeneity is beneficial for machine learning model development, as it enhances robustness and enables the models to generalize across a broader range of practical scenarios rather than being limited to a single experimental setup.

4.2 Frequency Distribution of Input Parameters

A comprehensive overview of the distribution characteristics of the input variables used in modeling the compressive strength of FA-based geopolymer concrete is shown in Fig. 2, revealing meaningful trends in mix design, chemical composition, and curing conditions. The FA content in Fig. 2a spans a wide range between approximately 200 and 700 kg/m^3, with frequencies distributed relatively evenly, indicating that the dataset captures both low- and high-powder mixes. The oxide compositions shown in Fig. 2b,c indicate that SiO_2 and Al_2O_3 contents predominantly fall within the mid-range, reflecting the typical variability of class-F FA while ensuring sufficient representation of both silica-rich and alumina-rich sources. The aggregate distributions in Fig. 2d,e show that coarse and fine aggregates vary widely across their respective ranges, highlighting the need to include mixes with different aggregate proportions and gradations, which is critical for capturing mechanical diversity in geopolymer systems.

The activator-related parameters exhibit distinct patterns: NaOH content in Fig. 2f and its molarity in Fig. 2g show clustered peaks, suggesting that certain molarity levels (e.g., 8–16 M) and dosage ranges are more commonly explored in practice. Na_2SiO_3 content (Fig. 2h) and the Na_2SiO_3/NaOH ratio (Fig. 2i) display broader and more dispersed distributions, indicating significant experimental variation in activator proportions—an important factor given the sensitivity of geopolymerization to silicate availability. Likewise, the AA/FA ratio in Fig. 2j shows a moderately uniform spread, capturing mixes with both low and high activator dosages relative to FA mass. Water content (Fig. 2k) also spans a broad range, reflecting differences in workability requirements and activator dilution practices across studies.

The curing parameters in Fig. 2l,m highlight controlled but diverse thermal regimes: curing temperatures primarily fall within 60°C–120°C, consistent with typical heat-curing protocols needed to accelerate geopolymerization, while curing duration ranges from 1 to 90 days, ensuring representation of both early-age behavior and longer-term strength development. Collectively, Fig. 2 demonstrates that the dataset encompasses broad yet realistic ranges of all influential parameters, ensuring variability sufficient for robust ML modeling while also reflecting practical mix design and curing practices in geopolymer concrete research.

4.3 Data Preprocessing

A data preprocessing pipeline was implemented to ensure the reliability and reproducibility of model outputs. The initial step involved completeness verification. Only records containing all 13 input variables along with the corresponding compressive strength values were retained. This eliminated the need for data imputation, which could introduce unwanted bias or noise into the learning process.

Subsequently, normalization was applied to standardize the input features. Given the varying units and magnitudes of the input variables, normalization is essential to avoid scale dominance, particularly in gradient- and distance-based algorithms. The StandardScaler method from the Scikit-learn library was used to transform each feature to have a zero mean and unit variance. This transformation ensures equitable contribution of each feature during model training and enhances convergence performance.

Following normalization, the dataset was divided into training (70%), testing (15%), and validation (15%) sets. This split was selected based on its demonstrated effectiveness in previous ML studies on concrete material properties [5,41–43]. Furthermore, the 70%-15%-15% division was adopted to balance model training, hyperparameter validation, and final independent testing. Since the dataset contains 145 samples, allocating 70% of the data for training ensured that the models had sufficient observations to learn the nonlinear relationships between mix design variables and compressive strength, while the validation and testing subsets provided independent checks for model tuning and final performance assessment. To reduce possible sampling bias, the dataset was stratified based on the compressive strength distribution before

splitting. The strength values were divided into low-, medium-, and high-strength ranges, and samples from each range were proportionally represented in the training, validation, and testing subsets. This procedure ensured that the testing data were not dominated by a particular strength level and that model performance was evaluated across the full range of compressive strength values.

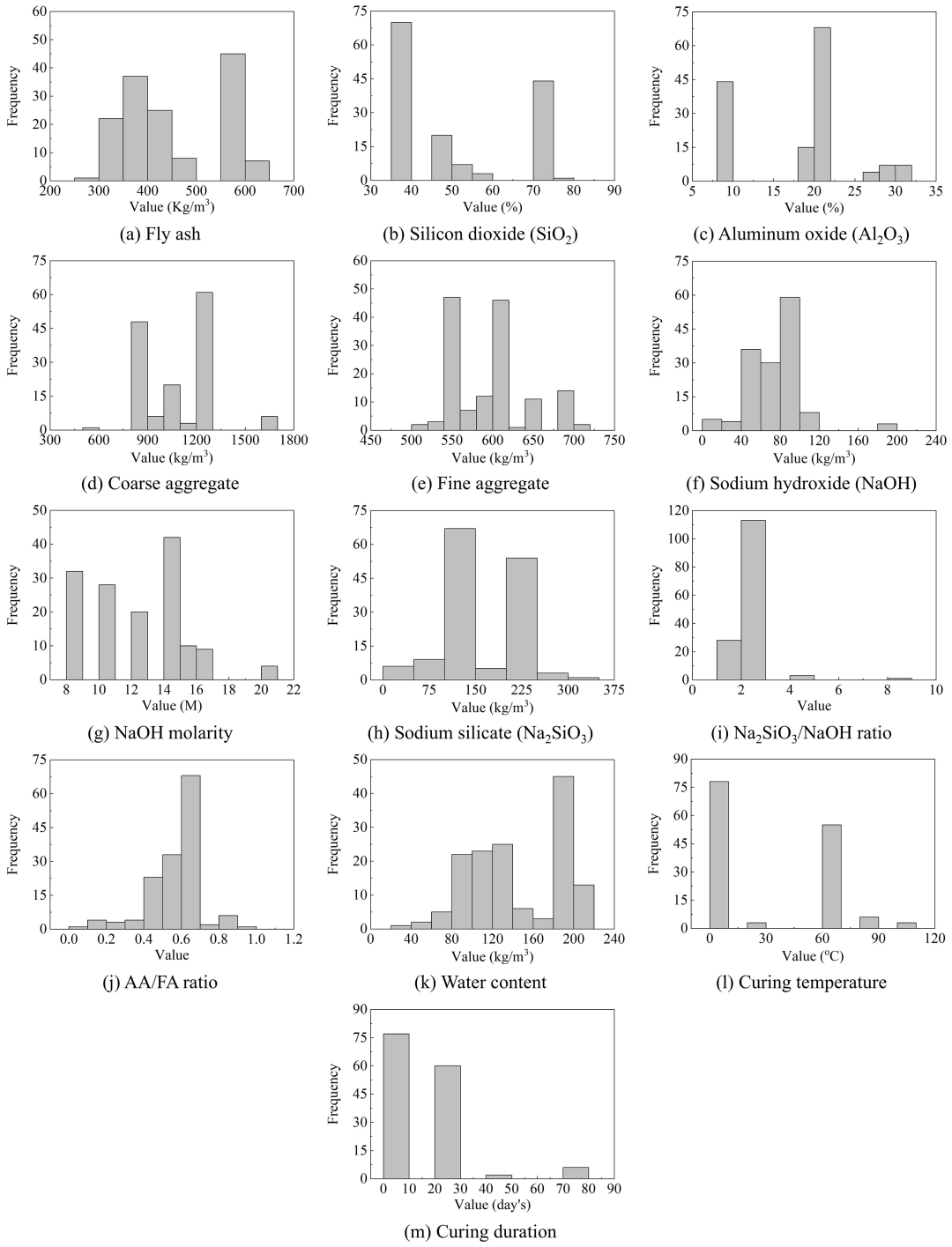

Figure 2: Frequency distribution of the thirteen input parameters used in the GPC strength prediction dataset.

This partitioning strategy is particularly well-suited for datasets of moderate size, such as the 145 entries used in this study. It maximizes the utilization of available data for model development while preserving sufficient observations for unbiased testing. Additionally, this approach aligns with best practices in civil engineering applications of ML, ensuring that the trained models retain generalizability and robustness when deployed in unseen real-world conditions.

4.4 Hyperparameter Tuning

Hyperparameter tuning is a critical process in ML model development that involves selecting the optimal set of hyperparameters to improve model performance. Hyperparameters are configuration settings used to control the learning process, such as the number of trees in a random forest, the learning rate in boosting algorithms, or the number of neighbors in KNN. Unlike model parameters, which are learned during training, hyperparameters must be set before training begins. Effective tuning of these hyperparameters can significantly enhance a model's accuracy, generalizability, and robustness.

In this study, randomized search hyperparameter tuning was employed to optimize the ML models. Randomized Search is an efficient method that samples a fixed number of hyperparameter combinations from a specified distribution or range, rather than exhaustively testing every possible combination, as in grid search. Randomized Search explores the hyperparameter space more broadly but with fewer iterations, making it computationally less expensive and faster, especially for complex models and large search spaces. The search space and final optimized hyperparameter values for all models are summarized in Table 5.

Table 5: Hyperparameter tuning details.

Model	Hyperparameter	Search Range	Optimized Value
XGBoost	n_estimators	50–300	150–250
	max_depth	3–8	4–6
	learning_rate	0.01–0.2	0.05–0.1
	subsample	0.6–1.0	0.8
LightGBM	n_estimators	50–300	200
	max_depth	3–8	5–7
	learning_rate	0.01–0.2	0.05
CatBoost	Iterations	50–300	200
	Depth	4–8	5–6
	learning_rate	0.01–0.2	0.1
RFR	n_estimators	50–300	150
	max_depth	5–15	8–12
KNN	n_neighbors	3–10	4–6
	Weights	Uniform/distance	Distance
Bagging	n_estimators	10–150	80–100
	max_samples	0.6–1.0	0.8

4.5 K-Fold Cross-Validation (CV)

K-fold Cross-Validation (CV) is a widely used technique in ML to evaluate a model's performance and generalizability, where K is the number of folds (e.g., 1–5). It helps ensure that the model's predictive

power is not limited to the specific data it was trained on but can also perform well on unseen data. Among various methods, CV is one of the most popular due to its balance between bias and variance in performance estimation.

In this study, a 5-fold CV was applied during model training and hyperparameter tuning. It offers a robust and efficient framework for model evaluation, particularly when dealing with limited datasets. In a 5-fold CV, the dataset is randomly partitioned into five equal-sized subsets or "folds". The model is trained five times, each time using 4-folds for training and the remaining 1-fold for validation [44]. This process cycles through all five folds, ensuring that each subset is used once for validation. The performance results are then averaged across all five runs to obtain a robust estimate of the model's generalization ability.

5 Results and Discussions

5.1 Comparison of ML Models

Fig. 3 illustrates the comparative performance of the six ML models—XGBoost, LightGBM, RFR, CatBoost, KNN, and Bagging—in predicting the compressive strength of geopolymer concrete, as measured by the R^2 across the training, validation, and testing phases. As shown in the figure, all models demonstrate high predictive capability, with training R^2 values clustering around 0.97–0.98, indicating that each algorithm is capable of learning the underlying relationships within the dataset. The validation and testing results exhibit a slight but expected reduction in R^2, reflecting each model's ability to generalize to unseen data. Among the models, XGBoost consistently maintains the highest R^2 across all three stages, followed closely by LightGBM and the ensemble-based RFR and CatBoost models, which show similar and stable behavior. KNN and Bagging display comparatively lower R^2 values, particularly in the validation and testing phases, suggesting greater sensitivity to noise and reduced robustness. Overall, Fig. 3 demonstrates that gradient-boosting and tree-based ensemble methods provide superior, more stable predictive performance, while distance-based and simple ensemble methods exhibit comparatively weaker generalizability.

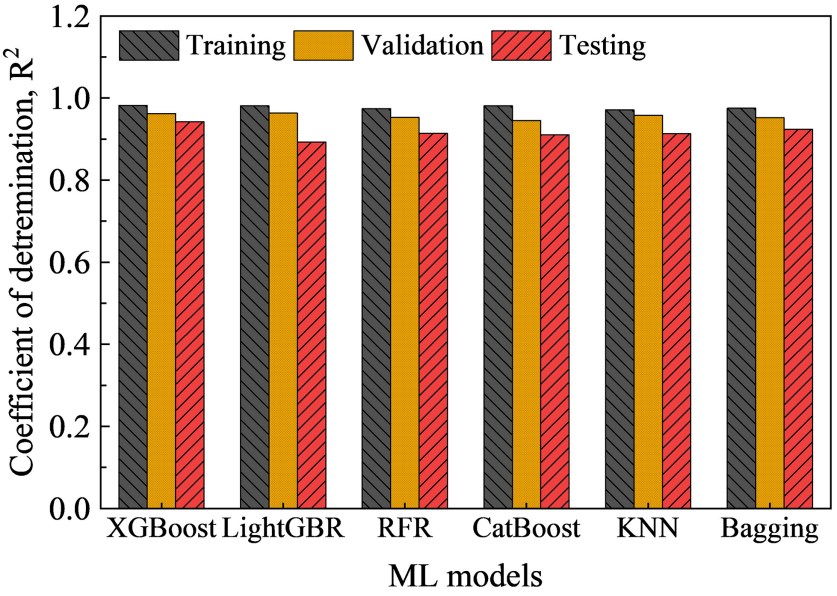

Figure 3: R^2 performance of the six machine learning models across training, validation, and testing sets.

The statistical performance of the evaluated models during the training, validation, and testing phases is presented in Table 6. The XGBoost model emerged as the most robust and accurate predictor among the

evaluated algorithms. It achieved the highest average R^2 value of 0.962 and, most importantly, the highest R^2 score on the test set (0.942). This superior performance on previously unseen data highlights its exceptional generalization capability. The robustness of the XGBoost model is further corroborated by its error metrics: it yielded the lowest average MSE (4.66) and MAE (1.42), as well as the lowest test MSE (6.23) and test MAE (1.89), indicating the highest predictive precision.

Table 6: Statistical performance of ML models and evaluation metrics.

Evaluation Metrics	Sets	Machine Learning Models					
		XGBoost	KNN	CatBoost	Bagging	RFR	Light GBM
R^2	Training	0.982	0.981	0.974	0.981	0.971	0.975
	Validation	0.962	0.963	0.953	0.945	0.958	0.952
	Testing	0.942	0.893	0.914	0.91	0.913	0.924
	Average	0.962	0.946	0.947	0.945	0.947	0.95
MSE	Training	2.39	2.51	3.5	2.53	3.79	3.38
	Validation	5.37	5.29	6.7	7.87	6.02	6.8
	Testing	6.23	11.48	9.24	9.67	9.35	8.21
	Average	4.66	6.43	6.48	6.69	6.38	6.13
MAE	Training	0.81	0.61	1.28	0.87	1.29	1.18
	Validation	1.57	1.52	1.61	1.84	1.47	1.79
	Testing	1.89	2.74	2.3	2.4	2.32	2.26
	Average	1.42	1.62	1.73	1.7	1.69	1.74

The LightGBM also demonstrated strong performance, securing the second-highest average R^2 (0.950) and test R^2 (0.924). This positions LGBM as another highly effective model for this prediction task. The other ensemble methods, including CatBoost, RFR, and Bagging, delivered commendable and closely clustered results, with test R^2 values of 0.914, 0.913, and 0.910, respectively. These models represent viable alternatives, though slightly less performant.

In contrast, the KNN model, despite achieving the highest R^2 score during the validation phase (0.963), exhibited significant performance degradation on the test set, where its R^2 value dropped to 0.893. This discrepancy suggests a degree of overfitting, where the model's instance-based learning did not generalize as effectively to the broader feature space represented by the test data. This is further evidenced by its markedly high test MSE of 11.48.

The performance summarized in Table 6 can also be viewed in the context of previously reported studies. For instance, prior works using XGBoost and ensemble models have reported R^2 values of approximately 0.94–0.95 with RMSE values typically ranging between 3.0 and 4.0 MPa. In comparison, the XGBoost model in this study achieves an R^2 of 0.962 with an RMSE of approximately 2.49 MPa, indicating competitive predictive performance within the range of state-of-the-art models. It is important to note that direct comparisons across studies are subject to differences in dataset composition and experimental conditions. However, the results confirm that the proposed model maintains strong predictive capability even when trained on a heterogeneous dataset compiled from multiple sources.

The superior performance of the gradient boosting models (XGBoost, LightGBM, and CatBoost) can be attributed to their algorithmic architecture. These models employ a sequential ensemble technique where

each subsequent tree is trained to correct the errors of its predecessors. This iterative refinement, combined with built-in regularization mechanisms in models like XGBoost, effectively mitigates overfitting and enhances the capacity to capture the intricate, non-linear dependencies between the mix design parameters and the compressive strength of GPC. In conclusion, while all models demonstrated significant predictive power, the XGBoost algorithm achieved the optimal balance of accuracy, robustness, and generalization, making it the most suitable model for reliable compressive strength prediction in this study.

5.2 Performance of ML Models

The comparison of predicted and actual compressive strength results for the evaluated ML models is comprehensively presented in Fig. 4a–f. These scatter plots, showing predicted values against actual values, provide a robust visualization of the models' predictive accuracy and generalization performance across training, validation, and test datasets. A consistent trend across all models is the proximity of the data points to the diagonal line of equality ($y = x$), indicating a strong correlation between the predicted and actual values.

XGBoost, RFR, LightGBM, CatBoost, and KNN demonstrate robust performance with the majority of their data values clustered exceptionally tightly around the central line and within the ±20 range. This strong alignment suggests minimal prediction error and excellent generalization capabilities. While Bagging exhibited poor performance compared to other models. As depicted in Fig. 4a, the XGBoost data points show the tightest clustering around the line of equality, indicating minimal prediction error and highly accurate predictions. This consistent alignment across subsets underscores XGBoost's robust generalization capabilities, making it the most effective algorithm for modeling compressive strength in this study. However, a closer examination of the LightGBM, RFR, CatBoost, KNN, and Bagging models' performance (Fig. 4b–f) reveals some limitations. Although these models displayed a reasonable fit on the training data, their performance on the test set showed greater error dispersion than that of the ensemble models. Overall, among the evaluated algorithms, the XGBoost model consistently emerged as the most robust and accurate predictor, demonstrating its superior capability for this application.

The efficacy of various ML algorithms in predicting the compressive strength of FA-based GPC is comprehensively presented in Fig. 5a–f. In each figure, the red data points indicate the predicted compressive strength values, while the dark gray data points represent the corresponding actual values. The distribution of prediction errors, calculated as the difference between actual and predicted values, is illustrated quantitatively by the black histograms beneath each plot. A consistent observation across all evaluated models is the remarkable congruence between the predicted and actual values, particularly evident within the training datasets. Ensemble-based algorithms, including XGBoost, LightGBM, RFR, and Bagging, consistently achieve high predictive accuracy, with their predictions closely tracking empirical observations. The error magnitudes across all models consistently remain below ±10 MPa, signifying a robust level of precision in their predictions. This minimal deviation from the actual compressive strength values underscores the excellent overall performance of these models. Among the evaluated algorithms, XGBoost stands out as the most promising, exhibiting superior consistency in its predictions and the tightest clustering of errors around zero, indicating its exceptional ability to model the compressive strength of FA-based GPC accurately.

Fig. 6a–f displays the error distribution for each of the six ML models, allowing for a detailed comparison of their prediction accuracy. In this analysis, error is reported as a percentage rather than an absolute value. This approach was chosen to ensure a fair comparison across all data points, as it prevents the error from being skewed by the wide range of compressive strength values in the dataset.

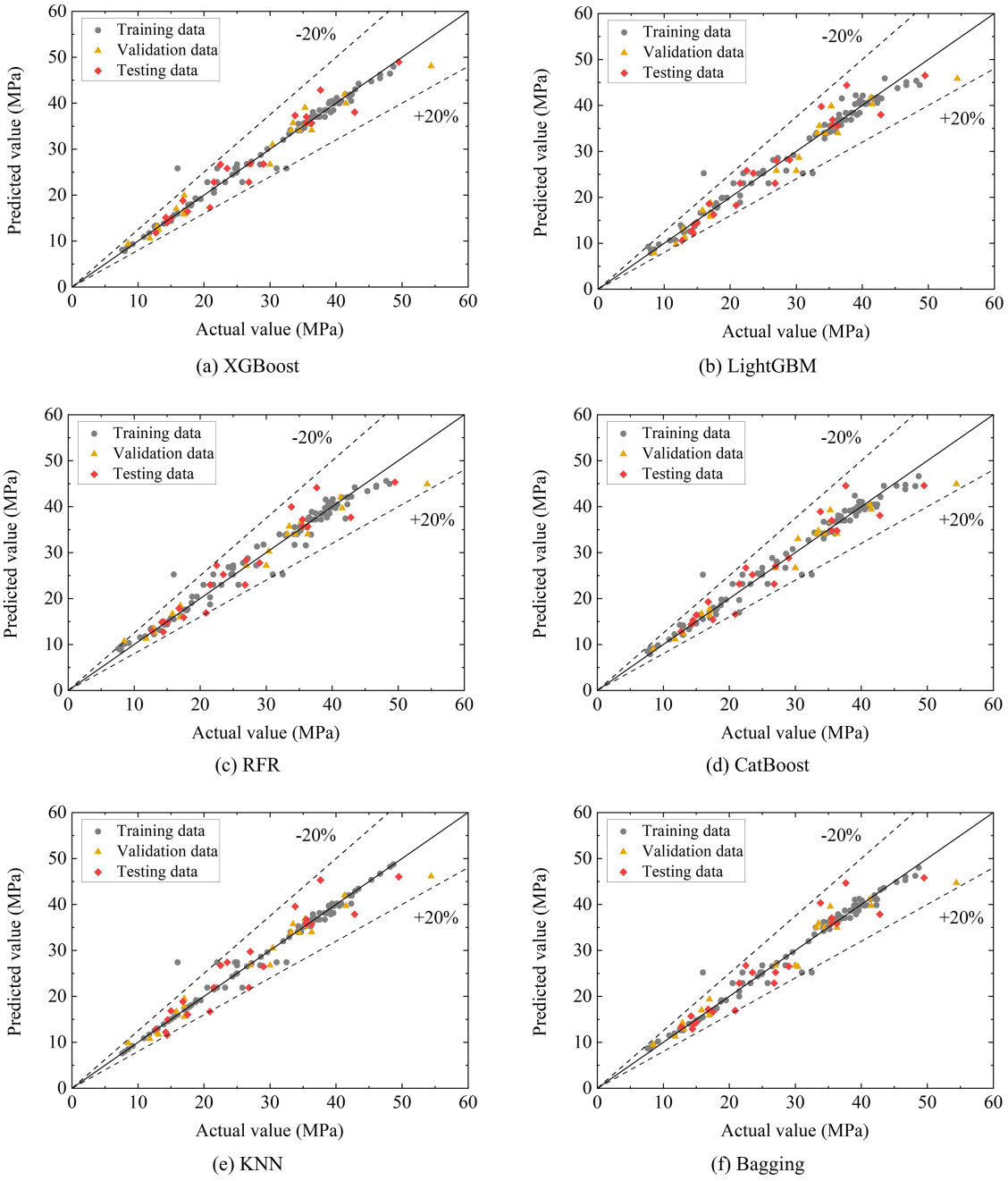

Figure 4: Predicted vs. actual compressive strength values for all six ML models.

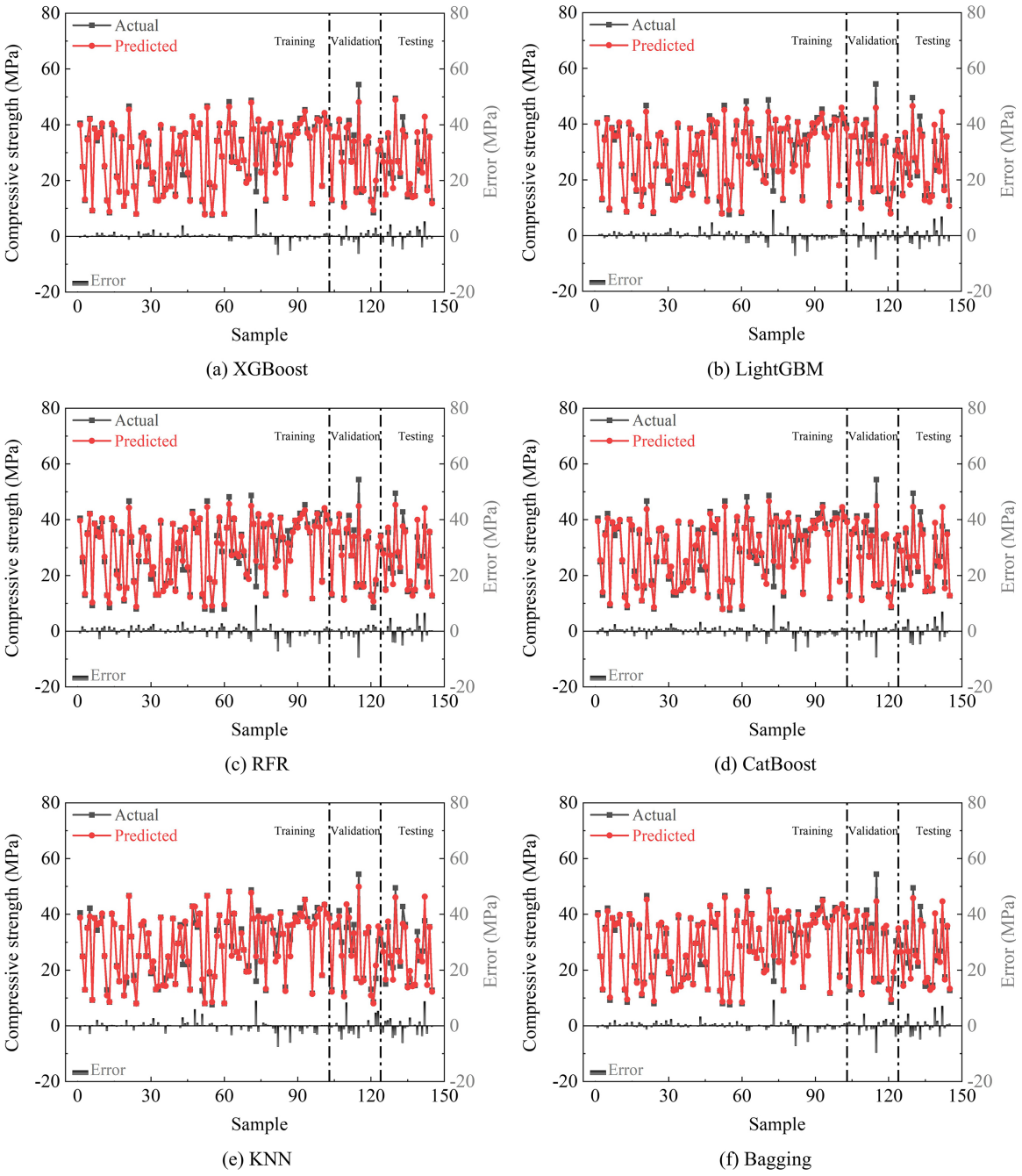

Figure 5: Predicted and actual compressive strength distributions with corresponding error plots for each model.

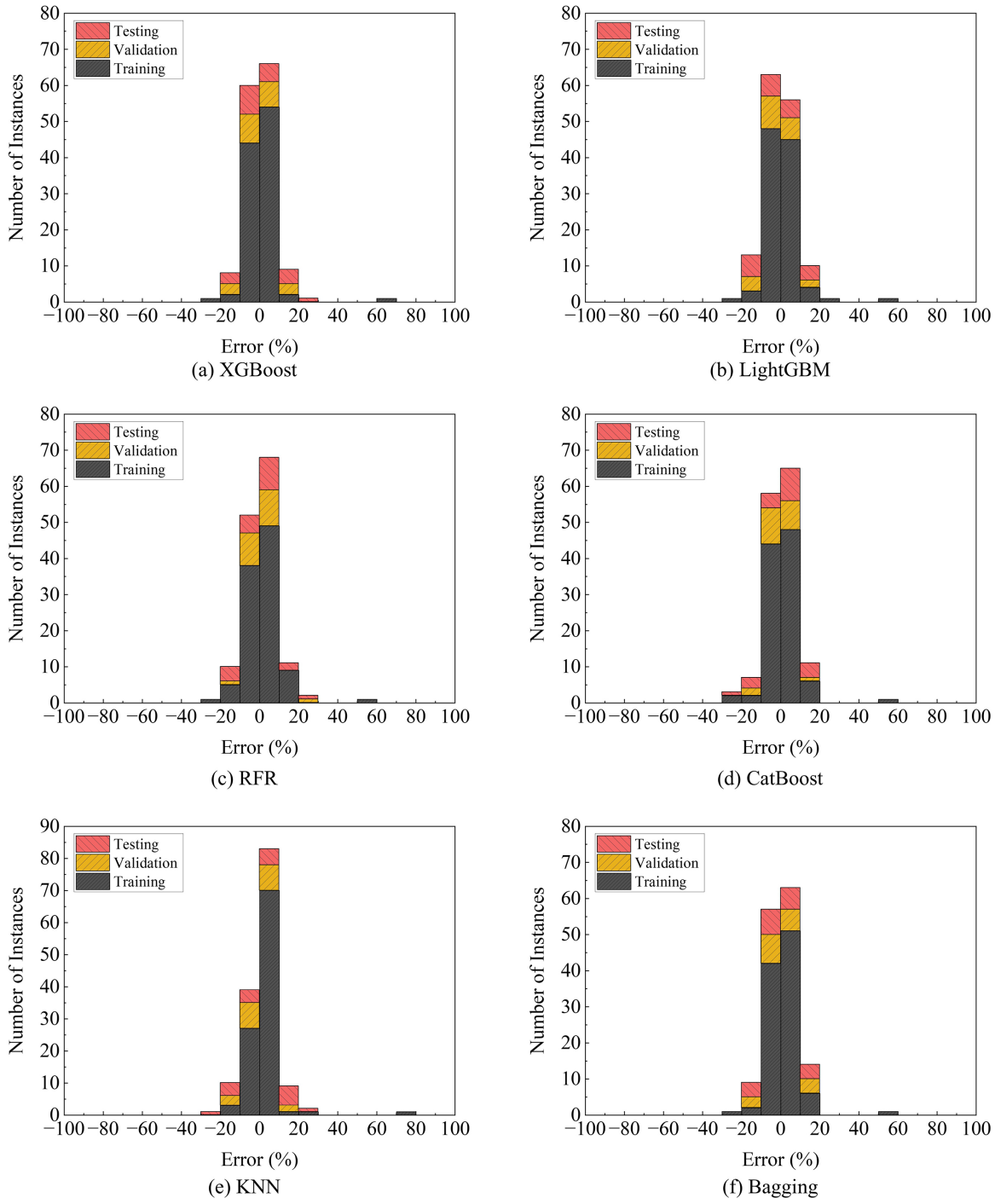

Figure 6: Percentage error distributions for the six ML models.

As shown in the histograms, most models produced errors that are clustered around zero, resembling a normal distribution. This indicates that the models do not systematically over- or under-predict the actual strength values. A distinct concentration of predictions within a narrow error range is particularly evident for XGBoost, LightGBM, RFR, and CatBoost, as shown in Fig. 6a–d. For these models, a large proportion

of predictions fall within ±20% of the actual value, resulting in a tall, narrow peak in their histograms. This shape signifies high precision and reliability.

On the other hand, the error distributions for KNN and Bagging models are more spread out than those for boosting models, indicating greater variance in their prediction accuracy, as highlighted in Fig. 6e,f. The visual evidence from these histograms supports the numerical findings of this study, confirming that gradient boosting models, particularly XGBoost, provide a more robust and accurate framework for predicting the compressive strength of FA-based GPC.

A closer examination of prediction performance indicates that larger errors are generally associated with samples located near the boundaries of the dataset or with extreme combinations of input parameters. These include mixes with very high or low activator content, unusual curing temperatures, or atypical aggregate proportions. Since such conditions are less represented in the dataset, the model has limited exposure to similar patterns during training, leading to higher prediction uncertainty.

Overall, the model demonstrates strong accuracy within the central range of the dataset, while slightly reduced performance is expected for outlier or sparsely represented conditions.

6 Analysis of Input Parameters and Their Effects

The influence of various mix design parameters on compressive strength is visualized in the correlation plots shown in Fig. 7. Overall, the findings reveal a high degree of data dispersion across most parameters, indicating that a single variable does not dominate compressive strength but rather results from complex, multi-variable interactions. This observation is consistent with the challenges of simple pairwise correlation analysis in highly multivariate systems.

Despite the general lack of a strong linear correlation, specific parameters exhibit discernible trends that impact compressive strength. A notable negative correlation is observed with increasing water content; specifically, a rise in water content above approximately 140 kg/m^3 appears to contribute to reduced strength. Conversely, the concentration of alkaline activators significantly impacts mechanical properties. The graphs show that higher compressive strengths are associated with higher molarity of NaOH and higher Na_2SiO_3 concentration, highlighting the critical role of activator composition. Furthermore, the curing conditions show expected correlations: compressive strength generally increases with both curing duration and temperature. The contribution of coarse and fine aggregates highlight their notable influence on strength. In contrast, the contributions of FA and the activator ratios, such as Na_2SiO_3/NaOH and AA/FA, appear highly dispersed, underscoring the need for optimized combinations rather than singular effects.

The influence of SiO_2 and Al_2O_3 on compressive strength can be understood in terms of geopolymerization mechanisms. SiO_2 contributes to the formation of a dense aluminosilicate gel network through dissolution and polycondensation reactions, while Al_2O_3 participates in charge balancing and structural stabilization. However, their effectiveness depends on sufficient alkaline activation and appropriate curing conditions. For example, higher NaOH molarity enhances the dissolution of silica and alumina species, while elevated curing temperature accelerates gel formation. Therefore, the contribution of these oxides to strength is not independent but strongly dependent on their interaction with activator concentration and curing parameters.

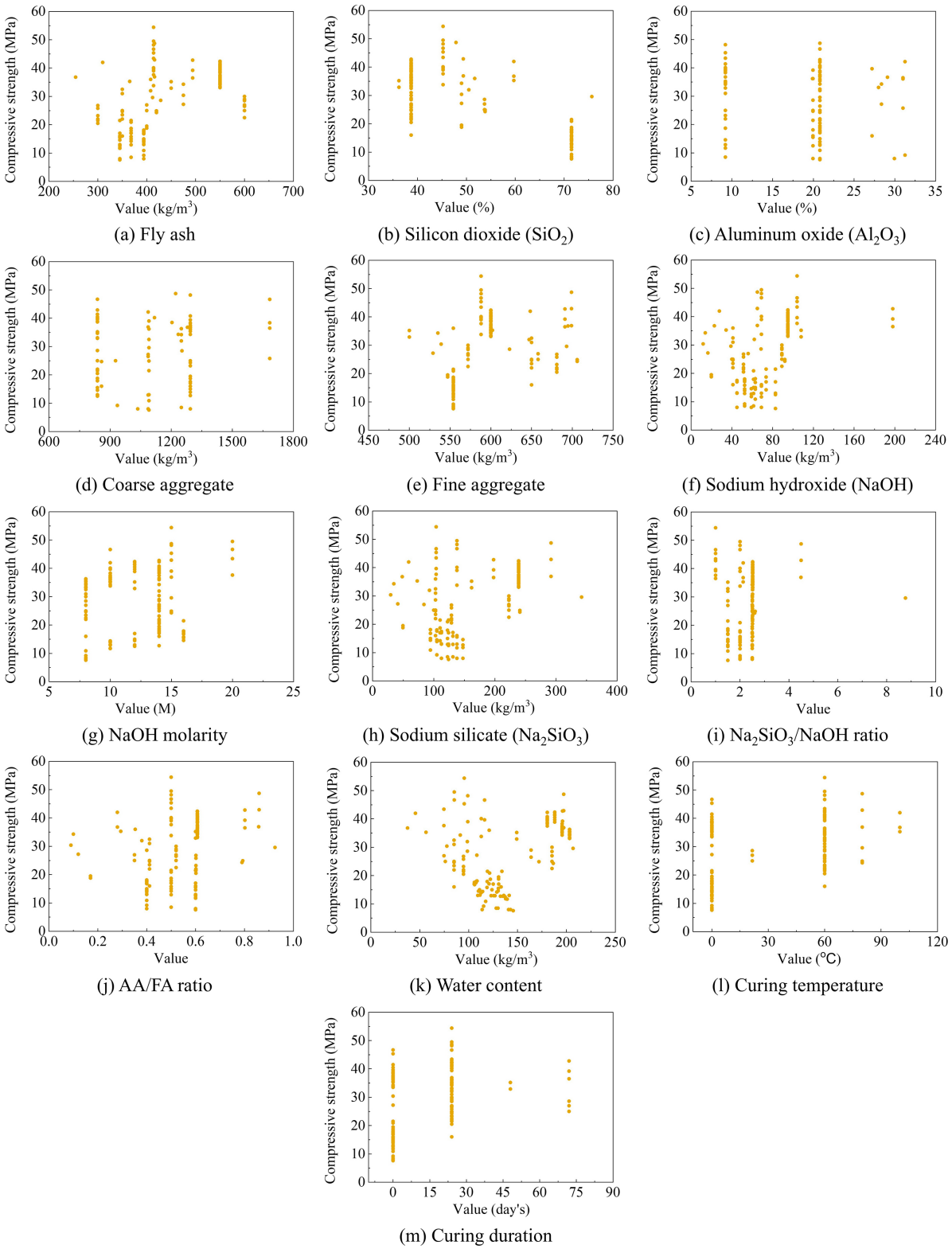

Figure 7: Correlation trends between compressive strength and individual input variables.

7 Sensitivity Analysis

Sensitivity analysis quantifies the influence of each input parameter on the predictive model's output. One method is the Morris Method, which is essential for interpreting complex, non-linear models, such as the one developed in this study, by identifying which input variables have the most significant impact on the predicted compressive strength of FA-based GPC. The Morris method was selected for its computational efficiency and its effectiveness in screening high-dimensional parameter spaces, making it exceptionally well-suited for engineering applications where numerous factors interact.

The Morris method functions by calculating the "elementary effect" (*EE*) for each input variable. An elementary effect measures the change in the model's output resulting from a small, discrete perturbation (Δ) in a single input variable, while all other variables are held constant. For a model function f with k input variables (x_1, x_2, ..., x_k), the elementary effect for the *i*th variable (x_i) is defined in Eq. (7).

$$EE_i = \frac{f(x_1, \ldots, x_i + \Delta, \ldots, x_{k)} - (x_1, \ldots, x_i, \ldots, x_k)}{\Delta} \tag{7}$$

To ensure a robust global analysis, this calculation is repeated multiple times across randomly sampled trajectories throughout the entire input space. The resulting distribution of elementary effects for each variable is then analyzed to derive two key sensitivity indices. One of them is the mean of the absolute elementary effects (μ^*) which quantifies the overall influence of an input variable on the model's output. The other is the standard deviation (σ) of the elementary effects, which quantifies the overall influence of an input variable on the model's output. A higher μ^* value indicates that the variable has a more significant impact on the predicted compressive strength. The σ of the elementary effects captures the degree of nonlinearity or interaction associated with a variable. A high value of σ suggests that the variable's effect is not constant but varies with the values of other input variables. By calculating these two indices for each input parameter, the Morris method provides a precise and quantitative ranking of their relative importance and interaction effects, thereby offering critical insights into the underlying mechanics of the GPC strength prediction model.

The sensitivity results, summarized in Table 7, clearly indicate that coarse aggregate content is the most influential input variable. This observation aligns well with the physical role of coarse aggregates as the primary load-carrying phase in concrete. Their size distribution, volume fraction, and surface characteristics significantly influence the concrete's packing density and interfacial transition zones, thereby impacting strength [8,45]. The large σ for this parameter suggests substantial interaction with other features and excellent aggregate and binder content, indicating a nonlinear influence.

Among the chemical parameters, the SiO_2 and Al_2O_3 contents in FA emerged as key contributors. Their prominent μ^* values confirm their dominant roles in the geopolymerization process, wherein they serve as the precursors for the aluminosilicate gel formation [11,14]. The compressive strength of GPC is highly dependent on the extent and nature of this gel network. The significant σ values for these oxides also indicate that their effectiveness depends on other input parameters, such as alkaline activator concentration and curing conditions. The sodium hydroxide molarity and curing temperature also showed a significant influence. NaOH concentration governs the dissolution of SiO_2 and Al_2O_3 from FA, directly affecting the reaction kinetics and matrix uniformity. Similarly, curing temperature accelerates the polycondensation process, promoting early strength gain and more complete gel development [8,15]. These parameters are therefore crucial to optimizing both early-age and long-term strength characteristics of GPC.

Table 7: Sensitivity analysis of input parameters using Morris method.

Input Parameter	$\mu^\star$	σ
Coarse aggregate	9.0553	4.89
Silicon dioxide (SiO_2)	5.0958	4.8766
Aluminum oxide (Al_2O_3)	4.8691	3.1264
Sodium hydroxide molarity	4.051	3.4889
Curing temperature	3.7489	2.9946
Sodium silicate (Na_2SiO_3)	3.2207	2.6781
Fine aggregate	2.7654	1.8823
Sodium hydroxide (NaOH)	2.2389	1.5802
Na_2SiO_3/NaOH ratio	2.0113	1.6321
AA/FA ratio	1.9641	1.5235
Water content	1.684	1.3749
Fly ash	1.4805	1.1202
Curing duration	1.3098	1.0634

On the other hand, other parameters, such as water content, fine aggregate, AA/FA ratio, and curing duration, showed relatively lower $\mu^\star$ values, indicating a less direct impact on compressive strength within the studied design space. Although these features affect workability, setting time, and microstructural development, their contribution to strength prediction was more subtle and context-dependent. Nonetheless, several of these variables exhibited moderate interaction terms, suggesting that their influence may be conditional on combinations with more dominant features.

The σ values provide insight into nonlinear effects and parameter interactions. High σ values for coarse aggregate and SiO_2 indicate that their influence on compressive strength depends on interactions with other variables rather than acting independently. Coarse aggregate affects strength through packing density, load transfer, and interfacial transition zone behavior, while SiO_2 controls geopolymer gel formation through dissolution and polycondensation. The variability in their effects suggests strong coupling with factors such as activator concentration, water content, and curing conditions. These results confirm that compressive strength is governed by combined physical and chemical interactions rather than single-parameter dominance.

The identification of coarse aggregate as the most influential parameter in this study requires careful interpretation in the context of existing literature, where chemical precursors and alkaline activators are often reported as dominant factors governing geopolymerization. This apparent discrepancy can be attributed to both dataset characteristics and the nature of the sensitivity analysis method. First, the dataset compiled in this study exhibits a relatively wider range of variation in coarse aggregate content compared to certain chemical parameters. Since the Morris method evaluates the influence of input variables based on their variability and impact on the model response, parameters with broader variation can exhibit higher sensitivity indices. From a mechanistic standpoint, coarse aggregates play a crucial role in determining the mechanical behavior of geopolymer concrete. They contribute to load transfer, influence packing density, and affect the quality of the interfacial transition zone (ITZ) between the aggregate and geopolymer matrix. Variations in aggregate content can therefore significantly alter stress distribution and failure behavior, leading to noticeable changes in compressive strength. Although certain findings, such as the importance of coarse aggregate, may align with general expectations, the analysis provides additional insight into their

relative influence within a multi-source dataset and their interaction with other governing parameters. In particular, the combined interpretation of μ^* and σ highlights how both physical factors (aggregate structure) and chemical processes (geopolymerization) interact to control strength development. This enhances the interpretability of the model beyond purely predictive outcomes and provides useful guidance for mix design considerations. It is also important to emphasize that chemical parameters such as SiO_2, Al_2O_3, and NaOH molarity were identified as highly influential in this study, consistent with established geopolymerization mechanisms. These parameters control the dissolution and polycondensation processes that govern binder formation. Overall, the results of the Morris sensitivity analysis confirm that the predictive model aligns with the physical and chemical mechanisms governing GPC behavior. These findings offer actionable insight for mix designers by emphasizing the importance of optimizing aggregate content and oxide composition, as well as controlling the alkali environment and thermal curing conditions. The observed sensitivity trends are also consistent with previous studies on alkali-activated materials [6,8,15], reinforcing the scientific credibility of the model and its outputs.

8 Conclusions

This study aimed to develop robust and interpretable ML models to predict the compressive strength of FA-based GPC. A total of 145 unique mix designs were extracted from eleven peer-reviewed publications, encompassing thirteen input variables, including chemical composition (e.g., SiO_2, Al_2O_3), mix proportions (e.g., FA, coarse and fine aggregates, water), activator characteristics (e.g., sodium hydroxide molarity, Na_2SiO_3 content, AA/FA ratio), and curing conditions (temperature and duration). Six advanced ML algorithms, XGBoost, LightGBM, CatBoost, RFR, KNN, and Bagging, were trained and evaluated using R^2, MAE, and MSE across training, validation, and test datasets.

The following key findings were drawn from this work:

1. All ML models demonstrated high predictive performance, with ensemble models outperforming simpler learners. Among them, XGBoost exhibited the highest accuracy and generalization ability, achieving an average R^2 of 0.962 and the lowest error values across all data splits. The model's stability and predictive precision make it especially suitable for practical implementation in GPC strength prediction tasks.
2. The error distribution and prediction plots confirmed that most of the predicted compressive strength values were closely aligned with the actual values, particularly in the case of XGBoost, LightGBM, and CatBoost. These models consistently delivered predictions within a 20% error margin, with only minor deviations observed during testing.
3. A Morris sensitivity analysis revealed that coarse aggregate content was the most influential parameter affecting compressive strength, due to its critical role in load transfer and matrix densification. This was followed by the SiO_2 and Al_2O_3 content in FA, which are key precursors in the geopolymerization process. Furthermore, the molarity of NaOH and curing temperature also emerged as strong contributors, reflecting their essential roles in activating chemical reactions and enhancing gel formation. In contrast, parameters like water content, AA/FA ratio, and curing duration had relatively lower impacts, although they demonstrated moderate interactions with other features.
4. The integration of ML in material prediction workflows provides a rapid, cost-effective, and sustainable alternative to traditional experimental methods. Instead of conducting exhaustive laboratory trials, practitioners can now predict strength outcomes with high confidence using data-driven approaches. This accelerates material design cycles and supports the broader adoption of GPC as a low-carbon construction material.

5. From a practical standpoint, the developed prediction models—particularly XGBoost—can serve as decision-support tools in both research and industry settings. They can be embedded in digital platforms to assist engineers in selecting optimal mix proportions to achieve desired performance outcomes, thereby reducing overdesign, minimizing waste, and enhancing quality assurance.

In conclusion, this research underscores the potential of ML to revolutionize predictive modeling in sustainable concrete technology. By capturing the nonlinear, multivariable relationships inherent in GPC behavior, ML models provide a scalable solution for reliable strength prediction. The findings support the movement toward innovative, sustainable, and performance-based material engineering, where data-driven insights enable the transition from empirical design to intelligent construction practices. Future research should focus on expanding datasets, incorporating real-time sensor data, and exploring explainable AI methods to further refine model transparency and industrial applicability.

It is important to mention that despite the promising results, this study has several limitations that should be acknowledged. First, the dataset used in this study consists of 145 samples, which, although compiled from multiple sources, remains relatively moderate for machine learning applications. Second, the study focuses exclusively on fly ash-based geopolymer concrete, and therefore the developed models may not be directly applicable to geopolymer systems incorporating other precursors such as slag or metakaolin. Third, compressive strength was considered as the primary performance indicator, while other important properties such as durability, tensile strength, and long-term behavior were not included. Future research should focus on expanding the dataset, incorporating multiple precursor systems, and integrating additional performance indicators to develop more comprehensive and generalizable predictive models.

Acknowledgement: Not applicable.

Funding Statement: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions: Arslan Qayyum Khan: conceptualization; data curation; formal analysis; investigation; methodology; visualization; writing—original draft. Muhammad Dawood Rasheed: data curation; methodology; visualization; writing—original draft. Muhammad Huzaifa Naveed: investigation; project administration; writing—review & editing. Amorn Pimanmas: project administration; resources; supervision; writing—review & editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The dataset supporting the findings of this study, including all input parameters and output variables required to reproduce the reported results, is publicly available in the Zenodo repository at: https://zenodo.org/records/18098943.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Barbhuiya S, Bhusan Das B, Adak D. Roadmap to a net-zero carbon cement sector: strategies, innovations and policy imperatives. J Environ Manage. 2024;359(5):121052. doi:10.1016/j.jenvman.2024.121052.
2. Dong H, Abdul Aziz N, Zulhaidi Mohd Shafri H, Arifin Bin Ahmad K. Computational fluid dynamics study on cemented paste backfill slurry: review. Constr Build Mater. 2023;369:130558. doi:10.1016/j.conbuildmat.2023.130558.
3. Cavalieri M, Ferrara PL, Finocchiaro C, Martorana MF. An economic analysis of the use of local natural waste: volcanic ash of Mt. Etna volcano (Italy) for geopolymer production. Sustainability. 2024;16(2):740. doi:10.3390/su16020740.

4. Wong LS. Durability performance of geopolymer concrete: a review. Polymers. 2022;14(5):868. doi:10.3390/polym14050868.
5. Khan AQ, Naveed MH, Rasheed MD, Miao P. Prediction of compressive strength of fly ash-based geopolymer concrete using supervised machine learning methods. Arab J Sci Eng. 2024;49(4):4889–904. doi:10.1007/s13369-023-08283-w.
6. Kumar A, Arora HC, Kapoor NR, Kumar K. Prognosis of compressive strength of fly-ash-based geopolymer-modified sustainable concrete with ML algorithms. Struct Concr. 2023;24(3):3990–4014. doi:10.1002/suco.202200344.
7. Verma M, Upreti K, Khan MR, Alam MS, Ghosh S, Singh P. Prediction of compressive strength of geopolymer concrete by using random forest algorithm. In: Advanced communication and intelligent systems. Berlin/Heidelberg, Germany: Springer; 2023. p. 170–9. doi:10.1007/978-3-031-25088-0_14.
8. Ahmad A, Ahmad W, Chaiyasarn K, Ostrowski KA, Aslam F, Zajdel P, et al. Prediction of geopolymer concrete compressive strength using novel machine learning algorithms. Polymers. 2021;13(19):3389. doi:10.3390/polym13193389.
9. Ahmad A, Ahmad W, Aslam F, Joyklad P. Compressive strength prediction of fly ash-based geopolymer concrete via advanced machine learning techniques. Case Stud Constr Mater. 2022;16:e00840. doi:10.1016/j.cscm.2021.e00840.
10. Gupta T, Rao MC. Prediction of compressive strength of geopolymer concrete using machine learning techniques. Struct Concr. 2022;23(5):3073–90. doi:10.1002/suco.202100354.
11. Song Y, Zhao J, Ostrowski KA, Javed MF, Ahmad A, Khan MI, et al. Prediction of compressive strength of fly-ash-based concrete using ensemble and non-ensemble supervised machine-learning approaches. Appl Sci. 2022;12(1):361. doi:10.3390/app12010361.
12. Mansouri E, Manfredi M, Hu JW. Environmentally friendly concrete compressive strength prediction using hybrid machine learning. Sustainability. 2022;14(20):12990. doi:10.3390/su142012990.
13. Cao R, Fang Z, Jin M, Shang Y. Application of machine learning approaches to predict the strength property of geopolymer concrete. Materials. 2022;15(7):2400. doi:10.3390/ma15072400.
14. Nguyen KT, Nguyen QD, Le TA, Shin J, Lee K. Analyzing the compressive strength of green fly ash based geopolymer concrete using experiment and machine learning approaches. Constr Build Mater. 2020;247:118581. doi:10.1016/j.conbuildmat.2020.118581.
15. Tanyildizi H. Deep learning–based prediction of compressive strength of eco-friendly geopolymer concrete. Environ Sci Pollut Res. 2024;31(28):41246–66. doi:10.1007/s11356-024-33853-2.
16. Zeng Y, Chen Y, Liu Y, Wu T, Zhao Y, Jin D, et al. Prediction of compressive and flexural strength of coal gangue-based geopolymer using machine learning method. Mater Today Commun. 2025;44:112076. doi:10.1016/j.mtcomm.2025.112076.
17. Jiang P, Zhao D, Jin C, Ye S, Luan C, Tufail RF. Compressive strength prediction and low-carbon optimization of fly ash geopolymer concrete based on big data and ensemble learning. PLoS One. 2024;19(9):e0310422. doi:10.1371/journal.pone.0310422.
18. Qayyum Khan A, Ahmad Awan H, Rasul M, Ahmad Siddiqi Z, Pimanmas A. Optimized artificial neural network model for accurate prediction of compressive strength of normal and high strength concrete. Clean Mater. 2023;10(2):100211. doi:10.1016/j.clema.2023.100211.
19. Khan AQ, Muhammad SG, Raza A, Pimanmas A. Advanced machine learning techniques for predicting mechanical properties of eco-friendly self-compacting concrete. J Road Eng. 2025;5(2):213–29. doi:10.1016/j.jreng.2024.12.002.
20. Microsoft. LightGBM's documentation!—LightGBM 4.6.0.99 documentation. [cited 2025 Jul 15]. Available from: https://lightgbm.readthedocs.io/en/latest/.
21. Khan AQ, Muhammad SG, Raza A, Chaimahawan P, Pimanmas A. Advanced machine learning techniques for predicting compressive strength of ultra-high performance concrete. Front Struct Civ Eng. 2025;19(4):503–23. doi:10.1007/s11709-025-1169-4.

22. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data. 2020;7(1):94. doi:10.1186/s40537-020-00369-8.
23. Khan AQ, Muhammad SG, Raza A, Pimanmas A. Machine learning models for predicting carbonation depth in fly ash concrete: performance and interpretability insights. J Road Eng. 2026;6(1):74–90. doi:10.1016/j.jreng.2025.07.002.
24. Kramer O. K-nearest neighbors. In: Dimensionality reduction with unsupervised nearest neighbors. Berlin/Heidelberg, Germany: Springer; 2013. p. 13–23. doi:10.1007/978-3-642-38652-7_2.
25. Khan AQ, Naveed MH, Rasheed MD, Pimanmas A. Enhancing predictive accuracy in shear strength of RC deep beams: a comprehensive analysis using ensemble machine learning models. Arab J Sci Eng. 2025. doi:10.1007/s13369-025-10763-0.
26. Bühlmann P, Yu B. Analyzing bagging. Ann Statist. 2002;30(4):927–61. doi:10.1214/aos/1031689014.
27. Buja A, Stuetzle W. Observations on bagging. 2002 [cited 2025 May 24]. Available from: http://stat.wharton.upenn.edu/~buja/PAPERS/sinica-bagging-buja-stuetzle.pdf.
28. Gunasekara C, Setunge S, Law DW. Long-term mechanical properties of different fly ash geopolymers. ACI Struct J. 2017;114(3):743–52. doi:10.14359/51689454.
29. Deb PS. Durability of fly ash based geopolymer concrete [master's thesis]. Perth, Australia: Curtin University; 2013.
30. Ghafoor MT, Khan QS, Qazi AU, Sheikh MN, Hadi MNS. Influence of alkaline activators on the mechanical properties of fly ash based geopolymer concrete cured at ambient temperature. Constr Build Mater. 2021;273:121752. doi:10.1016/j.conbuildmat.2020.121752.
31. Nath P, Sarker PK. Use of OPC to improve setting and early strength properties of low calcium fly ash geopolymer concrete cured at room temperature. Cem Concr Compos. 2015;55(12):205–14. doi:10.1016/j.cemconcomp.2014.08.008.
32. Kusbiantoro A, Nuruddin MF, Shafiq N, Qazi SA. The effect of microwave incinerated rice husk ash on the compressive and bond strength of fly ash based geopolymer concrete. Constr Build Mater. 2012;36(2):695–703. doi:10.1016/j.conbuildmat.2012.06.064.
33. Sarker PK, Haque R, Ramgolam KV. Fracture behaviour of heat cured fly ash based geopolymer concrete. Mater Des. 2013;44(6):580–6. doi:10.1016/j.matdes.2012.08.005.
34. Gunasekara C, Atzarakis P, Lokuge W, Law DW, Setunge S. Novel analytical method for mix design and performance prediction of high calcium fly ash geopolymer concrete. Polymers. 2021;13(6):900. doi:10.3390/polym13060900.
35. Aliabdo AA, Abd Elmoaty AEM, Salem HA. Effect of water addition, plasticizer and alkaline solution constitution on fly ash based geopolymer concrete performance. Constr Build Mater. 2016;121(6):694–703. doi:10.1016/j.conbuildmat.2016.06.062.
36. Deb PS, Nath P, Sarker PK. The effects of ground granulated blast-furnace slag blending with fly ash and activator content on the workability and strength properties of geopolymer concrete cured at ambient temperature. Mater Des. 2014;62(9):32–9. doi:10.1016/j.matdes.2014.05.001.
37. Topark-Ngarm P, Chindaprasirt P, Sata V. Setting time, strength, and bond of high-calcium fly ash geopolymer concrete. J Mater Civ Eng. 2015;27(7):04014198. doi:10.1061/(asce)mt.1943-5533.0001157.
38. Joseph B, Mathew G. Influence of aggregate content on the behavior of fly ash based geopolymer concrete. Sci Iran. 2012;19(5):1188–94. doi:10.1016/j.scient.2012.07.006.
39. Zhuang XY, Chen L, Komarneni S, Zhou CH, Tong DS, Yang HM, et al. Fly ash-based geopolymer: clean production, properties and applications. J Clean Prod. 2016;125(3):253–67. doi:10.1016/j.jclepro.2016.03.019.
40. Criado M, Palomo A, Fernández-Jiménez A. Alkali activation of fly ashes. Part 1: effect of curing conditions on the carbonation of the reaction products. Fuel. 2005;84(16):2048–54. doi:10.1016/j.fuel.2005.03.030.
41. Asteris PG, Skentou AD, Bardhan A, Samui P, Lourenço PB. Soft computing techniques for the prediction of concrete compressive strength using non-destructive tests. Constr Build Mater. 2021;303(1):124450. doi:10.1016/j.conbuildmat.2021.124450.
42. Ray S, Haque M, Ahmed T, Mita AF, Saikat MH, Alom MM. Predicting the strength of concrete made with stone dust and nylon fiber using artificial neural network. Heliyon. 2022;8(3):e09129. doi:10.1016/j.heliyon.2022.e09129.

43. Peng Y, Unluer C. Analyzing the mechanical performance of fly ash-based geopolymer concrete with different machine learning techniques. Constr Build Mater. 2022;316:125785. doi:10.1016/j.conbuildmat.2021.125785.
44. Sickit Learn. Cross-validation: evaluating estimator performance—scikit-learn 1.7.0 documentation. [cited 2025 Jul 15]. Available from: https://scikit-learn.org/stable/modules/cross_validation.html.
45. Tian Q, Su Z, Fiorentini N, Zhou J, Luo H, Lu Y, et al. Ensemble learning models to predict the compressive strength of geopolymer concrete: a comparative study for geopolymer composition design. Multiscale Multidiscip Model Exp Des. 2024;7(3):1793–806. doi:10.1007/s41939-023-00303-4.