Research Article

Muhammad Nasir Amin*, Asad Naeem*, Muhammad Iftikhar Faraz and Muhammad Tahir Qadir*

Equation-driven strength prediction of GGBS concrete: a symbolic machine learning approach for sustainable development

Abstract: The growing demand for low-carbon construction materials highlights the need to predict and optimise the performance of Ground Granulated Blast-Furnace Slag (GGBS) concrete with high accuracy. This study aims to develop reliable and interpretable predictive models for estimating the compressive strength (CS) of GGBS-based concrete to support sustainable mix design. Three machine learning approaches, Multilayer Perceptron (MLP), Adaptive Boosting (AdaBoost), and Gene Expression Programming (GEP) were trained on a diverse dataset collected from published experimental studies. Each model was evaluated using five-fold cross-validation and multiple statistical indicators (R^2, RMSE, and MAE), and interpretability was examined through LIME and permutation-based sensitivity analysis. The MLP model achieved the highest predictive accuracy ($R^2 = 0.89$), followed by AdaBoost ($R^2 = 0.88$) and GEP ($R^2 = 0.86$). The GEP model provided a transparent mathematical equation suitable for practical strength estimation. These findings reduce the need for repetitive laboratory testing and enable data-driven optimisation of GGBS concrete design. The study uniquely integrates explainable artificial intelligence with symbolic regression to combine predictive accuracy and interpretability, offering a reproducible framework for sustainable concrete design and analysis.

*Corresponding authors: Muhammad Nasir Amin, NUST Institute of Civil Engineering (NICE), School of Civil and Environmental Engineering (SCEE), National University of Sciences and Technology (NUST), Sector H-12, Islamabad, Pakistan, E-mail: mamin@nice.nust.edu.pk; **Asad Naeem**, Department of Structural Engineering, Military College of Engineering, National University of Sciences and Technology (NUST), Risalpur Campus, Risalpur, 24080, Pakistan, E-mail: asnaeem@mce.nust.edu.pk; and **Muhammad Tahir Qadir**, Department of Civil and Environmental Engineering, College of Engineering, King Faisal University, Al-Ahsa, 31982, Saudi Arabia, E-mail: mqdir@kfu.edu.sa
Muhammad Iftikhar Faraz, Department of Mechanical Engineering, College of Engineering, King Faisal University, Al-Ahsa, 31982, Saudi Arabia

1 Introduction

The construction industry is one of the most significant contributors to global carbon emissions, primarily due to the extensive use of ordinary Portland cement (OPC) in concrete production [1–3]. Cement manufacturing is an energy-intensive process that emits significant amounts of carbon dioxide (CO_2), accounting for nearly 7–8% of global CO_2 emissions [4]. Considering the ongoing climate crisis and the increasing pressure to reduce the environmental footprint of built infrastructure, there is a growing demand for sustainable construction materials that can deliver comparable or enhanced performance with reduced ecological impact [5–8].

One of the most promising strategies to address these environmental challenges is partially replacing OPC with supplementary cementitious materials (SCM's) [9–11]. SCM's are industrial byproducts or naturally occurring materials that exhibit cementitious or pozzolanic properties, enabling them to enhance concrete's long-term strength and durability [12–14]. More importantly, their use significantly reduces carbon emissions, energy consumption, and natural resource depletion associated with cement production [15, 16]. The use of SCMs in concrete aligns with international sustainability agendas, including the SDGs and the European Green Deal, which emphasise carbon reduction and circular construction [17–19]. Consequently, the shift toward SCM-based concrete systems is not only a technical innovation but a vital environmental imperative for a more sustainable and resilient built environment [20, 21].

Among various SCM's, ground granulated blast-furnace slag (GGBS) has emerged as one of the most effective and widely adopted alternatives to OPC [22, 23]. GGBS is a byproduct of the iron and steel manufacturing industry, obtained by rapidly cooling molten slag to form a fine, glassy powder with latent hydraulic properties [24, 25]. Its

incorporation in concrete significantly enhances performance characteristics such as long-term compressive strength, durability, and resistance to aggressive environmental conditions, making it particularly suitable for infrastructure exposed to marine or sulfate-rich environments [26–28]. From an environmental standpoint, the utilisation of GGBS contributes meaningfully to waste valorisation and industrial symbiosis, wherein waste from one industry becomes a valuable resource for another [29, 30]. This not only diverts substantial volumes of slag from landfills but also reduces the carbon footprint of concrete by decreasing the demand for energy-intensive clinker [31, 32]. Several life cycle assessment (LCA) studies have consistently shown that concrete containing high volumes of GGBS exhibits lower embodied carbon and environmental impact compared to traditional OPC concrete [33, 34].

However, despite its advantages, the use of GGBS in concrete is not without challenges [35]. Variability in raw material properties, differences in curing conditions, and slower early-age strength development are among the practical issues that complicate its broader adoption [36, 37]. These factors can introduce uncertainty into the structural performance and design process, particularly when conventional prediction models fall short in learning the complex, nonlinear connections between mix components and curing regimes [38, 39]. Therefore, understanding and accurately predicting the strength behaviour of GGBS-based concrete becomes critical for optimising its use in both sustainable construction and structural applications [40, 41].

The CS of concrete remains one of the most critical parameters in structural design, quality control, and long-term performance assessment [42, 43]. In GGBS-based concrete systems, however, predicting strength development is particularly challenging due to the complex and nonlinear interactions among various influencing factors. These include the proportions of cement, GGBS, and water, as well as curing time, water-to-binder ratio, and the specific chemical and physical characteristics of the slag [44]. Unlike conventional OPC concrete, where strength development follows more predictable patterns, GGBS-modified concrete can exhibit delayed hydration and strength gain, especially at early ages, complicating design timelines and structural evaluations. Furthermore, the effect of GGBS concrete is highly sensitive to environmental curing conditions and mix design variations, both of which influence the kinetics of slag activation and pozzolanic reactions [45]. These sensitivities introduce variability and uncertainty that make it difficult to establish universally reliable empirical or semi-empirical strength prediction models [46]. Traditional regression-based approaches often assume linear relationships and may not adequately capture the multivariate, nonlinear

nature of the strength development process in GGBS-containing concrete [47].

Given the increased interest in performance-based design approaches and the shift toward more sustainable concrete formulations, there is a pressing need for more accurate, flexible, and data-driven predictive models [48]. Such models should be capable of accommodating a high level of mix compositions and curing conditions while reliably estimating the CS at various ages [49]. This need has driven researchers toward exploring ML techniques, which have shown promising capabilities in modelling complex material behaviours where traditional methods have proven insufficient [50–52]. In recent years, ML has emerged as an impressive tool for addressing the complexity and uncertainty associated with predicting the properties of advanced construction materials. Unlike traditional regression models, ML algorithms do not rely on predefined functional relationships; instead, they learn directly from data to uncover hidden patterns and nonlinear interactions among variables [53–58]. This makes them particularly well-suited for modelling the CS of GGBS-based concrete, where numerous interdependent factors contribute to strength development in a nonlinear and often dataset-specific manner [59]. A wide range of ML techniques have been successfully applied in the domain of concrete technology, including artificial neural networks (ANNs), support vector machines (SVMs), random forests (RF), and gradient boosting machines [60–62]. These models have demonstrated superior predictive accuracy and generalizability compared to classical statistical methods, particularly when dealing with large, complex, and noisy datasets [63]. Among these, MLP networks, a type of deep neural network, have been widely adopted due to their strong capability in approximating nonlinear functions and learning intricate data structures [64].

Considering the outlined challenges, this study aims to develop and evaluate three distinct ML models, MLP, AdaBoost, and GEP, for predicting the CS of GGBS-based concrete. The models are applied to an experimental dataset encompassing varied mix designs and curing conditions to identify the most robust and accurate prediction framework. A key novelty lies in the comparative use of conventional (MLP), ensemble-based (AdaBoost), and symbolic evolutionary (GEP) approaches, offering both performance benchmarking and interpretability. In particular, GEP provides transparent, equation-based outputs that contrast with the black-box nature of MLP and AdaBoost. To enhance model trustworthiness and engineering relevance, the study further integrates interpretability techniques, LIME and permutation-based sensitivity analysis, to understand feature influence at both local and global levels. The findings
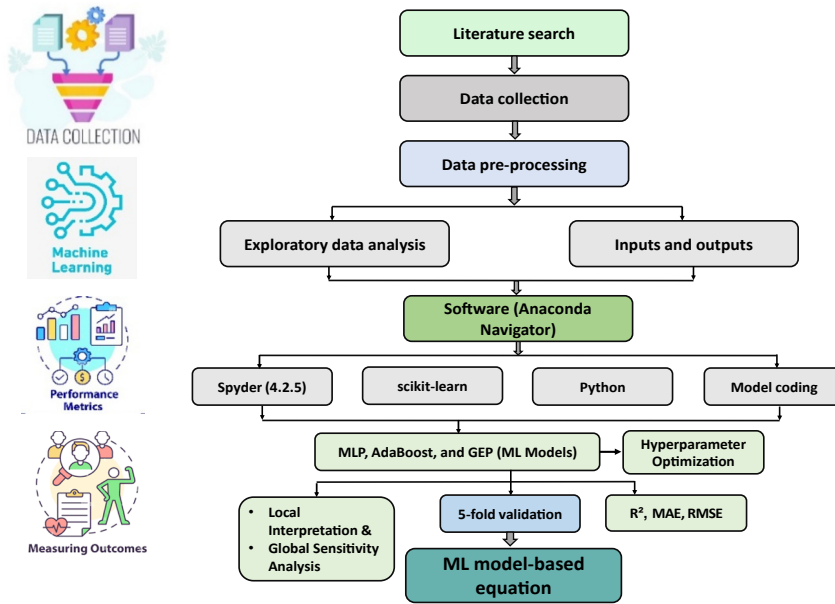
Figure 1: Research strategy and steps involved in the study. "Model coding, Python (Spyder), and scikit-learn" denote the implementation environment for model development rather than a separate step.

contribute to sustainable concrete design by reducing reliance on empirical testing and improving predictive insight.

It is noted that the "optimization" referenced in this study does not involve a separate numerical or heuristic optimization algorithm. Instead, it refers to a data-driven approach in which the trained machine learning models and their interpretability analyses (LIME and permutation sensitivity) are employed to identify the most influential mix parameters, such as water-to-binder ratio, age, and GGBS content, that maximise compressive strength. This predictive-interpretive framework enables rational adjustment of mix proportions in practice, thereby serving as an indirect but effective optimization tool for sustainable GGBS concrete design.

framework, (ii) contextualisation of the dataset and its relevance to sustainable construction, (iii) preprocessing and engineering of input features to enhance model learning, (iv) implementation of machine learning pipelines for model training and validation, (v) evaluation using multiple statistical performance metrics, and (vi) generation of an interpretable GEP-based mathematical equation that enables practical deployment for rapid strength estimation and mix adjustment. This methodology is designed not only to generate high-performance predictive models but also to support their real-world application in the optimisation of GGBS concrete mix designs, with the broader aim of contributing to sustainable construction practices, as shown in Figure 1.

2 Methodology

This study employed a structured and data-driven ML approach to address the complexities associated with predicting the CS of GGBS-based concrete. The methodology integrates experimental data, systematic preprocessing, advanced ML model development, and performance evaluation to create robust predictive tools for sustainable concrete design. By leveraging diverse algorithmic paradigms, including deep learning (MLP), ensemble learning (AdaBoost), and symbolic evolutionary modelling (GEP), the proposed framework ensures both high accuracy and interpretability in strength prediction. The workflow adopted in this research follows a six-stage process: (i) formulation of research objectives and conceptual

2.1 Research framework and objectives

This study employs a structured machine learning framework to address the challenge of accurately predicting the compressive strength (CS) of GGBS-based concrete. Given the nonlinear, multivariable nature of strength development in such sustainable mix designs, conventional empirical models often fail to deliver reliable results across different mix conditions. Therefore, the proposed framework integrates diverse algorithmic strategies to enable data-driven prediction with improved accuracy, generalizability, and practical utility.

The framework is centred around three core machine learning paradigms, such as MLP, a type of deep feedforward neural network capable of modelling highly nonlinear

relationships. AdaBoost is an ensemble learning method that enhances the performance of weak learners through sequential reweighting. GEP is an evolutionary algorithm that produces symbolic models offering mathematical interpretability. The objective is to not only compare the predictive capabilities of these models using a real-world dataset of GGBS-based concrete but also to explore their strengths in terms of accuracy, consistency, and model transparency. The research workflow includes data acquisition, preprocessing, model development and validation, performance evaluation using multiple statistical metrics, and practical deployment of the most effective model. By implementing this framework, the study aims to contribute a replicable, scalable methodology for ML-based prediction of concrete performance and to support the design of more sustainable and optimised cementitious composites.

2.2 Dataset and domain context

The final dataset comprises 787 experimental records collected from published studies [65–67]. Each record represents a unique mix design, incorporating Ground Granulated Blast Furnace Slag as a partial cement replacement, with the CS (in MPa) measured at different curing ages. The dataset includes a wide range of mix compositions and curing durations, offering a robust basis for machine learning model development and generalisation. Key input features extracted across the sources include cement content, GGBS content, water content, fine and coarse aggregates, water-to-binder ratio (W/B), and curing age. These variables are widely recognised as critical determinants of concrete strength, particularly when incorporating SCMs such as GGBS. The dataset reflects real-world practices in sustainable concrete mix design, particularly in contexts where reduced carbon emissions and improved durability are desired. All values were cross-checked for consistency and formatted for use in Python-based modelling workflows. Full references to the data sources are provided in the appropriate section to ensure transparency and reproducibility.

To explore the distributional characteristics of the input features, box plots were generated for all variables used in predicting CS (Figure 2). Most parameters, including cement, GGBS, and aggregate content, exhibit wide ranges with several outliers, particularly in superplasticizer and W/C ratios. These variations reflect the heterogeneity of mix designs compiled from literature and emphasise the need for robust model training to accommodate such diversity. Moreover, the pairplot of the input parameters used in the study is shown in Figure 3, which shows the distribution and pairwise relationships among the variables and CS. However, the descriptive statistics for the same parameters are depicted in Table 1.

The compiled records encompass a broad range of mix compositions and curing conditions reported across the selected studies, including GGBS replacement levels between 0 % and 80 %, curing ages of 3–365 days, and water-to-binder
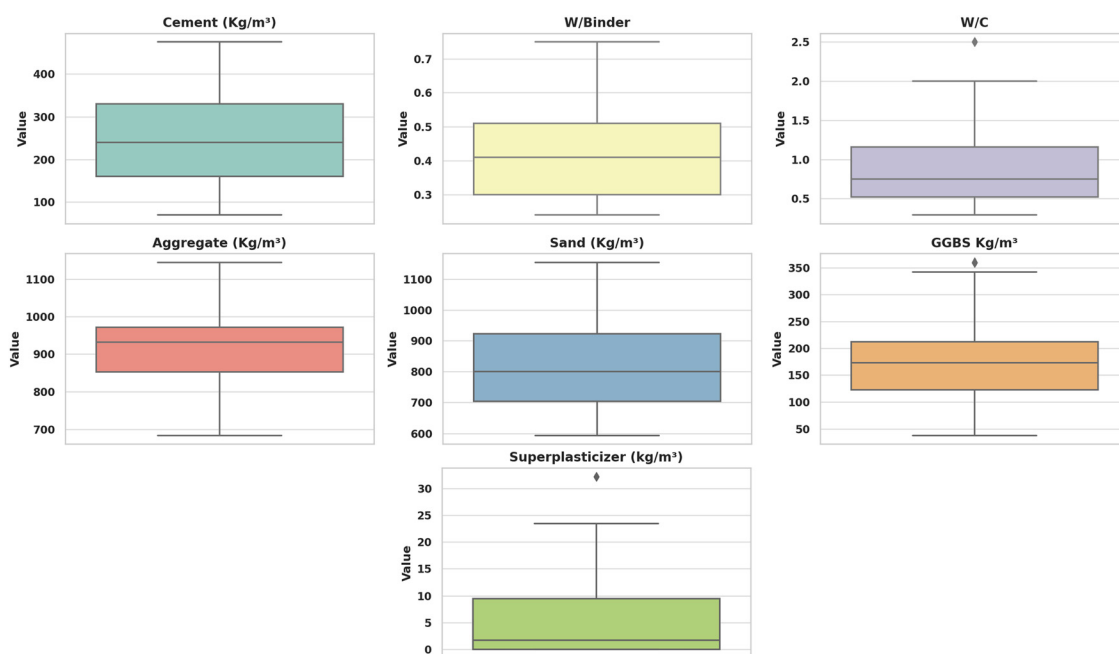
Figure 2: Box plots showing the distribution and variability of input parameters used for predicting CS of GGBS-based concrete.
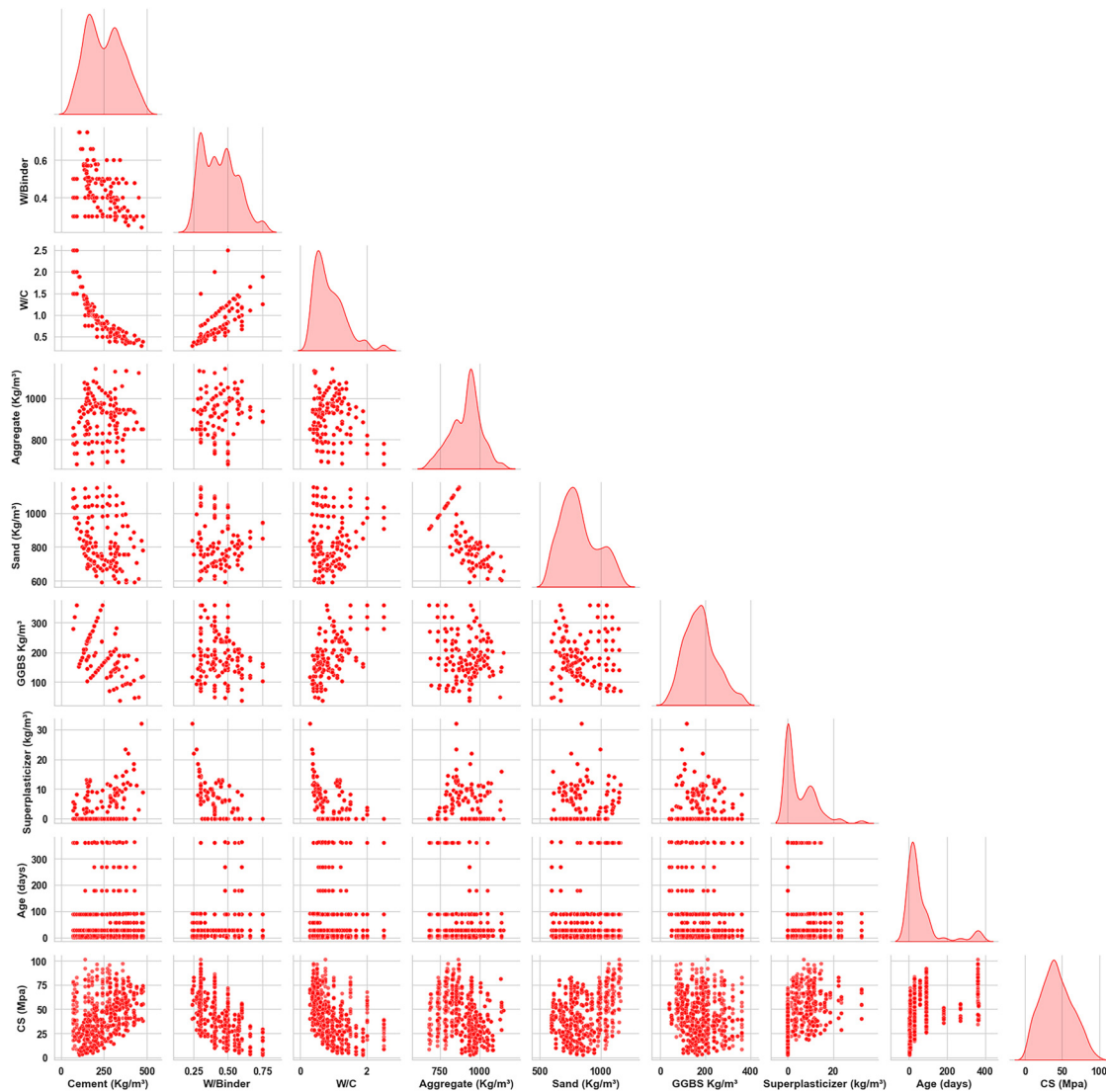
Figure 3: Pairplot of the GGBS concrete dataset showing the distribution and pairwise relationships among input variables and CS.

Table 1: Statistics obtained for the parameters used in forecasting the CS of concrete material.

Parameters	Cement (Kg/m³)	Water/Binder	W/Cement	Aggregate (Kg/m³)	Sand (Kg/m³)	GGBS Kg/m³	Admixture (kg/m³)	Age (days)
Mean	253.36	0.44	0.88	912.78	818.47	177.47	5.1	64.07
Standard error	3.72	0	0.02	3.34	5.33	2.53	0.24	3.42
Median	240	0.41	0.75	932	800	173	1.75	28
Mode	425	0.3	0.67	932	594	189	0	28
Standard deviation	104.39	0.13	0.47	93.73	149.59	71.09	6.65	96.06
Skewness	0.19	0.43	1.24	−0.26	0.5	0.47	1.5	2.27
Range	405	0.51	2.21	461.3	560.25	322	32.2	362
Lower	70	0.24	0.29	683.7	594	38	0	3
Higher	475	0.75	2.5	1,145	1,154.25	360	32.2	365
Confidence level (95.0 %)	7.3	0.01	0.03	6.56	10.47	4.97	0.47	6.72

ratios ranging from 0.24 to 0.75. Cement, aggregate, and admixture quantities also show high variability, with coefficients of variation exceeding 20 % for most parameters. Such dispersion confirms the dataset's suitability for capturing diverse mix behaviours and realistic material variability required for machine-learning-based modelling.

2.3 Data preprocessing and feature engineering

The raw dataset, retrieved from multiple published literature sources, was initially compiled and formatted for compatibility with Python-based ML workflows. Basic preprocessing steps were conducted to ensure consistency in units and variable representation across all records. No imputation or categorical encoding was required since all features were numerical and complete. It is essential to distinguish between the two water ratios used in this study: the water-to-binder ratio (W/B) includes cement and GGBS in the denominator, whereas the water-to-cement ratio (W/C) considers only cement. Including both ratios enables the models to capture how GGBS replacement modifies effective water demand and hydration behaviour compared with conventional mixes.

Exploratory data analysis (EDA) was performed using histograms and pair plots to assess each feature's distribution and identify potential correlations and outliers visually. Histograms revealed the spread and skewness of key input variables such as cement, GGBS, water content, and curing age. Pair plots enabled the identification of linear and nonlinear trends between predictors and the target variable (compressive strength), aiding in understanding variable importance and potential multicollinearity. Feature scaling or normalisation was not applied at this stage, as the selected machine learning models are either inherently robust to unscaled inputs (e.g., decision-tree-based AdaBoost) or can tolerate varied feature magnitudes with proper regularisation. No additional derived features were introduced beyond those provided in the dataset. However, attention was given to preserving meaningful predictors such as the water-to-binder ratio (W/B), which was explicitly included as an input feature. The dataset was randomly partitioned into training (70 %) and testing (30 %) subsets, ensuring representative distribution across the target variable. This split was used consistently across all machine learning models to allow fair and comparable evaluation of predictive performance. A 70/30 train test split was selected due to its proven balance between providing adequate data for training and retaining sufficient unseen data for reliable generalisation. Alternative ratios of 80/20 and 75/25 were also examined and yielded

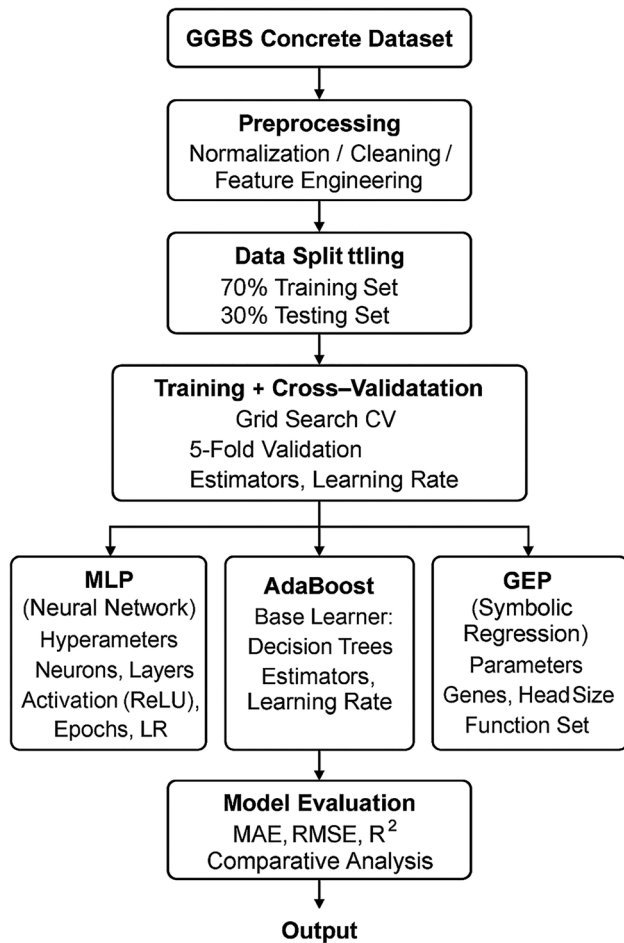consistent performance trends, confirming the stability of the developed models.

2.4 Machine learning model pipeline

The ML pipeline implemented in this study was designed to systematically develop, train, and evaluate three distinct algorithms: MLP, AdaBoost, and GEP. All modelling work was performed using Python 3.9, executed within the Spyder IDE as part of the Anaconda Navigator distribution. This environment offered robust package management and reproducibility, with core libraries including scikit-learn (for MLP and AdaBoost), NumPy, Pandas, Matplotlib, and gplearn (for GEP implementation). After data preprocessing, the dataset was split into training (70 %) and testing (30 %) subsets using randomised sampling to avoid sampling bias. Each model was trained exclusively on the training set, with hyperparameter tuning performed via grid search and cross-validation where applicable. MLP was configured as a feedforward neural network with multiple hidden layers. The number of neurons, activation functions (e.g., ReLU), learning rate, and epochs were optimised to prevent underfitting or overfitting. AdaBoost was implemented using a series of decision tree regressors as weak learners. The number of estimators and the learning rate were tuned to balance bias and variance. GEP, a symbolic regression approach, was employed to derive an interpretable mathematical expression representing the relationship between input features and CS. Parameters such as the number of genes, head size, and function set were configured to optimise model expressiveness and convergence.

All models were evaluated on the same testing set using consistent metrics for fair comparison. The modular pipeline structure allowed for easy replication, experimentation, and adaptation of models, ensuring flexibility for further research or real-world deployment. The graphical representation of the model pipeline for the presented research work is shown in Figure 4.

2.5 Evaluation metrics and performance analysis

Multiple evaluation metrics were used to assess the predictive accuracy and robustness of the ML models developed in this research. These metrics were selected to capture different aspects of model performance, ranging from overall fit to error magnitude, and to facilitate meaningful comparison between the Multilayer Perceptron (MLP),

GGBS Concrete Dataset

↓

Preprocessing
Normalization / Cleaning /
Feature Engineering

↓

Data Split ttling
70% Training Set
30% Testing Set

↓

Training + Cross–Validatation
Grid Search CV
5-Fold Validation
Estimators, Learning Rate

MLP
(Neural Network)
Hyperameters
Neurons, Layers
Activation (ReLU),
Epochs, LR

AdaBoost
Base Learner:
Decision Trees
Estimators,
Learning Rate

GEP
(Symbolic
Regression)
Parameters
Genes, Head Size
Function Set

Model Evaluation
MAE, RMSE, R^2
Comparative Analysis

↓

Output

Figure 4: Graphical representation of the model's pipeline adopted for the presented research work.

AdaBoost, and Gene Expression Programming (GEP) models. The primary performance indicators used include:

The model's predictive performance was evaluated using three key metrics. The Coefficient of Determination (R^2) quantifies how well the model explains variability in the target variable, with values approaching one indicating a strong fit. The Root Mean Square Error (RMSE) assesses prediction accuracy by emphasising larger deviations through the squared error component. Meanwhile, the Mean Absolute Error (MAE) offers a straightforward, scale-dependent measure by averaging the absolute differences between predicted and observed values. Table 2 presents a detailed overview of the performance indicators employed to evaluate the proposed ML models' predictive efficiency comprehensively. These metrics were computed using the test dataset to ensure unbiased evaluation, with all values rounded to three decimal places for clarity. In addition to quantitative metrics, residual plots, prediction versus actual scatter plots, and histograms of error distribution were

Table 2: Predictive accuracy measure for the employed models.

| Metric	Formula	Description
Mean Absolute Error (MAE)	$MAE = \frac{1}{n} \sum_{i=0}^{n} |y_i - \hat{y}_i|$	Average magnitude of prediction errors; measures absolute deviation between predicted and experimental values.
Mean Squared Error (MSE)	$MSE = \frac{1}{n} \sum_{i=0}^{n} (y_i - \hat{y}_i)^2$	Penalises larger deviations more strongly by squaring error terms.
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^{n} (y_i - \hat{y}_i)^2}$	Square root of MSE; interpretable in the same units as the target variable.
Adjusted R2	$Adj.R2 = 1 - \frac{(1-R^2)(n-1)}{n-p-1}$	Adjust R2 for the number of predictors (ρ) to account for model complexity and avoid overfitting.

where y_i = actual value, $\hat{y}_i$ = predicted value, $\bar{y}$ = mean of actual values, n = number of samples, and p = number of predictors.

generated to assess model behaviour and error patterns visually. Such plots help in identifying overfitting, heteroscedasticity, or systematic under- or over-prediction. Comparative analysis has been done to determine which model offered the most balanced performance across all metrics. Emphasis was placed not only on high R^2 values but also on minimising RMSE and MAE, thereby ensuring that the selected model was both accurate and practically reliable for real-world strength prediction tasks.

Table 2 presents a detailed overview of the performance indicators employed to evaluate the proposed ML models' predictive efficiency comprehensively [68].

Moreover, the execution process adopted in the study of the 5-fold validation approach is shown in Figure 5.

Interpretability analyses were performed post-model training to improve model transparency and provide
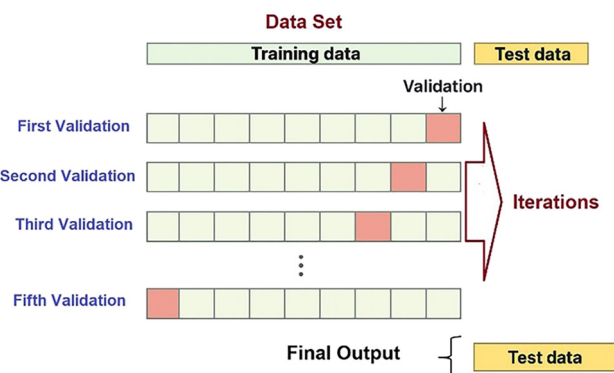
Figure 5: The adopted 5-fold validation approach for the research work indicates the process.
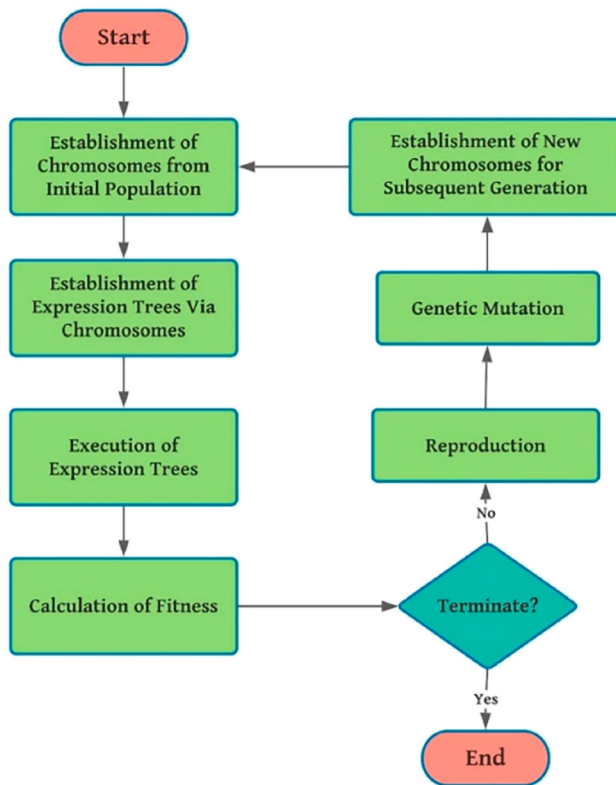
Figure 6: Flowchart indicating the execution process of the GEP model [69].

engineering-relevant insights. Local Interpretable Model-Agnostic Explanations (LIME) were used to assess the contribution of input variables to individual predictions. At the same time, permutation-based sensitivity analysis quantified the global impact of each feature across the dataset. These complementary techniques offered both localised and generalised understanding of model decisions, reinforcing the reliability of ML-driven strength predictions for GGBS concrete.

3 Machine learning models

This study utilised three ML algorithms, GEP, AdaBoost, and MLP, to develop predictive models for estimating the CS of GGBS-based concrete. Each model brings distinct advantages in capturing complex, nonlinear interactions among input features typically encountered in concrete mix design. The selection of three distinct algorithms, MLP, AdaBoost, and GEP, was deliberate to balance predictive performance and engineering interpretability. MLP offered the highest numerical accuracy, whereas GEP produced an explicit symbolic equation that enables direct analytical estimation of compressive strength. This complementary design allows

the framework to satisfy both research and practical implementation needs within sustainable concrete design.

3.1 Gene expression programming (GEP)

GEP is a nature-inspired evolutionary algorithm that evolves computer programs or mathematical expressions capable of modelling intricate relationships within a dataset. In the context of this research, GEP was implemented to derive symbolic expressions linking the mix design variables to the target output, compressive strength. The model operates by generating a population of chromosomes composed of genes, which subsequently evolve through selection, mutation, and crossover operations. The final model not only enables accurate prediction but also offers a transparent mathematical formulation, facilitating interpretability and integration into design frameworks. The prediction process by GEP is shown in Figure 6 [69].

3.2 Adaptive boosting (AdaBoost)

AdaBoost enhances predictive accuracy by iteratively combining multiple weak learners, such as decision stumps, into a single strong model. In this work, AdaBoost was employed to iteratively improve prediction accuracy by minimising the residual error of each subsequent learner. The model assigns higher weights to poorly predicted samples in each iteration, enabling the ensemble to focus on challenging data points. Its strength lies in enhancing generalisation while maintaining computational efficiency, making it suitable for structured datasets such as those derived from concrete mix designs. The steps involved in the
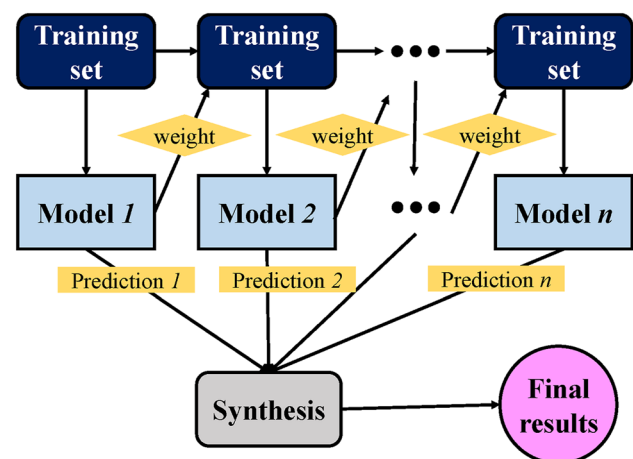
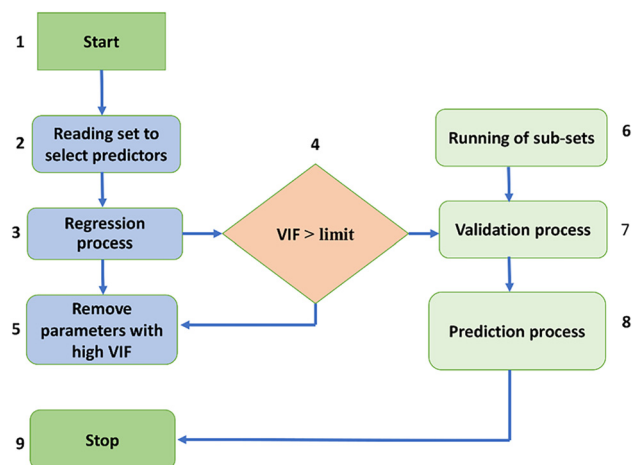**Figure 7:** Steps involved for the final prediction by AdaBoost model [70].

execution process for the predicted outcomes by AdaBoost is shown in Figure 7 [70].

3.3 Multilayer perceptron (MLP)

MLP is a class of feedforward artificial neural networks designed to approximate complex nonlinear functions. As illustrated in Figure 8, the MLP model consists of an input layer corresponding to the selected mix parameters, followed by two hidden layers with non-linear activation functions that capture the complex relationships between variables [71]. The model is trained using backpropagation, where weights are iteratively updated to minimise prediction error. The final single-neuron output layer produces the estimated compressive strength. This workflow ensures that both direct and interactive effects of input parameters are effectively learned by the network. The model's design includes an input layer for the input variables, followed by one or more hidden layers using ReLU activations, and a single-node output layer for compressive strength estimation. MLP was trained using the backpropagation algorithm with mean squared error as the loss function. Input features were normalised to improve convergence. Its data-driven nature and capability to learn hidden patterns without requiring predefined equations make MLP an effective tool for modelling material behaviour.

Together, these models offer complementary strengths: GEP provides symbolic interpretability, AdaBoost enhances predictive robustness through boosting, and MLP captures deep nonlinear associations in the data. Their collective application offers a multi-perspective approach to modelling the CS of GGBS-based concrete. The flowchart of the MLP model for predicting the outputs is depicted in Figure 8.
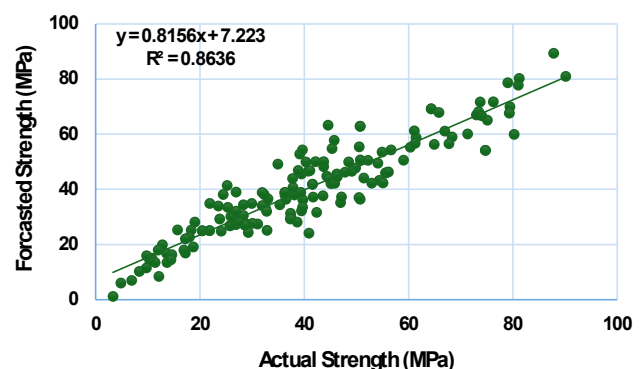
4 Results and discussions

4.1 GEP model outcome

The GEP model demonstrated a strong predictive capability for estimating the CS of GGBS-based concrete, as evidenced by the correlation between the experimental and forecasted results shown in Figure 9. The model yielded a high R^2 value of 0.8636, indicating that approximately 86.4 % of the variance in the actual CS values can be explained by the GEP model's predictions. The fitted regression line, defined by the equation $y = 0.8156x + 7.223$, further confirms the strong linear relationship, although a slight underestimation trend is noticeable at higher strength ranges.

To further assess the model's performance, the absolute differences among the predicted and actual CS values are illustrated in Figure 10. This residual error distribution provides deeper insight into the precision and generalisation ability of the GEP model. The analysis shows that the prediction error ranges from a minimum of 0.010 MPa to a maximum of 20.56 MPa. Notably, a substantial portion of the predictions fall within acceptable error bounds: 60.37 % of the data points exhibit an absolute error between 0 and 6 MPa, indicating a high degree of precision in the majority of the cases. An additional 27.67 % of predictions lie within an error range of six–12 MPa, reflecting moderate deviation. Only 11.32 % of the test set displayed errors exceeding 12 MPa, which could be attributed to a combination of data outliers, complex material interactions, or limitations in the model's sensitivity to specific input features.

These findings confirm that the GEP model can be considered a robust and reliable approach for modelling the

Figure 8: Schematic workflow of the MLP model showing the flow of data from input parameters through hidden layers to the output node responsible for compressive strength prediction [71].

Figure 9: Relationship of the forecasted and actual CS of the GGBS-based concrete using the GEP model.
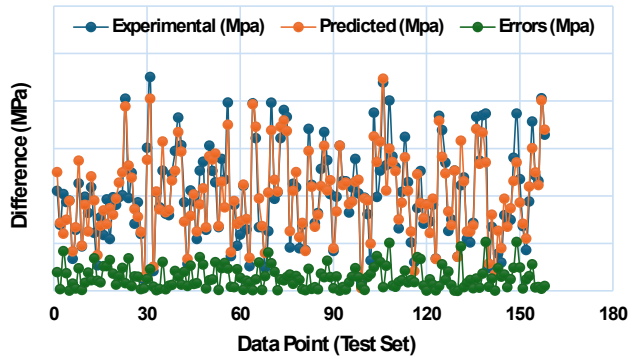
Figure 10: Evaluation of the GEP model performance through actual versus predicted CS of GGBS concrete.
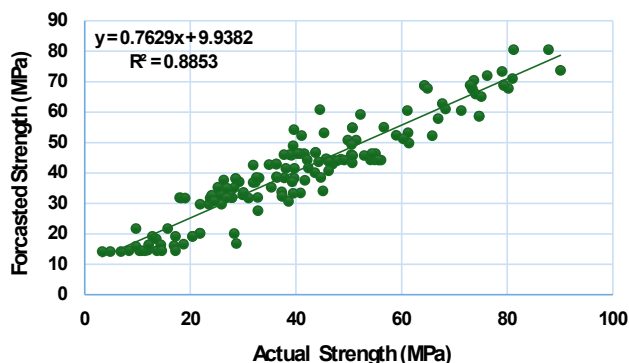

Figure 12: Evaluation of the AdaBoost model performance through actual versus predicted CS of GGBS concrete.

CS of sustainable concrete mixes incorporating GGBS as a partial cement replacement. The results support its suitability for preliminary strength prediction in practical engineering applications, particularly when interpretability and transparency of the underlying equation are valued.
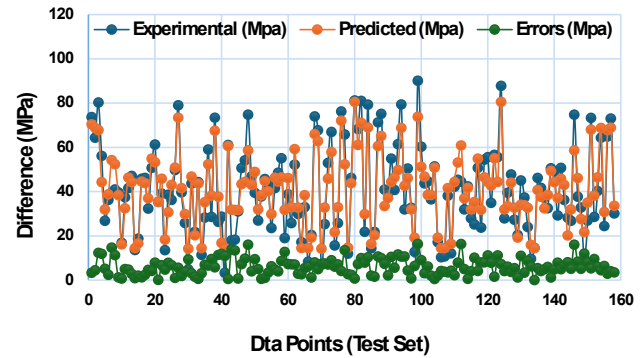
4.2 AdaBoost model outcome

The AdaBoost regression model also demonstrated strong predictive capability in estimating the CS of concrete incorporating GGBS. As illustrated in Figure 11, the relationship between the experimental and predicted CS values shows a solid linear trend with a high R^2 = 0.8853. This represent that the model can explain approximately 88.5 % of the variation in the actual strength values. The regression equation, $y = 0.7629x + 9.9382$, reveals a slight underestimation in the slope, implying that while the trend is accurately captured, the model tends to compress higher values toward the mean slightly.

A deeper examination of prediction reliability is presented in Figure 12, which visualizes the error distribution across the test dataset. The residual analysis reveals a
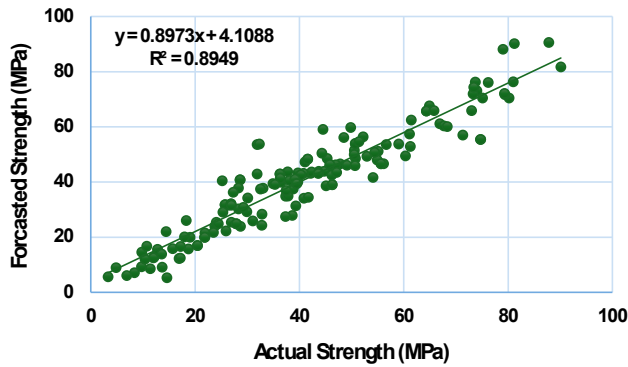
maximum error of 16.32 MPa and a minimum of 0.064 MPa, with the majority of predictions demonstrating high accuracy: Approximately 52.0 % of the data points fall within an absolute error range of 0–6 MPa, indicating high prediction precision for over half of the dataset. An additional 40.25 % of predictions fall within the six–12 MPa error range, suggesting moderate but acceptable deviations in complex cases. Only 6.91 % of the predictions exceeded a 12 MPa error, reflecting a relatively low frequency of significant outliers.

Compared to the GEP model, AdaBoost demonstrates slightly improved accuracy in terms of R^2 and a tighter error distribution, making it a promising candidate for practical applications where generalisation is critical. Its performance further underscores the utility of ensemble learning approaches in capturing nonlinear interactions in concrete mix constituents and their effect on compressive strength.

4.3 MLP model outcome

The MLP was the top-performing model in forecasting the CS of GGBS concrete mixes. As presented in Figure 13, a strong linear relationship is observed between the actual and predicted values, characterised by a high coefficient of determination (R^2 = 0.8949). The corresponding regression equation, $y = 0.8973x + 4.1088$, suggests that the MLP model effectively captures the underlying trends in the data with minimal bias, as evidenced by the slope approaching unity and a relatively low intercept.

The prediction residuals, detailed in Figure 14, offer a detailed view of the model's error distribution across the test set. The analysis reveals a maximum error of 21.74 MPa and a minimum of just 0.014 MPa, indicating that while outliers exist, the majority of the predictions are highly accurate. The distribution of the absolute errors is as follows: 62.26 % of the predictions have an absolute error between 0 and 6 MPa,

Figure 11: Relationship of the forecasted and actual CS of the GGBS-based concrete using the AdaBoost model.
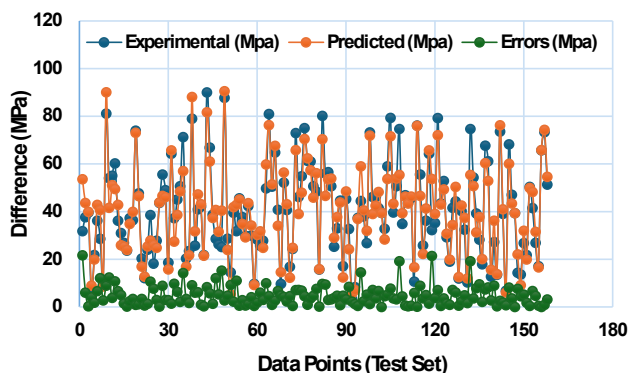
Figure 13: Relationship of the forecasted and actual CS of the GGBS-based concrete using the MLP model.

demonstrating excellent predictive precision for nearly two-thirds of the dataset. 26.55 % of the predictions fall within the six–12 MPa range, still representing acceptable performance given the variability inherent in concrete materials. Only 6.28 % of the predictions exhibit an error greater than 12 MPa, suggesting that significant deviations are relatively rare.

This level of accuracy indicates that the MLP model is particularly well-suited for capturing the nonlinear interactions between the mix constituents and their influence on the resulting CS. Compared to GEP and AdaBoost, MLP shows the best R^2 and the lowest percentage of high-error predictions, reinforcing its robustness and adaptability for predictive modelling in sustainable concrete solutions.

4.4 Interaction analysis using surface and contour plots

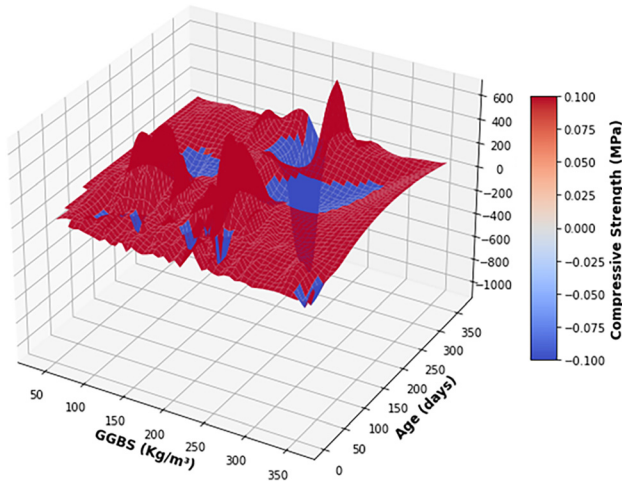To further validate and interpret the nonlinear trends captured by the machine learning models, both a 3D surface plot and a 2D contour plot were generated to explore the interactive effect of GGBS content and curing age on the CS of concrete. These plots provide visual support to the results obtained by the MLP, AdaBoost, and GEP models, particularly in highlighting the dominance of age and GGBS in strength development, which is further explained in Section 7 (Model Interpretability and Feature Analysis).

The 3D surface plot (Figure 15) reveals a distinct nonlinear increase in CS as both GGBS content and age increase. Notably, strength improvement becomes more significant beyond 150 kg/m³ of GGBS and 90 days of curing, indicating a synergistic effect between GGBS hydration and prolonged curing. This trend aligns with the MLP model's superior predictive performance ($R^2 = 0.8949$), which effectively captured such high-order interactions. The surface topography, with curved ridges and valleys, visually represents the complex relationships that the GEP model approximated symbolically and the AdaBoost model generalised through boosting iterations. Complementarily, the 2D contour plot (Figure 16) projects these interactions onto a planar format, where contour bands delineate regions of equal strength. This format simplifies the identification of strength thresholds and can aid practical mix design. The plot confirms that higher strengths are concentrated in the upper-right region of the graph, where both GGBS dosage and age are at elevated levels. These visual findings corroborate the importance rankings from permutation-based sensitivity analysis, which identified age and water-related ratios as the most influential features. Together, these plots not only validate the predictive insights from the ML models but also provide engineers with an interpretable, data-driven method for fine-tuning concrete mix compositions to achieve optimal performance and sustainability. They bridge the gap between black-box predictions and practical understanding, supporting transparent decision-making in GGBS-based concrete design.
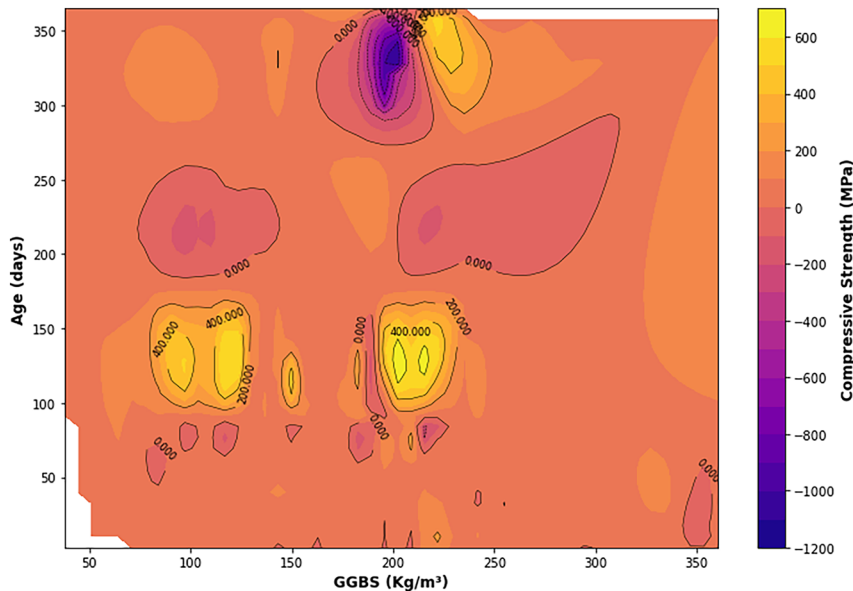
5 Cross-validation (5-fold)

Model robustness and generalisation were assessed using 5-fold cross-validation, with results reported in Table 3. The evaluation was based on three key statistical indicators: MAE, RMSE, and the R^2 for all three models: GEP, MLP, and AdaBoost.

Across all folds, the MLP model consistently outperformed the GEP and AdaBoost models, achieving the lowest average MAE and RMSE values and the highest R^2, particularly evident in folds three and 5, where it recorded R^2 values of 0.92 and 0.93, respectively. The average MAE and RMSE values for MLP across the folds were also the most

Figure 14: Evaluation of the MLP model performance through actual versus predicted CS of GGBS concrete.

Figure 15: 3D surface plot illustrating the nonlinear relationship between GGBS content, curing age, and compressive strength.

algorithm's superior capability to capture complex, nonlinear behaviour in the dataset. The AdaBoost model demonstrated moderate consistency, with MAE values between 5.86 and 7.52 MPa and RMSE values ranging from 6.92 to 9.04 MPa. While its performance was generally stable, it struggled slightly in the fourth fold, where it recorded the lowest R^2 value (0.81), suggesting reduced generalisation in that subset. In contrast, the GEP model, while advantageous for its symbolic interpretability, exhibited slightly less predictive power in comparison. It showed R^2 values in the range of 0.85–0.90, with somewhat higher RMSE and MAE values, particularly in the fourth and fifth folds. This suggests that although GEP can capture certain relationships, it may lack the flexibility required for modelling the nonlinear multivariate interactions commonly found in GGBS-concrete strength prediction.

Figure 16: 2D contour plot showing compressive strength zones as a function of GGBS content and age.

Table 3: Statistical results of the 5-fold cross-validation for all the employed models.

K-fold	GEP			MLP			AdaBoost		
	MAE (MPa)	RMSE (MPa)	R^2	MAE (MPa)	RMSE (MPa)	R^2	MAE (MPa)	RMSE (MPa)	R^2
1	5.64	7.35	0.86	4.74	6.25	0.90	5.86	6.92	0.88
2	5.16	6.81	0.90	5.43	7.76	0.87	6.11	8.09	0.86
3	5.41	7.16	0.90	4.90	6.22	0.92	6.69	8.06	0.87
4	6.40	8.02	0.85	4.68	6.03	0.92	7.52	9.04	0.81
5	5.93	7.76	0.86	4.37	5.54	0.93	6.70	8.09	0.84

favourable, with an overall MAE range of 4.37–5.43 MPa, and RMSE between 5.54 and 7.76 MPa. This reflects the MLP

These results are particularly relevant given the nonlinear and highly interdependent nature of the
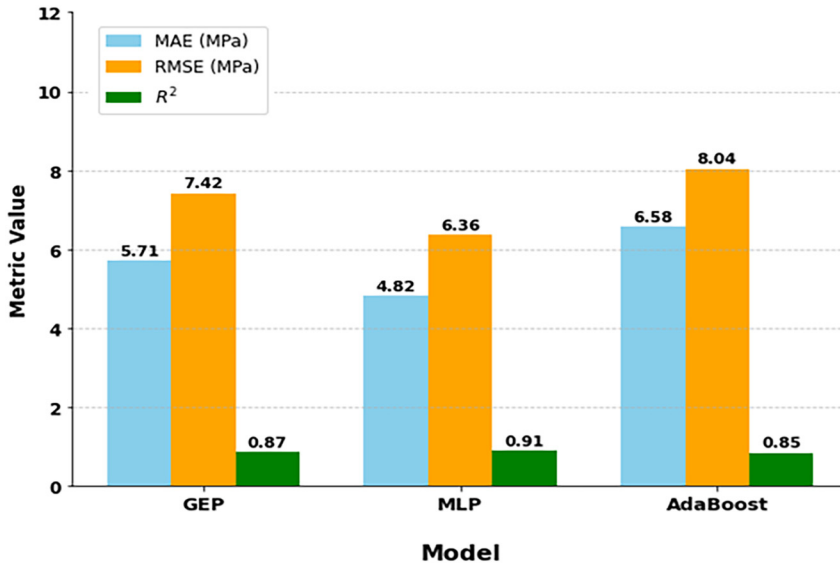
Figure 17: Comparison of average performance metrics (MAE, RMSE, and R^2) across the GEP, MLP, and AdaBoost models based on 5-fold cross-validation.

parameters influencing the CS of GGBS-based concrete. Factors such as binder content, water-to-binder ratio, curing time, and the proportion of GGBS replacement exhibit nonlinear synergistic effects on strength development. The MLP model, by virtue of its layered architecture and non-linear activation functions, is better equipped to approximate these interactions without prior assumptions about their structure. On the other hand, GEP, while advantageous for producing closed-form expressions, may not fully encapsulate the nuanced nonlinearities, and AdaBoost's iterative decision-tree boosting, though powerful, might be limited in extrapolative performance. Therefore, the cross-validation outcomes strongly support the MLP model as the most reliable and generalisable tool for predicting compressive strength in sustainable concrete incorporating GGBS, especially in scenarios with diverse and complex input-variable relationships. The summarised result of the models indicates the overall accuracy level using the selected metrics, as shown in Figure 17. To further visualise and compare the predictive capabilities of the developed models, a Taylor diagram, as shown in Figure 18, was constructed. The diagram simultaneously illustrates the correlation, standard deviation, and centred RMSE for the MLP, AdaBoost, and GEP models.

The small error margins obtained for the developed models, where more than 60 % of predictions fall within ± 6 MPa and less than 10 % exceed ± 12 MPa, demonstrate promising practical applicability. Such accuracy levels are acceptable for early-stage decision-making and optimisation of GGBS concrete mix designs, allowing practitioners to minimise the number of laboratory trials required. It is noted that the models are not intended to replace design-code validation but rather to serve as a

reliable, data-driven support tool for enhancing mix development efficiency in real-world construction workflows.

5.1 GEP-based Derived Equation

The final expression derived from the GEP model represents a significant analytical outcome of this study. The explicit functional form, as illustrated in Equation (4), establishes a robust empirical relationship between the concrete mix constituents and the resulting CS. The symbolic regression capability of GEP has enabled the extraction of a transparent, human-interpretable formula that encapsulates the
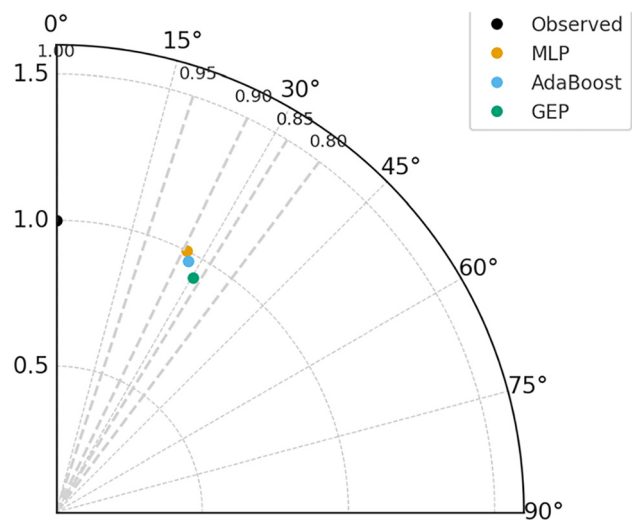
Figure 18: Taylor diagram comparing the predictive performance of the MLP, AdaBoost, and GEP models for compressive strength prediction of GGBS-based concrete.

complex nonlinear interactions among the input variables. This derived equation is not merely a regression surrogate, but a predictive model grounded in data-driven optimisation. Each term within the equation reflects meaningful interactions, such as the ratio of activator to binder and composite logarithmic transformations involving the GGBS-to-binder and activator concentration parameters. These terms highlight the multivariate dependencies and synergistic effects that govern the strength development in GGBS-based concrete.

$$f(X) = A + B + C + D \tag{4}$$

$$A = \frac{\sqrt{X_4}}{X_2}$$

$$B = 3.\, log\left(0.195\,\frac{X_7^2}{X_0}\right) + log\left[(X_1 - X_2)^2\right]$$
$$+ log\left[\left(\frac{X_5}{X_0} - X_2\right)(X_1 - X_2)\right]$$

$$C = \sqrt{\frac{X_5}{X_6}} + log(X_7) + log\left(X_1 - \frac{X_5}{X_0}\right) + log\left(\frac{X_7}{X_6}\right)$$

$$D = log\left[\left(X_7\,(-0.329)\left(X_1 - \frac{X_5}{X_0}\right)^2\right)\right] + log\left[x_7\,(-0.329)\,\frac{X_5}{X_0}\right]$$

Where,

    X0 = Cement.
    X1 = GGBS.
    X2 = Water.
    X3 = Aggregate.
    X4 = Sand.
    X5 = W/B Ratio.
    X6 = Superplasticizers.
    X7 = Age.

From a practical standpoint, this equation serves as a compact and deployable tool for engineers and material scientists, enabling rapid estimation of CS with no need for costly or time-consuming laboratory experiments. The predictive precision of the GEP model, reflected in an R^2 equal to 0.863, underscores the reliability and generalisability of the model across unseen data points. Importantly, the equation not only enhances model interpretability, a standard limitation in black-box machine learning models, but also contributes to the ongoing shift toward explainable AI in civil engineering material design. By encapsulating the empirical knowledge learned from the dataset into a closed-form expression, this model aids in both forward prediction and reverse engineering of mix proportions to achieve target strength requirements. In the context of sustainable construction, such

predictive modelling aligns with the objectives of optimising material usage, reducing experimental iterations, and accelerating the deployment of low-carbon cementitious systems such as GGBS concrete. Hence, the derived GEP equation represents a pivotal outcome of this research, demonstrating both technical rigour and practical utility.
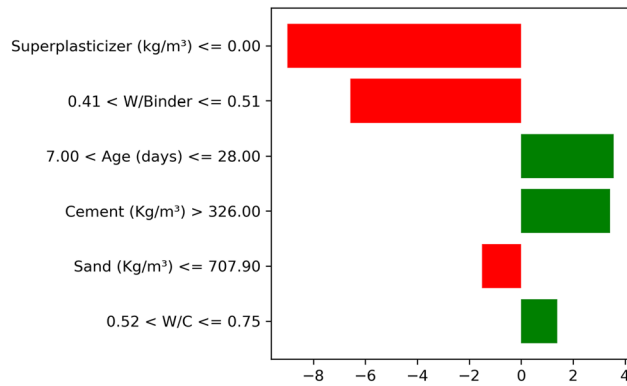
6 Model interpretability and feature analysis

To enhance transparency and support the reliability of the developed predictive model, interpretability techniques were employed to evaluate the relative influence of input variables. Both local and global analysis approaches were applied to gain a comprehensive understanding of the model's internal decision process and feature interactions. These insights are critical for validating machine learning predictions in the context of civil engineering, where input parameters directly relate to material behaviour.

It is acknowledged that post-hoc interpretability tools such as LIME and permutation sensitivity analysis do not replace formal design code verifications; however, they offer a transparent, quantitative understanding of variable influence that is consistent with established engineering judgement. These methods enable practitioners to trace the reasoning behind model outputs, compare feature contributions with known material behaviour, and verify that model decisions remain physically meaningful within regulatory safety margins.
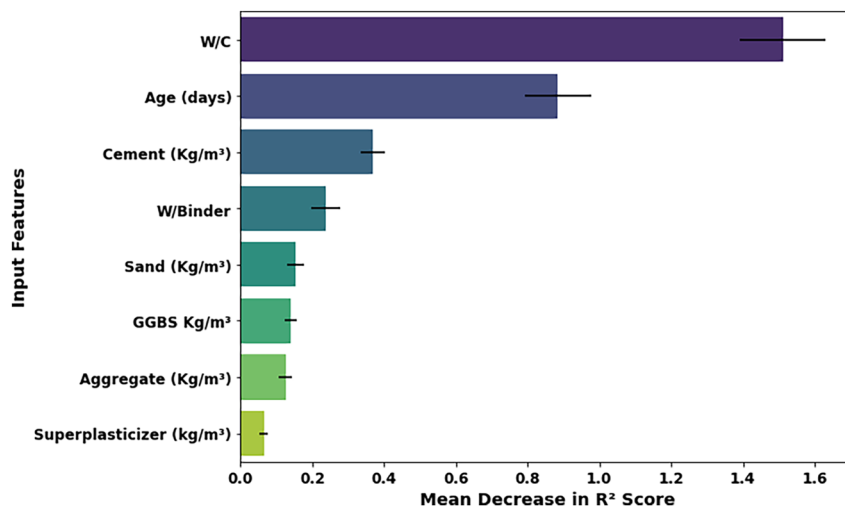
6.1 Local interpretation using LIME

Local interpretable model-agnostic explanations (LIME) were employed to examine how individual input features influenced the prediction of CS for a representative test instance. As shown in Figure 19, the model predicted a relatively lower compressive strength for this mix, primarily due to two dominant factors: the absence of superplasticizer (≤ 0.00 kg/m^3) and a water-to-binder (W/Binder) ratio between 0.41 and 0.51. Both features had strong negative contributions, consistent with the understanding that reduced workability and increased porosity adversely affect early-age strength.

Conversely, the model assigned positive weight to features such as specimen age (7–28 days) and cement content greater than 326 kg/m^3. These findings align with established concrete behaviour, where strength development is governed by the progression of hydration and sufficient binder

Figure 19: Local feature contribution to the predicted CS of a GGBS-based concrete mix using LIME.

also contributed meaningfully to model performance, while features like GGBS content, aggregate volume, and superplasticizer dosage had relatively lower global influence. The lower importance of GGBS may be attributed to its indirect and time-dependent effect, which is already captured through the age variable, or potentially due to limited variability within the dataset.

Together, these insights demonstrate that while local predictions may highlight case-specific sensitivities, the global analysis identifies parameters with consistent influence across the model's learning space. The combination of LIME and sensitivity analysis offers a well-rounded view of

Figure 20: Permutation sensitivity analysis with the MLP model revealed the global influence of each input feature.

availability. Notably, the influence of sand and water-to-cement (W/C) ratio was comparatively minor in this specific case. The LIME plot demonstrates the model's ability to learn physically meaningful relationships, providing confidence in its application to material design scenarios.

6.2 Global sensitivity analysis

To complement the local interpretation, global feature importance was evaluated using permutation-based sensitivity analysis. The results, illustrated in Figure 20, indicate that the water-to-cement (W/C) ratio and age were the most influential features across the entire dataset. This is consistent with the fundamental principles of concrete technology: the W/C ratio is a key driver of strength, durability, and porosity, while age governs the extent of hydration and pozzolanic activity, particularly relevant for GGBS-based systems. Interestingly, cement content and W/Binder ratio

the model's decision-making logic and affirms its compatibility with engineering expectations.

7 Conclusions

This research establishes a data-driven framework to predict the CS of GGBS-based concrete using advanced ML techniques. The modelling framework balances predictive accuracy and interpretability by integrating MLP and GEP approaches. The MLP model provides superior statistical precision, while the GEP model yields a transparent equation that can be readily applied in engineering calculations, ensuring the results are reliable and practically deployable.

Integrating MLP, AdaBoost, and GEP allows high predictive accuracy and practical usability through a symbolic, equation-based tool. The findings confirm that curing age and water-related parameters are dominant strength-governing factors, aligning with established material

behaviour while enhancing modelling transparency through explainable AI methods.

The proposed approach enables engineers to reduce trial-and-error experimentation and rapidly assess mix performance using the derived GEP equation. This supports resource-efficient design decisions and aligns with the broader goal of reducing carbon emissions by enabling higher, optimised use of SCMs such as GGBS.

While the current database is extensive (787 records), future work will incorporate broader durability-related properties and additional SCM combinations and develop a digital tool for interactive model use in industry. Overall, the presented framework provides a replicable and scalable path toward data-driven optimisation of low-carbon concrete.

Author contributions: M.N.A.: conceptualization, funding acquisition, supervision, project acquisition, methodology, writing, writing, reviewing, and editing. A.N.: data acquisition, methodology, software, investigation, writing-original draft. M.I.F.: resources, funding acquisition, visualization, formal analysis, writing, revieing, and editing. M.T.Q.: data acquisition, validation, funding acquisition, software, writing, reviewing, and editing. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Conflicts of interest: The authors state no conflict of interest.
Data availability statement: The datasets generated and/or analysed during the current study are available from the corresponding author upon reasonable request.

References

1. Cheng D, Reiner DM, Yang F, Cui C, Meng J, Shan Y, et al. Projecting future carbon emissions from cement production in developing countries. Nat Commun 2023;14:8213.
2. Shivaprasad KN, Yang H-M, Singh JK. A path to carbon neutrality in construction: an overview of recent progress in recycled cement usage. J CO2 Util 2024;83:102816.
3. Khan K, Ahmad W, Amin MN, Rafiq MI, Abu Arab AM, Alabdullah IA, et al. Evaluating the effectiveness of waste glass powder for the compressive strength improvement of cement mortar using experimental and machine learning methods. Heliyon 2023;9:e16288.
4. Wu S, Shao Z, Andrew RM, Bing L, Wang J, Niu L, et al. Global CO2 uptake by cement materials accounts 1930–2023. Sci Data 2024;11:1409.
5. Bouchelil L, Shah Bukhari SJ, Khanzadeh Moradllo M. Evaluating the performance of internally cured limestone calcined clay concrete mixtures. J Sustain Cement-Based Mater 2025;14:198–208.
6. Bukhari SJS, Khanzadeh Moradllo M. Multicriteria performance assessment of 'low w/c + low cement + high dosage admixture' concrete: environmental, economic, durability, and mechanical performance considerations. J Clean Prod 2025;523:146419.
7. Fahim AA, Bukhari SJS, Khanzadeh Moradllo M. Additive manufacturing of carbonatable ternary cementitious systems with cellulose nanocrystals. Constr Build Mater 2025;495:143753.
8. Ahmad W, McCormack SJ, Byrne A. Biocomposites for sustainable construction: a review of material properties, applications, research gaps, and contribution to circular economy. J Build Eng 2025;105:112525.
9. Gupta S, Chaudhary S. State of the art review on supplementary cementitious materials in India–II: characteristics of SCMs, effect on concrete and environmental impact. J Clean Prod 2022;357:131945.
10. Hu J, Ahmed W, Jiao D. A critical review of the technical characteristics of recycled brick powder and its influence on concrete properties. Buildings 2024;14:3691.
11. Ahmed W, Ye C, Lu G, Ng ST, Liu G, Wang Y. Low-carbon concrete comprising high-volume pozzolan and recycled aggregate: evaluating mechanical performance, microstructure, environmental impact, and cost efficiency. J Clean Prod 2025;518:145796.
12. Ndahirwa D, Zmamou H, Lenormand H, Leblanc N. The role of supplementary cementitious materials in hydration, durability and shrinkage of cement-based materials, their environmental and economic benefits: a review. Clean Mater 2022;5:100123.
13. Park S, Wu S, Liu Z, Pyo S. The role of supplementary cementitious materials (SCMs) in ultra high performance concrete (UHPC): a review. Materials 2021;14:1472.
14. Ahmed MM, Sadoon A, Bassuoni MT, Ghazy A. Utilizing agricultural residues from hot and cold climates as sustainable SCMs for low-carbon concrete. Sustainability 2024;16:10715.
15. Mohamad N, Muthusamy K, Embong R, Kusbiantoro A, Hashim MH. Environmental impact of cement production and solutions: a review., vol 48; 2022. p. 741–6. Mater Today Proc.
16. Barbhuiya S, Kanavaris F, Das BB, Idrees M. Decarbonising cement and concrete production: strategies, challenges and pathways for sustainable development. J Build Eng 2024;86:108861.
17. Scrivener KL, John VM, Gartner EM. Eco-efficient cements: potential economically viable solutions for a low-CO2 cement-based materials industry. Cement Concr Res 2018;114:2–26.
18. Miller SA, Horvath A, Monteiro PJM. Impacts of booming concrete production on water resources worldwide. Nat Sustain 2018;1:69–76.
19. Jamil M, Zhao W, He N, Gupta MK, Sarikaya M, Khan AM, et al. Sustainable milling of Ti–6Al–4V: a trade-off between energy efficiency, carbon emissions and machining characteristics under MQL and cryogenic environment. J Clean Prod 2021;281:125374.
20. Terán-Cuadrado G, Tahir F, Nurdiawati A, Almarshoud MA, Al-Ghamdi SG. Current and potential materials for the low-carbon cement production: life cycle assessment perspective. J Build Eng 2024;96:110528.
21. Orozco C, Babel S, Tangtermsirikul S, Sugiyama T. Comparison of environmental impacts of fly ash and slag as cement replacement materials for mass concrete and the impact of transportation. Sustain Mater Technol 2024;39:e00796.

22. Ahmad J, Kontoleon KJ, Majdi A, Naqash MT, Deifalla AF, Ben Kahla N, et al. A comprehensive review on the ground granulated blast furnace slag (GGBS) in concrete production. Sustainability 2022;14:8783.

23. Zhao Q, Pang L, Wang D. Adverse effects of using metallurgical slags as supplementary cementitious materials and aggregate: a review. Materials 2022;15:3803.

24. Rattanadecho P, Makul N, Pichaicherd A, Chanamai P, Rungroungdouyboon B. A novel rapid microwave-thermal process for accelerated curing of concrete: prototype design, optimal process and experimental investigations. Constr Build Mater 2016;123:768–84.

25. Shi C, Jiménez AF, Palomo A. New cements for the 21st century: the pursuit of an alternative to Portland cement. Cement Concr Res 2011; 41:750–63.

26. Meng T, Yu Y, Qian X, Zhan S, Qian K. Effect of nano-TiO2 on the mechanical properties of cement mortar. Constr Build Mater 2012;29: 241–5.

27. Meddah MS, Zitouni S, Belâabes S. Effect of content and particle size distribution of coarse aggregate on the compressive strength of concrete. Constr Build Mater 2010;24:505–12.

28. Zhang Y, Tang Z, Liu X, Zhou X, He W, Zhou X. Study on the resistance of concrete to high-concentration sulfate attack: a case study in jinyan bridge. Materials 2024;17:3388.

29. Shi C, Qian J. High performance cementing materials from industrial Slags—a review. Resour Conserv Recycl 2000;29:195–207.

30. Engida TG, Rao X, Berentsen PBM, Oude Lansink AGJM. Measuring corporate sustainability performance– the case of European food and beverage companies. J Clean Prod 2018;195:734–43.

31. Habert G, Miller SA, John VM, Provis JL, Favier A, Horvath A, et al. Environmental impacts and decarbonization strategies in the cement and concrete industries. Nat Rev Earth Environ 2020;1:559–73.

32. Bernal SA, Provis JL. Durability of alkali-activated materials: progress and perspectives. J Am Ceram Soc 2014;97:997–1008.

33. Thomas BS, Gupta RC. A comprehensive review on the applications of waste tire rubber in cement concrete. Renew Sustain Energy Rev 2016; 54:1323–33.

34. Turner LK, Collins FG. Carbon dioxide equivalent (CO2-e) emissions: a comparison between geopolymer and OPC cement concrete. Constr Build Mater 2013;43:125–30.

35. Lawrence A, Nehler T, Andersson E, Karlsson M, Thollander P. Drivers, barriers and success factors for energy management in the Swedish pulp and paper industry. J Clean Prod 2019;223:67–82.

36. Xu J, Jiang L, Wang J. Influence of detection methods on chloride threshold value for the corrosion of steel reinforcement. Constr Build Mater 2009;23:1902–8.

37. Pourjavadi A, Fakoorpoor SM, Hosseini P, Khaloo A. Interactions between superabsorbent polymers and cement-based composites incorporating colloidal silica nanoparticles. Cement Concr Compos 2013;37:196–204.

38. He Z-h., Li L-y., Du S-g. Creep analysis of concrete containing rice husk ash. Cement Concr Compos 2017;80:190–9.

39. Xu S, Wang W, Dong S, Li Q, Liu X, Peng Y. Thermal stability study of fiber-reinforced cementitious composites with high ductility under high-temperature casting. Constr Build Mater 2021;282:122700.

40. Anand P, Singh SD, Bhowmik PN, Kontoni D-PN. Optimizing concrete mix proportions with zeolite, GGBS, and CDW: a data-driven approach integrating experimental analysis and machine learning models. Eng Res Express 2025;7:015105.

41. Dash PK, Parhi SK, Patro SK, Panigrahi R. Efficient machine learning algorithm with enhanced cat swarm optimization for prediction of compressive strength of GGBS-based geopolymer concrete at elevated temperature. Constr Build Mater 2023;400:132814.

42. Valenza JJ, Thomas JJ. Permeability and elastic modulus of cement paste as a function of curing temperature. Cement Concr Res 2012;42: 440–6.

43. Ahmad W, Veeraghantla VSSCS, Byrne A. Advancing sustainable concrete using biochar: experimental and modelling Study for mechanical strength evaluation. Sustainability 2025;17:2516.

44. Poon CS, Ho DWS. A feasibility study on the utilization of r-FA in SCC. Cement Concr Res 2004;34:2337–9.

45. Yang K, Yang C, Zhu X. High volume ground granulated blast-furnace slag cement, high-volume mineral admixtures in cementitious binders. Cambridge, UK: Elsevier; 2025:1–29 pp.

46. Sezavar R, Shafabakhsh G, Mirabdolazimi SM. New model of moisture susceptibility of nano silica-modified asphalt concrete using GMDH algorithm. Constr Build Mater 2019;211:528–38.

47. Chou J-S, Pham A-D. Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength. Constr Build Mater 2013;49:554–63.

48. Yeh IC. Modeling of strength of high-performance concrete using artificial neural networks. Cement Concr Res 1998;28:1797–808.

49. Yang H, Bai L, Duan Y, Xie H, Wang X, Zhang R, et al. Upcycling corn straw into nanocelluloses via enzyme-assisted homogenization: application as building blocks for high-performance films. J Clean Prod 2023;390:136215.

50. Asteris PG, Skentou AD, Bardhan A, Samui P, Pilakoutas K. Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. Cement Concr Res 2021;145:106449.

51. Mansouri E, Manfredi M, Hu J-W. Environmentally friendly concrete compressive strength prediction using hybrid machine learning. Sustainability 2022;14:12990.

52. Farooq F, Ahmed W, Akbar A, Aslam F, Alyousef R. Predictive modeling for sustainable high-performance concrete from industrial wastes: a comparison and optimization of models using ensemble learners. J Clean Prod 2021;292:126032.

53. Cook R, Lapeyre J, Ma H, Kumar A. Prediction of compressive strength of concrete: critical comparison of performance of a hybrid machine learning model with standalone models. J Mater Civ Eng 2019;31: 04019255.

54. Kashem A, Karim R, Das P, Datta SD, Alharthai M. Compressive strength prediction of sustainable concrete incorporating rice husk ash (RHA) using hybrid machine learning algorithms and parametric analyses. Case Stud Constr Mater 2024;20:e03030.

55. Hameed MM, Abed MA, Al-Ansari N, Alomar MK. Predicting compressive strength of concrete containing industrial waste materials: novel and hybrid machine learning model. Adv Civ Eng 2022; 2022:5586737.

56. Joshi DA, Menon R, Jain RK, Kulkarni AV. Deep learning based concrete compressive strength prediction model with hybrid meta-heuristic approach. Expert Syst Appl 2023;233:120925.

57. Ly H-B, Nguyen T-A, Pham BT, Nguyen MH. A hybrid machine learning model to estimate self-compacting concrete compressive strength. Front Struct Civ Eng 2022;16:990–1002.

58. Tipu RK, Batra V. Development of a hybrid stacked machine learning model for predicting compressive strength of high-performance concrete. Asian J Civil Eng 2023;24:2985–3000. Suman

59. Mallikarjuna Rao G, Gunneswara Rao TD. Final setting time and compressive strength of fly ash and GGBS-Based geopolymer paste and mortar. Arabian J Sci Eng 2015;40:3067–74.

60. Gogineni A, Panday IK, Kumar P, Paswan Rk. Predictive modelling of concrete compressive strength incorporating GGBS and alkali using a machine-learning approach. Asian J Civil Eng 2024;25:699–709.

61. Philip S, Nidhi M, Ahmed HU. A comparative analysis of tree-based machine learning algorithms for predicting the mechanical properties of fibre-reinforced GGBS geopolymer concrete. Multiscale and Multidiscip Model Exp Des 2024;7:2555–83.

62. Philip S, Nidhi M. Performance comparison of artificial neural network and random forest models for predicting the compressive strength of fibre-reinforced GGBS-based geopolymer concrete composites. Mater Circular Economy 2024;6:34.

63. Shanmukh BSS, Sasirekha BSVB, Sampath VRS. Using artificial intelligence and machine learning model for prediction of uniaxial compressive strength of GGBS concrete. IOP Conf Ser Mater Sci Eng 2023;1273:012002.

64. Gupta P, Gupta N, Saxena KK, Goyal S. Multilayer perceptron modelling of geopolymer composite incorporating fly ash and GGBS for prediction of compressive strength. Adv Mater Process Technol 2022;8:1441–55.

65. Bilim C, Atiş CD, Tanyildizi H, Karahan O. Predicting the compressive strength of ground granulated blast furnace slag concrete using artificial neural network. Adv Eng Software 2009;40:334–40.

66. Yeh I-C. Concrete slump test; 2007. https://doi.org/10.24432/C5FG7D.

67. Maekawa OK. H, data base for mechanical properties of concrete; 2001. http://bme.t.u-tokyo.ac.jp/researches/detail/concreteDB/download.html.

68. Liu W, Liu G, Zhu X. Applicability of machine learning algorithms in predicting chloride diffusion in concrete: modeling, evaluation, and feature analysis. Case Stud Constr Mater 2024;21:e03573.

69. Ali Khan M, Zafar A, Akbar A, Javed MF, Mosavi A. Application of gene expression programming (GEP) for the prediction of compressive strength of geopolymer concrete. Materials 2021;14:1106.

70. Wang C, Xu S, Yang J. Adaboost algorithm in artificial intelligence for optimizing the IRI prediction accuracy of asphalt concrete pavement. Sensors 2021;21:5682.

71. Khan K, Ahmad A, Amin MN, Ahmad W, Nazar S, Arab AM. Comparative study of experimental and modeling of fly ash-based concrete. Materials 2022;15:3762.