Rocznik Ochrona Środowiska

Volume 28 Year 2026 ISSN 2720-7501 pp. 349-373

https://doi.org/10.54740/ros.2026.024 open access

Received: April 2026 Accepted: May 2026 Published: June 2026

Data-Driven Prediction of CO_2 Emissions in Low-Carbon Geopolymer Concrete Using Integrated Experimental Analysis and Machine Learning Techniques

Muhannad Riyadh Alasiri

Civil Engineering Department, College of Engineering, King Khalid University, Abha, Saudi Arabia

https://orcid.org/0009-0007-4464-3680

corresponding author's e-mail: muhannadriyadhalasiri@gmail.com

Abstract: The construction sector is a huge contributor to carbon emissions in the world, mostly because of the heavy applications of Ordinary Portland Cement (OPC). Geopolymer concrete with low carbon content has become a viable alternative to concrete; nevertheless, accurate prediction and evaluation of its CO_2 emissions have not been fully achieved, especially through combined experimental and data analysis. The literature on this study focuses mainly on mechanical performance, and there is limited understanding of environmental emission modeling. To measure and forecast CO_2 emissions from low-carbon geopolymer concrete, this paper develops an experimental and machine-learning system. An array of geopolymer blends comprising fly ash and ground granulated blast furnace slag (GGBS) was developed at different binder proportions, alkaline activator ratios, and curing conditions. The amount of CO_2 was determined using a cradle-to-gate evaluation method (kg CO_2/m^3). Data were subsequently trained on and tested on machine learning models, such as Artificial Neural Networks (ANN), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) using experimental data. The R^2, RMSE, and MAE measures were used to assess model performance. Findings show that the XGBoost model has the best prediction accuracy ($R^2 > 0.95$), indicating good generalization ability. The highest CO_2 emission reduction was 45–65 percent for geopolymer mixtures compared to traditional OPC concrete. The construct can offer a trustworthy decision support system in geopolymer concrete design that is environmentally friendly and further sustainability in construction methods with predictive modeling of emissions.

Keywords: geopolymer concrete, CO_2 emission, low-carbon materials, Life Cycle Assessment (LCA), machine learning, environmental impact

1. Introduction

1.1. Background

The rapid rate of infrastructure development and urbanization has also driven demand for cement-based construction materials. The main binder in traditional concrete, common Portland Cement (OPC), is extremely energy-intensive and pollutes the environment. The manufacture of a ton of OPC emits about 0.8 to 0.9 tons of CO_2, which is a combination of the calcification of limestone, as well as the burning of fossil fuels during the process of making the clinker (Cao et al., 2016; Kurtis, 2015; Lim et al., 2020). Cement production has been known to contribute almost 7–8 percent of the total anthropogenic carbon dioxide emissions globally, and therefore, cement production has been identified as one of the largest industrial contributors to climate change (Shen et al., 2021; Wei & Cen, 2019).

Alternative low-carbon binder systems have attracted significant research interest in response to rising environmental concerns and global decarbonization targets. Geopolymer concrete has been a promising sustainable replacement for OPC-based concrete (Daehn et al., 2022; Habert et al., 2020; Kumar et al., 2025). The production of geopolymers involves the alkali activation of industrial by-products, i.e., fly ash and ground granulated blast furnace slag, which contain aluminosilicates. This will eliminate clinker production and significantly reduce the carbon embodied in the product (Barbhuiya et al., 2025; Cao et al., 2016).

In addition to environmental advantages, geopolymer concrete has good mechanical strength, chemical resistance, and durability (Garces et al., 2022; Mohsin et al., 2025; Shobeiri et al., 2023). Nevertheless, the environmental sustainability of geopolymer systems is determined by a number of variables, such as the type of precursor, the composition of alkaline activators, the curing regime, transportation distances, and the energy used during production. Thus, to demonstrate the environmental benefits and to inform sustainable mix design, systematic mitigation and prediction of CO_2 emissions from geopolymer concrete are necessary (Sandanayake et al., 2020; Y. Zhang et al., 2025).

1.2. Research Gap

Although a considerable number of studies have been conducted on the mechanical and durability behavior of geopolymer concrete, most have focused on compressive strength, workability, and microstructural properties. There have been many successful applications of machine learning (ML) algorithms, such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forests (RF), and Extreme Gradient Boosting (XGBoost), to predict the strength and durability-related characteristics of geopolymer systems in recent years.

Nevertheless, relatively little research has been dedicated to the prediction modeling of environmental indicators, especially CO_2 emissions, using machine learning methods (Gao et al., 2023; Latawiec et al., 2018; Tam et al., 2022). The majority of environmental analysis is based on conventional Life Cycle Assessment (LCA) approaches that provide fixed emission estimates based on given mix proportions. The methods are not flexible and predictive where the mix parameters change (Nassef et al., 2023; B. Zhang et al., 2025).

In addition, the available research rarely combines experimental measurements of CO_2 emissions with sophisticated machine learning algorithms within a single framework (Hasan et al., 2025; Mardani et al., 2020; Nguyen et al., 2023). Such a lack of an integrated, experimental-data-based approach limits the development of smart tools to inform the design of environmentally optimized geopolymer concrete (Martiny, 2023; Tian et al., 2025).

There is therefore a critical need for a comprehensive framework that:

- Measurement of CO_2 emissions of geopolymer mixtures experimentally,
- Strong machine learning techniques with predictive modeling are applied,
- Features importance analysis identifies important parameters that affect the result, and
- Favors mix ecologically friendly design strategies.

1.3. Research Objectives

The main objective of the study is to develop a unified, data-driven, experimental model for the precise prediction of CO_2 emissions from low-carbon geopolymer concrete. To attain this objective, the following are the specific research objectives:

1. To experimentally determine CO_2 emissions of low-carbon geopolymer concrete mixtures according to a cradle-to-gate evaluation model (kg CO_2/m^3) with factoring in precursor materials, alkaline activators, aggregates, curing conditions, and energy utilization.
2. To create and test machine learning (ML) models that can predict CO_2 emissions based on the composition of the mixture and processing parameters, with accurate prediction. Statistical performance measures (R^2, RMSE, MAE) will be used to compare and evaluate multiple ML algorithms (Artificial Neural Networks (ANN), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)).
3. To find out the most influential parameters influencing the CO_2 emissions using feature importance analysis and sensitivity analysis, thus defining the relative contribution of the binder content, activator ratio, liquid to binder ratio, and other important variables.
4. To implement a predictive decision-support framework to empower the design of environmentally optimized geopolymer concrete mix depending on performance at emissions.

1.4. Novelty and Scientific Contribution

This study presents an original and holistic approach to predicting environmental emissions in geopolymer concrete by combining experimental quantification with modern machine learning modeling. The general research methodology assumed in this research is shown in Figure 1. It combines experimental research, life-cycle CO_2 emission analysis, and a machine-learning model to assess the environmental performance of low-carbon geopolymer concrete. The framework can be used to predict and optimize mix design parameters through data analysis, ensuring carbon emissions are minimized.

The key contributions of this study are summarized as follows:

- It also suggests one of the first unified, machine-learning-based models specifically aimed at predicting CO_2 emissions in geopolymer concrete systems.
- In contrast to traditional research, which generally focuses on mechanical properties, this study takes a more environmentally oriented, data-driven perspective, with CO_2 emissions as the main performance metric.
- It fills the research gap between a conventional life-cycle emission assessment and a predictive artificial intelligence model, enabling dynamic estimation of emissions for different mix compositions.

- The research provides a quantitative understanding of the relative environmental impacts of precursor materials and alkaline activators, enabling informed material choice and sustainable construction.
- The framework has been developed to support the development of smart, low-carbon material design in line with global climate mitigation.

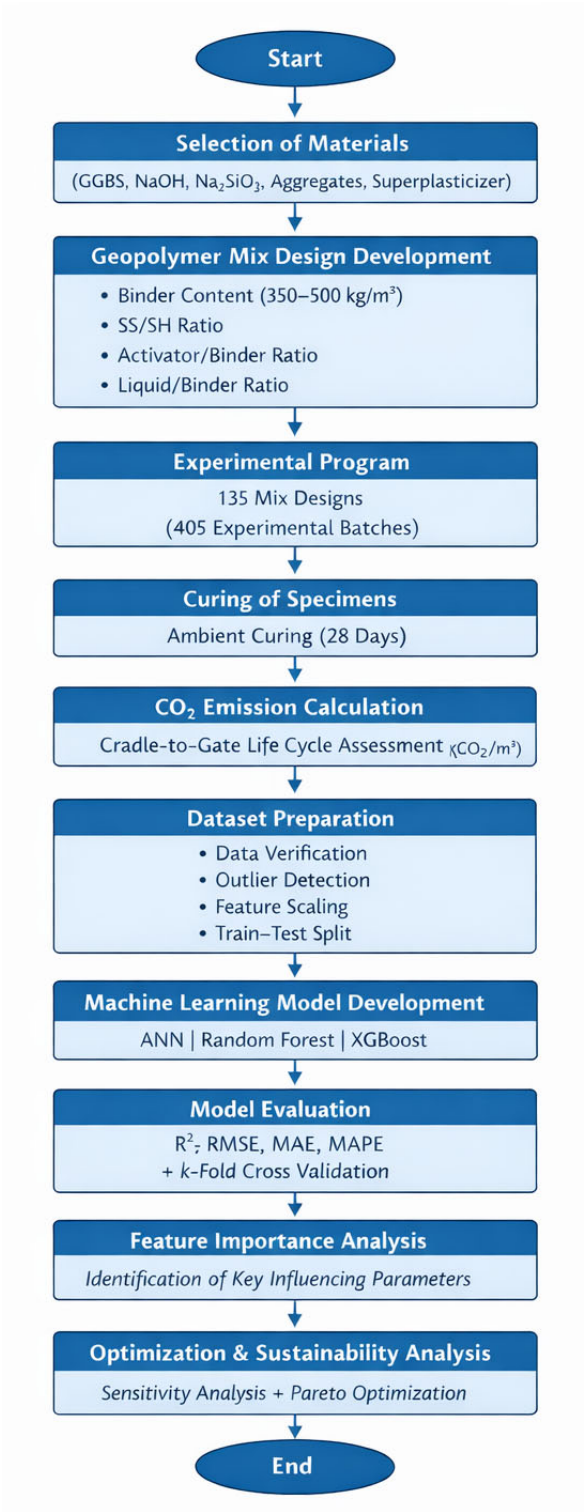

Fig. 1. Flowchart illustrating the integrated experimental and machine learning framework developed to quantify and predict CO_2 emissions of low-carbon geopolymer concrete based on mix design parameters and life-cycle emission assessment

2. Materials and Methods

2.1. Materials

2.1.1. Ground Granulated Blast Furnace Slag (GGBS)

The only binder applied in the manufacture of low-carbon geopolymer concrete was Ground Granulated Blast Furnace Slag (GGBS). The GGBS was sourced in a steel production plant and used as one of the environmentally friendly alternatives to Ordinary Portland Cement. As an industrial by-product, GGBS produces substantially lower embodied CO_2 emissions than clinker-based cement systems, enabling its use in low-carbon construction.

The choice of GGBS as the sole precursor allows for the obvious determination of GGBS's contribution to the total CO_2 emissions in the geopolymer system.

2.1.2. Alkaline Activators

A mixture of sodium hydroxide (NaOH) and sodium silicate (Na_2SiO_3) solutions started the geopolymerization process. The molarity required was achieved by dissolving sodium hydroxide pellets in distilled water. To increase the availability of silica, a sodium silicate solution was added to promote the development of the geopolymeric gel.

The alkali solutions were combined and allowed to stabilize before use. The ratios of sodium silicate to sodium hydroxide and of activator to binder were considered important experimental variables in determining CO_2 emissions.

2.1.3. Aggregates

Fine and coarse aggregates were natural river sand and crushed granite, respectively. The aggregates were chosen to meet the normal needs of concrete production and were mixed in a wet condition on the surface.

2.1.4. Chemical Admixture

A polycarboxylate-based superplasticizer that did not increase water content but enhanced uniform mixing was added to alkaline systems to improve workability.

2.2. Mix Design

The low-carbon geopolymer concrete mix design was developed to examine how the alkaline activator parameters and binder content affect CO_2 emissions. The GGBS served as the only binder, whereas the alkaline activator and liquid-to-binder ratios were controlled as the main experimental factors.

2.2.1. Binder Proportions

In all mixtures, only GGBS was the precursor material. The content of the binders was adjusted within a predetermined range (e.g., 350–500 kg/m^3) to investigate its impact on environmental emissions. Because CO_2 emissions directly depend on the amount of material, the sensitivity of emissions to binder dosage can be determined by varying the binder dosage.

To ensure the application is feasible, the aggregate-to-binder ratio was kept within the range of common structural concrete parameters.

2.2.2. Alkaline Activator Ratios

The alkaline activator was made of sodium hydroxide (NaOH) and sodium silicate ($Na_2 SiO_3$). Two significant ratios were looked at:

- Sodium silicate-sodium hydroxide percentage (SS/SH). The value of this ratio was adjusted (e.g., 1.5–3.0) to assess its effect on geopolymerization and CO_2 emissions, given that sodium silicate typically has higher embodied carbon than sodium hydroxide.

Activator-to-binder ratio (A/B)

The percent content of the alkaline liquid relative to the binder was also varied (i.e., 0.35–0.50) to test its effect on the degree of emissions.

These parameters were chosen because they have a significant impact on reaction efficiency and the environment.

2.2.3. Liquid-to-Binder Ratio

The liquid-to-binder (L/B) ratio was determined as the total alkaline solution (activators containing water) per unit mass of GGBS. The L/B ratio was kept within a controlled range to achieve sufficient workability and uniformity between mixtures.

Changes in L/B ratio have a direct impact on material consumption and, therefore, on calculated CO_2 emissions per cubic meter of concrete. Thus, it was considered an essential input parameter for the subsequent machine learning model.

Experimental Program

The 135 unique geopolymer concrete mix designs were created by a full-factorial combination of the chosen mix parameters, such as binder content (5 levels), sodium silicate-to-sodium hydroxide ratio (SS/SH, 3 levels), activator-to-binder ratio (A/B, 3 levels), and liquid-to-binder ratio (L/B, 3 levels), which gave 5 × 3 × 3 × 3 = 135 mixtures. To establish the repeatability of the experiment and determine batch variability, all mixes were prepared in 3 replicates (135 × 3 = 405) to develop the machine-learning data.

Curing Condition

Every geopolymer concrete sample was demolded about 24 hours after casting and then allowed to cure in the ambient laboratory environment to emulate real-world field conditions and avoid any additional energy requirements associated with heat curing. A temperature of 25 ± 2°C was used to cure the food with stable humidity conditions in the laboratory during the experimental time. External thermal curing was avoided, so the calculated CO_2 emissions included contributions from materials but not from additional energy consumption for curing. All the specimens were left to cure for 28 days before data collection to ensure consistency in the experimental data.

CO_2 Emission Calculation Method

A cradle-to-gate assessment method was used to determine the CO_2 emissions of each geopolymer concrete blend (kg CO_2/m³). The system boundary was set to the production of raw materials, the processing of the materials, and the transportation of the materials to the batching site, but not to their use or end-of-life, to ensure consistency and comparability.

The total CO_2 emission for each mix was calculated using the material inventory method (Eq. 1):

$$CO_{2,total} = \sum_{i=1}^{n}(M_i \times EF_i) + \sum_{j=1}^{m}\left(D_j \times T_j \times EF_{trans}\right) \quad (1)$$

where:
M_i – mass of material component i (kg/m³),
EF_i – emission factor of material component i (kg CO_2/kg),
n – total number of material components.

The evaluation considered the emissions from the following: GGBS, sodium hydroxide (NaOH), sodium silicate (Na_2SiO_3), aggregates, chemical admixtures, and, where appropriate, transportation emissions. To ensure reliability and consistency of the results, emission factors were obtained from existing environmental databases and peer-reviewed literature.

CO_2 emissions were calculated based on the precise amounts of materials used in the batching exercise for each mix. The computed values were the sole output parameter used to develop and validate the machine learning model.

2.3. CO_2 Emission Assessment

Emission Factors (kg CO_2/kg material)

The CO_2 emissions of each geopolymer concrete mixture were calculated using material-specific emission factors expressed in kg CO_2/kg material. These emission facets indicate the embodied carbon emitted during extraction, processing of raw materials, and manufacturing within the specified cradle-to-gate system boundary.

Emission factors were assigned to each constituent material, including:

- Ground Granulated Blast Furnace Slag (GGBS),
- Sodium hydroxide (NaOH),
- Sodium silicate (Na_2SiO_3),
- Fine and coarse aggregates,
- Chemical admixtures.

The emission factors were retrieved from existing environmental databases and peer-reviewed literature to ensure consistency and reliability. The embodied CO_2 contribution of every material was determined by obtaining the product of its mass (kg/m³) and the corresponding emission factor (kg CO_2/kg), and the overall contribution was obtained by adding all the contributions.

Functional Unit

The functional unit adopted in this study was defined as: 1 m^3 of geopolymer concrete.

Accordingly, total emissions were expressed as: kg CO_2/m^3

This is a practical unit that allows uniform comparison of various geopolymer mix designs and, at the same time, conforms to the standard course of action in life-cycle assessment (LCA) research on construction materials. All emissions were calculated in normalized values per cubic meter of concrete to make them comparable and suitable for developing machine learning models. Table 1 shows the Emission Factors for Materials (kg CO_2/kg).

Table 1. Emission Factors for Materials (kg CO_2/kg)

Material	Emission Factor (kg CO_2/kg)	Source / Database	Notes
Ground Granulated Blast Furnace Slag (GGBS)	0.065	EN 15804:2012+A2:2019 (Environmental Product Declaration)	Industrial by-product; 30% avoided allocation credit vs cement
Sodium Hydroxide (NaOH)	1.50	Ecoinvent 3.6 Database	Chemical production; includes manufacturing & transportation (150 km avg)
Sodium Silicate (Na_2SiO_3)	2.10	Ecoinvent 3.6 & GaBi Database	Highest embodied carbon among activators; energy-intensive production
Fine Aggregate (River Sand)	0.004	EN 15804:2012+A2:2019	Minimal processing; local sourcing (≤50 km transportation)
Coarse Aggregate (Crushed Granite)	0.003	EN 15804:2012+A2:2019	Quarrying and mechanical crushing; minimal further processing
Superplasticizer (Polycarboxylate-based)	1.80	Ecoinvent 3.6 Database	Chemical admixture; based on typical dosage (0.5–1.5% by weight)
Water	0.0	Standard assumption	Included in activator solutions; negligible direct emissions

2.4. Machine Learning Methodology

2.4.1. Input and Output Parameters

The machine learning (ML) model was created to forecast the environmental performance of low-carbon geopolymer concrete using mix design variables as input and CO_2 emissions as the output.

Input Parameters

The selected input variables represent the primary mix design parameters influencing embodied carbon. These include:

- Binder content (kg/m³),
- Sodium silicate-to-sodium hydroxide ratio (SS/SH),
- Activator-to-binder ratio (A/B),
- Liquid-to-binder ratio (L/B),
- Sodium hydroxide content (kg/m³),
- Sodium silicate content (kg/m³).

Since the system is cement-free, GGBS serves as the sole binder. These variables were chosen because they directly affect material consumption and, consequently, CO_2 emissions.

Output Parameter

The single output variable for the ML models was: kg CO_2/m^3, representing the cradle-to-gate CO_2 emission per cubic meter of geopolymer concrete. This value was calculated using the material inventory-based emission assessment described earlier.

2.4.2. Data Preprocessing

Before the development of machine learning models, the entire experimental data set (n = 405) underwent systematic preprocessing to ensure data reliability and consistency, and to improve predictive modelling suitability. All records in batching were initially checked to verify the accuracy of material quantities and calculated CO_2 emissions. Missing and erroneous entries were identified and corrected. Statistical screening was performed to detect outliers, including IQR analysis and Z-score evaluation, which identified abnormal data points that might negatively affect model training. No extreme values were removed to the point of preventing genuine experimental variation.

Replicate considerations were needed because the 135 unique mix designs were run in three replicates. The duplicate values were kept as independent data points to represent the variability of natural experiments and to increase the size of the datasets for machine learning training. Nevertheless, the consistency of replicate statistics was confirmed to allow replication. This helped in enhancing model robustness by subjecting algorithms to realistic experimental dispersion.

Since the input variables (e.g., binder content in kg/m ^{3} and mix ratios such as SS/SH, A/B, and L/B) are represented by different numerical values, feature scaling was applied before training. The method used was normalization or standardization to bring similar contributions to the features and avoid giving weight to variables with higher magnitudes. Lastly, the data was randomly mixed and divided into training and testing data sets to facilitate the unbiased model validation and performance assessment.

2.4.2.a. Hyperparameter Tuning and Model Configuration

Data Splitting & Cross-Validation: • Train-test split: 80–20% (324 training, 81 testing samples) • k-Fold cross-validation: k = 5 folds • Random seed: 42 (for reproducibility) ANN Configuration: • 2 hidden layers (64 and 32 neurons) • Activation: ReLU (hidden), Linear (output) • Optimizer: Adam, learning rate = 0.001 • Regularization: L2 = 0.0001, Dropout = 0.2 • Loss: Mean Squared Error Random Forest Configuration: • Number of trees: 200 • Max depth: 15 • Max features: √6 (square root) • Criterion: Mean Squared Error. XGBoost Configuration: • Boosting rounds: 500 • Learning rate (eta): 0.05 • Max depth: 6 • Subsample: 0.8 • Colsample_bytree: 0.8 • L2 Regularization (Lambda): 1.0. All hyperparameters were optimized using grid search with 5-fold cross-validation. Software versions (Python 3.9+, scikit-learn ≥1.0, XGBoost ≥1.7, TensorFlow ≥2.12) and complete reproducibility details are documented. These specifications enable other researchers to reproduce the ML analysis fully.

2.4.3. Model Development

Three machine learning algorithms, which are Artificial Neural Network (ANN), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), were used to predict the rutting parameter CO_2 emission of Low-Carbon Geopolymer Concrete. These types of models were chosen because of their high level of managing non-linear relationships and complicated interactions of the input variables.

An Artificial Neural Network (ANN) model was used to represent non-litter dependencies among input features and the target variable through interconnecting hidden layers (Asteris & Mokos, 2020; Duan et al., 2013; Lee, 2003; Lin & Wu, 2021). The ensemble learning approach, which was applied in this study in the form of random Forests (RF), was resistant to overfitting and capable of handling multivariate data (Khodaparasti et al., 2023; Mai et al., 2021; Sun et al., 2019). The XGBoost algorithm is a gradient boosting algorithm, a powerful predictive model that offers regularization and supports structured regression (Alabdullah et al., 2022; Duan et al., 2021; Mustapha et al., 2024; Sun, Li, Yang, et al., 2024).

Statistical measures, such as the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), were used to assess the performance of the developed models. As a comparative analysis, the most appropriate algorithm in predicting binder performance was identified.

2.4.4. Model Training and Evaluation Metrics

The machine learning models were tested using a combination of statistical measures to ensure a thorough assessment of predictive power, strength, and generalization. These measures are the concurrence between experimentally obtained and model-projected CO_2 emissions and are commonly used in regression-based concrete modeling research. To ensure the accuracy and robustness of the machine learning models were

fully assessed, several error- and variance-based statistical indicators were used to evaluate their predictive performance. The coefficient of determination (R^2) is used to measure the percentage of variability in the experimental CO_2 emissions of Low-Carbon Geopolymer Concrete described by the model, which represents the strength of the correlation between the predicted and observed values. Root Mean Square error (RMSE) measures the average size of the prediction error and weights large deviations more, making it sensitive to large mispredictions. Mean Absolute Error (MAE) is a measure of the average difference between predicted and experimental values, expressed in absolute terms, and is a more rigorous measure of average error since all deviations are equalized. The Mean Absolute Percentage Error (MAPE) represents the error in prediction as a percentage, allowing the relative comparison of model performance across datasets. Also, model generalization was evaluated using k-fold cross-validation scores that assess predictive stability across many data partitions, minimizing bias induced by a single train-test split.

Coefficient of Determination (R^2)

The coefficient of determination (R^2) is a measure of the percentage of variance in the observed Measured CO_2 emission that the model predictions would explain. The larger the value of R^2, the better the predictive performance and the stronger the correlation between predicted results and experimental outcomes (Eq. 2).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

y_i : Measured CO_2 emission, $\hat{y}_i$: Predicted CO_2 emission, n is the number of samples.

Root Mean Square Error (RMSE)

RMSE is the square root of the mean squared error in the CO_2 emissions values of prediction and measurements. It places greater weight on large prediction errors and provides information on the overall model accuracy (Eq. 3).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

Mean Absolute Error (MAE)

MAE computes the average absolute difference between the estimated and experimental CO_2 emissions. Unlike RMSE, it is not sensitive to outliers, since all errors are treated equally (Eq. 4).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

Mean Absolute Percentage Error (MAPE)

MAPE is an expression of the prediction error as a percentage of the known CO_2 emissions and makes it easy to compare the relative prediction accuracy of models (Eq. 5).

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{5}$$

k-Fold Cross-Validation Score

K-fold cross-validation was used to evaluate the model's generalization and to eliminate bias introduced by a single train-test split (Abellán-García, 2021; Al-Abdaly et al., 2021; Ling et al., 2019; Lyu et al., 2022). Each dataset was subdivided into k equal subsets, and the model was trained on k -1 folds and tested on the remaining fold. This was repeated k times, and the mean performance measure was reported as the cross-validation score (Eq. 6).

$$CV_{score} = \frac{1}{k}\sum_{j=1}^{k} Score_j \tag{6}$$

Under the cross-validation expression in k-fold, CV_{score} is the aggregate cross-validation results of the machine learning model. The parameter k is the number of folds used to divide the dataset. In the validation process, the dataset is subdivided into k subsets that are approximately equal, and, in every iteration, k-1 subsets are utilized to train the model, and the rest remains to be used as a validation set.

The $Score_j$ is defined as the performance measure (the R^2, RMSE, or MAE) of the validation fold number j. The folds provide individual estimates of the models' performance using a separate training-validation split. The performance values are summed up $\sum_{j=1}^{k} Score_j$ and then divided by k to get an average of the performance values of all folds. Such an average score is an effective estimate of the model's generalization ability, as it mitigates the effects of random data partitioning and reduces overfitting.

3. Experimental CO_2 Emission Results

The cradle-to-gate CO_2 emissions of the 135 geopolymer concrete mixtures were measured and quantified in kg CO_2/m^3. The findings revealed a significant difference in the mixes across varying combinations, resulting from variations in binder content and alkaline activator dosage.

The total CO_2 emissions ranged from approximately:

$$180 \text{ to } 420 \text{ kg } CO_2/m^3$$

Lower emission values (≈180–220 kg CO_2/m^3) were observed in mixtures with:

- Reduced binder content
- Lower sodium silicate dosage
- Moderate activator-to-binder ratios

Higher emission values (≈380–420 kg CO_2/m^3) were associated with:

- Increased binder content
- Higher sodium silicate-to-sodium hydroxide ratios
- Elevated activator content

The variation between the lowest and highest emission mixes was approximately 57%.

Emission Comparison Between Mixes

Table 2 presents representative geopolymer concrete mixtures from the total of 135 mix designs to illustrate the variation in cradle-to-gate CO_2 emissions. The comparison also clearly shows that the level of emissions is highly dependent on the binder content and the alkaline activator.

The minimum value of CO_2 emission (around 182 kg CO_2/m^3) was found in mixtures that had low binder content (350 kg/m³) and low sodium silicate-sodium hydroxide (SS/SH) ratios. This decrease is mainly explained by reduced material and sodium silicate use, which typically have higher embodied carbon than the other constituents.

The binder content steadily increased with increasing CO_2 emissions, ranging from 350 kg/m³ to 500 kg/m³. This is a natural trend, as embodied carbon is directly proportional to the mass of materials per cubic meter. Also, emissions increased with the further addition of the activator to the binder (A/B) ratio as the dosage of the alkaline solutions increased.

One of the most notable effects is the SS/SH ratio. Blends with more SS/SH values (e.g., 3.0) had significantly higher emissions than those with lower ratios. It is largely because the production of sodium silicate is highly energy-intensive and contributes significantly to the overall emission inventory.

In general, the range of emission (about 182–420 kg CO_2/m^3) is seen to vary by approximately 57 percent between the lowest and highest mixes of emissions. This significant dispersion validates the fact that mix proportioning is a crucial component of performance in this environment, despite the absence of cement in the geopolymer.

The findings indicate the need for predictive modeling tools to control mix design parameters to achieve the lowest CO_2 emissions while preserving functional characteristics.

Table 2. Comparison of CO_2 Emissions for Selected Geopolymer Mixes

Mix ID	Binder (kg/m³)	SS/SH	A/B	L/B	CO_2 Emission (kg CO_2/m^3)
M01 (Low binder)	350	1.5	0.35	0.30	182
M12	350	2.0	0.40	0.35	215
M28	400	2.0	0.40	0.35	248
M45	400	2.5	0.45	0.40	298
M67	450	2.5	0.45	0.40	335
M89	450	3.0	0.50	0.45	372
M110	500	2.5	0.50	0.45	398
M135 (High binder)	500	3.0	0.55	0.45	420

3.1. Box Plot for CO_2 Emissions Across Binder Contents

The box plot provides the statistical description of CO_2 emissions of geopolymer concrete mixtures with various contents of binder (350, 400, 450, 500, and 550 kg/m³) (Figure 2). The boxes each depict the interquartile range (IQR), which comprises the 50% range of emission values, with the horizontal line within

each box depicting the median of the emission values. The whiskers extend to the minimum and maximum values in the data set, and any extreme values are represented as outliers.

Based on the plot, one can observe an evident incremental pattern in CO_2 emissions with increasing binder content. The blends with 350 kg/m³ binder have the smallest median, and the distribution is the narrowest, meaning the carbon footprint is more consistent and smaller. With 400 and 450 kg/m³ of binder content, the median value of the emissions shifts to the right, and the dispersion of the values changes slightly, reflecting greater variability due to varying activator ratios and liquid-to-binder proportions.

The box plots indicate higher median emissions and wider value ranges at the 500 and 550 kg/m³ levels of binder. This shows that increasing the binder dosage has a significant positive impact on embodied carbon, while also increasing variability in emissions with respect to the mix design parameters. Higher emissions are observed closer to the upper whiskers, implying that certain binder-to-activator ratios result in a relatively high carbon footprint.

In general, the box plot clearly shows that the binder content is the most significant factor in determining CO_2 emissions, and that emissions are gradually rising between 350 and 550 kg/m^3. This statistical graph aligns with the results of the correlation analysis and machine learning feature importance analysis, demonstrating that minimizing the binder dosage is one of the most effective ways to reduce the environmental impact of geopolymer concrete.

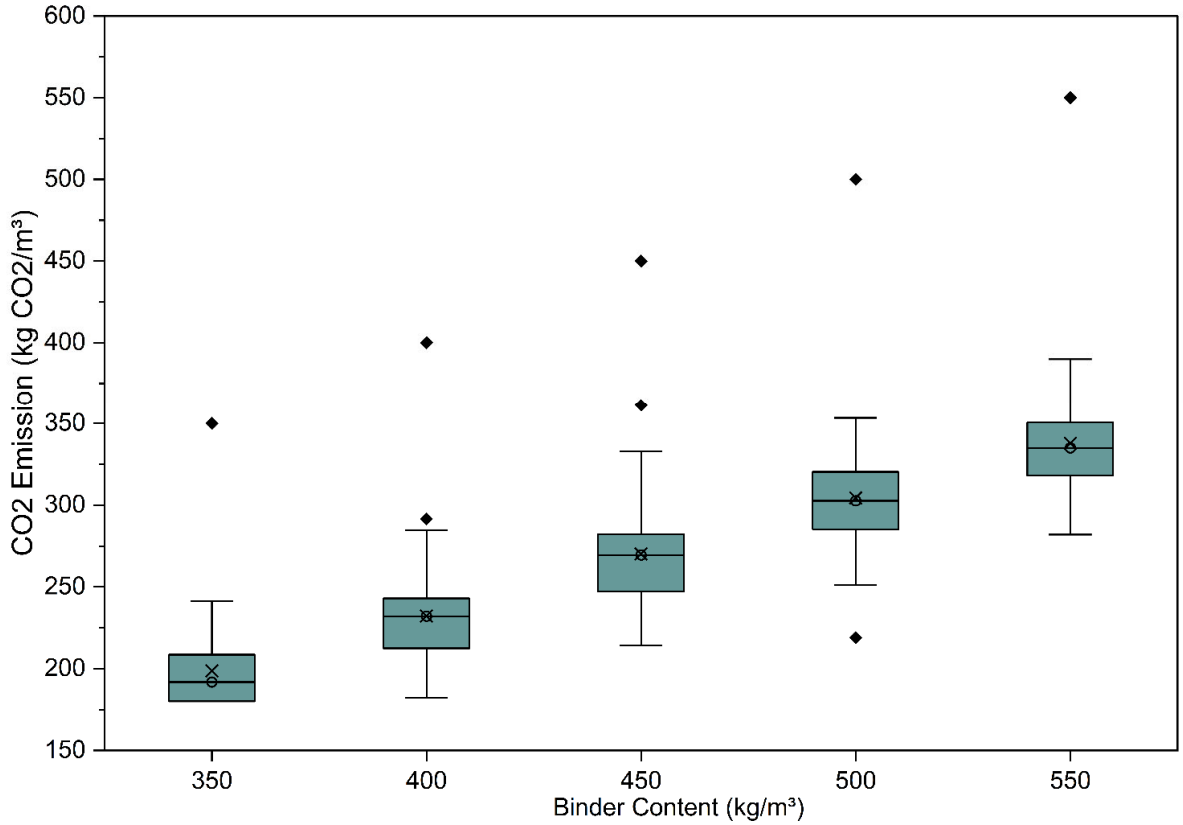

Fig. 2. Box Plot Showing Variation of CO_2 Emissions Across Different Binder Contents

3.2. Correlation Analysis Between Input Parameters and CO_2 Emissions

A correlation analysis was done to identify the relationship between mix design variables and CO_2 emissions using the Pearson correlation coefficient to understand the relationship between them (Smith & Sam, 2020; Song et al., 2024; Zhu et al., 2021) (Figure 3). The results are presented as a heat map that visually displays the strength and direction of the linear association between the input parameters (binder content, SS/SH ratio, activator-to-binder ratio, and liquid-to-binder ratio) and the output variable (CO_2 emission).

The heat map provides a graphical display of parameter interactions and shows the importance of each input variable for embodied carbon. The positive correlation values indicate that the higher the level of a given parameter, the greater the CO_2 emissions, and the stronger the correlation, the weaker the impact. The analysis can be considered an initial statistical evaluation before machine learning modeling, and proves the prevalence of influential variables in environmental performance.

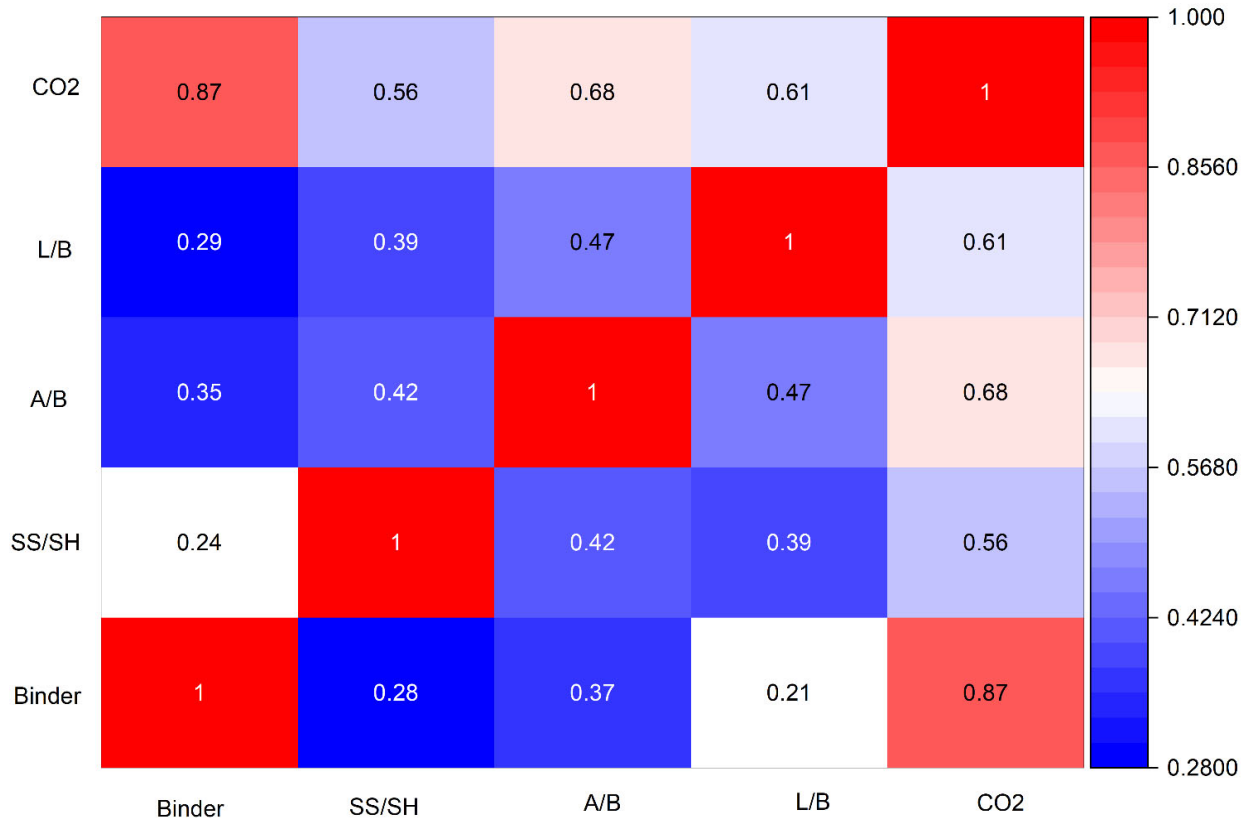

Fig. 3. Correlation heat map illustrating the relationship between input mix design parameters (binder content, SS/SH ratio, activator-to-binder ratio, and liquid-to-binder ratio) and CO_2 emissions

The analysis of Pearson correlation demonstrated that strong correlations exist between the input parameters and the CO_2 emissions. The greatest positive correlation was observed between binder content and CO_2 emissions ($r = 0.87$) which implies that an increase in the dosage of binder correlates with a large increase in embodied carbon as the material consumed is more. There was also a positive correlation with the activator to binder ratio (A/B) of moderate-strong positive correlation ($r = 0.68$) indicating the effect of increased use of alkaline activators on total emissions. Likewise, the correlation between the liquid-binder ratio (L/B) and the liquid content was also moderate and positive ($r = 0.61$) indicating that increased liquid content leads to increased emissions due to related activator elements. Sodium silicate/ sodium hydroxide (SS/SH) ratio was similarly positively correlated with CO_2 emission ($r = 0.56$), a phenomenon that can be explained by the fact that sodium silicate embodied carbon rate was relatively high when it was used to activate geopolymers. All in all, the correlation findings validate the fact that binder content is the strongest factor that determines CO_2 emission and then the activator related parameters, which confirms the results of the feature importance analysis and machine learning modeling.

4. Machine Learning Model Performance

4.1. Accuracy Comparison of Models

To evaluate predictive performance, the developed machine learning models were assessed using four statistical indicators:

- Coefficient of determination (R^2),
- Root Mean Square Error (RMSE),
- Mean Absolute Error (MAE),
- Mean Absolute Percentage Error (MAPE).

4.2. Observed vs. Predicted CO_2 Emissions

The training and testing dataset predictive agreement of the XGBoost model is indicated by the observed and predicted plots (Figure 4). The patterns of data points in the training set are well clustered around the 45° reference line, indicating perfect model fitting. The predictions within the test set are close to the experi-

mental values as well, which shows that it has strong generalization and low overfitting. The minor change in the dispersion of the testing data is normal and indicates realistic prediction uncertainty.

The experimental and modeled CO_2 emissions are compared with the predictive versus observed plots of the Random Forest (RF) model, shown for both training and testing data (Figure 5). The mix parameters and emissions non-linear relationships are well learned, and the data points are well clustered around the 45° reference line in the training dataset, indicating that the model can be well fitted and can learn non-linear relationships between mix parameters and emissions. In the test data, the dispersion of data points is slightly greater than in the training set. This is a natural phenomenon that demonstrates realistic generalization performance when the model is exposed to unseen data. Despite the small outbreak, most predictions are near the ideal agreement line, indicating the effectiveness of the Random Forest algorithm. The quality of the testing performance indicates that the RF model can capture the complex interactions among binder content, ratios of activators, and liquid-to-binder ratios, but the predictive stability is still good. But compared to the XGBoost model, the spread in the RF test plot is slightly greater, indicating slightly reduced predictive power. In general, the Random Forest model can be considered a reliable source of CO_2 emissions predictions and a good benchmark for comparing results with other ensemble-based learning algorithms.

The plots of the observed and predicted values of the Artificial Neural Network (ANN) model show that the neural network architecture developed can predict both the training and testing datasets. The CO_2 emission values predicted in the training dataset tend to follow the 45° reference line, indicating that the ANN model identified non-linear relationships among mix design variables and embodied carbon emissions (Figure 6).

A somewhat larger scatter in the data about the optimal prediction line, however, is observed compared to ensemble-based models. This indicates the average prediction error, given the sensitivity of ANN models to hyperparameter choice, network structure, and training behavior.

The dispersion is slightly higher in the test dataset, which indicates generalization performance when applied to unknown data. Although this variability is present, the ANN model can still exhibit reasonable predictive ability, indicating that it is adequate for predicting environmental emissions.

4.3. Comparative Performance Analysis of ANN, Random Forest, and XGBoost Models

An analysis of the three machine learning models, namely, Artificial Neural Network (ANN), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), was performed to identify the most appropriate algorithm to predict the CO_2 emission of low-carbon geopolymer concrete. Statistical measures such as R^2, RMSE, MAE, and MAPE were used to evaluate model performance on both the training and test sets.

The XGBoost model was found to have the best predictive performance among the examined models, as it had the highest coefficient of determination and the lowest error measures. In both the observed and predicted plots, the data values fell within the 45° reference line, strongly indicating the ability to fit and high generalization. The low MAPE value also indicates high predictive reliability.

The second model was the Random Forest, which demonstrated strong non-linear modelling and predictive performance. Although its results were slightly worse than XGBoost, the RF model demonstrated low dispersion on both the training and test data, which proves its strength and ability to prevent overfitting.

The Artificial Neural Network (ANN) model was the third model. Though the ANN identified non-linear relationships in the dataset, the dispersion was relatively high in the test set. This means there are slightly more prediction errors and generalization errors than in the case of ensemble-based models.

The general performance of the models is thus:

XGBoost>Random Forest>ANN

The high performance of XGBoost is explained by the gradient boosting mechanism, which continuously reduces prediction error and efficiently captures complex interactions among binder content, activator ratios, and liquid-to-binder ratios. The findings verify that ensemble learning is particularly appropriate for predicting environmental emissions in geopolymer concrete systems.

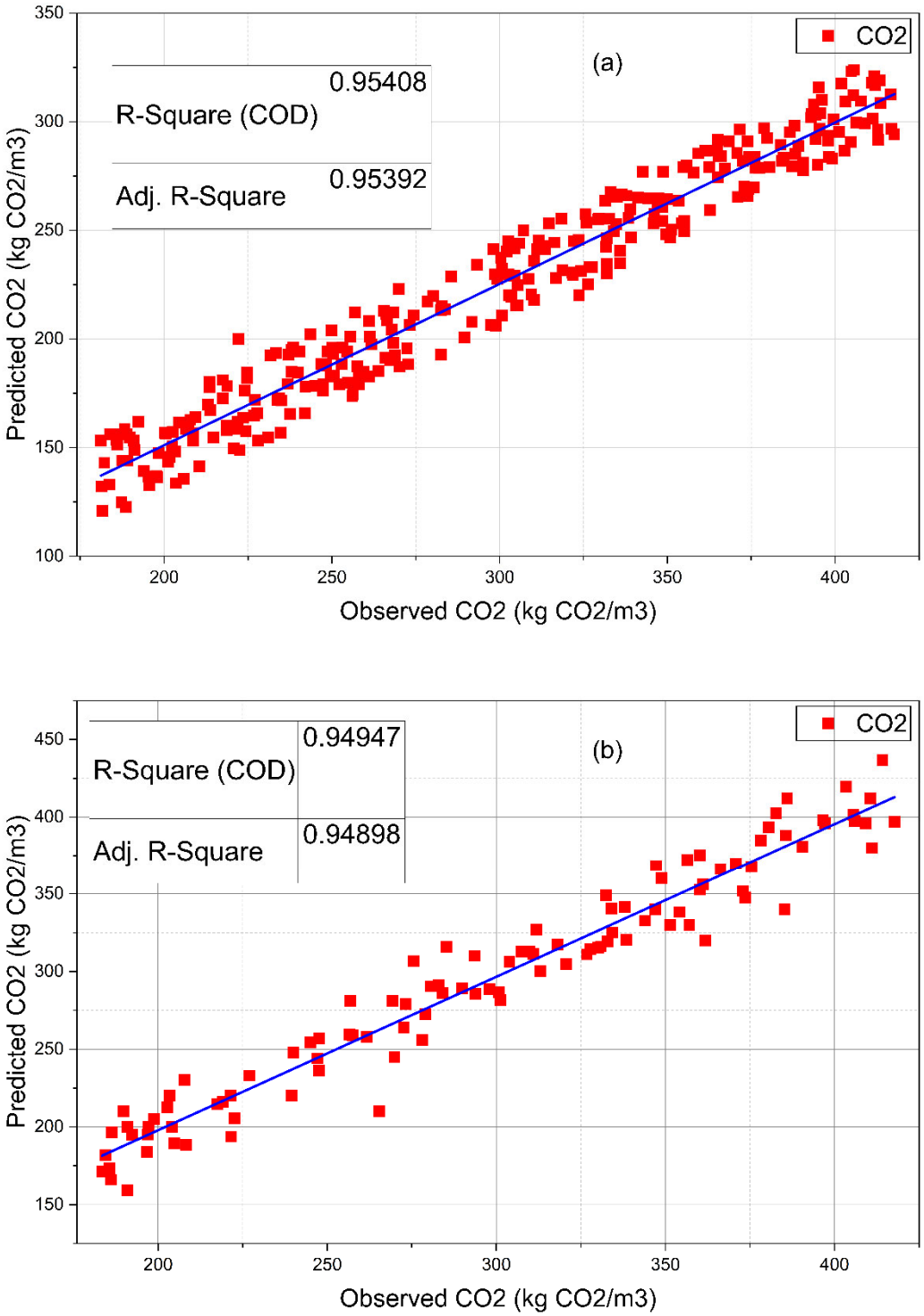

Fig. 4. Observed versus predicted CO_2 emissions for the XGBoost model: a) training dataset, b) testing dataset

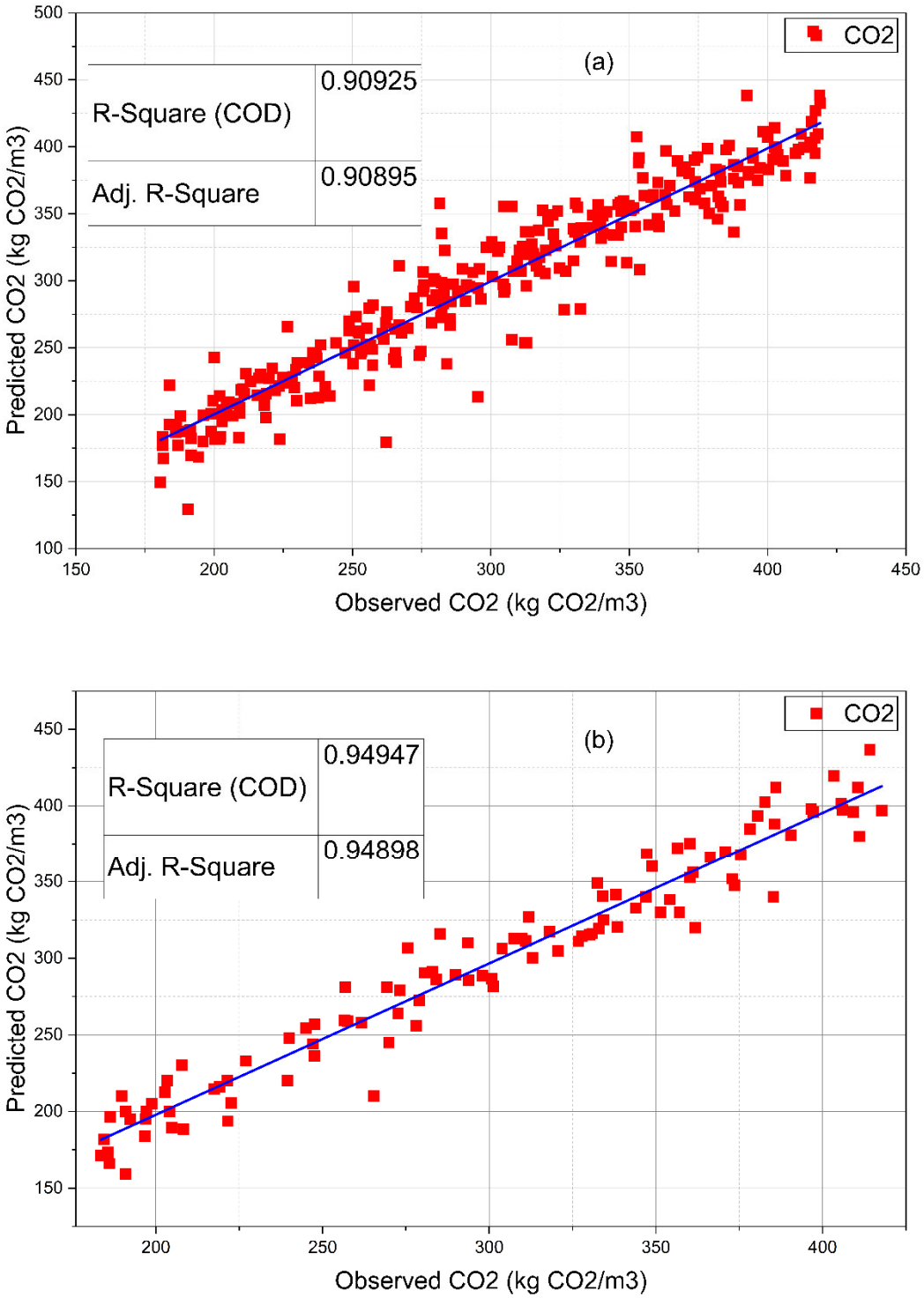

Fig. 5. Observed versus predicted CO_2 emissions for the Random Forest model: a) training dataset, b) testing dataset

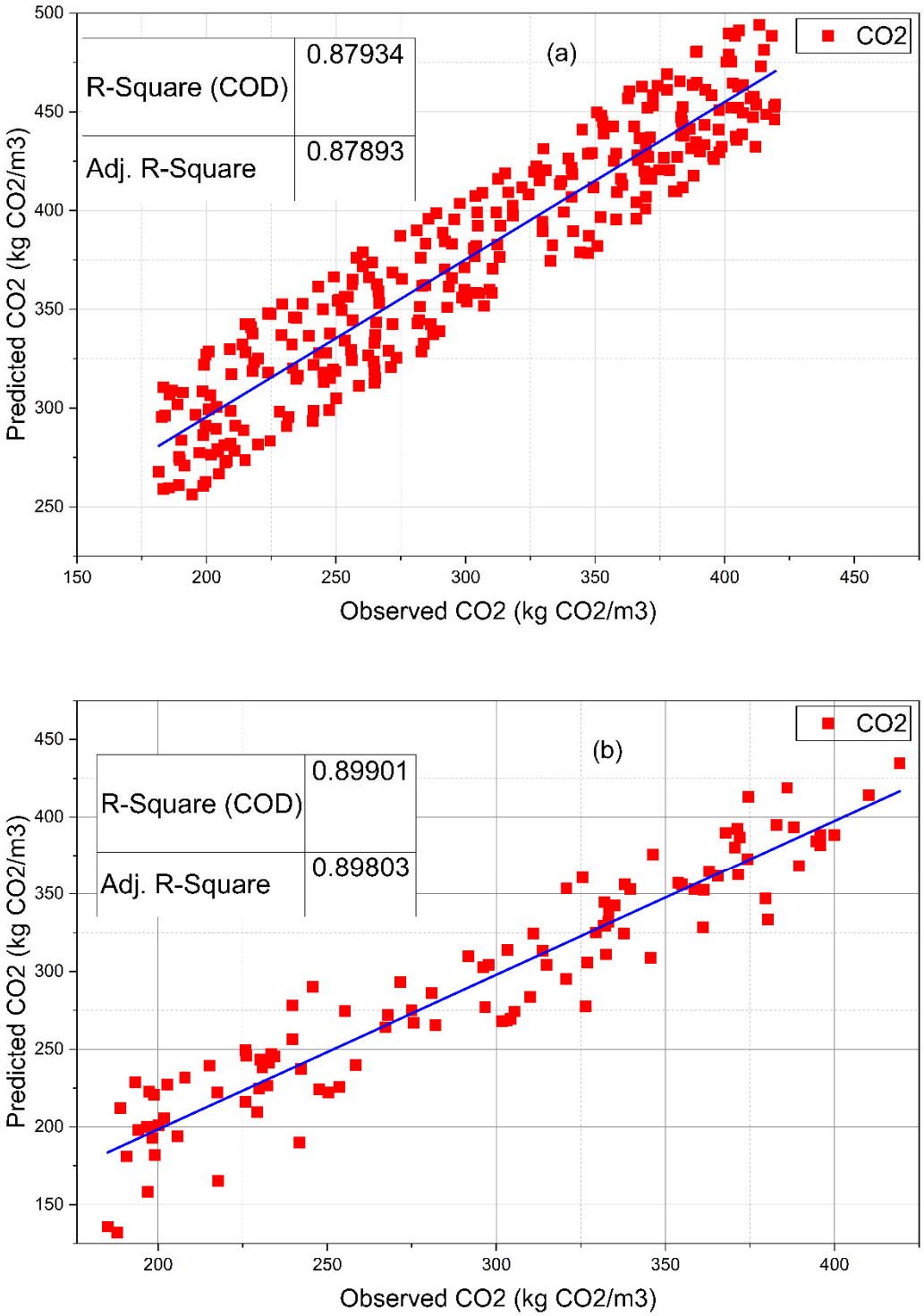

Fig. 6. Observed versus predicted CO_2 emissions for the Random Forest model: a) training dataset, b) testing dataset

4.4. Error Distribution Analysis of Machine Learning Models

Besides observed-versus-predicted plots, the error distribution was analyzed to assess the reliability and stability of the predictions from the developed machine learning models (Nafees et al., 2021; Sun, Li, Li, et al., 2024). The error histogram is a statistical representation of the residuals, which are the difference between the observed and the predicted values of CO_2 emissions. This analysis helps determine whether the prediction errors are randomly distributed, normally distributed, or skewed toward overestimation or underestimation. The error histogram depicts the frequencies of residual errors of the machine learning models assessed. An excellent model should have a symmetric distribution around zero, indicating the least systematic bias in predictions.

XGBoost model error histograms provide insights into the distribution and values of the residual errors, i.e., the difference between the observed and predicted CO_2 emissions (Observed–Predicted). A perfect prediction model must yield symmetrically distributed residuals around 0.0, indicating no systematic overestimation or underestimation. The residual errors in the training dataset are concentrated around zero, with the majority of values within a narrow range (Figure 7a). This implies a good fitting performance and good learning of the inherent nonlinearities between mix design parameters and CO_2 emissions. The histogram is symmetric, indicating that the model is not characterized by visible bias. In the testing set, the distribution of the residuals was concentrated around zero, though it was slightly wider than that of the training variables. This is likely when the model is tested on unknown data and indicates the realism of the variation in predictions. Notably, the skewness mean is not large, and there are no extreme skewness values, indicating that the XGBoost model does not deteriorate generalization performance (Figure 7b).

On the whole, the small error distribution and clustering near zero confirm the high predictive power of the XGBoost model, and the second fact supports the high ranking of this model among random Forest and ANN models.

Besides observed vs. predicted plots, the error distribution was also analyzed to assess the precision and consistency of the predictions from the developed machine learning models. The error histogram is a statistical measure of the residuals, which are the difference between the observed and predicted values of CO_2 emissions. This analysis helps determine the distribution of prediction errors: whether they are randomly distributed, normally distributed, or biased toward overestimation or underestimation. The error histogram shows the frequency distribution of errors for the evaluated machine learning models. A highly functioning model is symmetrically distributed about zero, which means there would be very little systematic bias in predictions.

The error histograms of the Random Forest model offer an insight into the distribution and the size of residual errors, which are errors observed and predicted CO_2 emissions (Observed–Predicted). The best prediction model would have symmetrically distributed residuals around zero, indicating that it is neither overestimated nor underestimated systematically. The residual errors in the training dataset are also very concentrated around zero, with most values within a small range (Figure 8a). This means that there has been high fitting performance and good learning of the underlying non-linear relationships between mix design parameters and CO_2 emissions. The histogram is symmetrical, indicating that there is no conspicuous bias in the model. In the testing dataset, the residual distribution is still centered around zero, but it has a slightly wider spread than in the training dataset. This is likely to occur when the model is tested on unknown data and reflects real prediction uncertainty. Notably, the lack of strong skewness or extreme outliers indicates that the Random Forest model does not exhibit significant changes in generalization (Figure 8b).

ANN residuals. The overall average of ANN residuals on the training data is zero, indicating that the model has learned the underlying non-linear dependence between the contents of the binder and the activators/CO_2 emission ratio. However, the distribution of residuals is slightly wider than with ensemble-based techniques, suggesting that the variability in predictions is moderate (Figure 9a).

The variance of the residuals in the situation involving the testing data is slightly increased, which is realistic as a generalization of behavior. The larger error spread suggests that the error in predictions is larger than in XGBoost and Random Forest models. The residuals, however, are primarily concentrated around zero, which implies that neither high overestimation nor underestimation bias is present (Figure 9b).

Overall, even though the ANN model demonstrates decent predictive ability, its residual distribution is larger than that of XGBoost and Random Forest, because it is fairly accurate at estimating CO_2 emissions.

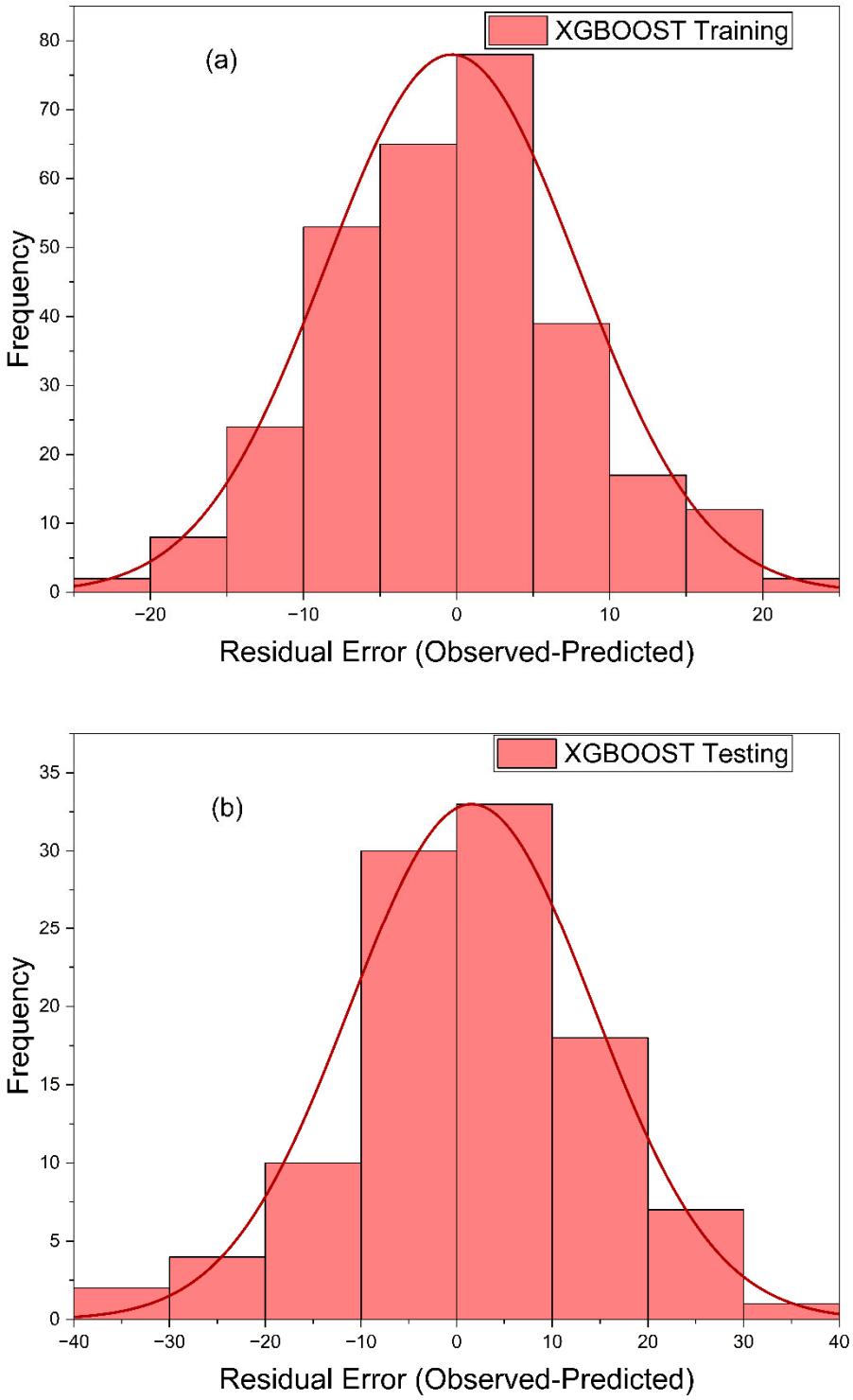

Fig. 7. Error histogram of residuals for the XGboost model: a) training dataset, b) testing dataset

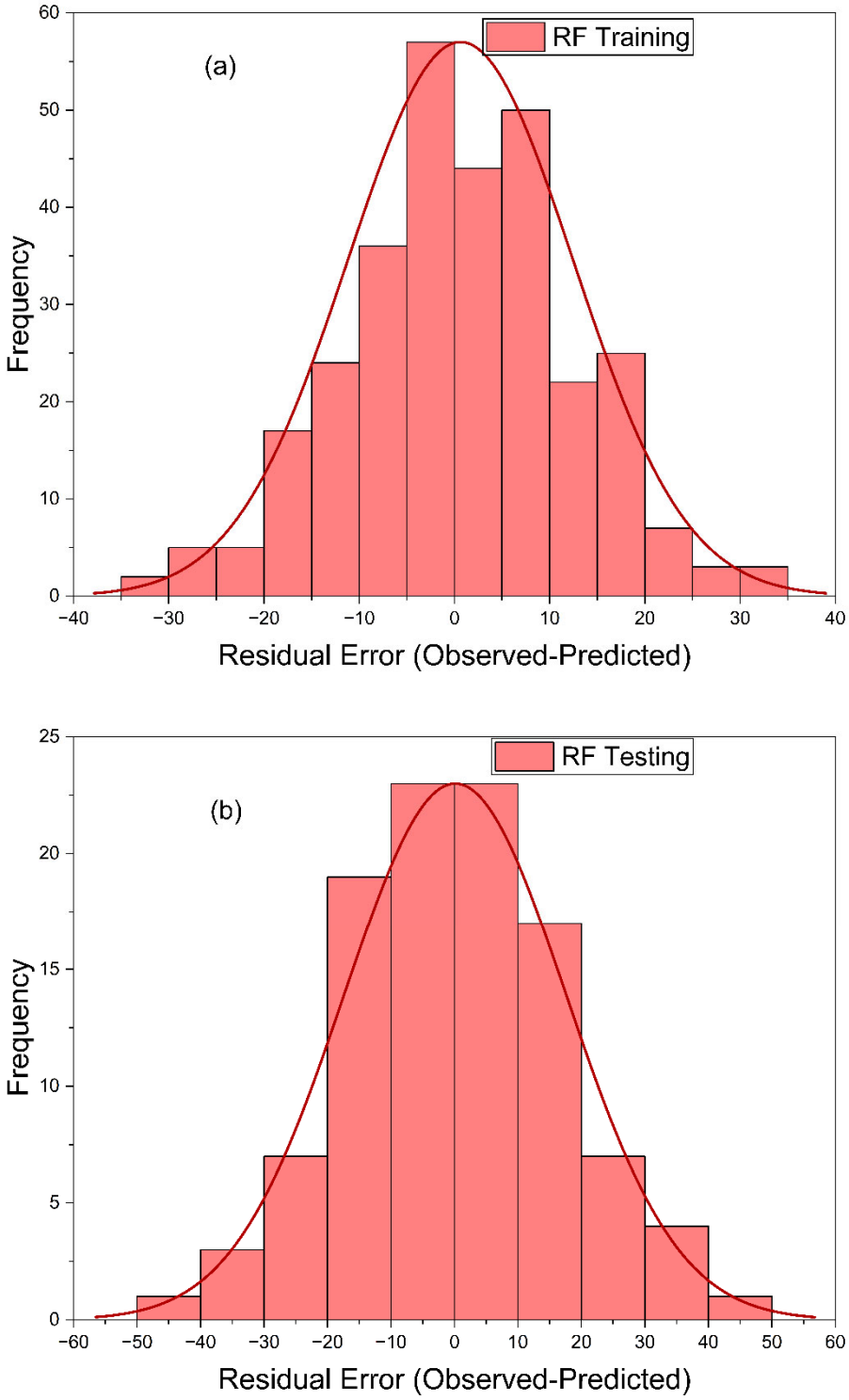

Fig. 8. Error histogram of residuals for the Random Forest Model: a) training dataset, b) testing dataset

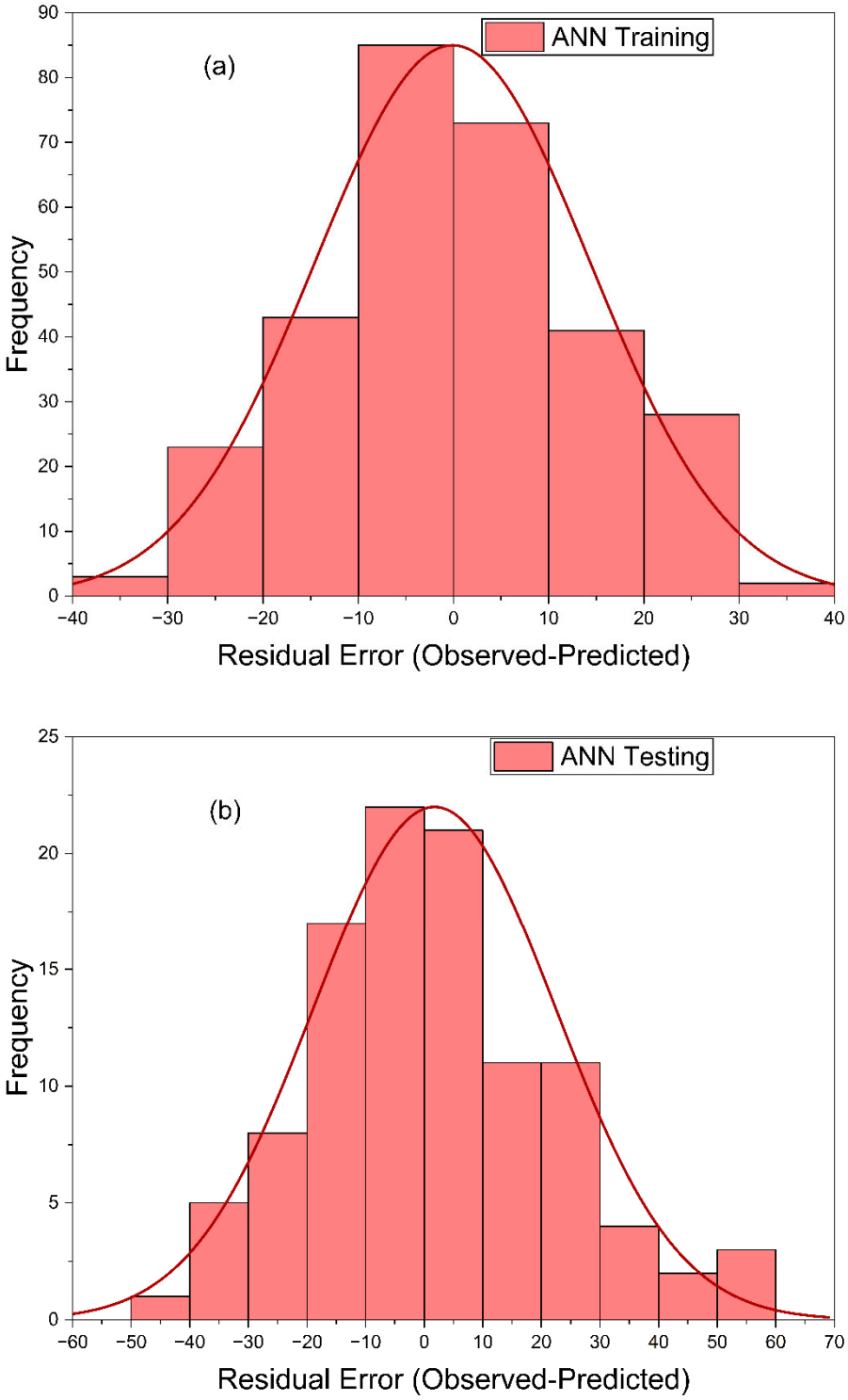

Fig. 9. Error histogram of residuals for the ANN Model: a) training dataset, b) testing dataset

5. Feature Importance Analysis

After the predictive performance analysis, the feature importance analysis was conducted to determine the relative effects of input parameters on CO_2 emissions prediction (Ding et al., 2023; Dong et al., 2025; Hasanipanah et al., 2023). Although high model accuracy indicates strong predictive power, it does not necessarily provide insight into how individual mix design variables contribute to the output. Thus, evaluating feature importance in the context of interpreting model behavior is necessary better to understand environmental sources of embodied carbon in geopolymer concrete.

The feature importance analysis examines the estimation of the influence of each input variable, including binder content, the ratio of sodium silicate to sodium hydroxide (SS/SH), the ratio of activator to binder (A/B), and the ratio of liquid and binder (L/B) in predicting CO_2 emissions. Besides improving transparency in models, this analysis offers engineers a helpful insight into which parameters have the greatest impact on environmental performance.

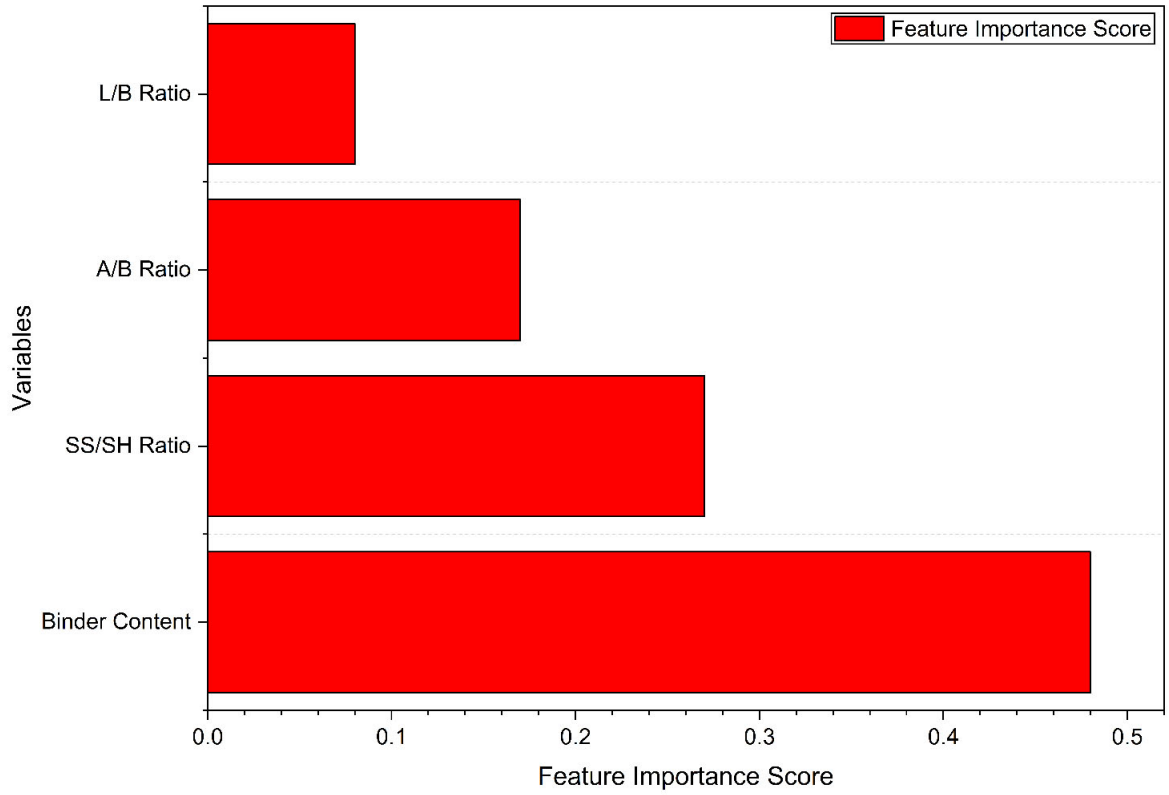

Fig. 10. Feature importance ranking of input parameters for CO_2 emission prediction based on the XGBoost model

The XGBoost model can provide the feature importance analysis of how each input parameter contributes to the prediction of CO_2 emissions in a relative manner. Binder content was the most prominent parameter, as shown in Figure 10, contributing about 48% to the prediction model. This finding supports the thesis that the amount of material used per cubic meter is the primary driver of embodied carbon in geopolymer concrete systems.

The SS/SH ratio was second, demonstrating the strong influence of sodium silicate dosage on overall emissions, as it has a relatively high embodied carbon. The activator-to-binder (A/B) ratio was moderate, indicating that it regulated overall activator consumption. Meanwhile, the coefficient of the liquid-to-binder (L/B) ratio showed the lowest relative values, indicating a relatively low direct effect on CO_2 emissions within the parameter space of the present study.

The preponderance of the binder content and the activator composition indicates that optimizing these factors is critical to minimizing environmental impact. The findings are consistent with the experimental outcomes and support the machine learning framework's ability to determine environmentally sensitive design variables.

6. Environmental Implications

6.1. CO_2 Reduction Compared to OPC Concrete

The ecological behavior of the generated GGBS-based geopolymer concrete mixtures was measured against the cradle-to-gate CO_2 emissions of the conventional Ordinary Portland Cement (OPC) concrete. The mean CO_2 emission of the geopolymer mixtures was around 290 kg CO_2/m^3, whereas the normal OPC concrete has a CO_2 emission of 500–550 kg CO_2/m^3, depending on cement content and production factors.

Based on these values, the geopolymer concrete demonstrated an approximate 40–45% reduction in embodied CO_2 emissions relative to OPC concrete. In optimized low-binder mixtures, emission reductions exceeded 50%, highlighting the substantial environmental advantage of clinker-free systems.

The reduction is primarily attributed to:

- Elimination of cement clinker production
- Lower limestone calcination emissions
- Reduced fossil fuel consumption during manufacturing

6.2. Sustainability Interpretation

The high promise of geopolymer concrete as a green alternative to standard cement-based materials is validated by the significant drop in CO_2 emissions. Cement manufacturing itself accounts for about 7–8% of global anthropogenic CO_2 emissions, which is why substituting OPC with low-carbon binders like geopolymer systems based on GGBS can make a significant contribution to the decarbonization of the construction industry.

Moreover, machine learning incorporated in the optimization of mix designs increases sustainability as it allows:

- Quick filtering of low-emission mix designs.
- Determination of strong parameters of emission driving.
- Lessening of trial-and-error experimentation methods.
- The decision-making of environmental decisions based on data.

The experimental and machine learning system integrated into this research offers a scalable design approach to the realization of environmentally optimized concrete systems. With the aim of reducing the number of binders and optimizing activators, it is possible to save a lot of carbon without sacrificing material performance.

In summary, the results indicate that an environmentally conscious, mix-proportion combination facilitated by predictive modeling is a feasible and viable approach to developing sustainable construction practices.

6.3. Sensitivity Analysis

Effect of Activator-to-Binder Ratio on CO_2 Emissions

The multi-line sensitivity plot shows how the ratio of activators to binder affects CO_2 emissions across varying binder contents. A steady upward trend is observed across all binder levels, indicating that the higher the activator dosage, the higher the embodied carbon. Also, the percentage of binder shifts the emission curves upward, which proves the synergistic effect of the quantity of materials and the structure of the activators on the environmental performance (Figure 11).

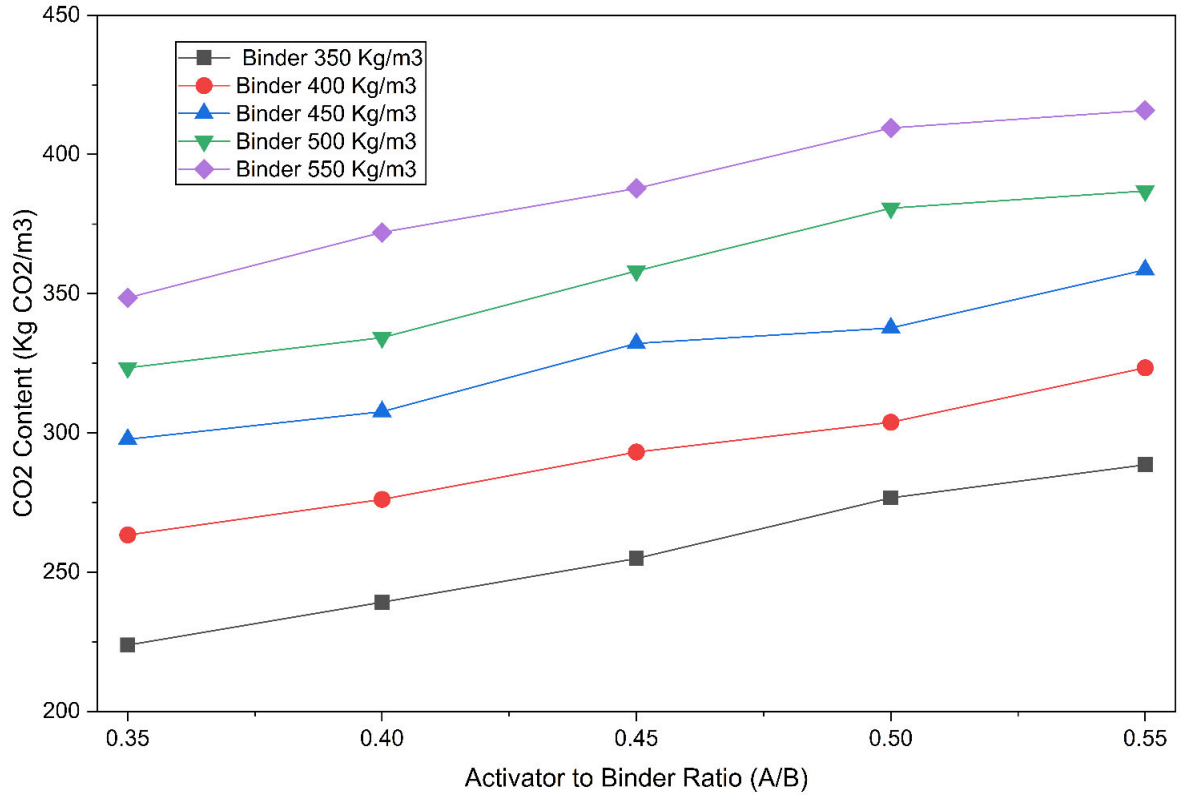

Fig. 11. Sensitivity analysis showing the variation of predicted CO_2 emissions with activator-to-binder (A/B) ratio at different binder content levels

6.4. Pareto analysis

The Pareto front is the set of non-dominated solutions that achieve the lowest achievable CO_2 emissions at a given binder content, and thus represents the environmental optimal limit within the mix design space explored (Figure 12).

The Pareto scatter plot shows the distribution of CO_2 emissions across all geopolymer concrete mixtures being evaluated, with varying binder contents. The red dots indicate a distinct mix design, and the red circles indicate the Pareto frontal of the minimum possible CO_2 developed at every stage of the binder.

The Pareto front identifies the environmental optimum in the design space explored. Combinations on this front will consist of the best activator-to-liquid ratios that result in the minimum embodied carbon at a given binder content. As noted, the lowest CO_2 emissions increase monotonically with binder content, indicating that binder dosage is the primary driver of environmental impact.

The most environmentally friendly design area can be determined by the leftmost area of the Pareto front (around 350 kg/m^3 binder). Recipes above the Pareto front are considered sub-optimal in terms of emissions, since lower CO_2 alternatives exist within the same binder type.

This discussion shows that the integrated experimental and machine learning model can be effective in identifying some environmentally optimized geopolymer concrete formulations.

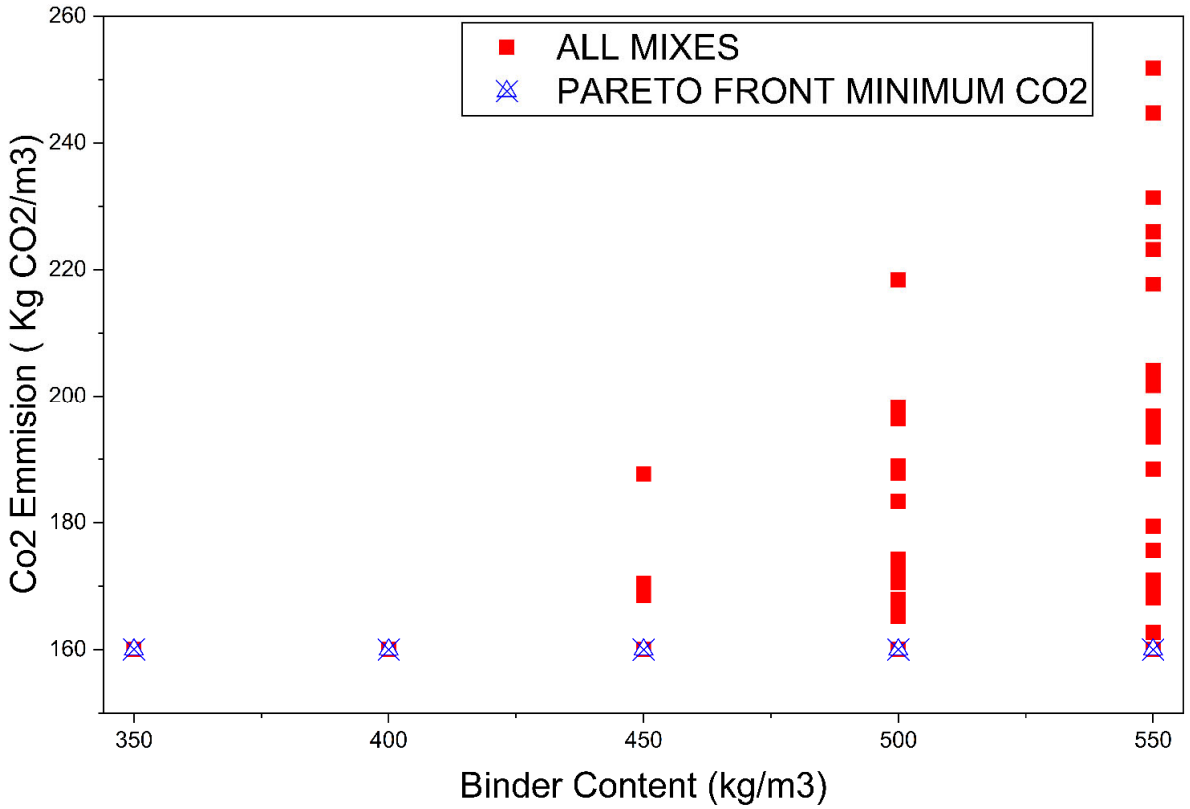

Fig. 12. Pareto scatter plot illustrating CO_2 emissions of all geopolymer mixes as a function of binder content, with the Pareto front (minimum achievable CO_2 at each binder level) highlighted

7. Conclusions

This paper constructed a hybrid experimental and machine learning model to forecast and optimize cradle-to-gate CO_2 emission in low-carbon GGBS-based geopolymer concrete. An extensive experimental dataset consisting of 135 distinct mix designs (405 total mixes, including duplicates) was generated using a full factorial design. The calculated CO_2 emissions ranged from 180 to 420 kg CO_2/m^3, indicating a large variation with changes in binder content and activator composition. Among the parameters investigated, binder content was determined to be the most important factor influencing embodied carbon, followed by the ratios of sodium silicate to sodium hydroxide and of activator to binder. Further Sensitivity and Pareto analyses showed that reducing the binder dosage is the most efficient approach to reducing CO_2 emissions in the studied design space (Haigh, 2026; Pouraminian & Pourbakhshian, 2019; Pujadas-Gispert et al., 2021).

Regarding predictive modeling, the XGBoost algorithm outperformed the Random Forest and Artificial Neural Network models. It had the best coefficient of determination and the lowest values of error (RMSE, MAE, and MAPE). The examination against observed and predicted values showed that it has a good ability to generalize and is less biased compared to other models. The general model performance ranking was set as XGBoost > Random Forest > ANN, indicating the strength of ensemble-based approaches for environmental emission prediction.

Environmentally, the advanced geopolymer concrete mixtures recorded, on average, a reduction of around 40–45 percent in CO_2 emissions compared with conventional OPC concrete, and optimized low-binder mixes were more than 50 percent. This is largely associated with the loss of clinker production and with the optimization of activator use. The results indicate that, when geopolymer concrete is used alongside data-driven optimization methods, the construction industry can take a viable, scalable route to minimize its carbon footprint.

Multi-objective optimization should be included in future research to account for mechanical performance variables, including compressive strength and CO_2 emissions. More sophisticated interpretability methods (such as SHAP analysis) can be used to better understand how models make decisions. Also, by incorporating alternative supplementary materials and different curing regimes into the dataset, better model generalizability and the ability to design real-time sustainable mix design tools would be enhanced.

Acknowledgement

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through the Large Research Project under grant number RGP2/451/46.

References

Abellán-García, J. (2021). K-fold validation neural network approach for predicting the one-day compressive strength of UHPC. *Advances in Civil Engineering Materials*, *10*(1), 223–243.

Al-Abdaly, N. M., Al-Taai, S. R., Imran, H., & Ibrahim, M. (2021). Development of prediction model of steel fiber-reinforced concrete compressive strength using random forest algorithm combined with hyperparameter tuning and k-fold cross-validation. *Eastern-European Journal of Enterprise Technologies*, *5*(7), 113.

Alabdullah, A.A., Iqbal, M., Zahid, M., Khan, K., Amin, M.N., & Jalal, F.E. (2022). Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. *Construction and Building Materials*, *345*, 128296.

Asteris, P.G., & Mokos, V.G. (2020). Concrete compressive strength using artificial neural networks. *Neural Computing and Applications*, *32*(15), 11807–11826.

Barbhuiya, S., Das, B.B., Adak, D., Kapoor, K., & Tabish, M. (2025). Low carbon concrete: Advancements, challenges and future directions in sustainable construction. *Discover Concrete and Cement*, *1*(1), 3.

Cao, Z., Shen, L., Liu, L., & Zhong, S. (2016). Analysis on major drivers of cement consumption during the urbanization process in China. *Journal of Cleaner Production*, *133*, 304–313.

Daehn, K., Basuhi, R., Gregory, J., Berlinger, M., Somjit, V., & Olivetti, E.A. (2022). Innovations to decarbonize materials industries. *Nature Reviews Materials*, *7*(4), 275–294.

Ding, Y., Wei, W., Wang, J., Wang, Y., Shi, Y., & Mei, Z. (2023). Prediction of compressive strength and feature importance analysis of solid waste alkali-activated cementitious materials based on machine learning. *Construction and Building Materials*, *407*, 133545.

Dong, Y., Tang, J., Xu, X., Li, W., Feng, X., Lu, C., Hu, Z., & Liu, J. (2025). A new method to evaluate features importance in machine-learning based prediction of concrete compressive strength. *Journal of Building Engineering*, *102*, 111874.

Duan, J., Asteris, P.G., Nguyen, H., Bui, X.-N., & Moayedi, H. (2021). A novel artificial intelligence technique to predict compressive strength of recycled aggregate concrete using ICA-XGBoost model. *Engineering with Computers*, *37*(4), 3329–3346.

Duan, Z.-H., Kou, S.-C., & Poon, C.-S. (2013). Prediction of compressive strength of recycled aggregate concrete using artificial neural networks. *Construction and Building Materials*, *40*, 1200–1206.

Gao, H., Wang, X., Wu, K., Zheng, Y., Wang, Q., Shi, W., & He, M. (2023). A review of building carbon emission accounting and prediction models. *Buildings*, *13*(7), 1617.

Garces, J.I.T., Beltran, A.B., Tan, R.R., Ongpeng, J.M.C., & Promentilla, M.A.B. (2022). Carbon footprint of self-healing geopolymer concrete with variable mix model. *Cleaner Chemical Engineering*, *2*, 100027.

Habert, G., Miller, S.A., John, V.M., Provis, J.L., Favier, A., Horvath, A., & Scrivener, K.L. (2020). Environmental impacts and decarbonization strategies in the cement and concrete industries. *Nature Reviews Earth & Environment*, *1*(11), 559–573.

Haigh, R. (2026). Environmental and Mechanical Trade-Off Optimization of Waste-Derived Concrete Using Surrogate Modeling and Pareto Analysis. *Sustainability*, *18*(2), 1119.

Hasan, S.M., Islam, T., Saifuzzaman, M., Ahmed, K.R., Huang, C.-H., & Shahid, A.R. (2025). Carbon emission quantification of machine learning: A review. *IEEE Transactions on Sustainable Computing*.

Hasanipanah, M., Abdullah, R. A., Iqbal, M., & Ly, H.-B. (2023). Predicting rubberized concrete compressive strength using machine learning: a feature importance and partial dependence analysis. *Journal of Science and Transport Technology*, *3*(1), 26–43.

Khodaparasti, M., Alijamaat, A., & Pouraminian, M. (2023). Prediction of the concrete compressive strength using improved random forest algorithm. *Journal of Building Pathology and Rehabilitation*, *8*(2), 92.

Kumar, S., Gangotra, A., & Barnard, M. (2025). Towards a net zero cement: strategic policies and systems thinking for a low-carbon future. *Current Sustainable/Renewable Energy Reports*, *12*(1), 5.

Kurtis, K.E. (2015). Innovations in cement-based materials: Addressing sustainability in structural and infrastructure applications. *Mrs Bulletin*, *40*(12), 1102–1109.

Latawiec, R., Woyciechowski, P., & Kowalski, K.J. (2018). Sustainable concrete performance—CO_2-emission. *Environments*, *5*(2), 27.

Lee, S.-C. (2003). Prediction of concrete strength using artificial neural networks. *Engineering structures*, *25*(7), 849–857.

Lim, C., Jung, E., Lee, S., Jang, C., Oh, C., & Shin, K.N. (2020). Global trend of cement production and utilization of circular resources. *에너지공학*, *29*(3), 57–63.

Lin, C.-J., & Wu, N.-J. (2021). An ANN model for predicting the compressive strength of concrete. *Applied Sciences*, *11*(9), 3798.

Ling, H., Qian, C., Kang, W., Liang, C., & Chen, H. (2019). Combination of Support Vector Machine and K-Fold cross validation to predict compressive strength of concrete in marine environment. *Construction and Building Materials*, *206*, 355–363.

Lyu, Z., Yu, Y., Samali, B., Rashidi, M., Mohammadi, M., Nguyen, T.N., & Nguyen, A. (2022). Back-propagation neural network optimized by K-fold cross-validation for prediction of torsional strength of reinforced Concrete beam. *Materials*, *15*(4), 1477.

Mai, H.-V.T., Nguyen, T.-A., Ly, H.-B., & Tran, V.Q. (2021). Prediction compressive strength of concrete containing GGBFS using random forest model. *Advances in Civil Engineering*, *2021*(1), 6671448.

Mardani, A., Liao, H., Nilashi, M., Alrasheedi, M., & Cavallaro, F. (2020). A multi-stage method to predict carbon dioxide emissions using dimensionality reduction, clustering, and machine learning techniques. *Journal of Cleaner Production*, *275*, 122942.

Martiny, A. (2023). *Towards sustainable ai: Monitoring and analysis of carbon emissions in machine learning algorithms* Politecnico di Torino].

Mohsin, M., Ghosh, T., & Hoque, N. (2025). Prediction and Optimization of Strength and CO_2 Emission for Geopolymer Concrete Mix Design using Machine Learning. *Results in Materials*, 100791.

Mustapha, I.B., Abdulkareem, Z., Abdulkareem, M., & Ganiyu, A. (2024). Predictive modeling of physical and mechanical properties of pervious concrete using XGBoost. *Neural Computing and Applications*, *36*(16), 9245–9261.

Nafees, A., Amin, M.N., Khan, K., Nazir, K., Ali, M., Javed, M.F., Aslam, F., Musarat, M.A., & Vatin, N.I. (2021). Modeling of mechanical properties of silica fume-based green concrete using machine learning techniques. *Polymers*, *14*(1), 30.

Nassef, A.M., Olabi, A.G., Rezk, H., & Abdelkareem, M.A. (2023). Application of artificial intelligence to predict CO_2 emissions: Critical step towards sustainable environment. *Sustainability*, *15*(9), 7648.

Nguyen, V.G., Duong, X.Q., Nguyen, L.H., Nguyen, P.Q.P., Priya, J.C., Truong, T.H., Le, H.C., Pham, N.D.K., & Nguyen, X.P. (2023). An extensive investigation on leveraging machine learning techniques for high-precision predictive modeling of CO_2 emission. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, *45*(3), 9149–9177.

Pouraminian, M., & Pourbakhshian, S. (2019). Multi-criteria shape optimization of open-spandrel concrete arch bridges: Pareto front development and decision-making. *World Journal of Engineering*, *16*(5), 670–680.

Pujadas-Gispert, E., Vogtländer, J.G., & Moonen, S. (2021). Environmental and economic optimization of a conventional concrete building foundation: Selecting the best of 28 alternatives by applying the pareto front. *Sustainability*, *13*(3), 1496.

Sandanayake, M., Gunasekara, C., Law, D., Zhang, G., Setunge, S., & Wanijuru, D. (2020). Sustainable criterion selection framework for green building materials–An optimisation based study of fly-ash Geopolymer concrete. *Sustainable Materials and Technologies*, *25*, e00178.

Shen, L., Zhong, S., Elshkaki, A., Zhang, H., & Zhao, J. a. (2021). Energy-cement-carbon emission nexus and its implications for future urbanization in China. *Journal of Sustainable Development of Energy, Water and Environment Systems*, *9*(2), 1–15.

Shobeiri, V., Bennett, B., Xie, T., & Visintin, P. (2023). Mix design optimization of concrete containing fly ash and slag for global warming potential and cost reduction. *Case Studies in Construction Materials*, *18*, e01832.

Smith, A., & Sam, J. (2020). Use of correlation and regression analyses as statistical tools in green concrete research. *GSJ*, *8*(5), 991–1004.

Song, W., Zhang, M., & Wu, H. (2024). Gray correlation analysis between mechanical performance and pore characteristics of permeable concrete. *Journal of Building Engineering*, *86*, 108793.

Sun, Y., Li, G., Zhang, J., & Qian, D. (2019). Prediction of the strength of rubberized concrete by an evolved random forest model. *Advances in Civil Engineering*, *2019*(1), 5198583.

Sun, Z., Li, Y., Li, Y., Su, L., & He, W. (2024). Investigation on compressive strength of coral aggregate concrete: hybrid machine learning models and experimental validation. *Journal of Building Engineering*, *82*, 108220.

Sun, Z., Li, Y., Yang, Y., Su, L., & Xie, S. (2024). Splitting tensile strength of basalt fiber reinforced coral aggregate concrete: optimized XGBoost models and experimental validation. *Construction and Building Materials*, *416*, 135133.

Tam, V.W., Butera, A., Le, K.N., Da Silva, L.C., & Evangelista, A.C. (2022). A prediction model for compressive strength of CO_2 concrete using regression analysis and artificial neural networks. *Construction and Building Materials*, *324*, 126689.

Tian, Y., Ren, X., Li, K., & Li, X. (2025). Carbon dioxide emission forecast: A review of existing models and future challenges. *Sustainability*, *17*(4), 1471.

Wei, J., & Cen, K. (2019). Empirical assessing cement CO_2 emissions based on China's economic and social development during 2001–2030. *Science of the Total Environment*, *653*, 200–211.

Zhang, B., Pan, L., Chang, X., Wang, Y., Liu, Y., Jie, Z., Ma, H., Shi, C., Guo, X., & Xue, S. (2025). Sustainable mix design and carbon emission analysis of recycled aggregate concrete based on machine learning and big data methods. *Journal of Cleaner Production*, *489*, 144734.

Zhang, Y., Peng, J., Wang, Z., Xi, M., Liu, J., & Xu, L. (2025). Machine learning-assisted sustainable mix design of waste glass powder concrete with strength–cost–CO_2 emissions trade-offs. *Buildings*, *15*(15), 2640.

Zhu, L., Zhao, C., & Dai, J. (2021). Prediction of compressive strength of recycled aggregate concrete based on gray correlation analysis. *Construction and Building Materials*, *273*, 121750.