Journal of Cleaner Production 514 (2025) 145727

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Predicting calcium carbonate yield from wet carbonation of recycled cement paste using interpretable ensemble machine learning

Yining Li [a], Ehecatl Antonio del Rio Chanona [b], Hong S. Wong [a], Rupert J. Myers [a,*]

[a] Department of Civil and Environmental Engineering, Imperial College London, London, SW7 2AZ, United Kingdom
[b] Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London, SW7 2AZ, United Kingdom

ARTICLE INFO

Keywords:
Recycled cement paste
Wet carbonation
Machine learning
Ensemble learning
Shapley additive explanations

ABSTRACT

Wet carbonation is a promising method for enhancing the performance of recycled cement paste as a low carbon supplementary cementitious material, however, its optimisation for yield is hindered by an incomplete understanding of its mechanisms and uncertainty about the relative importance of various influencing factors. This study applies advanced machine learning techniques to optimise the wet carbonation of recycled cement paste to maximise calcium carbonate yield, utilising a dataset of 227 experimental records covering diverse process conditions. Six machine learning models—support vector regression, decision tree, random forest, gradient boosting regressor, extreme gradient boosting, and stacking—were trained and evaluated using R^2, MAE, and RMSE as performance metrics. Model evaluation was performed using 10-fold cross-validation to enhance robustness and ensure generalisation across different data subsets. The stacking ensemble model achieved the highest predictive accuracy, with R^2 values of 0.997 (training) and 0.948 (testing), improving prediction accuracy by 19.85 % compared to the least accurate model (decision tree testing). Additionally, the stacking model achieved MAE values of 0.0101 (training) and 0.0393 (testing), and RMSE values of 0.0125 (training) and 0.0536 (testing), demonstrating its superior predictive capability. This model effectively addressed the challenge of limited data, a common issue in applying machine learning in materials science. Shapley additive explanations identified the fineness of cement paste particles and carbonation time as critical factors affecting the calcium carbonate yield variability, contributing 0.14 and 0.11, respectively. Therefore, these results show that finer particle size, extended carbonation time, and the addition of extra OH- ions in the reactor can be applied to develop more efficient wet carbonation processes, enhancing the production of $CaCO_3$ and CO_2 uptake. This research highlights the potential of ensemble learning to optimise concrete recycling techniques, offering a sustainable solution for the construction industry.

Abbreviation List

Abbreviation	Term	Abbreviation	Term
ANN	Artificial Neural Network	MLP	Multilayer Perceptron
C-(A-)S-H	Calcium-(Alumino-) Silicate-Hydrate	R^2	Coefficient of Determination
CV	Cross-Validation	RCP	Recycled Cement Paste
DT	Decision Tree	RF	Random Forest
GB	Gradient Boosting	RMSE	Root Mean Squared Error
GBDT	Gradient Boosting Decision Tree	SCM	Supplementary Cementitious Material
GBR	Gradient Boosting Regressor	SHAP	Shapley additive explanations
MAE	Mean Absolute Error	SVR	Support Vector Regression
ML	Machine Learning	XGBoost	Extreme Gradient Boosting

(*continued on next column*)

(*continued*)

1. Introduction

Recent advancements in the cement and concrete industries have been significantly driven by sustainability considerations. A notable recent development is the increased focus on utilising end-of-life

* Corresponding author.
E-mail address: r.myers@imperial.ac.uk (R.J. Myers).

https://doi.org/10.1016/j.jclepro.2025.145727
Received 3 November 2024; Received in revised form 9 May 2025; Accepted 14 May 2025
Available online 19 May 2025

concrete, especially end-of-life cement paste, often referred to as recycled cement paste (RCP). This development addresses the increasing importance of environmental protection and resource efficiency, and provides an opportunity for significant value creation. RCP has shown promise as a supplementary cementitious material (SCM) in composite Portland cement, especially when carbonated (Kwon et al., 2015; Li et al., 2021).

However, RCP in its untreated form has poor reactivity and unfavourable physical characteristics, which hinder its direct use in cementitious systems (Schoon et al., 2015; Wang et al., 2018). To overcome these limitations, carbonation treatment has been proposed as an effective method to enhance the reactivity and performance of RCP. Among various carbonation approaches, wet carbonation is particularly promising due to its high efficiency in converting reactive phases into calcium carbonate (Zajac et al., 2020a, 2023; Liu et al., 2021).

Predicting the calcium carbonate yield in carbonation processes is useful for optimising them since it characterises the reaction progress, and can be used for reactor design. However, this is a complex task due to the non-linear interactions among multiple variables, such as RCP composition, process conditions, and chemical additives.

Machine learning (ML) is increasingly recognised as an effective approach for modelling complex and multivariable relationships in cementitious systems. Its capacity to identify nonlinear relationships among diverse input factors makes it particularly suitable for predicting material properties and process outcomes in cementitious systems. However, applications of ML in optimising the carbonation of RCP remain limited, and ensemble approaches—particularly stacking—have not been explored in this context.

To address this gap, this study develops and applies a stacking-based ensemble learning model to predict calcium carbonate yield from wet carbonation of RCP, providing a data-driven approach to optimising the carbonation process. This paper includes the following aspects: (1) Five base learner models (SVR, DT, RF, GBR, and XGBoost) for individual predictions. (2) A stacking ensemble learning model established using RF, GBR, and XGBoost as base learner models and linear regression as the meta-learner. (3) Comparing the accuracy of predictions of the stacking ensemble learning model and other ML models using three statistical parameters: coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE) in training and test datasets. (4) Analysis of results using Shapley additive explanations (SHAP) to explore the significance and influence of each input parameter on the target parameter, which here is calcium carbonate yield. Overall, the aim of this paper is to create recommendations for enhancing the effectiveness of wet carbonation of RCP.

2. Literature review

RCP, as the hydrated fraction of end-of-life concrete, has been investigated as a potential SCM due to its favourable chemical composition and recyclability. However, its utilisation in cementitious systems is constrained by several intrinsic characteristics. RCP contains reacted cement and fine aggregates, making it a relatively poorly reactive SCM (Schoon et al., 2015; Wang et al., 2018). Its irregular shape, rough surface, and micropores increase water demand, negatively affecting concrete workability at increasing substitution extents of RCP for Portland clinker (hereafter 'clinker') (Rocha and Toledo Filho, 2023; Oh et al., 2021). Additionally, the mechanical properties of concrete worsen with increasing RCP content because it dilutes the main reactive phases in clinker (C_2S and C_3S) (Cantero et al., 2021; Sun et al., 2021; He et al., 2022). The typically larger particle sizes (>75 μm) and low reactivity of RCP complicate its incorporation into new concrete (Kim and Jang, 2022; Tang et al., 2020).

Carbonation treatment of RCP can be used to enhance its reactivity. This process converts its solid hydrated phases (mainly portlandite and calcium-(alumino-)silicate-hydrate, C-(A-)S-H) into calcium carbonate and a fine (alumino)silicate gel, increasing its reactivity (Li et al., 2021; Mao et al., 2023). Enhanced microstructural, mechanical, and durability properties of concrete containing carbonated RCP are reported (Mehdizadeh et al., 2021; Lu et al., 2018). Substituting PC with carbonated RCP can significantly lower CO_2 emissions, e.g., the carbon footprint of composite PC containing 57 mass% clinker, 38 mass% carbonated RCP, and 5 mass% gypsum is 0.53 kg CO_2-eq. per kg cement, which is lower than that for CEM II (0.63 kg CO_2-eq. per kg cement, clinker-to-cement mass ratio = 0.75) (Driver et al., 2024).

Accurate prediction of calcium carbonate yield is essential for evaluating the progress of carbonation reactions and for optimising reactor operation and process parameters. Wet carbonation of RCP has gained considerable attention to achieve higher degrees of carbonation than dry and semi-dry carbonation, reaching near-complete carbonation within 6 h (Zajac et al., 2020a, 2023; Liu et al., 2021). The efficiency of the wet carbonation of RCP and its performance as a clinker substitute is affected by various factors: cement composition of RCP (cement type, CaO composition in recycled cement, and particle fineness), carbonation process conditions (liquid-to-solid ratio, CO_2 flow rate and concentration, carbonation time, and temperature), and chemical concentrations in the reaction solution (Mg, Na, and OH ions) (Mao et al., 2023; Teune et al., 2023). The influence of combining several of these factors on RCP carbonation is generally poorly known.

Machine learning (ML) can facilitate process optimisation for cement and concrete materials. By analysing various factors such as composition, particle size, and curing conditions, ML models can help identify optimal processing parameters, significantly reducing time and labor involved in traditional methods. Traditional empirical models and numerical simulations often fall short of accurately predicting the behaviour of cement mixes due to the large amount of imperfect and uncertain information in these, and the high number of potentially affecting model parameters (e.g., CO_2 flowrate, temperature, and humidity) (You et al., 2022; Le et al., 2021). Furthermore, the complex interaction between the mechanical properties of cement-based materials and their mix composition, such as binder-to-aggregate ratio, water content, and chemical additives, makes it challenging to quantify certain properties using simple linear or nonlinear empirical models.

ML employs algorithms to identify patterns and deduce relationships within complex and potentially highly multivariable datasets, providing insights with higher accuracy and complexity than conventional regression methods (Jordan and Mitchell, 2015). In recent years, ML models have increasingly been implemented in construction materials research to predict properties such as chloride diffusion (Cai et al., 2020), mechanical properties (Zhang et al., 2020; Nguyen et al., 2021), elastic modulus (Han et al., 2020), slump flow (Gomaa et al., 2021), compressive strength (Abdellatief et al., 2024), and the coefficient of thermal expansion (Nilsen et al., 2019) of concrete. Despite the growing application of ML in concrete research, common challenges include small datasets and complex arrays of input variables. These challenges are significant and can hinder model generalisability and effective application (Li et al., 2022). Over 55 % of studies that have applied ML to concrete materials had sample sizes below 200, and only ~11 % had more than 1000 samples, a threshold often considered sufficient for robust ML models (Li et al., 2022; Teschendorff, 2019).

Ensemble ML models combine the predictions of multiple individual ML models to improve predictive accuracy and robustness (Sagi and Rokach, 2018). Unlike traditional ML approaches, such as single decision trees or linear regression models, which rely on a single predictive model, ensemble methods aggregate the predictions of several models to better capture complex relationships in high-dimensional data (Cai et al., 2020; Nguyen et al., 2021; Asteris et al., 2021). This approach is particularly effective when dealing with small datasets, where traditional models may struggle to generalise effectively due to limited data availability. Ensemble methods can reduce overfitting during training and enhance the ability of the model to generalise on unseen validation data, especially in problems involving complex datasets with numerous features or non-linear relationships. Techniques such as bagging,

boosting, and stacking help mitigate variance and bias, improving model performance (Ribeiro and dos Santos Coelho, 2020).

Bagging is a technique that aims to reduce variance by training multiple versions of a ML model on different subsets of the data, created by sampling with replacement (Ganaie et al., 2022). The predictions from these bagging ensemble models are then averaged to produce a final output, which generally improves accuracy by mitigating overfitting. The random forest (RF) model is the most used bagging algorithm. Boosting focuses on reducing bias and variance by sequentially training models. Each new model attempts to correct the errors made by the previous ones (Ganaie et al., 2022). The final model output is a weighted sum of the individual models, with greater emphasis on those that perform better. This iterative correction process enables boosting algorithms, such as Gradient Boosting Regressor (GBR) and Extreme Gradient Boosting (XGBoost), to achieve high accuracy.

Lyngdoh et al. (2022) assessed various ML models for predicting concrete strength and identified that XGBoost had the best performance (Lyngdoh et al., 2022). The iterative, gradient-based learning approach and the ensemble of multiple weak models in XGBoost allowed it to effectively learn the underlying distribution and provide lower errors (Lyngdoh et al., 2022). Nguyen et al. (2021) implemented support vector regression (SVR), multilayer perceptron (MLP), GBR, and XGBoost to predict the compressive and tensile strengths of high-performance concrete (Nguyen et al., 2021). They found that GBR and XGBoost models demonstrated superior predictive performance because of their ability to handle complex nonlinear relationships between input features, resulting in higher accuracy and computational efficiency. Gomaa et al. (2021) used the RF model to predict the slump flow and compressive strength of fly ash-based alkali-activated concrete, providing a means to optimise its properties (Gomaa et al., 2021). Abdellatief et al. (2025) demonstrated the effectiveness of ML models, including Gradient Boosting (GB), RF, and GPR, in predicting porosity and compressive strength of sustainable foam glass (Abdellatief et al., 2025). Their study highlights the broader applicability of ML techniques in construction materials research, reinforcing the potential of ensemble methods for optimising cementitious materials.

Stacking is an ensemble learning technique designed to improve the predictive performance of ML models by training multiple base models (individual ML algorithms trained on the original data to make predictions) and then combining their outputs using a meta-learner (a model that combines the predictions of base models to make the final prediction). This method capitalises on the complementary strengths of different base models to produce a more robust and accurate final prediction. Unlike other ensemble methods like bagging and boosting, stacking can integrate diverse types of models (e.g., SVR, RF, GBR) within a single framework. The meta learner is trained on the outputs of the base models to refine the final prediction (Ganaie et al., 2022). This method aims to minimise the generalisation error, which is the discrepancy between the performance on training and unseen data.

Li and Song (2023) developed a stacking ensemble learning model to determine the compressive strength of rice husk ash concrete, which outperformed other ML models such as artificial neural network (ANN), SVR, and gradient boosting decision tree (GBDT) (Li and Song, 2023). The stacking model showed superior performance because it effectively combined the strengths of different base models (RF and XGBoost) and used a meta-learner (linear regression) to improve prediction accuracy. Cui et al. (2021) created a stacking model for predicting earthquake casualties, demonstrating its superior prediction accuracy compared to individual models, such as SVR, RF, GBDT, and XGBoost, by effectively integrating the prediction results of multiple base learners (Cui et al., 2021).

3. Methodology

This study involves six ML models: two standalone models and four ensemble models, for predicting calcium carbonate yield from the wet carbonation of RCP. The models are trained using experimental data, and their predictive accuracy is assessed to determine the most effective approach. The methodology used is summarised in Fig. 1.

3.1. Stacking ensemble learning

In this study, the stacking approach involves two layers: the first layer consists of base learners, and the second layer is a meta-learner.

3.1.1. Base learners

The choice of base learners is crucial in stacking as it directly affects the diversity and accuracy of the ensemble. In this study, five common ML models were evaluated:

- Support Vector Regression (SVR) is effective for both linear and nonlinear data. It works by using the kernel trick to implicitly map the input data into a higher-dimensional space (Montesinos López et al.). This implicit mapping allows SVR to compute the dot products of the transformed feature vectors as if the data were transformed into that higher-dimensional space, without actually performing the explicit transformation. The goal of SVR is to find a hyperplane that best fits the data while allowing some flexibility (within a margin) for error. This makes SVR particularly useful for handling complex regression problems where relationships between variables may not be linear.
- Decision Tree (DT) is known for its hierarchical structure, which organises data based on the values of different features. At each step, the tree splits the data into branches using a specific feature, represented by a node. Each node acts as a decision point, determining how the data is divided. The goal is to create branches that lead to subsets of data where the target variable is more consistent or similar within each subset. This increased homogeneity within subsets helps improve the accuracy of classification or prediction (Pekel, 2020).
- Random Forest (RF) improves the performance of decision trees by aggregating the predictions from multiple trees, resulting in more accurate and reliable outcomes. It employs bootstrap sampling, where each tree is trained on a random subset of the data, promoting diversity among the trees (Rodriguez-Galiano et al., 2015). Moreover, RF randomly selects a subset of features for each tree, preventing over-reliance on any single feature. These strategies help mitigate overfitting, where a model might otherwise become too tailored to the training data and perform poorly on new data. Consequently, RF enhances the overall stability of the model, making its predictions more robust and consistent.
- Gradient Boosting Regressor (GBR) builds a series of decision tree models sequentially, where each new model is trained to correct the errors of the previous one (Natekin and Knoll, 2013). This approach focuses on reducing the residuals, which are the differences between the actual and predicted values. By targeting these residuals, each subsequent model aims to better predict the data points that were previously difficult to predict accurately, typically those with the largest errors. This iterative process gradually improves the overall accuracy of the predictions.
- Extreme Gradient Boosting (XGBoost) is an advanced implementation of the GBR with several key improvements (Chen and Guestrin, 2016). It uses both first and second-order derivatives, providing more precise optimisation. XGBoost incorporates L1 and L2 regularisation to control model complexity and prevent overfitting. Unlike the sequential nature of GBR, XGBoost employs parallel processing for faster training and can efficiently handle missing values. By focusing on both the residuals and the curvature of the loss function, XGBoost can adjust for previous errors and consider the rate of change in these errors, offering enhanced performance.

The general architecture of each model is shown in Fig. 2, respectively.

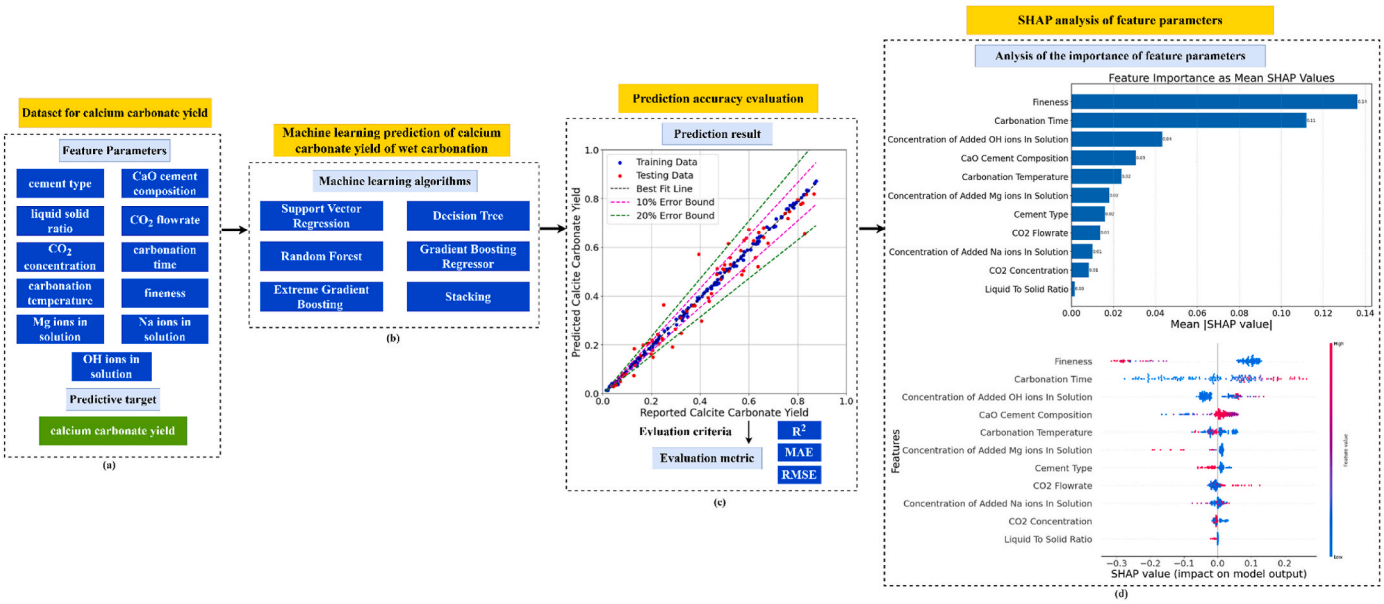

Fig. 1. Overview of methodology for predicting calcium carbonate yield of wet carbonation process.

3.1.2. K-fold cross-validation

K-fold cross-validation (CV) is used to estimate how well a model will generalise to testing data based on the training data. It is especially used when dealing with limited data. In this study, we employ ten-fold CV. The dataset is divided into ten equal-sized parts, called folds. The model is trained ten times, each time using nine folds for training and the remaining fold for validation, as illustrated in Fig. 3. This rotation ensures that each fold serves as a validation set exactly once.

This process helps in obtaining reliable performance metrics, such as R^2, MAE, and RMSE, by averaging the results across all folds, ensuring a comprehensive evaluation of the predictive capability of the model. This technique maximises the utility of the training dataset in validating the model used, thereby mitigating bias (i.e., the error introduced by approximating a real-world problem with a simplified model) and variance (i.e., the sensitivity of the model to small fluctuations in the training set) (Fushiki, 2011).

3.1.3. Meta learner

Following the CV, the predictions from each base model (RF, GBR, and XGBoost) are combined to form a new dataset. This new dataset serves as input for the meta-learner, as shown in Fig. 4. In this study, a linear regression model was chosen as the meta-learner due to its simplicity and effectiveness in combining relationships linearly. This approach leverages the strengths of each base learner to improve overall prediction accuracy and aims to minimise error when the model is applied to unseen data.

3.1.4. Stepwise implementation of the stacking algorithm

The stepwise implementation of the ML algorithms, as depicted in Fig. 5, illustrates the process for predicting calcium carbonate yield in this study.

The process starts with data collection and preprocessing, which serves as the input. This data is then randomly divided into two parts: 70 % for training and 30 % for testing.

The training data is further split into ten equal folds for 10-fold CV. Each base model, such as RF, GBR, and XGBoost, is trained using nine folds and validated on the remaining fold. This process is repeated ten times, and the predictions from each base model are combined to form a new dataset.

This new dataset is used as input for linear regression as the meta-learner. The meta-learner then produces the final predictions for calcium carbonate yield.

Finally, the models are evaluated using unseen testing data, providing an unbiased measure of their performance. The evaluation results determine the predicted calcium carbonate yield from the wet carbonation process as the final output.

3.2. Performance evaluation metric

To evaluate ML model performance in predicting calcium carbonate yield, three primary metrics are employed: coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE).

The proportion of variance in the dependent variable that is predictable from the independent variables is quantified by R^2, which provides a measure of explanatory power (Eq. (1)). An R^2 value close to 1 indicates that the model explains a large portion of the variance, while an R^2 value close to 0 indicates that the model does not explain much of the variance.

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

Where here y_i is the actual calcium carbonate content of the ith sample point in dataset, $\hat{y}_i$ is the predicted value by the ML model for the ith sample point, $\bar{y}$ is the average value of calcium carbonate yield, and n is the number of sample points.

The average of the absolute differences between the predicted values ($\hat{y}_i$) and the actual values (y_i) is measured by MAE (Eq. (2)). It reflects the typical magnitude of prediction errors. Generally, a lower MAE value indicates better model performance, with values closer to 0 being ideal.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (2)$$

The square root of the average squared differences between predicted and actual values is calculated by RMSE (Eq. (3)). It is particularly sensitive to large errors, providing insight into the error magnitude. Lower RMSE values indicate better model performance, with typical acceptable values depending on the specific context but generally lower values are preferred.

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

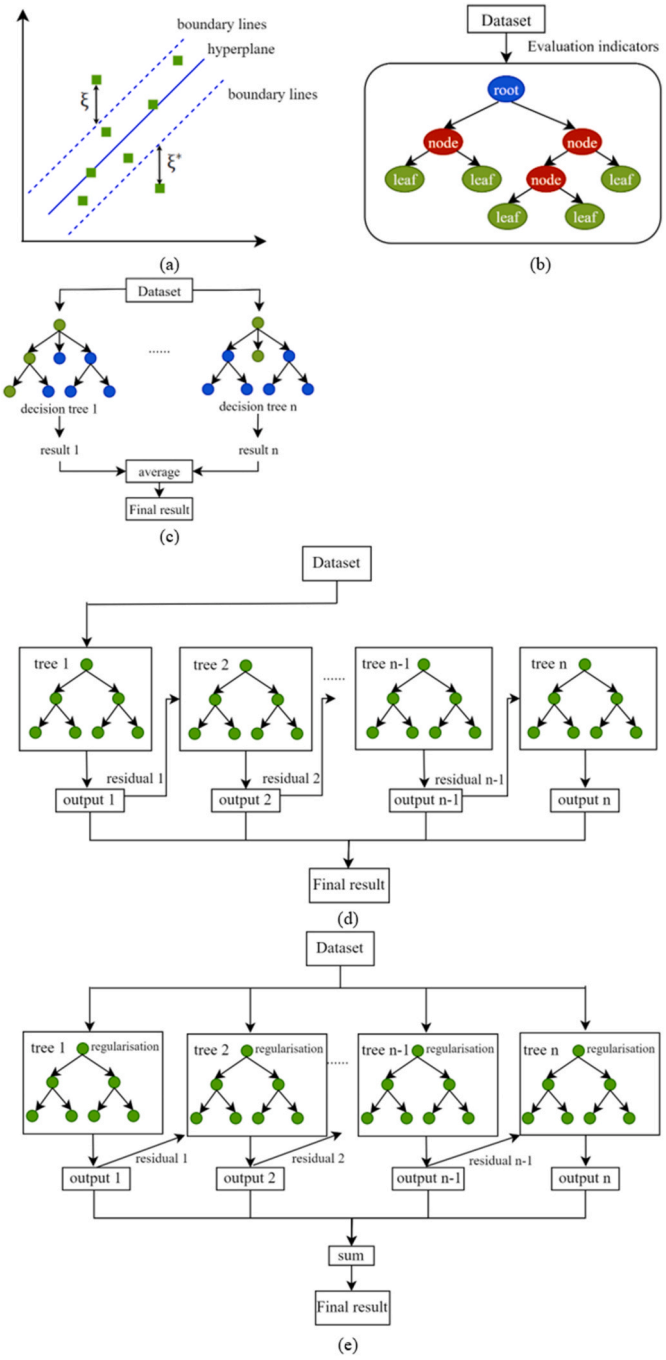

Fig. 2. Schematic diagram of the selected ML models (a) Support Vector Regression (SVR), (b) Decision Tree (DT), (c) Random Forest (RF), (d) Gradient Boosting Regressor (GBR), (e) Extreme Gradient Boosting (XGBoost).

3.3. Shapley additive explanations (SHAP)

Interpretability remains a critical challenge in ML. To address this, SHAP applies the concept of Shapley values from cooperative game theory (Shapley, 1953) to explain the contributions of each feature (such as cement fineness, carbonation time, and chemical concentrations in this study) to the target parameter values predicted by the ML model (Lundberg and Lee, 2017). SHAP values quantify the influence of each feature on the model prediction. A SHAP value of zero indicates that the feature has no effect on the predicted outcome. Positive SHAP values suggest the feature increases the prediction, while negative values indicate the feature decreases it. SHAP values ensure a 'fair' distribution of feature contributions based on the principles of efficiency (the total contribution of all features equals the difference between the actual and average predictions), dummy (a feature that does not influence the prediction has a SHAP value of zero), symmetry (features with equal influence have identical SHAP values), and additivity (the SHAP value of a feature in a combined model equals the sum of its SHAP values from individual models) (Keshk et al., 2023). Incorporating SHAP into ML model analysis provides both the overall importance of features and the contributions of features to individual predictions, which enhances model transparency and interpretability.

The SHAP value of a feature is determined using Eq. (4), which here considers the marginal contribution of the feature on the wet carbonation of RCP across all possible combinations of features.

$$\varphi_j(f,x) = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left[f\left(x_{s \cup \{j\}}\right) - f(x_s) \right] \quad (4)$$

Where $\varphi_j(f,x)$ denotes the SHAP value for the feature j, x represents a specific data point or instance from the dataset being explained, N is the set of all features, S is a subset of N excluding j (denoted by $S \subseteq N\backslash\{j\}$ in Eq. (4)), and f represents the prediction function of the model, which the ML model uses to generate predictions given a set of input features.

3.4. Data collection and preprocessing

3.4.1. Data collection

Achieving suitable quality, diversity, and volume of datasets is critical to the predictive and generalisation capabilities of ML models. For this study, we created a database of 227 datasets that were collected from various published sources (Zajac et al., 2020a, 2020b, 2021a, 2021b, 2022a, 2024; Mao et al., 2023, 2024; Teune et al., 2023; Shen et al., 2022a, 2022b, 2023a, 2023b; Jiang et al., 2022, 2023), which covers the range of factors influencing the RCP carbonation process. Detailed information about these datasets is provided in the Appendix (supplementary data).

3.4.2. Feature selection

We define eleven input features that are relevant to wet carbonation of RCP. These input features, and brief explanations describing their relevance, are provided below:

1. Cement type (CEM I and CEM III): Different cement types have varying chemical compositions and microstructural properties that influence carbonation. This study categorises the cement paste in RCP into two primary types, 'CEM I' and 'CEM III,' with further subcategorisation within each type. This classification ensures both computational efficiency and precise cement characterisation.
2. CaO composition in RCP (wt.%): The availability of calcium ions directly determines the extent of calcium carbonate formation during carbonation.
3. Fineness (mm): The fineness of RCP particles affects the surface area available for reaction. In this study, particle sizes were discretised into four distinct bins: 0–0.15 mm, 0.15–0.6 mm, 0.6–1.18 mm, and 1.18–2.36 mm, labelled by their upper size limits (0.15, 0.6, 1.18, and 2.36) following Jiang et al. (2022). Since different studies report particle sizes using various standards, such as D90, average size, or particle size distribution figure, the particle size bins in this study were simplified for consistency, and thus reflect typical RCP particle sizes for studies reported in the literature. The size bins were selected to represent distinct physical and chemical behaviours observed during carbonation processes. Finer particles (0–0.15 mm) tend to disintegrate and form calcium carbonate and silica gel composites. The 0.15–0.6 mm range is considered a transition zone, where a balance between decomposition and densification is observed. The coarser particles (0.6–2.36 mm) primarily exhibit surface densification and improve surface properties, but because of the differences in

Fig. 3. Ten-fold cross-validation (CV) schematic.

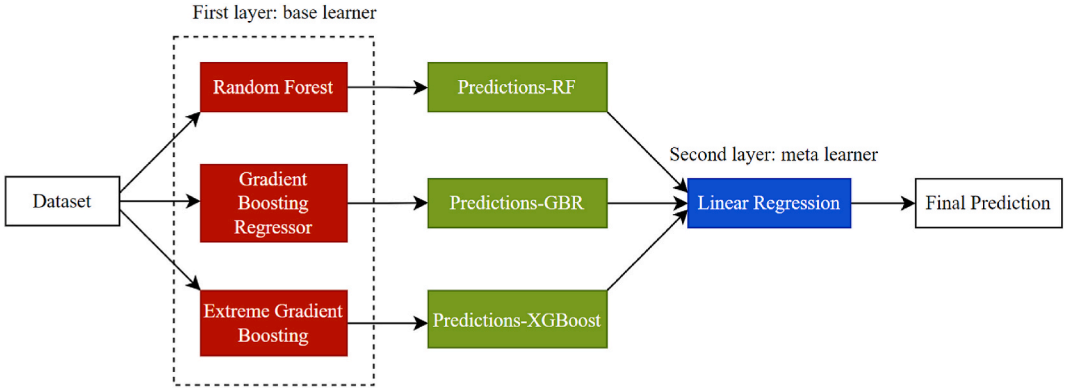

Fig. 4. Flowchart of the stacking-based ensemble learning method.

dissolution and carbonation kinetics, further splitting this range into 0.6–1.18 mm and 1.18–2.36 mm allows for a clearer representation of these differences.

4. Liquid-to-solid ratio (ml g^{-1}) in the carbonation reactor: This factor influences the CO_2 dissolution, Ca extraction, and $CaCO_3$ precipitation rates, particularly affecting the initial reaction dynamics and overall stability of the carbonation process.
5. CO_2 flowrate (L min^{-1} g^{-1}) and concentration (volume fraction) into the carbonation reactor: These factors affect the carbonation reaction by supplying CO_2 molecules for reaction with calcium ions.
6. Carbonation time (min): The duration during which the wet carbonation reaction occurs, representing the exposure time of RCP to CO_2 in the presence of water under controlled conditions in the reactor. The carbonation time significantly influences the extent of conversion and the characteristics of the carbonation products
7. Temperature (°C) in the carbonation reactor: Temperature affects the kinetics of the carbonation reaction.
8. Concentrations of added Mg, Na, and OH ions (mol L^{-1}) in solution in the carbonation reactor: These ions are added as chemical additives to the reaction solution, influencing the carbonation process by altering the solution chemistry, and impacting solubility and precipitation dynamics.

The output calcium carbonate yield measures the efficiency of the wet carbonation process by quantifying the amount of calcium carbonate formed relative to the theoretical maximum that could be achieved based on the available calcium sources. This efficiency is crucial in assessing the overall performance of the carbonation process. The calculation of the calcium carbonate yield is expressed in Eqs. (5) and (6).

$$relative\ yield = \frac{Mass\ of\ calcium\ carbonate\ precipitated\ in\ the\ carbonation\ process}{Theoretical\ mass\ of\ calcium\ carbonate\ in\ the\ carbonation\ process} \quad (5)$$

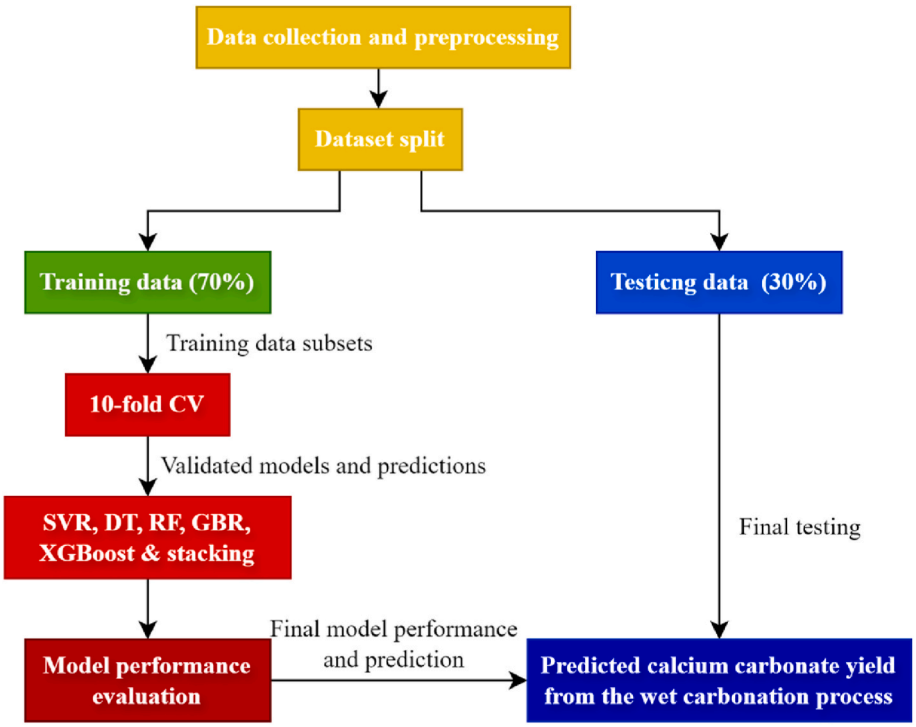

Fig. 5. Flowchart of the ML-based prediction process for calcium carbonate yield in wet carbonation.

Table 1
Properties of input and output parameters.

	Attribute	unit	Min.	Max.	Mean	Std. Dev.
Input	Cement type	N/A	1	3		
	CaO cement composition	N/A	0.40	0.67	0.59	0.084
	Liquid-to-solid ratio	ml g^{-1}	10.0	53.3	12.3	7.50
	CO_2 flowrate	L min^{-1} g^{-1}	0.002	0.167	0.03	0.04
	CO_2 concentration	volume fraction	0.10	1.00	0.45	0.41
	Carbonation time	min	1.50	400.2	96.8	119
	Carbonation temperature	°C	20.0	80.0	33.6	22.3
	Fineness	mm	0.15	2.36	0.52	0.77
	Concentrations of added Mg ions in solution	mol L^{-1}	0.00	0.80	0.03	0.10
	Concentrations of added Na ions in solution	mol L^{-1}	0.00	1.00	0.18	0.29
	Concentrations of added OH ions in solution	mol L^{-1}	0.00	1.00	0.10	0.19
Output	Calcium carbonate yield	N/A	0.015	0.87	0.40	0.24

$$relative\ yield = \frac{CaCO_3{}^{product} - CaCO_3{}^{input}}{\frac{CaO^{input} \times M_{CaCO_3}}{M_{CaO}} - CaCO_3{}^{input} - \frac{CaSO_4{}^{input} \times M_{CaCO_3}}{M_{CaSO_4}}} \quad (6)$$

Where $CaCO_3^{product}$ represents the final mass of $CaCO_3$ in the reactor product, $CaCO_3^{input}$, CaO^{input} and $CaSO_4^{input}$ are the initial masses of $CaCO_3$, CaO and $CaSO_4$ in the reactor, M_{CaCO_3}, M_{CaO} and M_{CaSO_4} represent the molecular masses of $CaCO_3$, CaO, and $CaSO_4$ respectively.

Each parameter is described based on its actual data range and frequency, without assuming a specific distribution type. Table 1 provides a statistical summary of the input and output variables.

4. Results and discussions

4.1. Prediction results and performance evaluation

Of the ML models used here, ensemble models, particularly the stacking ensemble model, are the most effective (Table 2). The stacking

Table 2
Prediction performance of ML models.

		Base learner	Base learner	Base learner/Ensemble	Base learner/Ensemble	Base learner/Ensemble	Ensemble (RF + GBR + XGBoost)
		SVR	DT	RF	GBR	XGBoost	Stacking
R^2	Training	0.850	0.833	0.889	0.907	0.923	**0.997**
	testing	0.803	0.791	0.867	0.890	0.904	**0.948**
MAE	training	0.0774	0.0646	0.0528	0.0443	0.0466	**0.0101**
	testing	0.0976	0.0734	0.0541	0.0539	0.0495	**0.0393**
RMSE	training	0.0927	0.0989	0.0731	0.0613	0.0624	**0.0125**
	testing	0.0974	0.109	0.0796	0.0801	0.0770	**0.0536**

model demonstrated high R^2 values (0.997 for training and 0.948 for testing) along with low MAE and RMSE values, indicating its proficiency in identifying complex patterns and enhancing prediction accuracy, even with limited data. This model significantly outperformed single-predictor models (Table 2; SVR, DT, RF, GBR), showing the advantage of combining multiple models to boost predictive accuracy.

The individual ML models for predicting calcium carbonate yield from wet carbonation of RCP show varied performances (Fig. 6).

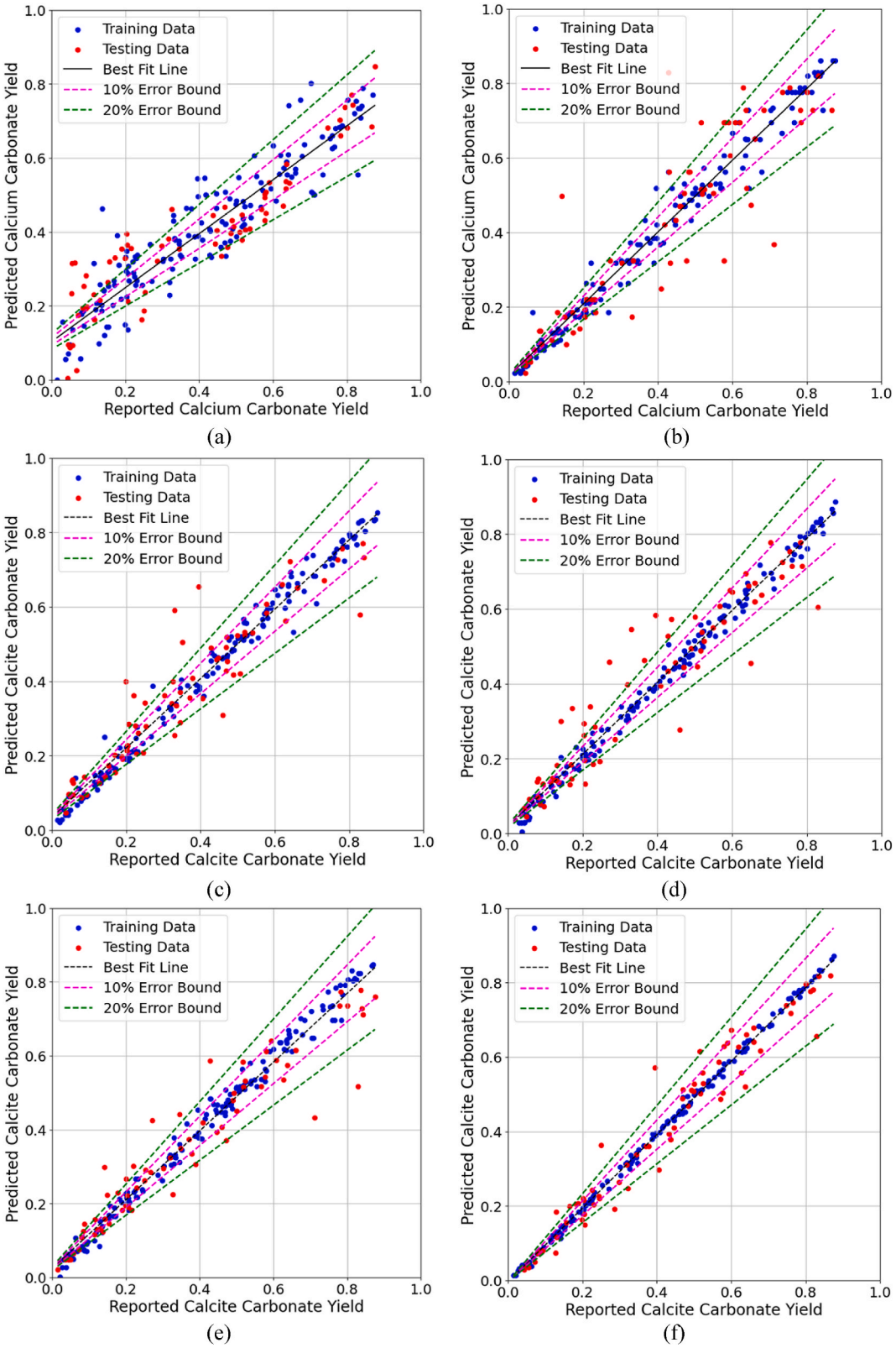

Fig. 6. Comparison of model prediction results using ML models: (a) Support Vector Regression (SVR), (b) Decision Tree (DT), (c) Random Forest (RF), (d) Gradient Boosting Regressor (GBR), (e) Extreme Gradient Boosting (XGBoost), (f) Stacking.

Ensemble models like GBR and XGBoost have higher accuracy, with the latter having slightly higher performance in terms of R^2 (higher) and MAE (lower) values. This result shows the strength of ensemble strategies in improving model generalisation by combining multiple decision trees. The RF model, although not as precise as GBR and XGBoost, effectively reduces overfitting, as indicated by closer training and testing R^2 values compared to SVR and DT. This result demonstrates the capability of RF in handling complex datasets.

The enhanced performance of ensemble models, particularly stacking, on this small dataset suggests that combining multiple algorithms provides resilience against the challenges typically posed by small sample sizes, such as overfitting and reduced generalisability.

The SVR model shows a moderate match between predicted and actual calcium carbonate yield (Fig. 6a), with many points exceeding the 20 % error margin. The RF model achieves greater precision (Fig. 6c), with more than 80 % data points densely packed within the ±20 % error margin. The GBR model more closely matches the actual values, such as SVR, and DT. This is due to its robust integration of decision trees with boosting techniques, which enhances its predictive accuracy and ability to generalise to unseen data. Similarly to the GBR model, the XGBoost model concentrates predictions around the ideal fit, with less scatter at higher values of calcium carbonate yield, indicating robustness — the ability of model to consistently make accurate predictions despite variations in the data. The stacking model is the most precise and its predictions mainly fall within the 10 % error range, highlighting the advantages of combining different ML algorithms for superior predictive capability and generalisability. Overall, ensemble models, particularly the stacking approach, demonstrated robust predictive capability in handling complex data relationships even within the constraints of a small dataset. This finding underscores the practical value of ensemble methods in domains where data collection is limited but predictive accuracy is critical, such as material science.

4.2. Global explanation for SHAP analysis

4.2.1. SHAP values for feature importance

Fig. 7 illustrates the feature importance derived from SHAP analysis, indicating the influence of each feature on maximising calcium

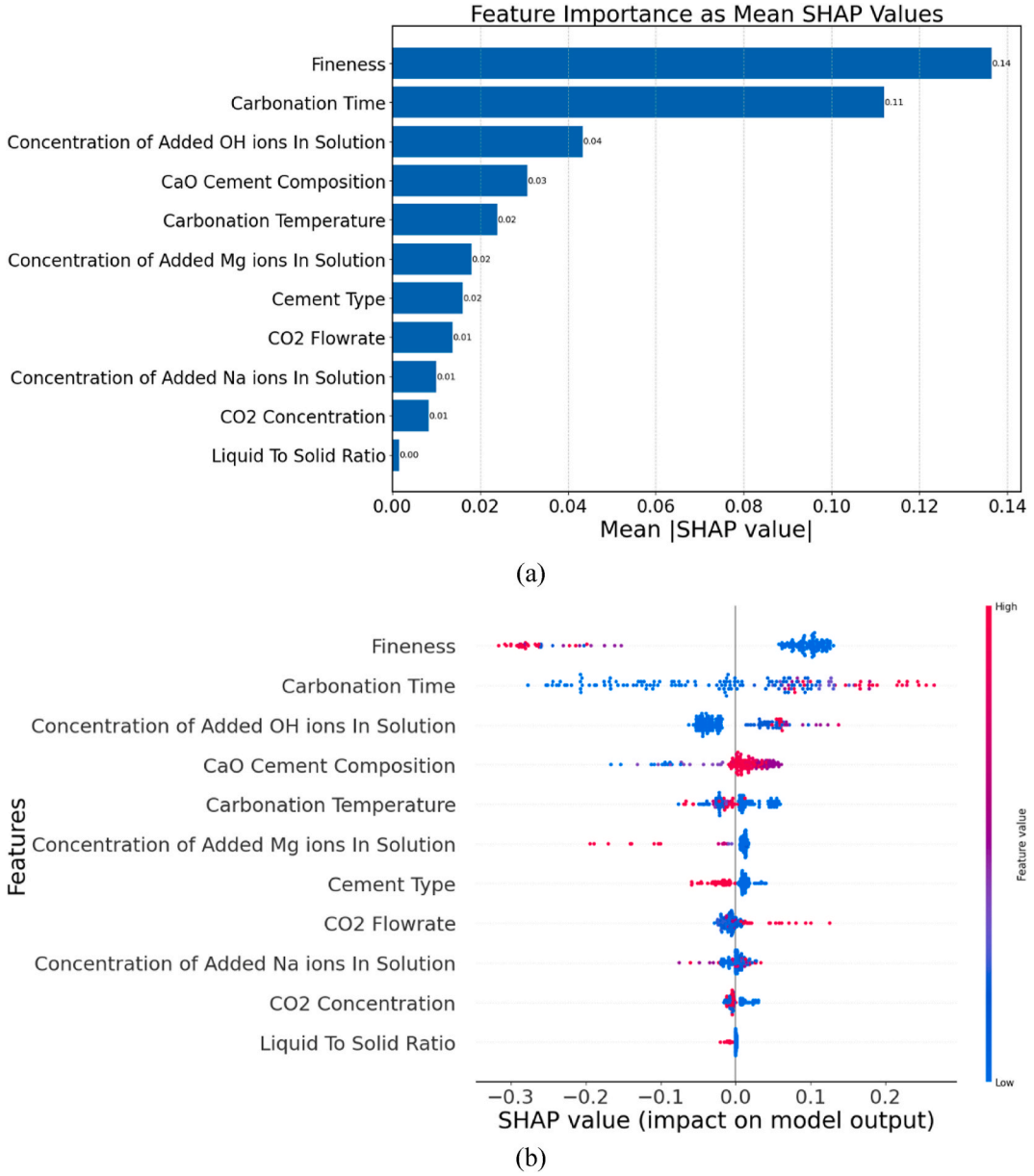

Fig. 7. Feature importance based on SHAP values, (a) the mean absolute SHAP values. (b) the summary plot of SHAP values.

carbonate yield in the wet carbonation process. Fineness and carbonation time are the most important features, which have mean SHAP values of 0.14 and 0.11 respectively. The concentration of added OH^- ions in solution is moderately important with a mean SHAP value of 0.04, while CaO cement composition has a notable influence at a mean SHAP value of 0.03. Carbonation temperature, the concentration of added Mg ions in solution, cement type and CO_2 flowrate show minor effects on calcium carbonate yield. CO_2 concentration and liquid to solid ratio have the least influence. Temperature is a significant factor in determining the relative stability of calcium carbonate polymorphs, however, our results here indicate that its influence on overall yield is relatively low (Chang et al., 2017). The minor effect of CO_2 flowrate on calcium carbonate precipitation is likely because all flow rates result in systems with excess CO_2, where the yield is limited by the availability of dissolved calcium rather than the amount of CO_2. Additionally, the impact of CO_2 flowrate varies across reactor types (Liendo et al., 2022).

Although the liquid-to-solid ratio has a lesser impact on wet carbonation, water content still significantly influences the reaction mechanisms and carbonation products compared to semi-dry carbonation. In wet carbonation, water promotes the separation of calcium carbonate and aluminosilica gel, creating an open microstructure, which accelerates the carbonation process and leads to a higher degree of carbonation (Liendo et al., 2022). Conversely, semi-dry carbonation causes the co-precipitation of products, resulting in denser cement grains and a slower carbonation rate (Zajac et al., 2022b). Utilising CO_2 concentration below 50 % led to only a slight reduction in the carbonation rate, while the properties of the carbonation products remained consistent (Shen et al., 2023a). This highlights the potential for directly sequestering industrial flue gases containing 10–30 % CO_2 in future applications (Shen et al., 2023a).

The sensitivity of SHAP values to data availability and feature magnitude may influence the assessment of feature importance (Lundberg and Lee, 2017). If the dataset used in the SHAP analysis lacks sufficient variability in certain features, their contribution to the model output may be underrepresented. Additionally, if other features exhibit a broader range of values, SHAP is likely to attribute greater importance to those features due to their wider magnitude distribution.

The bimodal SHAP distribution for fineness suggests distinct influences under different conditions. These dual effects likely arise from the differing behaviours of particle sizes during carbonation. Finer particles enhance carbonation due to their larger specific surface area, while coarser particles may have a slower or negative impact due to reduced specific surface area or carbonation induced surface densification.

The wide range of SHAP values for carbonation time highlights its complex, non-linear influence on the carbonation process, with varying influence across different samples or conditions. In some cases, longer carbonation times are highly beneficial (large positive SHAP values), while in others, the influence is less pronounced (smaller positive SHAP values). This variation could be due to factors such as the availability of dissolved calcium, particle size, or the presence of additional ions, which influence how much additional carbonation time contributes to the process.

4.2.2. *Dependency analysis using SHAP*

Dependency plots (Fig. 8) offer a detailed view of the influence of each feature varies across its range of values, revealing specific interactions and non-linear effects of each parameter on the wet carbonation process. Here, a negative SHAP value for a parameter means that it contributes to reducing the predicted calcium carbonate yield relative to the baseline prediction (e.g., the average yield).

In wet carbonation, finer particles (smaller fineness values) have a larger specific surface area, enhancing the dissolution rate and ultimately promoting calcium carbonate formation (Jiang et al., 2022). This leads to more efficient formation of calcium carbonate. As particle size increases, the surface area decreases, reducing the dissolution rate. This reduction likely explains the negative SHAP values associated with larger particles (Fig. 7b, red data points).

Carbonation time positively impacts calcium carbonate yield, as longer carbonation allows more CO_2 to react with calcium-bearing phases like $Ca(OH)_2$ and C-S-H, forming calcium carbonate (Zajac et al., 2020b). This is reflected in the SHAP values (Fig. 8b), where early carbonation times (0–50 min) show predominantly negative SHAP values, indicating that shorter carbonation times reduce the predicted calcium carbonate yield due to incomplete conversion. As carbonation time increases beyond 50 min, the SHAP values transition from negative to positive, showing that extended carbonation time has a stronger positive influence on calcium carbonate yield. During this initial stage (0–100 min), the rapid reaction of CO_2 with these phases results in a significant increase in calcium carbonate yield.

However, beyond approximately 100 min, the SHAP values begin to stabilise, indicating a slowdown in the carbonation rate as most accessible reactive phases are consumed, limiting further reactions (Shen et al., 2022b). This shift is evident in the flattening of SHAP values between 100 and 360 min in the plot, suggesting that the effect of additional carbonation time becomes less significant in increasing calcium carbonate yield. The slowdown is likely due to complete conversion being achieved, or the densely packed layer on the surface and slower decalcification rate of C-S-H (Shen et al., 2023b).

The influence of the concentration of added OH^- ions in solution on the carbonation process shows a transition from a negative to a positive effect as their concentration increases. At zero OH^- ion concentration added in solution, the SHAP plot shows a negative influence on carbonation. This might be because without sufficient OH^- ions, the pH of the solution is lower, and the solubility of calcium carbonate is higher at lower pH. Therefore, calcium carbonate yield is lower at low pH. As OH^- ion concentration increases, the SHAP values turn positive, indicating a beneficial effect on the carbonation process. Higher concentrations of OH^- ions increase the pH of the solution and stability of calcium carbonate, which is less soluble under higher pH conditions.

Higher CaO compositions in cement positively influence calcium carbonate formation during wet carbonation, as indicated by the SHAP plot (Fig. 8c). The increased CaO content leads to a greater availability of calcium ions dissolved from RCP into the solution, which in turn allows for more calcium carbonate to precipitate. This elevated calcium ion concentration enhances the yield of calcium carbonate by providing additional reactants for the precipitation process.

Mg ions in solution negatively affect the carbonation process, as indicated by decreasing SHAP values with increasing Mg concentration. Mg ions inhibit calcite formation by adsorbing onto growing crystals, which interferes with their growth. This promotes the formation of less stable carbonate phases like aragonite or amorphous calcium carbonate (Shen et al., 2023a). Additionally, Mg can react with carbonate ions to form magnesium carbonate phases, reducing carbonate availability and altering the solution pH (Shen et al., 2023b). The presence of Mg ions in solution can also lead to formation of brucite on the surface of RCP particles, affecting their dissolution behaviour (Shen et al., 2023a).

Regarding the moderately important parameters, the cement type in RCP significantly influences calcium carbonate formation during wet carbonation. In CEM I cement, a higher amount of clinker dissolves, resulting in more calcium available in solution to react and form carbonate. In contrast, CEM III contains less clinker, and SCMs are largely inert during short reaction times. Consequently, these SCMs dissolve less, leading to reduced calcium availability for carbonation. Additionally, SCMs contain inherently less calcium than clinker, further limiting the extent of carbonation in CEM III.

The SHAP plot indicates that while the CO_2 flow rate does influence calcium carbonate formation, its effect is relatively minor compared to other factors. Lower flowrates result in slower carbonation and less calcium carbonate formation, while higher flowrates enhance CO_3^{2-} production, promoting more extensive carbonation (Liu and Yates, 2006). However, once the CO_2 flow rate reaches a certain level, further

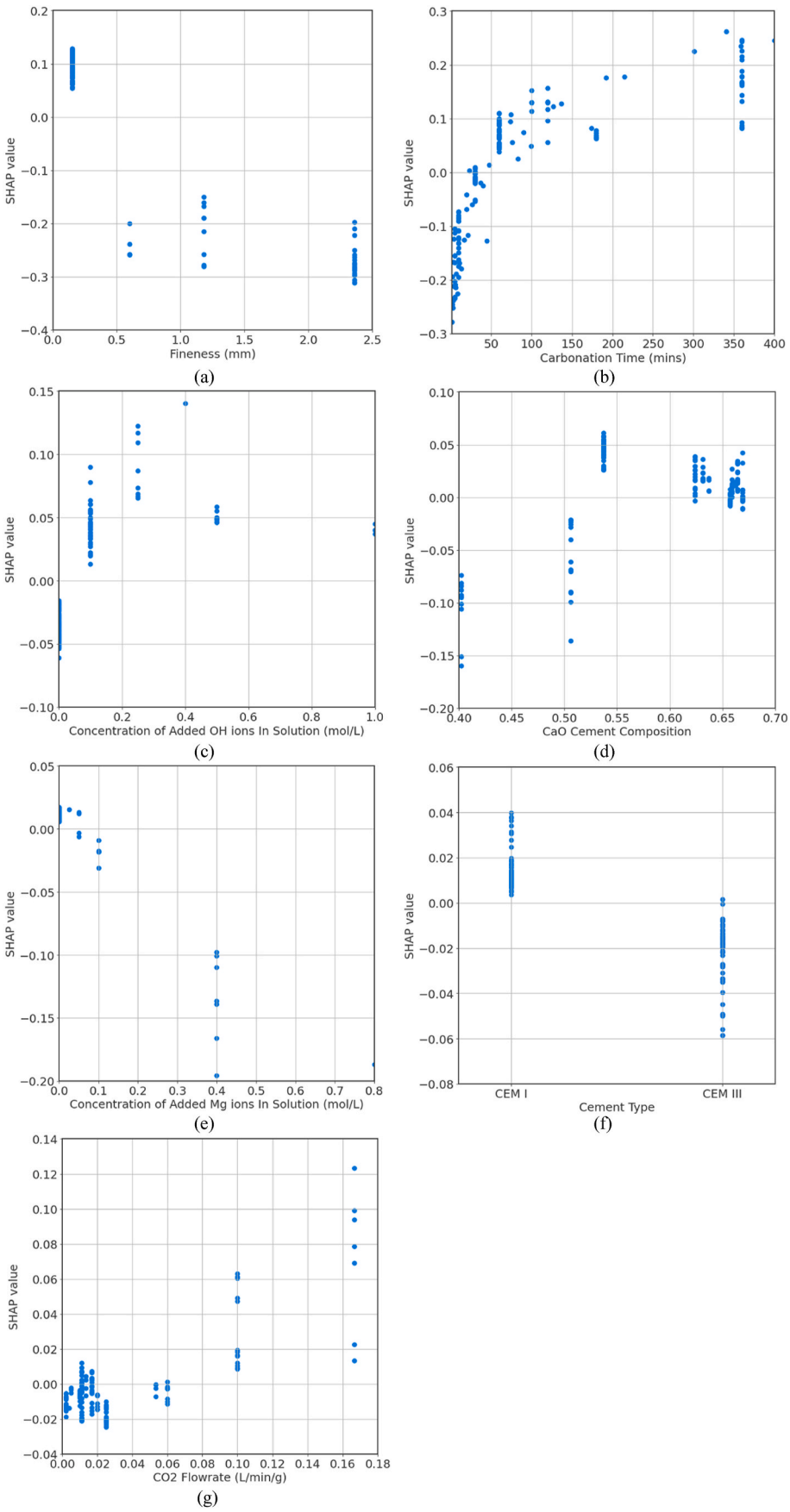

Fig. 8. SHAP values of each feature for calcium carbonate yield for wet carbonation: (a) Fineness, (b) Carbonation time, (c) Concentrations of added OH^- ions in solution, (d) CaO cement composition, (e) Concentrations of added Mg ions in solution, (f) Cement type (CEM I = 1, CEM III = 3), (g) CO_2 flowrate.

increases do not enhance calcium carbonate yield, as the system already has sufficient CO_2 for complete carbonation.

The SHAP analysis reveals the complex interdependence of several key factors influencing calcium carbonate yield in the carbonation process. To maximise yield, it is essential to employ finer particles, and to provide sufficient carbonation time to allow full reaction progress. A high CaO content in the material is crucial for promoting efficient carbonate formation, while maintaining an optimal OH^- ion concentration in the solution further facilitates this reaction. Conversely, minimising Mg ions is recommended, as their presence may inhibit calcium carbonate formation. Additionally, the use of RCP containing CEM I cement, with its high CaO availability, enhances carbonation efficiency. Together, these factors contribute to achieving an optimised calcium carbonate yield in the carbonation process.

4.3. Recommendations for future research

To further advance the applicability and reliability of ML models in predicting calcium carbonate yield under wet carbonation conditions, several areas warrant further investigation:

Expanding data scope. This study employs 227 data points, but their scope and variability are limited, restricting ML modelling of calcium carbonate yield under diverse wet carbonation conditions. Increasing the dataset size and incorporating a broader range of influencing factors, such as pH levels and CO_2 pressures, would broaden the scope of ML in this topic.

Experimental validation of ML predictions. Future research could experimentally validate ML model predictions to improve their reliability, by conducting controlled carbonation experiments and comparing measured outcomes with predicted values.

Industrial implementation and process optimisation. Applying the ML modelling approach developed here to industrial data can facilitate process optimisation and prediction, for example in real-time monitoring and process control of key parameters in carbonation reactors, improving process efficiency and reducing risk.

5. Conclusions

This study evaluated the performance of six ML models, including traditional algorithms (SVR, DT) and advanced ensemble methods (RF, GBR, XGBoost, and stacking), to predict the calcium carbonate yield in wet carbonation processes. The key findings are summarised as follows:

- The stacking ensemble model achieved the highest predictive accuracy, with an R^2 of 0.997 for training and 0.948 for testing, indicating its excellent capability to model calcium carbonate yield. It also exhibited the lowest error metrics, with MAE values of 0.0101 (training) and 0.0393 (testing), and RMSE values of 0.0125 (training) and 0.0536 (testing), demonstrating its superior precision and generalisation ability. Compared to XGBoost, the best-performing individual model, stacking improved test R^2 by 4.9 %, confirming its enhanced predictive capability. This improvement is attributed to the ability to capture complex nonlinear relationships between input variables and calcium carbonate yield, making it an effective tool for modeling wet carbonation processes.
- Among the individual models, XGBoost exhibited the highest predictive performance (R^2 is 0.923 in training, 0.904 in testing), followed by GBR and RF. Ensemble learning methods (RF, GBR, and XGBoost) consistently outperformed individual models, demonstrating improved accuracy, reduced overfitting, and enhanced generalisation. These results highlight the effectiveness of ensemble learning in capturing complex relationships within the dataset, making it particularly advantageous for modeling problems involving limited and highly variable data.
- SHAP analysis identified cement fineness (mean SHAP value = 0.14), carbonation time (mean SHAP value = 0.11), and added OH^- ion concentration (mean SHAP value = 0.04) as the most influential factors in determining calcium carbonate yield. These findings indicate that finer particle sizes, extended carbonation times, and optimised chemical conditions significantly enhance carbonation efficiency and CO_2 sequestration. Other factors, including carbonation temperature (mean SHAP value = 0.02), added Mg^{2+} concentration (mean SHAP value = 0.02), and cement type (mean SHAP value = 0.02), had minor effects. CO_2 concentration and liquid-to-solid ratio were the least significant, indicating that carbonation yield is primarily limited by calcium availability rather than CO_2 supply at the CO_2 concentrations studied (10–100 vol%).

In summary, the application of ensemble ML models, particularly the stacking approach, was proven effective in predicting calcium carbonate yield during wet carbonation, offering a useful tool for advancing cement production with lower environmental impact.

CRediT authorship contribution statement

Yining Li: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Ehecatl Antonio del Rio Chanona:** Writing – review & editing, Supervision. **Hong S. Wong:** Writing – review & editing, Supervision. **Rupert J. Myers:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding provided by Imperial College London is gratefully acknowledged. The research leading to this publication benefitted from funding through EPSRC grant EP/Y001370/1 (selected by the European Research Council (ERC) for a Starting Grant, funded by the EPSRC Frontier Research Guarantee).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2025.145727.

Data availability

All of the data used in the paper have been uploaded as electronic supporting information.

References

Abdellatief, M., Wong, L.S., Din, N.M., Mo, K.H., Ahmed, A.N., El-Shafie, A., 2024. Evaluating enhanced predictive modeling of foam concrete compressive strength using artificial intelligence algorithms. Mater. Today Commun. 40, 110022. https://doi.org/10.1016/j.mtcomm.2024.110022.

Abdellatief, M., Wong, L.S., Din, N.M., Ahmed, A.N., Hassan, A.M., Ibrahim, Z., Murali, G., Mo, K.H., El-Shafie, A., 2025. Sustainable foam glass property prediction using machine learning: a comprehensive comparison of predictive methods and techniques. Results Eng. 25, 104089. https://doi.org/10.1016/j.rineng.2025.104089.

Asteris, P.G., Skentou, A.D., Bardhan, A., Samui, P., Pilakoutas, K., 2021. Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. Cement Concr. Res. 145, 106449. https://doi.org/10.1016/j.cemconres.2021.106449.

Cai, R., Han, T., Liao, W., Huang, J., Li, D., Kumar, A., Ma, H., 2020. Prediction of surface chloride concentration of marine concrete using ensemble machine learning. Cement Concr. Res. 136, 106164. https://doi.org/10.1016/j.cemconres.2020.106164.

Cantero, B., Bravo, M., De Brito, J., Del Bosque, I.S., Medina, C., 2021. Water transport and shrinkage in concrete made with ground recycled concrete-additioned cement

and mixed recycled aggregate. Cem. Concr. Compos. 118, 103957. https://doi.org/10.1016/j.cemconcomp.2021.103957.
Chang, R., Kim, S., Lee, S., Choi, S., Kim, M., Park, Y., 2017. Calcium carbonate precipitation for CO_2 storage and utilization: a review of the carbonate crystallization and polymorphism. Front. Energy Res. 5, 17. https://doi.org/10.3389/fenrg.2017.00017.
Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. Proce.e 22nd acm sigkdd Intern. Confer. Knowl. Disc. Data Mining 785–794. https://doi.org/10.1145/2939672.2939785.
Cui, S., Yin, Y., Wang, D., Li, Z., Wang, Y., 2021. A stacking-based ensemble learning method for earthquake casualty prediction. Appl. Soft Comput. 101, 107038. https://doi.org/10.1016/j.asoc.2020.107038.
Driver, J.G., Bernard, E., Patrizio, P., Fennell, P.S., Scrivener, K., Myers, R.J., 2024. Global decarbonization potential of CO_2 mineralization in concrete materials. Proc. Natl. Acad. Sci. USA 121, e2313475121. https://doi.org/10.1073/pnas.2313475121.
Fushiki, T., 2011. Estimation of prediction error by using K-fold cross-validation. Stat. Comput. 21, 137–146. https://doi.org/10.1007/s11222-009-9153-8.
Ganaie, M.A., Hu, M., Malik, A.K., Tanveer, M., Suganthan, P.N., 2022. Ensemble deep learning: a review. Eng. Appl. Artif. Intell. 115, 105151. https://doi.org/10.1016/j.engappai.2022.105151.
Gomaa, E., Han, T., ElGawady, M., Huang, J., Kumar, A., 2021. Machine learning to predict properties of fresh and hardened alkali-activated concrete. Cem. Concr. Compos. 115, 103863. https://doi.org/10.1016/j.cemconcomp.2020.103863.
Han, T., Siddique, A., Khayat, K., Huang, J., Kumar, A., 2020. An ensemble machine learning approach for prediction and optimization of modulus of elasticity of recycled aggregate concrete. Constr. Build. Mater. 244, 118271. https://doi.org/10.1016/j.jclepro.2020.120665.
He, Z.H., Han, X.D., Zhang, M.Y., Yuan, Q., Shi, J.Y., Zhan, P.M., 2022. A novel development of green UHPC containing waste concrete powder derived from construction and demolition waste. Powder Technol. 398, 117075. https://doi.org/10.1016/j.powtec.2021.117075.
Jiang, Y., Li, L., Lu, J.-x., Shen, P., Ling, T.-C., Poon, C.S., 2022. Mechanism of carbonating recycled concrete fines in aqueous environment: the particle size effect. Cem. Concr. Compos. 133, 104655. https://doi.org/10.1016/j.cemconcomp.2022.104655.
Jiang, Y., Peng, L., Ma, Z., Lu, J.-x., Shen, P., Poon, C.S., 2023. Enhancing the treatment efficiency of recycled concrete fines with aqueous carbonation. Cement Concr. Res. 174, 107338. https://doi.org/10.1016/j.cemconres.2023.107338.
Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. Science. 349, 255–260. https://doi.org/10.1126/science.aaa8415.
Keshk, M., Koroniotis, N., Pham, N., Moustafa, N., Turnbull, B., Zomaya, A.Y., 2023. An explainable deep learning-enabled intrusion detection framework in IoT networks. Inf. Sci. 639, 119000. https://doi.org/10.1016/j.ins.2023.119000.
Kim, J., Jang, H., 2022. Closed-loop recycling of C&D waste: mechanical properties of concrete with the repeatedly recycled C&D powder as partial cement replacement. J. Clean. Prod. 343, 130977. https://doi.org/10.1016/j.jclepro.2022.130977.
Kwon, E., Ahn, J., Cho, B., Park, D., 2015. A study on development of recycled cement made from waste cementitious powder. Constr. Build. Mater. 83, 174–180. https://doi.org/10.1016/j.conbuildmat.2015.02.086.
Le, H.-B., Bui, Q.-B., Tang, L., 2021. Geopolymer recycled aggregate concrete: from experiments to empirical models. Materials 14, 1180. https://doi.org/10.3390/ma14051180.
Li, Q., Song, Z., 2023. Prediction of compressive strength of rice husk ash concrete based on stacking ensemble learning model. J. Clean. Prod. 382, 135279. https://doi.org/10.1016/j.jclepro.2022.135279.
Li, S., Gao, J., Li, Q., Zhao, X., 2021. Investigation of using recycled powder from the preparation of recycled aggregate as a supplementary cementitious material. Constr. Build. Mater. 267, 120976. https://doi.org/10.1016/j.conbuildmat.2020.120976.
Li, Z., Yoon, J., Zhang, R., Rajabipour, F., Srubar Iii, W.V., Dabo, I., Radlińska, A., 2022. Machine learning in concrete science: applications, challenges, and best practices. npj Comput. Mater. 8, 127. https://doi.org/10.1038/s41524-022-00810-x.
Liendo, F., Arduino, M., Deorsola, F.A., Bensaid, S., 2022. Factors controlling and influencing polymorphism, morphology and size of calcium carbonate synthesized through the carbonation route: a review, powder. Tech 398, 117050. https://doi.org/10.1016/j.powtec.2021.117050.
Liu, D., Yates, M.Z., 2006. Formation of rod-shaped calcite crystals by microemulsion-based synthesis. Langmuir 22, 5566–5569. https://doi.org/10.1021/la060612i.
Liu, S., Shen, P., Xuan, D., Li, L., Sojobi, A., Zhan, B., Poon, C.S., 2021. A comparison of liquid-solid and gas-solid accelerated carbonation for enhancement of recycled concrete aggregate. Cem. Concr. Compos. 118, 103988. https://doi.org/10.1016/j.cemconcomp.2021.103988.
Lu, B., Shi, C., Zhang, J., Wang, J., 2018. Effects of carbonated hardened cement paste powder on hydration and microstructure of Portland cement. Constr. Build. Mater. 186, 699–708. https://doi.org/10.1016/j.conbuildmat.2018.07.159.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30. https://doi.org/10.48550/arXiv.1705.07874.
Lyngdoh, G.A., Zaki, M., Krishnan, N.M.A., Das, S., 2022. Prediction of concrete strengths enabled by missing data imputation and interpretable machine learning. Cem. Concr. Compos. 128, 104414. https://doi.org/10.1016/j.cemconcomp.2022.104414.
Mao, Y., He, P., Drissi, S., Zhang, J., Hu, X., Shi, C., 2023. Effect of conditions on wet carbonation products of recycled cement paste powder. Cem. Concr. Compos. 144, 105307. https://doi.org/10.1016/j.cemconcomp.2023.105307.
Mao, Y., Drissi, S., He, P., Hu, X., Zhang, J., Shi, C., 2024. Quantifying the effects of wet carbonated recycled cement paste powder on the properties of cement paste. Cement Concr. Res. 175, 107381. https://doi.org/10.1016/j.cemconres.2023.107381.
Mehdizadeh, H., Ling, T.-C., Cheng, X., Mo, K.H., 2021. Effect of particle size and CO_2 treatment of waste cement powder on properties of cement paste. Can. J. Civ. Eng. 48, 522–531. https://doi.org/10.1139/cjce-2019-0574.
O.A. Montesinos López, A. Montesinos López, J. Crossa, Support Vector Machines and Support Vector Regression, Multivariate Statistical Machine Learning Methods for Genomic Prediction, Springer2022, pp. 337-378.
Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. Front. Neurorob. 7, 21. https://doi.org/10.3389/fnbot.2013.00021.
Nguyen, H., Vu, T., Vo, T.P., Thai, H.-T., 2021. Efficient machine learning models for prediction of concrete strengths. Constr. Build. Mater. 266, 120950. https://doi.org/10.1016/j.conbuildmat.2020.120950.
Nilsen, V., Pham, L.T., Hibbard, M., Klager, A., Cramer, S.M., Morgan, D., 2019. Prediction of concrete coefficient of thermal expansion and other properties using machine learning. Constr. Build. Mater. 220, 587–595. https://doi.org/10.1016/j.conbuildmat.2019.05.006.
Oh, D., Noguchi, T., Kitagaki, R., Choi, H., 2021. Proposal of demolished concrete recycling system based on performance evaluation of inorganic building materials manufactured from waste concrete powder. Renew. Sustain. Energy Rev. 135, 110147. https://doi.org/10.1016/j.rser.2020.110147.
Pekel, E., 2020. Estimation of soil moisture using decision tree regression. Theor. Appl. Climatol. 139, 1111–1119. https://doi.org/10.1007/s00704-019-03048-8.
Ribeiro, M.H.D.M., dos Santos Coelho, L., 2020. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. Appl. Soft Comput. 86, 105837. https://doi.org/10.1016/j.asoc.2019.105837.
Rocha, J.H.A., Toledo Filho, R.D., 2023. The utilization of recycled concrete powder as supplementary cementitious material in cement-based materials: a systematic literature review. J. Build. Eng., 107319 https://doi.org/10.1016/j.jobe.2023.107319.
Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M., 2015. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol. Rev. 71, 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001.
Sagi, O., Rokach, L., 2018. Ensemble learning: a survey. WIREs Data Min. Knowl. Discovery 8, e1249. https://doi.org/10.1002/widm.1249.
Schoon, J., De Buysser, K., Van Driessche, I., De Belie, N., 2015. Fines extracted from recycled concrete as alternative raw material for Portland cement clinker production. Cem. Concr. Compos. 58, 70–80. https://doi.org/10.1016/j.cemconcomp.2015.01.003.
Shapley, L.S., 1953. A Value for n-person Games.
Shen, P., Zhang, Y., Jiang, Y., Zhan, B., Lu, J., Zhang, S., Xuan, D., Poon, C.S., 2022a. Phase assemblance evolution during wet carbonation of recycled concrete fines. Cement Concr. Res. 154, 106733. https://doi.org/10.1016/j.cemconres.2022.106733.
Shen, P., Lu, J., Zhang, Y., Jiang, Y., Zhang, S., Poon, C.S., 2022b. Preparation aragonite whisker-rich materials by wet carbonation of cement: towards yielding micro-fiber reinforced cement and sequestrating CO_2. Cement Concr. Res. 159, 106891. https://doi.org/10.1016/j.cemconres.2022.106891.
Shen, P., Jiang, Y., Zhang, Y., Liu, S., Xuan, D., Lu, J., Zhang, S., Poon, C.S., 2023a. Production of aragonite whiskers by carbonation of fine recycled concrete wastes: an alternative pathway for efficient CO_2 sequestration. Renew. Sustain. Energy Rev. 173, 113079. https://doi.org/10.1016/j.rser.2022.113079.
Shen, P., Gu, Z., Lu, J., Zhang, Y., Jiang, Y., Xuan, D., Zhang, S., Poon, C.S., 2023b. Preparation of reactive urchin-like recycled concrete aggregate by wet carbonation: towards improving the bonding capability of the interfacial transition zone in recycled aggregate concrete. Cem. Concr. Compos. 143, 105235. https://doi.org/10.1016/j.cemconcomp.2023.105235.
Sun, C., Chen, L., Xiao, J., Singh, A., Zeng, J., 2021. Compound utilization of construction and industrial waste as cementitious recycled powder in mortar. Resour. Conserv. Recycl. 170, 105561. https://doi.org/10.1016/j.resconrec.2021.105561.
Tang, Q., Ma, Z., Wu, H., Wang, W., 2020. The utilization of eco-friendly recycled powder from concrete and brick waste in new concrete: a critical review. Cem. Concr. Compos. 114, 103807. https://doi.org/10.1016/j.cemconcomp.2020.103807.
Teschendorff, A.E., 2019. Avoiding common pitfalls in machine learning omic data science. Nat. Mater. 18, 422–427. https://doi.org/10.1038/s41563-018-0241-z.
Teune, I., Schollbach, K., Florea, M., Brouwers, H., 2023. Carbonation of hydrated cement: the impact of carbonation conditions on CO_2 sequestration, phase formation, and reactivity. J. Build. Eng. 79, 107785. https://doi.org/10.1016/j.jobe.2023.107785.
Wang, J., Mu, M., Liu, Y., 2018. Recycled cement. Constr. Build. Mater. 190, 1124–1132. https://doi.org/10.1016/j.conbuildmat.2018.09.181.
You, X., Hu, X., He, P., Liu, J., Shi, C., 2022. A review on the modelling of carbonation of hardened and fresh cement-based materials. Cem. Concr. Compos. 125, 104315. https://doi.org/10.1016/j.cemconcomp.2021.104315.
Zajac, M., Skibsted, J., Durdzinski, P., Bullerjahn, F., Skocek, J., Haha, M.B., 2020a. Kinetics of enforced carbonation of cement paste. Cement Concr. Res. 131, 106013. https://doi.org/10.1016/j.cemconres.2020.106013.
Zajac, M., Skibsted, J., Skocek, J., Durdzinski, P., Bullerjahn, F., Haha, M.B., 2020b. Phase assemblage and microstructure of cement paste subjected to enforced, wet carbonation. Cement Concr. Res. 130, 105990. https://doi.org/10.1016/j.cemconres.2020.105990.

Zajac, M., Skibsted, J., Ben Haha, M., 2021a. Effect of alkalis on enforced carbonation of cement paste: mechanism of reaction. J. Am. Ceram. Soc. 104, 1076–1087. https://doi.org/10.1111/jace.17481.

Zajac, M., Skibsted, J., Durdzinski, P., Haha, M.B., 2021b. Effect of alkalis on products of enforced carbonation of cement paste. Constr. Build. Mater. 291, 123203. https://doi.org/10.1016/j.conbuildmat.2021.123203.

Zajac, M., Skibsted, J., Lothenbach, B., Bullerjahn, F., Skocek, J., Haha, M.B., 2022a. Effect of sulfate on CO2 binding efficiency of recycled alkaline materials. Cement Concr. Res. 157, 106804. https://doi.org/10.1016/j.cemconres.2022.106804.

Zajac, M., Skibsted, J., Bullerjahn, F., Skocek, J., 2022b. Semi-dry carbonation of recycled concrete paste. J. CO2 Util. 63, 102111. https://doi.org/10.1016/j.jcou.2022.102111.

Zajac, M., Song, J., Skocek, J., Haha, M.B., Skibsted, J., 2023. Composite cements with aqueous and semi-dry carbonated recycled concrete pastes. Constr. Build. Mater. 407, 133362. https://doi.org/10.1016/j.conbuildmat.2023.133362.

Zajac, M., Song, J., Ullrich, P., Skocek, J., Haha, M.B., Skibsted, J., 2024. High early pozzolanic reactivity of alumina-silica gel: a study of the hydration of composite cements with carbonated recycled concrete paste. Cement Concr. Res. 175, 107345. https://doi.org/10.1016/j.cemconres.2023.107345.

Zhang, J., Li, D., Wang, Y., 2020. Toward intelligent construction: prediction of mechanical properties of manufactured-sand concrete using tree-based models. J. Clean. Prod. 258, 120665. https://doi.org/10.1016/j.jclepro.2020.120665.