Cement and Concrete Composites 167 (2026) 106436

Contents lists available at ScienceDirect

Cement and Concrete Composites

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

Natural carbonation of concrete: a data-driven analysis

T.C. Chan, M.H.N. Yio *, H.S. Wong

Department of Civil and Environmental Engineering, Imperial College London, SW7 2AZ, UK

A R T I C L E I N F O

Keywords:
Natural carbonation
Concrete
Durability
Statistical analysis
Machine learning
SHAP

A B S T R A C T

This study presents a data-driven analysis of long-term natural carbonation in concrete, using a newly compiled database comprising 1079 mixes and 8194 carbonation depth measurements over 65 years, representing one of the largest natural carbonation datasets assembled to date. Four different tree-based machine learning models (CatBoost, XGBoost, Random Forest and Decision Tree) were evaluated for predicting carbonation rate (k), with CatBoost emerging as the most effective. Partial dependence and SHAP analyses were used to quantify the relative importance of 13 features influencing k, including binder composition, mix proportion, curing and exposure conditions. The water-to-calcium oxide (w/CaO) ratio emerges as the most important feature, encapsulating the effects of porosity, carbonatable content, and supplementary cementitious material (SCM) type and content. Higher SCM replacement levels increase carbonation, while aggregate-related features show comparatively less pronounced effects. Carbonation environment is a more important feature than curing condition in predicting k. By combining machine learning with interpretable SHAP analysis, the model independently recognised trends consistent with literature findings. This study offers a comprehensive dataset and robustly selected features to facilitate future machine learning-based predictions of concrete carbonation.

1. Introduction

Carbonation of cement-based materials has attracted growing research interest due to its double-edged impact: it causes deterioration by inducing corrosion of embedded reinforcing steel but facilitates carbon capture and storage in the built environment. Many factors can influence carbonation in cementitious materials. Accelerated carbonation in high CO_2 environment is commonly employed to study carbonation mechanism by serving as a proxy for the much slower process of natural carbonation. However, fundamental differences exist between the mechanism that occur in accelerated and that in natural condition, particularly in the carbonation sequence of hydrates [1,2] and the types of carbonation products formed [3,4]. A detailed review of these differences is provided by the RILEM TC 281-CCC [5].

The square-root-time function [6], derived from Fick's first law, is commonly used to determine carbonation rate, represented by the carbonation coefficient (k). This coefficient serves as a single parameter that encapsulates the effects of composition (e.g. binder type, carbonatable content), microstructure, and exposure condition (CO_2 concentration, temperature, carbonation duration) [6,7]. Furthermore, carbonation coefficients derived from measurements in accelerated conditions are often used to estimate those under natural conditions by applying the square root ratio of CO_2 concentration between the two environments [8,9]. However, this simple approach tends to underestimate natural carbonation coefficients, particularly for blended systems containing supplementary cementitious materials (SCM) due to differences in internal water content and pore-blocking effects [10,11]. These critical differences and knowledge gaps underscore the need to independently study, understand, and predict natural carbonation, to accurately assess its long-term durability risk and carbon capture/storage potential.

Recently, the RILEM TC 281-CCC [5] evaluated the influence of a range factors on the carbonation of concrete, including mix design, SCM type and replacement level, curing and carbonation conditions. The study identified the water-to-reactive CaO ratio (w/$CaO_{reactive}$), proposed by Leeman et al. [12,13], as one of the most important factors. This ratio captures the combined influence of porosity (governed by water content) and chemical buffering capacity (determined by the amount of reactive CaO). The study also found that, at similar replacement levels (∼25 %), SCMs such as fly ash and silica fume tend to reduce carbonation resistance more than slag or limestone. The importance of adequate curing is strongly emphasised, particularly for blended concretes, which tend to hydrate more slowly than ordinary Portland cement. The study also noted that elevated CO_2 levels (>3 %) can alter

* Corresponding author.
E-mail address: marcus.yio@ic.ac.uk (M.H.N. Yio).

https://doi.org/10.1016/j.cemconcomp.2025.106436
Received 23 July 2025; Received in revised form 30 October 2025; Accepted 5 December 2025
Available online 5 December 2025

the carbonation mechanisms and products, favouring the formation of metastable vaterite and aragonite over calcite, and potentially resulting in incomplete carbonation of portlandite.

Determining the relative importance of variables influencing carbonation is experimentally challenging, as it is difficult to comprehensively evaluate multiple variables and isolate their individual effects. This challenge is further compounded by the growing variety of new low-carbon binder systems, which exhibit increasingly complex compositions, hydration chemistries, and largely unknown carbonation behaviours. Moreover, natural carbonation experiments often require years if not decades to produce meaningful results. In light of these challenges, machine learning (ML) models trained on existing experimental data offer significant potential to accelerate understanding and prediction of carbonation behaviour across a wide range of binder systems.

ML models, such as neural networks and tree-based algorithms, have shown strong capability in predicting various properties of cementitious materials, including pore solution composition [14], chloride concentration [15] and carbonation depth [16–19]. With the emergence of interpretable ML methods, such as SHapley Additive Explanations (SHAP) and partial dependence plots (PDP), recent studies [20,21] have moved beyond black-box predictions to quantify the contributions of different factors affecting the performance of concrete, thereby offering solutions for optimising mix design and durability. Such approaches are particularly suited to studying natural carbonation of concrete, which is a complex process influenced by multiple interacting factors, including the concrete composition, environmental exposure conditions, and time.

However, the effectiveness of these models is highly dependent on the quality and quantity of the training data. Most published models rely on limited datasets, with input variables often selected from a narrow and sometimes subjective pool, thereby limiting their interpretability and generalisability [22,23]. Furthermore, a common limitation is the lack of validation using truly unseen data, i.e. data collected independently from sources not involved in model development or training [24]. Therefore, to develop effective machine learning models (to predict carbonation for example), a comprehensive database and a robust approach to selecting variables based on their relative importance (to carbonation) are required.

In response to this, the RILEM TC 281-CCC compiled a database of 1044 concrete and mortar mixes sourced from 54 references, of which 26 of them were unpublished research and testing reports [11]. They conducted a statistical analysis of the database to investigate the influence of different variables. The study largely confirmed previous findings [5], notably that w/$CaO_{reactive}$ is a decisive variable in predicting k across a range of binders, aligning with the work of Leeman et al. [11,12]. The analysis also found no significant link between k and either compressive strength or the aggregate-to-binder ratio. Furthermore, they observed that accelerated carbonation tests systematically underestimate k under natural conditions, especially in concretes with high SCM content.

The study further applied probabilistic inference and eXtreme Gradient Boosting (XGBoost) machine learning model, combined with SHAP to evaluate the relative contribution of different variables (hereafter referred to as 'feature(s)' when discussed in the context of machine learning) to carbonation. Carbonation duration was revealed as the most significant factor, followed by CO_2 concentration and compressive strength. However, compressive strength was used as a proxy for water content, while aggregate content and curing condition were excluded from the analysis. Furthermore, accelerated and natural carbonation data were combined, potentially limiting the applicability of the findings to natural carbonation. In-depth evaluation of the effects of varying feature values on carbonation was lacking. Feature collinearity, e.g. simultaneous inclusion of overlapping features such as cement type and SCM type, also compromised the robustness of the model.

To address these limitations, the present study proposes a systematic framework for feature selection and the development of interpretable ('white box') ML models to manage multifactorial correlations within heterogeneous datasets and reduce the risks of feature collinearity. This marks a shift away from traditional accuracy-focused black-box approaches, aiming instead to elucidate the complex relationships between variables and k. To support this, a new database has been compiled, consisting entirely of data from published literature and comprising over 1079 mixes and 8194 carbonation depth measurements. This represents one of the largest natural carbonation datasets assembled to date, spanning 65 years of experimental evidence. It should be noted that by the time the RILEM TC 281-CCC database was published, the present compilation was already near completion, with only four overlapping references. Furthermore, the TC 281-CCC database lacks aggregate-related information, rendering it unsuitable for integration into the current database.

In short, this study has two complementary aims: (i) to demonstrate a methodological framework that integrates a curated literature database with interpretable ML analyses, and (ii) to use this framework to extract mechanistic insights into how different features influence natural carbonation of concrete. Guided by these aims, the research addresses the following questions using ML:

1. Can consistent trends be independently identified from large, heterogeneous literature datasets to enable accurate carbonation prediction?
2. What is the relative importance of different variables influencing natural carbonation?
3. How do these variables quantitatively affect the carbonation process?
4. Which variables should be prioritised to guide future predictions of natural carbonation?

The novelty lies in combining rigorous feature analysis and robust model interpretation, with a focus on model explainability in the context of natural carbonation. Specifically, the distribution of features and potential collinearity issues in datasets were examined (Section 3.2), and post-hoc model interpretation using SHAP analysis was performed to quantify the contribution of individual features to natural carbonation of concrete and to examine how these contributions change as a function of varying feature values (Sections 3.3 and 3.4).

2. Methodology

2.1. Database compilation

A new database for natural carbonation depth was compiled from 66 peer-reviewed journal articles [1,7,25–88] and 7 reports from VTT Technical Research Centre of Finland (VTT) and the UK Building Research Establishment (BRE) [89–95], covering research conducted over 65 years (1958–2022). All of the carbonation depth data was measured using the phenolphthalein spray test, which is the most widely adopted method recognised by national and international standards [96–102]. The reliability and consistency of this method have also been evaluated through round-robin experiments [9,103,104], confirming its suitability for comparative studies. While more sophisticated methods are available for measuring carbonation depth (e.g. thermogravimetric analysis (TGA), X-ray diffraction (XRD), and Fourier-transform infrared spectroscopy (FTIR)), the amount of carbonation depth data obtained using these alternative methods is considerably less compared to the phenolphthalein spray method, which remains the most widely used.

Table 1 summarises the variables (45 in total) recorded in the database, categorised into binder properties, aggregate properties, mix proportions, fresh properties, curing condition and carbonation condition. Most of the variables are well-defined and self-explanatory. Due to limited data availability, only binary systems are considered in the present study. The binder oxide composition was calculated based on Portland cement, SCM type and content and X-ray fluorescence (XRF)

Table 1
Summary of variables recorded in the database. Information in [], () and {} represents the unit, abbreviation, and categories or definitions of each variable, respectively.

Category	Variables
General information	Authors, year of publication, sample ID, system type {paste, mortar, concrete}
Binder properties	SCM type {neat ordinary Portland cement (OPC), class-F fly ash (FA-F), ground granulated blast-furnace slag (GGBS), limestone (LS), silica fume (SF)}, oxide and mineralogical composition [wt.%] Blaine fineness [cm^2/g], Brunauer-Emmett-Teller (BET) specific surface area [m^2/g], specific gravity Initial setting time of Portland cement [min]
Fine and coarse aggregate properties	Type {natural, crushed, recycled or mineralogy of aggregate}, maximum aggregate size [mm], fineness [cm^2/g], specific gravity, bulk density [kg/m^3], water absorption [wt.%]
Mix proportion[a]	Blend type {neat, binary, ternary} Unit content [kg/m^3] of Portland cement, SCM, water, CaO, coarse and fine aggregates Water to binder ratio (w/b), water to cement ratio (w/c), water to CaO ratio (w/CaO), SCM replacement level [wt.%], coarse to fine aggregate ratio (coarse/fine agg.), aggregate to binder ratio (agg./b), aggregate to paste {paste = binder + water} ratio (agg./paste) Type {air-entraining agent, superplasticiser, water reducing agent} and content of admixture [wt.% binder content]
Fresh properties	Slump [mm], flow [mm], air content [%] and density [kg/m^3]
Curing condition	Curing method {water, fog, sealed, air}, duration [day], temperature [°C], relative humidity [%]
Carbonation condition	Carbonation environment {indoor, outdoor sheltered, outdoor unsheltered}, duration [day], temperature [°C], relative humidity [%], carbonation depth [mm]
Supplementary information	Sample shape {prism, cylinder} and dimensions [mm]

[a] All mix design ratios are by mass. Fine and coarse aggregates are classified as reported in the literature, with a 4 mm threshold applied when not otherwise specified.

data. The Portland cement mineralogical composition was estimated using the modified Bogue calculation [105]. The unit content of raw materials and mix design ratios (e.g. water to binder (w/b), aggregate to paste (agg./paste), coarse to fine aggregate (coarse/fine agg.)) were calculated from provided information when not explicitly stated. Mix design ratios refer to the relative proportions of constituent materials in the concrete mixture by weight (e.g. water/cement ratio, aggregate to binder ratio, etc.), whereas unit contents refer to the weight of each constituent material per unit volume of concrete in kg/m^3.

$CaO_{reactive}$ is the primary component that hydrates and carbonates, and the water to $CaO_{reactive}$ mass ratio (w/$CaO_{reactive}$) is used to describe the combined effects of mass transport and carbonation buffer capacity of the system [12,13]. However, determining $CaO_{reactive}$ is challenging because the information required (i.e. initial limestone content, gypsum, unreacted clinker, and unreacted SCMs) [5] is not provided in most literature. Moreover, SCMs can contain small amounts of inert Ca-bearing phases (e.g. merwinite in GGBS or anorthite in FA-F) that further complicate these calculations [5,11]. Therefore, the water to total CaO content of the binder (w/CaO) was used in lieu of w/$CaO_{reactive}$ in this study to avoid making erroneous assumptions. Hereafter, all instances of w/CaO refer to the mass ratio of water to total CaO (i.e., all CaO in the system), whereas w/$CaO_{reactive}$ refers to the mass ratio of water to the portion of CaO that participates in hydration and carbonation. It will be shown later in Section 3.3.1 that using total CaO instead of $CaO_{reactive}$ does not significantly impact carbonation predictions.

Curing condition was classified as water, fog, sealed, or air curing, while carbonation environment was categorised as indoor, outdoor sheltered or outdoor unsheltered [13,106]. For indoor exposure, CO_2 concentration, temperature, and relative humidity were recorded. For outdoor exposure, these variables were often unclear or missing, although some studies [27,28,95] included detailed meteorological data, allowing them to be reported as ranges.

Tabular data were extracted directly, while graphical data were digitised using OriginPro [107] software and reported to two decimal places. Missing data were classified as either missing at random (MAR) or structurally missing (SMD). MAR data were labelled 'Null', for example, when the binder oxide composition is not available or the fine aggregate content in a mortar mix is unknown. SMD refers to inherently non-existent data and were labelled 'N/A', such as coarse aggregate content for mortars.

Imputation methods were considered to address missing data in the dataset. However, to enhance the reliability of ML predictions and model interpretability, and to mitigate any potential bias introduced by imputation [108], all data entries with MAR and SMD were excluded. This resulted in the removal of approximately 38 % of the data, including paste and mortar systems (which accounted for about 4 % of the total mixes) due to SMD in agg./paste and coarse/fine agg. ratios, leaving only concrete systems for further analyses.

As such, the final dataset can be regarded as complete for conventional OPC, SCM-blended binary systems, and concretes, thereby defining the scope of this study yet providing sufficient data quality and quantity to robustly develop and evaluate the proposed model framework. It is noted, however, that this may reduce representativeness for paste and mortar, as well as for more complex ternary and newer binder systems. The refined dataset contains over 1000 entries, which remains appropriately sized for tree-based models used in this study [109]. The exclusion of missing data entries ensures that the models were trained and tested on complete datasets, thereby enhancing the robustness and validity of the outcomes [110].

2.2. Data preprocessing

Carbonation depths were converted into carbonation coefficients (k) for each mix subjected to the same curing and carbonation conditions, using the square-root-time function ($d = k\sqrt{t} + d_0$), where d, t and d_0 represent the measured carbonation depth at time t and initial carbonation depth at time zero, respectively. This approach is well-established, widely accepted and has been extensively scrutinised for its use in predicting carbonation in research [111] and is recognised in standards (e.g. EN 12390-12) [112]. This approach standardises the effect of time, allowing k to serve as a consistent metric for comparing different mixes. It enables the study to focus on the influence of binder composition, mix design, curing condition and carbonation environment, without the confounding impact of carbonation duration. More importantly, it addresses data imbalances and reduces bias from studies with disproportionately more carbonation time points than others.

If carbonation depth was measured only at a single time point (t), the carbonation coefficient (k) was calculated by simply dividing the carbonation depth by the square root of the carbonation duration, assuming $d_0 = 0$. For mixes with carbonation depth measured over time, the data were fitted based on the least squares method, accounting for d_0 but without imposing constraints. Although it is physically impossible and meaningless for d_0 to be negative [9], the fitting accuracy improved when d_0 was not constrained to zero (with an average residual sum of squares between actual and predicted carbonation depths of 2.80 ± 12.5 mm vs. 4.49 ± 18.8 mm). Linear regression that produces negative d_0 value suggests errors or inconsistencies in the experiment [113], for example, when samples were not properly pre-conditioned prior to carbonation. Therefore, k obtained with negative d_0 were retained as-is for analysis. Further analysis and discussion on the impact of k with or without d_0 are provided in Sections 3.2 and 4. To ensure reliability, any fitted k values not meeting $R^2 \geq 0.9$ were excluded from further analyses. Furthermore, a quantile-quantile (Q-Q) plot was applied to evaluate the distribution of k before and after filtering of dataset (see Section 3.1), and to identify outliers, thereby preventing extreme (or anomalous) data from distorting the findings.

Categorical variables such as SCM type, curing method, and carbonation environment were encoded (Table 2) to enable numerical analysis and machine learning (Section 2.5). Note that this was applied only to the Decision Tree (DT), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) models, as the Categorical Boosting (CatBoost) model does not require an encoding process due to its inherent support for handling categorical data. Since SCM type has no intrinsic order, one-hot encoding [114] was employed, creating a binary sparse matrix to represent each SCM. Conversely, ordinal encoding [114] was applied to curing method and carbonation environment to preserve their known intrinsic effect on the properties of concrete. These variables were ranked based on their tendency to increase carbonation according to literature findings [38,60,94,115], considering their effects on hydration and microstructure development. An ordinal score of 5 represents the highest favourability for carbonation, while a score of 1 represents the lowest.

2.3. Correlation analysis

Spearman's rank correlation coefficient (ρ) and Grey Relational Analysis (GRA) were employed to examine the relationships between all identified variables and k. The main purpose was to facilitate feature selection and pre-processing by identifying highly correlated features and removing redundant ones for subsequent machine learning (Section 2.5). Spearman's correlation evaluates linear monotonic relationships without any distributional requirements, making it suitable for both continuous and discrete ordinal-encoded variables [116]. In contrast, GRA based on grey system theory [117,118], is particularly effective for capturing nonlinear correlations with implicit (or incomplete) numerical data, and for handling small or poorly informed datasets [119]. The method produces a Grey Relational Grade (GRG) and classifies the degree of correlation as high strength (GRG between 0.7 and 1), moderate (GRG between 0.5 and 0.7) or weak (GRG <0.5).

2.4. Effect quantification

The effect of a particular variable on carbonation depth was quantified using Eq. (1), which is an approach adapted from Ref. [120]. This was obtained by calculating the change in carbonation depth (d) relative to the reference case, assuming no interactions with other variables. This is expressed as an E-value:

$$E_t^{x,j} = \frac{d_t^{x,j} - d_t^{Ref,j}}{d_t^{Ref,j}} \tag{1}$$

Where $d_t^{x,j}$ and $d_t^{Ref,j}$ represent the carbonation depth under the influence of variable x and the carbonation depth of the reference, respectively, at the same carbonation period t in study j. For example, to quantify the effect of SCM replacement level, $d_t^{x,j}$ would be the carbonation depth of a mix at a particular SCM replacement level, while $d_t^{Ref,j}$ would be the carbonation depth of the respective reference (i.e. neat OPC system, SCM replacement level = 0 %). Therefore, a greater E-value (whether positive or negative) indicates that the variable has a stronger effect on carbonation.

2.5. Machine learning

Based on correlation analysis, 13 features were selected and divided into two subsets according to mix design ratios or unit contents of raw materials (Table 3) to mitigate overfitting caused by excessive dimensionality (e.g., w/CaO ratio already encompasses both water and binder contents) and high collinearity (e.g. water and cement contents are linearly correlated with w/b). For the 'mix design ratio' subset, w/b, SCM type and SCM replacement level were used instead of w/CaO when analysing the impact of SCM (Section 3.3.2) to avoid collinearity, as w/CaO encapsulates the combined effects of the three variables. As a result, the 'mix design ratio' subset contained 6 features whereas the 'unit content' subset contained 9 features for ML analysis. Note that curing method, curing duration and carbonation environment were included in both feature sets.

For each feature set, the data was further split into training and validation sets in a 9:1 ratio, using a random seed number of 42 to ensure reproducibility. For model training, K-fold cross-validation (5-fold) was applied, where each fold served as a validation set once, and the remaining folds formed the training set. This approach helps detect overfitting and ensures that the model generalises well to unseen data. In addition, model hyperparameters were tuned using the Optuna package [121], a Bayesian optimisation-based program, to further reduce the risk of overfitting and enhance the prediction accuracy.

Tree-based models were employed to enable reliable and meaningful feature interpretability, particularly through SHAP analysis [122], as described later. Four tree-based machine learning models, i.e. DT, RF, XGBoost and CatBoost, were evaluated. DT was included as a simple baseline, though it is prone to underfitting and noise sensitivity. RF improves generalisation through bagging but may miss rare patterns and

Table 2
One-hot encoding for SCM type and ordinal encoding for curing method and carbonation environment, ranked according to their tendency to increase carbonation. Higher values indicate conditions that favour carbonation.

	One-hot	Ordinal
SCM type		
Neat OPC	(1, 0, 0, 0, 0)	–
FA-F	(0, 1, 0, 0, 0)	–
GGBS	(0, 0, 1, 0, 0)	–
LS	(0, 0, 0, 1, 0)	–
SF	(0, 0, 0, 0, 1)	–
Curing method		
No curing	–	5
Air	–	4
Seal	–	3
Fog	–	2
Water	–	1
Carbonation environment		
Indoor	–	3
Outdoor sheltered	–	2
Outdoor unsheltered	–	1

Table 3
Division of features for machine learning to avoid high collinearity and excessive dimensionality.

Variables	Feature subsets	
	Mix design ratio	Unit content
w/b	✓[a]	
w/CaO	✓	
SCM type	✓[a]	✓
SCM replacement level (wt. %)	✓[a]	✓
Agg./paste	✓	
Coarse/fine agg.	✓	
Unit content of binder (kg/m^3)		✓
Unit content of water (kg/m^3)		✓
Unit content of coarse aggregate (kg/m^3)		✓
Unit content of fine aggregate (kg/m^3)		✓
Curing method	✓	✓
Curing duration (days)	✓	✓
Carbonation environment	✓	✓

[a] w/b, SCM type and SCM replacement level were used instead of w/CaO when analysing the impact of SCM (Section 3.3.2) to avoid collinearity, as w/CaO encapsulates the combined effects of all three variables.

is less interpretable than a single DT. XGBoost applies gradient boosting to iteratively correct errors to achieve higher predictive accuracy, though it is prone to overfitting if not carefully tuned. CatBoost also uses gradient boosting but incorporates ordered boosting, native handling of categorical features, reducing preprocessing and minimising data leakage, albeit with increased computational demand [123]. Unlike the other three models, CatBoost does not require ordinal or one-hot encoding for categorical features [124]. The best performing model for predicting *k*, defined by the lowest mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE) and the highest coefficient of determination (R^2), on the test dataset was selected for further analysis of feature contribution.

PDP [125,126] and SHAP [122,127] analyses were employed to quantify the relative importance of different features. PDP visualises the relationship between a selected feature and the predicted outcome by plotting the average model predictions as the feature varies, while keeping other features constant. SHAP, based on game theory, decomposes model predictions into contributions from each feature, by measuring how much each feature shifts the prediction away from the model's expected value (i.e. the average predicted *k* across the training dataset), thereby quantifying each feature's influence on the final prediction for every data point. These two methods are complementary, with PDP offering a global interpretation by showing the average effect of a feature, whereas SHAP providing local interpretations by allocating contributions of individual data points, thereby enabling identification of specific interactions and non-linear relationships that may be obscured by global averages.

Both PDP and SHAP have already been adopted in other important research fields. They have been used in meteorology to reveal the drivers and effects of seasonal asynchrony [128], in medical research to understand the interplay between the genome and transcriptome in influencing intratumor heterogeneity [129] and in hydrology to assess how human activities affect global patterns of decomposition rates in rivers [130]. Across these applications, they have demonstrated high reliability and strong explanatory power. While alternative methods, such as Gini importance, Local Interpretable Model-agnostic Explanations or Accumulated Local Effects are available, they were not adopted because they are often biased, unstable, or provide only approximate or oversimplified explanations [131–133].

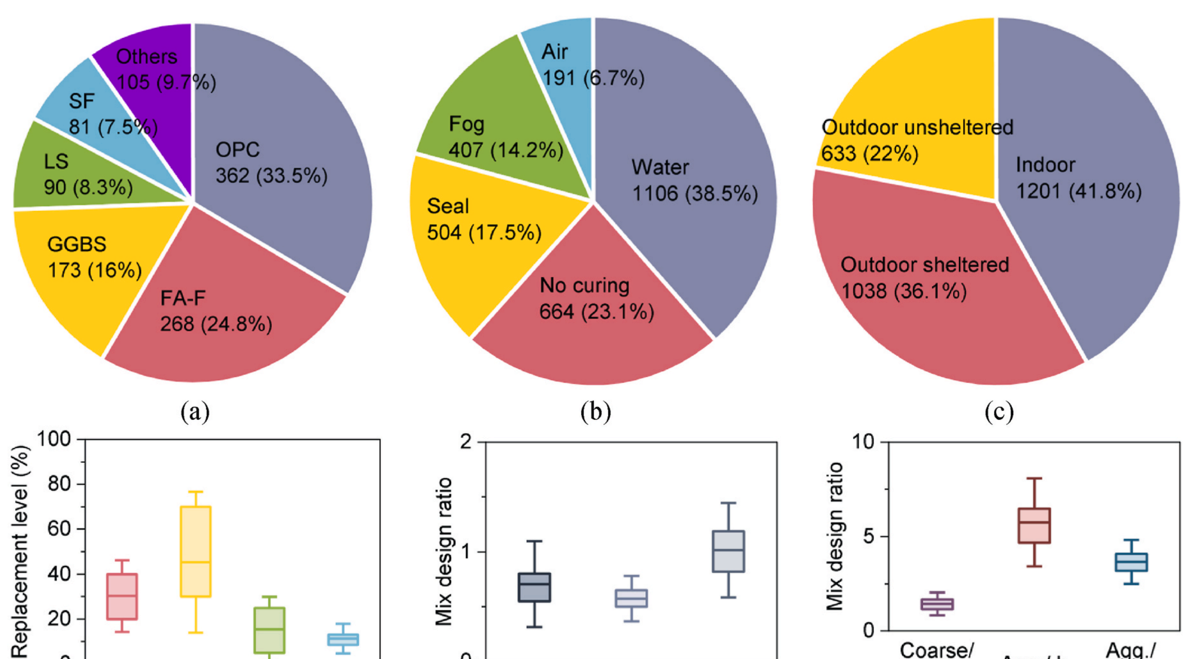
2.6. Computational resources

Python (v. 3.10.14) was used for all computational analyses, including statistical analysis and ML workflows, on an Apple Silicon M3 Max system with 64 GB unified memory in Visual Studio Code. Pandas (v. 2.2.3) [134], was used for dataset handling, while scikit-learn (v.1.5.2) [135] was used for data preprocessing and building machine learning models (DT and RF). XGBoost (v. 2.1.1) [136] and CatBoost (v1.2.8) [124] models were installed from their official repositories. All packages were installed via Python Package Index (PyPI) (v. 25.0.0).

3. Result

3.1. Overview of database

Fig. 1 provides an overview of the database. This consists of 1079 mix designs, from which 8194 carbonation depth measurements and 2872 *k* values were obtained. The distribution of binder types (Fig. 1a) shows that class-F fly ash (FA-F) was the most prevalent SCM, accounting for a quarter of the mixes. This was followed by systems incorporating ground granulated blastfurnace slag (GGBS), limestone fines (LS) and silica fume (SF). Less common SCMs, categorised under



	FA - F	GGBS	LS	SF
Median	30	50	13	10
Mean	30	45	15	11
S.D.	10.6	20.9	9.6	4.4

(d)

	w/c	w/b	w/CaO
Median	0.65	0.55	0.98
Mean	0.71	0.57	1.01
S.D.	0.26	0.14	0.29

(e)

	Coarse/ fine agg.	Agg./ b	Agg./ paste
Median	1.42	5.89	3.75
Mean	1.43	5.75	3.66
S.D.	0.41	1.55	0.78

(f)

Fig. 1. Pie charts showing the number and proportion of (a) binder types based on 1079 mix designs, and (b–c) curing methods and carbonation environments based on 2872 *k* values; box plots showing the distribution of (d) SCM replacement level, and (e–f) key mix design ratios, where the horizontal line inside each box represents the mean and the whiskers extend to 1.5 standard deviations.

"Others", included metakaolin, natural pozzolan, steel slag, rice husk ash, trass, etc.

Fig. 1b–c illustrate the distribution of curing methods and carbonation environments. Water curing was the most common method. Interestingly, nearly a quarter of the 2872 k values were from samples exposed to carbonation immediately after demoulding (no curing), while 42 % underwent indoor carbonation under controlled CO_2 concentration (200–600 ppm), relative humidity (40–95 %) and temperature (17.5–35 °C). The remaining were exposed to outdoor carbonation, either sheltered or unsheltered, with varying relative humidity (55–90 %) and temperature (6–34 °C). The curing and carbonation durations ranged from 0 to 360 days and from 7 days to 54 years, respectively. Preconditioning is an important step in ensuring moisture equilibrium [103,137], which may influence the carbonation process. However, most studies either do not implement this process or fail to specify it.

The distribution of SCM replacement levels (Fig. 1d) shows that most values were within the maximum limits specified by BS EN 197-1 [138], i.e. 55 %, 95 %, 35 % and 10 % for FA-F, GGBS, LS and SF, respectively. Mix design ratios (Fig. 1e–f) were also well-represented and centred around typical values, generally following a normal distribution. To mitigate potential biases from extreme values, outliers exceeding 1.5 times the interquartile range of mix design (~8 % of 1079 mixes) were excluded from subsequent analyses. Details of the mean, standard deviation, skewness and a box plot illustrating the distribution of each numerical variable for the 1079 unique mixes are provided in Table S1.

The maximum carbonation depths ranged from 3 to 46.5 mm. k values ranged from 0.1 to 34.3 mm/year$^{0.5}$, while d_0 averaged 0.31 ± 2.47 mm. Fig. 2a–b shows that k with consideration for initial carbonation depth follows a Weibull distribution (rather than a normal distribution, Fig. S1), while d_0 fits a normal distribution. The Weibull distributions of k before and after data filtering exhibit similar shape and scale parameters, with strong overlap of data points (Fig. S1a), indicating that the filtered datasets remain representative. Outliers identified in the Q-Q plot (red circles in Fig. 2a) were excluded, and the remaining 1960 k values were used for subsequent analyses. Table S2 summarises the criteria for excluding k, including sample type, exposure and curing conditions, and goodness-of-fit. Although this reduction from the initial total of 2872 may seem significant, the retained data is unbiased, ensuring robustness and reliability of the ML model, as well as SHAP and PDP analyses.

3.2. Feature importance

Table 4 presents the correlation between key variables and k computed by Spearman's and GRA methods. The complete results are visualised as heat maps in Figs. S2–S3. Spearman's methods consistently identify strong correlations between water-related mix design ratios (w/CaO, w/c, w/b) and k, followed by binder, carbonation, curing and aggregate-related variables. The relatively low ranks of water content, SCM replacement level (RL) and content are likely due to their non-monotonic relationships with k [139,140].

Table 4
Spearman's rank correlation coefficient (ρ) and Grey Relation Grade (GRG) between key variables and carbonation coefficient (k) ranked according to magnitude in descending order. Negative values in parentheses indicate an inverse relationship.

Rank	ρ	GRG
1	w/CaO (0.56)	w/c (0.81)
2	w/c (0.53)	w/CaO (0.79)
3	Carbonation environment (0.47)	Water content (0.76)
4	w/b (0.45)	SCM RL (0.75)
5	CaO content (−0.43)	SCM content (0.75)
6	Cement content (−0.41)	w/b (0.73)
7	Binder content (−0.39)	Curing duration (0.72)
8	Agg./b (0.37)	Binder content (0.71)
9	Curing duration (−0.35)	CaO content (0.71)
10	Coarse agg. content (0.3)	Agg./b (0.7)
11	Agg./paste (0.29)	Coarse/fine agg. (0.7)
12	Total agg. content (0.27)	Fine agg. content (0.69)
13	Coarse/fine agg. (0.24)	Cement content (0.69)
14	Water content (0.22)	Agg./paste (0.68)
15	Fine agg. content (−0.21)	Coarse agg. content (0.6)
16	Curing method (0.19)	Curing method (0.6)
17	SCM RL (0.14)	Carbonation environment (0.59)
18	SCM content (0.1)	Total agg. content (0.57)

The GRA results indicate that most variables exhibit moderate (GRG between 0.5 and 0.7) to high strength correlation (GRG between 0.7 and 1) with k. Generally, water and binder-related variables (w/c, w/CaO, water content, SCM RL, SCM content, w/b, binder content, CaO content) and curing duration show 'high strength' correlations, with GRG values ranging from 0.71 to 0.81. Aggregate-related variables generally exhibit 'moderate strength' correlations, with GRG values between 0.57 and 0.70. Interestingly, carbonation environment ranks near the bottom. This is due to the limited distinct categories in these encoded variables, which reduces the precision of GRA, when compared to continuous variables [141].

Both methods demonstrate the ability to evaluate variable importance, and they consistently highlight the importance of water-related mix design ratios to k. They also confirm strong intercorrelations (Figs. S2–S3) between (1) w/c, w/CaO and w/b, (2) mix design ratios and the unit content of raw materials, and (3) w/CaO, CaO content, SCM replacement level and content. However, inconsistencies persist for water content, SCM-related variables and carbonation environment.

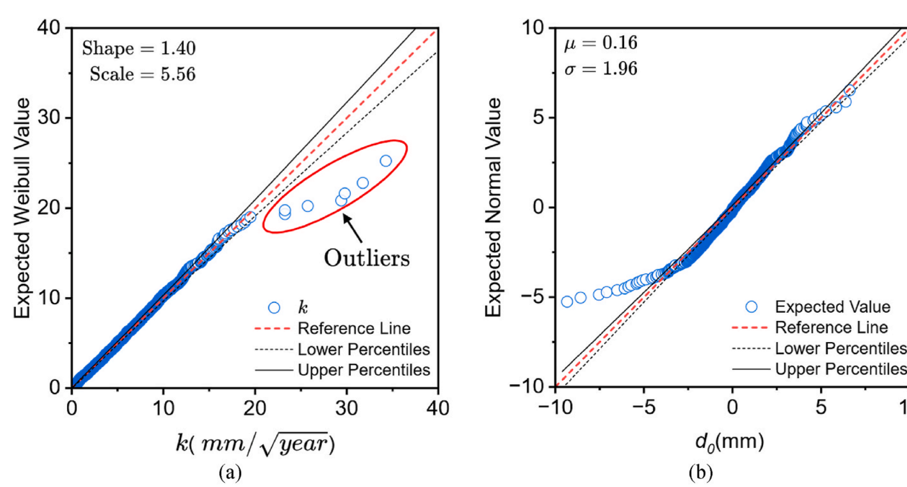


Fig. 2. Q–Q plots to evaluate the distribution of k and d_0 **using (a) Weibull distribution, and (b) normal distribution.**

This supports the need to split the data strategically (Table 3) to mitigate machine learning overfitting caused by excessive dimensionality and high collinearity.

Fig. 3 shows the permutation importance and mean absolute SHAP values, computed by DT, RF, XGBoost and CatBoost for variables in the 'mix design ratio' feature set (Table 3). The rankings consistently show that w/CaO and carbonation environment are the most important features, followed by curing duration and aggregate-related features. Curing method ranks the lowest. $CaO_{reactive}$ and CO_2 concentration have also been identified as two important features contributing to carbonation in the recent study by RILEM TC 281-CCC [11], despite differences in the input features used in their SHAP analysis. Both permutation and SHAP rankings align closely with those of Spearman's, suggesting that these variables are monotonically related to carbonation.

As mentioned in Section 2.2, k can be determined from the square-root-time function either with consideration for initial carbonation depth ($d_0 \neq 0$) or without (i.e. d_0 constrained to 0). The approach with $d_0 \neq 0$ was adopted because it improves linear regression, but this could affect feature importance analysis and therefore merits discussion. Table 5 shows the relative importance of key features determined by SHAP when d_0 is considered versus when it is not. The percentage differences are small (<10 %) in all cases except curing method and duration (~37 %), the effects of which are discussed in Section 3.4.

Overall, the CatBoost model performed the best in predicting k, with the lowest errors and highest R^2 on the test dataset (Table 6), largely due to its ability to handle categorical features through ordered boosting, without requiring explicit encoding. These findings align with those of McElfresh et al. [109], who reported superior performance by CatBoost gradient boosting models across 17 regression datasets with categorical variables. Although RF and XGBoost exhibited slightly higher training accuracy, they showed lower test accuracy compared to CatBoost (Table 6), likely due to data leakage introduced during the ordinal encoding process, which can lead to prediction shifts and reduced generalisation to unseen data [24,124,142]. Based on these findings, the CatBoost model was selected for further analysis of the effects of key features (Sections 3.3 and 3.4).

Table 5
Effect of d_0 on feature importance computed by SHAP analysis for all cases where multiple carbonation depths over time are available.

Feature	Mean (|SHAP|)	
	k with d_0	k without d_0
w/CaO	1.23	1.27 (+3.2 %)
Coarse/fine agg	0.42	0.38 (−9.1 %)
Agg./paste	0.55	0.49 (−11 %)
Curing method	0.18	0.25 (+37 %)
Curing duration	0.77	1.04 (+36 %)
Carbonation environment	1.06	1.01 (−5.1 %)

Table 6
Performance comparison of different tree-based ML models for predicting k.

	DT	RF	XGBoost	CatBoost
Training dataset				
MSE (mm/year$^{0.5}$)	2.58	0.93	0.85	1.12
RMSE (mm/year$^{0.5}$)	1.61	0.96	0.92	1.06
MAE (mm/year$^{0.5}$)	1.15	0.67	0.64	0.74
R^2	0.77	0.92	0.93	0.90
Test dataset				
MSE (mm/year$^{0.5}$)	2.91	3.64	3.15	1.73
RMSE (mm/year$^{0.5}$)	1.98	1.91	1.78	1.32
MAE (mm/year$^{0.5}$)	1.42	1.34	1.24	1.03
R^2	0.64	0.67	0.71	0.85

3.3. *Effect of mix proportion*

3.3.1. *w/b and w/CaO ratios*

Fig. 4a shows a positive linear correlation between w/b ratio and its quantified effect on carbonation depth (E value, Eq. (1)). The reference case here is w/b ratio ~0.5. It is worth noting that this trend is independent of binder type and other variables. The regression analysis indicates that carbonation depth increases by 36.5 % with a 0.1 increase in w/b ratio, in line with the findings of [5]. This trend is consistent with current understanding that higher w/b ratios lead to increased total porosity [1,35,58] and transport properties of concrete, thereby accelerating the carbonation process.

Fig. 4b shows a strong positive correlation between w/b and w/CaO across all binder systems. Blended systems have larger w/CaO than neat OPC systems at the same w/b ratio. This is attributed to the lower CaO content of blended systems, especially at high SCM replacement levels. Therefore, the use of w/CaO instead of w/b effectively captures the influence of two key variables, which controls the total porosity (water) and carbonatable content (CaO). The latter encapsulates the influence of SCM, as it varies with SCM type and replacement level. This highlights w/CaO as a reliable proxy for capturing the combined effects of w/b, SCM type and SCM replacement level in machine learning models.

The significance of $CaO_{reactive}$ content (i.e. CaO involved in hydration and carbonation) was evaluated by comparing the impact of w/CaO and w/$CaO_{reactive}$ on predicted k. This is shown in Fig. 4c for LS-blended systems, which represent a good test case since these systems have the highest contrast between total CaO and $CaO_{reactive}$. The $CaO_{reactive}$ was estimated by subtracting CaO from limestone filler, which is assumed to be inert [11,13]. For example, consider a concrete mix from Huy Vu Q

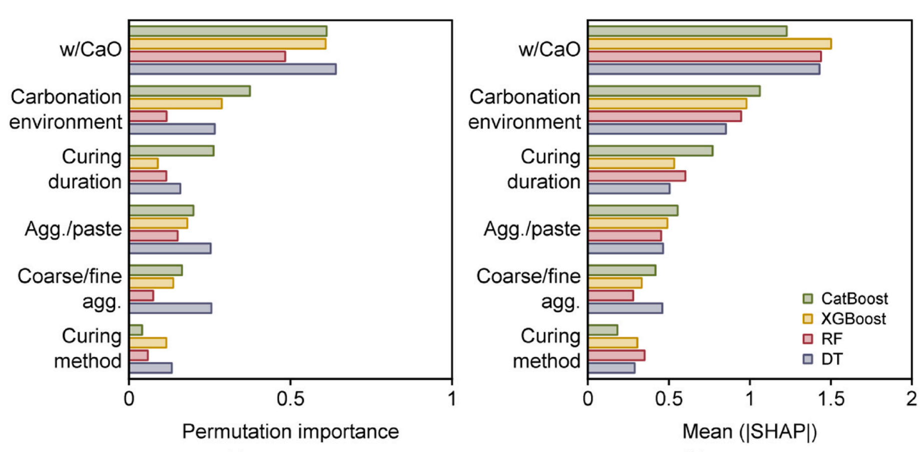


Fig. 3. Feature importance computed by different machine learning models for variables in the 'mix design ratio' feature set (Table 3).

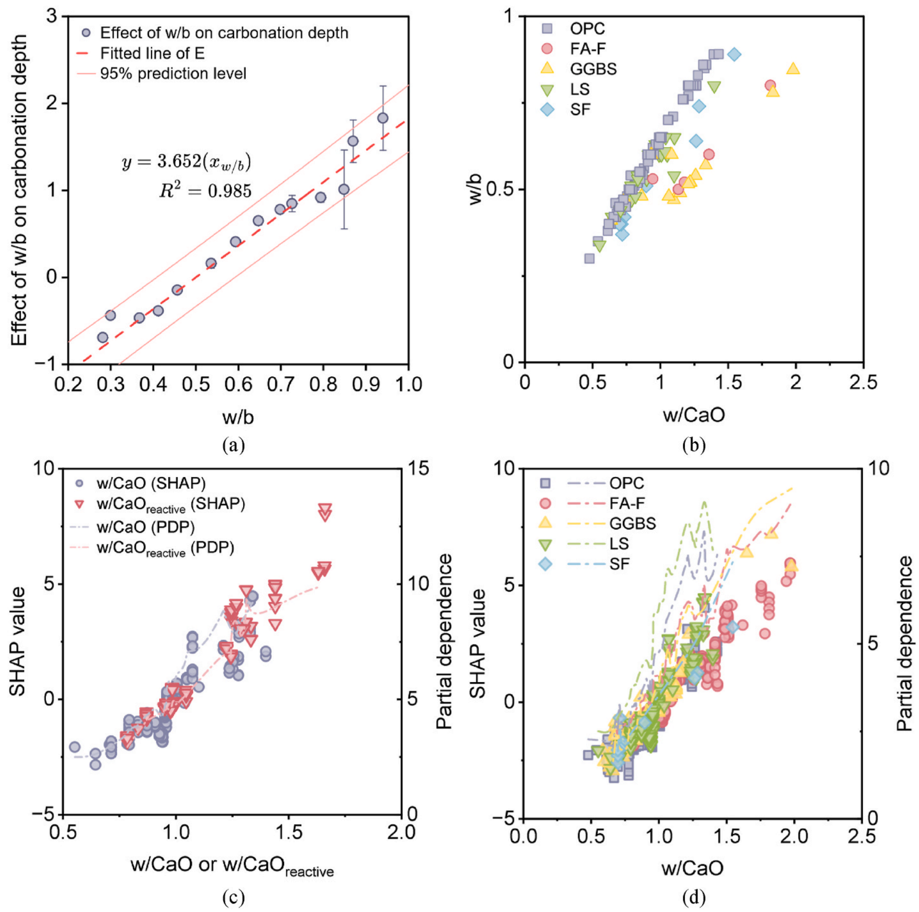


Fig. 4. (a) Quantified effect (based on Eq. (1)) of w/b on carbonation depth using w/b 0.5 as reference; (b) correlation between w/b and w/CaO; (c) impact of w/CaO and w/CaO$_{reactive}$ on predicted k for limestone blended systems by SHAP (scatter points indicating individual contributions) and PDP (dashed lines showing average trends) analyses; (d) impact of w/CaO on predicted k for all binder systems by SHAP (scattered points) and PDP (dashed lines) analyses.

et al. [27] with a binder containing 30 % LS replacement, a total binder content of 280 kg/m^3, and 64.5 % of CaO in the cement, the CaO$_{reactive}$ content is calculated as 280 kg/m^3 × (1–0.3) × 64.5 % = 126 kg/m^3.

Both SHAP and PDP analyses show similar predictive trends using w/CaO and w/CaO$_{reactive}$, suggesting that using total CaO instead of CaO$_{reactive}$ in the denominator does not significantly affect the prediction of k. Using the example above, at a w/b ratio of 0.65, w/CaO equals to 1.07 and w/CaO$_{reactive}$ equals to 1.44. In both cases, increasing water content coupled with reducing CaO content increases carbonation. The impact of w/CaO and w/CaO$_{reactive}$ in systems with higher inert CaO content (>30 %) on k remains unclear and will be revisited as more data become available. Nonetheless, these results demonstrate w/CaO is a robust feature for carbonation prediction in most binder systems compliant with EN-197-1 [138], with LS replacement levels of up to 35 %.

Fig. 4d depicts the impact of w/CaO on k for different binder types, evaluated via SHAP and PDP analyses. All systems, including neat OPC, show a linear proportional impact on k with higher w/CaO. These trends align with the direct correlations between k and w/CaO$_{reactive}$ observed in Ref. [11] for neat OPC, GGBS and FA-F-blended systems. LS-blended systems show a slightly larger impact on k, attributable to increased porosity [143].

3.3.2. SCM type and replacement level

Fig. 5a presents the quantified effect (E value, Eq. (1)) of different SCMs at various replacement levels (grouped according to BS EN197-1 classifications) on carbonation depth. The results show that, irrespective of the SCM type and other variables, carbonation depth increases with higher replacement levels. This increase is primarily due to a reduction in reactive CaO (LS, FA-F) and/or a decrease in portlandite content resulting from pozzolanic (FA-F, SF) or latent hydraulic (GGBS) reactions [5,144].

It is challenging to draw firm conclusions on the significance of different SCM types at replacement levels between 6 and 35 %. This could be due to the wide spread of replacement levels and the limitation of the E-value calculation, relying on single-factor statistical analysis which assumes no interactions between variables [120]. However, it is clear that SF has a significant effect at replacement levels >10 %, attributable to increased pore sizes [12] and micro-cracking induced by autogenous shrinkage [145], while GGBS shows a more pronounced impact than FA-F at replacement levels >36 %.

To further investigate the feature contribution of SCM, both SHAP and PDP analyses (Fig. 5b) were performed using the 'mix design ratio' feature set with w/b, SCM type and SCM RL rather than w/CaO (Table 3). The results confirm that k increases with higher SCM replacement levels. PDP analysis reveals that at the same replacement level of up to 30 %, LS exhibits the greatest impact, likely due to its effect on pore structure and dilution [12], followed by FA-F, GGBS and SF. A direct comparison of the impact of different SCMs on carbonation shows that a binder with 35 % FA-F exhibits a similar k value to a binder with 55 % GGBS, assuming all other variables remain constant. This aligns with the higher portlandite consumption of FA-F compared to GGBS, as shown with thermodynamic modelling, and the increase in k is governed by the rate of the pozzolanic or latent hydraulic reactions [5].

Fig. 5c presents a two-way partial dependence interaction plot between w/b and SCM replacement level for k, irrespective of the SCM

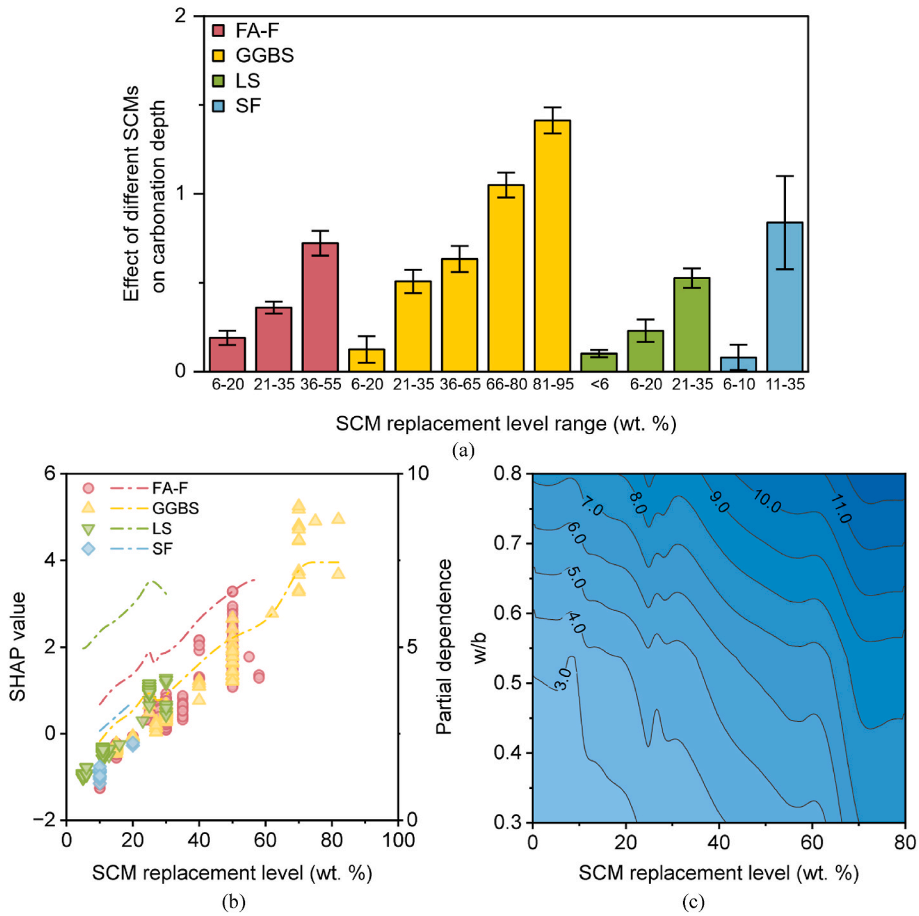


Fig. 5. (a) Quantified effect (based on Eq. (1)) of SCM replacement level on carbonation depth; (b) SHAP (scatter points indicating individual contributions) and PDP (dashed lines showing average trends) analyses of SCM replacement level on predicted k; and (c) two-way partial dependence interaction plot between SCM replacement level and w/b.

type. The plot demonstrates that increasing both w/b and SCM replacement level leads to a greater k. It also illustrates the equivalence between different combinations of w/b and SCM replacement levels in terms of their impact on k. For example, a w/b ratio of 0.5 combined with GGBS replacement level of 50 % results in a k value comparable to a neat OPC system at w/c ratio of 0.75.

3.3.3. Aggregate content

Fig. 6a–b shows the influence of agg./paste and coarse/fine agg. ratios on k. Compared to w/CaO and SCM replacement level, the aggregate-related features have a less pronounced impact on carbonation. Fig. 6a suggests a slight increase in k with higher agg./paste ratio. On one hand, an increase in aggregate fraction reduces the overall pH buffering capacity. On the other hand, higher aggregate content reduces the overall porosity and increases the tortuosity of cement paste, thereby limiting pathways for CO_2 ingress [146]. However, a higher aggregate fraction leads to more interfacial transition zones (ITZ), which can promote microcracking and facilitate CO_2 diffusion, particularly in mixes with high w/b ratio [147,148].

Fig. 6b shows that the coarse/fine agg. ratio (at constant agg./paste ratio) has no consistent impact on k. However, further analysis using the 'unit content' feature set (Fig. 6c) decouples the impact of fine and coarse aggregates on k, revealing a clear increase in k with coarse aggregate content, which becomes more pronounced beyond ~1250 kg/m^3. In contrast, the impact of fine aggregate content on k is less obvious.

Fig. 6d presents the two-way partial dependence interaction plot between w/CaO and agg./paste ratio for k. This shows w/CaO has a much greater impact on k compared to agg./paste. However, when agg./paste exceeds ~4, and is paired with w/CaO above 1, a notable increase in k is observed. This is likely due to the combined effects of high porosity and reduced CO_2 buffering capacity from low paste content and the less tortuous pathways created by high total aggregate content [6, 149].

3.4. Effect of curing and carbonation conditions

Fig. 7a presents the quantified effect of curing method and duration on carbonation depth. As expected, carbonation depth reduces with longer curing duration, regardless of the curing method. Water curing is most effective for limiting carbonation depth, followed by fog and seal curing, irrespective of the binder type. SHAP and partial dependence analyses (Fig. 7b) show that, for water and fog curing, k decreases with increasing curing duration up to 7 days, with no definitive trends observed thereafter due to limited data availability.

As mentioned in Section 3.2, curing influences d_0. To investigate this, the distribution of d_0 values is plotted for different curing methods (all with a curing duration of 28 days) in Fig. 7c. The results show that water, fog and seal curing produce lower d_0 values compared to air and no curing. Negative d_0 values, especially for water and fog curing, suggest that the initiation of carbonation may be delayed due to high internal water content, likely caused by the absence of preconditioning. In contrast, when air (or no curing) is applied, the sample surface is drier, allowing carbonation to initiate earlier, resulting in larger positive d_0 values.

Fig. 8a presents the impact of carbonation environment (indoor, outdoor sheltered, and outdoor unsheltered) based on the E-value, SHAP

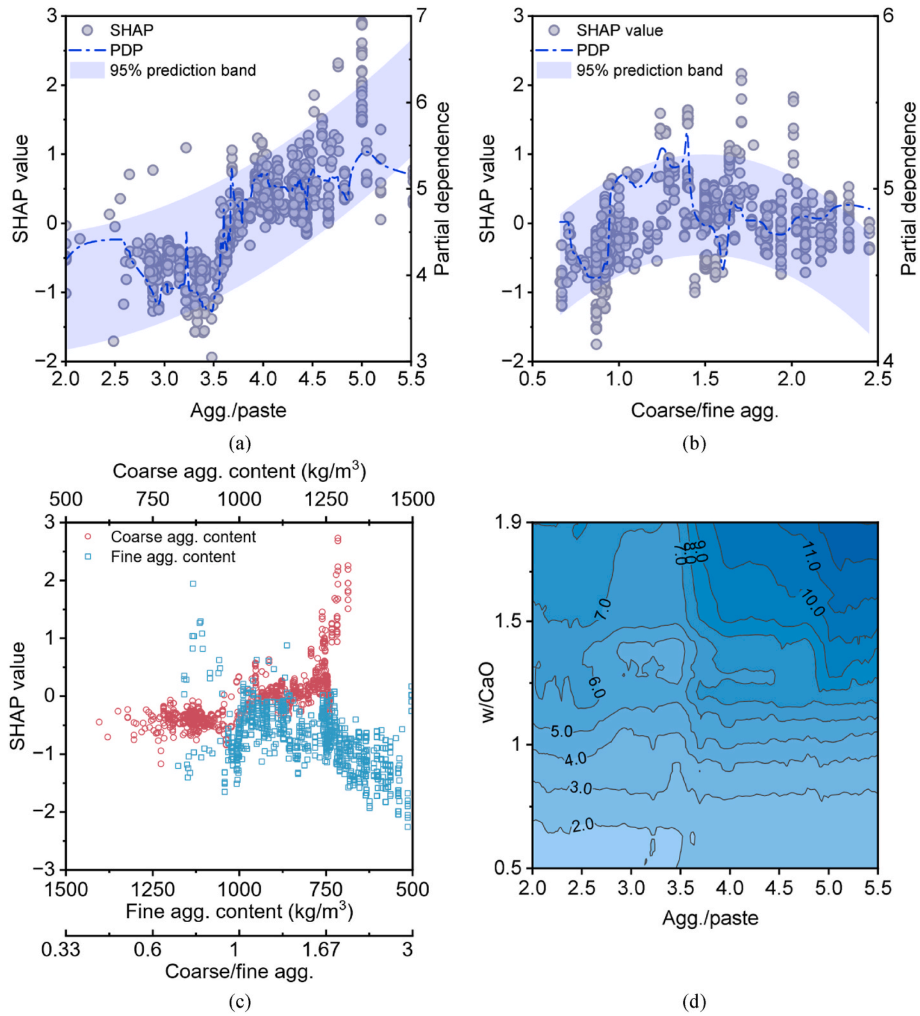


Fig. 6. SHAP (scattered points indicating individual contributions) and PDP (dashed lines showing average trends) analyses of (a) agg./paste and (b) coarse/fine agg. ratio on predicted *k*; (c) SHAP analysis of fine agg. and coarse agg. content on predicted *k*; and (d) two-way partial dependence interaction plot between w/CaO and agg./paste.

and partial dependence analyses. All three methods show consistent trends. Carbonation is most severe under indoor exposure, followed by outdoor sheltered and unsheltered environments. Indoor exposure conditions are well-controlled (e.g. constant relative humidity of 50–65 %) throughout the test [26,94] and set to maximise carbonation. In contrast, outdoor conditions experience fluctuations in relative humidity and temperature, and wet/dry cycles, in particular unsheltered exposure. Carbonation decreases or suspends when specimens become excessively dry or wet [5,26,94].

Fig. 8b shows the two-way partial dependence interaction between curing method and carbonation environment on *k*. Regardless of the curing method, indoor exposure has the largest impact on *k*, followed by outdoor sheltered and outdoor unsheltered conditions. However, within the same carbonation environment, no definitive conclusions can be drawn about the impact of curing method, particularly for outdoor sheltered and unsheltered exposures. Overall, curing method has a lower impact on *k* compared to carbonation environment.

3.5. Comparison with literature findings from RILEM TC 281-CCC

The SHAP values presented in Section 3.3 were converted to *k* values by adding them to the average predicted *k*, enabling estimation of the average factor by which *k* changes as each feature varies within its range. This approach allows for direct comparison with findings from TC 281-CCC [5], which reported the effects on *k* based on single-factor literature analysis. For categorical features, including SCM type, curing method and carbonation condition, predicted *k* values were compared against selected baselines, i.e. a 25 % replacement level for SCM type, a 28-day curing age for curing method and indoor exposure for carbonation condition.

Table 7 shows generally good agreement between the SHAP-based analysis in this study and the single-factor, literature-based analysis from TC 281-CCC. In both studies, w/CaO emerges as the most influential feature. SHAP analysis indicates that *k* increases by a factor of 3.7, broadly aligning with the 5-fold increase reported by TC 281-CCC for w/$CaO_{ractive}$.

However, a discrepancy arises in the ranking of SCM types. SHAP analysis, consistent with the PDP results (Fig. 5b), suggests that LS and FA-F lead to higher *k* than SF and GGBS, whereas TC 281-CCC suggested that SF should rank ahead of LS at a 25 % replacement level. This difference is likely due to the different approaches adopted rather than to data imbalance in the present study. TC 281-CCC's findings are based on qualitative literature review, while the results here are derived from a data-driven analysis using ML and interpretable methods. Some degree of data imbalance is common in real-world datasets of this type, where research on and applications of different SCMs vary depending on local availability. Nevertheless, the CatBoost model applied in this study is relatively robust to such imbalance due to its decision-tree algorithms and embedded techniques, including ordered boosting, effective handling of categorical features (such as SCM type), which help mitigate

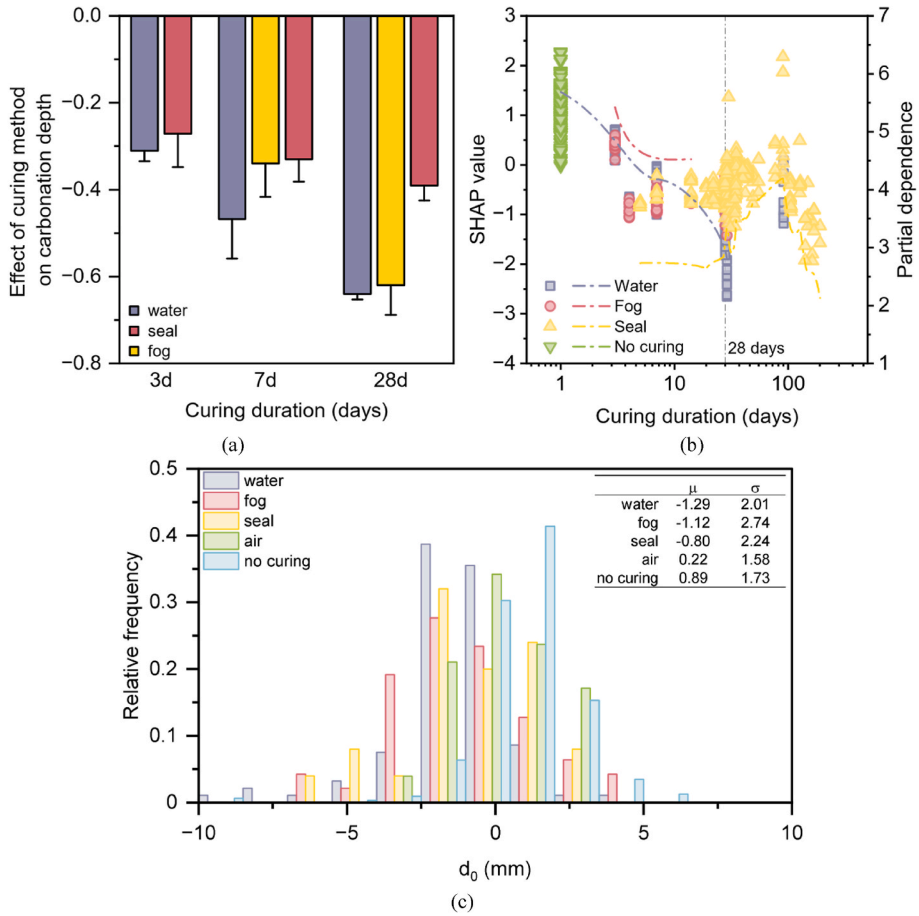


Fig. 7. (a) Quantified effect (based on Eq. (1)) of curing method and duration on carbonation depth; (b) SHAP (scatter points indicating individual contributions) and PDP (dashed lines showing average trends) analyses of curing duration and methods on predicted *k*; and (c) Distribution of d_0 for different curing methods under a 28-day curing duration.

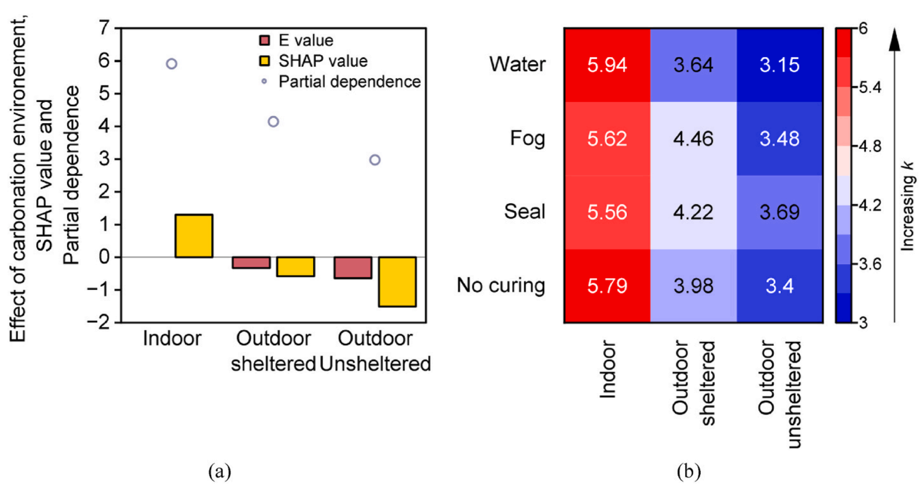


Fig. 8. (a) Quantified effect (based on Eq. (1)) of carbonation environment on carbonation depth, SHAP and PDP analyses on predicted *k*; and (b) two-way partial dependence interaction plot between curing method and carbonation environment.

the risk of bias [4].

TC 281-CCC did not provide a quantitative estimate for the influence of SCM replacement level but qualitatively identified an increasing trend for *k* with higher replacement across all SCM types. SHAP analysis quantifies this effect as a factor of 2.49, confirming that higher replacement levels lead to increased *k*. For aggregate-related features, TC 281-CCC reported a 1.1–1.5-fold increase in *k* with a higher aggregate-to-cement volume ratio, depending on binder type. SHAP analysis identifies a similar effect, with a factor of 1.36 averaged across all binder types. SHAP analysis further shows that increasing coarse/fine agg. from 0.6 to 2.5 results in a slight increase in *k*, by a factor of 1.08.

In terms of curing, both the TC 281-CCC and SHAP analyses confirm that water curing is more effective than air curing or no curing. SHAP analysis further shows that prolonging curing to 28 days reduces *k* by a

Table 7
Influence of key features on *k*, comparing data-driven SHAP analysis in this study with single-factor analyses performed by TC-281 based on literature review.

Feature category	Feature	Observed range	Factor by which *k* changes across the observed range of a feature[a]	
			Single-factor analysis in TC-281[b]	SHAP analysis in this study
Binder composition and mix design	w/CaO	0.5 → 2 (SHAP) 0.4 → 0.9 (TC-281, w/ $CaO_{reactive}$)	5.0	3.7, across all binder types
	Type of SCM (at a replacement level of 25 %)	OPC, FA-F, SF, LS, GGBS	FA-F ≈ SF (2.3) > LS (1.9) > GGBS (1.3) > OPC[c]	LS (1.46) > FA-F (1.38) > SF (1.37) > GGBS (1.37) > OPC[c]
	SCM replacement level	0 → 85 % of OPC	No quantitative conclusion was drawn, but a qualitative increasing trend was identified across all SCM types	2.49, across all SCM types
	Agg./paste	2 → 5.5 (SHAP) 1.7 → 2.3 (TC-281, aggregate-to-cement volume ratio)	1.1 (OPC) – 1.5 (blended with pozzolan or slag)	1.36, across all binder types
	Coarse/fine agg.	0.6 → 2.5	n/a	1.08
Curing	Curing method (with a curing duration of 28 days)		Air > water (28days)	No curing > seal (−0.73) > fog (−0.65) > water (−0.61)[d]
	Curing duration	1 → 28 days	−1.1 to −2.5 (vary depending on SCM type, degree of hydration and curing method)	−0.65 (1 → 28 days), irrespective of curing method and binder type
Carbonation	Carbonation condition		Indoor > outdoor sheltered > outdoor unsheltered	Indoor > outdoor sheltered (−0.69) > outdoor unsheltered (−0.54)[e]

[a] A positive value indicates a contributing effect that increases *k*, whereas a negative value indicates a reducing effect.
[b] Data for all features were from Ref. [5], except for carbonation condition, which was taken from [9].
[c] Values in parentheses are relative to OPC.
[d] Values in parentheses are relative to no curing.
[e] Values in parentheses are relative to indoor.

factor of 0.65, which aligns with TC 281-CCC's reported reduction in *k* by factors ranging from 1.1 to 2.5. For carbonation conditions, both approaches agree on the same ranking, with indoor > sheltered outdoor > unsheltered outdoor. These findings highlight the potential of SHAP-based analysis to validate and complement existing literature. While traditional single-factor analyses consider each parameter in isolation, SHAP provides quantitative estimates that reflect the combined and interacting effects of features, offering a more comprehensive understanding of their influence on *k*.

4. Discussion

The modelling framework presented in this study, based on a large dataset and machine learning analyses, provides quantitative insights into which features are most important and how they influence natural carbonation of concrete. These insights can potentially be used to provide indicative assessments of durability of existing concrete structures, guide future binder selection and mix design, inform or refine design codes, and support the development of new binders, ultimately aiming to achieve more carbonation-resistant and durable concrete structures. However, to fully exploits its practical potential, further work as detailed below is required.

A challenging aspect of this study was the occurrence of missing data in the datasets. To avoid potential bias and preserve model interpretability, all entries with missing data were excluded (Section 2.1), which significantly reduced the dataset. While this approach helped ensure the robustness and validity of the ML predictions for *k*, it may have limited the models' full predictive potential, especially for paste and mortar, as well as more complex ternary and newer binder systems. In future work, systematic imputation techniques, such as regression imputation, distribution-based imputation and k-nearest neighbour imputation [150,151], will be explored to evaluate whether a more complete but imputed dataset can further improve the reliability of carbonation predictions.

Inevitably, some binder systems are less well studied than others. These were recorded in the database but excluded from analyses due to limited data availability. Examples include (1) ternary blended systems, which exhibit complex chemical interactions and synergistic effects on carbonation [152–154], (2) marine-exposed or chloride-contaminated systems, which show altered pH and ion equilibria in the pore solution [155,156], (3) systems containing recycled or coated aggregates, which modifies carbonation capacity and overall porosity [80,157], (4) self-compacting systems with high dosages of water-reducing admixtures [158–161], and (6) cored specimens from in-service structures, where structural loading exacerbates carbonation under tension and vice-versa [162]. Future work should revisit these when their data availability improves.

A potential limitation of this study is that it only considers carbonation depths obtained via phenolphthalein spray. This is a simple operator-dependent test, and its pH-induced colour change may underestimate the true carbonation front. But it is the mostly widely used standardised method. Its reliability has been assessed through round-robin testing [9,103], and confirmed with high-resolution techniques such as Raman microscopy [104]. Studies have also used other methods such as TGA, XRD, and FTIR to profile carbonation depth. Although these techniques provide detailed information on phase and chemical composition, they typically require destructive sampling and preparation (e.g. cutting, grinding, and hydration stoppage) at intervals along the carbonation depth. Such procedures can affect measurement accuracy and may create an artificial transition zone containing both carbonated and uncarbonated material, which, according to previous work [104], does not exist. Furthermore, these methods can produce different carbonation depth values depending on the sampling intervals and interpretation of results, making it difficult to achieve fair comparisons across datasets, unlike the standardised phenolphthalein spray method.

The w/CaO ratio has emerged as the most important feature impacting *k* as it encapsulates the combined effects of porosity, carbonatable content, and indirectly, the influence of SCMs. This feature has potential to predict more complex systems such as ternary blends containing uncommon SCMs. To illustrate this, the CatBoost model trained with the 'mix design ratio' feature set with w/CaO was tested on unseen blended systems containing uncommon SCMs such as metakaolin, class-C fly ash, natural pozzolan, steel slag, rice husk ash, trass and diatomite. The results (Fig. 9) show reasonably good accuracy (R^2: 0.62, RMSE: 1.82 mm/year$^{0.5}$) and indicate the applicability of w/CaO-trained model

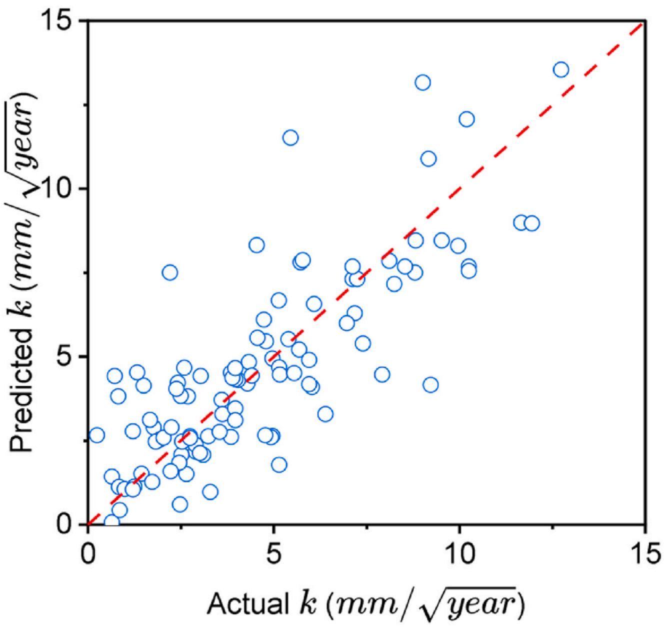


Fig. 9. Predicted vs. actual k values for unseen blended systems containing metakaolin, class-C fly ash, natural pozzolan, steel slag, rice husk ash, trass and diatomite, generated using the CatBoost model trained with the 'mix design ratio' feature set with w/CaO, encompassing OPC, FA-F, GGBS, LS, SF-blended systems.

for carbonation prediction. It is important to note that these are preliminary findings. The scope of the dataset, limited to conventional SCM and binary binders, may reduce representativeness for these less-studied systems. Further studies are ongoing to validate this across a broader range of data.

Overall, this study shows that the relative importance of different variables can be predicted using a data-driven approach, leveraging machine learning methods such as partial dependence and SHAP analyses. Further investigations are required to confirm the influence of aggregate-related variables, including agg./paste and coarse/fine agg. on carbonation. While ordinal encoding has limitations in capturing the precise impact magnitude of categorical data (carbonation environment and curing method) [114], the predicted impact on carbonation is generally consistent with experimental findings in the literature. Future research could explore more advanced methods such as scoring techniques, logit models, or latent variable models [163] to quantify the influence of these categorical features.

5. Conclusions

This study presents a comprehensive analysis of the relative importance of binder composition, mix proportions, curing and carbonation conditions on the natural carbonation of concrete using a data-driven approach. By combining correlational analyses, effect quantification, and ML methods, the study quantitatively evaluates the influence of various features on natural carbonation. To support the investigation, an extensive database has been compiled, comprising 1079 mixes and 8194 carbonation depth measurements collected over a period of 65 years, representing one of the largest natural carbonation datasets assembled to date.

A systematic framework for feature selection and the development of interpretable ML models was established. Trends consistent with literature findings were independently identified across large, heterogeneous literature datasets. The most influential features affecting the natural carbonation of concrete were ranked and prioritised for future prediction models, and their effects were quantified. The key findings are summarised as follows:

a A total of 2872 carbonation coefficient (k) values were obtained using the square-root-time function ($d = k\sqrt{t} + d_0$). Allowing d_0 as a free parameter could improve regression fit, reducing the average residual sum of squares from 4.49 ± 18.8 mm to of 2.80 ± 12.5 mm.

b. Spearman's correlation and GRA indicate strong collinearity among mix design ratios, highlighting the need for careful feature selection and dataset splitting to prevent overfitting and maintain model interpretability.

c. CatBoost outperformed other tree-based models (XGBoost, Random Forest and Decision Tree) in predicting k ($R^2 = 0.85$, RMSE = 1.32 mm/year$^{0.5}$) due to its native handling of categorical features, eliminating the need for explicit encoding and enabling more effective capture of category-based patterns.

d. Data-driven ML combined with explainable SHAP analysis, shows potential to independently capture trends consistent with experimental findings in the literature, supporting the validity of both the constructed database and the predictive model for natural carbonation depth in concrete.

e. The most important features influencing k under natural carbonation based on SHAP analysis, ranked in descending order are (1) w/CaO ratio, (2) carbonation environment, (3) curing duration, (4) agg./paste ratio, (5) coarse/fine agg. ratio, and (6) curing method. Other highly intercorrelated features should be excluded to minimise collinearity.

f. The w/CaO ratio can capture the effects of porosity, carbonatable content, and SCM, with higher w/CaO associated with increasing k. For binders low in inert CaO (e.g. LS < 30 wt.%), total CaO may be used in lieu of reactive CaO to predict carbonation, although this remains uncertain for binders with higher inert CaO content. According to PDP and SHAP analysis, k generally increases with SCM replacement level <25 wt.%, with LS having the greatest observed effect, followed by FA-F, SF, and GGBS. Further investigation is warranted to confirm these.

g. Aggregate-related variables seem to have a less pronounced impact on k, with no clear trends observed for the agg./paste and coarse/fine agg. ratios. However, a closer analysis suggests k may increase with coarse aggregate content, while fine aggregate appears to have a weaker effect. Further investigation is required to confirm these.

h. Carbonation environment appears to have a greater impact on k than curing, with indoor exposure under controlled conditions showing the most noticeable effect. Among curing methods, water curing seems most effective at limiting carbonation, and extended curing durations are generally associated with reducing k.

Overall, the study provides a robust foundation for advancing ML research in concrete durability, supporting the development of transfer-learning models to extend carbonation predictions to novel or multi-SCM binder systems and the integration of other durability factors, thereby broadening model applicability.

CRediT authorship contribution statement

T.C. Chan: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **M.H.N. Yio:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Conceptualization. **H.S. Wong:** Writing – review & editing, Visualization, Supervision, Resources, 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.

Acknowledgement

The research benefitted from EPSRC funding under grant No. EP/R010161/1 and from support from the UKCRIC Coordination Node, EPSRC grant number EP/R017727/1, which funds UKCRIC's ongoing coordination. We also acknowledge the helpful discussion with Mr Xiangdong Geng.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cemconcomp.2025.106436.

Data availability

Data will be made available on request.

References

[1] N. Hyvert, A. Sellier, F. Duprat, P. Rougeau, P. Francisco, Dependency of C–S–H carbonation rate on CO2 pressure to explain transition from accelerated tests to natural carbonation, Cement Concr. Res. 40 (2010) 1582–1589, https://doi.org/10.1016/j.cemconres.2010.06.010.

[2] H. Cui, W. Tang, W. Liu, Z. Dong, F. Xing, Experimental study on effects of CO2 concentrations on concrete carbonation and diffusion mechanisms, Constr. Build. Mater. 93 (2015) 522–527, https://doi.org/10.1016/j.conbuildmat.2015.06.007.

[3] M. Thiery, G. Villain, P. Dangla, G. Platret, Investigation of the carbonation front shape on cementitious materials: effects of the chemical kinetics, Cement Concr. Res. 37 (2007) 1047–1058, https://doi.org/10.1016/j.cemconres.2007.04.002.

[4] M. Auroy, S. Poyet, P. Le Bescop, J.-M. Torrenti, T. Charpentier, M. Moskura, X. Bourbon, Comparison between natural and accelerated carbonation (3% CO2): impact on mineralogy, microstructure, water retention and cracking, Cement Concr. Res. 109 (2018) 64–80, https://doi.org/10.1016/j.cemconres.2018.04.012.

[5] S. von Greve-Dierfeld, B. Lothenbach, A. Vollpracht, B. Wu, B. Huet, C. Andrade, C. Medina, C. Thiel, E. Gruyaert, H. Vanoutrive, I.F. Saéz del Bosque, I. Ignjatovic, J. Elsen, J.L. Provis, K. Scrivener, K.-C. Thienel, K. Sideris, M. Zajac, N. Alderete, Ö. Cizer, P. Van den Heede, R.D. Hooton, S. Kamali-Bernard, S.A. Bernal, Z. Zhao, Z. Shi, N. De Belie, Understanding the carbonation of concrete with supplementary cementitious materials: a critical review by RILEM TC 281-CCC, Mater. Struct. 53 (2020) 136, https://doi.org/10.1617/s11527-020-01558-w.

[6] V.G. Papadakis, C.G. Vayenas, M.N. Fardis, Fundamental modeling and experimental investigation of concrete carbonation, ACI Mater. J. 88 (1991) 363–373, https://doi.org/10.14359/1863.

[7] L. De Ceukelaire, D. Van Nieuwenburg, Accelerated carbonation of a blast-furnace cement concrete, Cement Concr. Res. 23 (1993) 442–452, https://doi.org/10.1016/0008-8846(93)90109-M.

[8] K. Sisomphon, L. Franke, Carbonation rates of concretes containing high volume of pozzolanic materials, Cement Concr. Res. 37 (2007) 1647–1653, https://doi.org/10.1016/j.cemconres.2007.08.014.

[9] H. Vanoutrive, N. Alderete, N. De Belie, M. Etxeberria, C. Grengg, I. Ignjatović, T.-C. Ling, Z. Liu, I. Garcia-Lodeiro, C. Medina Martínez, J. Sanchez, A. Palomo, N. Rebolledo, M. Sakoparnig, K. Sideris, C. Thiel, P. Van den Heede, A. Vollpracht, S. von Greve-Dierfeld, J. Wei, M. Zając, E. Gruyaert, Report of RILEM TC 281-CCC: outcomes of a round robin on the resistance to natural carbonation of Portland, Portland-fly ash and blast-furnace cements and its relation to accelerated carbonation, Mater. Struct. 57 (2024) 209, https://doi.org/10.1617/s11527-024-02464-1.

[10] P. Van den Heede, C. Thiel, N. De Belie, Natural and accelerated carbonation behaviour of high-volume fly ash (HVFA) mortar: effects on internal moisture, microstructure and carbonated phase proportioning, Cement Concr. Compos. 113 (2020) 103713, https://doi.org/10.1016/j.cemconcomp.2020.103713.

[11] A. Vollpracht, G.J.G. Gluth, B. Rogiers, I.D. Uwanuakwa, Q.T. Phung, Y. Villagran Zaccardi, C. Thiel, H. Vanoutrive, J.M. Etcheverry, E. Gruyaert, S. Kamali-Bernard, A. Kanellopoulos, Z. Zhao, I.M. Martins, S. Rathnarajan, N. De Belie, Report of RILEM TC 281-CCC: insights into factors affecting the carbonation rate of concrete with SCMs revealed from data mining and machine learning approaches, Mater. Struct. 57 (2024) 206, https://doi.org/10.1617/s11527-024-02465-0.

[12] A. Leemann, P. Nygaard, J. Kaufmann, R. Loser, Relation between carbonation resistance, mix design and exposure of mortar and concrete, Cement Concr. Compos. 62 (2015) 33–43, https://doi.org/10.1016/j.cemconcomp.2015.04.020.

[13] A. Leemann, F. Moro, Carbonation of concrete: the role of CO2 concentration, relative humidity and CO2 buffer capacity, Mater. Struct. 50 (2017) 30, https://doi.org/10.1617/s11527-016-0917-2.

[14] C. Albert, O.B. Isgor, U. Angst, Exploring machine learning to predict the pore solution composition of hardened cementitious systems, Cement Concr. Res. 162 (2022) 107001, https://doi.org/10.1016/j.cemconres.2022.107001.

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

[16] I. Nunez, M.L. Nehdi, Machine learning prediction of carbonation depth in recycled aggregate concrete incorporating SCMs, Constr. Build. Mater. 287 (2021) 123027, https://doi.org/10.1016/j.conbuildmat.2021.123027.

[17] J. Wu, G. Zhao, M. Wang, Y. Xu, N. Wang, Concrete carbonation depth prediction model based on a gradient-boosting decision tree and different metaheuristic algorithms, Case Stud. Constr. Mater. 21 (2024) e03864, https://doi.org/10.1016/j.cscm.2024.e03864.

[18] Z. Chen, J. Lin, K. Sagoe-Crentsil, W. Duan, Development of hybrid machine learning-based carbonation models with weighting function, Constr. Build. Mater. 321 (2022) 126359, https://doi.org/10.1016/j.conbuildmat.2022.126359.

[19] N.R. Buenfeld, N.M. Hassanein, A.J. Jones, An artificial neural network for predicting carbonation depth in concrete structures, in: Artificial Neural Networks for Civil Engineers: Advanced Features and Applications, Flood I, 1998, pp. 77–117. Kartam N, USA.

[20] A. Abdulalim Alabdullah, M. Iqbal, M. Zahid, K. Khan, M. Nasir Amin, F.E. Jalal, Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis, Constr. Build. Mater. 345 (2022) 128296, https://doi.org/10.1016/j.conbuildmat.2022.128296.

[21] I.U. Ekanayake, D.P.P. Meddage, U. Rathnayake, A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP), Case Stud. Constr. Mater. 16 (2022) e01059, https://doi.org/10.1016/j.cscm.2022.e01059.

[22] P. Xu, X. Ji, M. Li, W. Lu, Small data machine learning in materials science, npj Comput. Mater. 9 (2023) 1–15, https://doi.org/10.1038/s41524-023-01000-z.

[23] Z. Li, J. Yoon, R. Zhang, F. Rajabipour, W.V. Srubar III, I. Dabo, A. Radlińska, Machine learning in concrete science: applications, challenges, and best practices, npj Comput. Mater. 8 (2022) 1–17, https://doi.org/10.1038/s41524-022-00810-x.

[24] S. Kapoor, A. Narayanan, Leakage and the reproducibility crisis in machine-learning-based science, Patterns 4 (2023) 100804, https://doi.org/10.1016/j.patter.2023.100804.

[25] J. Bijen, G.V.D. Wegen, R.V. Selst, Carbonation of Portland bast furnace slag cement concrete with fly ash, in: Special Publication-114: Fly Ash, Silica Fume, Slag, and Natural Pozzolans in Concrete: Proceedings of the Third International Conference, ACI, 1989, pp. 645–668, https://doi.org/10.14359/2209.

[26] C. Ewertson, P.E. Petersson, The influence of curing conditions on the permeability and durability of concrete. Results from a field exposure test, Cement Concr. Res. 23 (1993) 683–692, https://doi.org/10.1016/0008-8846(93)90019-6.

[27] Q. Huy Vu, G. Pham, A. Chonier, E. Brouard, S. Rathnarajan, R. Pillai, R. Gettu, M. Santhanam, F. Aguayo, K.J. Folliard, M.D. Thomas, T. Moffat, C. Shi, A. Sarnot, Impact of different climates on the resistance of concrete to natural carbonation, Constr. Build. Mater. 216 (2019) 450–467, https://doi.org/10.1016/j.conbuildmat.2019.04.263.

[28] H. Al-Khaiat, N. Fattuhi, Carbonation of concrete exposed to hot and arid climate, J. Mater. Civ. Eng. 14 (2002) 97–107, https://doi.org/10.1061/(ASCE)0899-1561(2002)14:2(97).

[29] T. Soda, K. Yamazaki, Long time Study (20 years) on the neutralization of concrete and the rusting of reinforcing bars, J. Japan Soc. Test. Mater. 7 (1958) 680–690, https://doi.org/10.2472/jsms1952.7.680.

[30] M. Adolf, Investigations on the carbonation of concrete, in: Chemistry of Cement, Cement & Concrete Assoc, 1968, pp. 394–401. Tokyo, Japan, https://www.iccc-online.org/archive/.

[31] Y. Kasai, I. Matsui, Y. Fukushima, H. Kamohara, Air Permeability and Carbonation of Blended Cement Mortars, 79, ACI Special Publication, 1983, pp. 435–452, https://doi.org/10.14359/6706.

[32] 0ystein Vennesland, O.E. Gjorv, Silica Concrete - Protection Against Corrosion of Embedded Steel, 79, ACI Symposium Publication, 1983, pp. 719–730, https://doi.org/10.14359/6721.

[33] J.D. Matthews, Carbonation of ten year old concrete with and without pulverized-fuel ash, in: Proc., 2nd Int. Conf. on Ash Technology and Marketing, Ashtech, 1984, p. 398A.

[34] K. Byfors, Carbonation of concrete with silica fume and fly ash, Nord. Concr. Res. 4 (1985) 26–35.

[35] Th.A. Bier, Influence of type of cement and curing on carbonation progress and pore structure of hydrated cement pastes, MRS Proceedings 85 (1986) 123, https://doi.org/10.1557/PROC-85-123.

[36] R.J. Collins, Porous aggregates in concrete: jurassic limestones, in: Building Research Establishment, 1986. https://cis.ihs.com//CIS/document/14117. (Accessed 4 December 2022).

[37] G.G. Litvan, A. Meyer, Carbonation of granulated blast furnace slag cement concrete during twenty years of field exposure, in: Special Publication-91: Fly Ash, Silica Fume, Slag, and Natural Pozzolans in Concrete, ACI, 1986, pp. 1445–1462, https://doi.org/10.14359/10130.

[38] S. Nagataki, H. Ohga, Eun Kyum Kim, effect of curing conditions on the carbonation of concrete with fly ash and the corrosion of reinforcement in long-term tests, in: Special Publication-91: Fly Ash, Silica Fume, Slag, and Natural Pozzolans in Concrete, ACI, 1986, pp. 521–540, https://doi.org/10.14359/10086.

[39] O. Skjolsvold, Carbonation depths of concrete with and without Condensed Silica Fume, in: Special Publication-91: Fly Ash, Silica Fume, Slag, and Natural Pozzolans in Concrete, ACI, 1986, pp. 1031–1048, https://doi.org/10.14359/10111.

[40] D.W.S. Ho, R.K. Lewis, Carbonation of concrete and its prediction, Cement Concr. Res. 17 (1987) 489–504, https://doi.org/10.1016/0008-8846(87)90012-3.
[41] J. Gebauer, R.J. Torrent, Effect of curing on some pore structure-dependent properties of concrete, MRS Proceedings 137 (1988) 301–306, https://doi.org/10.1557/PROC-137-301.
[42] D.W. Hobbs, Carbonation of concrete containing pfa, Mag. Concr. Res. 40 (1988) 69–78, https://doi.org/10.1680/macr.1988.40.143.69.
[43] R.K. Dhir, P.C. Hewlett, Y.N. Chan, Near-surface characteristics of concrete: prediction of carbonation resistance, Mag. Concr. Res. 41 (1989) 137–143, https://doi.org/10.1680/macr.1989.41.148.137.
[44] M. Kokubu, S. Nagataki, Carbonation of concrete with fly ash and corrosion of reinforcements in 20 year tests, in: Special Publication-114: Fly Ash, Silica Fume, Slag, and Natural Pozzolans in Concrete: Proceedings of the Third International Conference, ACI, 1989, pp. 315–330, https://doi.org/10.14359/2007.
[45] G.J. Osborne, Carbonation and permeability of Blastfurnace slag cement concretes from field structures, in: Special Publication-114: Fly Ash, Silica Fume, Slag, and Natural Pozzolans in Concrete: Proceedings of the Third International Conference, ACI, 1989, pp. 1209–1238, https://doi.org/10.14359/2599.
[46] L.J. Parrott, Carbonation, corrosion and standardization, in: Protection of Concrete, E and F Spon Limited, UK, 1990, pp. 1009–1023. https://trid.trb.org/view/367172. (Accessed 4 December 2022).
[47] M.D.A. Thomas, G.J. Osborne, J.D. Matthewst, J.B. Cripwell, A comparison of the properties of OPC, PFA and ggbs concretes in reinforced concrete tank walls of slender section, Mag. Concr. Res. 42 (1990) 127–134, https://doi.org/10.1680/macr.1990.42.152.127.
[48] O.S.B. Al-Amoudi, Rasheeduzzafar, M. Maslehuddin, Carbonation and corrosion of rebars in salt contaminated OPC/PFA concretes, Cement Concr. Res. 21 (1991) 38–50, https://doi.org/10.1016/0008-8846(91)90029-H.
[49] J. Bijen, R.V. SelstI, Effects of fly ash on carbonation of concrete with Portland blast-furnace slag cement, in: Special Publication-126: Durability of Concrete: Second International Conference, ACI, Montreal, Canada, 1991, pp. 1001–1030, https://doi.org/10.14359/2479, 1991.
[50] G.G. Carette, V.M. Malhotra, Longterm strength development of Silica Fume concrete, in: Special Publication-132: Fly Ash, Silica Fume, Slag, and Natural Pozzolans and Natural Pozzolans in Concrete - Proceedings Fourth International Conference, ACI, 1992, pp. 1017–1044, https://doi.org/10.14359/2383.
[51] T. Hakkinen, THE PERMEABILITY OF HIGH STRENGTH BLAST FURNACE SLAG CONCRETE, NORDIC CONCRETE RESEARCH, NO 11, PUBLICATION, 1992. https://trid.trb.org/view/376229. (Accessed 4 December 2022).
[52] L.J. Parrott, Variations of water absorption rate and porosity with depth from an exposed concrete surface: effects of exposure conditions and cement type, Cement Concr. Res. 22 (1992) 1077–1088, https://doi.org/10.1016/0008-8846(92)90038-W.
[53] K. Trill, M. Kawamura, Pore structure and chloride permeability of concretes containing fly ash, blast furnace slag and Silica Fume, in: Special Publication-132: Fly Ash, Silica Fume, Slag, and Natural Pozzolans and Natural Pozzolans in Concrete - Proceedings Fourth International Conference, ACI, 1992, pp. 135–150, https://doi.org/10.14359/2020.
[54] T. Häkkinen, The influence of slag content on the microstructure, permeability and mechanical properties of concrete, Cement Concr. Res. 23 (1993) 518–530, https://doi.org/10.1016/0008-8846(93)90002-Q.
[55] M.N. Haque, M. Kawamura, Carbonation and chloride-induced corrosion of reinforcement in fly ash concrete, ACI Mater. J. 89 (1993) 41–48, https://doi.org/10.14359/1243.
[56] A.P. Barker, J.D. Matthews, Concrete durability specification by Water/Cement or compressive strength for European cement types, ACI Spec. Publ. 145 (1994) 1135–1160, https://doi.org/10.14359/4433.
[57] D.W. Hobbs, Carbonation of concrete containing pfa, Mag. Concr. Res. 46 (1994) 35–38, https://doi.org/10.1680/macr.1994.46.166.35.
[58] Ch. Malami, V. Kaloidas, G. Batis, N. Kouloumbi, Carbonation and porosity of mortar specimens with pozzolanic and hydraulic cement admixtures, Cement Concr. Res. 24 (1994) 1444–1454, https://doi.org/10.1016/0008-8846(94)90158-9.
[59] M.D.A. Thomas, J.D. Matthews, Effect of curing on durability of fly ash concrete, Transp. Res. Rec. (1994) 99–108.
[60] J.P. Balayssac, Ch.H. Détriché, J. Grandet, Effects of curing upon carbonation of concrete, Constr. Build. Mater. 9 (1995) 91–95, https://doi.org/10.1016/0950-0618(95)00001-V.
[61] T. SasataniI, K. TorlII, M. Kawamura, Five-Year exposure Test on long-term properties of concretes containing fly ash, blast-furnace slag, and Silica Fume, in: Special Publication-153: Fly Ash, Silica Fume, Slag, and Natural Pozzolans in Concrete Proceedings Fifth International Conference Milwauk, ACI, US, 1995, pp. 283–296, https://doi.org/10.14359/1075.
[62] J. Wolsiefer , Sr, V. Sivasundaram, V.M. Malhotra, G.G. Carette, Performance of concretes incorporating various forms of Silica Fume, in: Special Publication-153: Fly Ash, Silica Fume, Slag, and Natural Pozzolans in Concrete Proceedings Fifth International Conference Milwauk, ACI, US, 1995, pp. 591–656, https://doi.org/10.14359/1090.
[63] N.H. Olsen, G.R. Summers, Performance of reinforced blast furnace slag cement concrete specimens after 10-Years' exposure in Bahrain, in: Special Publication-170: Fourth CANMET/ACI International Conference on Durability of Concrete, ACI, 1997, pp. 285–308, https://doi.org/10.14359/6827. US.
[64] G.J. Osborne, J.L. Hardcastle, Changes in properties of Portland cement and their effects on concrete durability, in: Special Publication-170: Fourth CANMET/ACI International Conference on Durability of Concrete, ACI, 1997, pp. 1411–1432, https://doi.org/10.14359/6884. US.
[65] A.A. Ramezanianpour, Hadikhanloo Radfar, Moslehi, Maghsoodi, Performance of a different pozzolanic cement concretes under cyclic wetting and drying, in: Special Publication-178: Sixth CANMET/ACI/JCI Conference: Fly Ash, Silica Fume, Slag & Natural Pozzolans in Concrete, ACI, US, 1998, pp. 759–778, https://doi.org/10.14359/6007.
[66] A.M. Dunster, D.J. Bigland, I.R. Holton, Rates of carbonation and reinforcement corrosion in high alumina cement concrete, Mag. Concr. Res. 52 (2000) 433–441, https://doi.org/10.1680/macr.2000.52.6.433.
[67] G.J. Osborne, M.D. Connell, Effects of cementitious slag content and aggregate type on properties and durability of Portland blast-furnace slag cement concretes, in: Special Publication-192: 2000 Canmet/ACI Conference on Durability of Concrete, ACI, 2000, pp. 119–140, https://doi.org/10.14359/5745. US.
[68] M.D.A. Thomas, J.D. Matthews, C.A. Haynes, Carbonation of fly ash concrete, in: Special Publication-192: 2000 Canmet/ACI Conference on Durability of Concrete, ACI, 2000, pp. 539–556, https://doi.org/10.14359/5770. US.
[69] M.R. Jones, M.D. Newlands, A.M.O. Abbas, R.K. Dhir, Comparison of 2 year carbonation depths of common cement concretes using the modified draft CEN test, Mater. Struct. 34 (2001) 396–403, https://doi.org/10.1007/BF02482285.
[70] Y.F. Houst, F.H. Wittmann, Depth profiles of carbonates formed during natural carbonation, Cement Concr. Res. 32 (2002) 1923–1930, https://doi.org/10.1016/S0008-8846(02)00908-0.
[71] M.A. Sanjuán, C. Andrade, M. Cheyrezy, Concrete carbonation tests in natural and accelerated conditions, Adv. Cement Res. 15 (2003) 171–180, https://doi.org/10.1680/adcr.2003.15.4.171.
[72] M. Collepardi, S. Collepardi, E.S.J.J. Ogoumah Olagot, The influence of slag and fly ash on the carbonation of concrete, in: Special Publication-221: Eighth CANMET/ACI International Conference on Fly Ash, Silica Fume, Slag, and Natural Pozzolans in Concrete, ACI, US, 2004, pp. 483–494, https://doi.org/10.14359/13272.
[73] N. Bouzoubaa, B. Tamtsia, M.H. Zhang, R.L. Chevrier, A. Bilodeau, V. M. Malhotra, Carbonation of concrete incorporating high volumes of fly ash, in: Special Publication-234: Seventh CANMET/ACI International Conference on Durability of Concrete, ACI, 2006, pp. 283–304, https://doi.org/10.14359/15943. US.
[74] E. Rozière, A. Loukili, F. Cussigh, A performance based approach for durability of concrete exposed to carbonation, Constr. Build. Mater. 23 (2009) 190–199, https://doi.org/10.1016/j.conbuildmat.2008.01.006.
[75] B. Sabet Divsholi, Three Years Natural Carbonation of PC and Blended Cement Concrete in Singapore Exposure Condition, CI-Premier PTE Ltd., Singapore, 2009.
[76] N. Bouzoubaâ, A. Bilodeau, B. Tamtsia, S. Foo, Carbonation of fly ash concrete: laboratory and field data, Can. J. Civ. Eng. 37 (2010) 1535–1549, https://doi.org/10.1139/L10-081.
[77] B. Chatveera, P. Lertwattanaruk, Durability of conventional concretes containing black rice husk ash, J. Environ. Manag. 92 (2011) 59–66, https://doi.org/10.1016/j.jenvman.2010.08.007.
[78] T.A. Harrison, R. Jones, S. Kandasami, G. Khanna, Effectiveness of the traditional parameters for specifying carbonation resistance, Mag. Concr. Res. 64 (2012) 487–497, https://doi.org/10.1680/macr.10.00186.
[79] Y. Jia, B. Aruhan, P. Yan, Natural and accelerated carbonation of concrete containing fly ash and GGBS after different initial curing period, Mag. Concr. Res. 64 (2012) 143–150, https://doi.org/10.1680/macr.10.00134.
[80] S.-C. Kou, C.-S. Poon, Long-term mechanical and durability properties of recycled aggregate concrete prepared with the incorporation of fly ash, Cement Concr. Compos. 37 (2013) 12–19, https://doi.org/10.1016/j.cemconcomp.2012.12.011.
[81] S. Liu, Z. Wang, X. Li, Long-term properties of concrete containing ground granulated blast furnace slag and steel slag, Mag. Concr. Res. 66 (2014) 1095–1103, https://doi.org/10.1680/macr.14.00074.
[82] R. Bucher, P. Diederich, G. Escadeillas, M. Cyr, Service life of metakaolin-based concrete exposed to carbonation, Cement Concr. Res. 99 (2017) 18–29, https://doi.org/10.1016/j.cemconres.2017.04.013.
[83] M. Kumar, N. Malay, J. Kujur, Study of natural carbonation of concrete incorporating marble dust, Proc. Institut. Civil Eng. Construct. Mater. 171 (2018) 85–92, https://doi.org/10.1680/jcoma.17.00001.
[84] A.B. Ribeiro, T. Santos, A. Gonçalves, Performance of concrete exposed to natural carbonation: use of the k-value concept, Constr. Build. Mater. 175 (2018) 360–370, https://doi.org/10.1016/j.conbuildmat.2018.04.206.
[85] M.Á. Sanjuán, E. Estévez, C. Argiz, D. del Barrio, Effect of curing time on granulated blast-furnace slag cement mortars carbonation, Cement Concr. Compos. 90 (2018) 257–265, https://doi.org/10.1016/j.cemconcomp.2018.04.006.
[86] J. Zhao, E.D. Shumuye, Z. Wang, G.A. Bezabih, Performance of GGBS cement concrete under natural carbonation and accelerated carbonation exposure, J. Eng. 2021 (2021) 1–16, https://doi.org/10.1155/2021/6659768.
[87] S. Rathnarajan, B.S. Dhanya, R.G. Pillai, R. Gettu, M. Santhanam, Carbonation model for concretes with fly ash, slag, and limestone calcined clay - using accelerated and five - year natural exposure data, Cement Concr. Compos. 126 (2022) 104329, https://doi.org/10.1016/j.cemconcomp.2021.104329.
[88] A. Younsi, P. Turcry, A. Aït-Mokhtar, Quantification of CO2 uptake of concretes with mineral additions after 10-year natural carbonation, J. Clean. Prod. 349 (2022) 131362, https://doi.org/10.1016/j.jclepro.2022.131362.
[89] R.J. Currie, Carbonation Depths in structural-quality Concrete: an Assessment of Evidence from Investigations of Structures and from Other Sources - the Construction Information Service, Building Researh Establishment, UK, 1986. https://cis.ihs.com//CIS/document/83802. (Accessed 4 December 2022).

[90] J.D. Matthews, Performance of limestone-filled Cements, Building Research Establishment, UK, 1993. https://cis.ihs.com//CIS/document/84645. (Accessed 5 December 2022).
[91] M.D.A. Thomas, J.D. Matthews, Durability of Pfa Concrete, Building Research Establishment, UK, 1994. https://cis.ihs.com//CIS/document/255994. (Accessed 5 December 2022).
[92] J.D. Matthews, Performance of Pfa Concrete in Aggressive Conditions. 4: Carbonation, Building Research Establishment, UK, 1995. https://cis.ihs.com//CIS/document/255990. (Accessed 5 December 2022).
[93] CEN/TC 51, Measurement of the Carbonation Depth of Hardened Concrete, EU, 1997.
[94] A. Ali, A. Dunster, Durability of reinforced concrete: effects of concrete composition and curing on carbonation under different exposure conditions - the construction information service, Build. Res. Establish. (BRE) British Cement Assoc. (BCA) (1998). UK, https://cis.ihs.com//CIS/document/247154. (Accessed 5 December 2022).
[95] H. Kuosa, Concrete Durability Field Testing in DuraInt-project. Field and Laboratory Results 2007 - 2010, VTT, Finland, 2011.
[96] International Organization for Standardization, Testing of Concrete Part 12: Determination of the Carbonation Resistance of Concrete — Accelerated Carbonation Method, 2015.
[97] China Academy of Building Research, Standard for Test Methods of long-term Performance and Durability of Ordinary Concrete, 2009.
[98] NORDTEST, Concrete, Repairing Materialsand Protective Coating: Carbonation Resistance, 1989.
[99] British Standards Institution, Testing Hardened Concrete. Determination of the Carbonation Resistance of Concrete at Atmospheric Levels of Carbon Dioxide, 2018.
[100] Asociacion Espanola de Normalizacion, Testing Hardened Concrete - Part 10: Determination of the Carbonation Resistance of Concrete at Atmospheric Levels of Carbon Dioxide, 2019.
[101] Association Francaise de Nor malisation, Tests for hardened concrete, Accelerated Carbonation Test - Measurement of the Thickness of Carbonated Concrete, 2022.
[102] RILEM, CPC-18, Measurement of hardened concrete carbonation depth, Mater. Struct. 21 (1988) 453–455, https://doi.org/10.1007/BF02472327.
[103] S.A. Bernal, Y. Dhandapani, Y. Elakneswaran, G.J.G. Gluth, E. Gruyaert, M.C. G. Juenger, B. Lothenbach, K.A. Olonade, M. Sakoparnig, Z. Shi, C. Thiel, P. Van den Heede, H. Vanoutrive, S. von Greve-Dierfeld, N. De Belie, J.L. Provis, Report of RILEM TC 281-CCC: a critical review of the standardised testing methods to determine carbonation resistance of concrete, Mater. Struct. 57 (2024) 173, https://doi.org/10.1617/s11527-024-02424-9.
[104] K. Zhang, M. Yio, H. Wong, N. Buenfeld, Development of more accurate methods for determining carbonation depth in cement-based materials, Cement Concr. Res. 175 (2024) 107358, https://doi.org/10.1016/j.cemconres.2023.107358.
[105] H.F.W. Taylor, Modification of the Bogue calculation, Adv. Cement Res. 2 (1989) 73–77, https://doi.org/10.1680/adcr.1989.2.6.73.
[106] S.O. Ekolu, A review on effects of curing, sheltering, and CO2 concentration upon natural carbonation of concrete, Constr. Build. Mater. 127 (2016) 306–320, https://doi.org/10.1016/j.conbuildmat.2016.09.056.
[107] Originlab Corporation, Originpro 2023b, 2023.
[108] Z. Zhang, Missing data imputation: focusing on single imputation, Ann. Transl. Med. 4 (2016) 9, https://doi.org/10.3978/j.issn.2305-5839.2015.12.38.
[109] D. McElfresh, S. Khandagale, J. Valverde, V.P. C, B. Feuer, C. Hegde, G. Ramakrishnan, M. Goldblum, C. White, When do neural nets outperform boosted trees on tabular data?, arXiv.Org, https://arxiv.org/abs/2305.02997v4, 2023. (Accessed 19 June 2025).
[110] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, O. Tabona, A survey on missing data in machine learning, J. Big Data 8 (2021) 140, https://doi.org/10.1186/s40537-021-00516-9.
[111] J.C. Forsdyke, J.M. Lees, Model fitting to concrete carbonation data with non-zero initial carbonation depth, Mater. Struct. 56 (2023) 22, https://doi.org/10.1617/s11527-023-02104-0.
[112] BSI, BS EN 12390-12, 2020 testing hardened concrete - determination of the carbonation resistance of concrete. Accelerated Carbonation Method, 2020.
[113] S. Tasaka, M. Shinozuka, S. ray chaudhuri, U. Na, Bayesian Inference for Prediction of Carbonation Depth of Concrete Using MCMC, Memoirs of Akashi National College of Technology, 2009.
[114] A. Agresti, Categorical Data Analysis, third ed., Wiley-Interscience, Hoboken, New Jersey, 2013.
[115] H. Vanoutrive, P. Minne, Ö. Cizer, E. Gruyaert, Influence of curing type and duration on the resistance to accelerated carbonation and its relation to natural indoor and outdoor sheltered carbonation, in: A. Jędrzejewska, F. Kanavaris, M. Azenha, F. Benboudjema, D. Schlicke (Eds.), RILEM Bookseries, Springer Nature Switzerland, 2023, pp. 1104–1112, https://doi.org/10.1007/978-3-031-33187-9_102.
[116] A. Lehman (Ed.), JMP for Basic Univariate and Multivariate Statistics: a step-by-step Guide, SAS Press, Cary, NC, 2005.
[117] D. Ju-Long, Control problems of grey systems, Syst. Control Lett. 1 (1982) 288–294, https://doi.org/10.1016/S0167-6911(82)80025-X.
[118] Y. Kuo, T. Yang, G.-W. Huang, The use of grey relational analysis in solving multiple attribute decision-making problems, Comput. Ind. Eng. 55 (2008) 80–93, https://doi.org/10.1016/j.cie.2007.12.002.
[119] S. Liu, Y. Yang, J.Y.-L. Forrest, Grey Systems Analysis: Methods, Models and Applications, Springer, Singapore, 2023.
[120] U.M. Angst, M.R. Geiker, M.C. Alonso, R. Polder, O.B. Isgor, B. Elsener, H. Wong, A. Michel, K. Hornbostel, C. Gehlen, R. François, M. Sanchez, M. Criado, H. Sørensen, C. Hansson, R. Pillai, S. Mundra, J. Gulikers, M. Raupach, J. Pacheco, A. Sagüés, The effect of the steel–concrete interface on chloride-induced corrosion initiation in concrete: a critical review by RILEM TC 262-SCI, Mater. Struct. 52 (2019) 88, https://doi.org/10.1617/s11527-019-1387-0.
[121] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A Next-generation Hyperparameter Optimization Framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Association for Computing Machinery, New York, NY, USA, 2019, pp. 2623–2631, https://doi.org/10.1145/3292500.3330701.
[122] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (2020) 56–67, https://doi.org/10.1038/s42256-019-0138-9.
[123] L. Grinsztajn, E. Oyallon, G. Varoquaux, Why do tree-based models still outperform deep learning on typical tabular data? Adv. Neural Inf. Process. Syst. 35 (2022) 507–520.
[124] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, CatBoost: unbiased boosting with categorical features, in: Advances in Neural Information Processing Systems, Curran Associates, Inc., 2018, in: https://proceedings.neurips.cc/paper_files/paper/2018/hash/14491b756b3a51daac41c248632 85549-Abstract.html. (Accessed 6 June 2025).
[125] Y. Peng, C. Unluer, Interpretable machine learning-based analysis of hydration and carbonation of carbonated reactive magnesia cement mixes, J. Clean. Prod. 434 (2024) 140054, https://doi.org/10.1016/j.jclepro.2023.140054.
[126] A. Majlesi, H. Khodadadi Koodiani, O. Troconis de Rincon, A. Montoya, V. Millano, A.A. Torres-Acosta, B.C. Rincon Troconis, Artificial neural network model to estimate the long-term carbonation depth of concrete exposed to natural environments, J. Build. Eng. 74 (2023) 106545, https://doi.org/10.1016/j.jobe.2023.106545.
[127] J. Wang, Z. Zhang, X. Liu, Y. Shao, X. Liu, H. Wang, Prediction and interpretation of concrete corrosion induced by carbon dioxide using machine learning, Corros. Sci. 233 (2024) 112100, https://doi.org/10.1016/j.corsci.2024.112100.
[128] D.E. Terasaki Hart, T.-N. Bùi, L. Di Maggio, I.J. Wang, Global phenology maps reveal the drivers and effects of seasonal asynchrony, Nature 645 (2025) 133–140, https://doi.org/10.1038/s41586-025-09410-3.
[129] C. Martínez-Ruiz, J.R.M. Black, C. Puttick, M.S. Hill, J. Demeulemeester, E. Larose Cadieux, K. Thol, T.P. Jones, S. Veeriah, C. Naceur-Lombardelli, A. Toncheva, P. Prymas, A. Rowan, S. Ward, L. Cubitt, F. Athanasopoulou, O. Pich, T. Karasaki, D.A. Moore, R. Salgado, E. Colliver, C. Castignani, M. Dietzen, A. Huebner, M. Al Bakir, M. Tanić, T.B.K. Watkins, E.L. Lim, A.M. Al-Rashed, D. Lang, J. Clements, D.E. Cook, R. Rosenthal, G.A. Wilson, A.M. Frankell, S. de Carné Trécesson, P. East, N. Kanu, K. Litchfield, N.J. Birkbak, A. Hackshaw, S. Beck, P. Van Loo, M. Jamal-Hanjani, N. McGranahan, C. Swanton, M.A. Bakir, E.L. Lim, A. M. Frankell, K. Litchfield, N.J. Birkbak, P. Van Loo, J.F. Lester, A. Bajaj, A. Nakas, A. Sodha-Ramdeen, K. Ang, M. Tufail, M.F. Chowdhry, M. Scotland, R. Boyles, S. Rathinam, C. Wilson, D. Marrone, S. Dulloo, D.A. Fennell, G. Matharu, J. A. Shaw, J. Riley, L. Primrose, E. Boleti, H. Cheyne, M. Khalil, S. Richardson, T. Cruickshank, G. Price, K.M. Kerr, S. Benafif, K. Gilbert, B. Naidu, A.J. Patel, A. Osman, C. Lacson, G. Langman, H. Shackleford, M. Djearaman, S. Kadiri, G. Middleton, A. Leek, J.D. Hodgkinson, N. Totten, A. Montero, E. Smith, E. Fontaine, F. Granato, H. Doran, J. Novasio, K. Rammohan, L. Joseph, P. Bishop, R. Shah, S. Moss, V. Joshi, P. Crosbie, F. Gomes, K. Brown, M. Carter, A. Chaturvedi, L. Priest, P. Oliveira, C.R. Lindsay, F.H. Blackhall, M.G. Krebs, Y. Summers, A. Clipson, J. Tugwood, A. Kerr, D.G. Rothwell, E. Kilgour, C. Dive, H.J.W.L. Aerts, R.F. Schwarz, T.L. Kaufmann, Z. Szallasi, J. Kisistok, M. Sokac, M. Diossy, A. Bunkum, A. Stewart, A. Magness, A. Karamani, B. Chain, B. B. Campbell, C. Bailey, C. Abbosh, C.E. Weeden, C. Lee, C. Richard, C.T. Hiley, D. R. Pearce, D. Karagianni, D. Biswas, D. Levi, E. Hoxha, E. Nye, E. Grönroos, F. Gálvez-Cancino, F. Gimeno-Valiente, G. Kassiotis, G. Stavrou, G. Mastrokalos, H. Zhai, H.L. Lowe, I.G. Matos, J. Goldman, J.L. Reading, J. Herrero, J.K. Rane, J. Nicod, J.M. Lam, J.A. Hartley, K.S. Peggs, K.S.S. Enfield, K. Selvaraju, K.W. Ng, K. Chen, K. Dijkstra, K. Grigoriadis, K. Thakkar, L. Ensell, M. Shah, M.V. Duran, M. Litovchenko, M.W. Sunderland, M. Leung, M. Escudero, M. Angelova, M. Sivakumar, O. Chervova, O. Lucas, O. Al-Sawaf, P. Hobson, P. Pawlik, R. K. Stone, R. Bentham, R.E. Hynds, R. Vendramin, S. Saghafinia, S. López, S. Gamble, S.K.A. Ung, S.A. Quezada, S. Vanloo, S. Zaccaria, S. Hessey, S. Boeing, S.K. Bola, T. Denner, T. Marafioti, T.P. Mourikis, V. Spanswick, V. Barbè, W.-T. Lu, W. Hill, W.K. Liu, Y. Wu, Y. Naito, Z. Ramsden, C. Veiga, G. Royle, C.-A. Collins-Fekete, F. Fraioli, P. Ashford, T. Clark, M.D. Forster, S.M. Lee, E. Borg, M. Falzon, D. Papadatos-Pastos, J. Wilson, T. Ahmad, A.J. Procter, A. Ahmed, M. N. Taylor, A. Nair, D. Lawrence, D. Patrini, N. Navani, R.M. Thakrar, S.M. Janes, E.M. Hoogenboom, F. Monk, J.W. Holding, J. Choudhary, K. Bhakhri, M. Scarci, M. Hayward, N. Panagiotopoulos, P. Gorman, R. Khiroya, R.C.M. Stephens, Y.N. S. Wong, S. Bandula, A. Sharp, S. Smith, N. Gower, H.K. Dhanda, K. Chan, C. Pilotti, R. Leslie, A. Grapa, H. Zhang, K. AbdulJabbar, X. Pan, Y. Yuan, D. Chuter, M. MacKenzie, S. Chee, A. Alzetani, J. Cave, L. Scarlett, J. Richards, P. Ingram, S. Austin, E. Lim, P. De Sousa, S. Jordan, A. Rice, H. Raubenheimer, H. Bhayani, L. Ambrose, A. Devaraj, H. Chavan, S. Begum, S.I. Buderi, D. Kaniu, M. Malima, S. Booth, A.G. Nicholson, N. Fernandes, P. Shah, C. Proli, M. Hewish, S. Danson, M.J. Shackcloth, L. Robinson, P. Russell, K.G. Blyth, C. Dick, J. Le Quesne, A. Kirk, M. Asif, R. Bilancia, N. Kostoulas, M. Thomas, C. Swanton, N. McGranahan, Genomic–transcriptomic evolution in lung cancer and metastasis, Nature 616 (2023) 543–552, https://doi.org/10.1038/s41586-023-05706-4.
[130] S.D. Tiegs, K.A. Capps, D.M. Costello, J.P. Schmidt, C.J. Patrick, J.J. Follstad Shah, C.J. LeRoy, V. Acuña, R. Albariño, D.C. Allen, C. Alonso, P. Andino,

C. Arango, J. Aroviita, M.V.M. Barbosa, L.A. Barmuta, C. Baxter, B. Bellinger, L. Boyero, L. Bragina, L.E. Brown, A. Bruder, D.A. Bruesewitz, F. Burdon, M. Callisto, A. Camacho, C. Canhoto, M.M. Castillo, E. Chauvet, J. Clapcott, F. Colas, C. Colón-Gaud, J. Cornut, V. Crespo-Pérez, W.F. Cross, J. Culp, M. Danger, O. Dangles, E. de Eyto, A.M. Derry, V.D. Villanueva, M.M. Douglas, A. Elosegi, A.C. Encalada, S. Entrekin, R. Espinosa, V. Ferreira, C. Ferriol, K. M. Flanagan, A.S. Flecker, T. Fleituch, A. Frainer, N. Friberg, P.C. Frost, E. A. Garcia, L. García-Lago, P.E.G. Soto, M.O. Gessner, S. Ghate, D.P. Giling, A. Gilmer, J.F. Gonçalves, R.K. Gonzales, M.A.S. Graça, M. Grace, N.A. Griffiths, H.-P. Grossart, F. Guérold, V. Gulis, P.E. Gutiérrez-Fonseca, L.U. Hepp, S. Higgins, T. Hishi, J. Huddart, J. Hudson, M. Imberger, C. Iñiguez-Armijos, M.W. Isken, T. Iwata, D.J. Janetski, A.E. Kirkwood, A.A. Koning, S. Kosten, K.A. Kuehn, H. Laudon, P.R. Leavitt, A.L. Lemes da Silva, S. Leroux, P.J. Lisi, R. MacKenzie, A. M. Marcarelli, F.O. Masese, P.B. McIntyre, B.G. McKie, A. Medeiros, K. Meissner, M. Miliša, S. Mishra, Y. Miyake, A. Moerke, S. Mombrikotb, R. Mooney, T. Moulton, T. Muotka, J. Negishi, V. Neres-Lima, M.L. Nieminen, J. Nimptsch, J. Ondruch, R. Paavola, I. Pardo, E.T.H.M. Peeters, J. Pozo, A. Prussian, E. Quenta, B. Reid, J.S. Richardson, A. Rigosi, J. Rincón, G. Risnoveanu, C. T. Robinson, L. Rodríguez-Gallego, T.V. Royer, J.A. Rusak, A.C. Santamans, G. B. Selmeczy, G. Simiyu, A. Skuja, J. Smykla, R. Sponseller, K.R. Sridhar, A. Stoler, C.M. Swan, F.T. de Mello, J.D. Tonkin, S. Uusheimo, A.M. Veach, S. Vilbaste, L.B.-M. Vought, C.-P. Wang, J.R. Webster, P.B. Wilson, S. Woelfl, G. Woodward, M. A. Xenopoulos, A.G. Yates, C. Yoshimura, C.M. Yule, Y. Zhang, J.A. Zwart, Human activities shape global patterns of decomposition rates in rivers, Science 384 (2024) 1191–1195, https://doi.org/10.1126/science.adn1262.

[131] C. Molnar, Interpretable machine learning: a guide for making black box models explainable, Leanpub, Victoria, British Columbia (2020).

[132] S. Ahmed, M.S. Kaiser, M. Shahadat Hossain, K. Andersson, A comparative analysis of LIME and SHAP interpreters with explainable ML-Based diabetes predictions, IEEE Access 13 (2025) 37370–37388, https://doi.org/10.1109/ACCESS.2024.3422319.

[133] C. Molnar, T. Freiesleben, Supervised machine learning for science: how to stop worrying and love your black box, Chistoph Molnar c/o DRA, München, Germany (2024).

[134] W. McKinney, Data structures for statistical computing in python, in: S. van der Walt, J. Millman (Eds.), Proceedings of the Python in Science Conference, 2010, pp. 56–61, https://doi.org/10.25080/Majora-92bf1922-00a.

[135] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[136] T. Chen, C. Guestrin, XGBoost, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[137] F.G. da Silva, P. Helene, P. Castro-Borges, J.B. Liborio, Sources of variations when comparing concrete carbonation results, J. Mater. Civ. Eng. 21 (2009) 333–342, https://doi.org/10.1061/(ASCE)0899-1561(2009)21:7(333).

[138] BSI, BS EN 197-1:2011 Cement - Composition, Specifications and Conformity Criteria for Common Cements, 2019.

[139] C.-Q. Lye, R.K. Dhir, G.S. Ghataora, Carbonation resistance of fly ash concrete, Mag. Concr. Res. 67 (2015) 1150–1178, https://doi.org/10.1680/macr.15.00204.

[140] C.-Q. Lye, R.K. Dhir, G.S. Ghataora, Carbonation resistance of GGBS concrete, Mag. Concr. Res. 68 (2016) 936–969, https://doi.org/10.1680/jmacr.15.00449.

[141] M. Azzeh, D. Neagu, P.I. Cowling, Fuzzy grey relational analysis for software effort estimation, Empir. Softw. Eng. 15 (2010) 60–90, https://doi.org/10.1007/s10664-009-9113-0.

[142] Q. Qiu, H. Liu, Numerical embedding of categorical features in tabular data: a survey, 2023, International Conference on Machine Learning and Cybernetics (ICMLC) (2023) 446–451, https://doi.org/10.1109/icmlc58545.2023.10327921.

[143] P. Tennis, M.D.A. Thomas, W. Weiss, State-Of-The-Art Report on Use of Limestone in Cements at Levels of up to 15%, Portland cement association, UK, 2011.

[144] K.L. Scrivener, B. Lothenbach, N. De Belie, E. Gruyaert, J. Skibsted, R. Snellings, A. Vollpracht, TC 238-SCM: hydration and microstructure of concrete with SCMs, Mater. Struct. 48 (2015) 835–862, https://doi.org/10.1617/s11527-015-0527-4.

[145] M.H.N. Yio, M.J. Mac, Y.X. Yeow, H.S. Wong, N.R. Buenfeld, Effect of autogenous shrinkage on microcracking and mass transport properties of concrete containing supplementary cementitious materials, Cement Concr. Res. 150 (2021) 106611, https://doi.org/10.1016/j.cemconres.2021.106611.

[146] J. Han, W. Liu, S. Wang, D. Du, F. Xu, W. Li, G. De Schutter, Effects of crack and ITZ and aggregate on carbonation penetration based on 3D micro X-ray CT microstructure evolution, Constr. Build. Mater. 128 (2016) 256–271, https://doi.org/10.1016/j.conbuildmat.2016.10.062.

[147] H.S. Wong, M. Zobel, N.R. Buenfeld, R.W. Zimmerman, Influence of the interfacial transition zone and microcracking on the diffusivity, permeability and sorptivity of cement-based materials after drying, Mag. Concr. Res. 61 (2009) 571–589, https://doi.org/10.1680/macr.2008.61.8.571.

[148] N.R. Buenfeld, E. Okundi, Effect of cement content on transport in concrete, Mag. Concr. Res. 50 (1998) 339–351, https://doi.org/10.1680/macr.1998.50.4.339.

[149] V.G. Papadakis, C.G. Vayenas, M.N. Fardis, Experimental investigation and mathematical modeling of the concrete carbonation problem, Chem. Eng. Sci. 46 (1991) 1333–1338, https://doi.org/10.1016/0009-2509(91)85060-B.

[150] P.C. Austin, I.R. White, D.S. Lee, S. van Buuren, Missing data in clinical research: a tutorial on multiple imputation, Can. J. Cardiol. 37 (2021) 1322–1331, https://doi.org/10.1016/j.cjca.2020.11.010.

[151] A. Jadhav, D. Pramod, K. Ramanathan, Comparison of performance of data imputation methods for numeric dataset, Appl. Artif. Intell. 33 (2019) 913–933, https://doi.org/10.1080/08839514.2019.1637138.

[152] S. Hussain, D. Bhunia, S.B. Singh, J.S. Yadav, A study on the carbonation of binary and ternary blended cement mortar and concrete, J. Struct. Integrity Main. 7 (2022) 46–60, https://doi.org/10.1080/24705314.2021.1971892.

[153] R.A. Rivera, M.Á. Sanjuán, D.A. Martín, J.L. Costafreda, Performance of ground granulated blast-furnace slag and coal fly ash ternary Portland cements exposed to natural carbonation, Materials 14 (2021) 3239, https://doi.org/10.3390/ma14123239.

[154] K. De Weerdt, M.B. Haha, G. Le Saout, K.O. Kjellsen, H. Justnes, B. Lothenbach, Hydration mechanisms of ternary Portland cements containing limestone powder and fly ash, Cement Concr. Res. 41 (2011) 279–291, https://doi.org/10.1016/j.cemconres.2010.11.014.

[155] C. Andrade, Carbonation effect on the chloride profile, XV international conference on durability of building materials and components. Ebook of Proceedings, 2018, https://doi.org/10.23967/dbmc.2020.232.

[156] S. Nagataki, H. Ohga, Combined Effect of Carbonation and Chloride on Corrosion of Reinforcement in Fly Ash Concrete, 132, ACI Special Publication, 1992, pp. 227–244, https://doi.org/10.14359/2069.

[157] M.A. Sanjuán, C. del Olmo, Carbonation resistance of one industrial mortar used as a concrete coating, Build. Environ. 36 (2001) 949–953, https://doi.org/10.1016/S0360-1323(00)00045-7.

[158] D.W.S. Ho, R.K. Lewis, Carbonation of Concrete Incorporating Fly Ash or a Chemical Admixture, 79, ACI Special Publication, 1983, pp. 333–346, https://doi.org/10.14359/6700.

[159] G. Heirman, L. Vandewalle, V. Boel, K. Audenaert, G. De Schutter, J. D'Hemricourt, B. Desmet, J. Vantomme, Chloride penetration and carbonation in self-compacting concrete, in: Proc. of the Int. RILEM/JCI Seminar on Concrete Durability and Service Life Planning, RILEM Publications SARL, Bagneux, 2006, pp. 13–23.

[160] M. Sonebi, V. O'Donughue, G. Keogh, Effect of the Type of Supplementary Materials and Viscosity Enhancing Admixture on the Durability of Self-Compacting Concrete, 2008.

[161] M. Valcuende, C. Parra, Natural carbonation of self-compacting concretes, Constr. Build. Mater. 24 (2010) 848–853, https://doi.org/10.1016/j.conbuildmat.2009.10.021.

[162] Y. Yao, L. Wang, J. Li, N. De Belie, X. Shi, P. Van den Heede, C. Zhang, Z. Liu, V. Talakokula, Z. Jin, C. Xiong, J. Lu, S. Kamali-Bernard, T. Bansal, B. Li, Z. Wang, Y. Huang, Report of RILEM TC 281-CCC: effect of loading on the carbonation performance of concrete with supplementary cementitious materials — an interlaboratory comparison of different test methods and related observations, Mater. Struct. 56 (2023) 110, https://doi.org/10.1617/s11527-023-02190-0.

[163] D.A. Powers, Y. Xie, Statistical Methods for Categorical Data Analysis, second ed., Emerald, Bingley, 2008.