



**HAL**  
open science

# Interpretable machine learning model for autogenous shrinkage prediction of low-carbon cementitious materials

Benoit Hilloulin, van Quan Tran

► **To cite this version:**

Benoit Hilloulin, van Quan Tran. Interpretable machine learning model for autogenous shrinkage prediction of low-carbon cementitious materials. *Construction and Building Materials*, 2023, 396, pp.132343. 10.1016/j.conbuildmat.2023.132343 . hal-04153200

**HAL Id: hal-04153200**

**<https://hal.science/hal-04153200>**

Submitted on 6 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Interpretable Machine Learning Model for Autogenous Shrinkage Prediction of Low-Carbon Cementitious Materials

Benoît Hilloulin<sup>1</sup>, Van Quan Tran<sup>2</sup>, \*

<sup>1</sup>Nantes Université, Ecole Centrale Nantes, CNRS, GeM, UMR 6183, F-44000 Nantes, France – e-mail: benoit.hilloulin@ec-nantes.fr

<sup>2</sup>University of Transport Technology, Hanoi 100000, Vietnam

\*Corresponding author: quantv@utt.edu.vn

## Highlights

- A large database about the autogenous shrinkage of concrete with SCM is built
- Ensemble machine learning models can predict the autogenous shrinkage efficiently
- ML models can perform better than B4 and CEB models modified versions.
- SCM influence can be interpreted using SHAP analysis
- Insights into blended cement concrete shrinkage behavior can be envisioned.

## Abstract

Supplementary cementitious materials (SCM) are key components of low-carbon cementitious materials. However, their effects, especially for the emerging high cement replacement ratio formulations, are sometimes very challenging to predict, leading to restrictions in most standards. Through advanced Machine Learning techniques, this study provides novel insights into the autogenous shrinkage properties of concrete containing various kinds of supplementary cementitious materials such as slag, fly ash, silica fume, calcined clay, and filler. Four machine learning models are optimized and their predictions are compared on a dedicated database, including ternary and quaternary cement blends. The Extreme Gradient Boosting (XGBoost) model exhibited the highest accuracy among these models. Interpretative tools such as SHAP analysis are then used to obtain novel information about the relative influence of the SCM on the autogenous shrinkage depending on their dosage. The model is finally compared to the recently modified versions of the B4 and the CEB analytical models to predict the autogenous shrinkage of unknown concrete formulations. The model showed significantly better predictions than the analytical models, and a web application is proposed, paving the way for precise and

30 interpretable autogenous shrinkage predictions of cementitious materials incorporating SCM.

31

32 Keywords: Concrete; Autogenous Shrinkage; Supplementary Cementitious Material (SCM);  
33 Machine Learning; XGBoost; Shrinkage Models.

34

## 35 **1 Introduction**

36 Decreasing cementitious materials' carbon footprint is necessary to limit the overall  
37 environmental impact of the construction industry. This decrease can be achieved through several  
38 techniques such as the construction and demolition waste (CDW) usage, for example, the use of  
39 crushed concrete waste into Recycled Aggregate Concrete (RAC) [1], or the extensive use of  
40 supplementary cementitious materials, some of them being industrial byproducts [2–4]. However,  
41 these additions considerably impact the behavior of cementitious materials, and, because of these  
42 rapid evolutions, such novel components are rarely very well understood nor precisely considered  
43 in the construction standards. More specifically, it is generally admitted that concrete early age  
44 deformations can induce severe structural evolutions. Apart from environmental degradations,  
45 these deformations are mostly due to shrinkage and creep mechanisms [5]. For example, high-  
46 performance and ultra-high performance concrete (HPC and UHPC) exhibit significant early age  
47 deformations because of their low water-to-cement ratio. Indeed, while deformation of normal  
48 strength concrete is primarily governed by drying shrinkage, which occurs some weeks or months  
49 after setting due to a water loss, deformation in low water-to-cement concrete can develop very  
50 quickly at an early age because of the rapid evaporation of water from the surface of the specimen,  
51 namely plastic shrinkage, or because of the consumption of water during the hydration directly  
52 causing chemical shrinkage, and generating capillary depressions leading to macroscopic  
53 deformations referred to as autogenous shrinkage. It has been found that autogenous shrinkage can  
54 represent more than half of the total deformation in ultra-high performance concrete [6], creating  
55 considerable stresses at a microscale that can lead to cracking. As cracking should generally be  
56 avoided, either because of structural concerns or aesthetic reasons[7,8], specific attention should  
57 be paid to early-age deformation of low water-to-cement ratio high and ultra-high performance  
58 concrete.

59 Several analytical models have been introduced during the last decades to predict shrinkage.  
60 The major construction codes provide quick ways to estimate the shrinkage amplitude of several

61 concrete mixes based on their composition and, in some cases, by considering standard strength  
62 characteristics such as the 28-days compressive strength. Power and exponential laws have been  
63 introduced to describe the shrinkage amplitude evolution relatively to time [9,10]. Thanks to these  
64 laws, the quick increase in autogenous shrinkage amplitude during the first weeks can be grasped.  
65 Most of these models are a combination between empirical laws and advanced theories. For  
66 example, B4 model [10] is supported by the solidification theory, the theory of microprestress  
67 relaxation, activation energy concepts, the moisture diffusion theory, and damage models for  
68 microcracking, and proved very good results in predicting autogenous and drying shrinkage of  
69 structural elements [11]. Other models have been proposed in reference codes such as Eurocode 2,  
70 fib CEB MC 10 [9,12,13]. Recent improvements of both the B4 and the CEB MC 10 model have  
71 been proposed to take into consideration the influence of some supplementary cementitious  
72 materials [14,15]. These models have been found particularly effective in predicting long-term  
73 predictions deformation of concrete, but several shortcomings have been noticed concerning early-  
74 age prediction and regarding actual concrete formulations, which can include a high proportion of  
75 additions [15,16]. Moreover, the growing use of various supplementary cementitious materials  
76 (SCM), leading to ternary or quaternary concrete mixes in which synergistic effects may arise, is  
77 rarely taken into consideration.

78 Indeed, due to economic and environmental considerations, an increasing amount of SCM is  
79 being used in recent concrete formulations. Various SCM such as calcined clay, fly ash, ground  
80 granulated blast furnace slag, later referred to as slag, and silica fume commonly replace cement,  
81 while additions like fillers may partially replace cement or only play a physical role in increasing  
82 the overall compacity of the mix. These additions influence shrinkage mechanisms to various  
83 extents, especially autogenous shrinkage. It is commonly admitted that silica fume, commonly  
84 used in UHPC due to its filling properties and high pozzolanic activity, increases shrinkage even  
85 for small replacement ratios [5]. Conversely, fly ash alone has been reported to decrease shrinkage  
86 in most studies because of the slower hydration reaction [17,18]. In most published studies, slag  
87 increased autogenous shrinkage, especially long-term autogenous shrinkage, because its fineness  
88 leads to high reactivity and higher chemical shrinkage than pure cement concrete [19–22]. Some  
89 uncertainty remains concerning the influence of slag on the autogenous shrinkage of UHPC since  
90 it has been reported that the addition of slag increased the autogenous shrinkage in some cases  
91 [15]. With the demand for low-cement concrete, calcined clay, which is readily available around  
92 the globe, has been proved to be an interesting product to cut CO<sub>2</sub> emissions due to concrete

93 formulation. Though calcined clay, generally associated with limestone to produce LC<sup>3</sup> cement,  
94 induces novel chemical reactions during hydration, and leads to a fine porosity [23], its influence  
95 on the autogenous shrinkage is rarely reported and might not be prejudicial [24–26]. Finally, fillers  
96 have been found to decrease shrinkage as they act like small aggregates restricting deformations  
97 [27,28], but their interactions with other additions are often only considered from a chemical point  
98 of view and are rarely addressed from a shrinkage point of view [27,29].

99 Therefore, with the increasing complexity of concrete formulations, a better understanding of  
100 fundamental evolutions such as autogenous shrinkage is needed. Recent studies have proposed  
101 adaptation to the aforementioned analytical models to consider the use of SCM and their eventual  
102 combination [15,30]. In [15], CEB Model Code 2010 has been improved by introducing correction  
103 factors using an extensive experimental campaign. However, such parameters might not represent  
104 other concrete formulation behaviors, and more extensive shrinkage databases should be built.

105 Recent developments in Machine Learning (ML) have provided novel insights into concrete  
106 properties-related issues [31]. Since the publication of the first Artificial Neural Network model  
107 (ANN) to predict concrete compressive strength [32], numerous research works have been  
108 published to predict concrete compressive strength with ML-based models [33–36]. ML-based  
109 models have also proved their interest in predicting other concrete properties such as fresh  
110 properties [37], creep [38,39], chloride [40], carbonation resistance [41], frost resistance [42] and  
111 shrinkage [43,44]. They can also be used for microscale properties determination and  
112 enhancement, such as during indentation [45,46], image analysis [47–50], or large-scale structural  
113 monitoring [51,52]. While ML models might have initially appeared as black-box models, recent  
114 advances such as features analysis and dubbed SHapley Additives exPlanations (SHAP)  
115 introduced novel ways to explore the impacts of the model inputs on the final predictions [53].  
116 Such tools have been used recently in several studies related to concrete [44,54] and might help  
117 build advanced autogenous shrinkage prediction models comparable to the existing most  
118 performant analytical models such as B4 and CEB models.

119 This study develops an interpretable ensemble ML model to predict the autogenous shrinkage  
120 evolution in low-carbon cementitious materials with SCM. An extensive dedicated database is  
121 built, including ternary and quaternary formulations, and optimized ML models are compared to  
122 the classic and improved versions of B4 and CEB Model Code models using interpretative tools  
123 such as SHAP. Then, a web application is proposed to provide fast and precise shrinkage

124 predictions to the Civil Engineering community. The manuscript is structured as follows: a first  
125 section describes the new shrinkage database developed during this study, then a section presents  
126 the numerical methods employed before analyzing and discussing the results and a conclusion.

127

## 128 **2 Autogenous Shrinkage Database description and analysis**

129 To develop an accurate model for autogenous shrinkage prediction of concrete incorporating  
130 SCM, a comprehensive experimental database is required. To this end, we adopted the method  
131 published in [44]. The original NU database containing various creep and shrinkage evolutions of  
132 cementitious materials has been extended by low water-to-cement ratio cement paste, mortar or  
133 concrete samples autogenous shrinkage curves. This database has been further extended to include  
134 more autogenous shrinkage data of various concrete of binary, ternary or quaternary formulations  
135 from various sources [15,26,55–60]. In total, a database comprising 347 autogenous shrinkage  
136 observations and 28d compressive strength of various concrete mixes has been built. To the authors  
137 knowledge, this is the most comprehensive autogenous shrinkage database to date. The final  
138 database covers a wide range of water-to-cement and water-to-binder ratios (from normal strength  
139 to ultra-high strength concrete) as well as a wide variety of binary, ternary and quaternary  
140 formulations. These shrinkage curves were interpolated to obtain 1d, 2d, 7d, 14d, and 28d  
141 shrinkage values. Twelve input variables have been considered: water-cement ratio (W/C), water-  
142 binder ratio (W/B), aggregate-cement ratio (A/C), cement content ( $\text{kg}/\text{m}^3$ ), silica fume content (%  
143 cement mass), fly ash content (% cement mass), slag content (% cement mass), calcined clay  
144 content (% cement mass) (denoted as metakaolin content), filler content ( $\text{kg}/\text{m}^3$ ), amount of  
145 superplasticizer (% cement mass), time since the beginning of shrinkage measurements (days) and  
146 28-d compressive strength (MPa). After removing data points with missing data, a database with  
147 1482 datapoints has been finally obtained. The descriptions and statistical attributes of the input  
148 variables are given in Table 1, and Fig. 1 displays their statistical distributions. It can be seen that  
149 a wide range of compositions has been investigated as water-to-binder ratios range from 0.16 to  
150 0.86, cement content ranges from  $167.4 \text{ kg}/\text{m}^3$  to  $1632 \text{ kg}/\text{m}^3$  (cement paste) with a large  
151 proportion of values between  $167 \text{ kg}/\text{m}^3$  and  $700 \text{ kg}/\text{m}^3$  for concretes. Supplementary cementitious  
152 materials contents reflect the actual wide variety of formulations, some of them with important  
153 substitution amounts. Indeed silica fume content, fly ash content, slag content, calcined clay and  
154 filler content up to 50%, 150%, 400%, 57.4% and 125% relatively to cement mass were

155 considered. A maximum shrinkage value of around  $-2089 \mu\epsilon$  has been obtained and a median  
 156 shrinkage value of  $-114.7 \mu\epsilon$  can be calculated. Common autogenous shrinkage features can be  
 157 retrieved from the 2D scatter plots exposed in Fig. 1 by plotting linear regressions: shrinkage  
 158 increases with smaller W/B and A/C values, while it generally increases with higher cement or  
 159 silica fume content as well as higher compressive strength as they are generally associated to low  
 160 W/B values.

161 Correlations between the model parameters were calculated too. Fig. 2 gives a graphical  
 162 representation of the correlation matrix. As expected, strong positive correlations can be observed  
 163 between, W/B, A/C and shrinkage, and strong negative correlations can be observed between  
 164 shrinkage and  $fc_{28}$ , cement content or silica fume content. However, no strong correlations  
 165 between input variables, and as such, all these variables were used as the input for ML.

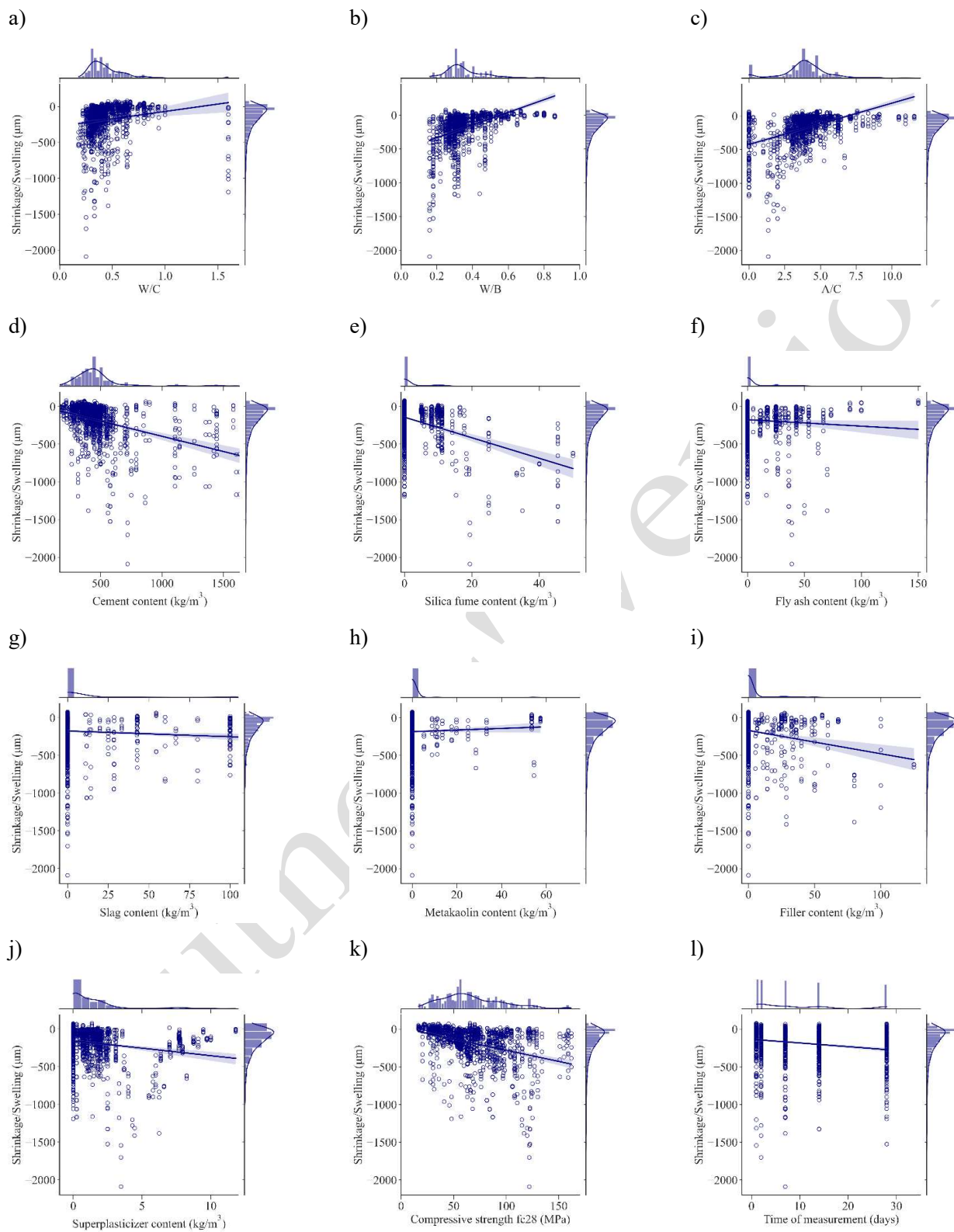
166

167 **Table 1.** Description of the database used in this study

	Unit	Min	Q25%	Mean	Median	Q75%	Max	Std	Skw
W/C	-	0.180	0.320	0.444	0.390	0.500	1.600	0.188	2.807
W/B	-	0.160	0.290	0.360	0.336	0.400	0.860	0.117	1.295
A/C	-	0.00	3.28	3.90	3.90	4.73	11.56	1.77	0.36
Cement content	kg/m <sup>3</sup>	167.4	360.0	472.1	440.0	496.0	1632.0	222.3	2.9
Silica fume content	% cem	0.0	0.0	3.0	0.0	0.0	50.0	7.2	3.7
Fly Ash	% cem	0.0	0.0	8.2	0.0	0.0	150.0	18.3	3.1
Slag content	% cem	0.0	0.0	9.8	0.0	0.0	400.0	40.6	6.7
Metakaolin content	% cem	0.0	0.0	1.6	0.0	0.0	57.4	8.2	5.7
Filler content	% cem	0.0	0.0	4.7	0.0	0.0	125.0	14.5	4.1
Superplasticizer content	% cem	0.0	0.0	1.5	0.8	1.8	11.8	2.2	2.3
Compressive strength $fc_{28}$	MPa	16.5	45.9	67.1	63.0	85.7	162.0	29.4	0.8
Time of measurement	days	1.0	2.0	9.7	7.0	14.0	28.0	9.6	1.0
Shrinkage/Swelling	$\mu\epsilon$	-2089.600	-243.000	-182.832	-114.700	-28.650	78.900	237.617	-2.576

168 Skw=Skewness; Std=Standard deviation;

169



170 **Fig. 1.** Data distribution of each input variable and the output variable (shrinkage/swelling).

171



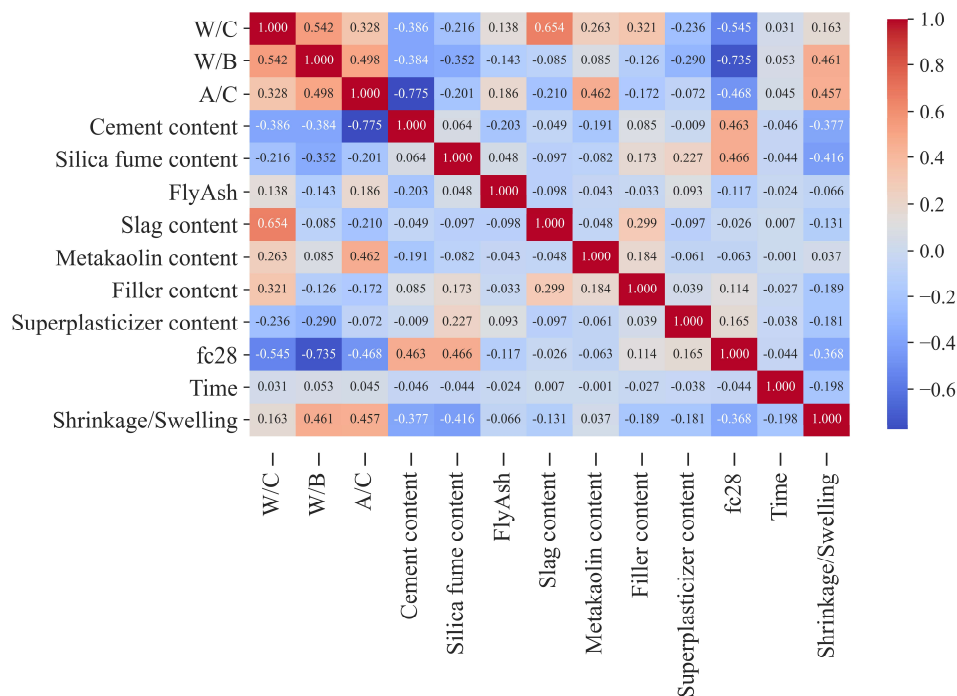


Fig. 2. Pearson correlation of all feature and Shrinkage/Swelling

### 3 Shrinkage prediction using Machine Learning and Analytical models

#### 3.1 Modeling strategy

The modeling strategy is four steps presented in Figure 3:

- Step 1: Collection a database consisting 12 input variables and one output variable “shrinkage/swelling”.
- Step 2: Determination the best performance of ML model

All the observation data from the original database have been classically split into training and testing datasets. 70% of whole dataset is used to building the four ML models including SVM, ANN, GB and XGB with aided of metaheuristic optimization algorithm Simulated Annealing (SA) Optimization algorithm which tunes the hyperparameters of four ML algorithms. The K-Fold Cross Validation (CV) is integrated in the tuning hyperparameter process for validating the cost function  $R^2$  of optimization process. The evaluation ML performances are carried out with aided of three metrics, such as  $R^2$ , RMSE and MAE, in prediction of Shrinkage/Swelling. Based on the best performance, the best ML model is proposed for the next steps.

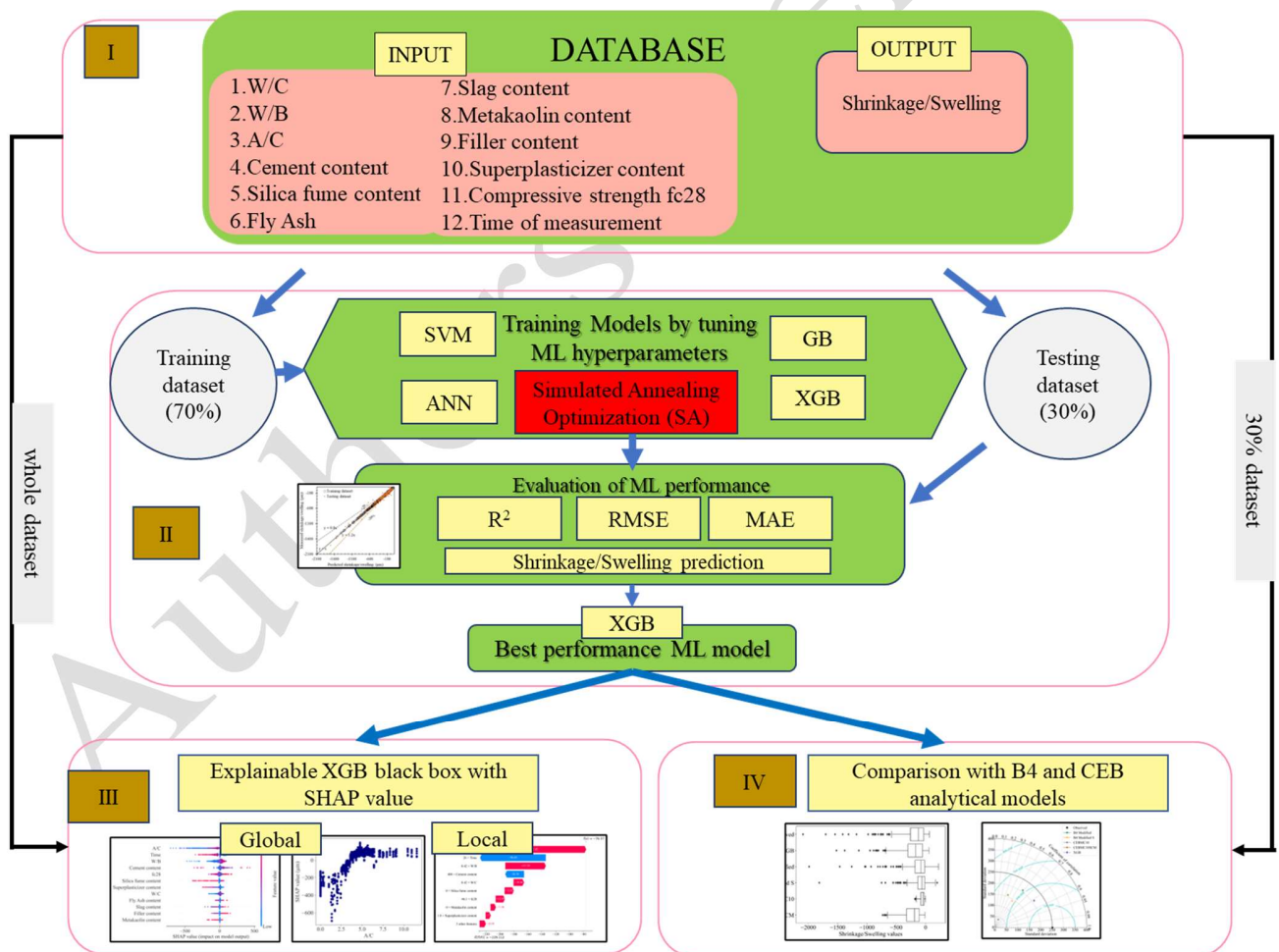
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201

- Step 3: Interpretation of XGB black box with SHAP value

In this step, the SHAPley Additive exPlanations (SHAP) [61] based on the best ML model is used for the interpretation of predicted Shrinkage/swelling values. The effect of each feature is quantified by SHAP global and SHAP local.

- Step 4: Comparison of predicted value based on the best ML model with B4 and CEB analytical models

The concrete formulations in the original database have been divided into a training and a test set to guarantee that no shrinkage value from any concrete with all time of measurement in the test set is known. Then the best Machine Learning model is fitted on this training set, and its performance is compared to the performance of the analytical models on the test set.



202  
203  
204

Fig 3. Conception of shrinkage/swelling investigation

## 205 **3.2 Machine learning models**

### 206 **3.2.1 Support Vector Machine (SVM)**

207 The support vector machine algorithm introduced by Vapnik [62] can be used to address  
208 classification difficulties. For the solution of regression and prediction issues, the support vector  
209 machine was further modified and dubbed support vector machine regression (SVR) [63]. The SVR  
210 method sought to reduce the generalization error and get the best solution using the structural risk  
211 minimization concept. The Gaussian radial basis function was used in this work to transition from  
212 a lower-dimensional feature space to a higher one, allowing for linear regression and improved  
213 fitting of the training dataset. The following linear function can be used to represent the output  
214 variable:

$$f(x) = \langle \omega \cdot \varphi(x) \rangle + b \quad (1)$$

215 where  $\langle \cdot \rangle$  denotes the dot function;  $\varphi(x)$  is a transformation that translates the input feature to a  
216 higher dimensional feature space;  $b$  is the function's parameter vector (model bias);  $\omega$  are the  
217 minimum values derived from the following equation:

$$\text{minimise } \left[ \frac{1}{2} \|\omega\|^2 + C \sum_{j=1}^n (\xi_j + \xi_j^*) \right] \quad (2)$$

$$\text{considering } \begin{cases} y_j - \omega \cdot \varphi(x_j) - b \leq \varepsilon + \xi_j \\ \omega \cdot \varphi(x_j) + b - y_j \leq \varepsilon + \xi_j^* \\ \xi_j, \xi_j^* \geq 0 \end{cases} \quad (3)$$

218 where  $\xi_j, \xi_j^*$  are the slack variables;  $C$  is the regularization parameter;  $\varepsilon$  is the insensitive loss function  
219 (error between the test and forecast values);  $y_j$  is the test result; and  $n$  is the sample size.

220

### 221 **3.2.2 Artificial Neural Network (ANN)**

222 The most widely used machine learning model is ANN, which is based on biological neural  
223 networks. It is made up of input neurons (similar to synapses), multiplied by weights representing  
224 the intensities of the relevant impulses, and then determined by the neuron's activation function.  
225 Finally, the required result will be computed by another function (perhaps an identity function). It  
226 is feasible to acquire the output matching to certain inputs by altering the weight of the artificial  
227 neuron. However, when there are too many neurons, calculating these weights becomes problematic.  
228 At this time, numerous ways are being developed to modify the weights of the neurons in order to  
229 get the output as rapidly as feasible. In this study, the solver for weight optimization contains three  
230 methods including 'LBFGS' Limited-Memory Broyden-Fletcher-Goldfarb-Shanno method of  
231 quasi-Newton methods, 'SGD' refers to stochastic gradient descent and the last 'ADAM' referring

232 to a Stochastic Gradient-Based Optimizer proposed by Kingma et al. [64]. In this investigation, the  
 233 best optimization solver will be selected by the metaheuristic algorithm in the process of tuning  
 234 hyperparameters.

235

### 236 3.2.3 Gradient Boosting (GB)

237 Gradient Boosting (GB) improves predictions by using a method called as augmentation.  
 238 The algorithm in boosting attempts to examine and correct predictions that are not acceptable on the  
 239 first iteration [65]. The algorithm's iterative process for correction then continues until the desired  
 240 prediction is obtained. Increased assistance for decreasing bias and variation in order to strengthen  
 241 poor learners.

242 Gradient Boosting is the process of combining several weak predictive models into a single  
 243 strong model [66]. In this case, a poor prediction model may perform somewhat better than random  
 244 guesses.

245 The model using the Boosting ensemble learning strategy can be expressed as follows:

$$F_N(x) = \sum_{n=1}^M \alpha_n f_n(x) \quad (4)$$

246 The ensemble learning model  $F_N(x)$  is generated by the linear combination of N basic  
 247 learning machine models, where  $f_n(x)$  represents the mth basic learning machine and n represents  
 248 the weight of the nth basic learning machine.

249 The gradient boosting algorithm operates on the gradient descent approach, which is based  
 250 on the functional. To begin, the goal function is as follows:

$$E = \sum_{t=1}^T L(y_t, F_N(x_t)) = \sum_{t=1}^T L(y_t, \alpha_n \times f_n(x_t)) \quad (5)$$

251 The number of samples is T, the number of basic learning machines is N, and the loss  
 252 function is L. The training procedure is to train each basic learning machine  $f_n$ , n-1, 2,..., N in turn  
 253 so that the objective function value E is the smallest, that is, the training outcomes for the nth basic  
 254 learning machine are as follows.

$$\sum_{t=1}^T L(y_t, F_{n-1}(x_t) + \alpha_n \times f_n(x_t)) \cong \sum_{t=1}^T L\left(y_t, F_{n-1}(x_t) + \sum_{t=1}^T \frac{\partial L}{\partial F_{n-1}(x_t)} \times f_n(x_t)\right) \quad (6)$$

255 According to the principle of gradient descent, the optimal model should be the negative  
 256 gradient direction of the objective function, that is:

$$f_n^*(x) = \arg \min_{f, \alpha} \sum_{t=1}^T L \left( -\frac{\partial L}{\partial F_{n-1}(x_t)} - \alpha \times f(x_t) \right)^2 \quad (7)$$

257 After obtaining the best A, coefficient S can be found by minimizing the objective  
 258 function, namely:

$$\alpha_n^* = \arg \min_{\alpha_n} \sum_{t=1}^T L(y_t, F_{n-1}(x_t) + \alpha_n \times f_n^*(x_t)) \quad (8)$$

259 The preceding steps are repeated until all N basic learning machines and coefficients have  
 260 been trained, and a learning model based on Equation 1 has been generated. The gradient boosting  
 261 approach is used in the learning model, and the next fundamental machine learning model may  
 262 further increase the accuracy based on the previous underlying machine learning models, reducing  
 263 the model's deviation. This is the distinction with the Bagging. Basic machine learning can be a low-  
 264 accuracy model; through the improvement approach, the accuracy of the entire model can be  
 265 steadily increased, but there is a certain degree of fit at the same time.

266

### 267 3.2.4 Extreme Gradient Boosting (XGB)

268 The Extreme Gradient Boosting (XGB) algorithm is a cutting-edge method for solving  
 269 supervised learning problems with high precision [26]. If Deep Learning only accepts raw data in  
 270 numerical form (which must typically be converted to n-vectors in the real number space), XGB  
 271 accepts tabular datasets of any size, including categorical data. Furthermore, XGB has a rapid  
 272 training speed, can scale for parallel computing on several servers, and can be GPU-accelerated,  
 273 therefore Big Data is not an issue for this model. The earliest and most well-known model of the  
 274 most recent generation of tree booster ensembles is the XGB; a representation of this approach is  
 275 the Gradient Boosting Decision Tree (GB) or Multiple Additive Regression Tree (MART).

276 The basic idea behind XGB is to first train a tree with the training set and the real value of  
 277 the sample (i.e. the standard response), and then use that tree to forecast the training set to acquire  
 278 the proper value. The residuals are the predicted value of each sample, as well as the difference  
 279 between the forecast and the real value. The second tree is then trained, but this time instead of using  
 280 the real number, use the residual as the standard response. After training two trees, the residuals of  
 281 the sample can be used to train more trees. To stop training, the total number of trees can be selected  
 282 explicitly or based on particular indicators.

283 The XGB algorithm can be thought of as a K-tree additive model:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (9)$$

284 The gradient boost model does not learn the weights of the d-dimensional space but directly  
 285 learns the decision tree set, where F is a function space consisting of all trees (here the regression  
 286 tree is also a partial function and the varied values of the different portions form a tree). The  
 287 objective function of XGB is a sum of a particular loss function assessing overall predictions and a  
 288 sum of all predictors' regularization terms (K trees).

289 Mathematically:

$$obj(\theta) = \sum_i^n l(y_i - \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (10)$$

290 Where  $l$  is the loss function, which is the difference between the expected and actual outputs  $\hat{y}_i$  and  
 291  $y_i$ . While  $\Omega$  is a measure of how complicated the model is, it aids in preventing model over-fitting and  
 292 is determined using:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (11)$$

293 The number T in the preceding equation reflects the number of leaves on the tree, and w  
 294 represents the weight of each leaf.

295

### 296 3.2.5 Simulated Annealing Optimization algorithm (SA)

297 Kirkpatrick et al. [67] pioneered simulated annealing (SA) in 1983. This metaheuristic  
 298 method is based on physical phenomena seen in continuum mechanics. To avoid becoming stuck in  
 299 local minima, SA employs a random search engine represented by Markov chains. SA employs  
 300 solution variants to increase the objective function while still maintaining the solutions, albeit the  
 301 least efficient, within specified constraints. Any adjustment that improves the objective function, J,  
 302 is permitted in the minimization problem; however, candidates that raise J are also included in the  
 303 solution set with probability p, which is known as the transition probability [68]:

$$p = \exp\left[-\frac{\Delta E}{k_B T}\right] \quad (12)$$

304 where  $k_B$  is Boltzmann's constant, T is the annealing process temperature, and E is the energy level,  
 305 is connected with the goal function modification by the constant, [68].

$$\Delta E = \gamma \Delta J \quad (13)$$

306 A random threshold,  $r$ , such that  $p > r$ , determines whether a solution modification is  
 307 acceptable.  $T$  lowers and the system cools down during optimization. A high  $T$  indicates a high-  
 308 energy system that is difficult to reduce, whereas a low  $T$  indicates a low-energy system that may  
 309 be trapped in a local minimum [69]. The proper decision  $T$  may be to first search the space for the  
 310 global minimum and then, as the system cools down, converge to a solution with sufficient accuracy  
 311 [68]. Stopping conditions, such as limiting the number of repeats or focusing on the improvement  
 312 of the goal function, are incorporated in  $t_f$  [70].

313

### 314 3.2.6 Performance evaluation of models

315 In order to evaluate the performance of ML models, three popular statistical metrics, such as  
 316 coefficient of determination  $R^2$ , Root Mean Square Error (RMSE), and Mean Absolute Error  
 317 (MAE), are introduced in following:

$$R^2 = \frac{\sum_{j=1}^N (p_{0,j} - \bar{p}_0)(p_{t,j} - \bar{p}_t)}{\sqrt{\sum_{j=1}^N (p_{0,j} - \bar{p}_0)^2 \sum_{j=1}^N (p_{t,j} - \bar{p}_t)^2}} \quad (16)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{j=1}^N (p_{0,j} - p_{t,j})^2} \quad (17)$$

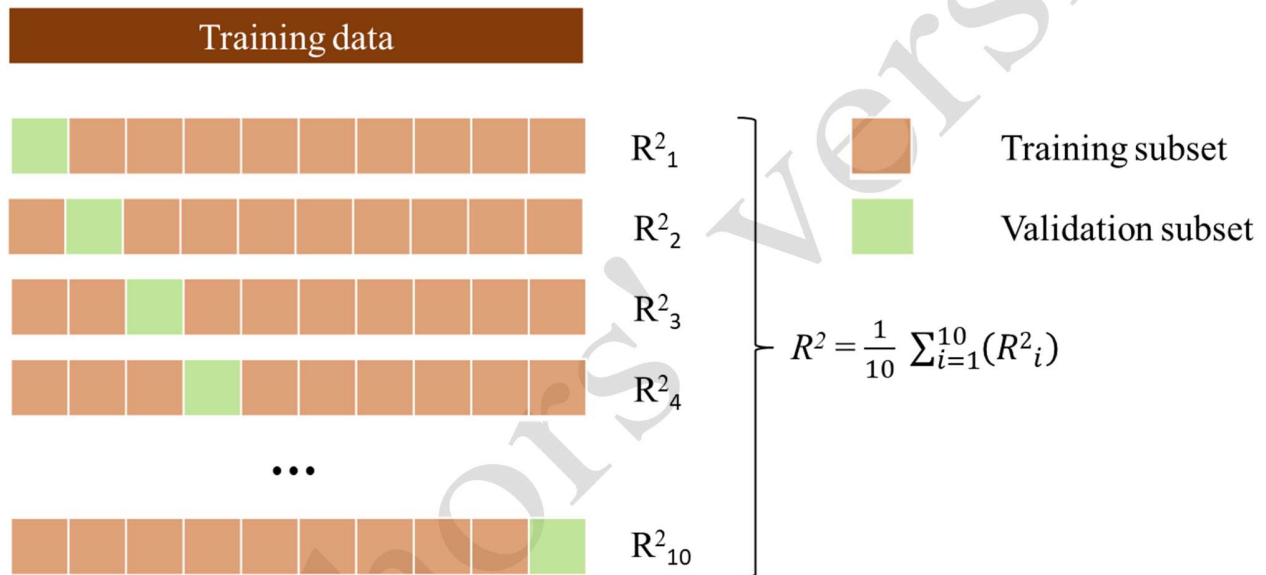
$$\text{MAE} = \frac{1}{N} \sum_{j=1}^N |p_{0,j} - p_{t,j}| \quad (18)$$

318 Where  $p_{0,j}$  is the shrinkage/expansion value of the  $i$ -th sample point in the database;  $p_{t,j}$  is  
 319 the machine learning model prediction value for the  $i$ -th sample point;  $\bar{p}_0$  is the averaged  
 320 experimental value of shrinkage/expansion; and  $\bar{p}_t$  is the mean predicted value. Both MAE and  
 321 RMSE explicitly quantify the residual error at each sample point and thus provide an accurate  
 322 assessment of model performance.  $R^2$  normalizes the squared residual error with the database  
 323 variance and generates dimensionless values ranging from 0 to 1. Because RMSE is often seen to  
 324 be more understandable because it is equivalent to measured values and handy for comparing the  
 325 performance of different models, it was chosen as the primary metric index.

326 **3.2.7 K-Fold Cross Validation technique**

327 To tune hyperparameters in machine learning algorithms, K-fold cross-validation (CV) is  
 328 utilized. Choosing k affects the variance; a low k value results in significant variance or bias.  
 329 Furthermore, while deciding on k, the data quantity and computational capabilities of the system  
 330 must be taken into account. The value of k in this study is set to 10 [71,72], indicating that the number  
 331 of training and testing rounds is ten. The dataset is split into two parts: training and testing. The  
 332 training data area is partitioned into k sections of equal size. The model is then trained with different  
 333 hyperparameters on k-1 subsets and evaluated on the last subset. The process is repeated k times  
 334 before averaging the evaluation results as illustrated in Figure 4.

335



336

337 **Fig 4.** 10-Fold CV of training dataset with cost function of  $R^2$  coefficient of determination

338

339 **3.3 Comparison with Analytical Models**

340 Three analytical models are also adopted for comparison to examine the prediction accuracy  
 341 of the machine learning models detailed above. These models are the B4 model [10], CEB MC 10  
 342 [9] model and a modified version of CEM MC 10 model to include the effect of SCM which has  
 343 been introduced recently [15] and will be referred to as 'CEBMC10SCM' model in this study. The  
 344 autogenous shrinkage model prediction are given as follows:

- 345 • Modified B4 model [73]:

346 Using this composition-based model, the autogenous shrinkage can be calculated at any time:



$$\epsilon_{au}(t) = k_{\gamma} k_s C t^n \quad (19)$$

347

348 where  $k_{\gamma}$  depends on the cement type,  $k_s = \left\{ 1 + 3 \left( \frac{SiO_2}{Cement} \right) \right\} \left\{ 1 + 2 \left( \frac{Slag}{Cement} \right) \right\}$  takes into

349 account the cement substitution by silica fume or slag, the prefactor  $C$  depends on the water-to-

350 cement and aggregate-to-cement ratios :  $C = \frac{100}{(w/c)^{2.5} + \left(\frac{a/c}{10}\right)^{1.5}}$  and the exponent factor  $n$ , close to 0.2,

351 reads  $n = 1.2 - 0.1(a/c) + (-0.14 + 0.005(a/c)) \ln C$ .

352 • Modified B4 S model [73]:

353 The autogenous shrinkage value can be calculated using the 28d compressive strength

354 without any information regarding the formulation and reads:

$$\epsilon_{au}(t) = 12 f_c (1 - g)^{1.7} t^{0.2} \quad (20)$$

355 where  $f_c$  is the 28d compressive strength and  $g = (\text{volume of aggregates}) /$

356  $(\text{volume of concrete})$  (default values of 0.7 and 0.0 can be used for concrete and cement pastes

357 respectively).

358 • CEB MC 10 model:

$$\epsilon_{cas}(t) = \epsilon_{cas0}(f_{cm}) \cdot \beta_{as}(t) \quad (21)$$

359 where  $\beta_{as}(t) = 1 - \exp(-0.2\sqrt{t})$  the time function of autogenous shrinkage and

360  $\epsilon_{cas0}(f_{cm}) = -\alpha_{as} \left( \frac{0.1 f_{cm}}{6 + 0.1 f_{cm}} \right)^{2.5} \cdot 10^{-6}$  the notional autogenous shrinkage coefficient, where  $\alpha_{as}$

361 is a coefficient depending on cement type and  $f_{cm}$  is the 28-d compressive strength.

362 • CEB MC 10 modified for SCM consideration [15]:

$$\epsilon_{cas}(t) = k_{SCM} \cdot \epsilon_{cas0}(f_{cm}) \cdot \beta_{as}(t) \quad (22)$$

363 where  $k_{SCM}$  is a correction factor to included the impact of SCM which, in the case of

364 autogenous shrinkage, reads  $k_{SCM} = (1 + 0.24f + 0.4m - 0.35s) \cdot 4M$  where  $f$  is the replacement

365 ratio of fly ash (%),  $m$  the replacement ratio of calcined clay (%),  $s$  the replacement ratio of slag

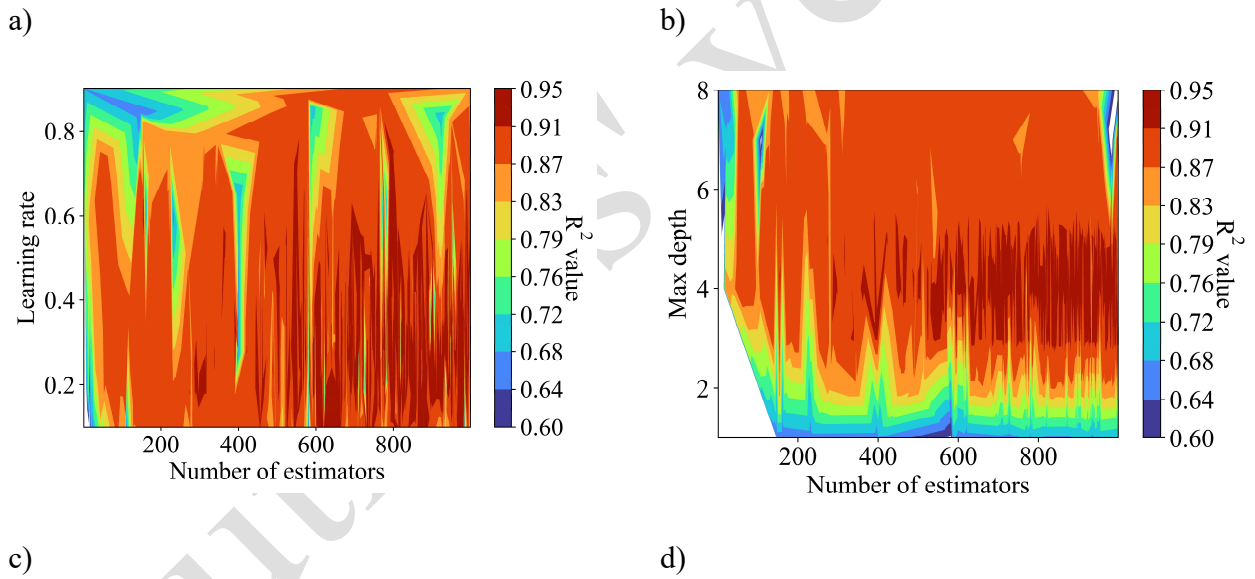
366 and  $M$  the multi-SCM coefficient (binary  $M = 0.95$ , single  $M = 1.0$ ).

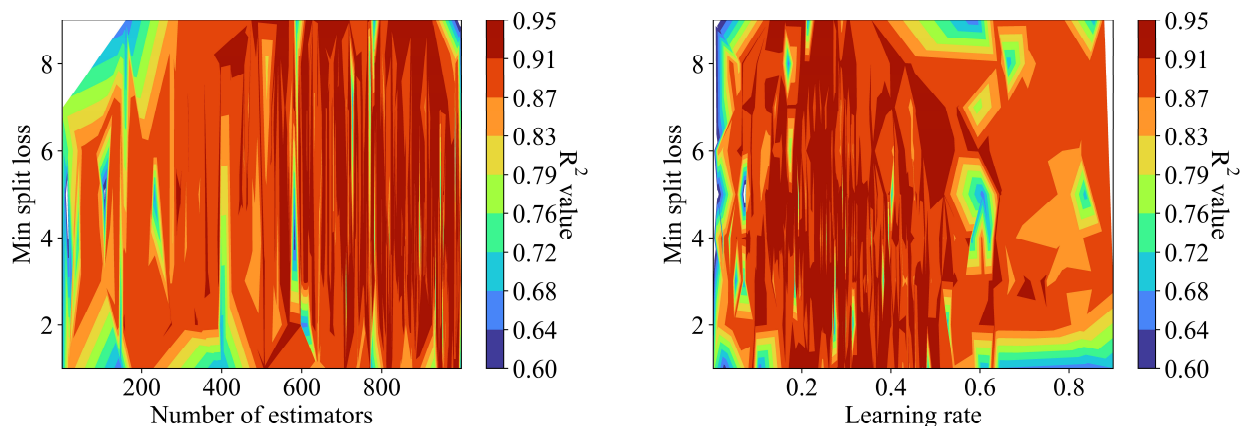
367

368 **4 Results and Discussion**

369 **4.1 Models hyperparameters tuning**

370 The ML model parameters have been tuned using the Simulated Annealing (SA) optimization  
371 to obtain the most performant model considering  $R^2$  metric. Surface plots of the  $R^2$  values for  
372 several hyperparameter values of the XGB model are reported in Fig. 5. As illustrated in these  
373 figures, extended zones of optimized parameter values could be found in most cases. For example,  
374 as illustrated in Fig. 5 a),  $R^2$  values higher than 0.91 could be obtained for learning rates lower than  
375 0.5 and more than 500 iterations. A maximum depth of 4 led to the best results as illustrated in  
376 Fig. 5 b) while various values of the minimum loss reduction required to make a partition on a leaf  
377 node of the tree (min split loss) could generate good results, as illustrated in Fig. 5 c) and d). Based  
378 on this optimization analysis carried out by SA, the best hyperparameters values of the four ML  
379 models were selected. These values are reported in Table 2.





380 **Fig 5.** Tuning XGBoost hyperparameters: surface plots of  $R^2$ -values using 10-fold cross-validation.

381

382 **Table 2.** Tunned hyperparameters of the ML models.

<b>XGB</b>	Nb of trees	Learning rate	Max depth	Min split loss			<b><math>R^2</math></b>
	1197	0.6680	10	4			<b>0.9311</b>
<b>GB</b>	Nb of trees	Learning rate	Max depth	Max features	Min samples split	Min samples leaf	<b><math>R^2</math></b>
	887	0.2740	10	10	0.0560	0.0580	<b>0.9284</b>
<b>SVM</b>	C	$\gamma$	$\epsilon$	Kernel			<b><math>R^2</math></b>
	1491	0.3610	0.0610	RBF			<b>0.8089</b>
<b>ANN</b>	Nb of neurons	Activation	Solver	Max iteration			<b><math>R^2</math></b>
	(19,9)	RELU	LBFGS	4000			<b>0.8054</b>

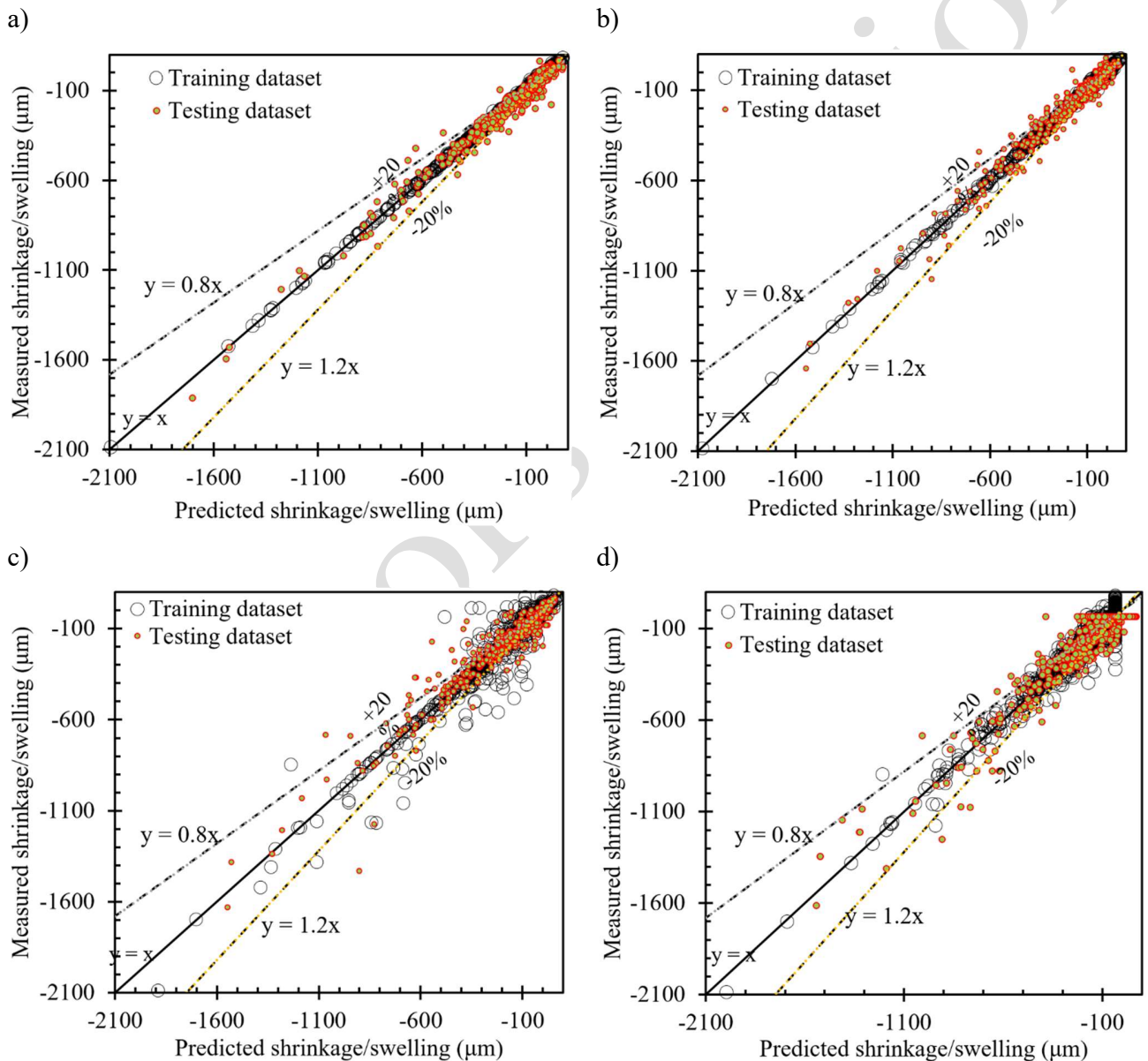
383

#### 384 4.2 Machine learning models typical predictions

385 The results obtained by the four models on the training and test sets are displayed in Fig. 6. As  
 386 illustrated in the figure, all models performed well on the training sets with  $R^2$  values higher than 0.94.  
 387 Most training set predictions fall within the interval between 0.8 times and 1.2 times the measurement  
 388 shrinkage value, except for very small shrinkage values. However, Gradient Boosting and XGBoost  
 389 algorithms outperformed SVM and ANN models. The first two models provided high-quality predictions  
 390 for various shrinkage values, while the latter two models were relatively prone to errors, even for small

391 shrinkage values. The detailed metrics corresponding to the model results are reported in Table 3. XGB  
 392 model performed better than the other models, and the  $R^2$  value associated with its prediction of the test  
 393 set was 0.9733, which corresponds to an RMSE of around  $38.9 \mu\text{e}$ .

394 The mean absolute error (MAE) distribution has been analyzed for the four models, as illustrated in  
 395 Fig. 7. It can be seen that the MAE remained limited for the XGB model as compared to the SVM and  
 396 ANN models. Moreover, the maximum MAE measured for the XGB model, close to  $200 \mu\text{e}$  was almost  
 397 twice as small as the maximum MAE obtained by the SVM and the ANN model.

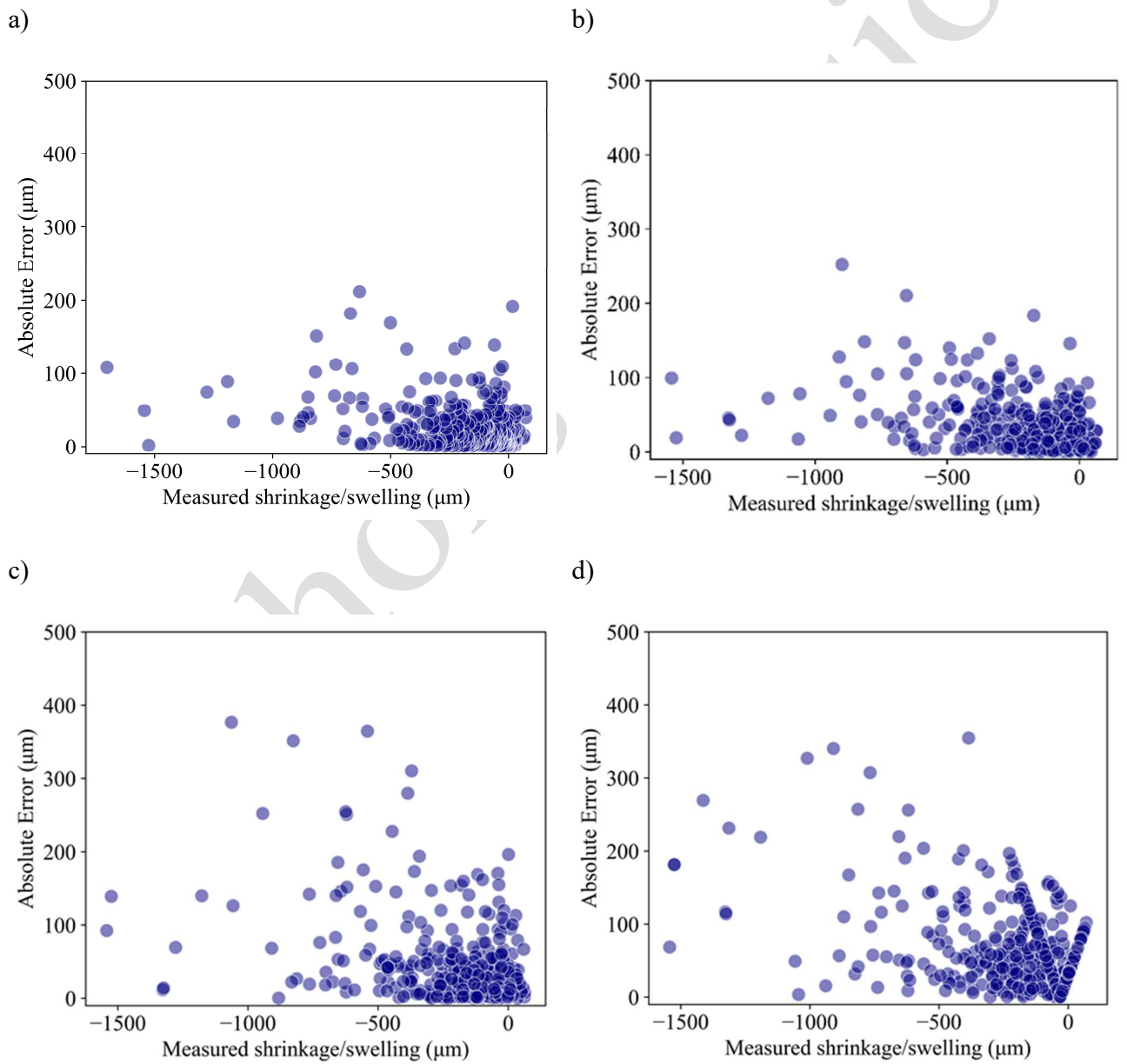


398 **Fig. 6.** Shrinkage/Swelling predicted by different ML models: a) XGB, b) GB, c) SVM and d) ANN  
 399 models

400 **Table 3.** Comparison of machine learning algorithms for autogenous shrinkage prediction of  
 401 concrete incorporating SCM.

	Train			Test		
	R <sup>2</sup>	RMSE (μϵ)	MAE (μϵ)	R <sup>2</sup>	RMSE (μϵ)	MAE (μϵ)
XGB	0.999	6.341	3.993	0.973	38.909	25.249
GB	0.998	7.816	4.620	0.965	45.797	31.039
SVM	0.941	57.085	21.974	0.915	71.775	40.764
ANN	0.949	50.695	37.303	0.910	79.288	56.594

402



403 **Fig 7.** Distribution of the absolute error: a) XGB, b) GB, c) SVM, and d) ANN models.

## 404 4.3 Feature importance and sensitivity analysis

### 405 4.3.1 Global interpretation

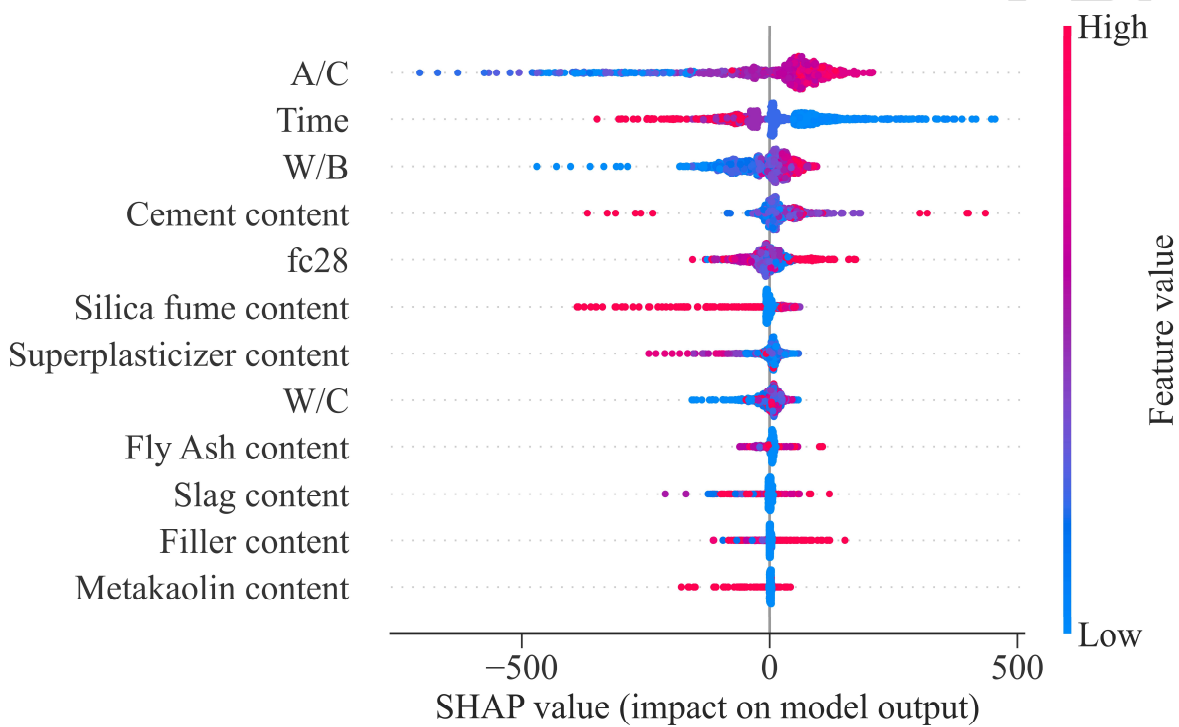
406 The most influential features on the shrinkage predictions have been obtained from the model  
407 using SHAP. The global SHAP values are reported in Fig. 8, the features classified in descending  
408 order based on their influence from top to bottom. The results align with the previous study  
409 published by the authors dedicated to autogenous shrinkage prediction of cementitious materials  
410 incorporating SAP and SCM even if the database has been largely extended and the model results  
411 improved [44]. Indeed, the most influential parameters are the A/C ratio, time, W/B ratio, cement  
412 content, the 28d compressive strength, silica fume content, superplasticizer content, W/C ratio, fly  
413 ash, slag, filler, and last calcined clay content. The relative influence of the most influential  
414 parameters can still be easily determined:

- 415 • increasing A/C ratio increases SHAP value, e.g., decrease shrinkage
- 416 • increasing time values correspond to higher shrinkage values
- 417 • increasing W/B and W/C ratios tend to decrease shrinkage
- 418 • high silica fume replacement ratio or cement content predominantly induces higher  
419 shrinkage

420 However, the global influence of the other parameters, mainly SCM other than silica fume,  
421 might not be that clear at first sight. Most fly ash highest values are associated with positive SHAP  
422 values, which means that a high fly ash content globally decreases shrinkage. Similarly, the highest  
423 filler contents are associated with positive SHAP values, i.e., decreased shrinkage. Since the  
424 highest calcined clay contents are associated with the most negative SHAP values, its adverse  
425 effect on autogenous shrinkage can be confirmed for high replacement ratios. Last, slag and 28d  
426 compressive strength effects cannot be analyzed easily from this global perspective. To this end,  
427 more detailed insights are needed using the features dependence plots.

428 SHAP can also give information about the feature dependency of the results, as illustrated in  
429 Fig. 9. Indeed, feature dependence plots reveal how the SHAP values change depending on the  
430 values of the features. As illustrated in Fig. 9 a), the SHAP values largely increase when the A/C  
431 values increase, especially when A/C values increase from around 1.0 to 4.0, corresponding to the  
432 transition from mortar to concrete. This increase of around 800  $\mu\epsilon$  highlights the relative influence

433 of the A/C ratio on the autogenous shrinkage phenomenon. For A/C values higher than 4.0, SHAP  
 434 value is almost constant, meaning there is no interest in further increasing the A/C ratio to reduce  
 435 the autogenous shrinkage. The water-to-binder ratio's influence can be analyzed using Fig. 9 b).  
 436 From this figure, it can be seen that the SHAP value increases from the minimum w/b considered  
 437 in this study, e.g. 0.16, to approximately 0.5. This gradual increase from around  $-300 \mu\epsilon$  to around  
 438  $50 \mu\epsilon$  confirms that autogenous shrinkage arises in cementitious materials of W/B ratio smaller  
 439 than 0.5. However, it is worth noting that the W/B relative influence is around twice as small as  
 440 the A/C influence.



441  
 442 **Fig. 8.** Global SHAP values of the tuned XGB model

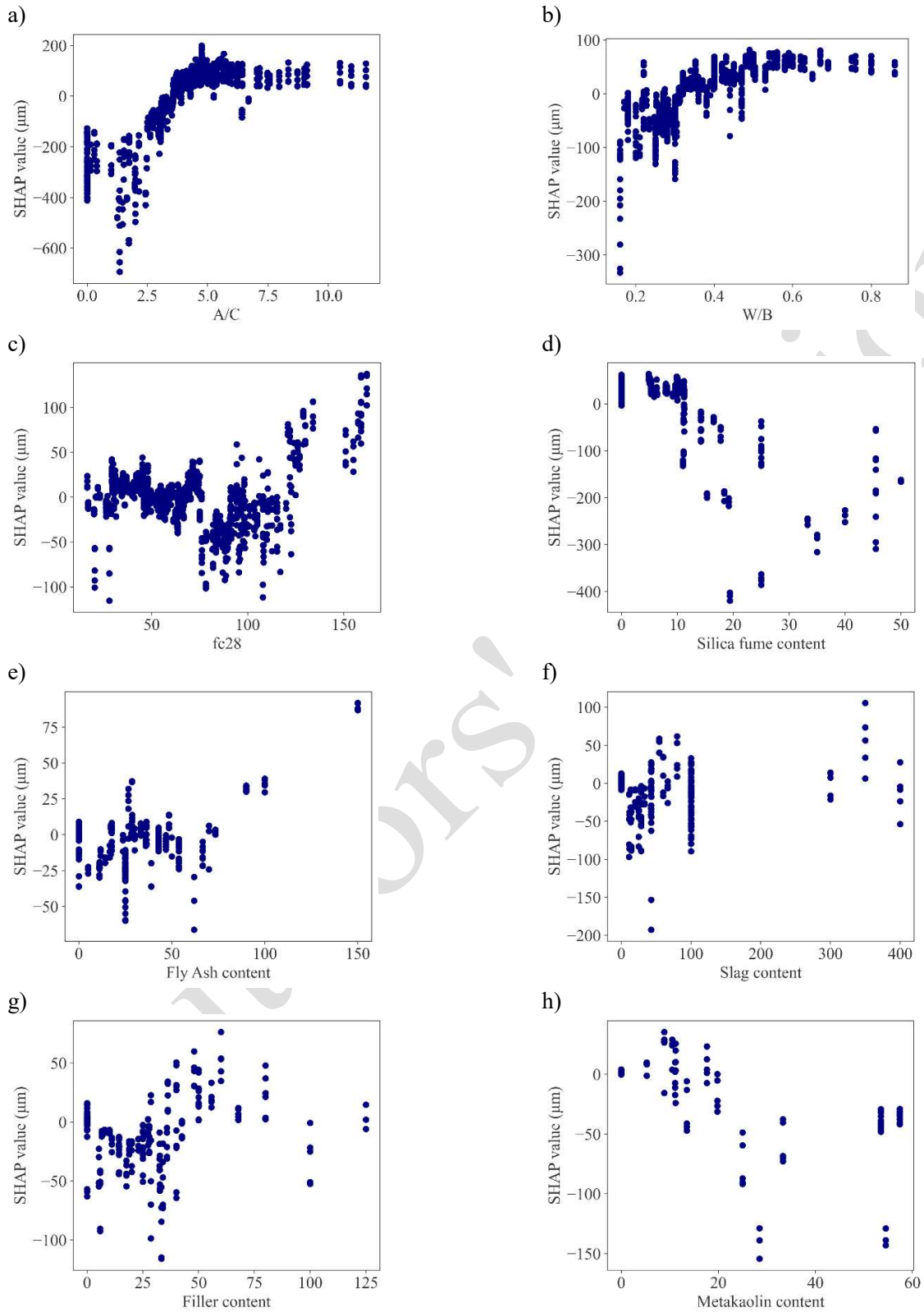
443 Based on Fig. 9 c), it can be seen that the increasing strength values up to around 100MPa  
 444 negatively influence SHAP values, which means that transitioning from normal strength to high  
 445 strength concrete leads to more autogenous shrinkage, independently from the other parameters  
 446 such as the W/B ratio. Conversely, SHAP values increase for strength higher than around 100MPa  
 447 up to final values around  $140 \mu\epsilon$ , which means that autogenous shrinkage for this type of high  
 448 performance and ultra-high performance concrete would have been more pronounced if their  
 449 strength had been smaller. This novel insight might correlate to the quick strength and Young's  
 450 modulus increase of these concretes, limiting early age autogenous shrinkage.

451 The influence of silica fume can be analyzed using Fig. 9 d). Based on this figure, it can be  
452 observed that increasing silica fume values, from 0% to 50% cement replacement, largely  
453 decreases SHAP values. However, a spread should be noticed when the silica fume replacement  
454 ratio is larger than 20%, and it should be noticed that a 50% replacement ratio can influence SHAP  
455 values less than a 20% replacement ratio in the case of some concretes. Therefore, although silica  
456 fume is known to increase considerably autogenous shrinkage, novel routes might be found to limit  
457 its influence. The other supplementary cementitious materials have a smaller influence on SHAP  
458 values. As illustrated in Fig. 9 e), fly ash influence on SHAP values becomes visible on this feature  
459 dependence plot for replacement ratios higher than 50%. In this case, cement substitution by fly  
460 ash leads to smaller autogenous shrinkage values explained by the slower hydration. Based on Fig.  
461 9 f), it can be observed that slag influence on SHAP values is relatively limited as most of the  
462 values belong to the  $[-100 \mu\epsilon, 50 \mu\epsilon]$  interval. However, it is worth noting that, according to this  
463 detailed analysis, slag negatively influences shrinkage for all the substitution ratios smaller than  
464 around 50%, which agrees with the literature [19,74]. On the opposite, different behaviors can be  
465 observed for replacement ratios higher than 50%: slag can positively affect shrinkage sometimes,  
466 while it still negatively influences shrinkage for other concrete mixes. It can be hypothesized that  
467 two competing interests are at stake: the refined pore structures that increase shrinkage and the  
468 slower hydration that decrease shrinkage at a given time. As illustrated in Fig. 9 g), filler globally  
469 slightly increases shrinkage for substitution rates smaller than 25%, which can be explained by the  
470 increased hydration speed. However, substitution ratios higher than 25% systematically positively  
471 influence SHAP values, highlighting the restraining effect of filler on the autogenous shrinkage.

472 Finally, though the effect of calcined clay is relatively small according to the model, it can be  
473 observed that calcined clay does not influence SHAP values a lot for replacement ratios smaller  
474 than 20%, while increasing the substitution rate, the SHAP values decrease. This observation  
475 agrees with the literature [24,58]. To conclude, since SHAP gives the influence rules of the  
476 different features on the autogenous shrinkage predictions of the model, it can be used to provide  
477 novel insights into the shrinkage phenomenon.

478





479 **Fig. 9.** Feature dependence plots associated with SHAP for the influential features: a) A/C, b)  
 480 W/B, c) fc28, d) silica fume, e) fly ash, f) slag, g) filler, and h) calcined clay.

### 481 4.3.2 Local interpretation

482 Besides global interpretations, SHAP also provides useful information to explain the model  
483 predictions on individual samples as SHAP decomposes the final prediction into a sum of  
484 contributions by all the input variables. Various 28d autogenous shrinkage predictions were thus  
485 analyzed for several unknown concrete samples containing various types of SCM. Fig. 10  
486 illustrates explanations for four typical samples.

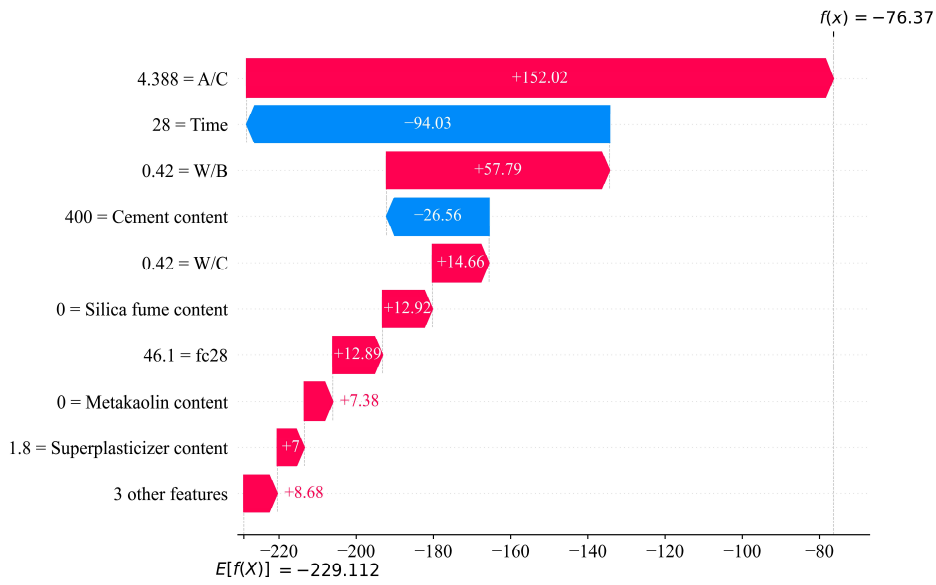
487 As shown in Fig. 10 a), the base prediction on the entire set is  $-229 \mu\epsilon$ . For this sample, the  
488 most significant reason for the final autogenous shrinkage prediction of  $-76 \mu\epsilon$  is the A/C ratio,  
489 whose individual contribution amounts to  $+152 \mu\epsilon$ , meaning this value decreases shrinkage  
490 compared to lower A/C ratios. Then time negatively influences the autogenous shrinkage  
491 prediction, which can be explained by the fact that 28 days is the final prediction associated with  
492 larger shrinkage deformations. The water to binder ratio is the third most influential parameter, as  
493 expected based on the global interpretation.

494 For concrete incorporating silica fume, its relative influence on the outputs can be quantified  
495 as illustrated in Fig. 10 b). The presence of silica fume can considerably influence the model  
496 prediction. Indeed, as illustrated in the figure, silica fume content becomes the most influential  
497 parameter, and, at 28 days, the SHAP value associated to silica fume is around  $-160 \mu\epsilon$ .

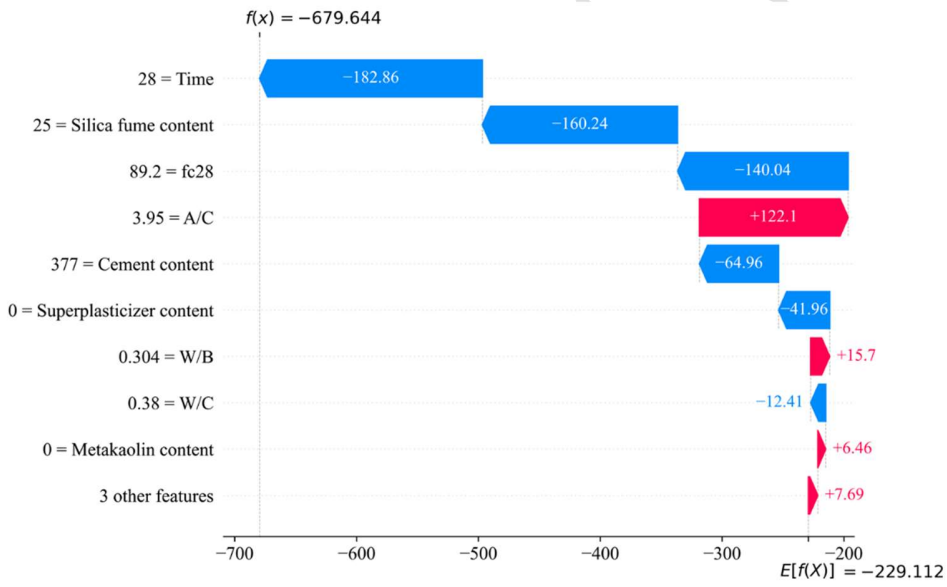
498 In most analyses, small replacement ratios of other SCM were found only slightly to affect  
499 shrinkage values as the A/C and the W/B ratios predominantly influenced the latter, as illustrated  
500 in Fig. 10 c) represents the inputs contributions on the autogenous shrinkage prediction of a  
501 concrete sample incorporating fly ash with the mass of fly ash equal to 11.1% of the one of cement.  
502 However, when larger proportions of cement are replaced, the influence of the SCM can become  
503 predominant. Indeed, as illustrated in Fig. 10 d), when half of the cement is replaced by slag, this  
504 slag proportion increases shrinkage by around  $52 \mu\epsilon$  according to the model.

505

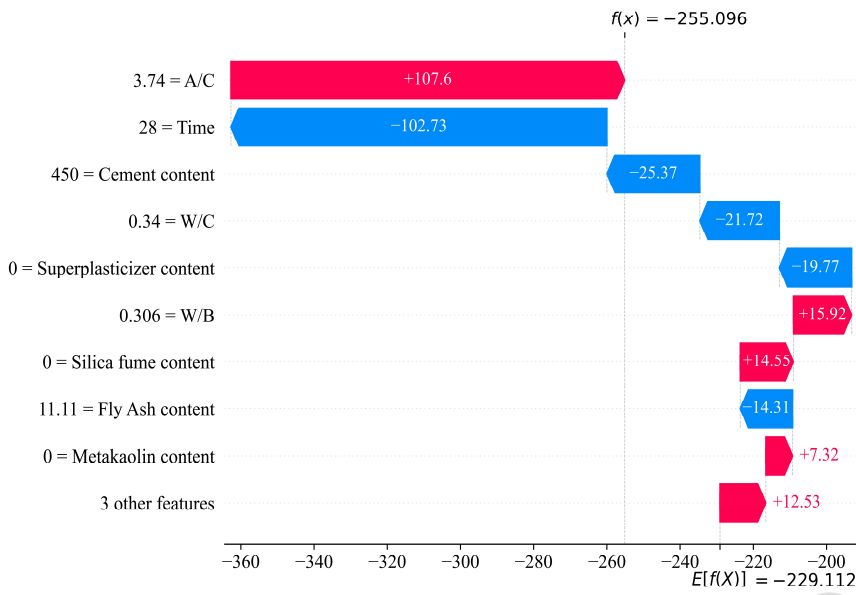
a)



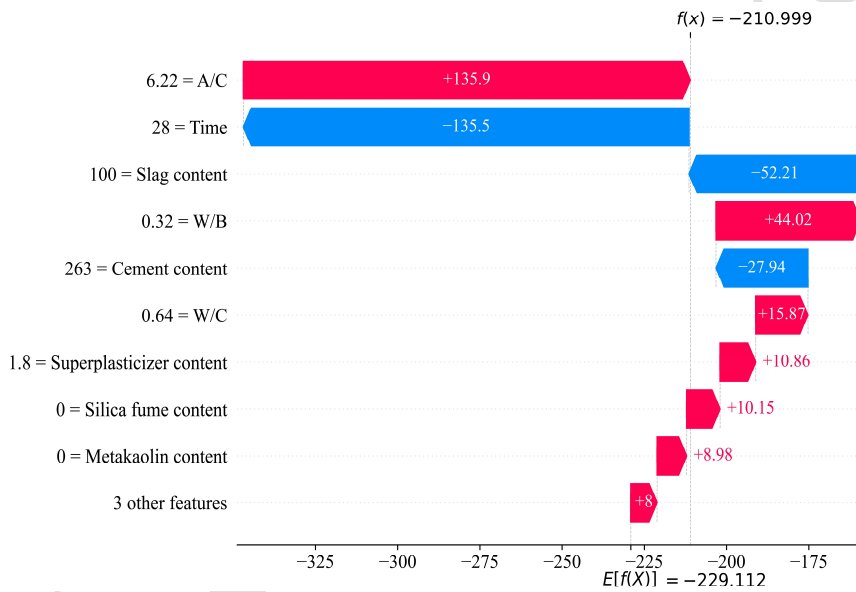
b)



c)



d)



506 **Fig. 10.** Local interpretation using SHAP of some of the XGB model autogenous shrinkage  
 507 predictions: a) normal strength concrete without SCM, b) high strength concrete with silica fume,  
 508 c) normal strength concrete with fly ash, d) high-strength concrete with 50% of cement replaced  
 509 by slag.

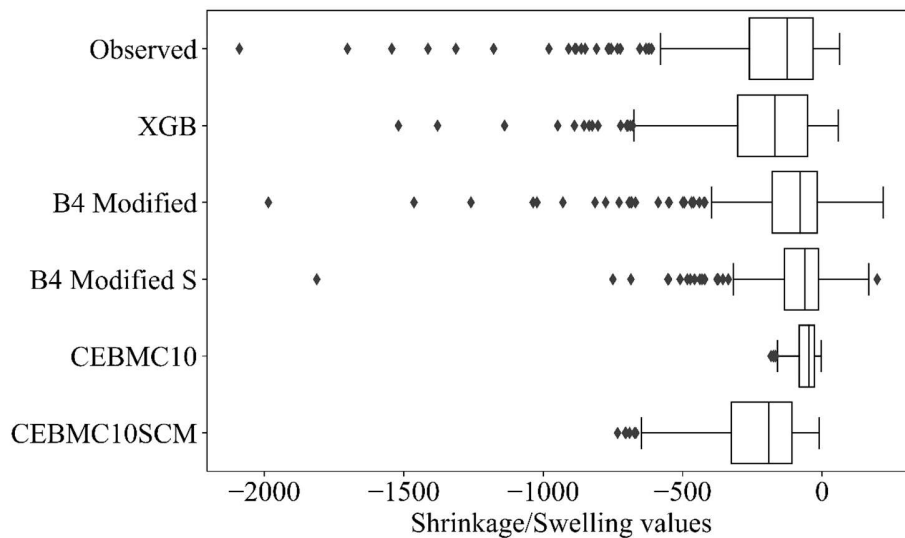
510

#### 511 4.4 Comparison with B4 and CEB analytical models

512 The developed XGBoost model, the modified B4 and B4s, the CEB MC10, and the CEB MC

513 10 model's modified version can be compared globally using their respective predictions on the  
514 test set composed of unknown concrete formulations. The performance of these four models has  
515 been compared with the observed shrinkage values, as illustrated in Fig. 11. It can be observed  
516 that, overall, the distribution of the shrinkage values predicted by the developed XGB model is in  
517 close agreement with the observed distribution, and the box plots are very similar. Indeed,  
518 interquantiles ranges and outliers present similar patterns. The XGB model slightly overestimates  
519 the overall shrinkage values, which is good regarding safety issues. Last but not least, the XGB  
520 model can predict small swelling values that might occur at an early age, while analytical models  
521 cannot be captured correctly this trend.

522 The analytical models' performance considerably varies. Based on the overall distribution of  
523 the shrinkage values, the modified B4 model provides a distribution of the predictions that is in  
524 good agreement with the observed distribution. However, one can observe that the mean shrinkage  
525 value predicted by the modified B4 model slightly underestimates the observed shrinkage  
526 distribution, which might represent a risk regarding safety. The same conclusion applies to the  
527 modified version of the B4 S model. Moreover, the predictions made by this model cannot capture  
528 the experimental variability, probably due to its simplicity. CEB MC 10 model results clearly  
529 underestimate shrinkage, and the size of the overall distribution is considerably smaller than the  
530 observed distribution, highlighting the shortcomings of this analytical model. The modified  
531 version of the CEB MC 10 model performed better, and the overall distribution of the shrinkage  
532 values looks like the observed distribution as the interquartile range is close to the experimental  
533 one. However, an apparent overestimation of the shrinkage values and a disability to predict large  
534 shrinkage values can be noticed.



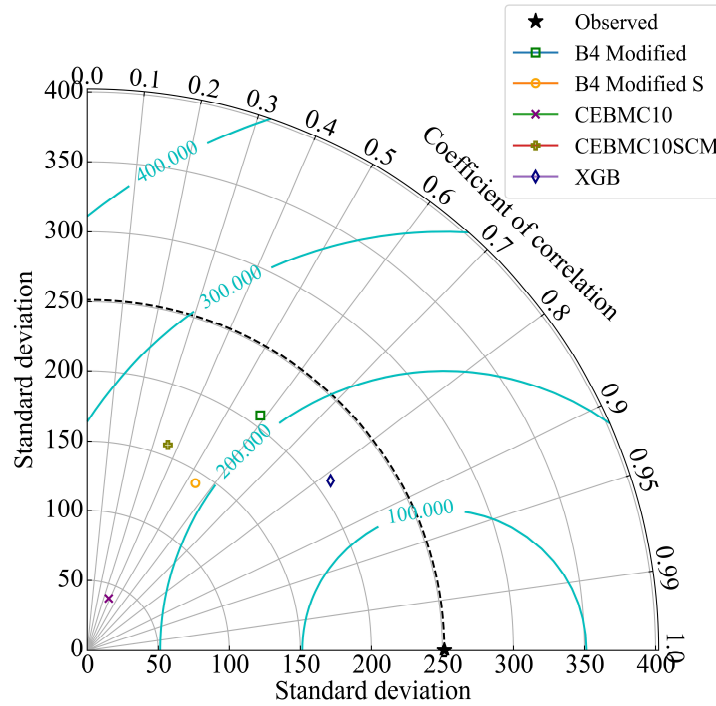
535

536 **Fig. 11.** Box plots of the shrinkage/swelling values predicted by the XGB model compared to  
 537 the analytical models and the observed values

538

539 The model results and the test data can be further compared using the Taylor diagram (Fig.  
 540 12), which summarizes three valuable characteristics of the models compared to the measured  
 541 values: the standard deviation associated with the model's predictions, the correlation between the  
 542 predictions and the experimental values, and the centered root-mean-square (RMS) difference. As  
 543 mentioned in the previous paragraph, the standard deviation of the simulated patterns reported on  
 544 the X-axis largely differs between the models. The standard deviation of the experimental data is  
 545 almost reproduced by the XGB and the modified version of the B4 model. Indeed, the standard  
 546 deviations associated with these models are higher than  $200 \mu\epsilon$ , close to the  $250 \mu\epsilon$  experimental  
 547 value. Then, the modified B4 S and modified version of the CEB MC 10 model exhibit a standard  
 548 deviation of around  $150 \mu\epsilon$ , which is in relatively good agreement with the observed values. Last,  
 549 the standard deviation of the classic CEB Model 10 is smaller than  $40 \mu\epsilon$ , clearly underestimating  
 550 the experimental variation and highlighting the model's shortcomings in predicting shrinkage of  
 551 concrete incorporating SCM. Concerning the correlation of the models with the experimental  
 552 values, the best correlation is obtained by the XGB model (0.82). This correlation is largely higher  
 553 than the second-best correlation obtained by the modified B4 model (0.58). The correlation of the  
 554 modified B4 S model follows. The CEB MC 10 and the modified version of the CEB MC 10 model  
 555 exhibit correlations close to 0.3-0.4, which is relatively small. Finally, the RMS value of the XGB

556 model (around  $160 \mu\epsilon$ ) is clearly better than the ones of the other models, which are higher than  
 557  $200 \mu\epsilon$ . For all these reasons, it can be concluded that the optimized XGB model significantly  
 558 outperforms the best analytical models that have been proposed up to date.



559  
 560 **Fig. 12.** Taylor diagram representing XGB, B4 and B4s modified, CEB MC 10 and CEB MC  
 561 10 modified version models performance.

562  
 563  
 564 Based on the XGB model, the autogenous shrinkage predictions of the eight concrete  
 565 compositions containing various types of SCM have been compared to the analytical models'  
 566 predictions, as shown in Fig. 13. The detailed formulations are given in Table 4. As illustrated in  
 567 these figures, the prediction accuracy of the XGB model globally outperforms that of the analytical  
 568 models in most cases which agrees with Taylor's analysis.

569  
 570  
 571  
 572  
 573

574 **Table 4.** Mortar and concrete compositions of some typical test samples.

Ref	C1	C2	C3	C4	C5	C6	C7	C8
W/C	0.42	0.38	0.34	0.64	0.29	0.43	0.31	0.44
W/B	0.42	0.304	0.306	0.32	0.28	0.3	0.21	0.3
A/C	4.39	3.95	3.74	6.22	4.21	3.68	3.94	3.74
Cement (kg/m <sup>3</sup> )	400	377	450	263	428	408	497	416
Silica fume (%)	-	25	-	-	-	14.2	10	-
Fly Ash (%)	-	-	11.1	-	-	28.7	-	-
Slag (%)	-	-	-	100	-	-	-	-
Metakaolin (%)	-	-	-	-	5.3	-	-	10.5
Filler (%)	-	-	-	-	-	-	34	36.2
Sp (%)	1.8	-	-	1.8	3.1	2.1	2.07	1.46
fc28 (MPa)	46.1	89.2	53.4	66.7	91.5	69.7	93	84
Time (days)	1-28	1-28	1-28	1-28	1-28	1-28	1-28	1-28

575

576 In the case of a normal strength concrete with a water-to-binder ratio of 0.42 and without any  
 577 SCM, the model's prediction agrees with the experimental values, and only a 20  $\mu\epsilon$  difference is  
 578 obtained after 28 days, as illustrated in Fig. 13 a). However, the B4 model better predicted the  
 579 long-term 28-d shrinkage value. The CEB MC 10 model's modified version largely overestimated  
 580 the autogenous shrinkage values. On the opposite, as observed in the global statistics, the CEB  
 581 MC 10 model underestimates the autogenous shrinkage at 28 days.

582 The effect of the cement substitution by silica fume is illustrated in Fig. 13 b). Both the XGB  
 583 model and the modified version of the CEB MC 10 model can provide good predictions, while the  
 584 modified versions of the B4 model largely underestimate autogenous shrinkage by around 300  $\mu\epsilon$   
 585 and the CEB MC 10 model underestimates the shrinkage value by approximately 400  $\mu\epsilon$  .

586 As illustrated in Fig. 13 c), the XGB model prediction of the shrinkage values of concrete with  
 587 a relatively low W/B around 0.3 incorporating fly ash closely matches the experimental values,  
 588 even at an early age. The analytical models' predictions are relatively far from the experimental  
 589 values at an early age, which might be attributed to these models' absent or relatively recent  
 590 considerations of fly ash effects. It should be noticed that the 28d shrinkage values predicted by  
 591 the modified versions of the B4 and CEB MC10 model are in relatively good agreement with the  
 592 experimental values, but their prediction might deviate from the measured values for more  
 593 extended periods.

594 Fig. 13 d) displays the autogenous shrinkage predictions of the models in the case of high-



595 strength concrete, with a water-to-binder ratio of 0.32, in which 50% of cement has been replaced  
596 by slag. In this case, exemplifying modern concretes with high substitution rates, the XGB model  
597 also performed well, though slightly underestimating the 28d autogenous shrinkage value. The  
598 modified B4 model provides the best prediction, but the modified version of the B4 S model  
599 underestimated the shrinkage values as well. The CEB MC 10 models underestimated the  
600 shrinkage values, while the modified version of the CEB MC 10 whose prediction can be qualified  
601 as correct, still overestimated the shrinkage value by around  $75 \mu\epsilon$ . It would be interesting to  
602 combine analytical and machine learning models in such a case.

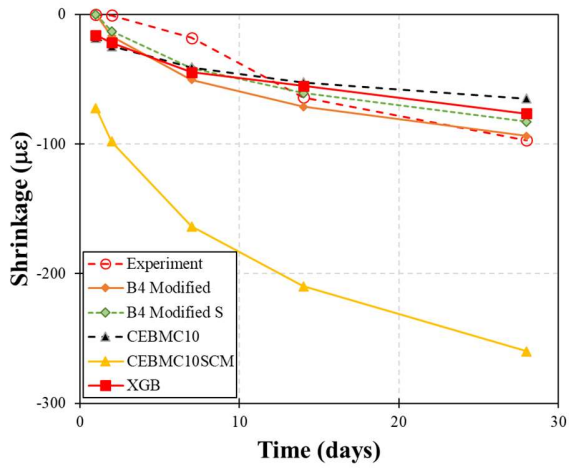
603 Though more experimental data is needed regarding the influence of calcined clay on drying  
604 and autogenous shrinkage, the developed XGB model showed better results than the analytical  
605 models in predicting the autogenous shrinkage evolution of concrete mixes incorporating calcined  
606 clay, as illustrated in Fig. 13 e). The analytical models which do not consider the effect of calcined  
607 clay - modified B4, modified B4 S and CEB MC 10 models - largely underestimated the shrinkage  
608 values, while the modified version of the CEB MC 10 models overestimated its influence.

609 The prediction quality of the models can also be calculated in the case of ternary cement  
610 blends. As illustrated in Fig. 13 f), the XGB model captured the combined effect of silica fume  
611 and fly ash satisfactorily, as the 28d autogenous shrinkage prediction differs by around  $60 \mu\epsilon$  from  
612 the experimental value, but overestimates early age shrinkage. In this case, the modified version  
613 of the B4 model outputs good yet slightly underestimated results. However, the prediction of the  
614 autogenous shrinkage of some ternary formulations might be improved in the future. As illustrated  
615 in concrete in Fig. 13 g), the XGB and analytical models predictions are pretty far from the  
616 measured values for concrete in which cement is partially substituted by silica fume and filler. In  
617 this particular case, the modified version of the CEB MC 10 model provided the best estimation  
618 of the final 28-day shrinkage value.

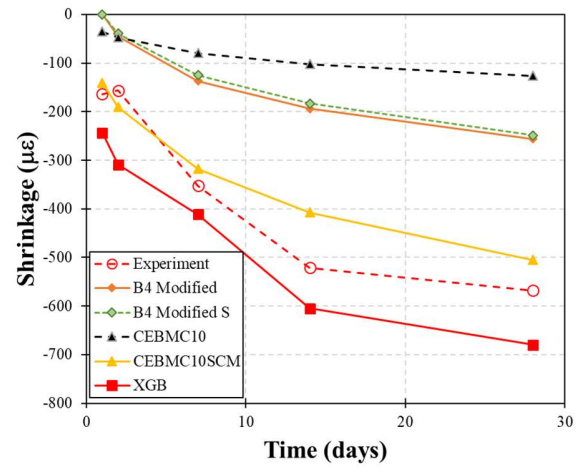
619 The XGB model's predictions are better in the case of LC<sup>3</sup> concrete incorporating calcined  
620 clay and filler, as illustrated in Fig. 13 h) since several formulations of this type are present in the  
621 database. In that case, the model predictions are good, especially after 10 days, while the early  
622 shrinkage values are overestimated. Other models cannot capture the experimental shrinkage  
623 evolution.

624

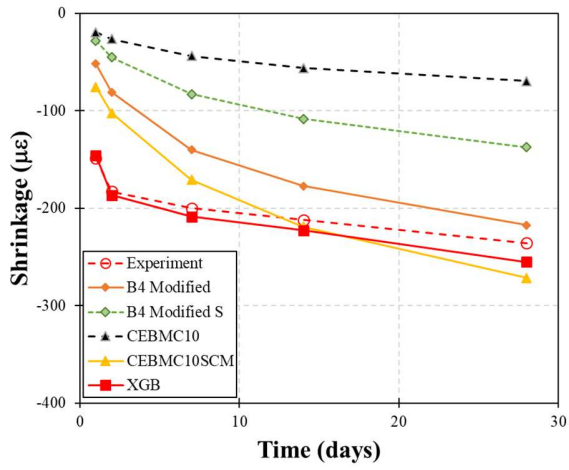
a)



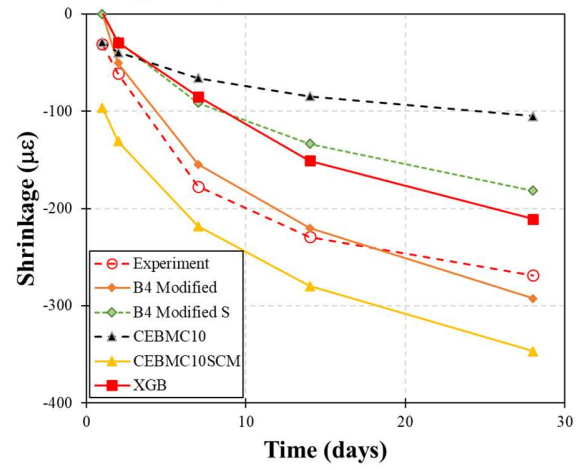
b)



c)

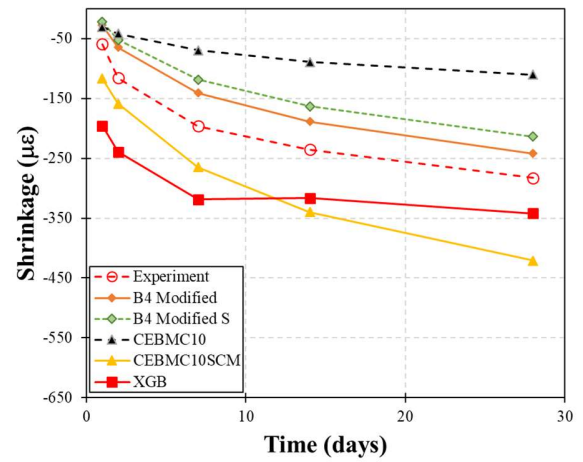
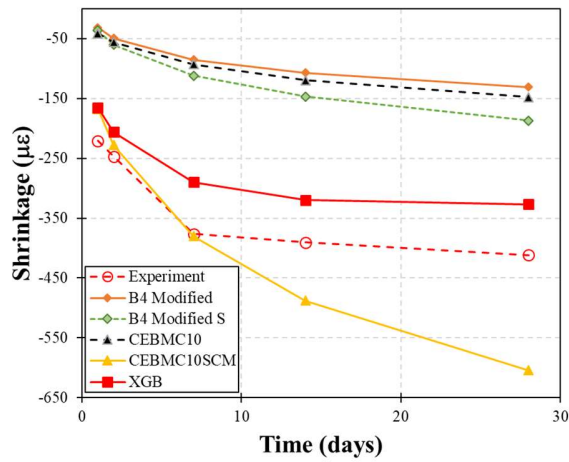


d)

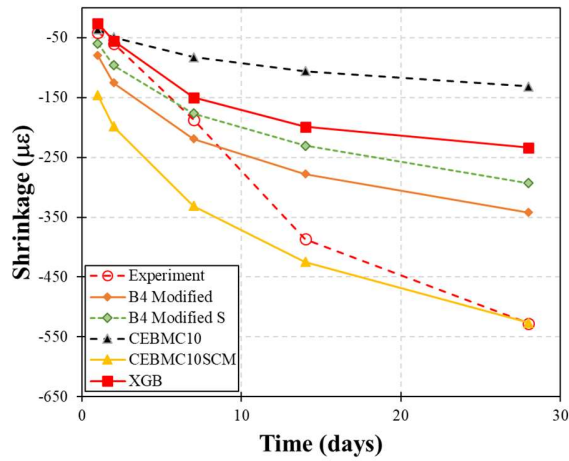


e)

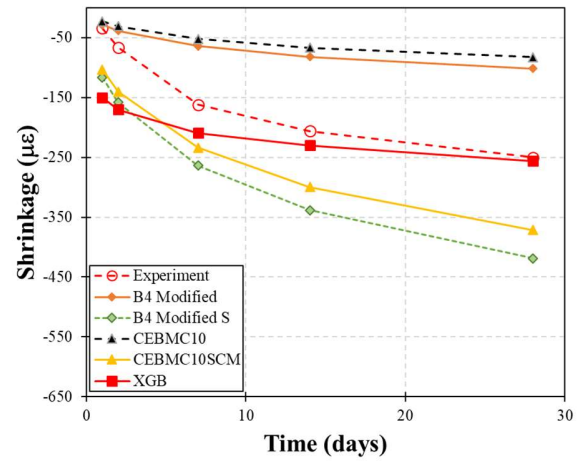
f)



g)



h)



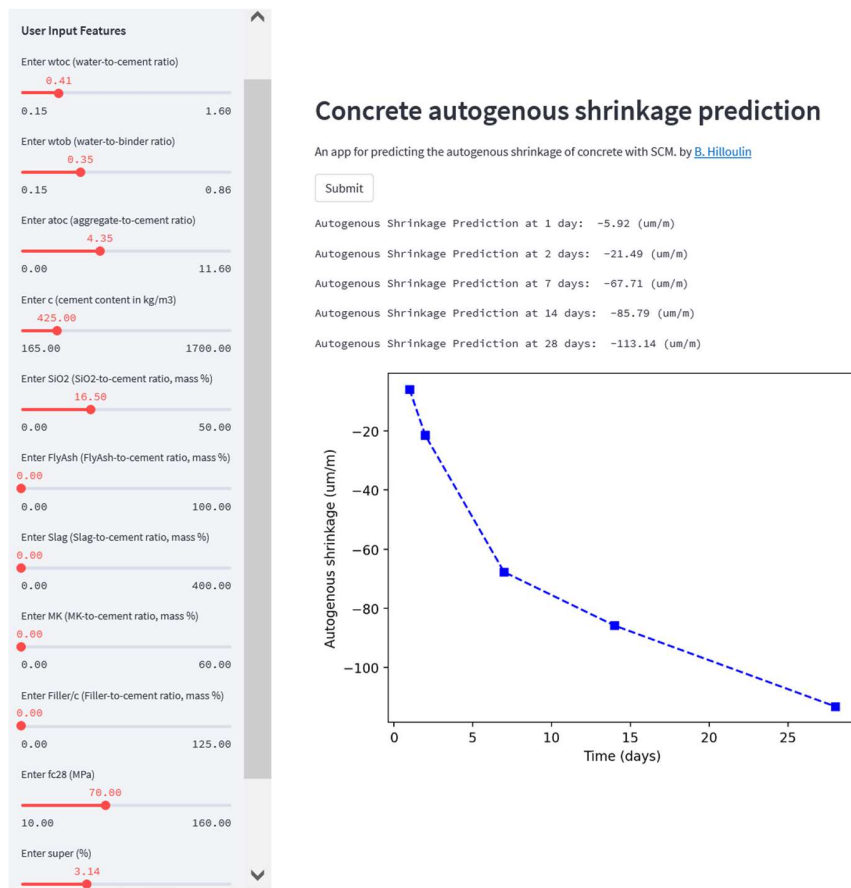
625 **Fig. 13.** Comparison of the autogenous shrinkage predictions of the XGB model and the  
 626 analytical models for several unknown concrete formulations detailed in Table 4: a) normal  
 627 strength concrete without SCM, b) high strength concrete with silica fume, c) normal strength  
 628 concrete with fly ash, d) high-strength concrete with 50% of cement replaced by slag, e) high-  
 629 strength concrete incorporating calcined clay, f) high-strength concrete with cement replaced by  
 630 silica fume and fly ash, g) high-strength concrete with silica fume and filler, h) normal strength  
 631 concrete incorporating calcined clay and filler (LC<sup>3</sup> concrete).

632

#### 633 4.5 Web application

634 A web application has been built using the first model trained on the whole database to provide  
 635 autogenous shrinkage predictions of various concrete formulations. The user can input the concrete

636 formulation using several sliders whose minimum and maximum values corresponds to the ones  
 637 from the database, and then predictions of the shrinkage values at 1 day, 2 days, 7 days, 14 days  
 638 and 28 days are computed online based on the model developed in this study. Default values  
 639 corresponding to a high-strength concrete formulation incorporating silica fume are considered  
 640 from a study not included in the database [75]. A graph is plotted to visualize the predictions as  
 641 illustrated in Fig.14. The web application is available at [https://share.streamlit.io/bhilloul/concrete-](https://share.streamlit.io/bhilloul/concrete-auto-shrinkage-ai-streamlit/main)  
 642 [auto-shrinkage-ai-streamlit/main](https://share.streamlit.io/bhilloul/concrete-auto-shrinkage-ai-streamlit/main) .



643 **Fig. 14.** Web application for autogenous shrinkage prediction of concrete incorporating SCM.  
 644  
 645

## 646 5 Conclusions

647 This study examined the usage and potential of several Machine Learning (ML) algorithms to  
 648 precisely predict the autogenous shrinkage of low-carbon cementitious materials with various  
 649 types of supplementary cementitious materials (SCM) and compared the results with analytical  
 650 models, e.g., the modified version of the B4 and B4 S models, the CEB MC 10, and the CEB MC

651 10 modified version. To this end, an extensive database of autogenous shrinkage measurements  
652 has been built based on a previous study. First, Gradient Boosting (GB), XG Boost (XGB), Support  
653 Vector Machine (SVM), and an Artificial Neural Network (ANN) model have been optimized  
654 using grid search hyperparameter optimization, and their performance has been assessed using k-  
655 fold cross-validation. The best model has been further analyzed using the Shapley Additives  
656 Explanations (SHAP) to facilitate interpretations and provide novel insights into the autogenous  
657 shrinkage mechanism. Finally, the developed model has been compared to the aforementioned  
658 analytical models, some recently adapted to consider the effect of SCM. Based on the results, the  
659 following main conclusions can be drawn:

- 660 • An extensive database composed of 1482 autogenous shrinkage curves of various  
661 binary, ternary and quaternary cement paste, mortar, and concrete incorporating silica  
662 fume, fly ash, slag, calcined clay, and filler, containing also concrete 28-day  
663 compressive strength, has been built using data from the literature.
- 664 • ML models can accurately predict the autogenous shrinkage, and their  
665 hyperparameters can be optimized through grid search. Indeed,  $R^2$  values of around  
666 0.805 to 0.931 have been obtained on the test set. The best model identified in this  
667 study is XGB.
- 668 • The relative influence of the model inputs on the predictions can be analyzed using  
669 SHAP values. Global SHAP values agreed with the experimental observations, and  
670 the A/C ratio, time, W/B ratio, and silica fume content are particularly impacting.  
671 Local SHAP analysis helped identify some evolutions of the shrinkage behavior with  
672 increasing amounts of SCM, for example, the competition between the shrinkage  
673 increase due to the pore refinement and the shrinkage delay due to slower hydration.
- 674 • The model predictions of completely unknown shrinkage behavior of concretes  
675 incorporating SCM are significantly better than the predictions of analytical models  
676 such as the modified B4 models, the CEB MC 10, and the modified version CEB MC  
677 10, even though some analytical models can perform better predictions for specific  
678 formulations. Detailed comparisons of the models' shrinkage predictions of various  
679 concrete compositions have been studied thoroughly.
- 680 • A web application has been developed to provide fast and precise predictions to the

681 scientific and industrial community and drive further low-carbon cementitious  
682 materials developments.

683 The study might open up novel research and industrial perspectives on shrinkage prediction  
684 of eco-friendly cementitious materials using advanced Machine Learning techniques.

#### 685 **CRedit authorship contribution statement**

686 **Benoit Hilloulin:** Conceptualization, Methodology, Investigation, Software, Formal analysis, Writing -  
687 original draft, Writing - review & editing, Validation.

688 **Van Quan Tran:** Conceptualization, Methodology, Software, Visualization. Writing -original draft,  
689 Writing - review & editing.

690

691 **Conflict of Interest:** The authors declare that there is no conflict of interest.

692

693 **Funding:** This research did not receive any specific grant from funding agencies in the public,  
694 commercial, or not-for-profit sectors.

695

#### 696 **References**

- 697 [1] S.A. Moghaddas, M. Nekoei, E. Mohammadi Golafshani, M. Nehdi, M. Arashpour, Modeling carbonation  
698 depth of recycled aggregate concrete using novel automatic regression technique, *Journal of Cleaner*  
699 *Production*. (2022) 133522. <https://doi.org/10.1016/j.jclepro.2022.133522>.
- 700 [2] D.N. Huntzinger, T.D. Eatmon, A life-cycle assessment of Portland cement manufacturing: comparing the  
701 traditional process with alternative technologies, *Journal of Cleaner Production*. 17 (2009) 668–675.  
702 <https://doi.org/10.1016/j.jclepro.2008.04.007>.
- 703 [3] B. Lothenbach, K. Scrivener, R.D. Hooton, Supplementary cementitious materials, *Cement and Concrete*  
704 *Research*. 41 (2011) 1244–1256. <https://doi.org/10.1016/j.cemconres.2010.12.001>.
- 705 [4] J. de Brito, R. Kurda, The past and future of sustainable concrete: A critical review and new strategies on  
706 cement-based materials, *Journal of Cleaner Production*. 281 (2021) 123558.  
707 <https://doi.org/10.1016/j.jclepro.2020.123558>.
- 708 [5] L. Wu, N. Farzadnia, C. Shi, Z. Zhang, H. Wang, Autogenous shrinkage of high performance concrete: A  
709 review, *Construction and Building Materials*. 149 (2017) 62–75.  
710 <https://doi.org/10.1016/j.conbuildmat.2017.05.064>.
- 711 [6] E. Tazawa, S. Miyazawa, Experimental study on mechanism of autogenous shrinkage of concrete, *Cement*  
712 *and Concrete Research*. 25 (1995) 1633–1638. [https://doi.org/10.1016/0008-8846\(95\)00159-X](https://doi.org/10.1016/0008-8846(95)00159-X).
- 713 [7] O. Bernard, E. Brühwiler, Influence of autogenous shrinkage on early age behaviour of structural elements  
714 consisting of concretes of different ages, *Materials and Structures*. 35 (2002) 7.  
715 <https://doi.org/10.1007/BF02483123>.
- 716 [8] J. Kheir, A. Klausen, T.A. Hammer, L. De Meyst, B. Hilloulin, K. Van Tittelboom, A. Loukili, N. De Belie,  
717 Early age autogenous shrinkage cracking risk of an ultra-high performance concrete (UHPC) wall: Modelling  
718 and experimental results, *Engineering Fracture Mechanics*. 257 (2021) 108024.

- 719 <https://doi.org/10.1016/j.engfracmech.2021.108024>.
- 720 [9] International Federation for Structural Concrete, fib Model Code for Concrete Structures, Lausanne,  
721 Switzerland, 2010.
- 722 [10] RILEM Technical Committee TC-242-MDC (Zdeněk P. Bažant, chair), RILEM draft recommendation: TC-  
723 242-MDC multi-decade creep and shrinkage of concrete: material model and structural analysis\*: Model B4  
724 for creep, drying shrinkage and autogenous shrinkage of normal and high-strength concretes with multi-  
725 decade applicability, *Mater Struct.* 48 (2015) 753–770. <https://doi.org/10.1617/s11527-014-0485-2>.
- 726 [11] M.H. Hubler, R. Wendner, Z.P. Bažant, Statistical justification of Model B4 for drying and autogenous  
727 shrinkage of concrete and comparisons to other models, *Mater Struct.* 48 (2015) 797–814.  
728 <https://doi.org/10.1617/s11527-014-0516-z>.
- 729 [12] EN 2004, ed., General rules and rules for building, Eurocode 2 –Design of concrete structures, Part 1, n.d.
- 730 [13] American Concrete Institute, ACI Committee 209--Creep and Shrinkage, eds., Guide for modeling and  
731 calculating shrinkage and creep in hardened concrete, American Concrete Institute, Farmington Hills, MI,  
732 2008.
- 733 [14] M. Rasoolnejad, Prediction of autogenous shrinkage in concrete from material composition or strength  
734 calibrated by a large database, as update to model B4, *Materials and Structures.* (2019) 17.
- 735 [15] J.-R. Weng, W.-C. Liao, Microstructure and shrinkage behavior of high-performance concrete containing  
736 supplementary cementitious materials, *Construction and Building Materials.* 308 (2021) 125045.  
737 <https://doi.org/10.1016/j.conbuildmat.2021.125045>.
- 738 [16] C. Youssef Namnoum, B. Hilloulin, F. Grondin, A. Loukili, Determination of the origin of the strength  
739 regain after self-healing of binary and ternary cementitious materials including slag and metakaolin, *Journal*  
740 *of Building Engineering.* 41 (2021) 102739. <https://doi.org/10.1016/j.jobe.2021.102739>.
- 741 [17] P. Termkhajornkit, T. Nawa, M. Nakai, T. Saito, Effect of fly ash on autogenous shrinkage, *Cement and*  
742 *Concrete Research.* 35 (2005) 473–482. <https://doi.org/10.1016/j.cemconres.2004.07.010>.
- 743 [18] J.M. Khatib, Performance of self-compacting concrete containing fly ash, *Construction and Building*  
744 *Materials.* 22 (2008) 1963–1971. <https://doi.org/10.1016/j.conbuildmat.2007.07.011>.
- 745 [19] K.M. Lee, H.K. Lee, S.H. Lee, G.Y. Kim, Autogenous shrinkage of concrete containing granulated blast-  
746 furnace slag, *Cement and Concrete Research.* 36 (2006) 1279–1285.  
747 <https://doi.org/10.1016/j.cemconres.2006.01.005>.
- 748 [20] A. Itim, K. Ezziane, E.-H. Kadri, Compressive strength and shrinkage of mortar containing various amounts  
749 of mineral additions, *Construction and Building Materials.* 25 (2011) 3603–3609.  
750 <https://doi.org/10.1016/j.conbuildmat.2011.03.055>.
- 751 [21] M. Bouasker, N.E.H. Khalifa, P. Mounanga, N. Ben Kahla, Early-age deformation and autogenous cracking  
752 risk of slag–limestone filler-cement blended binders, *Construction and Building Materials.* 55 (2014) 158–  
753 167. <https://doi.org/10.1016/j.conbuildmat.2014.01.037>.
- 754 [22] C. Jiang, Y. Yang, Y. Wang, Y. Zhou, C. Ma, Autogenous shrinkage of high performance concrete  
755 containing mineral admixtures under different curing temperatures, *Construction and Building Materials.* 61  
756 (2014) 260–269. <https://doi.org/10.1016/j.conbuildmat.2014.03.023>.
- 757 [23] K. Scrivener, F. Martirena, S. Bishnoi, S. Maity, Calcined clay limestone cements (LC3), *Cement and*  
758 *Concrete Research.* 114 (2018) 49–56. <https://doi.org/10.1016/j.cemconres.2017.08.017>.
- 759 [24] P.J.P. Gleize, M. Cyr, G. Escadeillas, Effects of metakaolin on autogenous shrinkage of cement pastes,  
760 *Cement and Concrete Composites.* 29 (2007) 80–87. <https://doi.org/10.1016/j.cemconcomp.2006.09.005>.
- 761 [25] Y. Dhandapani, T. Sakthivel, M. Santhanam, R. Gettu, R.G. Pillai, Mechanical properties and durability  
762 performance of concretes with Limestone Calcined Clay Cement (LC3), *Cement and Concrete Research.* 107  
763 (2018) 136–151. <https://doi.org/10.1016/j.cemconres.2018.02.005>.
- 764 [26] Q.D. Nguyen, S. Afroz, Y. Zhang, T. Kim, W. Li, A. Castel, Autogenous and total shrinkage of limestone  
765 calcined clay cement (LC3) concretes, *Construction and Building Materials.* 314 (2022) 125720.  
766 <https://doi.org/10.1016/j.conbuildmat.2021.125720>.
- 767 [27] M. Valcuende, E. Marco, C. Parra, P. Serna, Influence of limestone filler and viscosity-modifying admixture  
768 on the shrinkage of self-compacting concrete, *Cement and Concrete Research.* 42 (2012) 583–592.  
769 <https://doi.org/10.1016/j.cemconres.2012.01.001>.
- 770 [28] A. Alrifai, S. Aggoun, A. Kadri, S. Kenai, E. Kadri, Paste and mortar studies on the influence of mix design

- 771 parameters on autogenous shrinkage of self-compacting concrete, *Construction and Building Materials*. 47  
772 (2013) 969–976. <https://doi.org/10.1016/j.conbuildmat.2013.05.024>.
- 773 [29] B. Craeye, G. De Schutter, B. Desmet, J. Vantomme, G. Heirman, L. Vandewalle, Ö. Cizer, S. Aggoun, E.H.  
774 Kadri, Effect of mineral filler type on autogenous shrinkage of self-compacting concrete, *Cement and*  
775 *Concrete Research*. 40 (2010) 908–913. <https://doi.org/10.1016/j.cemconres.2010.01.014>.
- 776 [30] T. Sakthivel, R. Gettu, R. Pillai, ADJUSTMENT OF RILEM B4 MODEL PARAMETERS FOR BETTER  
777 PREDICTION OF THE SHRINKAGE RESPONSE OF BLENDED CEMENT CONCRETE, *Indian*  
778 *Concrete Journal*. 95 (2021) 51–57.
- 779 [31] Z. Li, J. Yoon, R. Zhang, F. Rajabipour, W.V. Srubar III, I. Dabo, A. Radlińska, Machine learning in  
780 concrete science: applications, challenges, and best practices, *Npj Comput Mater*. 8 (2022) 127.  
781 <https://doi.org/10.1038/s41524-022-00810-x>.
- 782 [32] I.-C. Yeh, Modeling of strength of high-performance concrete using artificial neural networks, *Cement and*  
783 *Concrete Research*. 28 (1998) 1797–1808. [https://doi.org/10.1016/S0008-8846\(98\)00165-3](https://doi.org/10.1016/S0008-8846(98)00165-3).
- 784 [33] J.-S. Chou, C.-F. Tsai, A.-D. Pham, Y.-H. Lu, Machine learning in concrete strength simulations: Multi-  
785 nation data analytics, *Construction and Building Materials*. 73 (2014) 771–780.  
786 <https://doi.org/10.1016/j.conbuildmat.2014.09.054>.
- 787 [34] D.K. Bui, T. Nguyen, J.S. Chou, H. Nguyen-Xuan, T.D. Ngo, A modified firefly algorithm-artificial neural  
788 network expert system for predicting compressive and tensile strength of high-performance concrete,  
789 *Construction and Building Materials*. 180 (2018) 320–333.  
790 <https://doi.org/10.1016/j.conbuildmat.2018.05.201>.
- 791 [35] M.J. Munir, S.M. Saleem Kazmi, Y.-F. Wu, X. Lin, M.R. Ahmad, Development of a novel compressive  
792 strength design equation for natural and recycled aggregate concrete through advanced computational  
793 modeling, *Journal of Building Engineering*. (2022) 104690. <https://doi.org/10.1016/j.job.2022.104690>.
- 794 [36] H.S. Ullah, R.A. Khushnood, J. Ahmad, F. Farooq, Predictive modelling of sustainable lightweight foamed  
795 concrete using machine learning novel approach, *Journal of Building Engineering*. 56 (2022) 104746.  
796 <https://doi.org/10.1016/j.job.2022.104746>.
- 797 [37] W.E. Elemam, A.H. Abdelraheem, M.G. Mahdy, A.M. Tahwia, Optimizing fresh properties and compressive  
798 strength of self-consolidating concrete, *Construction and Building Materials*. 249 (2020) 118781.  
799 <https://doi.org/10.1016/j.conbuildmat.2020.118781>.
- 800 [38] J. Karthikeyan, A. Upadhyay, N.M. Bhandari, Artificial Neural Network for Predicting Creep and Shrinkage  
801 of High Performance Concrete, *ACT*. 6 (2008) 135–142. <https://doi.org/10.3151/jact.6.135>.
- 802 [39] P. Chen, W. Zheng, Y. Wang, W. Chang, Creep model of high-strength concrete containing supplementary  
803 cementitious materials, *Construction and Building Materials*. 202 (2019) 494–506.  
804 <https://doi.org/10.1016/j.conbuildmat.2019.01.005>.
- 805 [40] R. Cai, T. Han, W. Liao, J. Huang, D. Li, A. Kumar, H. Ma, Prediction of surface chloride concentration of  
806 marine concrete using ensemble machine learning, *Cement and Concrete Research*. 136 (2020) 106164.  
807 <https://doi.org/10.1016/j.cemconres.2020.106164>.
- 808 [41] I. Nunez, M.L. Nehdi, Machine learning prediction of carbonation depth in recycled aggregate concrete  
809 incorporating SCMs, *Construction and Building Materials*. 287 (2021) 123027.  
810 <https://doi.org/10.1016/j.conbuildmat.2021.123027>.
- 811 [42] X. Wu, S. Zheng, Z. Feng, B. Chen, Y. Qin, W. Xu, Y. Liu, Prediction of the frost resistance of high-  
812 performance concrete based on RF-REF: A hybrid prediction approach, *Construction and Building Materials*.  
813 333 (2022) 127132. <https://doi.org/10.1016/j.conbuildmat.2022.127132>.
- 814 [43] L. Bal, F. Buyle-Bodin, Artificial neural network for predicting drying shrinkage of concrete, *Construction*  
815 *and Building Materials*. 38 (2013) 248–254. <https://doi.org/10.1016/j.conbuildmat.2012.08.043>.
- 816 [44] B. Hilloulin, V.Q. Tran, Using machine learning techniques for predicting autogenous shrinkage of concrete  
817 incorporating superabsorbent polymers and supplementary cementitious materials, *Journal of Building*  
818 *Engineering*. 49 (2022) 104086. <https://doi.org/10.1016/j.job.2022.104086>.
- 819 [45] B. Hilloulin, M. Robira, A. Loukili, Coupling statistical indentation and microscopy to evaluate  
820 micromechanical properties of materials: Application to viscoelastic behavior of irradiated mortars, *Cement*  
821 *and Concrete Composites*. 94 (2018) 153–165. <https://doi.org/10.1016/j.cemconcomp.2018.09.008>.
- 822 [46] B. Hilloulin, M. Lagrange, M. Duvillard, G. Garioud,  $\epsilon$ -greedy automated indentation of cementitious



- 823 materials for phase mechanical properties determination, *Cement and Concrete Composites*. (2022) 104465.  
824 <https://doi.org/10.1016/j.cemconcomp.2022.104465>.
- 825 [47] B. Hilloulin, I. Bekrine, E. Schmitt, A. Loukili, Open-source deep learning-based air-voids detection  
826 algorithm for concrete microscopic images, *Journal of Microscopy*. (2022) jmi.13098.  
827 <https://doi.org/10.1111/jmi.13098>.
- 828 [48] M. Liang, Y. Gan, Z. Chang, Z. Wan, E. Schlangen, B. Šavija, Microstructure-informed deep convolutional  
829 neural network for predicting short-term creep modulus of cement paste, *Cement and Concrete Research*. 152  
830 (2022) 106681. <https://doi.org/10.1016/j.cemconres.2021.106681>.
- 831 [49] P. Teck, R. Snellings, J. Elsen, Method for quantifying the reaction degree of slag in alkali-activated cements  
832 using deep learning-based electron microscopy image analysis, *Journal of Microscopy*. (2022) jmi.13094.  
833 <https://doi.org/10.1111/jmi.13094>.
- 834 [50] B. Hilloulin, I. Bekrine, E. Schmitt, A. Loukili, Modular Deep Learning Segmentation Algorithm for  
835 Concrete Microscopic Images, *Construction and Building Materials*. (2022).  
836 <https://doi.org/10.1016/j.conbuildmat.2022.128736>.
- 837 [51] R. Zhang, Y. Liu, H. Sun, Physics-guided convolutional neural network (PhyCNN) for data-driven seismic  
838 response modeling, *Engineering Structures*. 215 (2020) 110704.  
839 <https://doi.org/10.1016/j.engstruct.2020.110704>.
- 840 [52] B.K. Oh, H.S. Park, B. Glisic, Prediction of long-term strain in concrete structure using convolutional neural  
841 networks, air temperature and time stamp of measurements, *Automation in Construction*. 126 (2021) 103665.  
842 <https://doi.org/10.1016/j.autcon.2021.103665>.
- 843 [53] S.M. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, in: *Proceedings of the 31st*  
844 *International Conference on Neural Information Processing Systems*, Curran Associates Inc., Red Hook, NY,  
845 USA, 2017: pp. 4768–4777.
- 846 [54] M. Liang, Z. Chang, Z. Wan, Y. Gan, E. Schlangen, B. Šavija, Interpretable Ensemble-Machine-Learning  
847 models for predicting creep behavior of concrete, *Cement and Concrete Composites*. (2021) 104295.  
848 <https://doi.org/10.1016/j.cemconcomp.2021.104295>.
- 849 [55] Y. Sun, R. Yu, Z. Shui, X. Wang, D. Qian, B. Rao, J. Huang, Y. He, Understanding the porous aggregates  
850 carrier effect on reducing autogenous shrinkage of Ultra-High Performance Concrete (UHPC) based on  
851 response surface method, *Construction and Building Materials*. 222 (2019) 130–141.  
852 <https://doi.org/10.1016/j.conbuildmat.2019.06.151>.
- 853 [56] H.K. Lee, K.M. Lee, B.G. Kim, Autogenous shrinkage of high-performance concrete containing fly ash,  
854 *Magazine of Concrete Research*. (2003) 9.
- 855 [57] P.Y. Yan, Z.C. Chen, J.C. Wang, F. Zheng, Autogenous shrinkage of concrete prepared with the binders  
856 containing different kinds of mineral admixtures, in: *ICCC2007, 2007*.
- 857 [58] J.J. Brooks, M.A.M. Johari, Effect of metakaolin on creep and shrinkage of concrete, *Cement and Concrete*  
858 *Composites*. (2001) 8.
- 859 [59] A. Duran-Herrera, J. De-León-Esquivel, D.P. Bentz, P. Valdez-Tamez, Self-compacting concretes using fly  
860 ash and fine limestone powder: Shrinkage and surface electrical resistivity of equivalent mortars,  
861 *Construction and Building Materials*. 199 (2019) 50–62. <https://doi.org/10.1016/j.conbuildmat.2018.11.191>.
- 862 [60] M.-Y. Xuan, Y. Han, X.-Y. Wang, The Hydration, Mechanical, Autogenous Shrinkage, Durability, and  
863 Sustainability Properties of Cement–Limestone–Slag Ternary Composites, *Sustainability*. 13 (2021) 1881.  
864 <https://doi.org/10.3390/su13041881>.
- 865 [61] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal,  
866 S.-I. Lee, From local explanations to global understanding with explainable AI for trees, *Nat Mach Intell*. 2  
867 (2020) 56–67. <https://doi.org/10.1038/s42256-019-0138-9>.
- 868 [62] C. Cortes, V. Vapnik, Support-vector networks, *Mach Learn*. 20 (1995) 273–297.  
869 <https://doi.org/10.1007/BF00994018>.
- 870 [63] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, *Statistics and Computing*. 14 (2004) 199–  
871 222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>.
- 872 [64] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, arXiv, 2017.  
873 <https://doi.org/10.48550/arXiv.1412.6980>.
- 874 [65] M.M. Ahmed, M. Abdel-Aty, Application of Stochastic Gradient Boosting Technique to Enhance Reliability

- 875 of Real-Time Risk Assessment: Use of Automatic Vehicle Identification and Remote Traffic Microwave  
876 Sensor Data, *Transportation Research Record*. 2386 (2013) 26–34. <https://doi.org/10.3141/2386-04>.
- 877 [66] J. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, *Annals of Statistics*. 29 (2001)  
878 1189–1232. <https://doi.org/10.2307/2699986>.
- 879 [67] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, *Science*. 220 (1983) 671–  
880 680. <https://doi.org/10.1126/science.220.4598.671>.
- 881 [68] S. Koziel, X.-S. Yang, eds., *Computational Optimization, Methods and Algorithms*, Springer-Verlag, Berlin  
882 Heidelberg, 2011. <https://doi.org/10.1007/978-3-642-20859-1>.
- 883 [69] C.M. Martínez, D. Cao, 2 - Integrated energy management for electrified vehicles, in: C.M. Martínez, D. Cao  
884 (Eds.), *Ihorizon-Enabled Energy Management for Electrified Vehicles*, Butterworth-Heinemann, 2019: pp.  
885 15–75. <https://doi.org/10.1016/B978-0-12-815010-8.00002-8>.
- 886 [70] W. Gao, C. Mi, Hybrid vehicle design using global optimisation algorithms, *IJEHV*. 1 (2007) 57.  
887 <https://doi.org/10.1504/IJEHV.2007.014447>.
- 888 [71] V.Q. Tran, Hybrid gradient boosting with meta-heuristic algorithms prediction of unconfined compressive  
889 strength of stabilized soil based on initial soil properties, mix design and effective compaction, *Journal of*  
890 *Cleaner Production*. (2022) 131683. <https://doi.org/10.1016/j.jclepro.2022.131683>.
- 891 [72] V.Q. Tran, L.Q. Nguyen, Using machine learning technique for designing reinforced lightweight soil, *Journal*  
892 *of Intelligent & Fuzzy Systems*. 43 (2022) 1633–1650. <https://doi.org/10.3233/JIFS-212621>.
- 893 [73] M. Rasoolnejad, S. Rahimi-Aghdam, Z.P. Bažant, Prediction of autogenous shrinkage in concrete from  
894 material composition or strength calibrated by a large database, as update to model B4, *Mater Struct*. 52  
895 (2019) 33. <https://doi.org/10.1617/s11527-019-1331-3>.
- 896 [74] A. Darquennes, S. Staquet, M.-P. Delplancke-Ogletree, B. Espion, Effect of autogenous deformation on the  
897 cracking risk of slag cement concretes, *Cement and Concrete Composites*. 33 (2011) 368–379.  
898 <https://doi.org/10.1016/j.cemconcomp.2010.12.003>.
- 899 [75] M. Mazloom, A.A. Ramezani-pour, J.J. Brooks, Effect of silica fume on mechanical properties of high-  
900 strength concrete, *Cement and Concrete Composites*. 26 (2004) 347–357. [https://doi.org/10.1016/S0958-](https://doi.org/10.1016/S0958-9465(03)00017-9)  
901 [9465\(03\)00017-9](https://doi.org/10.1016/S0958-9465(03)00017-9).
- 902