Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

UNIVERSITY OF TRANSPORT TECHNOLOGY

An ensemble approach to predict the compressive strength of ultra-high-performance concrete

May Huu Nguyen[1], Hai-Van Thi Mai[2,*], Son Hoang Trinh[2]
[1]Civil and Environmental Engineering Program, Graduate School of Advanced Science and Engineering, Hiroshima University, 1-4-1, Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan
[2]University of Transport Technology, Hanoi 100000, Vietnam

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2026.en.6.3.161-195

[*]**Corresponding author:**
Email address:
vanmth@utt.edu.vn

Received: 18/11/2025
Received in Revised Form: 05/05/2026
Accepted: 07/08/2026

Abstract: Ultra-High-Performance Concrete (UHPC) is a state-of-the-art concrete technology with exceptional qualities, including high compressive strength (CS) and durability. The CS, an essential property of UHPC, is determined through costly, time-consuming studies that require large amounts of material. To overcome these constraints, this work aimed to estimate the CS of UHPC using a range of single- and hybrid-machine learning (ML) methods. For this, five ML models, including Decision Tree (DT), Gradient Boosting (GB), Light Gradient Boosting Machine (LightGBM), CatBoost, and a stacking Ensemble model combining these base models, were developed. The input space includes cement, silica fume, slag, fly ash, quartz powder, limestone powder, nano-silica, water, fine and coarse aggregate, fiber, superplasticizer, temperature, relative humidity, and age. The findings demonstrate that the Ensemble model achieved the best overall predictive performance across 20 Monte Carlo simulations, with a mean RMSE of 6.751 ± 0.316 MPa, MAE of 4.943 ± 0.219 MPa, and R^2 of 0.972 ± 0.002. According to the findings of 1D, 2D partial dependence plot (PDP), and SHAP analyses, age, cement, silica fume, water, sand, fiber, and superplasticizer were the primary variables influencing UHPC's CS.

Keywords: Ultra-High-Performance Concrete; Ensemble learning; Compressive Strength; Machine Learning.

1. Introduction

Ultra-high performance concrete (UHPC) is a more advanced development of traditional concrete with high durability and strength [1]. UHPC offers many advantages over conventional concrete, such as reducing the required volume of concrete and decreasing costs and construction time. However, to achieve a dense texture and shallow voids, UHPC minimizes particle size and increases binder content, i.e., approximately 1000 kg/m^3 [2]. That is one of the main reasons UHPC is costly and considered environmentally harmful. Cement production generates significant amounts of toxic gases (accounting for 5% of global emissions) that contribute to the greenhouse effect. It becomes a barrier to the widespread application of UHPC in construction. Therefore, it is necessary to research and develop UHPC with similar technical characteristics but lower cement content, offering enormous economic and environmental benefits [3].

Fly ash (FA), ground blast furnace slag

(GGBS), silica fume (SF), metakaolin (MK), and natural pozzolans (NP) are often blended with clinker in Portland cement production or used as partial cement replacements in concrete. These materials, often called Supplementary Cementitious Materials (SCMs), are added to concrete more often than clinker, and more than 60 percent of premix concrete in the United States currently contains SCMs [4]. Numerous studies have shown that SCMs can densify the microstructure, improve durability, and enhance later-age strength through pozzolanic processes. For instance, E. Güneyisi et al. [5] showed that HSC containing 5-15% MK has higher compressive strength (CS) at 28 days (39–44%) than non-MK-based concrete. The strength improvement is attributed to enhanced bonding between the cement binder and particles, integration with available MK, and increased cement-matrix density. Skaropoulou et al. [6] studied the sulfate resistance of limestone concrete (LC) soaked in 1.8% $MgSO_4$ solution. The results showed that regular limestone concrete lost its strength after 6-12 months and completely lost it after 2 years. Similarly, Sharma et al. [7] used GGBS with 35% CaO content, which enhanced CSH formation and reduced hydration temperature, thus improving strength development. As GGBS accounted for almost 24% of the binder, the 7-day CS increased by 3.3% due to enhanced intergranular density and higher cement hydration. Ogawa et al. [8] determined that the long-term sulfate resistance of cement mortars may be enhanced by combining SCM with an adequate proportion of limestone powder (LP) as opposed to employing just SCMs such as FA, GGBS, and MK. Additional research has demonstrated that nano-silica, microsilica, FA, and GGBS can alleviate expansion induced by ettringite and gypsum production, thus improving dimensional stability [9, 10]. Oertel et al. [11] summarized that some SCMs with small particle sizes could accelerate cement hydration to achieve early strength. These findings collectively underscore the significant potential of SCMs to enhance mechanical performance, durability, and cost-effectiveness in cement-based materials.

SCMs are also used in UHPC for various purposes. SCMs can be used as fillers and additives to reduce the amount of cement required. Ghafari et al. produced UHPC with 950 kg/m^3 of cement and 250 kg/m^3 of SF [12]. Aldahdooh et al. [13] designed UHPC with CS of 120 MPa using 638 kg/m^3 and 214.25 kg/m^3 SF. Chong Wang et al. [14] used W/B of 0.16 and 450 kg cement, 90 kg SF, 180 kg GGBS, and 20% LP for UHPC 165 MPa. In addition to reducing cement content, SCMs enhance both the fresh and hardened properties. For example, GGBS improves workability owing to its smooth, low-absorption particles [15]. Moreover, incorporating LP into GGBFS-containing UHPC has been shown to improve workability and reduce both autogenous and drying shrinkage [16]. Moreover, enhanced particle packing through ultrafine powders and pozzolanic processes allows UHPC to achieve exceptional compressive strengths, ranging from 150 to 180 MPa at 28 days when adding rice husk ash (RHA) with particle sizes of 3.6 to 8 µm, in contrast to 160 MPa for mixtures containing 10% SF [17]. Other studies have reported 28-day strengths of 100–150 MPa for UHPC utilizing LP, GGBS, SF, FA, and other SCMs [14, 18, 19]. Rong et al. [20] also found an increase in the 28-day CS of UHPC to 75-95 MPa for UHPC with 0 and 3% NS as partial cement replacement, respectively. These studies collectively affirm that SCMs are essential for improving UHPC performance while substantially reducing cement consumption.

Using a combination of SCMs in varying ratios often complicates the composition and proportions of UHPC, though different effects can be expected [18]. The combined effects pose a major barrier to predicting the behavior of UHPCs using SCMs [21]. For example, GGBS and FA can retard the pozzolanic reaction, prolong setting time, and reduce early-age strength, but can also improve other properties, such as increasing workability, increasing waterproofing strength, and enhancing chemical attack resistance [22]. Small SF particles may fill the larger intergranular spaces, and the added LP and MK accelerate hydration; therefore, this combination is considered to boost the pozzolanic reaction at an early stage [23]. In addition, curing conditions also significantly affect the UHPC containing SCMs [24]. Rößler et al. [25] concluded that it was possible to enhance the 7-day CS of RHA–GGBS–UHPC by using temperatures of 65°C. As with all types of concrete, CS is a significant concern and is often the focus of the design phase. The experiments often determine the CS of concrete because of the complex combined effects of the various ingredients and proportions used to produce the final product. The downside is that laboratory experiments are often expensive, time-consuming, and labor-intensive. To reduce lab time, statistical models may be used to estimate the CS of conventional concrete [26, 27]. However, these conventional models are typically insufficient for UHPC, as they assume linear relationships and fail to account for the many SCMs and curing conditions that significantly affect UHPC performance. Consequently, more sophisticated and reliable forecasting methodologies are needed to capture the intricate relationships that influence UHPC compressive strength.

With the rapid development of artificial intelligence and soft computing techniques, several researchers have developed ML models to predict the CS of UHPC. Ghafari et al. [28] applied both backpropagation neural network (BPNN) and traditional design methods to predict the CS and slump of UHPC based on 53 samples with seven inputs. The results indicate that a BPNN can predict CS and slump with higher accuracy than a conventional statistical method. Omar R. Abuodeh [29] applied an artificial neural network (ANN) approach to predict the CS of UHPC using 110 samples and eight input variables, including cement, sand, SF, FA, FB, QP, water, and superplasticizer. Further, neural interpretation diagrams (NIDs) and Sequential feature selection (SFSs) were used to identify the most important factors shaping the ANN's output. Results showed that combining ANN with SFS and NID greatly enhanced the model's accuracy and provided useful insights into the ANN's ability to forecast the CS of UHPC with various composites. It is evident that ML has significant potential to predict the CS of UHPC. Although previous models in the literature achieved high predictive accuracy, they used a small dataset and a limited number of mixture components. The nonlinear relationship between the different inputs and CS of UHPC may be better understood with suitable, comprehensive experimental data for model training. Therefore, it is desirable to develop powerful tools to deal with complex data for predicting the CS of UHPC.

The goal of this research is to use recently implemented models for the first time to make accurate predictions of UHPC's CS from a large dataset (810 samples) with more complex input variables, adding the variables of age, humidity, and curing temperature besides the variables on material composition, which show generality and diversity. Specifically, 15 input variables were used: cement, GGBS/slag, SF, LP, QP, FA, NS, water, sand, gravel, fiber, SP, RH, temperature, and age. Accordingly, the models, namely CatBoost, DT, GB, LightGBM, and the stacking Ensemble model, were used to estimate UHPC's compression strength. These powerful ML models are widely used to predict material properties and structural behavior [30–33]. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2) were used to assess the models' efficacy. The impact of the inputs on UHPC's CS was further investigated through a sensitivity analysis.

2. Database description and analysis

Building ML models requires a substantial and trustworthy dataset. Thus, 810 experimental data from 28 experimental studies were gathered through an exhaustive literature search [1–3, 12, 14, 34–48] and summarized by Marani et al. [49].

Here, a sizable dataset encompassing different mixture components was first assembled from many experimental investigations that used different materials in UHPC design and employed various curing regimes. The resulting dataset has 15 input characteristics following preprocessing. Data preprocessing adhered to the methodology established by Marani et al. [49]. The assembled dataset contains no missing values for the chosen 15 inputs; entries with incomplete data were omitted during compilation. Apparent anomalous values were removed during preprocessing, and all variables were scrutinized for non-physical ranges before model training. The inputs include the contents, in kg/m^3, of cement (C), silica fume (SF), slag (S), fly ash (FA), quartz powder (QP), limestone powder (LP), nano silica (NS), water (W), fine aggregate (Sand), coarse aggregate (Gravel), fiber (Fi), superplasticizer (SP), and temperature (T, °C), relative humidity (RH, %), and age (days). The models' output is UHPC's compressive strength (CS, MPa). Table 1 shows the detailed input and output variables of this study. Except for X_2, X_4, X_5, X_6, and X_{10}, which are dispersed in a short range, input variables are often scattered over a wide range. For instance, cement content (X_1) ranges from 270 to 1251 (kg/m^3), with 771 (kg/m^3) being the median. The amount of water and sand (X_8 and X_9) varied from 90 and 0 to 273 and 1503 (kg/m^3). The age of the sample (X_{15}) ranges from 1 to 365 days. The compressive strength (Cs) (Y) range is 29 to 221 MPa.

Table 1. The inputs and output used to develop the ML models in this study

Var.	Unit	Symbol	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
C	kg/m^3	X_1	270	472	558	637	712	771	800	850	863	960	1251
S	kg/m^3	X_2	0	0	0	0	0	0	0	0	0	140	375
SF	kg/m^3	X_3	0	0	0	50	90	144	187	216	240	260	434
LP	kg/m^3	X_4	0	0	0	0	0	0	0	0	0	235	1058
QP	kg/m^3	X_5	0	0	0	0	0	0	0	0	0	176	397
FA	kg/m^3	X_6	0	0	0	0	0	0	0	0	0	165	356
NS	kg/m^3	X_7	0	0	0	0	0	0	0	0	5	16	48
W	kg/m^3	X_8	90	160	162	165	176	177	178	182	200	220	273
Sand	kg/m^3	X_9	0	616	739	815	873	1021	1079	1231	1273	1359	1503
Gravel	kg/m^3	X_{10}	0	0	0	0	0	0	0	0	0	923	1195
Fi	kg/m^3	X_{11}	0	0	0	0	0	0	20	117	156	156	234
SP	kg/m^3	X_{12}	1	10	18	21	28	30	34	41	45	46	57
RH	%	X_{13}	50	95	100	100	100	100	100	100	100	100	100
T	°C	X_{14}	20	20	20	21	21	21	23	23	23	23	210
Age	days	X_{15}	1	3	7	7	28	28	28	28	56	90	365
CS	MPa	Y	29	68	86	103	114	122	130	146	161	174	221

To assess dependency, the existing relationships between the databases used to estimate CS are examined, as shown in Fig. 1. Most correlations between inputs and outputs were weak, ranging from -0.5 to 0.5. The study shows that the input and output parameters are treated as independent variables. In other words, any input variable can be used to build ML models and further assess the feature's significance.

3. Methods used

3.1. Machine learning methods

In this study, tree-based models, namely, Decision Tree (DT), Gradient Boosting (GB), LightGBM, and CatBoost, were selected due to their ability to (i) capture nonlinear interactions among mixture components and curing conditions,

(ii) manage mixed-scale tabular data without stringent distributional assumptions, and (iii) deliver competitive accuracy alongside moderate interpretability. These qualities are crucial for the current UHPC database, where certain combination variables, including slag, limestone powder, quartz powder, fly ash, nano-silica, and gravel, exhibit sparse or zero-inflated distributions because these elements are absent in many mixtures. Tree-based models adeptly handle such variables by learning threshold-based splits, such as distinguishing between combinations with zero dosage and those with non-zero dosage, without assuming linearity or normality of the input distribution. Boosting algorithms capture nonlinear and dose-dependent effects when adequate data are available in the relevant feature regions. DT offers a clear baseline; however, it exhibits significant variance. GB is a powerful additive learner but may overfit without proper regularization; LightGBM improves efficiency through histogram and leaf-wise splitting; and CatBoost incorporates ordered boosting and effective regularization, often mitigating overfitting. Stacking combines complementary error patterns to improve generalization.

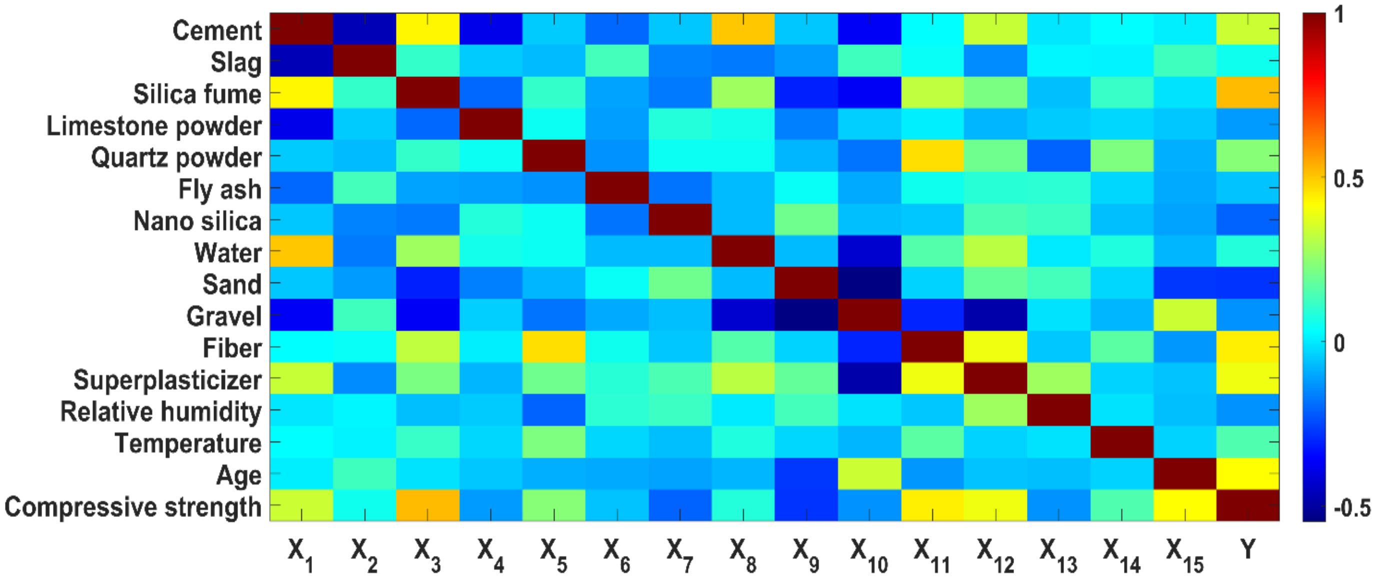

Fig. 1. Correlation analysis of the inputs and output of the current UHPC dataset

3.1.1. Categorical boosting

Categorical boosting (CatBoost) is a powerful technique for supervised ML problems. Yandex researchers and engineers developed it. This algorithm, in principle, is based on gradient boosted DTs (GBDT) and provides better performance than the original GBDT. CatBoost is built on oblivious trees with minimal parameters, supports categorical variables, and fits the GBDT framework with good accuracy. The primary challenge is making effective and reasonable use of category features. CatBoost incorporates a boost in addition to category factors. By fixing gradient bias and prediction shift, this method improves the algorithm's flexibility and adaptability in new situations. CatBoost is an algorithm for analyzing categorical features that incorporates all training datasets into the final model. As a further step, CatBoost randomly permutes these sample datasets and eliminates duplicates by classifying each characteristic separately. Each sample's target value is established, and its attributes are then numerically modified with the appropriate weight and priority. An a priori value is provided to significantly reduce the noise introduced by low-frequency features, thereby effectively limiting overfitting and increasing the model's generalization capacity. Recently, this algorithm has become universal and can be used across a wide range of areas and problems [33, 50]. In this study, CatBoost and some other supervised ML models are used to predict the CS of UHPC.

3.1.2. Decision Tree (DT)

DT is a popular ML method that can be applied to solve many complex problems proposed by Quinlan [51]. DT can solve both classification and regression problems. The primary idea behind DT is to use a set of rules to define regions where the output variables match the input variables, and to calculate a constant for each zone. The goal is to build a model that accurately predicts the value of a target variable by inferring fundamental decision rules from the properties of the data. The DT is used to categorize items according to a set of criteria. The DT generates rules for predicting the class of unseen objects based on data from objects with attributes and classes (training data). The components of a DT are the root node, mid-level nodes, leaves, and branches. At the top of the DT is a node whose output is the value of the first variable used to categorize the data. The attributes, which are the data values used to determine which branch to test next, are stored in the internal nodes. Leaf nodes hold the final classification variable's value. Values of independent variables are represented by branches that show how they are related to those of dependent variables. DT has several advantages, including the ability to express information graphically and to visualize and leverage effects. DT can handle missing data and outliers, model nonlinear interactions between dependent and independent variables, and require no mathematical assumptions about the relationship between these two sets of data [52]. On the other hand, DT has significant drawbacks, such as difficulties modeling functions and a tree structure that is sensitive to sample data, since even tiny changes in the training data can result in drastically different outcomes. [53] In this study, the author uses the DT model for regression to predict the CS of UHPC.

3.1.3. Ensemble learning

Ensemble learning is a critical ML technique in which multiple learners are trained to achieve the same classification or regression objective [54]. Unlike ordinary ML models, which aim to learn a single hypothesis from training data, ensemble learning methods create an ensemble of hypotheses and combine them. A fundamental goal of an ensemble model is to reduce variance or bias, thereby minimizing generalization error [55]. This study utilized a stacking ensemble technique to predict the CS of UHPC. The base learners included CatBoost, LightGBM, DT, and GB. Their predictions were subsequently utilized as secondary inputs for a regularized linear meta-learner, specifically RidgeCV. RidgeCV performs ridge regression with cross-validated selection of the regularization parameter, thereby enhancing the stability of the linear combination of base-model outputs and mitigating the risk of overfitting. Consequently, the stacking ensemble was formulated to integrate the complementary predictive patterns of the basic learners while enhancing the robustness of the final prediction.

3.1.4. Light Gradient Boosting Machine (LightGBM)

LightGBM is a gradient boosting framework developed by Microsoft that uses decision-tree algorithms to improve model efficiency and memory utilization. It employs DT algorithms for ranking, classification, and regression. Performance and scalability are the primary development goals. Many of XGBoost's characteristics are carried over into LightGBM. These characteristics include sparse optimization, parallel training, a variety of loss functions, regularization, bagging, and early stopping. However, LightGBM differs in how it constructs trees. In contrast to most other implementations, LightGBM does not build a tree by adding nodes one by one (row by row). LightGBM builds a tree from the bottom up, starting with the leaves. It picks the leaf that it thinks would minimize losses the most. Contrary to XGBoost and other implementations, LightGBM does not use the prevalent sorted-based DT learning approach, which seeks the best split point among sorted feature values. Instead, LightGBM uses a highly optimized histogram-based DT learning approach, which offers considerable advantages in both speed and memory usage. This improvement helps LightGBM increase training speed while reducing memory usage. The LightGBM method uses two unique approaches, namely Gradient-Based One-Side Sampling and Exclusive Feature Bundling, to accelerate computation while retaining a high degree of accuracy. While leaf-wise is desirable for limited datasets, leaf-wise trees often lead to premature overfitting. Therefore, LightGBM uses an additional hyperparameter (max depth) to limit this. Even so, LightGBM is still recommended when the dataset is large enough.

3.1.5. Gradient Boosting (GB)

Gradient Boosting (GB) is one of the most popular and widely used ML techniques and can solve regression and classification problems [56]. The benefits of GB include its flexibility, ability to model feature interactions, and built-in feature selection. The principle of this algorithm is to successively plant numerous DTs using data from previously planted trees. In an additive model, each tree (weak learner) is added to improve the efficiency of the learners that came before it. If the learner model is effective, prediction accuracy rises, and the procedure ends. Therefore, gradient descent on the loss function continuously adjusts the weak learner's weight, improving it and making it a good learner. While other statistical learning systems (such as NN and SVM) may achieve equivalent accuracy, GB's outputs are more easily interpretable without sacrificing precision.

3.2. Cross-validation and Random Sampling

3.2.1. Cross-validation

Cross-validation (CV) is a technique used in ML to counteract overfitting and improve the prediction performance of ML models. In this research, the data is split into a training dataset (70% of the total data) and a testing dataset (30%). To test the model, the test dataset is set aside. Each training iteration selects at random one of the K equal subsets of the training dataset to serve as validation data and the other subsets (K-1) as training data. For this reason, K rounds of cross-validation are conducted in each simulation to guarantee that the training dataset is thoroughly tested. The outcome will be based on a weighted average of K independent evaluations. K must be chosen appropriately; if it is too large, the training dataset will be much larger than the validation dataset. As a consequence, the assessment findings will not accurately represent the underlying nature of the ML approach, especially with large datasets. In this study, the number of cross-validations with K = 10 was selected, which is consistent with the choice of researchers worldwide [57].

3.2.2. Random Sampling

For developing ML models, datasets are very important and should not be biased. Otherwise, the model will produce inaccurate results and be costly and time-consuming. To obtain an objective, generalized dataset, random sampling techniques must be used. Random sampling is a technique in which each data point in a real data set is selected with equal probability. In this technique, many factors must be considered to minimize bias. There are generally four types of random sampling (RS) techniques: simple RS, stratified RS, cluster RS, and systematic RS. Simple RS uses randomly generated integers to select a sample. The first step is to gather all data samples in a sample frame, list, or database. The first n samples may be drawn at random, and Excel could be used to generate the random numbers. For instance, if a sample size of three is needed, select random numbers between 1 and 3. Twenty different randomized data sets were generated in this work for use in twenty different simulated ML model evaluations.

3.3. Partial dependence plots (PDP)

Determining the importance of inputs to a predictor is a key task in any supervised ML algorithm. However, determining the magnitude of the variables' influence is only part of the problem.

Once a subset of important features has been identified, it is necessary to assess how those features (or a subset) affect the predicted result. This may be accomplished in various ways, but in machine learning, PDP is often used. The PDP is a visual representation of the relationships between one or more input variables (typically one to three) and predicted values. [58] PDP enables you to examine how changes in one variable affect the other. Each PDP curve's integral is computed and used as a relevance indicator to assess each input variable's impact. Because it is based on the minimal impact of each predictor on the outcome, this method may be applied to any black-box model. Fundamentally, it is based on the following:

Let Y^j be the considered variable, as well as a list of values established at regular intervals along that range $Y^j = y_k^j$. The formula determines the mean value of predictions:

$$\bar{F}(y_k^j) = \frac{1}{M}\sum_{i=1}^{M} F(y_k^j, y_i^{jc}) \tag{1}$$

where y_i^{jc} is all inputs' value, Y^j is for the observation i, and M is the number of observations.

Similarly, a PDP plot representing interactions among multiple inputs can be constructed. Then, the mean values are calculated for the pairs y_k^j, where j takes two different values. As a result, a 2D Partial dependence plot (2D PDP) can be built. The 2D PDP represents pairwise interactions for all considered inputs. In this study, 2D PDP is used to study the influence of 15 input variables (Cement, Slag, Silica fume, Limestone powder, Quartz powder, Fly ash, Nano Silica, Water, Sand, Gravel, Fiber, Superplasticizer, Relative humidity, Temperature, Age) on the output variable, which is the CS of UHPC.

3.4. Performance indices of models

This study used the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) to evaluate the predictive performance of ML models. Precisely, R^2 is used to estimate the correlation between the actual and predicted values. The value of R^2 varies from -∞ to 1, and the higher the R^2, the more accurate the model's prediction. The Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) both quantify the typical deviation from the predicted value. Smaller RMSE and MAE values indicate higher accuracy and greater performance. The following formulas calculate the RMSE, MAE, and R^2:

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{k=1}^{m}\left(s_k - \hat{s}_k\right)^2} \tag{2}$$

$$\mathrm{MAE} = \frac{1}{m}\sum_{k=1}^{m}\left|s_j - \hat{s}_j\right| \tag{3}$$

$$R^2 = 1 - \frac{\sum_{k=1}^{m}\left(s_j - \hat{s}_j\right)^2}{\sum_{k=1}^{m}\left(s_j - \bar{s}_j\right)^2} \tag{4}$$

where s_k is the experimental value, $\hat{s}_k$ is the model's output value, $\bar{s}_k$ is the average of experimental values, and m is the total number of samples.

4. Methodology flowchart

The four key phases of the technique used in the current study are described in detail after an introductory section, a section on database development, and a section on methods (see Fig. 2):

Dataset preparation: In this step, the authors collected 810 data points from reputable journals, including 15 input parameters and 1 outcome parameter. The random sampling approach is used to randomly divide 20 data sets in a 7/3 ratio, with 70% of the total dataset being utilized to train the models (training set), and 30% of the whole dataset remaining being used to test the models (testing set).

Building model: During this period, only the training set is used to train the models. The trained models do not know the testing dataset. The model is validated using 10-fold cross-validation. Five methods are used, including CatBoost, DT, GB, LightGBM, and the stacking Ensemble model.

Testing model: The remaining 30% of the dataset (the test dataset) is used to test the models. The predictability of the proposed model is assessed using the performance indices mentioned earlier.

Sensitivity analysis: Partial dependence plots (PDP) are used to assess the impact of 15 input parameters on UHPC's CS, with the findings shown in subsection 5.3.

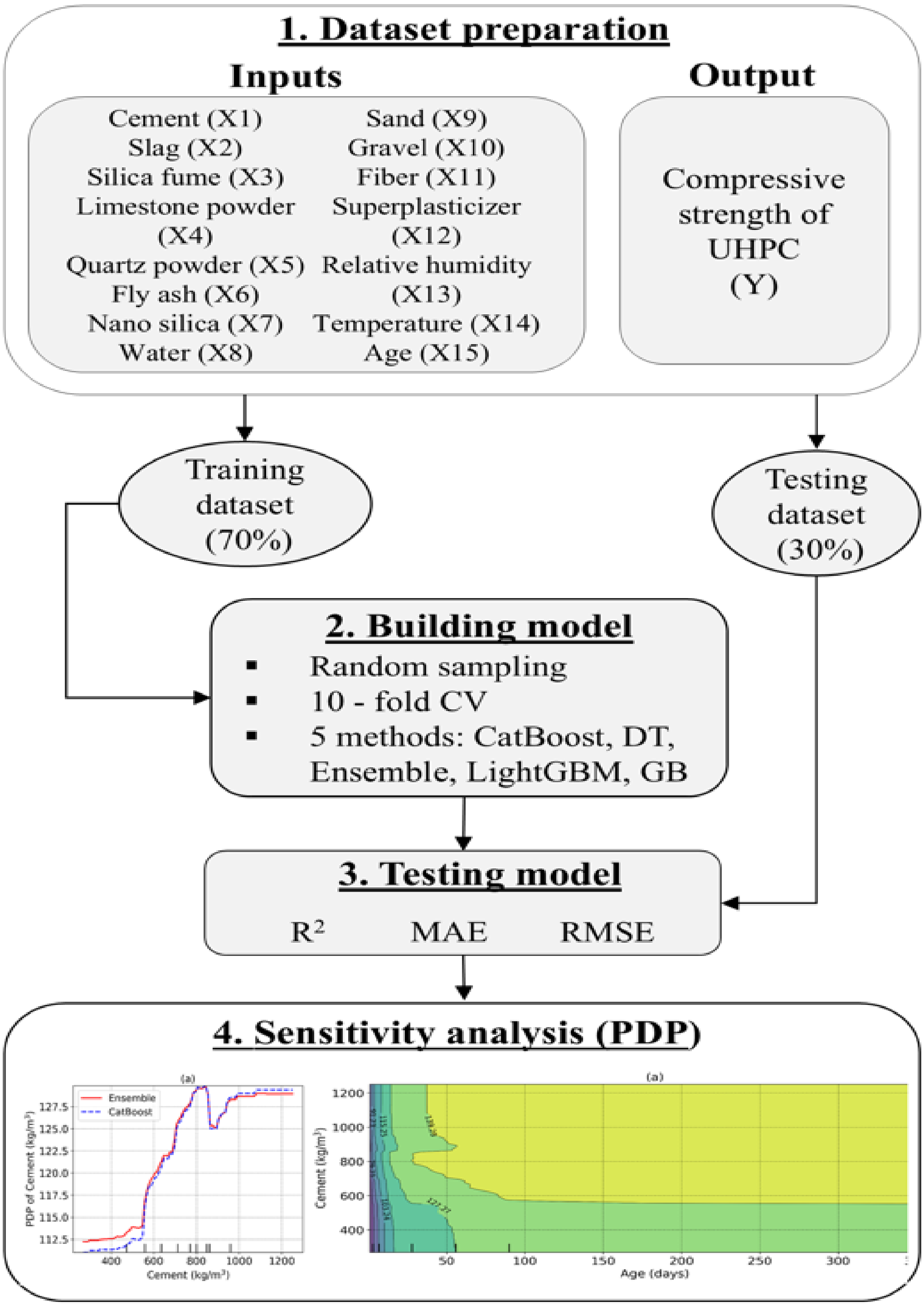

Fig. 2. Methodology flowchart of the present work

5. Results and Discussion

5.1. Performance evaluation of machine learning algorithms

In this study, Monte Carlo simulation (MC) and K-Fold Cross-Validation (CV) were used to assess the models' generalization and predictive performance. The Monte Carlo approach generates 20 data sets in a 7:3 ratio, each with various disturbances. 70% of the total data is used to train the model (567 samples, the training set), and 30% is used to test the model (243 samples, the test set). Here, the 7/3 ratio for generating the dataset was chosen based on recommendations from previous work [59, 60]. After that, the 10-fold cross-validation (CV10-Fold) is employed to control overfitting, a common issue in ML model training. CV10–Fold is applied only to the training and validation sets during the training and validation periods, and the models are uninformed of the test set. In detail, the training is divided into 10 equal parts. One portion will be utilized as validation data (the validation dataset) during each training session, while the remaining nine parts will be used to train models. A total of ten runs will be required to guarantee that each element of the training dataset is verified only once. The average values of 10 times are the outcomes of the model's predictive assessment criteria for the training and validation periods. Hyperparameters significantly affect the performance of machine learning models. Thus, this study implemented a systematic trial-based search (manual grid) as directed by the notebook code: for each model, essential hyperparameters were adjusted within specified ranges (see Table 2), while other parameters remained at their default settings. The final parameters were chosen to minimize the 10-fold cross-validation RMSE on the training dataset and reduce the training-validation disparity. This technique is replicable because the search ranges and selected values are documented in detail. Table 2 lists the parameters of the ML algorithms used in this study.

Fig. 3 presents the changes in RMSE, MAE, and R2 across 20 MC simulations in the training and validation sets. In the training set, low RMSE and MAE and high R2 indicate that the five proposed ML models perform well. According to Figs. 3a, c, and e, GB achieved the best training performance, with average R^2, RMSE, and MAE values of 0.994, 3.159 MPa, and 1.143 MPa, respectively. The predictive ability of ML models can be arranged based on the R^2 values from the training phase: GB > Ensemble > CatBoost > DT > LightGBM. From both RMSE and MAE perspectives, a similar sequence was seen.

Table 2. Parameters chosen for different models used in this study

General hyperparameters	Description	General hyperparameters	Description
DT		**CatBoost**	
max_depth	20	depth	6
min_samples_leaf	2	iterations	1000
min_samples_split	2	grow_policy	SymmetricTree
min_impurity_decrease	0	bootstrap_type	Bayesian
max_leaf_nodes	None	**Ensemble Stacking**	
LightGBM		base estimators	CatBoost, LightGBM, DT, GB
boosting_type	gbdt	meta-learner/final estimator	RidgeCV
num_leaves	31	regularization strategy	Ridge regression with cross-validated regularization strength
n_estimators	100		
GB			
Learing_rate	0.1		
n_estimators	100		
subsample	1		

However, the exceptional training performance of GB should not be construed as indicative of higher generalization capability. A distinct training–validation disparity was observed for this model: the mean R^2 decreased from 0.994 during training to 0.940 in validation, whilst RMSE increased from 3.159 MPa to 9.473 MPa. This signifies a high propensity for overfitting in the GB model. Overfitting can be ascribed to the considerable flexibility of sequential tree boosting, which is susceptible to tree depth, the number of estimators, and inadequate regularization when modeling intricate tabular data. In contrast, CatBoost and the stacking ensemble model exhibited a reduced training-validation discrepancy, indicating superior resilience and generalization. The Ensemble model exhibited superior validation performance, achieving an average R^2 of 0.961 and an RMSE of 7.673 MPa, indicating that integrating complementary base learners via a regularized meta-learner effectively mitigated overfitting.

Consequently, while GB closely aligns with the training data, the validation results suggest that Ensemble and CatBoost yield more reliable predictive performance for novel UHPC mixes. The mean, standard deviation, minimum, and maximum values of R^2, RMSE, and MAE for five methods, obtained from training and validation, are listed in Table 3.

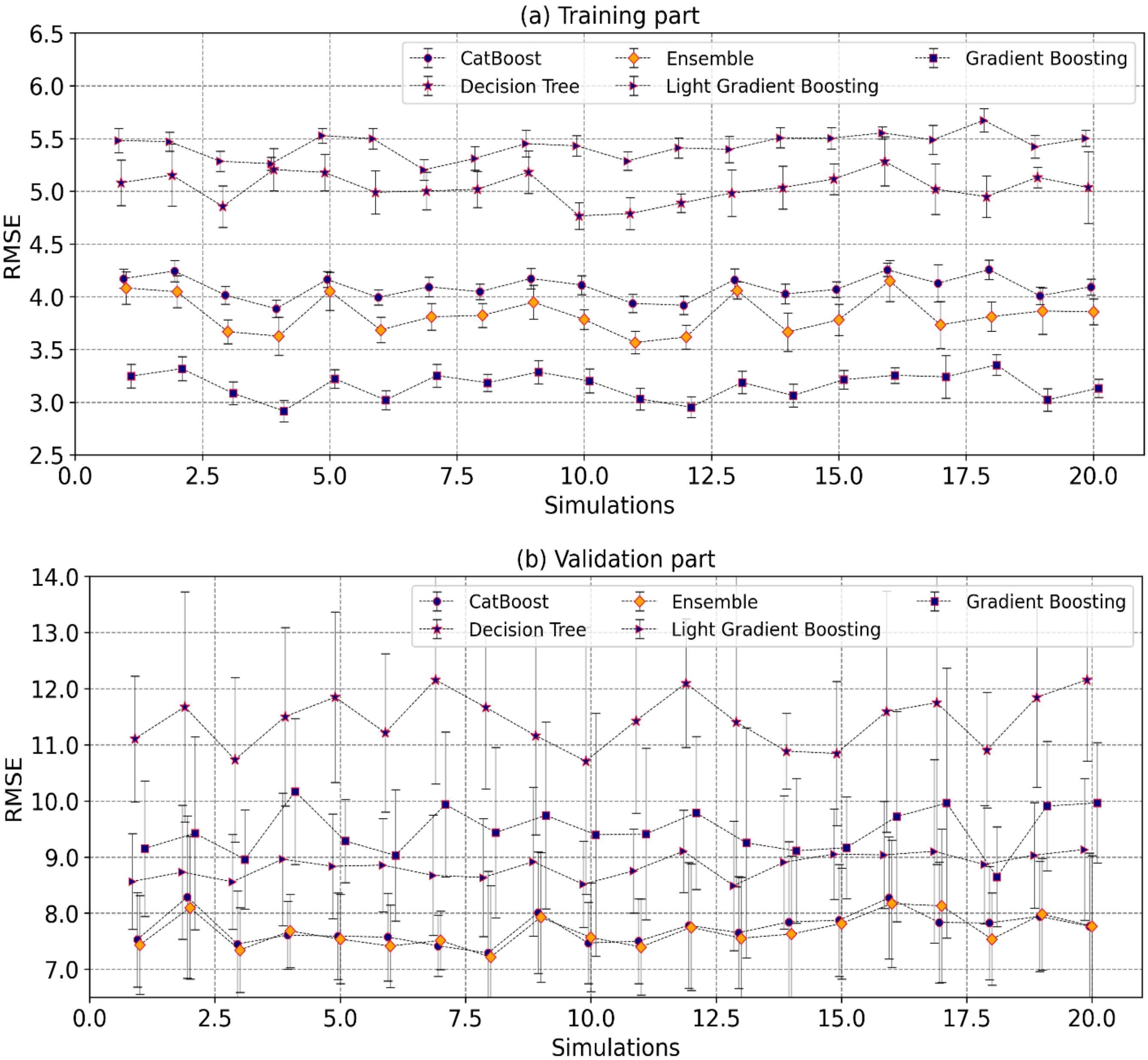

Fig. 3. Results of 20 Monte Carlo runs of training and validation parts (a, b) RMSE; (c, d) MAE; (e, f) R^2

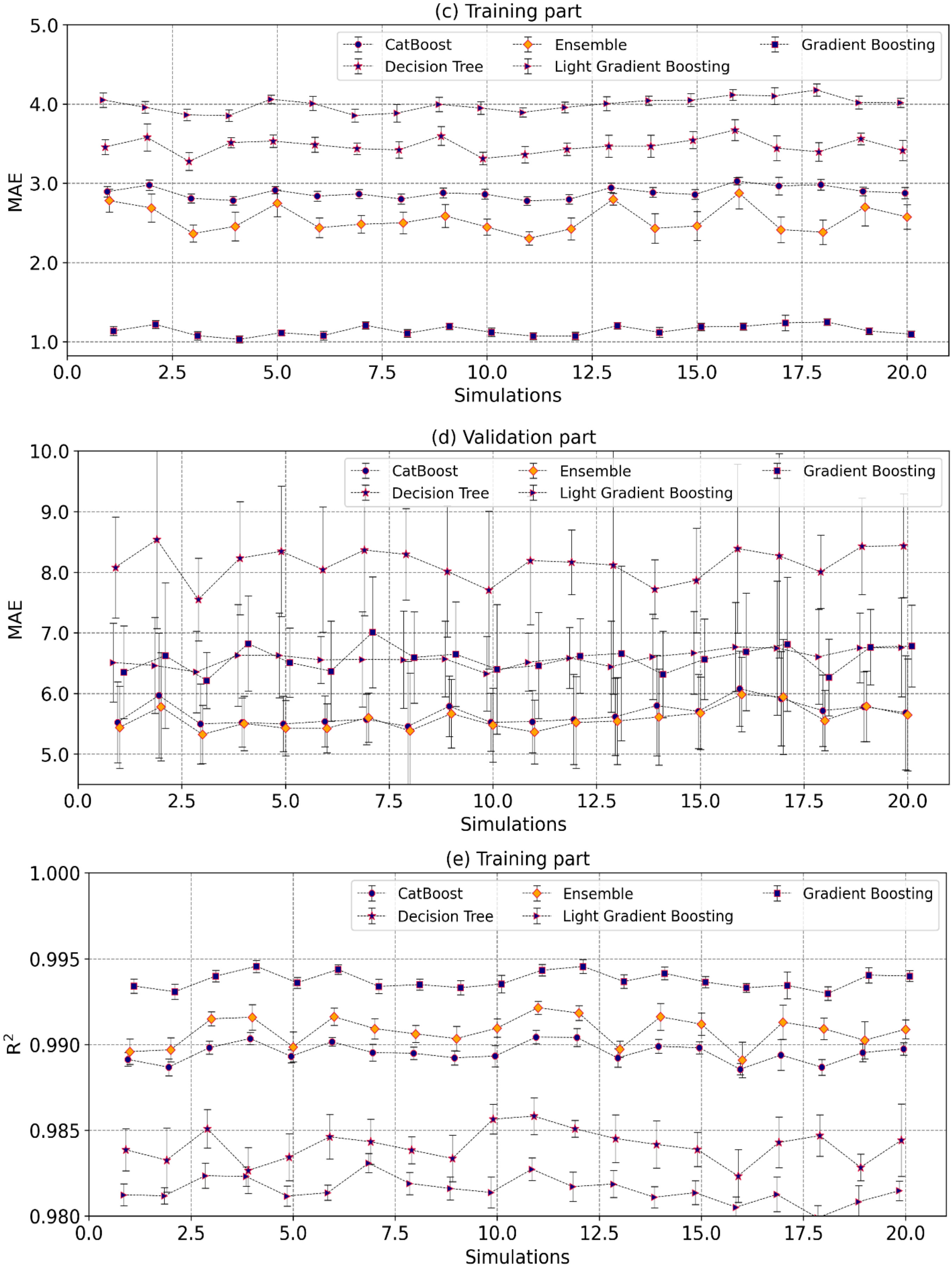

Fig. 3. (continued)

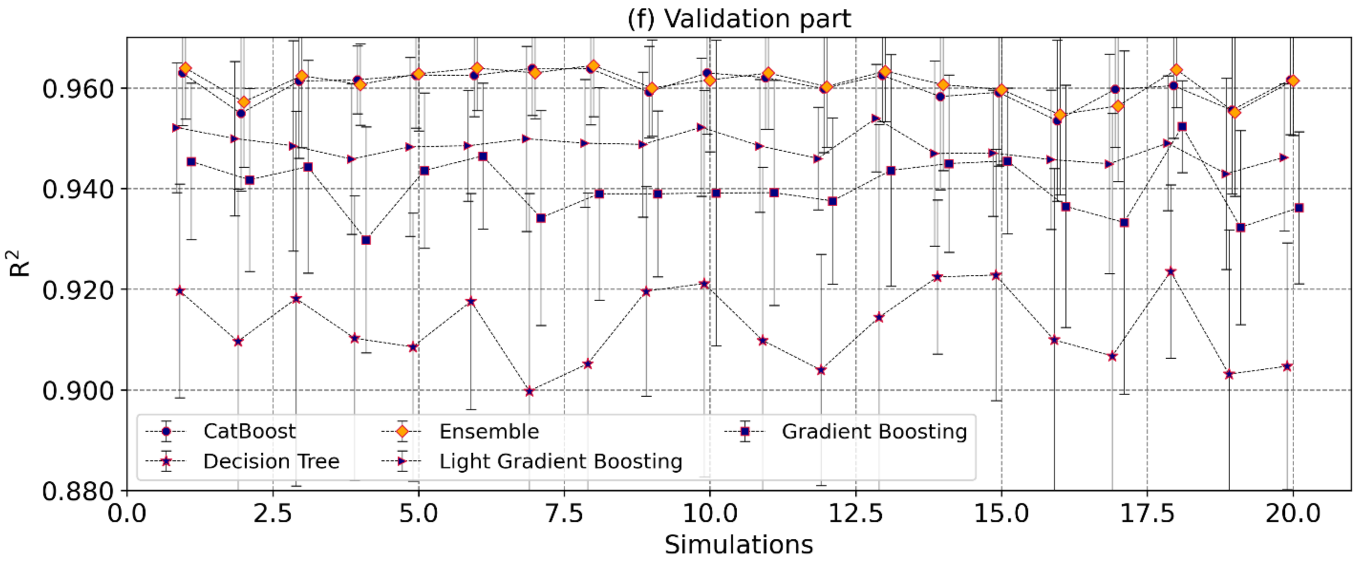

Fig. 3. (continued)

Table 3. Summary of prediction results of training and validation parts

	Metrics	Dataset	Mean	Std	Min	Max
CatBoost	RMSE	Training	4.087	0.106	3.886	4.257
		Validation	7.726	0.263	7.290	8.287
	MAE	Training	2.882	0.069	2.777	3.027
		Validation	5.665	0.172	5.457	6.078
	R^2	Training	0.990	0.001	0.989	0.990
		Validation	0.960	0.003	0.953	0.964
DT	RMSE	Training	5.033	0.135	4.767	5.283
		Validation	11.434	0.461	10.710	12.157
	MAE	Training	3.469	0.094	3.277	3.673
		Validation	8.138	0.262	7.551	8.538
	R^2	Training	0.984	0.001	0.982	0.986
		Validation	0.913	0.007	0.900	0.924
Ensemble	RMSE	Training	3.832	0.170	3.566	4.149
		Validation	7.673	0.268	7.216	8.169
	MAE	Training	2.544	0.161	2.306	2.876
		Validation	5.583	0.179	5.324	5.984
	R^2	Training	0.991	0.001	0.989	0.992
		Validation	0.961	0.003	0.955	0.964
LightGBM	RMSE	Training	5.433	0.112	5.203	5.673
		Validation	8.836	0.206	8.488	9.134
	MAE	Training	3.993	0.088	3.854	4.177
		Validation	6.580	0.122	6.326	6.767
	R^2	Training	0.982	0.001	0.980	0.983
		Validation	0.948	0.003	0.943	0.954
GB	RMSE	Training	3.159	0.121	2.917	3.353
		Validation	9.473	0.398	8.647	10.169
	MAE	Training	1.143	0.063	1.031	1.249
		Validation	6.574	0.206	6.215	7.009
	R^2	Training	0.994	0.000	0.993	0.995
		Validation	0.940	0.005	0.930	0.952

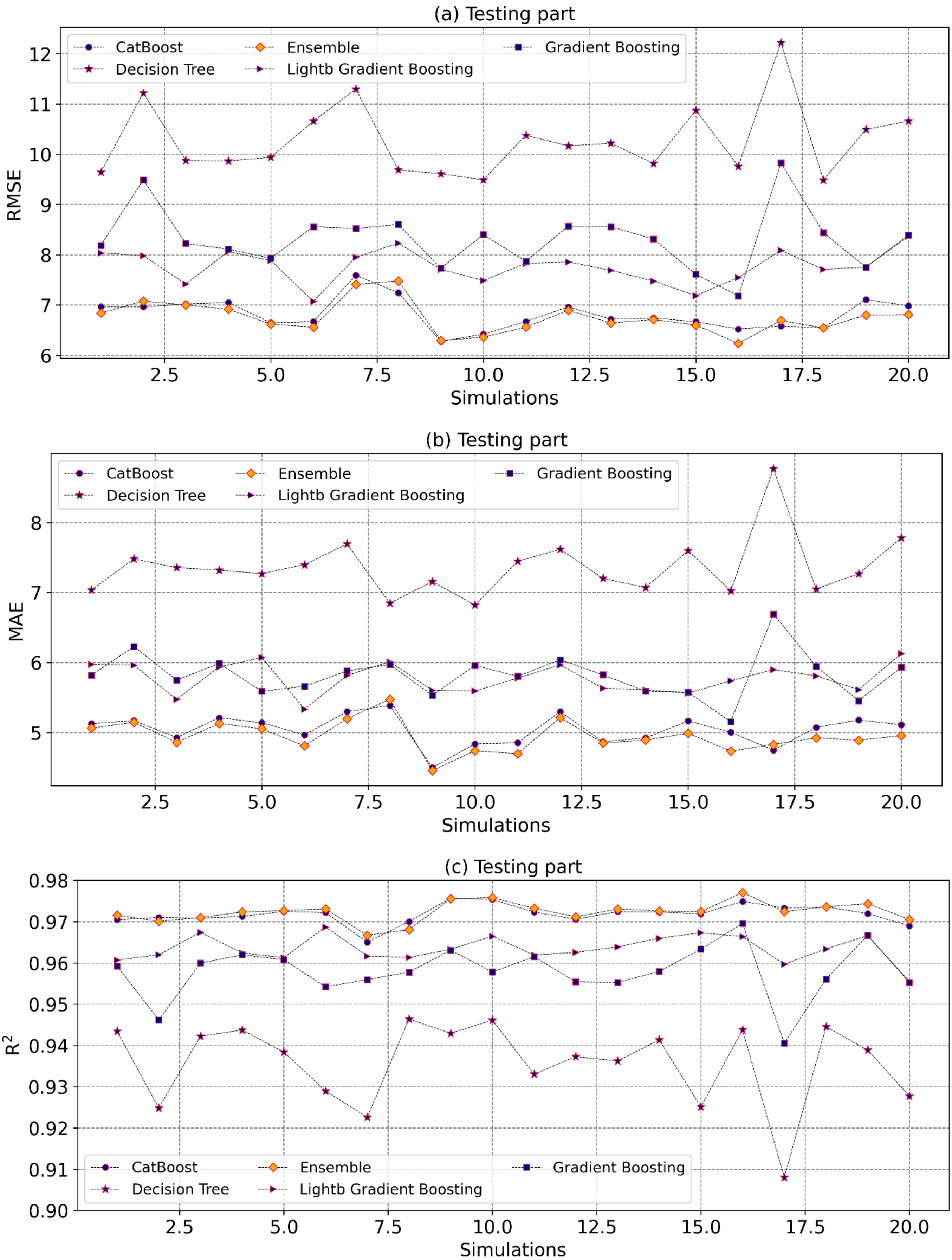

Fig. 4. Results of 20 Monte Carlo runs of testing parts (a) RMSE; (b) MAE; (c) R^2

Table 4. Summary of prediction accuracy on the testing part of different models considered in this study

	Metrics	Mean	Std	Min	Max
CatBoost	RMSE	6.817	0.303	6.276	7.591
	MAE	5.037	0.209	4.495	5.384
	R^2	0.972	0.002	0.965	0.976
DT	RMSE	10.268	0.700	9.485	12.224
	MAE	7.360	0.416	6.821	8.770
	R^2	0.936	0.010	0.908	0.946
Ensemble	RMSE	6.751	0.316	6.236	7.479
	MAE	4.943	0.219	4.455	5.473
	R^2	0.972	0.002	0.967	0.977
LightGBM	RMSE	7.766	0.325	7.071	8.368
	MAE	5.774	0.211	5.333	6.126
	R^2	0.963	0.003	0.955	0.969
GB	RMSE	8.313	0.585	7.182	9.828
	MAE	5.818	0.311	5.153	6.691
	R^2	0.958	0.006	0.941	0.970

The results suggested that the predictability of the testing set varied. As shown in Fig. 4, the five models' stability and predictability are as follows: Ensemble > CatBoost > LightGBM > GB > DT. An overview of the test-set prediction results from 20 simulations of the five suggested models is provided in Table 4. The results indicated that the DT model's prediction accuracy was lower and less consistent than that of the other four models. For instance, the average R^2 value is approximately 0.936 and ranges substantially, with R^2 = 0.908 and R^2 = 0.946 representing the lowest and greatest values, respectively. In addition, DT yielded higher RMSE and MAE values than the other models, suggesting lower predictive accuracy and weaker robustness. In contrast, after 20 runs, the Ensemble model showed small variations between the minimum and maximum values for each evaluation metric. The testing results showed R^2 values ranging from 0.967 to 0.977, RMSE values between 6.236 and 7.479 MPa, and MAE values between 4.455 and 5.473 MPa, indicating consistent predictive capability. The testing results corroborate the overfitting diagnosis identified for GB in the training-validation study. Despite GB attaining the highest training accuracy, its testing performance was inferior to that of Ensemble, CatBoost, and LightGBM, with an average R^2 of 0.958 and an RMSE of 8.313 MPa. This indicates that the robust training fit of GB did not result in optimal generalization performance. The Ensemble and CatBoost models exhibited the most dependable equilibrium between accuracy and resilience. The enhanced stability of the Ensemble model is due to the stacking method, which combines the predictions of various base learners using the regularized RidgeCV meta-learner.

To determine whether the supplied dataset was adequate for stable model building, a learning curve analysis was also conducted (Appendix Fig. A1). The stacking Ensemble model exhibited a fairly consistent bias–variance trade-off across the examined data domain, as evidenced by the validation RMSE declining with increasing training size and approaching a plateau at larger sample sizes.

5.2. Representative prediction results

These representative results are based on the testing set used for the selected representative run. In Figs. 5a, b, c, d, and e, the typical regression charts for the CatBoost, DT, Ensemble, LightGBM, and GB models are shown. Here, the regression model shows the relationship between the ML model's predicted output (from the training dataset) and the target value (from the experiment). The fit line, composed of points whose predictions match experimental findings, is the ideal line. The model's predictive accuracy increases as the data points approach the ideal fit line. As shown, the values derived from the CatBoost (Fig. 5a) and Ensemble (Fig. 5c) models are quite similar to the target values. Some results in the DT model are relatively far from the fit line (Fig. 5b). These findings demonstrate the ability of Ensemble and CatBoost to establish a link between input and output parameters and to produce accurate predictions.

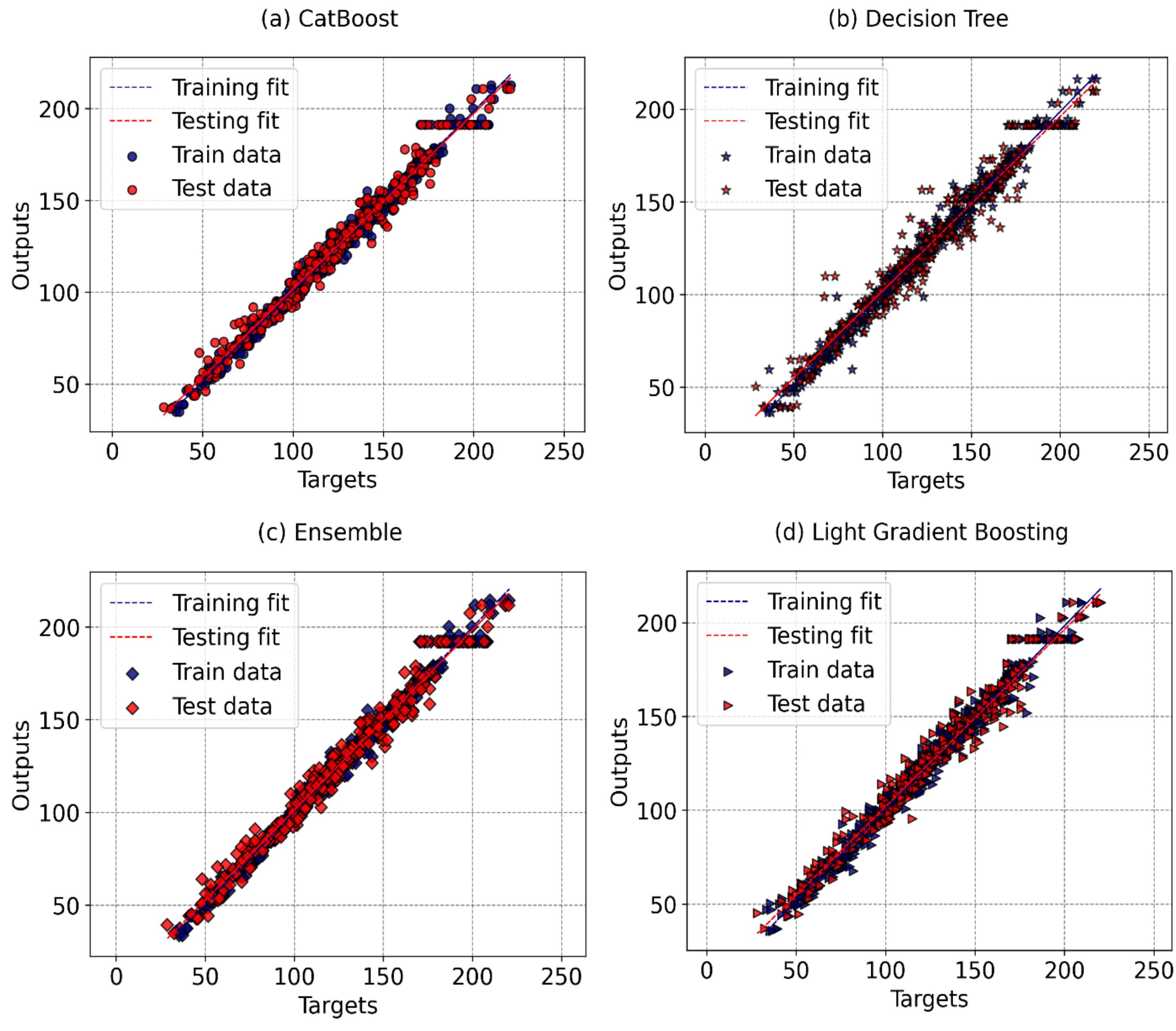

Fig. 5. Regression analysis of ML models considered in this study: (a) CatBoost, (b) DT, (c) Ensemble, (d) LightGBM, and (e) GB.

Fig. 6 shows the relative errors, defined as (abs(Targets-Outputs)/Targets), for the five ML models considered in this study. It is clear that the DT model performs the least well in terms of prediction, and its training and testing phases exhibit a wider relative error range than those of the other models. Most data points with the DT model have errors between [0, 0.2]. The testing portion included 8 samples (around 3.29%) with errors greater than 0.2, whereas the training portion included 22 samples (around 3.88%) with errors greater than 0.2, with a maximum value of 0.78. Compared with DT, the Ensemble, CatBoost, LightGBM, and GB models generally exhibited lower relative errors for most samples. However, several individual samples showed substantially larger errors, particularly near the boundaries of the dataset. As shown, most errors in the training data set for each of the five models are minor, demonstrating high training accuracy as previously indicated. When using the DT model, the prediction error is slightly higher. Except for a few abnormal samples, the Ensemble and CatBoost models consistently yield the lowest error. The results show that the Ensemble and CatBoost models outperform the others in this test section. Samples with the highest relative errors were selected to investigate failure scenarios in more detail. An examination of Fig. 6c reveals that the largest relative errors exceed 0.3 for samples near the dataset boundaries. The database infrequently represents specimens from very early ages, mixes cured at severe temperatures, and combinations of supplementary cementitious materials. On the other hand, in the main part of the dataset where there are many samples, the relative error is close to zero, showing that the Ensemble model is trustworthy within the range of data used.

(e) Gradient Boosting

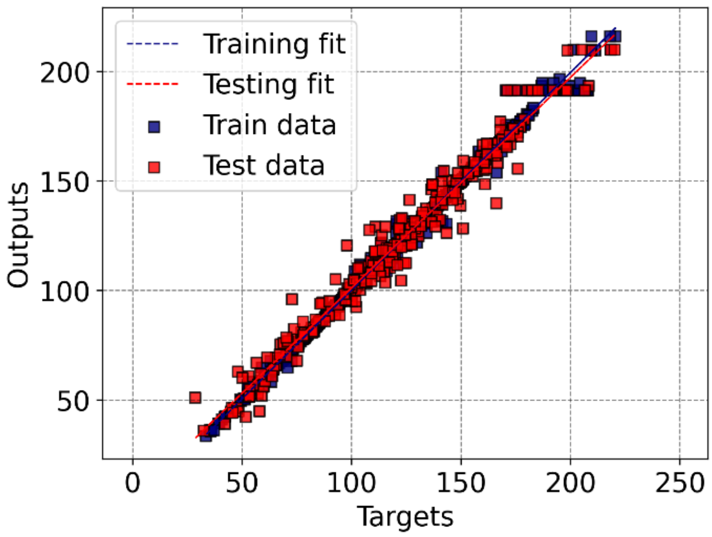

Fig. 5. (continued)

(a) CatBoost

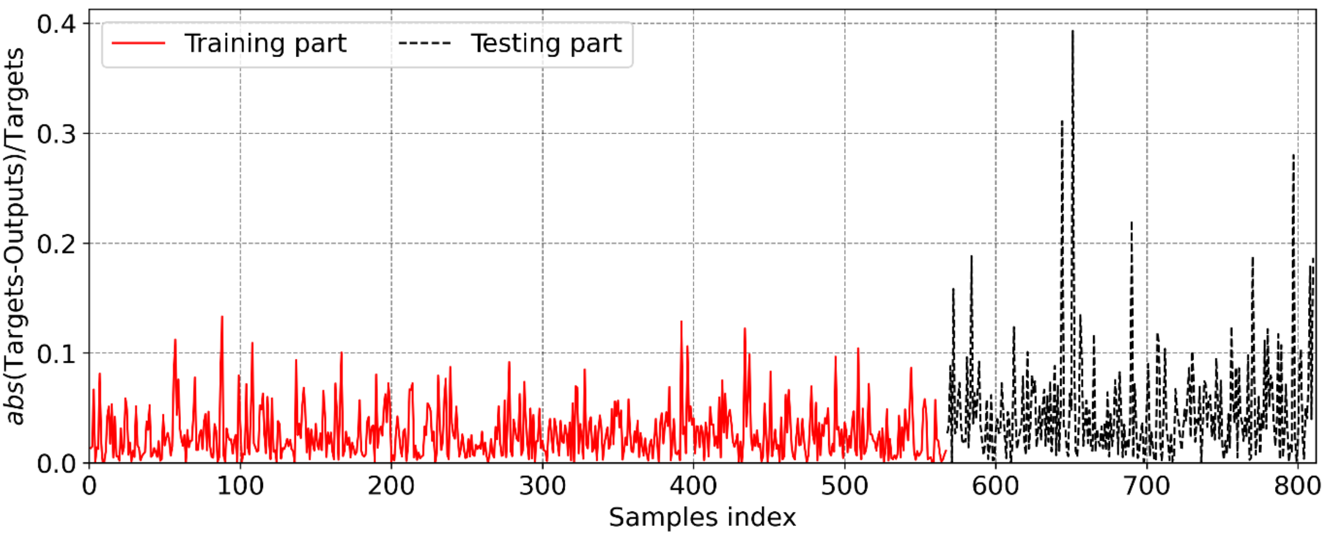

Fig. 6. Error graph of ML models considered in this study: (a) CatBoost, (b) DT, (c) Ensemble, (d) LightGBM, and (e) GB.

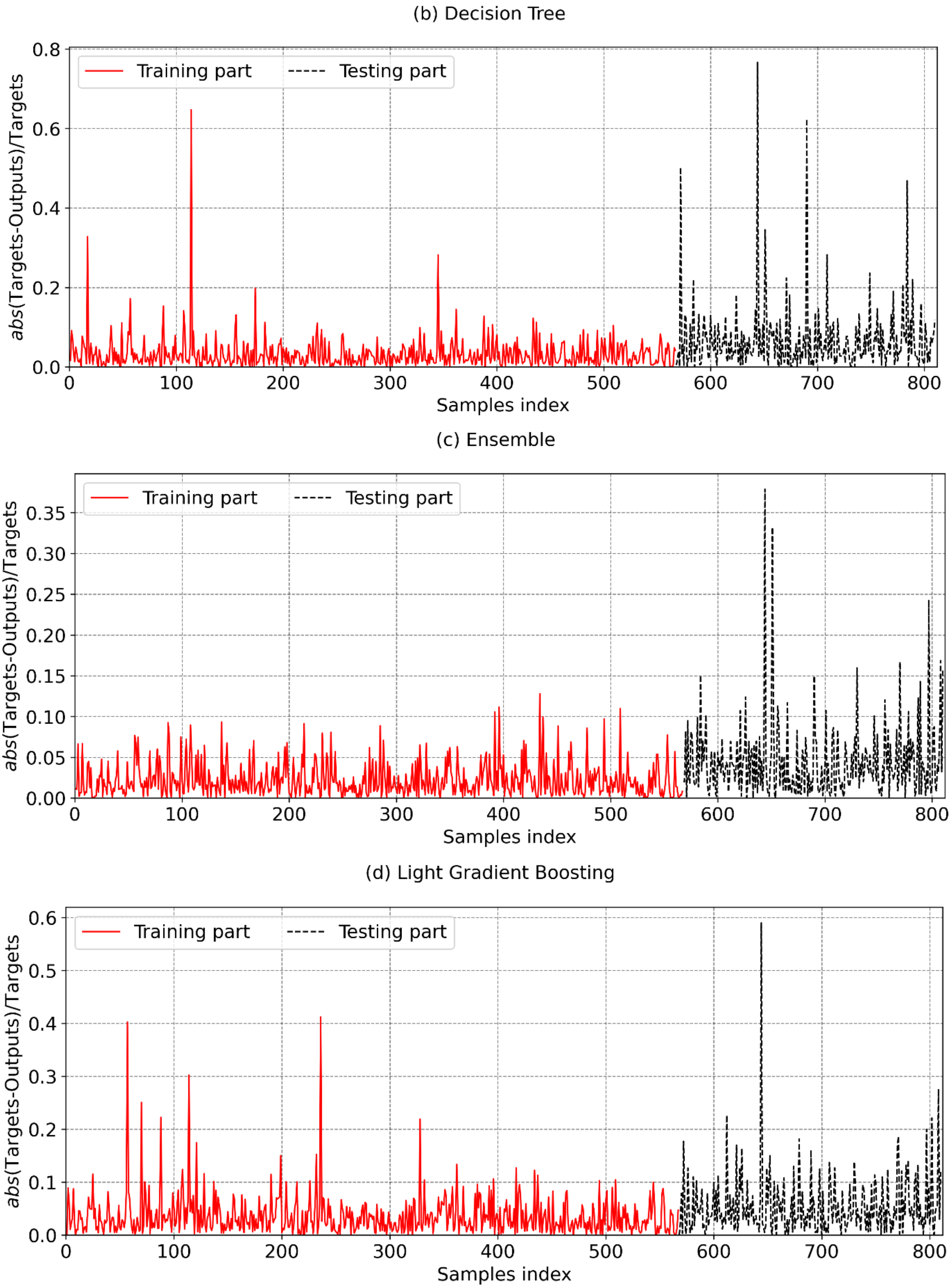

Fig. 6. (continued)

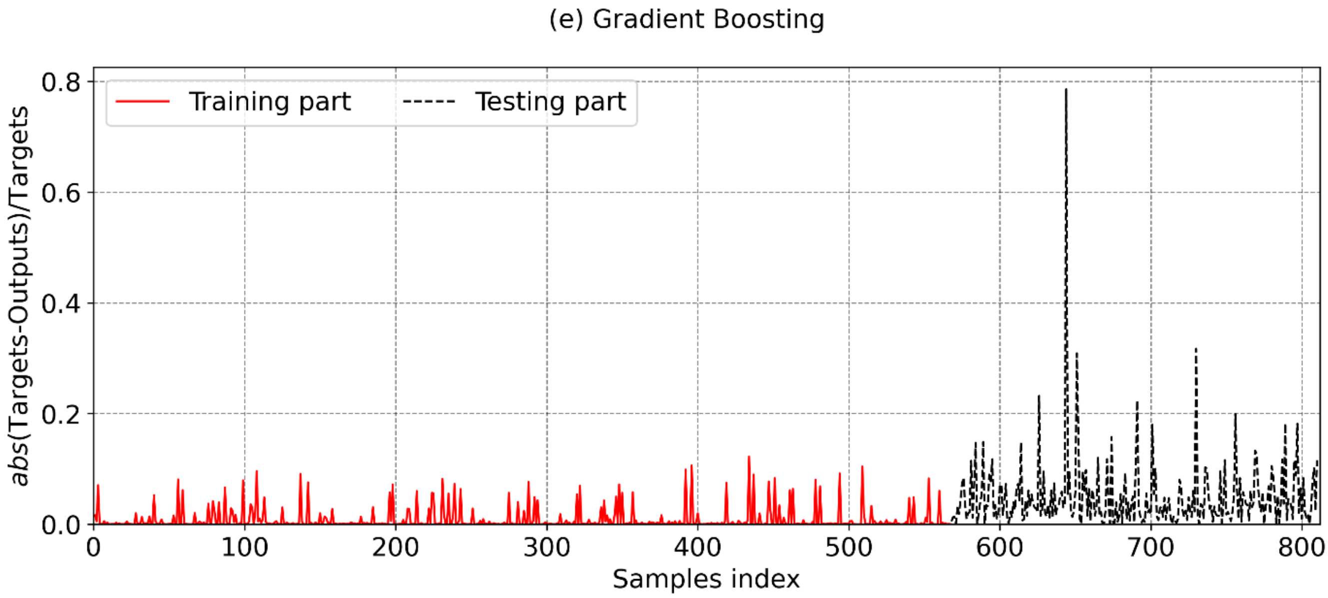

Fig. 6. (continued)

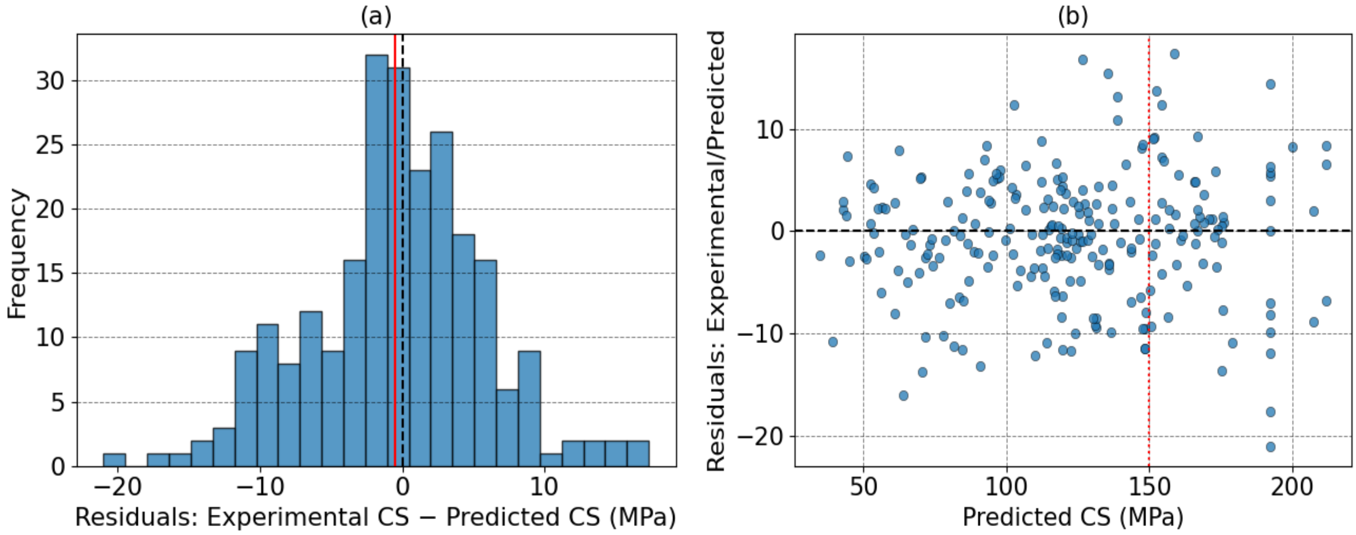

Fig. 7. Residual diagnostic analysis of the stacking Ensemble model on the testing set: (a) distribution of residuals and (b) residuals versus predicted compressive strength.

Table 5. The evaluation criteria for the representative prediction results for the training and testing parts

	Dataset	RMSE (MPa)	MAE (MPa)	R^2
CatBoost	Training	4.311	3.068	0.988
	Testing	6.522	5.001	0.975
DT	Training	5.302	3.670	0.982
	Testing	9.760	7.023	0.944
Ensemble	Training	4.003	2.679	0.990
	Testing	6.236	4.733	0.977
LightGBM	Training	5.466	4.011	0.981
	Testing	7.546	5.738	0.966
GB	Training	3.302	1.246	0.993
	Testing	7.182	5.153	0.970

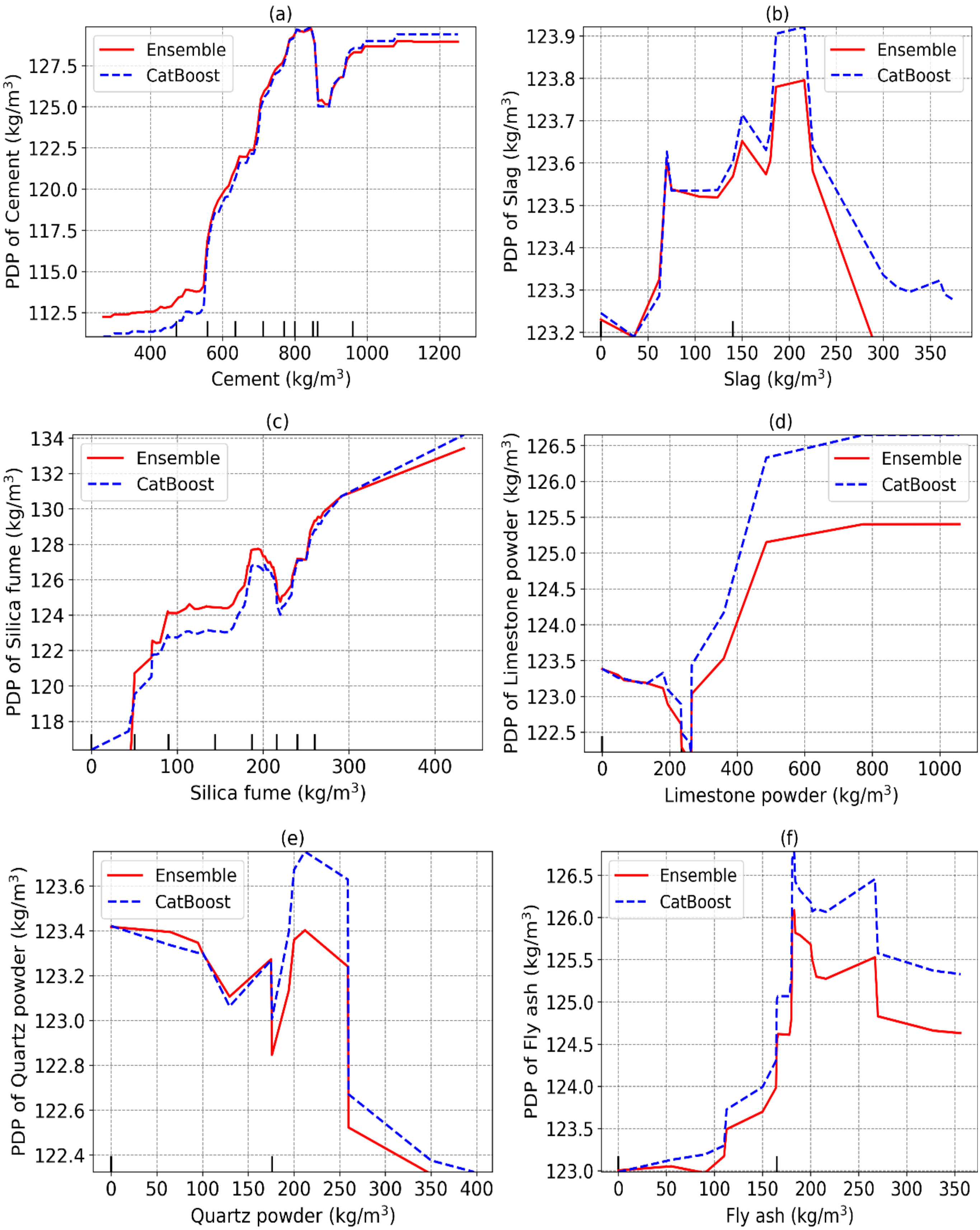

Fig. 8. PDP analysis of inputs considered in this work: (a) Cement; (b) Slag; (c) Silica fume; (d) Limestone powder; (e) Quartz; (f) Fly ash; (g) Nano silica; (h) Water; (i) Sand; (j) Gravel; (k) Fiber; (l) Superplasticizer; (m) Relative humidity; (n) Temperature; and (o) Age.

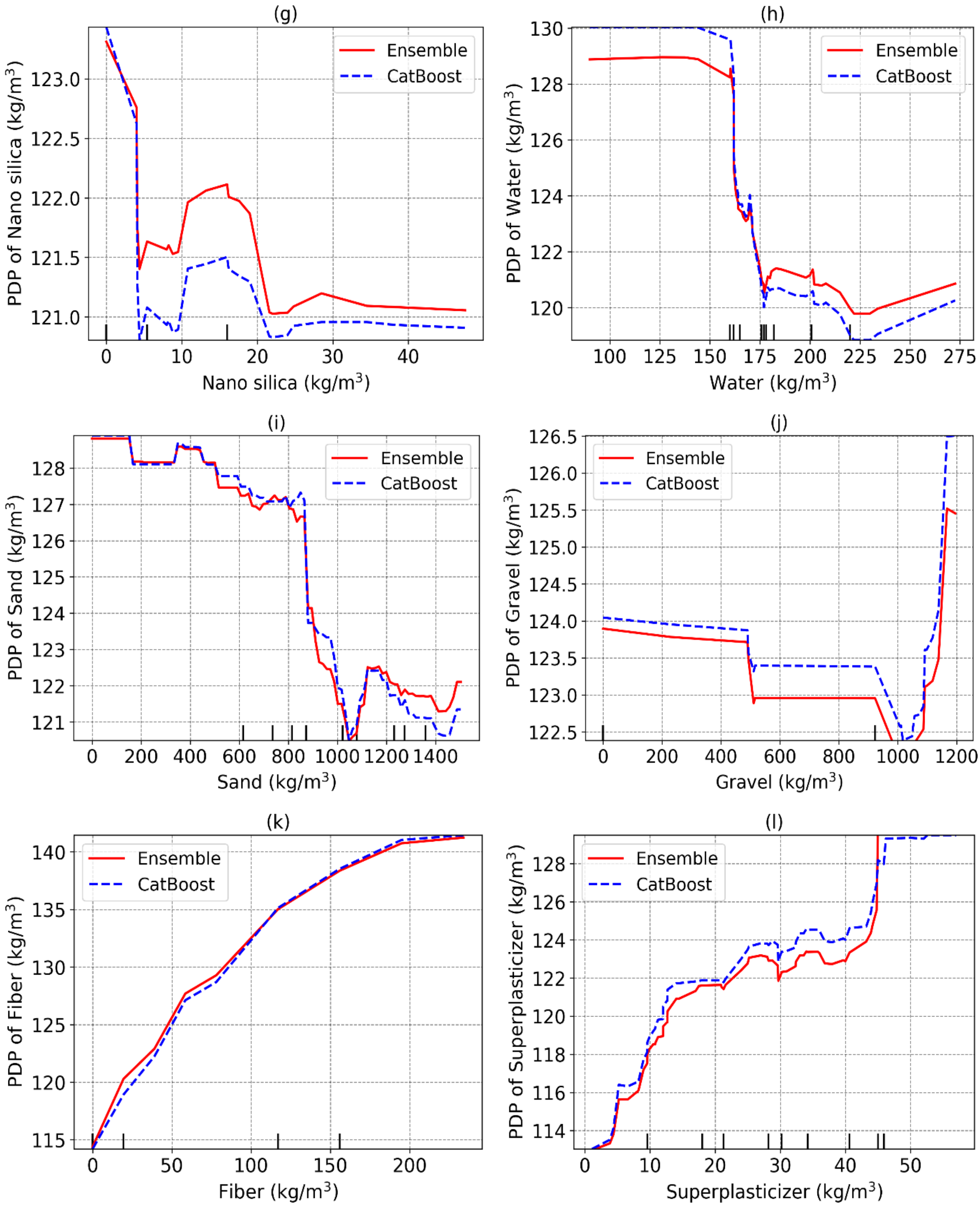

Fig. 8. (continued)

Residual diagnostics were performed further to investigate the error structure of the stacking ensemble model, as illustrated in Fig. 7. The residuals were determined by subtracting the predicted CS from the experimental CS. Fig. 7a shows that the residuals were predominantly centered at zero, indicating no systematic overprediction or underprediction bias, though a minor inclination towards overprediction was observed. Fig. 7b illustrates the residuals in relation to the expected CS. The residuals were dispersed around zero without a distinct funnel-shaped pattern, indicating that severe heteroscedasticity was not apparent from visual examination. Nevertheless, the residual spread increased marginally in the high-strength range, especially for anticipated compressive strength values exceeding roughly 150 MPa. This suggests that predicting very high-strength UHPC mixes is more difficult, possibly due to their sparse representation and the intricate interplay of binder composition, fiber content, superplasticizer dosage, and curing circumstances. Consequently, while the Ensemble model demonstrates consistent overall predictive ability, prediction in the high-strength regime should be approached with caution and ideally corroborated through experimental validation.

Finally, Table 5 presents the prediction metrics for each model. It is important to note that only one typical prediction result is shown in this section. The computed RMSE, MAE, and R^2 values for the Ensemble were 6.236, 4.733, and 0.977, respectively. For CatBoost, the values were 6.522, 5.001, and 0.975. It is reasonable to conclude that Ensemble and CatBoost are effective methods for precisely predicting the CS of UHPC.

5.3. Sensitivity analysis

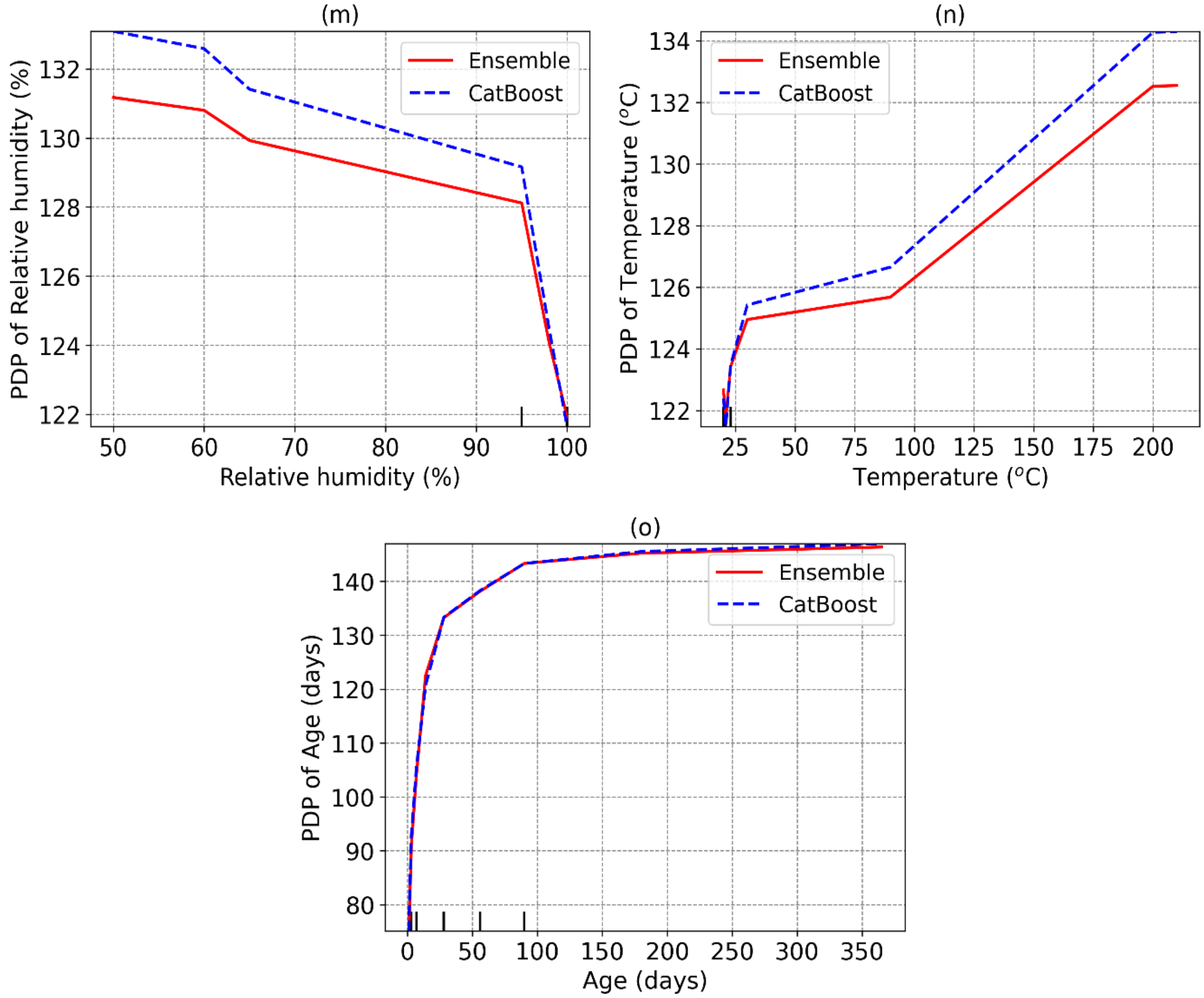

Fig. 8. (continued)

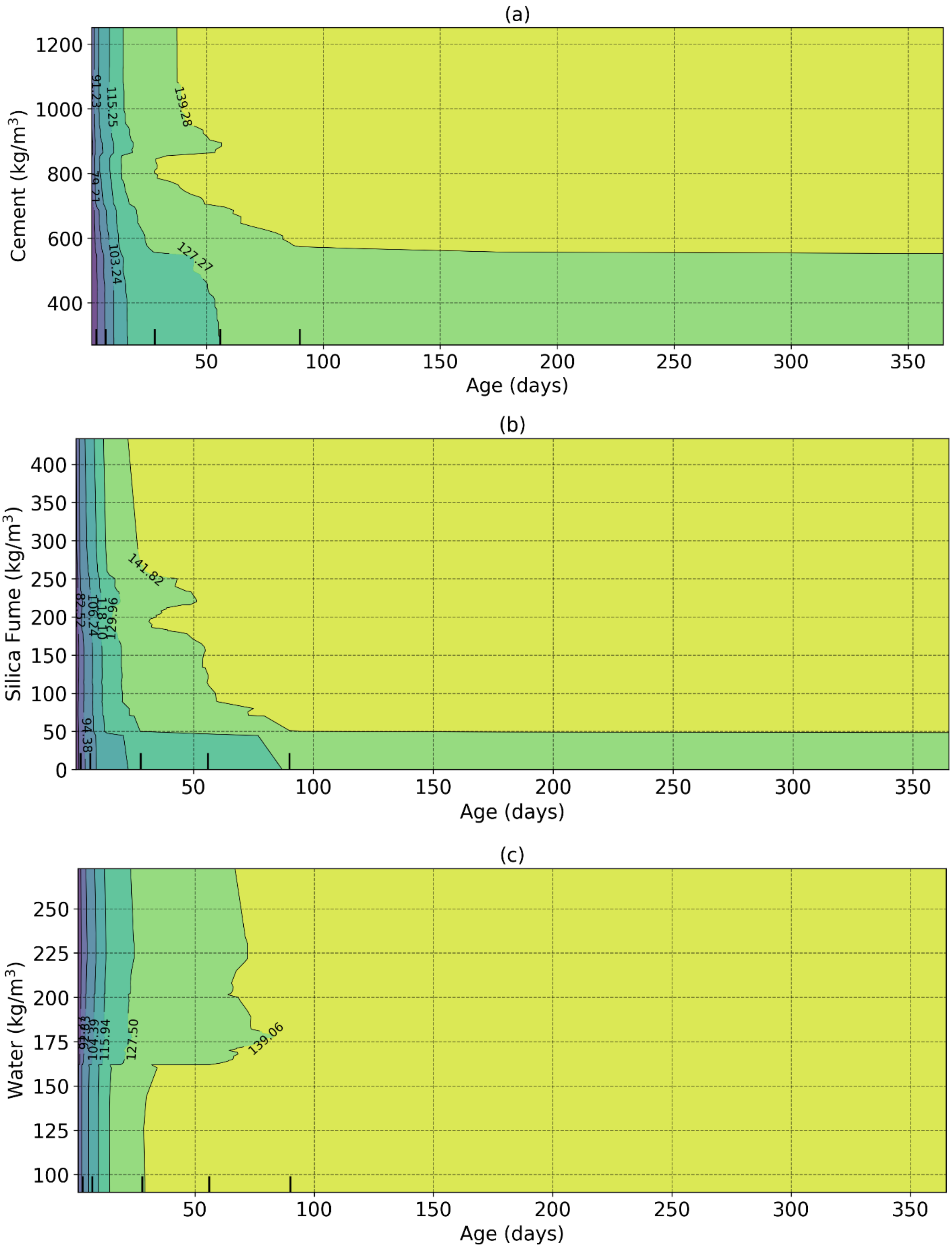

Fig. 9. Feature interaction effect analysis using 2D PDP: (a) Cement–Age, (b) Silica fume–Age, (c) Water–Age, (d) Sand–Age, (e) Fiber–Age, (f) Superplasticizer–Age, and (g) Silica fume–Nano-silica.

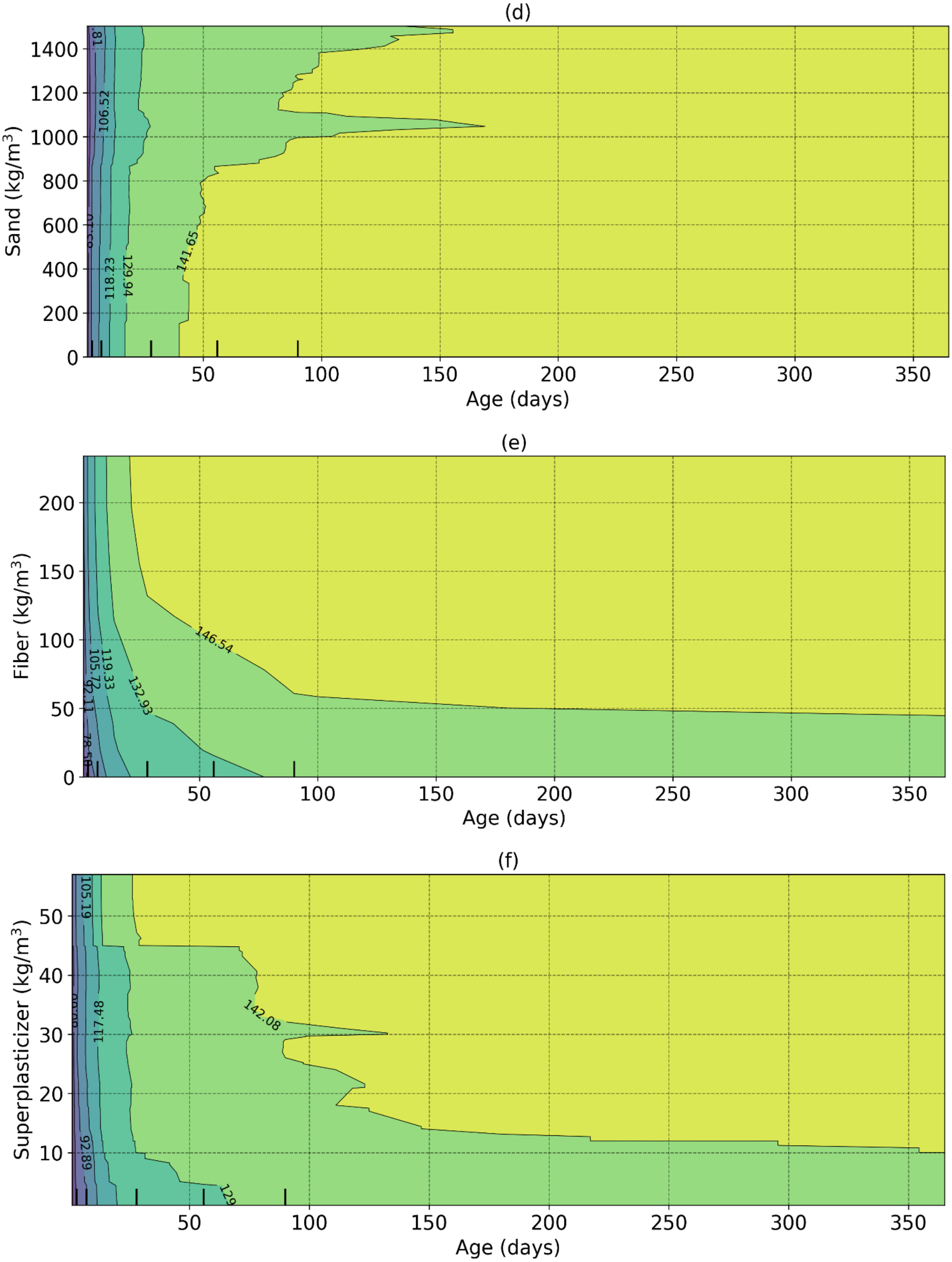

Fig. 9. (continued)

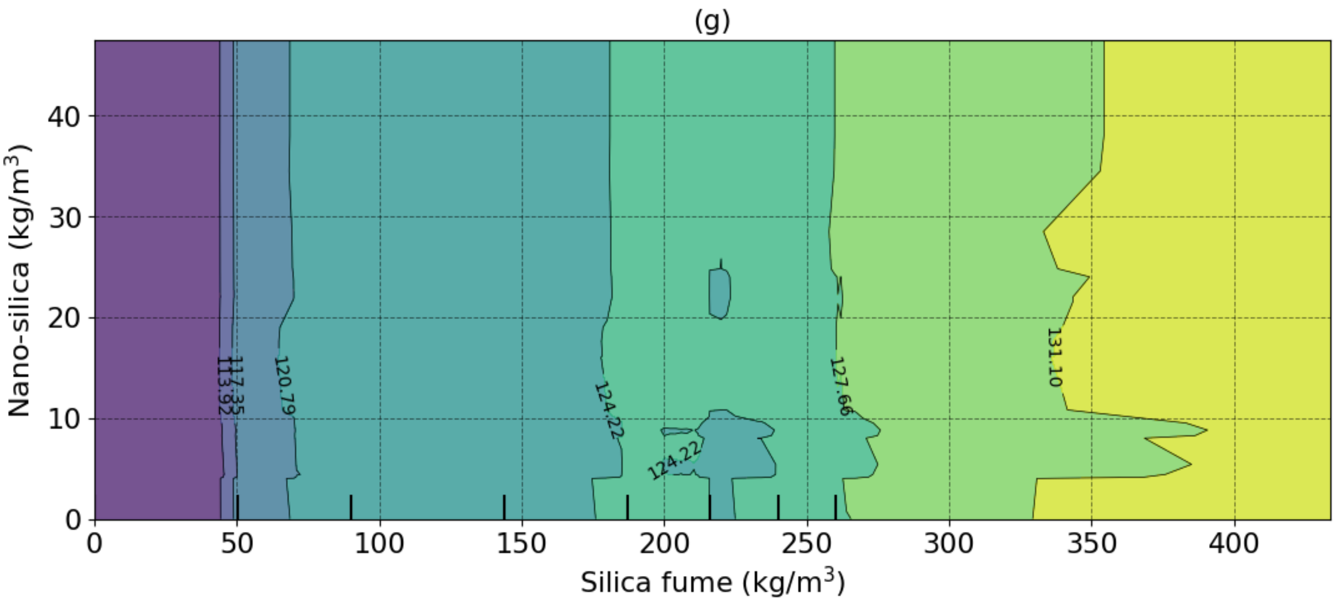

Fig. 9. (continued)

A PDP analysis was established (Fig. 8) to evaluate the impact of input variables on the CS of UHPC. In this technique, when analyzing the influence of each variable, the others were kept constant. The partial dependence of each affected input variable can be determined from the PDP results. The changes in the PDP of input parameters obtained from CatBoost and Ensemble models were generally similar. In particular, the tendency and values of PDP in the two methods were almost the same for Cement (Fig. 8a), Silica fume (Fig. 8c), Sand (Fig. 8i), Fiber (Fig. 8k), Temperature (Fig. 8n), and Age (Fig. 8o). The influence of each parameter on the CS of UHPC was observed and can be grouped into three main groups. The first group includes Cement (Fig. 8a), Silica fume (Fig. 8c), Fiber (Fig. 8k), Superplasticizer (Fig. 8l), Temperature (Fig. 8n), and Age (Fig. 8o), which have positive effects on the CS of UHPC. On the contrary, the second group consists of Quartz (Fig. 8e), Sand (Fig. 8i), and Relative humidity (Fig. 8m), which have an adverse impact on the CS of UHPC. The effect of the remaining input parameters in the third group on UHPC compression strength was inconsistent. For instance, regarding slag (Fig. 8b), in the range [0, 200] kg/m^3, the CS of UHPC increased with increasing slag content. Herein, slag replacement has a positive physical influence due to its filler feature, stimulating hydration (i.e., quickly producing portlandite) [61]. However, the CS of UHPC decreased as the slag content exceeded 200 kg/m^3, in agreement with previous findings [62, 63]. The CS variation for each parameter can also be calculated from the maximum and minimum PDP values for each input parameter. Each input parameter's value reflects the parameter's impact on the expected output. Based on Fig. 8, the impact of parameters reduced from the order of Age, Cement, Silica fume, Water, Sand, Fiber, and Superplasticizer that were employed to establish PDP 2D in the next part.

Fig. 9 shows the contribution and combined effect of inputs on the CS of UHPC. PDP 2D analysis is applied to examine the effect of input variable interactions on the CS of UHPC, using the Ensemble model selected in the previous section. Cement–Age (Fig. 9a), Silica fume–Age (Fig. 9b), Water–Age (Fig. 9c), Sand–Age (Fig. 9d), Fiber–Age (Fig. 9e), Superplasticizer–Age (Fig. 9f), and Silica fume–Nano-silica (Fig. 9g) are the seven primary input-variable interactions identified by 2D PDP analysis. In general, the effects of cement, silica fume, water, sand, and superplasticizer were insignificant, as evidenced by small changes up to 28 days. From the ages of later 28 days to around

50 days, the CS of samples gradually increased and reached a stable value at later ages. Here, changes in cement hydration products over time could be the main reason [64, 65].

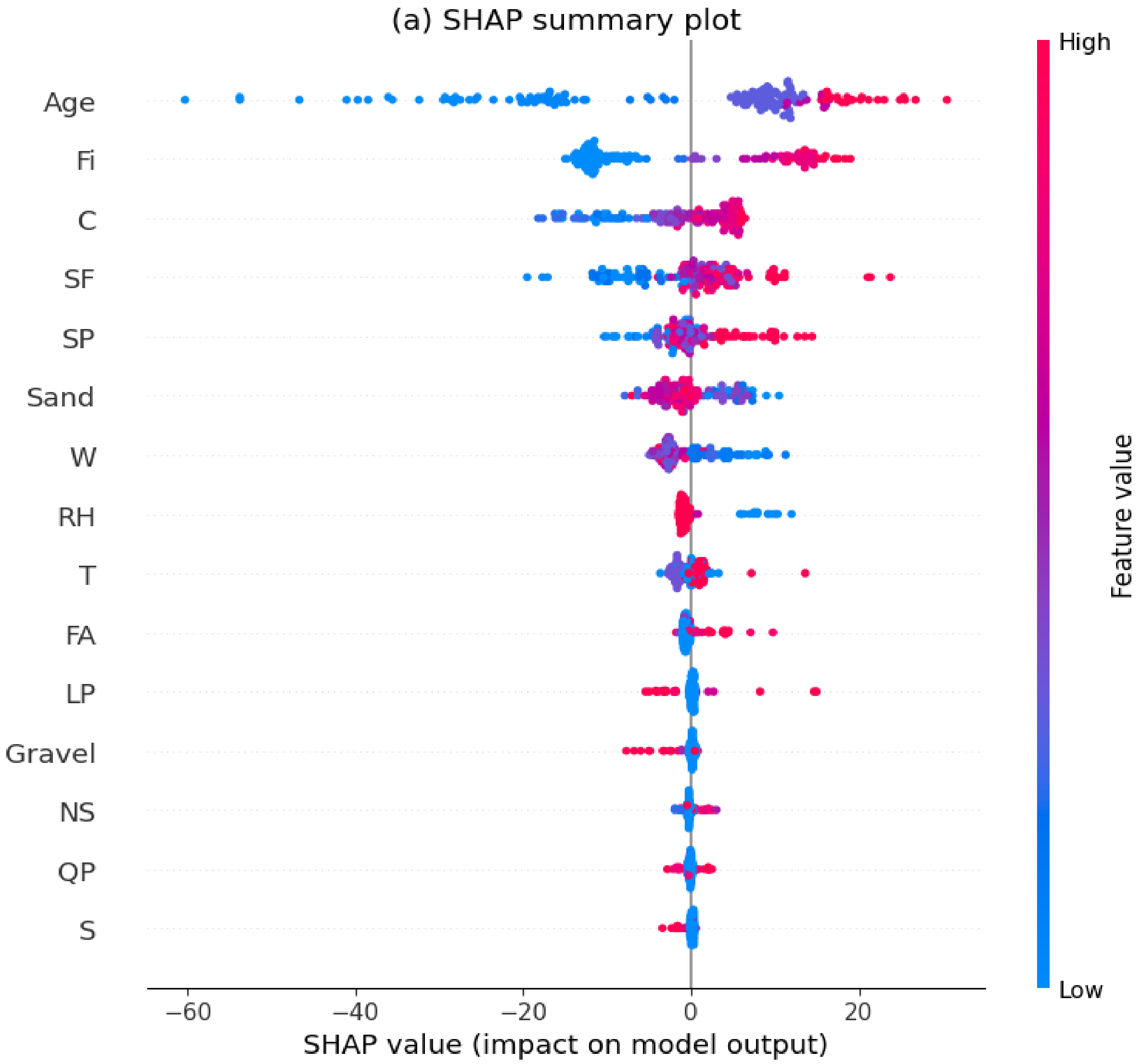

Fig. 10. SHAP-based interpretability analysis of the stacking Ensemble model: (a) SHAP summary plot and (b) mean absolute SHAP values

For instance, Fig. 9a indicates that, from 0 to around 20 days of exposure, the influence of cement (i.e., from 300 to 1250 kg/m^3) on the CS of UHPC was insignificant. At 28 to around 50 days, the CS of samples with a cement content of 300 to 600 kg/m^3 increased gradually to around 127.27 MPa and remained stable at later ages. Whereas for samples with a cement content greater than 600 kg/m3, the CS increased rapidly and reached a very high value (exceeding 139.28 MPa) at 40 days, as often observed in UHPC investigations [49, 63, 66]. Fig. 9g presents the supplementary SF–NS interaction plot, offering enhanced understanding of the synergistic function of siliceous additives in UHPC. The PDP surface indicates that the SF concentration mostly influences the predicted CS: as SF content rises to approximately 300–400 kg/m³, the predicted CS escalates from roughly 113–120 MPa to about 131 MPa, while the variance along the NS axis remains relatively constrained. This indicates that the SF–NS interaction does not exhibit a robust monotonic synergy but instead reflects an SF-dominated, dosage-sensitive complementing effect. This trend is physically justified, as SF primarily enhances particle packing and facilitates pozzolanic reactions, whereas NS may provide additional nucleation sites and expedite early C–S–H formation. The limited impact of NS suggests that its supplementary effect may reach saturation when adequate SF is already available, and agglomeration, heightened water/superplasticizer requirements, or diminished workability may also contribute to excessive NS. Given the scarcity of NS-containing combinations in the current database, this PDP should be seen as a model-derived inclination within the examined data area rather than as direct evidence of a universal synergistic process.

To enhance the PDP-based interpretation, a model-agnostic SHAP analysis was performed for the stacking ensemble model, as illustrated in Fig. 10.

Fig. 10a displays the SHAP summary plot, where each point corresponds to a UHPC mixture, and its x-axis location indicates whether a feature enhances or diminishes the predicted compressive strength relative to the baseline prediction. The hue denotes the feature value, ranging from low to high. This plot indicates that elevated Age values often increase the CS prediction. In contrast, diminished Age values significantly decrease it, aligning with the positive age effect illustrated in Fig. 8o and the Age-related interaction trends depicted in Fig. 9. Cement and silica fume have predominantly beneficial effects at elevated concentrations, corroborating the PDP-based conclusion that binder hydration, particle packing, and pozzolanic reactions are critical determinants of UHPC strength.

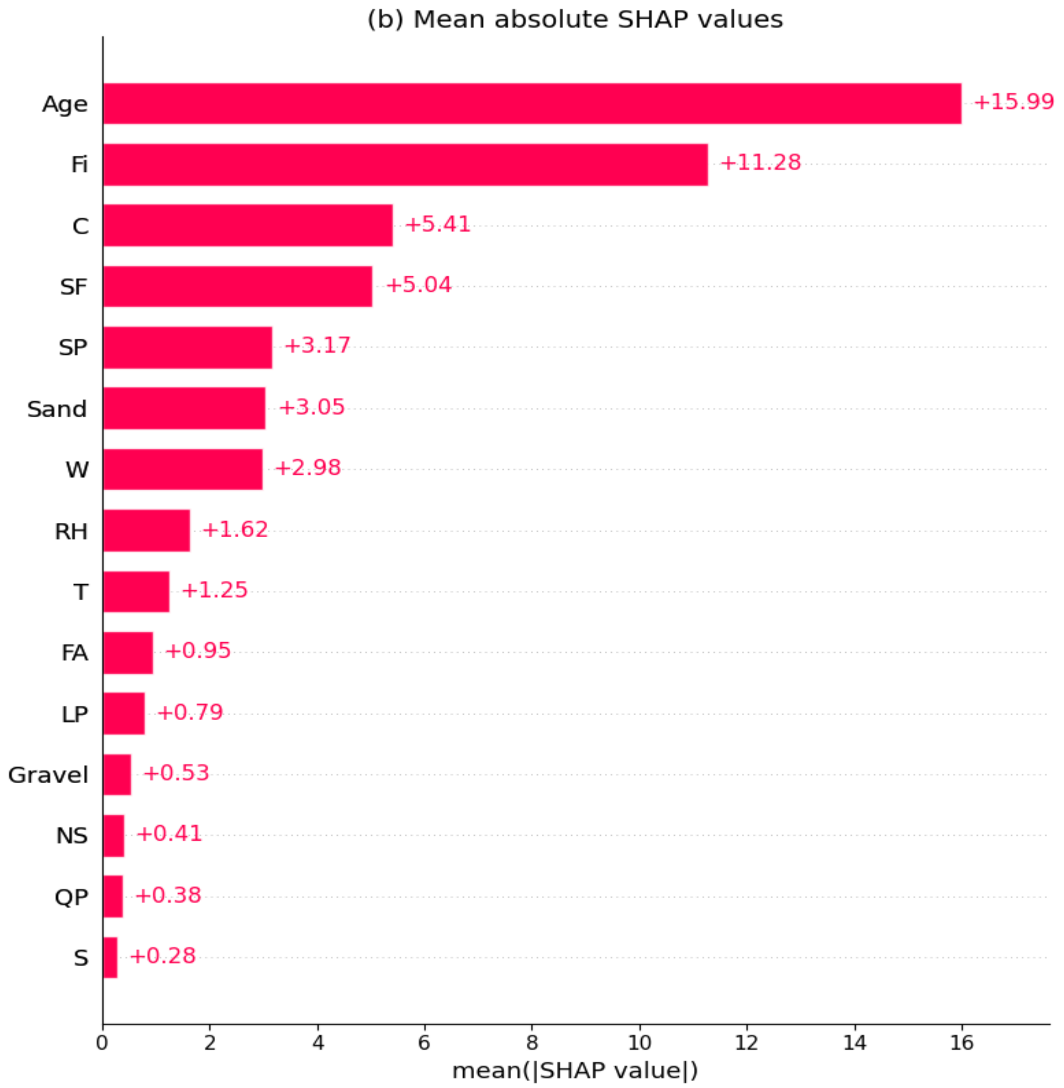

Fig. 10. (continued)

Fig. 10b encapsulates the mean absolute SHAP values and presents a comprehensive ranking of feature significance. The primary variables were Age, Fiber, Cement, Silica fume, Superplasticizer, Sand, and Water, which align closely with the significant variables reported in the PDP analysis presented in Figs. 8 and 9. The significant contribution of Fiber underscores its role in fracture mitigation and its correlation with high-performance UHPC mixtures, whereas Superplasticizer, Sand, and Water highlight the need for dispersion, workability, and packing equilibrium. Conversely, Nano-silica, Quartz powder, Slag, and Gravel exhibited reduced mean absolute SHAP values, indicating minimal model contribution within the current dataset. SHAP corroborates the primary themes identified in the PDP analysis and provides localized, instance-specific elucidations. These results should be interpreted as contributions from model-derived features rather than as direct causal effects, especially for variables with limited distributions. It is essential to note that variables exhibiting sparse or zero-inflated distributions, including S, FA, LP, QP, NS, and Gravel, necessitate careful interpretation. Tree-based models can handle zero-inflated variables via threshold-based splitting; however, the effects learned in high-dosage regions depend on the density of the available data. Consequently, the trends of PDP and SHAP for these variables should be regarded as model-derived patterns within the current data domain, rather than as impartial causal effects throughout the entire potential dose range.

5.4. Domain of applicability and limitations

While the proposed Ensemble model demonstrated significant predictive accuracy, its predictions must be understood within the context of the application established by the training dataset. This work utilized 810 UHPC combinations from the literature for model training, including diverse binder compositions, aggregate quantities, fiber amounts, admixture dosages, curing durations, and environmental variables. Consequently, the model is most reliable for UHPC combinations with input variables within the ranges and distributions delineated in Table 1.

Specific observation is required regarding environmental variables, including temperature and RH. The database indicates a nominal temperature range of 20–210°C; however, most data points cluster between 20°C and 23°C, with high-temperature curing conditions represented by a small number of samples. Likewise, RH ranges from 50% to 100%, with most measurements clustered around 95–100%. As a result, the model predictions are anticipated to be more reliable for normal-temperature, high-humidity curing conditions, while extrapolation to dry curing environments, intermediate relative humidity ranges that are inadequately represented, steam curing, or extremely high-temperature curing should be approached with caution.

The same consideration applies to mixture variables with sparse or zero-inflated distributions, including GGBS/slag, fly ash, limestone powder, quartz powder, nano-silica, and gravel. For these factors, numerous combinations in the database exhibit negligible or minimal contents, while regions with high dosages are sparsely populated. Consequently, PDP patterns and model predictions in these sparse regions may reflect data-driven tendencies rather than universal material principles.

Accordingly, the proposed model should be used primarily as a decision-support and virtual-screening tool for UHPC mixture design within the investigated data domain. Predictions for combinations or curing regimes beyond this domain must be experimentally confirmed prior to practical application. The future augmentation of the database with more equitable distributions of temperature, relative humidity, and SCM dose would enhance the model's extrapolative robustness and usefulness.

6. Conclusions and Perspectives

This study investigated the performance of five ML models, including Decision Tree (DT), Gradient Boosting (GB), LightGBM, CatBoost, and a stacking Ensemble model. A total of 810 experimental results from the literature cover 15 inputs: cement, silica fume, Slag, Fly Ash, Quartz powder, Limestone powder, Nano silica, Water, Fine and Coarse aggregate, Fiber, Superplasticizer, Temperature, Relative humidity, and Age. The output parameter is the CS of UHPC. A summary of the observations and findings from this study can be found below:

Based on R^2, RMSE, and MAE, the DT model had the lowest performance and the least stability among the examined models. The wide range of means and variances produced by 20 Monte Carlo simulations demonstrates this. In contrast, the Ensemble model discloses great performance, as evidenced by the small differences between the min and max values of each evaluation criterion after 20 simulations: R^2_{max} = 0.977, R^2_{min} = 0.967, $RMSE_{max}$ = 7.479, $RMSE_{min}$ =6.236, and MAE_{max} = 5.473, MAE_{min} = 4.455, respectively. Furthermore, the models can be ranked by predictive performance as follows: Ensemble > CatBoost > LightGBM > GB > DT.

Based on PDP 1D, the impact of each parameter on the CS of UHPC was discussed. Accordingly, Cement, Silica fume, Fiber, Superplasticizer, Temperature, and Age positively affect the compression strength of UHPC. Antithesis, Quartz, Sand, Relative humidity negatively affects the compression strength of UHPC. The effect of the remaining input parameters on the CS of UHPC was inconsistent.

Seven parameters that strongly affect the CS of UHPC were also identified, including Age, Cement, Silica fume, Water, Sand, Fiber, and Superplasticizer, which were employed to establish and discuss in PDP 2D and SHAP.

The proposed ensemble model is a strong decision-support tool for engineers to quickly screen UHPC mixtures in a virtual setting. The model greatly reduces the need for expensive and time-consuming "trial-and-error" laboratory batches by providing accurate strength estimates. However, its utilization should be confined to the scope delineated by the training data. Specifically, because the data are concentrated around standard curing temperatures and elevated relative humidity, predictions for dry curing, intermediate relative humidity ranges with insufficient data, steam curing, extremely high-temperature curing, or significantly different raw materials should be approached with caution and empirically validated. Future research must prioritize integrating specimen geometry and size effects as explicit input variables, augment the dataset with a more balanced range of environmental conditions and SCM dosage levels, and investigate advanced learning architectures to elucidate the intricate hydration kinetics of next-generation UHPC more effectively.

Acknowledgments

Not Applicable.

Credit author statement

May Huu Nguyen: Investigation, Software, Data Curation, Formal analysis, Writing - Original Draft, Validation, Visualization; **Hai-Van Thi Mai**: Methodology, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft; **Son Hoang Trinh**: Investigation, Methodology, Writing - Original Draft, Visualization;

Conflicts of interest or competing interests

The authors declare no conflict of interest.

Data and code availability

Data will be made available on request.

Supplementary information

Not Applicable.

Ethical approval

Not Applicable.

Declaration of Generative AI

During the preparation of this work, the authors did not use Generative AI or AI-assisted technologies to generate the manuscript content.

References

[1] R. Yu, P. Spiesz, H.J.H. Brouwers. (2014). Mix design and properties assessment of ultra-high performance fibre reinforced concrete (UHPFRC). *Cement and Concrete Research*,

56, 29–39. https://doi.org/10.1016/j.cemconres.2013.11.002

[2] K. Wille, C. Boisvert-Cotulio. (2015). Material efficiency in the design of ultra-high performance concrete. *Construction and Building Materials*, 86, 33–43. https://doi.org/10.1016/j.conbuildmat.2015.03.087

[3] R. Yu, P. Spiesz, H.J.H. Brouwers. (2015). Development of Ultra-High Performance Fibre Reinforced Concrete (UHPFRC): Towards an efficient utilization of binders and fibres. *Construction and Building Materials,* 79, 273–282. https://doi.org/10.1016/j.conbuildmat.2015.01.050

[4] S.H. Kosmatka, M.L. Wilson. (2011). Design and Control of Concrete Mixtures. *The Portland Cement Association, USA.*

[5] E. Güneyisi, M. Gesoğlu, S. Karaoğlu, K. Mermerdaş. (2012). Strength, permeability and shrinkage cracking of silica fume and metakaolin concretes. *Construction and Building Materials*, 34, 120–130. https://doi.org/10.1016/j.conbuildmat.2012.02.017

[6] A. Skaropoulou, K. Sotiriadis, G. Kakali, S. Tsivilis. (2013). Use of mineral admixtures to improve the resistance of limestone cement concrete against thaumasite form of sulfate attack. *Cement and Concrete Composites*, 37, 267–275. https://doi.org/10.1016/j.cemconcomp.2013.01.007

[7] D. Sharma, Y. Agrawal, T. Gupta, R. Sharma. (2021). A Review of the Hardened Properties of Eco-Friendly Concrete Containing Ground Granulated Blast-furnace Slag. *Journal of Scientific Research and Reports*, 27(3), 37–49. https://doi.org/10.9734/jsrr/2021/v27i330368

[8] S. Ogawa, T. Nozaki, K. Yamada, H. Hirao, R.D. Hooton. (2012). Improvement on sulfate resistance of blended cement with high alumina slag. *Cement and Concrete Research*, 42(2), 244–251. https://doi.org/10.1016/j.cemconres.2011.09.008

[9] A.M. Ramezanianpour, R.D. Hooton. (2013). Thaumasite sulfate attack in Portland and Portland-limestone cement mortars exposed to sulfate solution. *Construction and Building Materials*, 40, 162–173. https://doi.org/10.1016/j.conbuildmat.2012.09.104

[10] H.N. Atahan, D. Dikme. (2011). Use of mineral admixtures for enhanced resistance against sulfate attack. *Construction and Building Materials*, 25(8), 3450–3457. https://doi.org/10.1016/j.conbuildmat.2011.03.036

[11] T. Oertel, F. Hutter, R. Tänzer, U. Helbig, G. Sextl. (2013). Primary particle size and agglomerate size effects of amorphous silica in ultra-high performance concrete. *Cement and Concrete Composites*, 37, 61–67. https://doi.org/10.1016/j.cemconcomp.2012.12.005

[12] E. Ghafari, H. Costa, E. Júlio, A. Portugal, L. Durães. (2014). The effect of nanosilica addition on flowability, strength and transport properties of ultra high performance concrete. *Materials & Design*, 59, 1–9. https://doi.org/10.1016/j.matdes.2014.02.051

[13] M.A.A. Aldahdooh, N.M. Bunnori, M.A.M. Johari. (2013). Evaluation of ultra-high-performance-fiber reinforced concrete binder content using the response surface method. *Materials & Design (1980-2015),* 52, 957–965. https://doi.org/10.1016/j.matdes.2013.06.034

[14] C. Wang, C. Yang, F. Liu, C. Wan, X. Pu, (2012). Preparation of ultra-high performance concrete with common technology and materials. *Cement and Concrete Composites*, 34(4), 538–544. https://doi.org/10.1016/j.cemconcomp.2011.11.005

[15] J. Yang, J. Zeng, X. He, H. Hu, Y. Su, H.

Bai, H. Tan. (2021). Eco-friendly UHPC prepared from high volume wet-grinded ultrafine GGBS slurry. *Construction and Building Materials*, 308, 125057. https://doi.org/10.1016/j.conbuildmat.2021.125057

[16] Y. Sharma, M. Yeluri, S. Allena, J. Owusu-Danquah. (2024). Incorporating limestone powder and ground granulated blast furnace slag in ultra-high performance concrete to enhance sustainability. *International Journal of Concrete Structures and Materials*, 18, 79. https://doi.org/10.1186/s40069-024-00723-7

[17] N.V. Tuan, G. Ye, K. van Breugel, A.L.A. Fraaij, B.D. Dai. (2011). The study of using rice husk ash to produce ultra high performance concrete. *Construction and Building Materials*, 25(4), 2030–2035. https://doi.org/10.1016/j.conbuildmat.2010.11.046

[18] E. Ghafari, S.A. Ghahari, H. Costa, E. Júlio, A. Portugal, L. Durães. (2016). Effect of supplementary cementitious materials on autogenous shrinkage of ultra-high performance concrete. *Construction and Building Materials*, 127, 43–48. https://doi.org/10.1016/j.conbuildmat.2016.09.123

[19] M.C.G. Juenger, R. Siddique. (2015). Recent advances in understanding the role of supplementary cementitious materials in concrete. *Cement and Concrete Research*, 78, 71–80. https://doi.org/10.1016/j.cemconres.2015.03.018

[20] Z. Rong, W. Sun, H. Xiao, G. Jiang. (2015). Effects of nano-SiO_2 particles on the mechanical and microstructural properties of ultra-high performance cementitious composites. *Cement and Concrete Composites*, 56, 25–31. https://doi.org/10.1016/j.cemconcomp.2014.11.001

[21] S. Kumar, B. Rai. (2018). A review on wider application of supplementary cementitious materials on the development of high-performance concrete. *International Journal of Civil Engineering and Technology (IJCIET)*, 9(8), 187–204.

[22] S. Park, S. Wu, Z. Liu, S. Pyo. (2021). The Role of Supplementary Cementitious Materials (SCMs) in Ultra High Performance Concrete (UHPC): A Review. *Materials*, 14(6), 1472. https://doi.org/10.3390/ma14061472

[23] J. Zhang, Y. Zhao. (2017). The mechanical properties and microstructure of ultra-high-performance concrete containing various supplementary cementitious materials. *Journal of Sustainable Cement-Based Materials*, 6(4), 254–266. https://doi.org/10.1080/21650373.2016.1262798

[24] Z. Wu, C. Shi, W. He. (2017). Comparative study on flexural properties of ultra-high performance concrete with supplementary cementitious materials under different curing regimes. *Construction and Building Materials*, 136, 307–313. https://doi.org/10.1016/j.conbuildmat.2017.01.052

[25] V.-T.-A. Van, C. Rößler, D.-D. Bui, H.-M. Ludwig. (2014). Rice husk ash as both pozzolanic admixture and internal curing agent in ultra-high performance concrete. *Cement and Concrete Composites*, 53, 270–278. https://doi.org/10.1016/j.cemconcomp.2014.07.015

[26] M.A.A. Abd Elaty. (2014). Compressive strength prediction of Portland cement concrete with age using a new model. *HBRC Journal*, 10(2), 145–155. https://doi.org/10.1016/j.hbrcj.2013.09.005

[27] J. Namyong, Y. Sangchun, C. Hongbum. (2004). Prediction of compressive strength of in-situ concrete based on mixture proportions. *Journal of Asian Architecture and Building Engineering*, 3(1), 9–16. https://doi.org/10.3130/jaabe.3.9

[28] E. Ghafari, M. Bandarabadi, H. Costa, E. Júlio. (2012). Design of UHPC using artificial neural networks. *Brittle Matrix Composites 10*, 2012, 61–69. https://doi.org/10.1533/9780857099891.61

[29] O.R. Abuodeh, J.A. Abdalla, R.A. Hawileh. (2020). Assessment of compressive strength of ultra-high performance concrete using deep machine learning techniques. *Applied Soft Computing*, 95, 106552. https://doi.org/10.1016/j.asoc.2020.106552

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

[31] A. Karbassi, B. Mohebi, S. Rezaee, P. Lestuzzi. (2014). Damage prediction for regular reinforced concrete buildings using the decision tree algorithm. *Computers & Structures*, 130, 46–56. https://doi.org/10.1016/j.compstruc.2013.10.006

[32] V.Q. Tran, H.Q. Do. (2021). Prediction of California bearing ratio (CBR) of stabilized expansive soils with agricultural and industrial waste using light GB machine. *Journal of Science and Transport Technology*, 1(1), 1–8. https://doi.org/10.58845/jstt.utt.2021.en.1.1.1-8

[33] B. Kim, D.-E. Lee, G. Hu, Y. Natarajan, S. Preethaa, A.P. Rathinakumar. (2022). Ensemble Machine Learning-Based Approach for Predicting of FRP–Concrete Interfacial Bonding. *Mathematics*, 10(2), 231. https://doi.org/10.3390/math10020231

[34] R. Yu, P. Spiesz, H.J.H. Brouwers. (2014). Effect of nano-silica on the hydration and microstructure development of Ultra-High Performance Concrete (UHPC) with a low binder amount. *Construction and Building Materials*, 65, 140–150. https://doi.org/10.1016/j.conbuildmat.2014.04.063

[35] N. Randl, T. Steiner, S. Ofner, E. Baumgartner, T. Mészöly. (2014). Development of UHPC mixtures from an ecological point of view. *Construction and Building Materials*, 67, 373–378. https://doi.org/10.1016/j.conbuildmat.2013.12.102

[36] X. Zhang, S. Zhao, Z. Liu, F. Wang. (2019). Utilization of steel slag in ultra-high performance concrete with enhanced eco-friendliness. *Construction and Building Materials*, 214, 28–36. https://doi.org/10.1016/j.conbuildmat.2019.04.106

[37] A. Alsalman, C.N. Dang, W.M. Hale. (2017). Development of ultra-high performance concrete with locally available materials. *Construction and Building Materials*, 133, 135–145. https://doi.org/10.1016/j.conbuildmat.2016.12.040

[38] Z. Wu, C. Shi, K.H. Khayat, L. Xie. (2018). Effect of SCM and nano-particles on static and dynamic mechanical properties of UHPC. *Construction and Building Materials*, 182, 118–125. https://doi.org/10.1016/j.conbuildmat.2018.06.126

[39] R. Yang, R. Yu, Z. Shui, X. Gao, X. Xiao, X. Zhang, Y. Wang, Y. He. (2019). Low carbon design of an Ultra-High Performance Concrete (UHPC) incorporating phosphorous slag. *Journal of Cleaner Production*, 240, 118157. https://doi.org/10.1016/j.jclepro.2019.118157

[40] X. Liang, C. Wu, Y. Su, Z. Chen, Z. Li. (2018). Development of ultra-high performance concrete with high fire resistance. *Construction and Building Materials*, 179, 400–412. https://doi.org/10.1016/j.conbuildmat.2018.05.241

[41] D.-Y. Yoo, H.-O. Shin, J.-M. Yang, Y.-S. Yoon. (2014). Material and bond properties of ultra high performance fiber reinforced concrete with

micro steel fibers. *Composites Part B: Engineering*, 58, 122–133. https://doi.org/10.1016/j.compositesb.2013.10.081

[42] Z. Wu, C. Shi, W. He, D. Wang. (2017). Static and dynamic compressive properties of ultra-high performance concrete (UHPC) with hybrid steel fiber reinforcements. *Cement and Concrete Composites*, 79, 148–157. https://doi.org/10.1016/j.cemconcomp.2017.02.010

[43] Q. Song, R. Yu, Z. Shui, X. Wang, S. Rao, Z. Lin. (2018). Optimization of fibre orientation and distribution for a sustainable Ultra-High Performance Fibre Reinforced Concrete (UHPFRC): Experiments and mechanism analysis. *Construction and Building Materials*, 169, 8–19. https://doi.org/10.1016/j.conbuildmat.2018.02.130

[44] S.-H. Kang, Y. Jeong, K.H. Tan, J. Moon. (2018). The use of limestone to replace physical filler of quartz powder in UHPFRC. *Cement and Concrete Composites*, 94, 238–247. https://doi.org/10.1016/j.cemconcomp.2018.09.013

[45] A. Rajasekar, K. Arunachalam, M. Kottaisamy. (2019). Assessment of strength and durability characteristics of copper slag incorporated ultra high strength concrete. *Journal of Cleaner Production*, 208, 402–414. https://doi.org/10.1016/j.jclepro.2018.10.118

[46] D.-Y. Yoo, M.-J. Kim. (2019). High energy absorbent ultra-high-performance concrete with hybrid steel and polyethylene fibers. *Construction and Building Materials*, 209, 354–363. https://doi.org/10.1016/j.conbuildmat.2019.03.096

[47] Y. Li, K.H. Tan, E.-H. Yang. (2019). Synergistic effects of hybrid polypropylene and steel fibers on explosive spalling prevention of ultra-high performance concrete at elevated temperature. *Cement and Concrete Composites*, 96, 174–181. https://doi.org/10.1016/j.cemconcomp.2018.11.009

[48] S.-H. Kang, S.-G. Hong, J. Moon. (2019). The use of rice husk ash as reactive filler in ultra-high performance concrete. *Cement and Concrete Research*, 115, 389–400. https://doi.org/10.1016/j.cemconres.2018.09.004

[49] A. Marani, A. Jamali, M.L. Nehdi. (2020). Predicting Ultra-High-Performance Concrete Compressive Strength Using Tabular Generative Adversarial Networks. *Materials*, 13(21), 4757. https://doi.org/10.3390/ma13214757

[50] C. Zhang, X. Chen, S. Wang, J. Hu, C. Wang, X. Liu. (2021). Using CatBoost algorithm to identify middle-aged and elderly depression, national health and nutrition examination survey 2011–2018. *Psychiatry Research*, 306, 114261. https://doi.org/10.1016/j.psychres.2021.114261

[51] J.R. Quinlan. (1986). Induction of decision trees. *Machine Learning*, 1, 81–106. https://doi.org/10.1007/BF00116251

[52] M. Akin, C. Hand, E. Eyduran, B.M. Reed. (2018). Predicting minor nutrient requirements of hazelnut shoot cultures using regression trees. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 132, 545–559. https://doi.org/10.1007/s11240-017-1353-x

[53] T. Hastie, R. Tibshirani, J. Friedman. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. *Springer*. https://doi.org/10.1007/978-0-387-84858-7

[54] Z.-H. Zhou. (2009). Ensemble Learning. *Encyclopedia of Biometrics. Springer, pp. 270–273*. https://doi.org/10.1007/978-0-387-73003-5_293

[55] G. Valentini, T.G. Dietterich. (2004). Bias-Variance Analysis of Support Vector Machines

for the Development of SVM-Based Ensemble Methods. *Journal of Machine Learning Research*, 5, 725–775.
[56] L. Guelman. (2012). Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Systems with Applications*, 39(3), 3659–3667. https://doi.org/10.1016/j.eswa.2011.09.058
[57] V. Vakharia, R. Gujar. (2019). Prediction of compressive strength and portland cement composition using cross-validation and feature ranking techniques. *Construction and Building Materials*, 225, 292–301. https://doi.org/10.1016/j.conbuildmat.2019.07.224
[58] J.H. Friedman. (2001). Greedy function approximation: A gradients boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451
[59] M.S. Khorsheed, A.O. Al-Thubaity. (2013). Comparative evaluation of text classification techniques using a large diverse Arabic dataset. *Language Resources & Evaluation*, 47, 513–538. https://doi.org/10.1007/s10579-013-9221-8
[60] H.-B. Ly, M.H. Nguyen, B.T. Pham, (2021). Metaheuristic optimization of Levenberg–Marquardt-based artificial neural network using particle swarm optimization for prediction of foamed concrete compressive strength. *Neural Computing and Applications*, 33, 17331-17351. https://doi.org/10.1007/s00521-021-06321-y
[61] O.M. Abdulkareem, A.B. Fraj, M. Bouasker, L. Khouchaf, A. Khelidj. (2021). Microstructural investigation of slag-blended UHPC: The effects of slag content and chemical/thermal activation. *Construction and Building Materials*, 292, 123455. https://doi.org/10.1016/j.conbuildmat.2021.123455
[62] Z. Liu, S. El-Tawil, W. Hansen, F. Wang. (2018). Effect of slag cement on the properties of ultra-high performance concrete. *Construction and Building Materials*, 190, 830–837. https://doi.org/10.1016/j.conbuildmat.2018.09.173
[63] E. Fehling, M. Schmidt, S. Stürwald (edited). (2008). Ultra High Performance Concrete: (UHPC). *Proceedings of the Second International Symposium on Ultra High Performance Concrete, Germany*.
[64] D. Wang, C. Shi, Z. Wu, J. Xiao, Z. Huang, Z. Fang. (2015). A review on ultra high performance concrete: Part II. Hydration, microstructure and properties. *Construction and Building Materials*, 96, 368–377. https://doi.org/10.1016/j.conbuildmat.2015.08.095
[65] W. Huang, H. Kazemi-Kamyab, W. Sun, K. Scrivener. (2017). Effect of cement substitution by limestone on the hydration and microstructural development of ultra-high performance concrete (UHPC). *Cement and Concrete Composites*, 77, 86–101. https://doi.org/10.1016/j.cemconcomp.2016.12.009
[66] M. Shafieifar, M. Farzad, A. Azizinamini. (2017). Experimental and numerical study on mechanical properties of Ultra High Performance Concrete (UHPC). *Construction and Building Materials*, 156, 402–411. https://doi.org/10.1016/j.conbuildmat.2017.08.170

Appendix A

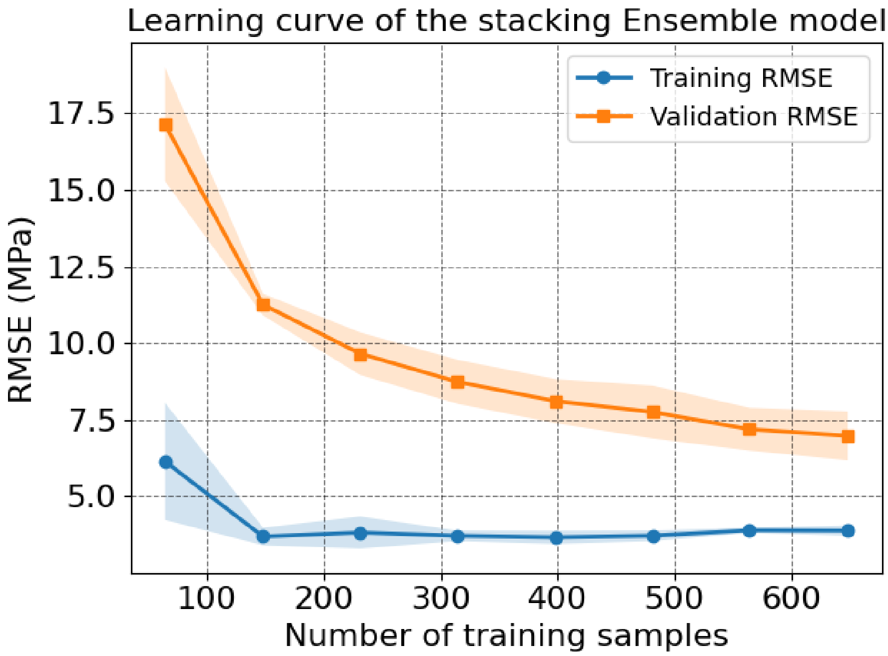

Fig. A1. Learning curve of the stacking Ensemble model showing training and validation RMSE as functions of the number of training samples. The shaded regions represent one standard deviation across cross-validation folds.