

## Article

# Explainable Ensemble Learning Models for the Rheological Properties of Self-Compacting Concrete

Celal Cakiroglu <sup>1</sup>, Gebrail Bekdas <sup>2,\*</sup>, Sanghun Kim <sup>3</sup> and Zong Woo Geem <sup>4,\*</sup><sup>1</sup> Department of Civil Engineering, Turkish-German University, 34820 Istanbul, Turkey<sup>2</sup> Department of Civil Engineering, Istanbul University-Cerrahpasa, 34320 Istanbul, Turkey<sup>3</sup> Department of Civil and Environmental Engineering, Temple University, Philadelphia, PA 19122, USA<sup>4</sup> Department of Smart City & Energy, Gachon University, Seongnam 13120, Korea

\* Correspondence: bekdas@iuc.edu.tr (G.B.); geem@gachon.ac.kr (Z.W.G.)

**Abstract:** Self-compacting concrete (SCC) has been developed as a type of concrete capable of filling narrow gaps in highly reinforced areas of a mold without internal or external vibration. Bleeding and segregation in SCC can be prevented by the addition of superplasticizers. Due to these favorable properties, SCC has been adopted worldwide. The workability of SCC is closely related to its yield stress and plastic viscosity levels. Therefore, the accurate prediction of yield stress and plastic viscosity of SCC has certain advantages. Predictions of the shear stress and plastic viscosity of SCC is presented in the current study using four different ensemble machine learning techniques: Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost), random forest, and Categorical Gradient Boosting (CatBoost). A new database containing the results of slump flow, V-funnel, and L-Box tests with the corresponding shear stress and plastic viscosity values was curated from the literature to develop these ensemble learning models. The performances of these algorithms were compared using state-of-the-art statistical measures of accuracy. Afterward, the output of these ensemble learning algorithms was interpreted with the help of SHapley Additive exPlanations (SHAP) analysis and individual conditional expectation (ICE) plots. Each input variable's effect on the predictions of the model and their interdependencies have been illustrated. Highly accurate predictions could be achieved with a coefficient of determination greater than 0.96 for both shear stress and plastic viscosity.

**Keywords:** plastic viscosity; self-compacting concrete; yield stress; V-funnel flow; slump flow; L-Box; XGBoost; LightGBM; CatBoost; SHAP



**Citation:** Cakiroglu, C.; Bekdas, G.; Kim, S.; Geem, Z.W. Explainable Ensemble Learning Models for the Rheological Properties of Self-Compacting Concrete. *Sustainability* **2022**, *14*, 14640. <https://doi.org/10.3390/su142114640>

Academic Editors: Woubishet Zewdu Taffese and Sandra Barbosa Nunes

Received: 6 October 2022

Accepted: 3 November 2022

Published: 7 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Self-compacting concrete (SCC) is broadly used in the construction industry due to its good mechanical properties, high fluidity, and ability to pass through and fill the gaps between reinforcing bars without vibrations [1,2]. Self-compactibility and resistance to segregation can be achieved by using superplasticizers, lowering the water-cement ratio, and reducing coarse aggregate content [3]. The flowability of SCC is correlated with the plastic viscosity, yield stress, and the outcome of empirical test procedures. The workability of concrete can be defined as its ability to properly fill its molding while having sufficient strength in its final hardened form [4]. To have good workability, a balance must be maintained between the mechanical properties of concrete and its fluidity. Therefore, it is important to have accurate procedures for the prediction of yield stress and plastic viscosity as these properties determine the workability of concrete. The rheological properties of fresh concrete can be investigated using various test procedures, including slump-flow, L-box, and V-funnel tests. Past studies have aimed at finding correlations between the properties, such as the correlation of slump flow diameter and V-funnel flow time with the yield stress and plastic viscosity of fresh concrete. The yield stress of fresh concrete

is defined as the minimum amount of stress that causes permanent deformation and flows [5]. In addition to yield stress, another property of fresh concrete that determines its workability is the plastic viscosity, which is defined as the resistance of concrete to flow when the shear stress is higher than yield stress. Due to the practical significance of these rheological properties, there have been studies aimed at developing models that predict these properties. Neophytou et al. [6] proposed an empirical model that related the yield stress of SCC to the non-dimensional final spread of concrete obtained from a slump flow test. A linear relationship between the non-dimensional final spread and yield stress was demonstrated. Schowalter and Christensen [7] developed an analytical equation that related the slump of fresh concrete to its yield stress. Pashias et al. [4] investigated the relationship between slump height and yield stress in flocculated materials. An approximating equation that related yield stress to slump height was proposed. Le et al. [8] showed that the yield stress of self-consolidating concrete could be predicted by considering concrete as an aggregate suspension in cement paste. Based on excess paste theory and percolation theory, it was shown that the yield stress could be predicted as a function of the excess paste layer thickness, volume fraction of the aggregates, and percolation.

In recent years, machine learning techniques have been applied in various areas of engineering such as bridge construction [9–13], reinforced concrete and steel frames [14–16], modeling of concrete and masonry structures [17–20], modeling of pavement foundations [21], and design of columns [22–24]. Machine learning models have also been applied to plastic viscosity and yield stress prediction for concrete. Different predictive models have been proposed that relates the yield stress and plastic viscosity of concrete to its rheological properties. Benaicha et al. [25] developed an artificial neural network and multi-variable regression models that predicted the yield stress and plastic viscosity of concrete. The predictive models were developed using a dataset of 59 samples. The slump flow diameter, V-funnel flow time, and L-Box ratio were the input variables in these models. The increase in the slump flow diameter and V-funnel flow time was found to positively impact yield stress and plastic viscosity. Alyamac and Ince [26] carried out an experimental study including slump-flow, L-box, and V-funnel tests on SCC with the addition of three different types of marble powder. Compression strength and split tension tests were conducted on hardened concrete specimens. Based on these experiments, a concrete mix design monogram was created that described the relationships between compressive strength, water-cement ratio, aggregate-cement ratio, and cement content.

Data-driven prediction methodologies have also been applied to estimating the chloride permeability and mechanical properties of SCC. Carbon dioxide and chloride penetration is a major factor leading to corrosion in reinforced concrete structures [27]. Yuan et al. [28] investigated the chloride penetration in SCC using single and hybrid regression methods. Cement content, fly ash, silica fume, fine and coarse aggregate percentages, and temperature were the input variables in this study. The predictive model accuracies were measured using the root mean square error, mean absolute error, mean absolute percentage error, and performance index. Kumar et al. [29] demonstrated the applicability of the multi-variate adaptive regression spline and minimax probability machine regression models to predict the results of rapid chloride penetration tests. The effects of fly ash and silica fume contents and temperature on the test results were investigated. Ge et al. [30] presented optimized random forest models developed using particle swarm optimization, whale optimization algorithm, and Harris hawks optimization. Input features included cement, fly ash, silica fume, fine and soft aggregate contents, water-to-cement ratio, and temperature. Amin et al. [31] used a gene expression programming algorithm to investigate the effects of fine and coarse aggregate contents, water-to-binder ratio, compressive strength, and metakaolin content on rapid chloride penetration. Aggarwal et al. [32] developed predictive models using random forest, random tree, multilayer perceptron, M5P, and support vector regression algorithms, based on the contents of cement, fine and coarse aggregates, metakaolin, rice husk ash, water, and superplasticizers as input features to predict the 28-day compressive strength of SCC. The models were trained using a dataset

of 159 samples. Farooq et al. [33] applied artificial neural network, support vector machine, and gene expression programming techniques to predict the mechanical properties of SCC. The coefficient of determination was used as the metric for accuracy. Zhu et al. [34] developed hybrid predictive models by combining random forest, support vector regression, and multi-layer perceptron techniques with the grey wolf optimization algorithm. The splitting tensile strength of concrete was predicted based on a dataset of 168 samples. De-Prado-Gil et al. [35] compared the performances of extra trees regression, gradient boosting, CatBoost, and XGBoost algorithms in predicting the splitting tensile strength of SCC. The performances of these algorithms were compared on a database consisting of 381 samples using coefficient of determination, root mean square error, and mean absolute error as the metrics for accuracy. The XGBoost algorithm was reported to perform best according to all error metrics.

The current study presents the application of four different ensemble machine learning models to predict the yield stress and plastic viscosity of SCC. The ensemble ML models have been trained on a newly curated database that combines different experimental results from past literature. Furthermore, the SHapley Additive exPlanations (SHAP) methodology has been utilized to make the ensemble learning models explainable. The impact of each rheological property on the predicted yield stress and viscosity was visualized according to the SHAP algorithm. Finally, the effect of changing individual input variables on the output of the ML models were shown using individual conditional expectation (ICE) plots.

## 2. Materials and Methods

The procedures for developing explainable predictive ensemble learning models for the rheological properties of SCC are presented in this section. After a summary of the test techniques used for generating the machine learning dataset, the applied ensemble learning methods are described.

### 2.1. Test Procedures

This section briefly describes the experiments for the measurement of the rheological properties of SCC.

#### 2.1.1. Slump Flow Test

The slump flow test quantifies the filling ability of SCC. The test procedure followed the EFNARC guidelines for SCC [36]. A mold with the shape of a truncated cone was placed on a flow table, as shown in Figure 1, and filled with SCC. The mold had a top and bottom diameter of 100 mm and 200 mm, respectively. The height of the cone was 300 mm. A circle of 500 mm diameter with the cone was drawn on the flow table. After the mold was lifted, SCC started to spread. The time needed for concrete to reach 500 mm diameter was recorded as  $T_{500}$ . Once SCC reached its final shape, the diameter of the spread was measured in two perpendicular directions (D1 and D2, shown in Figure 1), and the average value of these two measurements was recorded as the slump flow diameter.

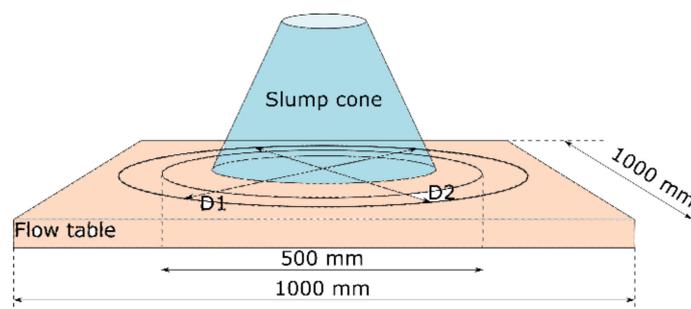


Figure 1. Slump flow test setup.

### 2.1.2. V-Funnel Test

The V-funnel test quantifies the ability of SCC to pass through narrow openings. The test was carried out using a V-shaped funnel (Figure 2), through which SCC passed under its weight. The standard dimensions of the equipment, shown in Figure 2, were adopted from the JSCE guidelines [37].

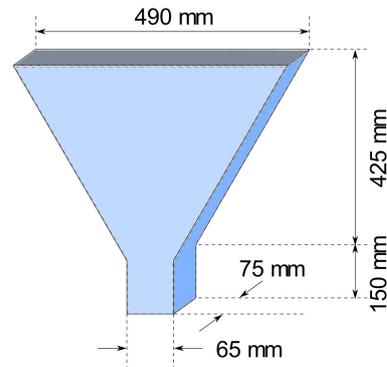


Figure 2. V-funnel test equipment [1].

The V-funnel test was carried out with 11.2 L of concrete that filled the V-shaped funnel. The time it took for this volume of SCC to completely flow out of the V-funnel was recorded. V-funnel flow time is an indication of how quickly SCC can fill narrow voids. Furthermore, greater V-funnel flow times indicate greater plastic viscosity [38].

### 2.1.3. L-Box Test

The equipment for the L-Box test consists of a vertical and horizontal section. Initially, the vertical section is filled with SCC, and then the SCC is allowed to flow into the horizontal section, passing through steel bars. A description of this test setup is shown in Figure 3.

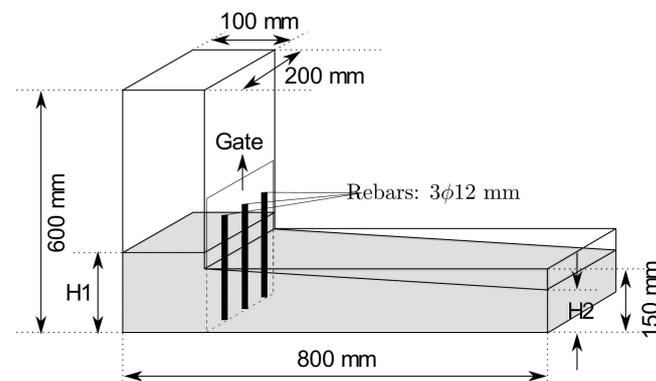


Figure 3. L-Box test setup [11].

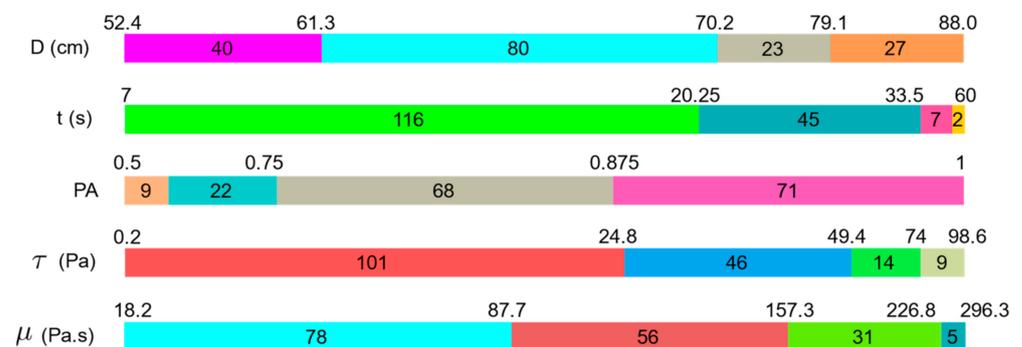
After stabilization of the concrete, the heights  $H_1$  and  $H_2$  are measured. The ratio of these two values ( $H_1/H_2$ ) is used as an indicator of the ability of SCC to pass through narrow spaces. The ratio  $PA = H_1/H_2$  is used as the passing ability parameter [39].

## 2.2. Ensemble Machine Learning Process

A dataset consisting of 170 samples was collected from the literature [3,25,39–41]. In this dataset, each sample contained the measurements for SCC from the V-funnel, slump flow, and L-Box tests. Furthermore, the shear stress and plastic viscosity values corresponding to each data sample was recorded using a rheometer. In this dataset, the slump flow diameter, V-funnel flow time, and passing ability parameter were the input features for the machine learning models, whereas shear stress and plastic viscosity were the predicted output variables. In addition to these three input features, some models have

been developed where the input data has been augmented with the plastic viscosity values when predicting shear stress, and vice versa. As the algorithms of predictive modeling, Extreme Gradient Boosting (XGBoost), Categorical Gradient Boosting (CatBoost), random forest, and Light Gradient Boosting Machine (LightGBM) were selected. The models were implemented using the Scikit-learn package available in the Python programming language. This selection was based on various studies in the recent years that reported these algorithms as some of the best-performing algorithms among state-of-the-art machine learning techniques. Rahman et al. [42] presented the results of eleven different machine learning algorithms, including linear regression, ridge regression, lasso regression, decision tree, random forest, support vector regression, k-nearest neighbors (KNN), artificial neural network (ANN), XGBoost, AdaBoost, and CatBoost, in the prediction of the shear strength of steel fiber-reinforced concrete beams. The XGBoost algorithm was shown to deliver the best accuracy among these methods, followed by random forest, AdaBoost, and CatBoost. Cakiroglu et al. [24] developed machine learning models for the prediction of the axial load carrying capacity of concrete filled steel tubular stub columns. The performances of Lasso regression, random forest, AdaBoost, LightGBM, Gradient Boosting Machine, XGBoost, and CatBoost models were compared. The CatBoost, LightGBM, and XGBoost models were demonstrated to perform better than the remaining models. Somala et al. [43] demonstrated the application of linear regression, KNN, support vector machine, random forest, and XGBoost models in predicting the peak ground acceleration and peak ground velocity during an earthquake. The best predictions were obtained using the random forest and XGBoost models for the prediction of peak ground acceleration and peak ground velocity, respectively. Degtyarev and Naser [44] developed predictive models for the estimation of the shear strength and elastic shear buckling load of cold-formed steel channels, using the gradient boosting regressor, XGBoost, LightGBM, AdaBoost, and CatBoost methods. The CatBoost method was observed to deliver the highest accuracy among these models. Sun et al. [9] used the ANN, KNN, decision tree, random forest, AdaBoost, gradient boosting regression tree, and XGBoost algorithms for the prediction of tuned mass damper accelerations. Among these algorithms, the random forest model was shown to achieve the best accuracy.

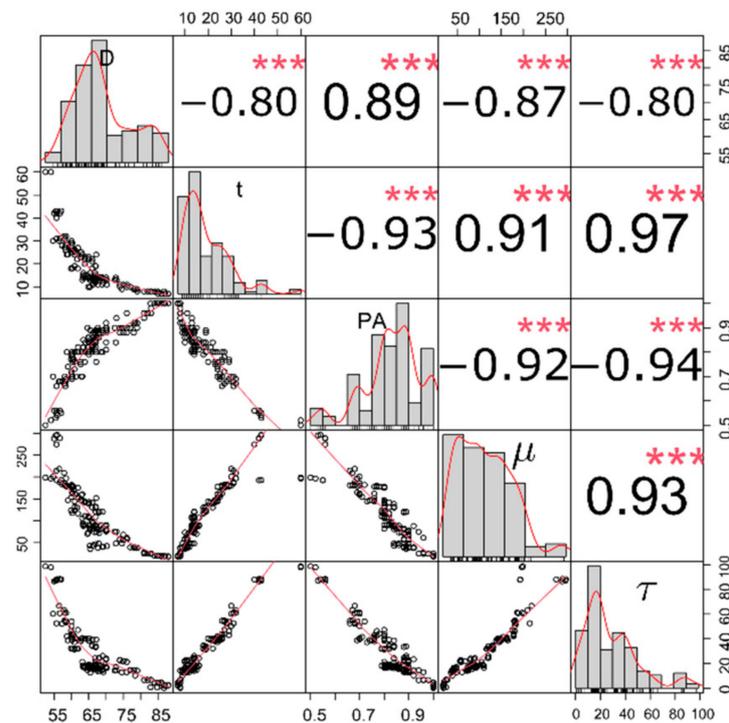
The maximum and minimum values, as well as the distribution of the input and output features used in this study, have been visualized in Figure 4. In Figure 4, each feature has been split into four compartments, with different colors based on the magnitude of the features. The length of each compartment depends on the number of samples in that compartment. The minimum and maximum values of each compartment are written above the boundaries of the compartments.



**Figure 4.** Variable ranges and distributions.

In Figure 4, the slump flow diameter, V-funnel flow time, L-Box H2/H1 ratio, shear stress, and plastic viscosity are shown with  $D$ ,  $t$ , PA,  $\tau$ , and  $\mu$ , respectively. The majority of cases (68%) have a V-funnel flow time between 7 and 20.25 s, whereas only 5.3% of the samples have a V-funnel flow time between 33.5 and 60 s. In addition, 70.6% of the samples have a slump flow diameter of less than 70.2 cm and 54.1% of the samples have a plastic

viscosity higher than 87.7 Pa.s. A visualization of the correlation between the different variables in this study is shown in Figure 5.



**Figure 5.** The correlation matrix with variable distributions (three stars indicate the significance of correlation).

The correlation of all the variables in the dataset and their statistical distributions are shown in Figure 5. Each tile on the diagonal shows the frequency distribution of a variable, whereas the lower triangular area contains bivariate scatter plots with regression lines. The upper triangular part of the correlation matrix contains the Pearson correlation values between the variables, where the font size and number of stars indicate the strength of the correlation. The Pearson correlation value  $r_{xy}$  was computed as shown in Equation (1), and an  $r_{xy}$  value close to 1 indicates a strong correlation between the variables  $x$  and  $y$ .

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \quad (1)$$

In Equation (1),  $x$  and  $y$  are two sets of values containing  $n$  samples each. For each variable in Figure 5, a scale showing the magnitude of this variable is available on both the horizontal and vertical axes. According to Figure 5, the passing ability (PA) and V-funnel flow time ( $t$ ) were correlated with a Pearson correlation coefficient of  $-0.93$ . The strongest correlation between any two variables was observed between the V-funnel flow time ( $t$ ) and shear stress ( $\tau$ ). On the other hand, relatively weak correlations were found between the slump flow diameter ( $D$ ) and  $t$ , as well as between  $D$  and  $\tau$ .

The dataset was randomly split into a training and test set for the training and testing of the predictive models. Based on a consensus in the machine learning-related literature, a 70% to 30% ratio between the training and test sets was adopted. Some of the notable examples were studies carried out by Feng et al. [45] and Nguyen et al. [46]. Nguyen et al. [46] carried out a comprehensive parametric study to demonstrate the effect of changing training/test set split ratio on the predictive model output. A total of nine split ratios ranging between 10/90 and 90/10 were tested, and the 70/30 split ratio was demonstrated to be the most suitable split ratio. Table 1 shows the statistical properties of the training and test sets where SD, As, and K denote standard deviation, skewness, and

kurtosis, respectively. The grid search approach was adopted to tune the hyperparameters of the predictive models using the training sets. Table 2 provides an overview of the hyperparameters used in the predictive models. The hyperparameters of Table 2 were obtained based on 10-fold cross-validation using the training set. In this process, the training set was split into 10 equal-sized segments, and each of these segments was used as a test set once, while the model was trained on the remaining part of the training set.

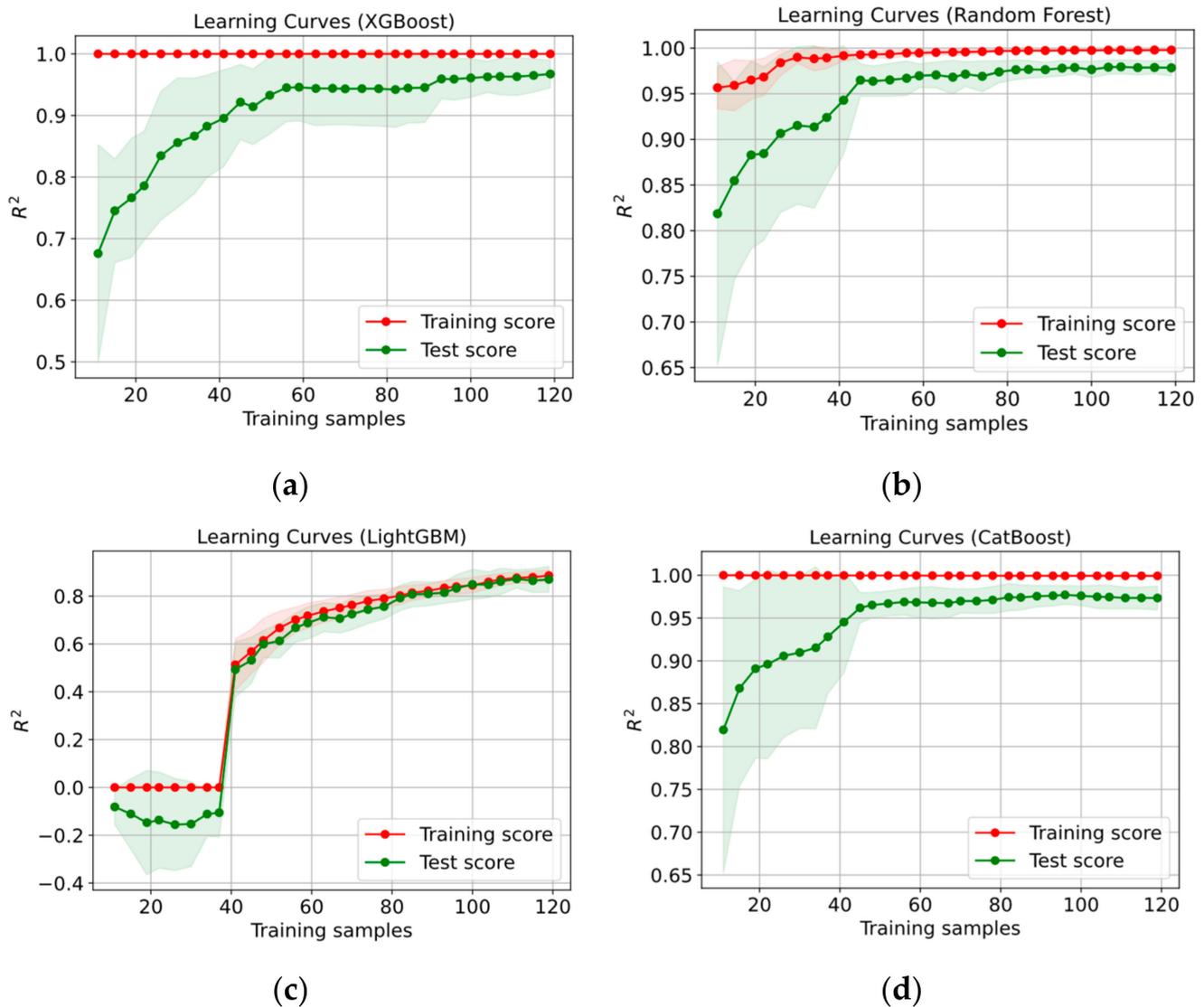
**Table 1.** Statistical properties of the training and test sets.

Dataset	Property	D	t	PA	$\mu$	$\tau$
Training (119 samples)	Unit	cm	s	-	[Pa · s]	[Pa · s]
	Min	52.4	7.0	0.5	18.2	0.2
	Max	88.0	60.0	1.0	296.3	98.6
	Mean	68.19	18.10	0.83	107.40	29.23
	SD	9.08	9.94	0.11	67.34	21.54
	As	0.65	1.30	−0.69	0.57	1.18
	K	−0.54	2.09	0.35	−0.10	1.28
Test (51 samples)	Min	54.0	7.0	0.52	18.3	0.8
	Max	88.0	60.0	1.0	274.65	97.8
	Mean	69.30	16.04	0.85	93.26	25.34
	SD	9.12	9.57	0.11	58.22	20.71
	As	0.59	2.38	−0.83	0.81	1.63
	K	−0.55	7.45	0.99	0.52	2.78

**Table 2.** Hyperparameters for the predictive models.

Model	Parameter	Value
Random Forest	Number of estimators (n_estimators)	5
	Minimum samples for split (min_samples_split)	3
	Minimum samples of leaf node (min_samples_leaf)	1
	Maximum tree depth (max_depth)	None
	Number of features (max_features)	“sqrt”
XGBoost	Number of estimators (n_estimators)	50
	Step size shrinkage (eta)	0
	Learning rate	0.1
	Subsample ratio of the training instances (subsample)	0.5
	Maximum depth of a tree	6
LightGBM	Number of estimators (n_estimators)	500
	Maximum number of decision leaves (num_leaves)	5
	Maximum depth of a tree (max_depth)	4
	Learning rate	0.2
	use extremely randomized trees (extra_trees)	True
CatBoost	Bagging temperature (bagging_temperature)	10
	Learning rate	0.3
	Depth	8
	Tree growing policy (grow_policy)	“Depthwise”

Figure 6 shows the learning curves of the predictive models. In each subplot of Figure 6, the development of the  $R^2$  scores obtained from the training and test sets are shown in red and green colors, respectively. For each model, the training samples were fed into the algorithm in 30 batches and the performance of the model was plotted after the model parameter updates. Figure 6 shows that the prediction performances of the test set improves as the number of training samples increases. The learning curves converge to their best performance, which indicates that the models have a good fit.



**Figure 6.** Learning curves of the (a) XGBoost, (b) random forest, (c) LightGBM, and (d) CatBoost models.

Figures 7 and 8 show the variations of different variables used in the predictions for shear stress ( $\tau$ ) and plastic viscosity ( $\mu$ ). In these plots, the colors of the dots represent the changes in the magnitudes of the output variables. In Figure 7, as the magnitude of  $\tau$  increases, the colors of the dots change from blue to red. Similarly, in Figure 8, the colors of the dots lighten as the magnitude of  $\mu$  increases. Both Figures 7 and 8 show that as the V-funnel flow time ( $t$ ) decreases, the slump flow diameter rapidly increases for  $t < 15$  s. A similar relationship can be observed between  $D$  and passing ability (PA). Particularly, for  $PA > 0.9$ , a rapid increase in  $D$  takes place. Both  $\tau$  and  $\mu$  are observed to decrease with increasing  $D$  values, and increase with increasing  $t$  values. A nearly linear relationship is observed between these variables ( $\tau$  and  $\mu$ ) and the V-funnel flow time  $t$ , particularly for  $t < 30$  s.

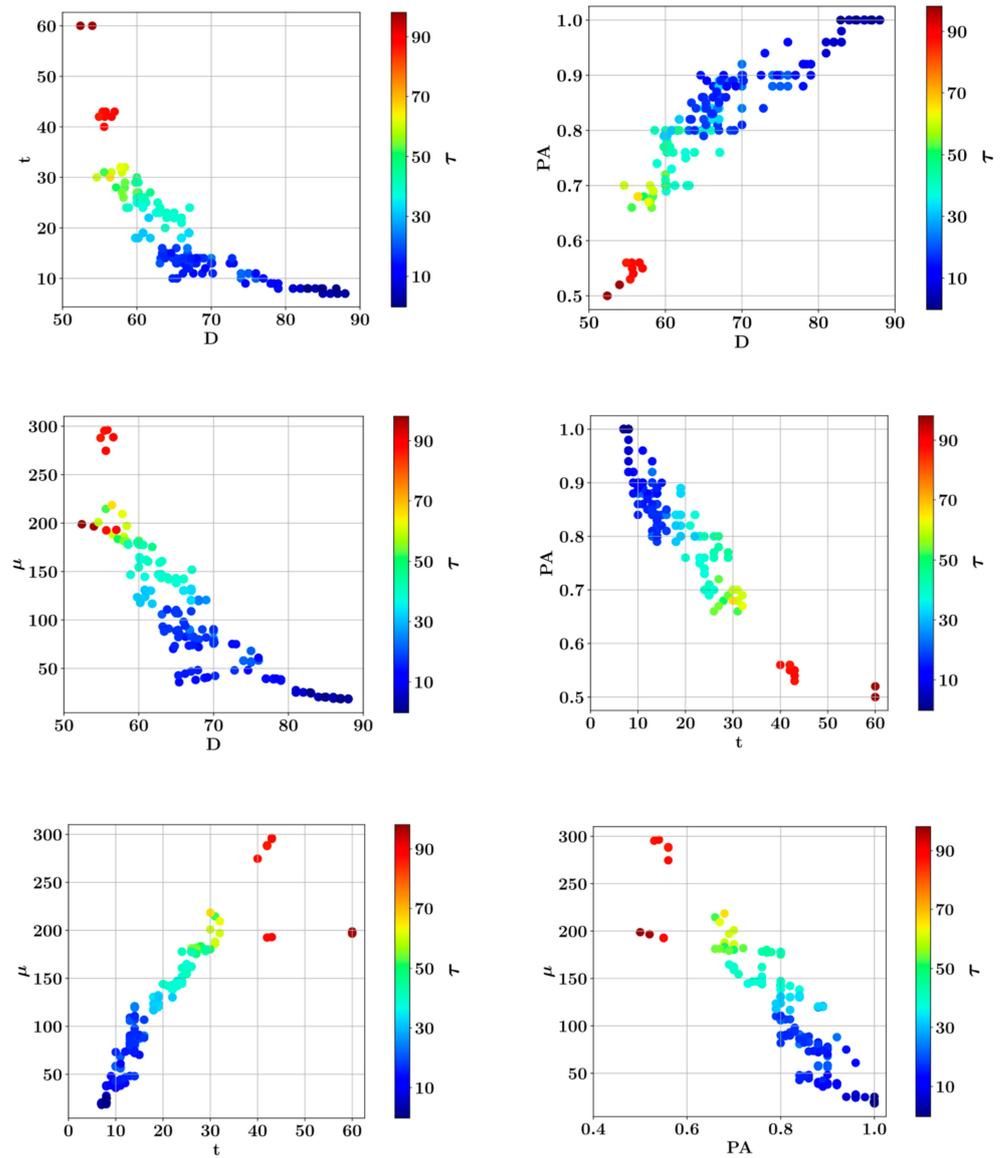


Figure 7. Relationships between the input variables and dependent variable  $\tau$ .

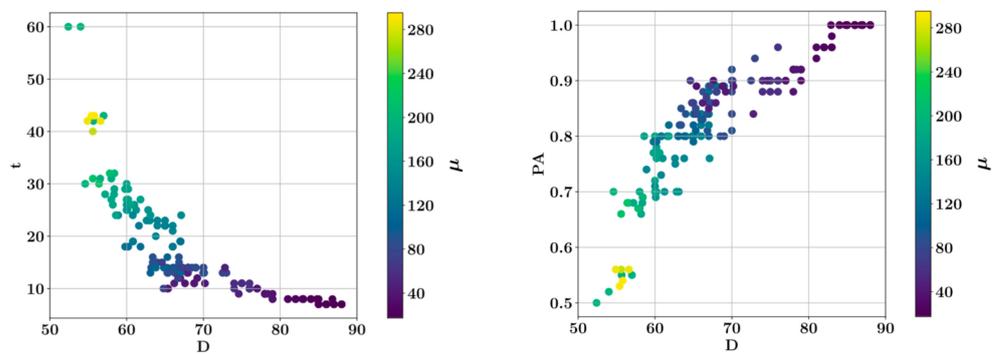
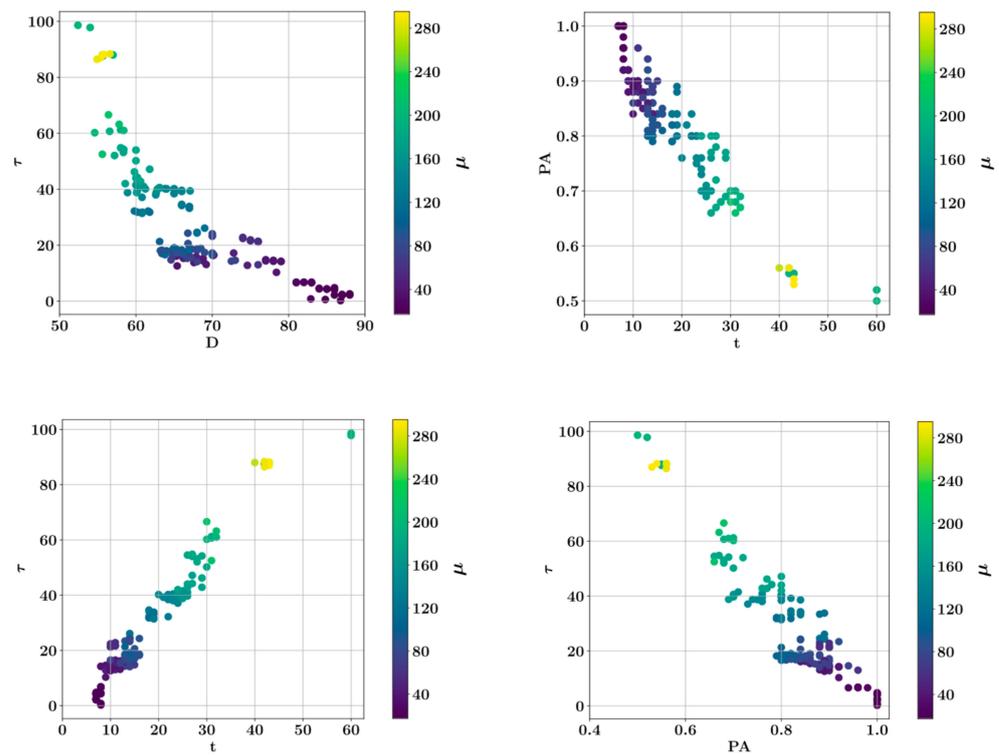


Figure 8. Cont.



**Figure 8.** Relationships between the input variables and dependent variable  $\mu$ .

### 2.3. Gradient Boosting Algorithms

The gradient boosting algorithms in this study were built on the technique of combining weak decision trees to generate strong learners. This procedure is also called ensemble learning. Among the algorithms in this category, the eXtreme Gradient Boosting (XGBoost) algorithm stands out as one of the most successful and frequently used algorithms. Equation (2) describes the process of generating the predicted values,  $\hat{y}_i$ , for the  $i$ -th data point in an XGBoost model. In Equation (2), the predictive model output is described as a linear combination of the outputs of individual regression trees,  $f_k(x_i)$ , where  $K$  is the total number of regression trees. Here,  $x_i$  denotes a vector of input features for a data point with index  $i$ . Equation (2) also shows the regularized objective function,  $\mathcal{L}(\Phi)$ , the minimization of which yields the individual regression trees,  $f_k$ . The leaf weights of these regression trees are combined in the vectors  $w_k$ , the total number of the leaves is denoted with  $T$ , and the loss function, which depends on the difference between the target and predicted values, is shown with  $L(y_i, \hat{y}_i)$ . In Equation (2),  $w_j^*$  stands for the optimal values of the leaf weights that minimize the loss function, and  $I_j$  is the set that contains the sample indices of the  $j$ -th leaf [47,48].

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i)$$

$$\mathcal{L}(\Phi) = \sum_i L(y_i, \hat{y}_i) + \sum_k \Omega(f_k) = \sum_i L(y_i, \hat{y}_i) + \sum_k \gamma T + \frac{1}{2} \lambda \|w_k\|^2 \quad (2)$$

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}, \quad g_i = \frac{\partial L(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}}, \quad h_i = \frac{\partial^2 L(y_i, \hat{y}_i^{(t-1)})}{\partial (\hat{y}_i^{(t-1)})^2}$$

Another algorithm that combines weak decision trees to generate strong learners is the random forest algorithm. Bagging and random feature selection methodologies are utilized to train each decision tree on a randomly selected subset of the training set. The random forest model forecast is determined by the average value of the individual decision tree predictions, as indicated in Equation (3). In Equation (3),  $\hat{m}_j(x)$  and  $\hat{m}(x)$  denote the

predictions of a single decision tree and the entire random forest model, respectively, for the input vector  $\mathbf{x}$ , where  $K$  is the total number of the decision trees in the model [49].

$$\hat{m}(\mathbf{x}) = \frac{1}{K} \sum_{j=1}^K \hat{m}_j(\mathbf{x}) \quad (3)$$

A more developed version of the gradient boosting machine algorithm with better accuracy, more efficient memory usage, and increased speed is called the LightGBM algorithm. Due to gradient-based one-side sampling (GOSS), parallel learning, and exclusive feature bundling (EFB) techniques, the LightGBM algorithm is capable of processing large datasets. The GOSS algorithm is based on the ranking of data instances to the magnitudes of their gradients. Using this methodology, the predictive model can be trained on a smaller training subset with greater gradient magnitudes, which increases model efficiency [50].

The CatBoost algorithm was mainly designed for the efficient processing of datasets with categorical features. Better prediction accuracy is achieved through the ordered boosting algorithm. Furthermore, using ordered target statistics, CatBoost eliminates the prediction shift that is observed in the other ensemble learning methods [51].

### 3. Results

In this section, the predictions of the ensemble learning algorithms are compared with the actual experimental values of shear stress and plastic viscosity. The accuracy and computational speed of the predictive models were quantified and tabulated. For each predictive model, the coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute error were used as the metrics for accuracy. Separate predictive models were developed for the shear stress and plastic viscosity of SCC. The impact of each input variable on model prediction was investigated using SHAP methodology. Individual conditional expectation (ICE) plots were generated for each input feature.

Figure 9 shows the comparison between the predicted and actual experimental shear stress values for the four different ensemble machine learning models. The predictions of the training and test sets are shown separately, with different colors and symbols. The predictive models in Figure 9 were developed using the slump flow diameter, V-funnel flow time, L-Box passing ability, and plastic viscosity as the input features affecting the shear stress  $\tau$ . In Table 3, the best accuracy values are shown in bold font. Table 3 shows that the XGBoost model delivered the best accuracy on the test set in terms of the  $R^2$  score, MAE, VAF, and RMSE metrics, followed by the random forest and CatBoost algorithms. On the other hand, the LightGBM algorithm was observed to have the worst performance in terms of all four metrics. A comparison of the model performances on the test set revealed that the XGBoost model performed best according to the  $R^2$ , MAE, and RMSE metrics, whereas the random forest algorithm performed best according to the VAF metric. The XGBoost, random forest, and LightGBM algorithms fared almost equally in terms of processing speed, whereas the CatBoost technique was significantly slower. Figure 10a shows the comparison of the predicted and target shear stress values for the training and test sets, where a close overlap between these two quantities can be observed. Figure 10b shows the prediction errors in the training and test sets in terms of percentages. The error percentages were reduced to near zero values for the entire training set, whereas up to 40% of over- or underestimation of the shear stress could be observed for certain samples in the test set. Figure 10c,d shows the distributions of the error percentages for the training and test sets, respectively. Figure 10c shows that the error percentages of the training set are clustered around the zero value, whereas a wider distribution can be observed with error percentages that are an order of magnitude greater, as shown in Figure 10d.

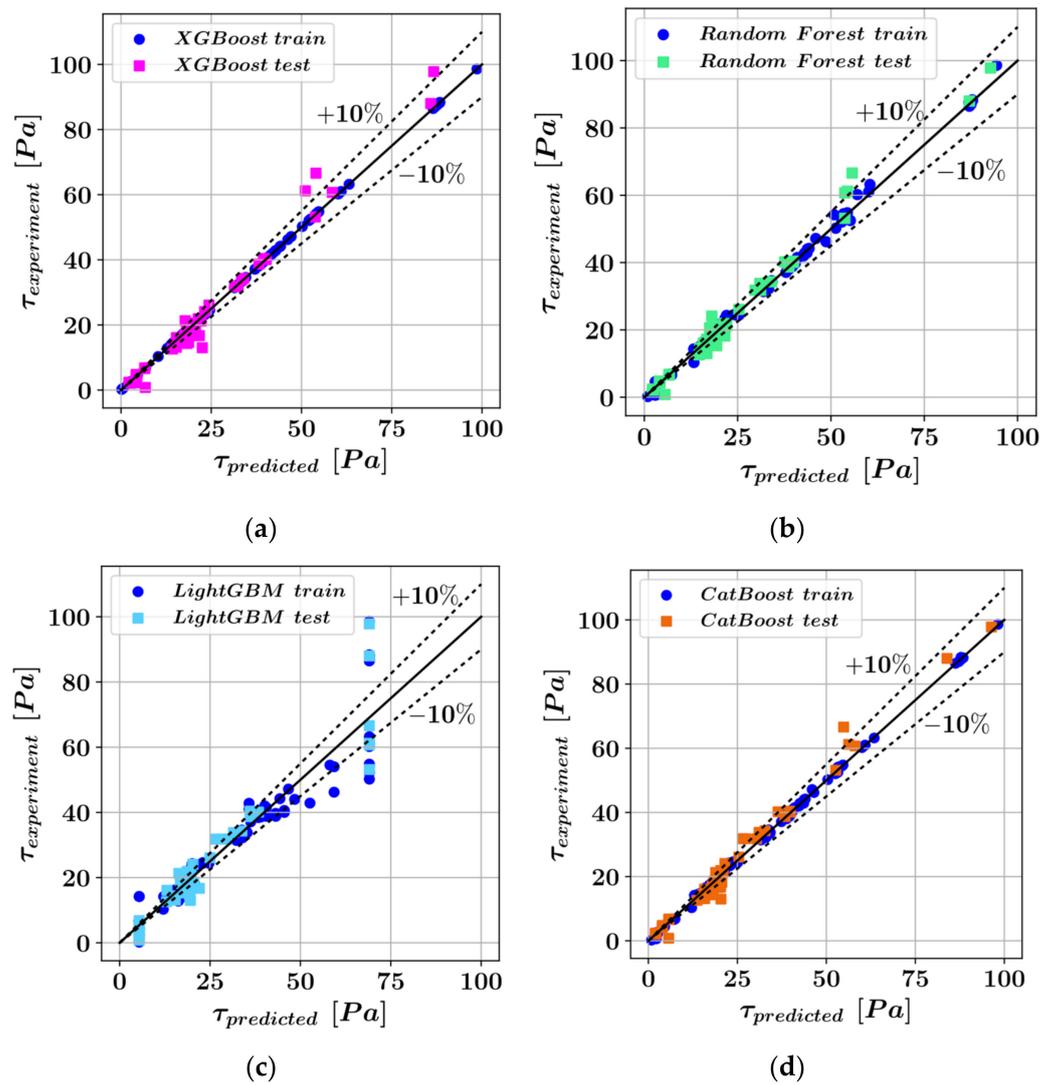


Figure 9. Comparison of the experimental and predicted shear stress values for (a) XGBoost, (b) random forest, (c) LightGBM, and (d) CatBoost models.

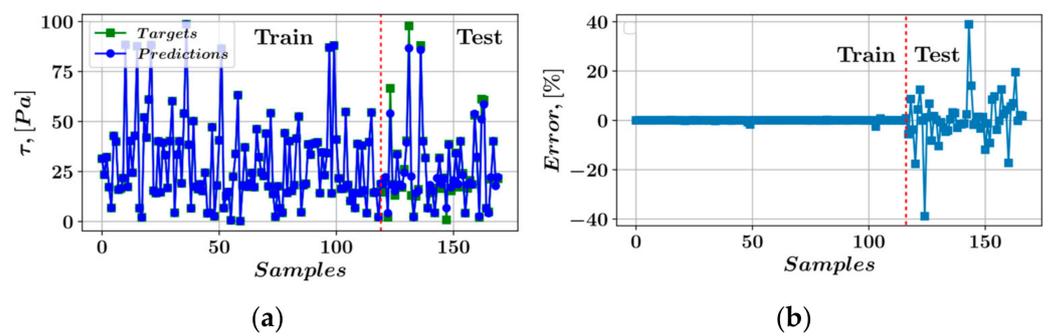
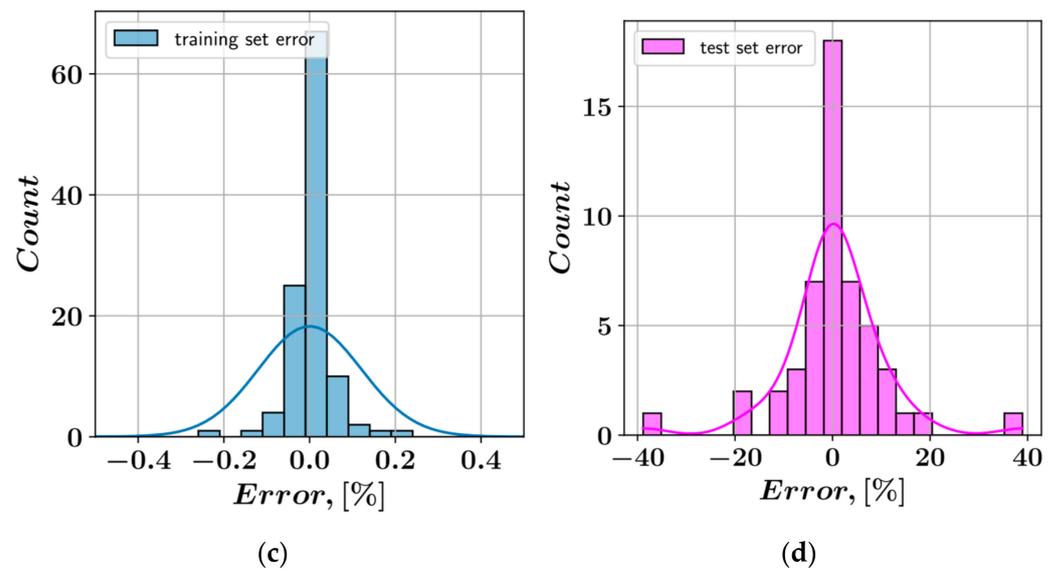


Figure 10. Cont.

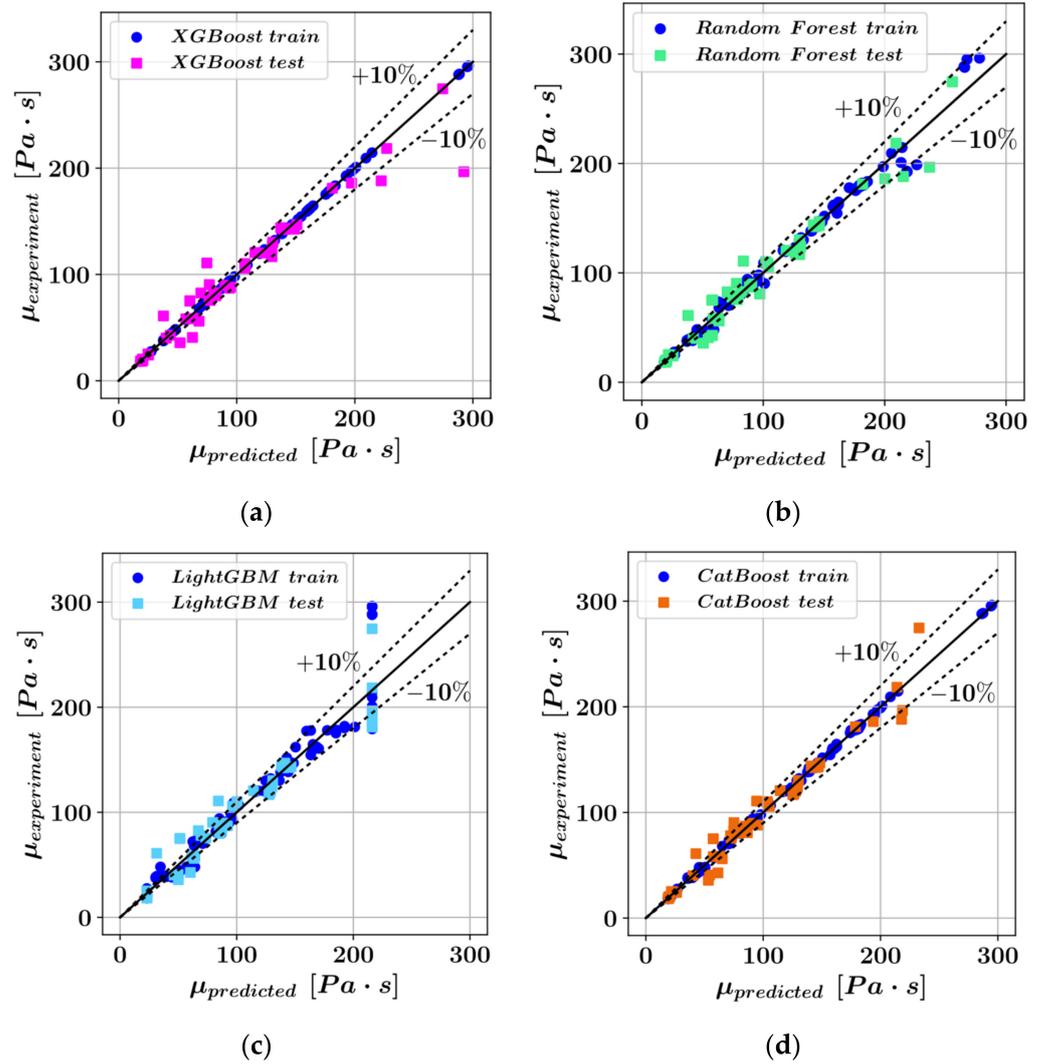


**Figure 10.** (a) Predictions of the shear stress, (b) error percentages, (c) error distribution of the training set, and (d) error distribution of the test set for the XGBoost model.

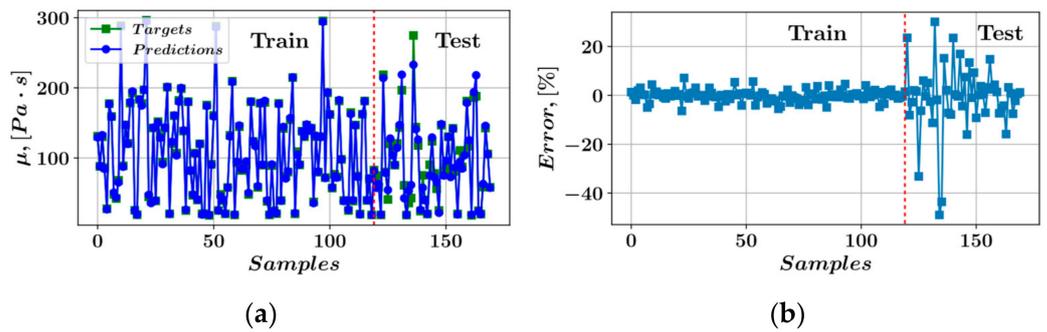
**Table 3.** Prediction accuracy of the machine learning models for shear stress.

Algorithm	R <sup>2</sup>		MAE		VAF		RMSE		Duration [s]
	Train	Test	Train	Test	Train	Test	Train	Test	
XGBoost	<b>0.9997</b>	<b>0.9802</b>	<b>0.094</b>	<b>1.712</b>	<b>99.99</b>	97.04	<b>0.397</b>	<b>2.885</b>	4.54
Random Forest	0.9977	0.9797	0.658	1.795	99.76	<b>98.05</b>	1.037	2.924	3.24
LightGBM	0.8968	0.9111	4.104	3.624	90.08	90.80	6.888	6.114	4.04
CatBoost	0.9988	0.9779	0.572	2.120	99.92	97.98	0.747	3.047	22.69

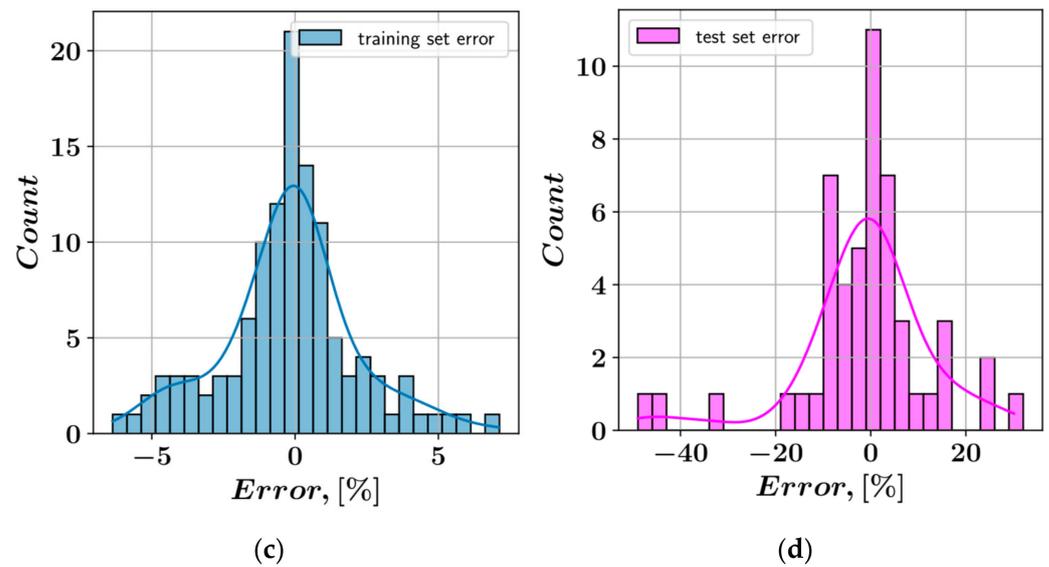
Figure 11 shows the comparison between the predicted and target plastic viscosity values. Similar to Figure 9, the accuracies of the models in predicting the plastic viscosity values are visualized by the positions of the predicted/target value pairs. The slump flow diameter, V-funnel flow time, L-Box passing ability, and shear stress were used as input variables to predict the plastic viscosities ( $\mu$ ) in Figure 11. The model performances in predicting plastic viscosity were quantified using the R<sup>2</sup> score, MAE, VAF, and RMSE metrics, as shown in Table 4. In Table 4, the best accuracy values have been shown in bold font. The performance values in Table 4 show that the CatBoost model was the most accurate for all four metrics, followed by random forest. On the other hand, the random forest algorithm was the fastest in terms of computational speed, whereas the CatBoost algorithm was significantly slower. Figure 12 shows the variation in the prediction error percentages throughout the training and test sets. An overlap of the predicted and target plastic viscosity values can be observed in Figure 12a. Error percentages increased in the transition from training set to test set, shown in Figure 12b. The distribution of these error percentages are shown as histogram plots in Figure 12c,d for the training and test sets, respectively. According to Figure 12c, the error percentages are mainly clustered in the  $\pm 1\%$  range, whereas the error percentages are distributed on a wider range in Figure 12d.



**Figure 11.** Comparison of the experimental and predicted plastic viscosity values for (a) XGBoost, (b) random forest, (c) LightGBM, and (d) CatBoost models.



**Figure 12.** Cont.



**Figure 12.** (a) Predictions of the plastic viscosity, (b) error percentages, (c) error distribution of the training set, and (d) error distribution of the test set for the CatBoost model.

**Table 4.** Prediction accuracy of the machine learning models for plastic viscosity.

Algorithm	R <sup>2</sup>		MAE		VAF		RMSE		Duration [s]
	Train	Test	Train	Test	Train	Test	Train	Test	
XGBoost	<b>0.9999</b>	0.9132	<b>0.041</b>	8.274	<b>99.99</b>	91.56	<b>0.084</b>	16.986	4.66
Random Forest	0.9896	0.9570	3.665	7.703	98.74	95.44	6.846	11.961	3.15
LightGBM	0.9324	0.9387	9.527	9.286	93.61	93.18	17.437	14.270	3.65
CatBoost	0.9986	<b>0.9654</b>	1.764	<b>7.602</b>	99.95	<b>96.32</b>	2.487	<b>10.727</b>	19.82

### SHAP Analysis

The SHAP methodology is an effective way of clarifying the impacts different input features have on the predictions of a model. The SHAP technique explains complex ML models using simpler explanation models that approximate the original model. This method can be summarized through Equation (4), where  $s$  is the explanation model,  $x' \in \{0, 1\}^M$  is a binary variable connected to the actual input feature values  $x$  through a mapping function, such that  $x = h_x(x')$ , and  $M$  is the total number of input features [52].

$$s(x') = \Phi_0 + \sum_{j=1}^M \Phi_j x'_j \quad (4)$$

$$\Phi_j = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)]$$

In Equation (4),  $\Phi_j$  represents the effect of each input feature. The set of all features is denoted with  $F$ , and  $S$  is a subset of  $F$  that does not contain the  $i$ -th feature. The values of the input features in the subset  $S$  are contained in the vector  $x_S$ . The output of the explanation model for when all input features are missing is denoted with  $\Phi_0$ . In Equation (4), the function  $f$  represents the actual predictive model. The Shapley regression values are computed based on the differences between the model predictions with and without the  $i$ -th input feature. The SHAP process can also be visually explained, as in Figure 13, where  $\Phi_1$  and  $\Phi_2$  have an increasing effect and  $\Phi_3$  has a decreasing effect on the model prediction. Further details of this procedure can be found in [52].

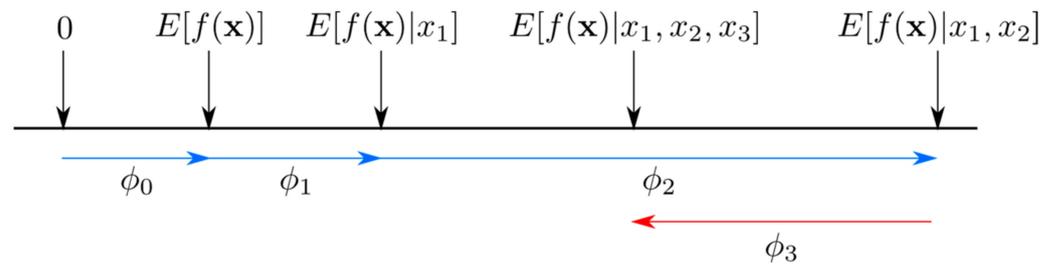


Figure 13. SHAP (SHapley Additive exPlanation) values [53].

The SHAP summary plots in Figures 14 and 15 visualize the impact of each input variable on predicted shear stress and plastic viscosity, respectively. In Figures 14 and 15, each data sample is shown with a dot for each input feature. The horizontal positions of these dots are determined by their SHAP values. The positive SHAP values indicate an increasing impact of the input feature on the model output, whereas negative SHAP values indicate a decreasing impact. Furthermore, the magnitude of each feature in a certain sample is represented by the color of its dot in the diagram, where shades of red correspond to high feature values and shades of blue correspond to low feature values. Figure 14 is based on the XGBoost model, whereas Figure 15 is generated using the CatBoost model, as these were the models with the highest accuracy in the prediction of  $\tau$  and  $\mu$ , respectively. According to Figure 14, the plastic viscosity had the highest impact on the XGBoost model prediction, followed by V-funnel flow time, passing ability, and slump flow diameter. Figure 15 shows that shear stress had the highest impact on the predicted plastic viscosity, and the remaining features had the same ranking ( $D < PA < t$ ) as in Figure 14. Figure 14 shows that high values of  $t$  and  $\mu$  have an increasing effect on the predicted value of  $\tau$ , whereas high values of  $PA$  and  $D$  have a decreasing effect on  $\tau$ . Similarly, Figure 15 shows that high values of  $\tau$  and  $t$  have an increasing effect on  $\mu$ , and high values of  $PA$  and  $D$  have a decreasing effect on  $\mu$ . These observations are also supported by Figures 7 and 8.

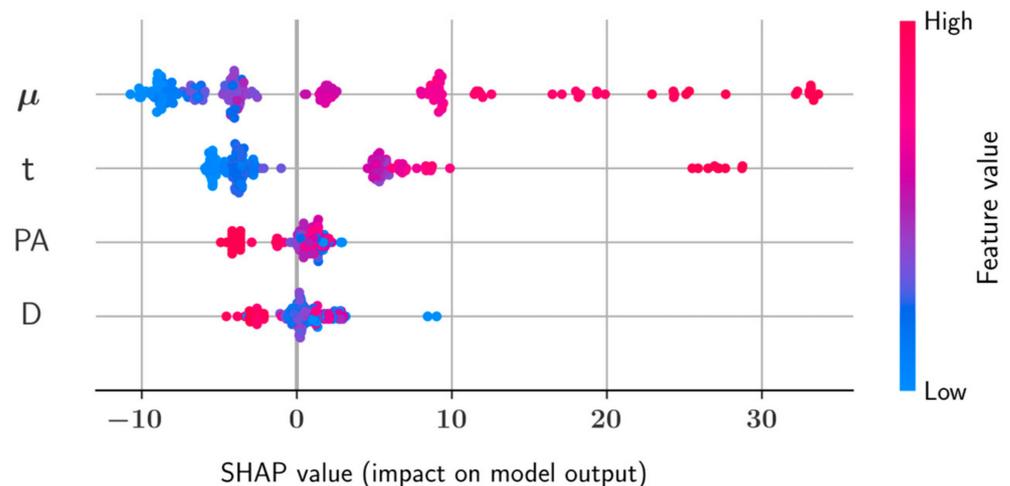
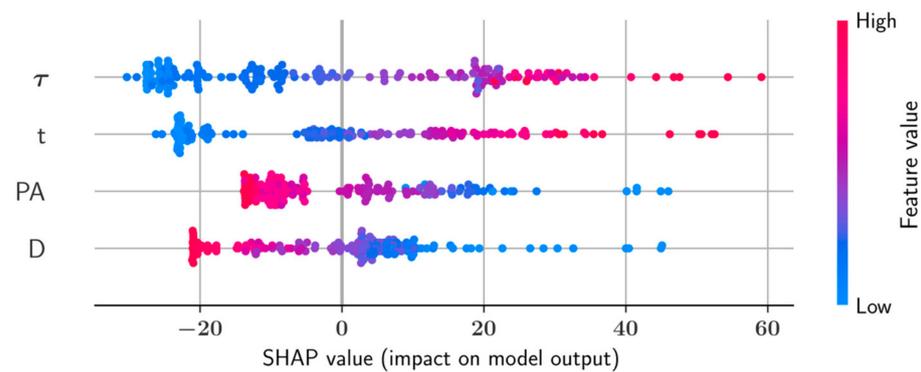
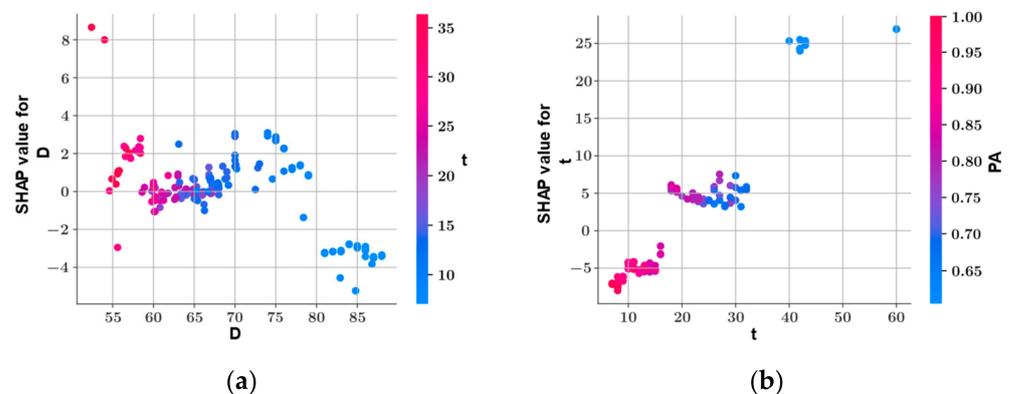


Figure 14. SHAP summary plot for the prediction of shear stress ( $\tau$ ) (XGBoost).



**Figure 15.** SHAP summary plot for the prediction of plastic viscosity ( $\mu$ ) (CatBoost).

The feature dependence plots in Figures 16 and 17 were generated from the XGBoost and CatBoost models, respectively. The feature dependence plots were generated for each input feature, and each feature dependence plot contains information about the variation of the SHAP values of a feature, with respect to the magnitude of that feature. For each input feature, another feature most dependent on that feature was included in the plots. The magnitudes of the most dependent features are presented with colors. Figure 16a shows that as the value of D increases, this variable begins to have negative SHAP values and the V-funnel flow time  $t$ , which is the most dependent variable on D, begins to have smaller values. From Figure 16b, as the value of  $t$  increases, passing ability (PA) decreases and the SHAP value for  $t$  increases, which indicates that increases in  $t$  increases the predicted shear stress ( $\tau$ ). Figure 16c shows that increasing the PA variable also increases D. On the other hand, increases in both D and PA have a decreasing effect on the predicted  $\tau$ . According to Figure 16d, increasing the value of the plastic viscosity ( $\mu$ ) decreases D, but has an increasing effect on  $\tau$ . Figure 17 shows the feature dependence plots obtained during the prediction of  $\mu$ . Figure 17a shows that for the larger values of D, the SHAP value becomes negative and shear stress decreases. The variation of  $t$  and corresponding SHAP values in Figure 17b resembles Figure 16b, such that PA appears to be the input feature most dependent on  $t$  in both plots, and is adversely affected by  $t$ . Figure 17c shows that an increase in the passing ability of concrete reduces plastic viscosity (negative SHAP values) and corresponds to lower  $\tau$  values. According to Figure 17d, the SHAP values of  $\tau$  almost linearly increase, up to when  $\tau$  reaches around 50 MPa. From this point, more irregular variation in the SHAP value is observed, whereas the D values are on the lower side of the spectrum for these larger values of  $\tau$ .



**Figure 16.** Cont.

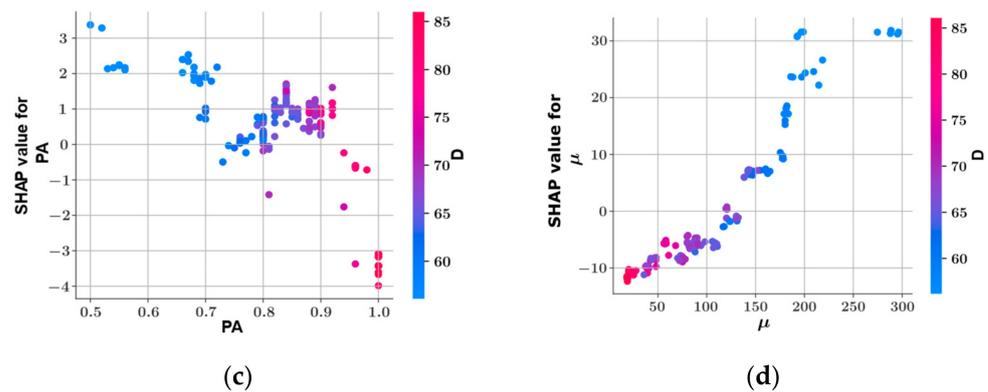


Figure 16. Feature dependence plots for the variables in the prediction of  $\tau$ : (a) D, (b) t, (c) PA, and (d)  $\mu$ .

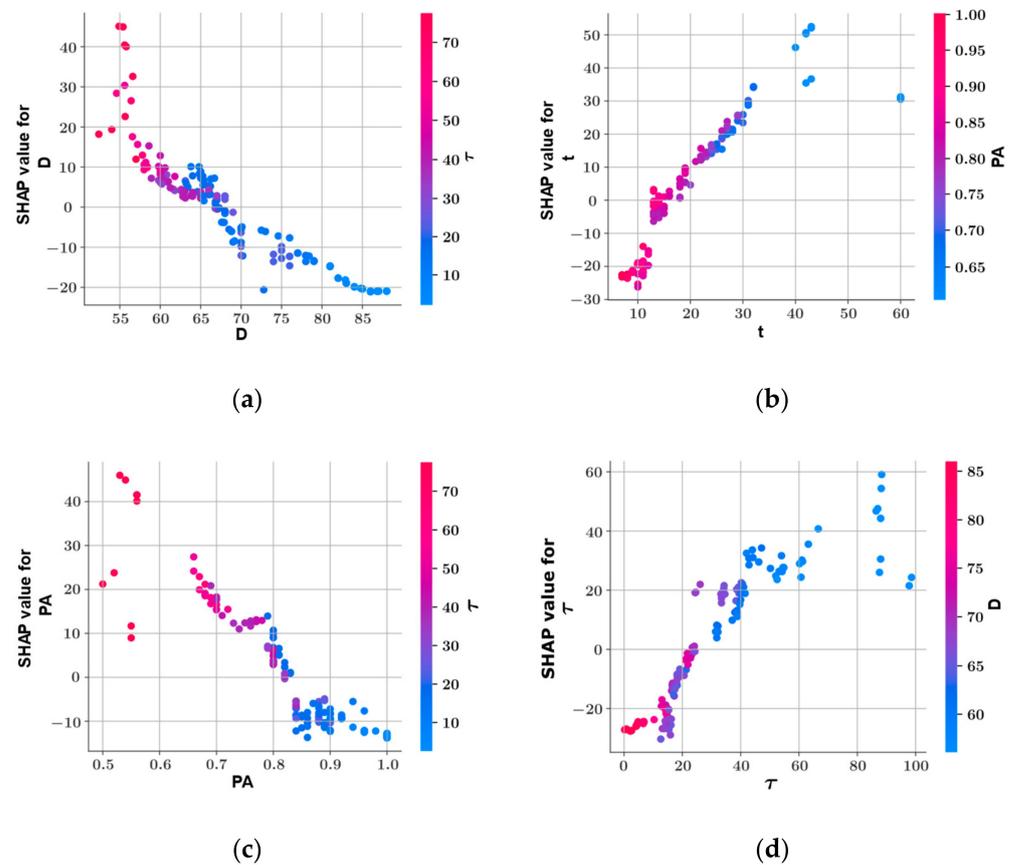
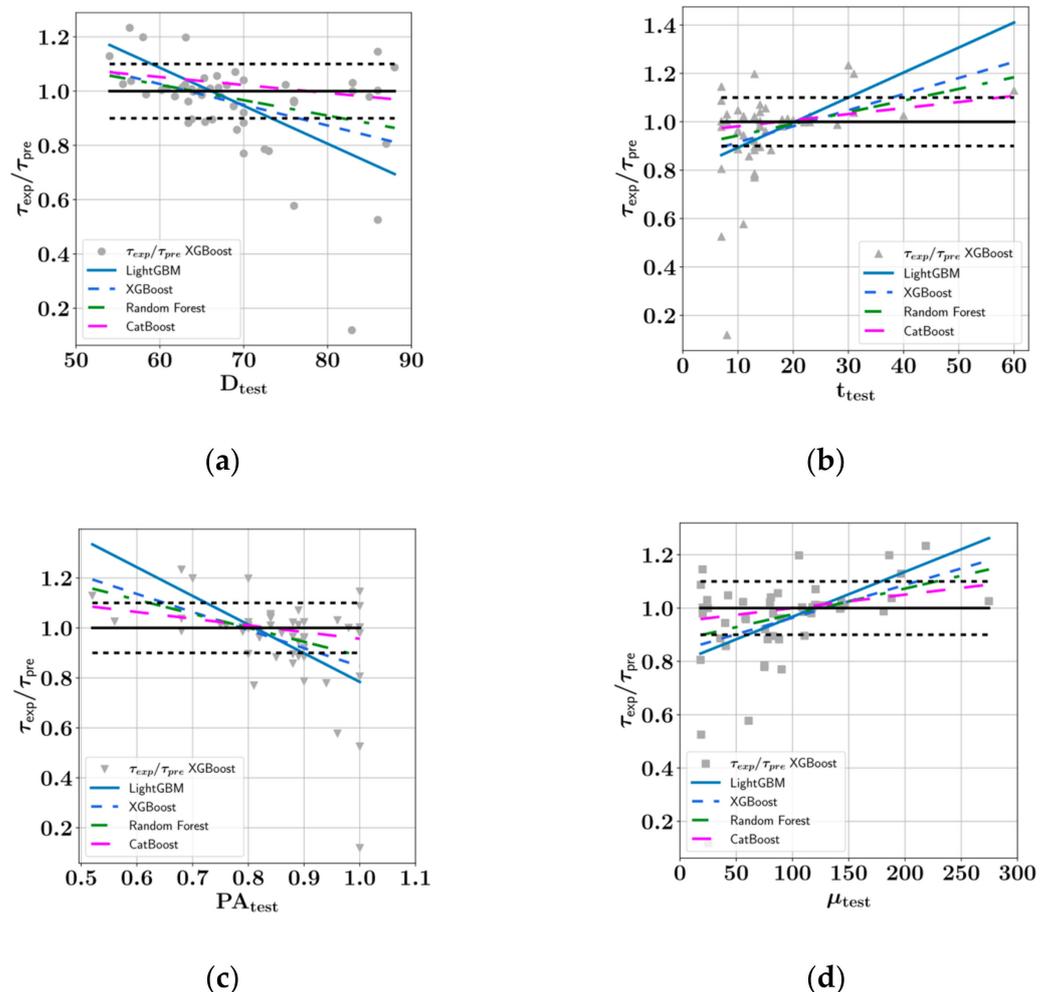


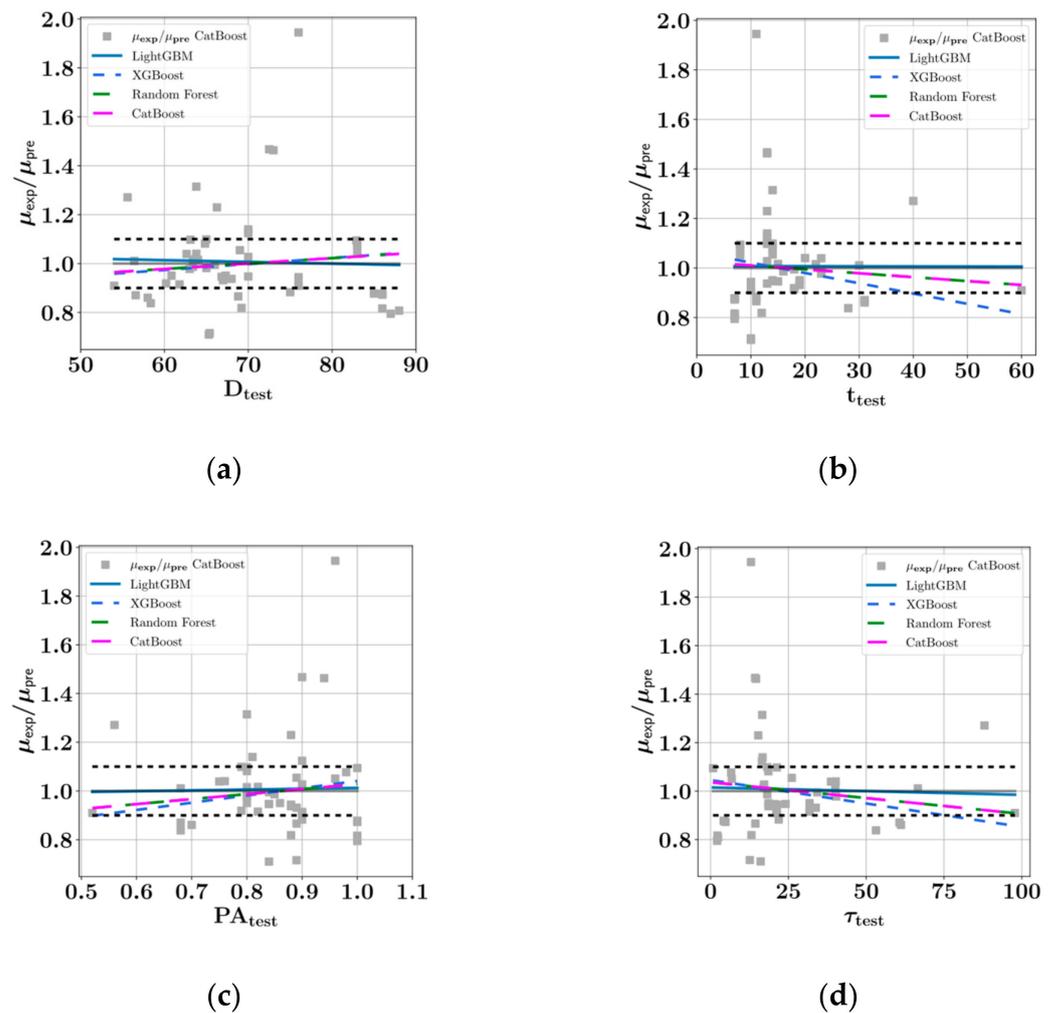
Figure 17. Feature dependence plots for the variables in the prediction of  $\mu$ : (a) D, (b) t, (c) PA, and (d)  $\tau$ .

The plots of prediction consistency in Figures 18 and 19 were generated using the predictions of the models on the test sets, which comprised 30% of the entire data set. The vertical axes in these plots show the ratio of the actual experimental values to predicted values. For each predictive model, a linear curve was fitted onto the  $\tau_{\text{exp}}/\tau_{\text{pre}}$  and  $\mu_{\text{exp}}/\mu_{\text{pre}}$  values in Figures 18 and 19, respectively. In addition to the linear curves, the  $\tau_{\text{exp}}/\tau_{\text{pre}}$  and  $\mu_{\text{exp}}/\mu_{\text{pre}}$  values from the XGBoost and CatBoost plots have been presented as scatter plots. The perfect match between the experimental and predicted values, as well as  $\pm 10\%$  deviation from the perfect match, has been shown with horizontal solid and dashed lines in black in each subplot of Figures 18 and 19. The consistency plots present information about the tendencies of each predictive model to overestimate or underestimate the target

value. In this sense, a perfectly horizontal curve would indicate that the model predictions are equally good for the entire range of input features. Figure 18 shows that the LightGBM predictions shown with the solid blue line had the least consistency among the predictive models. For  $D$  values less than 65 cm, the LightGBM predictions underestimated the  $\tau$  values, and for  $D$  values less than 58 cm, the LightGBM model underestimated  $\tau$  over 10%. Similarly, for  $D$  values greater than 75 cm, the LightGBM model overestimated  $\tau$  more than 10%. A similar pattern of consistency could also be observed for other input features during the estimation of  $\tau$ . On the other hand, Figure 18 shows that the most consistent model in the prediction of  $\tau$  was the CatBoost model. A different pattern of consistency was observed in the prediction of  $\tau$  such that LightGBM was observed to be the most consistent model. The curve fit for the  $\mu_{\text{exp}}/\mu_{\text{pre}}$  values obtained through the LightGBM model exhibit a near horizontal course for all input features. On the other hand, the least consistent model was the XGBoost model in the prediction of  $\mu$ . Overall, the predictive models exhibited better consistency in the prediction of  $\mu$ . XGBoost was the only model that over- or underestimated the target value by more than 10% for  $\tau > 75$  MPa and  $t > 40$  s.

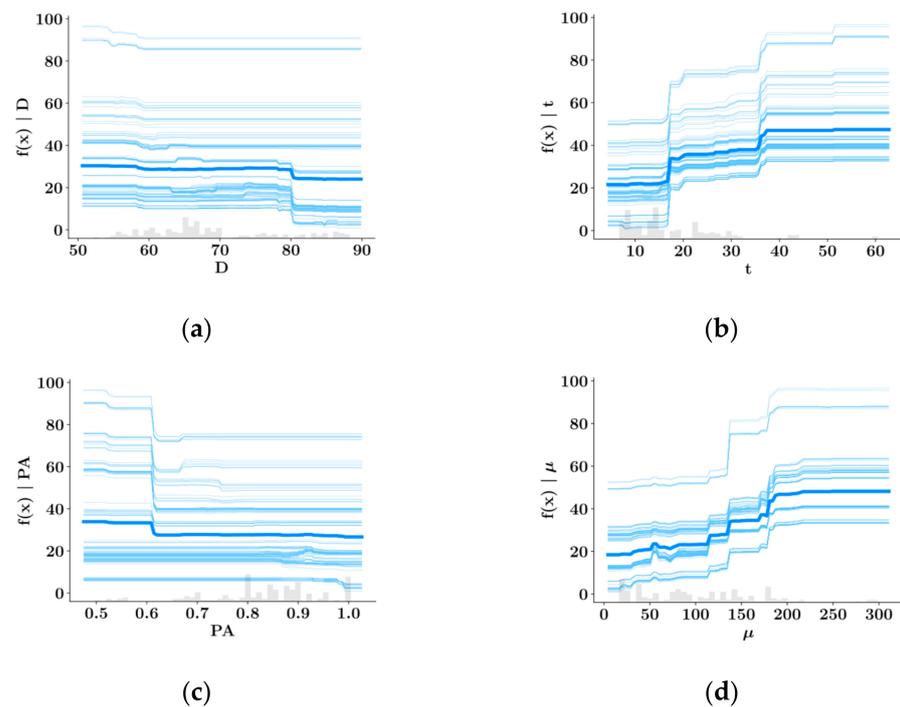


**Figure 18.** Prediction consistency plots for the variables in the prediction of  $\tau$ : (a)  $D$ , (b)  $t$ , (c)  $PA$ , and (d)  $\tau$ .

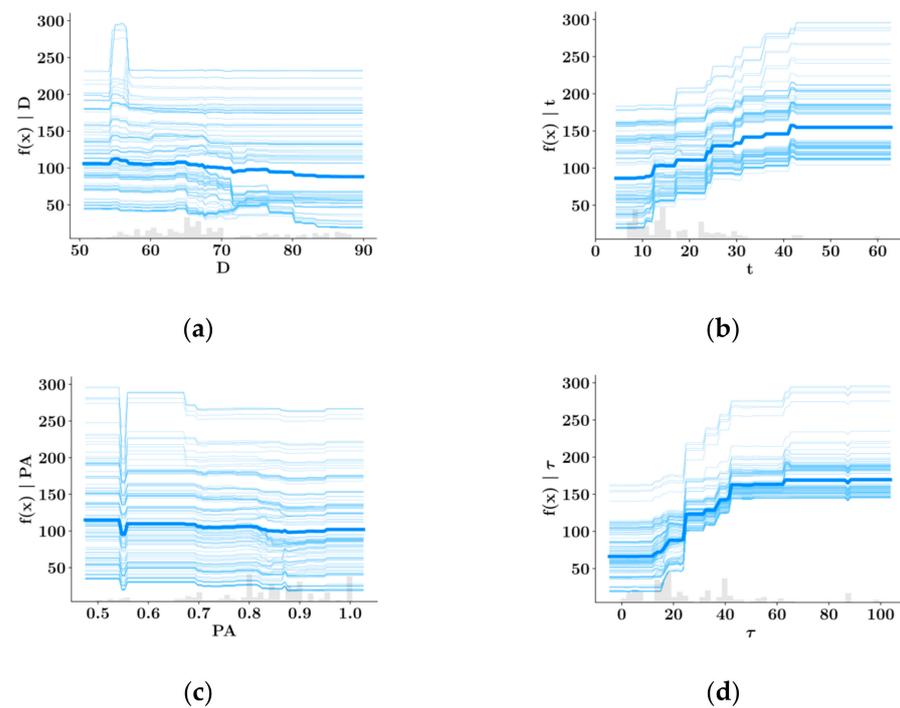


**Figure 19.** Prediction consistency plots for the variables in the prediction of  $\mu$ : (a)  $D$ , (b)  $t$ , (c)  $PA$ , and (d)  $\tau$ .

The ICE plots in Figures 20 and 21 show the variations in the predictive model output to each input feature for each data sample in the dataset. For any given input feature and data sample, all the values of the remaining input features were kept constant while the model predictions (denoted by  $f(x)$ ) were calculated for different values of that particular input feature. As a result, for each data sample, a different curve was generated in the ICE plots. These plots show whether there are differences in the interactions between the feature values and model predictions for different samples. In addition to the individual curves, Figures 20 and 21 also contain darker blue curves, which represent the average of all the curves in a plot. Both Figures 20 and 21 show that most curves have a similar course in all subplots, whereas only minor deviations from the average curve pattern can be observed in some of the curves. For example, in Figure 21 for  $D > 70$  cm, most curves exhibit a near horizontal course, whereas in some of the samples, a slight increase in predicted  $\mu$  values can be observed for  $70 < D < 75$  cm. Similarly, in Figure 20, most of the curves exhibit a near horizontal course for  $PA > 0.8$ , whereas some of the samples in the  $0.88 < PA < 0.92$  range show a slight increase in the predicted  $\tau$  value. Also, for  $PA > 0.95$ , a slight drop in the predicted  $\tau$  value can be observed in some of the samples.



**Figure 20.** Individual conditional expectation (ICE) plots for the variables in the prediction of  $\tau$ : (a)  $D$ , (b)  $t$ , (c)  $PA$ , and (d)  $\mu$ .



**Figure 21.** Individual conditional expectation (ICE) plots for the variables in the prediction of  $\mu$ : (a)  $D$ , (b)  $t$ , (c)  $PA$ , and (d)  $\tau$ .

#### 4. Discussion and Conclusions

Yield stress and plastic viscosity are significant indicators of the workability of SCC. Therefore, having the appropriate tools to accurately predict its material properties is a great advantage. In recent years, machine learning techniques have been increasingly used to investigate different engineering systems to predict their structural behavior. The current

study demonstrates the performances of four ensemble learning techniques on a newly curated data set that contains information about the relationships between the slump flow diameter, V-funnel flow time, passing ability from L-Box tests, plastic viscosity, and shear strength of SCC. By splitting the dataset into training and testing sets, the ensemble learning models were developed on the training set, and then their performances were measured on the test set. As the metrics for predictive model accuracy, the commonly used coefficient of determination, root mean square error, and mean absolute error were adopted. The current work includes extensive model interpretations based on the SHAP algorithm, consistency plots, and individual conditional expectation (ICE) plots. The interactions between the different input features and predictive model output values have been presented in detail using feature dependence plots. The SHAP analysis revealed that plastic viscosity had the greatest influence on model output when predicting shear stress, and shear stress had the highest impact when predicting plastic viscosity. Furthermore, slump flow diameter was found to have the lowest impact on the model output for the predictions of yield stress and plastic viscosity of SCC. Among all ensemble learning models, the CatBoost model was the most consistent model in the prediction of shear stress, and LightGBM was the most consistent model in the prediction of plastic viscosity. The relationships between the predictive model outputs and different input features were further clarified using ICE plots, in which the model predictions were visualized for the entire range of input features for each data sample. The average values of these predictions are presented in the ICE plots. The main conclusions of the study can be summarized as follows:

- The XGBoost model performed best on the test set during the prediction of shear stress as a function of the variables  $D$ ,  $t$ ,  $PA$ , and  $\mu$ , with an  $R^2$  score of 0.9802, followed by random forest ( $R^2 = 0.9797$ ), CatBoost ( $R^2 = 0.9779$ ), and LightGBM ( $R^2 = 0.9111$ ).
- The CatBoost model performed best on the test set during the prediction of plastic viscosity as a function of  $D$ ,  $t$ ,  $PA$ , and  $\tau$ , with an  $R^2$  score of 0.9654, followed by random forest ( $R^2 = 0.9570$ ), LightGBM ( $R^2 = 0.9387$ ), and XGBoost ( $R^2 = 0.9132$ ).
- Shear strength and plastic viscosity features were found to have the highest impact on the predictive model output during prediction of each other, based on the SHAP analysis. In the prediction of both shear stress and plastic viscosity, the slump flow diameter was found to have the lowest impact on the model output.
- In the prediction of shear stress, the most consistent predictions were made by the CatBoost model, whereas the LightGBM model was most consistent in predicting plastic viscosity.

Future research should include the study of the compressive strength and split tensile strength of SCC as a function of different rheological properties. The results of sieve segregation resistance tests could be included in predictive model development. Furthermore, closed-form equations could be developed with the help of optimization techniques that relate different rheological parameters to compressive strength, split tension strength, shear stress, and plastic viscosity of SCC. Overall, the availability of open-source machine learning techniques and predictive models is a great benefit for practical engineers and researchers working in the field of concrete research.

**Author Contributions:** Writing—original draft preparation, C.C.; conceptualization, C.C. and G.B.; data curation, C.C.; visualization, C.C.; writing—review and editing, G.B., S.K. and Z.W.G.; supervision G.B., S.K. and Z.W.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data will be made available upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bartos, P.J.M.; Sonebi, M.; Tamimi, A.K. *Report of Rilem Technical Committee TC145 WSM, Compendium of Tests, Workability and Rheology of Fresh Concrete*; RILEM (The International Union of Testing and Research Laboratories for Materials and Structures): Paris, France, 2002.
2. Ben Aicha, M.; Burtschell, Y.; Alaoui, A.H.; El Harrouni, K.; Jalbaud, O. Correlation between Bleeding and Rheological Characteristics of Self-Compacting Concrete. *J. Mater. Civ. Eng.* **2017**, *29*, 05017001. [[CrossRef](#)]
3. Okamura, H.; Ouchi, M. Self-compacting high performance concrete. *Prog. Struct. Eng. Mater.* **1998**, *1*, 378–383. [[CrossRef](#)]
4. Pashias, N.; Boger, D.V.; Summers, J.; Glenister, D.J. A fifty cent rheometer for yield stress measurement. *J. Rheol.* **1996**, *40*, 1179–1189. [[CrossRef](#)]
5. Roussel, N. Correlation between Yield Stress and Slump: Comparison between Numerical Simulations and Concrete Rheometers Results. *Mater. Struct.* **2006**, *39*, 501–509. [[CrossRef](#)]
6. Neophytou, M.K.A.; Pourgouri, S.; Kanellopoulos, A.D.; Petrou, M.F.; Ioannou, I.; Georgiou, G.; Alexandrou, A. Determination of the rheological parameters of self-compacting concrete matrix using slump flow test. *Appl. Rheol.* **2010**, *20*. [[CrossRef](#)]
7. Schowalter, W.R.; Christensen, G. Toward a rationalization of the slump test for fresh concrete: Comparisons of calculations and experiments. *J. Rheol.* **1998**, *42*, 865–870. [[CrossRef](#)]
8. Lee, J.H.; Kim, J.H.; Yoon, J.Y. Prediction of the yield stress of concrete considering the thickness of excess paste layer. *Constr. Build. Mater.* **2018**, *173*, 411–418. [[CrossRef](#)]
9. Sun, Z.; Feng, D.-C.; Mangalathu, S.; Wang, W.-J.; Su, D. Effectiveness Assessment of TMDs in Bridges under Strong Winds Incorporating Machine-Learning Techniques. *J. Perform. Constr. Facil.* **2022**, *36*, 04022036. [[CrossRef](#)]
10. Todorov, B.; Billah, A.M. Post-earthquake seismic capacity estimation of reinforced concrete bridge piers using Machine learning techniques. *Structures* **2022**, *41*, 1190–1206. [[CrossRef](#)]
11. Todorov, B.; Billah, A.M. Machine learning driven seismic performance limit state identification for performance-based seismic design of bridge piers. *Eng. Struct.* **2022**, *255*, 113919. [[CrossRef](#)]
12. Xu, J.; Feng, D.; Mangalathu, S.; Jeon, J. Data-driven rapid damage evaluation for life-cycle seismic assessment of regional reinforced concrete bridges. *Earthq. Eng. Struct. Dyn.* **2022**, *51*, 2730–2751. [[CrossRef](#)]
13. Chen, M.; Mangalathu, S.; Jeon, J.-S. Machine Learning–Based Seismic Reliability Assessment of Bridge Networks. *J. Struct. Eng.* **2022**, *148*, 06022002. [[CrossRef](#)]
14. Somala, S.N.; Karthikeyan, K.; Mangalathu, S. Time period estimation of masonry infilled RC frames using machine learning techniques. *Structures* **2021**, *34*, 1560–1566. [[CrossRef](#)]
15. Hwang, S.-H.; Mangalathu, S.; Shin, J.; Jeon, J.-S. Estimation of economic seismic loss of steel moment-frame buildings using a machine learning algorithm. *Eng. Struct.* **2022**, *254*, 113877. [[CrossRef](#)]
16. Hwang, S.-H.; Mangalathu, S.; Shin, J.; Jeon, J.-S. Machine learning-based approaches for seismic demand and collapse of ductile reinforced concrete building frames. *J. Build. Eng.* **2020**, *34*, 101905. [[CrossRef](#)]
17. Moghaddas, S.A.; Nekoei, M.; Golafshani, E.M.; Nehdi, M.; Arashpour, M. Modeling carbonation depth of recycled aggregate concrete using novel automatic regression technique. *J. Clean. Prod.* **2022**, *371*, 133522. [[CrossRef](#)]
18. Safayenikoo, H.; Nejati, F.; Nehdi, M.L. Indirect Analysis of Concrete Slump Using Different Metaheuristic-Empowered Neural Processors. *Sustainability* **2022**, *14*, 10373. [[CrossRef](#)]
19. Shah, H.A.; Nehdi, M.L.; Khan, M.I.; Akmal, U.; Alabduljabbar, H.; Mohamed, A.; Sheraz, M. Predicting Compressive and Splitting Tensile Strengths of Silica Fume Concrete Using M5P Model Tree Algorithm. *Materials* **2022**, *15*, 5436. [[CrossRef](#)]
20. Zhao, G.; Wang, H.; Li, Z. Capillary water absorption values estimation of building stones by ensembled and hybrid SVR models. *J. Intell. Fuzzy Syst.* **2022**, 1–13. [[CrossRef](#)]
21. Benemaran, R.S.; Esmaili-Falak, M.; Javadi, A. Predicting resilient modulus of flexible pavement foundation using extreme gradient boosting based optimised models. *Int. J. Pavement Eng.* **2022**. [[CrossRef](#)]
22. Bekdaş, G.; Cakiroglu, C.; Kim, S.; Geem, Z.W. Optimization and Predictive Modeling of Reinforced Concrete Circular Columns. *Materials* **2022**, *15*, 6624. [[CrossRef](#)] [[PubMed](#)]
23. Cakiroglu, C.; Islam, K.; Bekdaş, G.; Kim, S.; Geem, Z.W. Interpretable Machine Learning Algorithms to Predict the Axial Capacity of FRP-Reinforced Concrete Columns. *Materials* **2022**, *15*, 2742. [[CrossRef](#)] [[PubMed](#)]
24. Cakiroglu, C.; Islam, K.; Bekdaş, G.; Isikdag, U.; Mangalathu, S. Explainable machine learning models for predicting the axial compression capacity of concrete filled steel tubular columns. *Constr. Build. Mater.* **2022**, *356*, 129227. [[CrossRef](#)]
25. BEN Aicha, M.; Al Asri, Y.; Zaher, M.; Alaoui, A.H.; Burtschell, Y. Prediction of rheological behavior of self-compacting concrete by multi-variable regression and artificial neural networks. *Powder Technol.* **2022**, *401*, 117345. [[CrossRef](#)]
26. Alyamaç, K.E.; Ince, R. A preliminary concrete mix design for SCC with marble powders. *Constr. Build. Mater.* **2009**, *23*, 1201–1210. [[CrossRef](#)]
27. Taffese, W. Data-Driven Method for Enhanced Corrosion Assessment of Reinforced Concrete Structures. *arXiv* **2020**, arXiv:2007.01164. [[CrossRef](#)]
28. Yuan, J.; Zhao, M.; Esmaili-Falak, M. A comparative study on predicting the rapid chloride permeability of self-compacting concrete using meta-heuristic algorithm and artificial intelligence techniques. *Struct. Concr.* **2022**, *23*, 753–774. [[CrossRef](#)]
29. Kumar, S.; Rai, B.; Biswas, R.; Samui, P.; Kim, D. Prediction of rapid chloride permeability of self-compacting concrete using Multivariate Adaptive Regression Spline and Minimax Probability Machine Regression. *J. Build. Eng.* **2020**, *32*, 101490. [[CrossRef](#)]

30. Ge, D.-M.; Zhao, L.-C.; Esmaeili-Falak, M. Estimation of rapid chloride permeability of SCC using hyperparameters optimized random forest models. *J. Sustain. Cem. Mater.* **2022**, 1–19. [[CrossRef](#)]
31. Amin, M.N.; Raheel, M.; Iqbal, M.; Khan, K.; Qadir, M.G.; Jalal, F.E.; Alabdullah, A.A.; Ajwad, A.; Al-Faiad, M.A.; Abu-Arab, A.M. Prediction of Rapid Chloride Penetration Resistance to Assess the Influence of Affecting Variables on Metakaolin-Based Concrete Using Gene Expression Programming. *Materials* **2022**, *15*, 6959. [[CrossRef](#)]
32. Aggarwal, S.; Bhargava, G.; Sihag, P. Prediction of compressive strength of scc-containing metakaolin and rice husk ash using machine learning algorithms. In *Computational Technologies in Materials Science*; CRC Press: Boca Raton, FL, USA, 2021; pp. 193–205. [[CrossRef](#)]
33. Farooq, F.; Czarnecki, S.; Niewiadomski, P.; Aslam, F.; Alabduljabbar, H.; Ostrowski, K.A.; Śliwa-Wieczorek, K.; Nowobilski, T.; Malazdrewicz, S. A Comparative Study for the Prediction of the Compressive Strength of Self-Compacting Concrete Modified with Fly Ash. *Materials* **2021**, *14*, 4934. [[CrossRef](#)]
34. Zhu, Y.; Huang, L.; Zhang, Z.; Bayrami, B. Estimation of splitting tensile strength of modified recycled aggregate concrete using hybrid algorithms. *Steel Compos. Struct.* **2022**, *44*, 375–392. [[CrossRef](#)]
35. De-Prado-Gil, J.; Palencia, C.; Jagadesh, P.; Martínez-García, R. A Comparison of Machine Learning Tools That Model the Splitting Tensile Strength of Self-Compacting Recycled Aggregate Concrete. *Materials* **2022**, *15*, 4164. [[CrossRef](#)]
36. EFNARC. Specification and Guidelines for Self-Compacting Concrete. 2002. Available online: <https://www.feb.unesp.br/pbastos/c.especiais/Efnarc.pdf> (accessed on 16 September 2022).
37. JSCE, Japan Society of Civil Engineers. Recommendations for Self-Compacting Concrete, Concrete Library of JSCE. 1999, 31. 77p. Available online: <http://www.jsce.or.jp/committee/concrete/e/newsletter/newsletter01/recommendation/selfcompact/4.pdf> (accessed on 16 September 2022).
38. Yang, S.; Zhang, J.; An, X.; Qi, B.; Li, W.; Shen, D.; Li, P.; Lv, M. The Effect of Sand Type on the Rheological Properties of Self-Compacting Mortar. *Buildings* **2021**, *11*, 441. [[CrossRef](#)]
39. Sahraoui, M.; Bouziani, T. Effects of fine aggregates types and contents on rheological and fresh properties of SCC. *J. Build. Eng.* **2019**, *26*, 100890. [[CrossRef](#)]
40. EL Asri, Y.; Benaicha, M.; Zaher, M.; Alaoui, A.H. Prediction of plastic viscosity and yield stress of self-compacting concrete using machine learning technics. *Mater. Today: Proc.* **2022**, *59*, A7–A13. [[CrossRef](#)]
41. Benaicha, M.; Roguiez, X.; Jalbaud, O.; Burtschell, Y.; Alaoui, A.H. Influence of silica fume and viscosity modifying agent on the mechanical and rheological behavior of self compacting concrete. *Constr. Build. Mater.* **2015**, *84*, 103–110. [[CrossRef](#)]
42. Rahman, J.; Ahmed, K.S.; Khan, N.I.; Islam, K.; Mangalathu, S. Data-driven shear strength prediction of steel fiber reinforced concrete beams using machine learning approach. *Eng. Struct.* **2021**, *233*, 111743. [[CrossRef](#)]
43. Somala, S.N.; Chanda, S.; Karthikeyan, K.; Mangalathu, S. Explainable Machine learning on New Zealand strong motion for PGV and PGA. *Structures* **2021**, *34*, 4977–4985. [[CrossRef](#)]
44. Degtyarev, V.; Naser, M. Boosting machines for predicting shear strength of CFS channels with staggered web perforations. *Structures* **2021**, *34*, 3391–3403. [[CrossRef](#)]
45. Feng, D.-C.; Cetiner, B.; Kakavand, M.R.A.; Taciroglu, E. Data-Driven Approach to Predict the Plastic Hinge Length of Reinforced Concrete Columns and Its Application. *J. Struct. Eng.* **2021**, *147*, 04020332. [[CrossRef](#)]
46. Nguyen, Q.H.; Ly, H.-B.; Ho, L.S.; Al-Ansari, N.; Van Le, H.; Tran, V.Q.; Prakash, I.; Pham, B.T. Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Math. Probl. Eng.* **2021**, *2021*, 1–15. [[CrossRef](#)]
47. Bakouregui, A.S.; Mohamed, H.M.; Yahia, A.; Benmokrane, B. Explainable extreme gradient boosting tree-based prediction of load-carrying capacity of FRP-RC columns. *Eng. Struct.* **2021**, *245*, 112836. [[CrossRef](#)]
48. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
49. Feng, D.-C.; Wang, W.-J.; Mangalathu, S.; Hu, G.; Wu, T. Implementing ensemble learning methods to predict the shear strength of RC deep beams with/without web reinforcements. *Eng. Struct.* **2021**, *235*, 111979. [[CrossRef](#)]
50. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates Inc.: New York, NY, USA, 2017.
51. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montréal, QC, Canada, 3–8 December 2018; Curran Associates Inc.: New York, NY, USA, 2018.
52. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017. [[CrossRef](#)]
53. Mangalathu, S.; Hwang, S.-H.; Jeon, J.-S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Eng. Struct.* **2020**, *219*, 110927. [[CrossRef](#)]