applied sciences

MDPI

Article

Evaluating Parameter Influences on Planted Concrete Properties via Explainable Machine Learning Models

Xiansheng Duan [1], Ming Zhang [1,*] and Runjuan Zhou [2]

[1] School of Architecture and Civil Engineering, Anhui Polytechnic University, Wuhu 241000, China
[2] School of Chemical and Environmental Engineering, Anhui Polytechnic University, Wuhu 241000, China
* Correspondence: zhangming@ahpu.edu.cn

Abstract

To investigate the complex functional relationships between pH, effective porosity, and compressive strength of planted concrete and their corresponding mixing ratios, a comprehensive database was developed from the relevant published literature. In this study, four machine learning (ML) algorithms were employed: a single algorithm—Multi-Layer Perceptron (MLP), and three ensemble algorithms—Gradient Boosting Regression (GBR), Extreme Gradient Boosting (XGBoost), and Random Forest Regression (RFR)—to predict the pH, effective porosity, and compressive strength of planted concrete. Additionally, the interpretable algorithm Shapley Additive Explanations (SHAP) was used to evaluate both global and local interpretations independent of the ML algorithms, providing insight into the decision-making process. The results demonstrate that the RFR algorithm achieved the highest R^2 values of 0.93 (pH), 0.97 (effective porosity), and 0.94 (compressive strength) in predicting planted concrete properties, demonstrating optimal predictive performance. Furthermore, cement content was identified as the most influential factor affecting pH, while design porosity and maximum coarse aggregate size were the primary factors influencing effective porosity, in that order. For compressive strength, the two most critical factors were the water reducer and cement content.

Keywords: planted concrete; machine learning; SHAP analysis; pH prediction; effective porosity; compressive strength; mix design parameters

1. Introduction

With the rapid advancement of the global economy, infrastructure development is experiencing unprecedented growth. In water conservancy projects, traditional riverbank stabilization methods typically employ ordinary concrete and other rigid materials to meet industry standards for stability and drainage [1]. However, this approach often prioritizes mechanical properties such as slope stability and structural compressive strength, resulting in rough, gray surfaces that obstruct plant growth due to low porosity [2]. These rigid structures contribute to ecological degradation, disrupting the natural balance and hindering harmonious interactions between humans and the environment. In the era of global environmental change, the adoption of innovative slope protection methods is therefore imperative to enhance urban climate resilience and mitigate the urban heat island effect [3]. Planted concrete serves as a critical nature-based solution (NbS), offering the safety and stability of rigid materials while fostering ecological benefits similar to natural grass-covered slopes [4,5]. By providing habitats for local flora and fauna, it plays a vital role in restoring biodiversity within fragmented riparian ecosystems [6]. This approach

Check for updates

Academic Editor: Asterios Bakolas

Received: 7 December 2025
Revised: 6 January 2026
Accepted: 8 January 2026
Published: 12 January 2026

Copyright: © 2026 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.

represents an inevitable trend in ecological restoration, promoting balanced development and enhancing the environmental resilience of water management infrastructure [7].

Planted concrete, designed with porous structures to support vegetation, requires balancing compressive strength, porosity, and pH through material optimization to achieve both ecological and engineering goals. Peiyuan Zhuang et al. [8] developed a response surface methodology to optimize the performance of planted concrete, identifying that a water-binder ratio of 0.3, fly ash content of 26%, and a designed porosity of 22% could produce concrete with 10.32 MPa compressive strength, 20.00% porosity, and a pH of 11. Chen Jianguo et al. [9] systematically investigated the aggregate size effects, revealing that smaller aggregates significantly influence the porosity–strength relationship, which follows an exponential model. Their work also established important correlations between early-age and 28-day strength development. Cheng Yang et al. [10] found that 20% phosphogypsum replacement could effectively enhance the performance of planted concrete. Their comparative study of six alkali reduction methods demonstrated that magnesium sulfate treatment achieved the best balance between strength preservation (only 26% reduction) and pH reduction. Ganesh Prabhu Ganapathy et al. [11] further confirmed that supplementary cementitious materials play crucial roles, with silica fume being particularly effective in strength development and pH control, while fly ash contributes to creating a more favorable environment for plant growth.

Planted concrete performance prediction faces challenges due to simplified models and inefficient handling of multi-factor interactions, limiting accuracy and practicality in real-world applications. Kong Jiafeng et al. [12] used orthogonal experimental design to optimize the mixture design and used three alkali reduction treatments to reduce alkalinity and improve plant growth performance. The compressive strength, connected porosity, and pH value were analyzed to obtain the optimal level. Cai Jiashi et al. [13] used different designed porosities and cement–aggregate ratios to study the compressive strength and plant performance of planted concrete. Liu et al. [14] used the orthogonal test method to carry out the design of the slope vegetative concrete proportion by adopting low-alkalinity sulfur-aluminate cement and compared the effects of the concrete water–cement ratio, aggregate particle size, and rich slurry ratio on the compressive strength, porosity, and pH value. Pei Yuanzhuang et al. [8] established a response surface regression model of plant growth in ecological concrete. Based on factors such as the water–binder ratio, fly ash content, and designed porosity, and with 28-day compressive strength, connected porosity, and pH value as response variables, they studied the ecological performance of planted concrete. Navaratnarajah Sathiparan and Feipeng Liu found [15,16] that these models often fail to adequately represent the complex functional relationship between mixing ratios and concrete properties and have limitations in dealing with the complex interactions between multiple factors. Consequently, the methods currently in use often fall short in providing efficient and accurate predictions of planted concrete performance.

ML is increasingly recognized as a powerful alternative to traditional linear or nonlinear regression methods for predicting key properties of planted concrete, such as pH, effective porosity, and compressive strength. Unlike conventional empirical or semi-empirical models, ML techniques offer superior adaptability to diverse datasets by avoiding rigid function expressions [17,18]. The recent literature underscores the transformative potential of ML in the field of sustainable materials; for instance, Wu et al. [19] utilized neural networks to predict both the mechanical and hydraulic properties of pervious concrete. Similarly, Song et al. [20] leveraged ensemble learning to optimize fly-ash-based sustainable mixes, while Asteris et al. [21] introduced hybrid surrogate models to enhance strength prediction accuracy. These advancements are further supported by recent reviews: Xie et al. [22] analyzed the reuse of solid waste in pervious concrete; Sathiparan et al. [15]

established a comprehensive state-of-the-art framework for ML applications in pervious cementitious composites. Among the various ML models, MLP, XGBoost, GBR, and RFR are prominent and widely applied techniques. MLP, a type of artificial neural network, consists of interconnected layers of nodes that mimic the information processing of biological neural systems [23]. XGBoost and GBR leverage ensemble learning through boosted decision trees to enhance predictive accuracy [24], while RFR employs an ensemble of decision trees to achieve robust regression performance [25]. To further interpret the predictions of these models, the SHAP framework is utilized, providing insight into the contribution of each input feature to the model's output [26,27]. Crucially, SHAP interpretability bridges the gap between complex algorithms and practical application [28], as it quantifies the contribution of factors like the water–binder ratio and aggregate size, thereby providing direct guidance for the engineering design and mix proportioning [29,30].

The primary objective of this study is to develop a practical model and analyze the relationships between key factors and outcomes. Utilizing a dataset compiled from the published literature, four ML models (MLP, GBR, XGBoost, and RFR) were applied to predict performance. Given the inherently non-transparent nature of ML, this study also implemented SHAP to enhance interpretability. By elucidating the impact of each input variable, this approach enables engineers to move beyond "black-box" predictions and make informed design decisions based on the identified joint variable effects and predictor importance, ultimately optimizing the application of planted concrete in sustainable infrastructure.

2. Materials and Methods

2.1. Data Acquisition and Preprocessing

To ensure effective ML computations, datasets must be carefully structured beforehand [31]. A well-constructed dataset should not only include a large sample size (i.e., a larger sample set enables the ML model to more comprehensively and accurately capture potential associations between variables) but must also encompass all relevant factors that could influence outcomes. Only by meeting these prerequisites can the proposed research question be addressed with sufficient rigor [32,33].

This study investigates the influence of varying mix ratios of planted concrete on key properties, including pH, effective porosity, and compressive strength [34]. Despite an extensive review of nearly a decade's worth of literature, only 100 initial datasets were identified from minority publications meeting the study's criteria [8,35–42].

The literature search was conducted across major databases (Web of Science, Scopus, and Google Scholar) using keywords such as 'planted concrete', 'porous concrete for vegetation', and 'recycled aggregate planted concrete'. The inclusion criteria were strictly defined: (i) the study must provide a complete mix design (cement, aggregates, water, etc.); (ii) it must report at least two of the three target properties (pH, porosity, strength); and (iii) the testing methods must comply with international or national standards to ensure data consistency. This rigorous selection ensures that the initial 100 datasets represent high-quality, reproducible experimental results.

2.2. Data Augmentation and Physical Constraints

Despite an extensive literature review, only 100 initial experimental datasets were identified. To meet the requirements of high-dimensional machine learning and prevent overfitting, a Gaussian noise-based data augmentation technique was implemented to expand the dataset to 2100 samples. While the conceptual motivation follows the principles of sample expansion in cementitious materials research, we specify the algorithmic details applied in this study: Specifically, for each original experimental data point (X, Y), 20 synthetic samples were generated. A stochastic perturbation was applied to the input features

(X) using the formula: $X_{aug} = X \times (1 + \epsilon)$, where $\epsilon \sim N(0, 0.05^2)$. This 5% noise factor was selected to introduce sufficient variance while ensuring that the synthetic data points remain within the local manifold of the actual physical experimental space. The output labels (Y) remained unchanged to preserve the deterministic functional relationships [43,44].

To ensure the physical validity of the expanded dataset, a Physical-Constraint Filter was applied: any synthetic sample that violated concrete technology principles (e.g., negative compressive strength, connectivity porosity exceeding 1.0, or impossible water–binder ratios) was strictly excluded [45]. This process ensures that the model learns real-world physical laws rather than algorithmic noise [46]. The standard dataset was randomly divided into a training set, validation set, and test set in the ratio of 70%, 15%, and 15% (the corresponding number of datasets is 1400, 300, and 300, respectively) [47].

2.3. Mechanistic Rationale for Variable Selection

Based on the standardized dataset, ten input parameters were identified: (1) minimum particle size (mm), (2) maximum particle size (mm), (3) coarse aggregate content (kg/m^3), (4) water (kg/m^3), (5) cement (kg/m^3), (6) fly ash content (kg/m^3), (7) silica fume content (kg/m^3), (8) mineral powder content (kg/m^3), (9) water-reducing agent dosage (kg/m^3), and (10) designed porosity (%). The corresponding output parameters include (1) pH, (2) effective porosity (%), and (3) compressive strength (MPa). The distribution properties of input and output parameters are listed in Table 1 [47,48].

Table 1. Distribution properties of input and output parameters.

Variable	Unit	Category	Statistics			
			Min	Max	Average	STDEV
Minimum particle size	mm	Input	4.18	22.76	11.50	4.99
Maximum particle size	mm	Input	8.81	45.05	21.58	7.65
Coarse aggregate content	kg/m^3	Input	1243.86	1942.85	1546.43	113.08
Water	kg/m^3	Input	34.67	150.65	80.18	24.45
Cement	kg/m^3	Input	107.11	439.53	234.30	68.35
Fly ash content	kg/m^3	Input	0	152.53	37.36	34.46
Silica fume content	kg/m^3	Input	0	34.95	3.67	8.14
Mineral powder content	kg/m^3	Input	0	98.45	10.60	20.06
Water-reducing agent dosage	kg/m^3	Input	0	4.87	1.53	1.50
Designed porosity	-	Input	0.17	0.35	0.26	0.03
pH	-	Output	9.30	12.96	10.82	0.71
Effective porosity	-	Output	0.11	0.35	0.25	0.04
Compressive strength	MPa	Output	5.11	15.11	9.55	1.97

The ranges of input parameters shown in Table 1 (e.g., cement content from 107 to 439 kg/m^3 and designed porosity from 17% to 35%) were determined based on the standard engineering specifications for planted concrete. These ranges cover the transition from low-strength/high-permeability to high-strength/low-permeability configurations, representing the typical operational window for ecological slope protection and urban greening projects. By capturing these specific intervals, the model is trained to be applicable to most real-world engineering scenarios.

The selection of the ten input variables is grounded in the fundamental physical and chemical mechanisms of concrete science, ensuring the ML models capture realistic material behavior rather than purely statistical correlations:

(1) Chemical Mechanism (pH Control): Cement content is the primary source of alkalinity due to the release of $Ca(OH)_2$ during hydration. The inclusion of supplementary cementitious materials (SCMs) such as fly ash, silica fume, and mineral powder is essential

because they consume $Ca(OH)_2$ through pozzolanic reactions, which is the key mechanism for neutralizing planted concrete to support vegetation growth [49].

(2) Physical Mechanism (Pore Structure): Planted concrete is characterized by its skeleton–pore structure. The coarse aggregate content and its particle size distribution (minimum and maximum sizes) define the geometric arrangement of the granular skeleton. These factors, combined with the designed porosity, directly determine the connectivity and distribution of the macro-pores (effective porosity), which governs the permeability and root penetration capacity [50].

(3) Mechanical Mechanism (Compressive Strength): The mechanical performance is dictated by the strength of the cement paste and the bond quality at the interfacial transition zone (ITZ). The water content and water-reducing agent dosage are critical in determining the effective water–binder ratio and the thickness of the paste coating the aggregates. Furthermore, the filling effect of fine SCMs like silica fume enhances the matrix density, significantly contributing to the overall compressive strength [51].

A Pearson correlation coefficient of 1 indicates a perfect positive correlation, while −1 represents a perfect negative correlation, and 0 signifies no correlation [52]. Based on the statistical characteristics of each parameter in the dataset—including type, mean, maximum, minimum, and standard deviation, as listed in Table 1—the Pearson correlation coefficients for each pair of variables were calculated, as shown in Figure 1 [53].

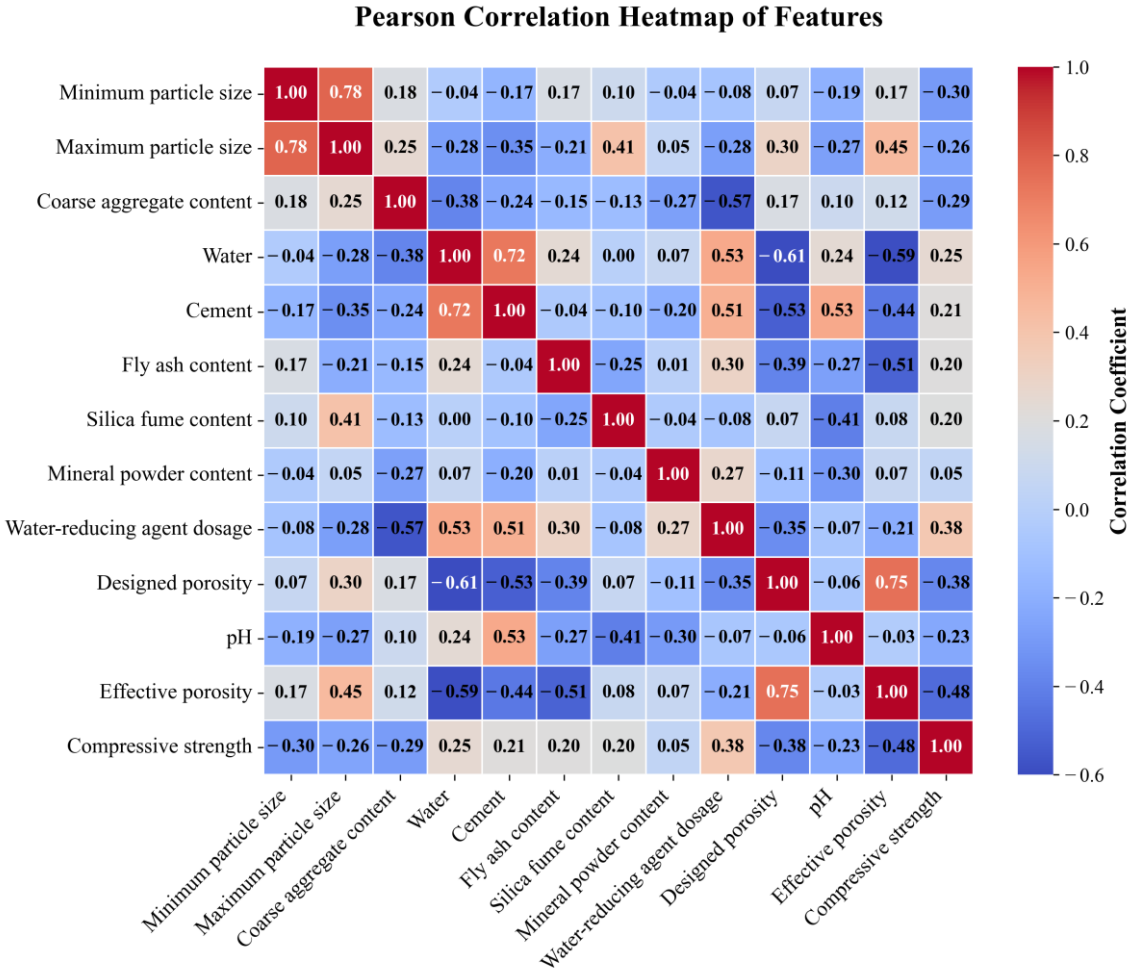


Figure 1. Thermal diagram of correlation of different characteristics of planted concrete.

The results reveal a significant negative correlation between effective porosity and minimum particle size (−0.30). Conversely, the strong positive correlation between maximum

and minimum particle sizes (0.78) suggests a grade control linkage effect. From an engineering optimization perspective, efforts can be directed toward regulating the pore structure (given the strong correlation between design porosity and effective porosity), adjusting the lower bound of aggregate grading (to prevent excessive accumulation of fine particles), and optimizing material composition. These insights contribute to a deeper understanding of how various parameters influence the overall performance of planted concrete [54].

2.4. Machine Learning Algorithms

MLP is a type of feedforward artificial neural network (ANN) that generates outputs based on a set of input features. Between the input and output layers, multiple hidden layers of neurons are interconnected through a directed graph, allowing the network to learn complex patterns [55]. MLPs are trained using the backpropagation algorithm, which iteratively adjusts the network's weights to minimize the prediction error. As a neural network architecture, an MLP consists of multiple layers where signals propagate in a unidirectional manner across the nodes. Each node, except for those in the input layer, applies a nonlinear activation function, enabling the model to capture intricate relationships within the data [56]. Due to the presence of multiple layers of neurons, MLP is often categorized as a deep learning approach. It is widely applied in supervised learning tasks, including machine translation, image recognition, and speech processing, as well as in computational science and parallel distributed processing studies. In MLP-based regression tasks, the algorithm initially selects predictors and evaluates collinearity using the Variance Inflation Factor (VIF) [57]. The VIF measures the extent to which the variance of a regression coefficient is inflated due to multicollinearity among predictors. To enhance model performance and stability, the algorithm iteratively eliminates variables with high VIF values, ensuring an optimal predictive framework [58].

GBR consists of stagewise additive ensemble models, where each iteration introduces a new weak learner trained to reduce the model's cumulative error. Unlike traditional ensemble approaches, previously added weak learners remain unchanged as new ones are incorporated [59]. Typically, GBR utilizes regression trees as weak learners, optimized through iterative, gradient-based algorithms such as Stochastic Gradient Descent (SGD) to minimize loss as each learner is added. In the first iteration, the algorithm develops an initial weak learner—a regression tree—to minimize the training error. In the subsequent iteration, a second tree is trained to reduce the error residuals from the first tree, iteratively continuing this process until an acceptable model quality, defined by a desired loss threshold, is achieved [60]. In this study, the GBR, a variant of GBMs designed for regression tasks, along with XGBoost—a highly scalable enhancement of GBR—are employed to address the high-performance computing (HPC) regression tasks at hand. It is worth noting that the XGBoost Regressor has gained significant recognition for achieving state-of-the-art (SOTA) performance in various ML and data mining competitions [61].

Random Forest Regression is an ensemble learning algorithm designed for regression tasks, which achieves robust performance by constructing multiple decision trees and integrating their predictions. In a Random Forest, each decision tree is independently trained on a randomly selected subset of the data, effectively reducing the risk of overfitting [62]. Additionally, at each node within a decision tree, only a randomly selected subset of features is considered for determining the optimal split, preventing any single feature from disproportionately influencing the model and enhancing its robustness. Using these randomized sub-samples, a decision tree is constructed based on algorithms such as CART (classification and regression trees), where the algorithm recursively selects the optimal splitting feature, partitioning the data to minimize impurity within each subset [63]. For new input samples, Random Forest Regression provides final predictions by averaging

or taking a weighted average of the predictions from all decision trees in the forest. This approach allows Random Forest Regression to balance bias and variance, offering high accuracy and robustness across diverse regression tasks [64,65].

2.5. Model Implementation and Hyperparameters

All data preprocessing and machine learning computations were implemented in the Python 3.9 environment. To ensure full replicability, the core libraries and optimized hyperparameters (determined via grid search) [66] are summarized in Table 2.

Table 2. Summary of optimized hyperparameters, software environment, and data preprocessing configurations.

Category	Item	Specifications
ML Models	RFR	n_estimators = 100, max_depth = None, random_state = 42, bootstrap = True
	XGBoost	learning_rate = 0.1, max_depth = 6, objective = 'reg:squarederror'
	GBR	n_estimators = 100, learning_rate = 0.1, loss = squared_error
	MLP	Hidden layers: (64, 128); Activation: ReLU; Optimizer: Adam
Environment	Programming	Python 3.9+
	Core Libraries	Scikit-learn, XGBoost, TensorFlow, Pandas, NumPy
Preprocessing [67]	Augmentation	Gaussian Noise (5% factor); Sample expansion: 100 to 2100
	Feature Selection	VIF < 10

2.6. Performance Assessment

The experimental dataset is commonly divided into three subsets, as is typical in the majority of previous studies: the training set, the validation set, and the testing set. The training set is used to optimize the model's parameters. The validation set serves to evaluate the model's performance and convergence during the training process, often guiding adjustments to hyperparameters [68]. Finally, the testing set is employed to assess the model's generalizability—its ability to accurately predict outcomes on unseen data after hyperparameter optimization [69]. The predictive performance of three ML models was evaluated using five key metrics: coefficient of determination (R^2), root mean squared error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), and the RMSE-to-observation's standard deviation ratio (RSR) [70].

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i' - y_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i' - y_i\right)^2}{n}} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left|y_i' - y_i\right| \quad (3)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left|\frac{y_i' - y_i}{y_i}\right| \times 100 \quad (4)$$

$$RSR = \frac{RMSE}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2}} \quad (5)$$

$$CPI = \frac{1}{N} \sum_{j=1}^{n} \frac{P_j - P_{min,j}}{P_{max,j} - P_{min,j}} \quad (6)$$

2.7. Shapley Additive Explanations (SHAP)

The Shapley Value, widely applied in cooperative game theory, is recognized for its ability to fairly and transparently distribute benefits among stakeholders, ensuring equitable outcomes. In the context of predictive modeling, the Shapley Value can be employed to allocate the total prediction inaccuracy—viewed as "income"—resulting from the collective contributions of individual prediction techniques [71]. This allocation is guided by the principle of "cooperative contribution," wherein the weight of each prediction method in the ensemble model reflects its relative significance. Consequently, the Shapley Value provides a robust framework for quantifying the contribution of each technique to the overall model performance [72].

3. Results and Discussion

3.1. Predictive Models for Planted Concrete pH

The performance results of the four presenting algorithms (GBR, MLP, XBR, and RFR) used to predict the concrete pH are given in Table 3, where the best trained model of each algorithm is highlighted. The outcomes are compared with those reported using the same dataset but a different approach.

Table 3. Results of different ML models on pH statistical metrics.

pH		R^2	RMSE	MAE	MAPE	RSR
MLP	Training	0.8920	0.2357	0.1743	1.6022%	0.3281
	Validation	0.8499	0.2874	0.2129	1.9789%	0.3873
	Testing	0.8198	0.2834	0.2073	1.9228%	0.4254
	All	0.8557	0.2517	0.1851	1.7068%	0.3524
GBR	Training	0.9233	0.1987	0.1438	1.3208%	0.2769
	Validation	0.8785	0.2586	0.1910	1.7698%	0.3485
	Testing	0.8432	0.2637	0.1866	1.7189%	0.3959
	All	0.9056	0.2193	0.1573	1.4478%	0.3072
XGBoost	Training	0.9995	0.0160	0.0107	0.0985%	0.0223
	Validation	0.9300	0.1964	0.1199	1.1139%	0.2646
	Testing	0.8863	0.2246	0.1219	1.1183%	0.3372
	All	0.9735	0.1163	0.0438	0.4038%	0.1629
RFR	Training	0.9929	0.0602	0.0321	0.2948%	0.0840
	Validation	0.9629	0.1428	0.0795	0.7340%	0.1924
	Testing	0.9334	0.1719	0.0864	0.7938%	0.2580
	All	0.9803	0.1002	0.0474	0.4355%	0.1403

Table 3 indicates that the performance of the three ensemble algorithms (GBR, XGBoost, and RFR) surpasses that of the single algorithm (MLP). The correlation coefficients (R^2) for the MLP on the training, validation, and test sets are 0.8920, 0.8499, and 0.8198, respectively. Compared to the validation and test sets, the training set exhibits a higher R^2 value and lower RMSE and MAE values, suggesting that the MLP performs well on the training set. Among the three ensemble models, RFR, based on the Bagging approach, outperforms GBR and XGBoost, which are based on the Boosting approach. The R^2 values of the training sets for GBR, XGBoost, and RFR are 0.9233, 0.9995, and 0.9929, respectively. While XGBoost demonstrates superior training performance in terms of R^2, it exhibits a significant decline in the validation and test sets, indicating pronounced overfitting. Furthermore, the R^2 values for RFR on the validation and test sets are 0.9629 and 0.9334, respectively, which are higher than those of GBR and XGBoost. Therefore, RFR demonstrates the best overall performance among the evaluated models.

Figure 2 presents a comparison of predicted versus actual values based on the selected four ML algorithms. The scatter plot of the single algorithm (MLP) shows numerous data points with substantial deviations from the central line, indicating suboptimal training performance compared to the ensemble models (GBR, XGBoost, and RFR). The XGBoost model demonstrates nearly perfect alignment between predicted and actual values on the training set. However, a pronounced dispersion of data points is observed on the validation and test sets, indicating significant overfitting. Despite its excellent training performance, the XGBoost model's reliability for practical applications is limited. In contrast, the scatter plot of the RFR model shows a high degree of agreement between predicted and actual values across the training, validation, and test sets, with tightly clustered data points. This highlights RFR as the most robust and generalizable model among the four algorithms.

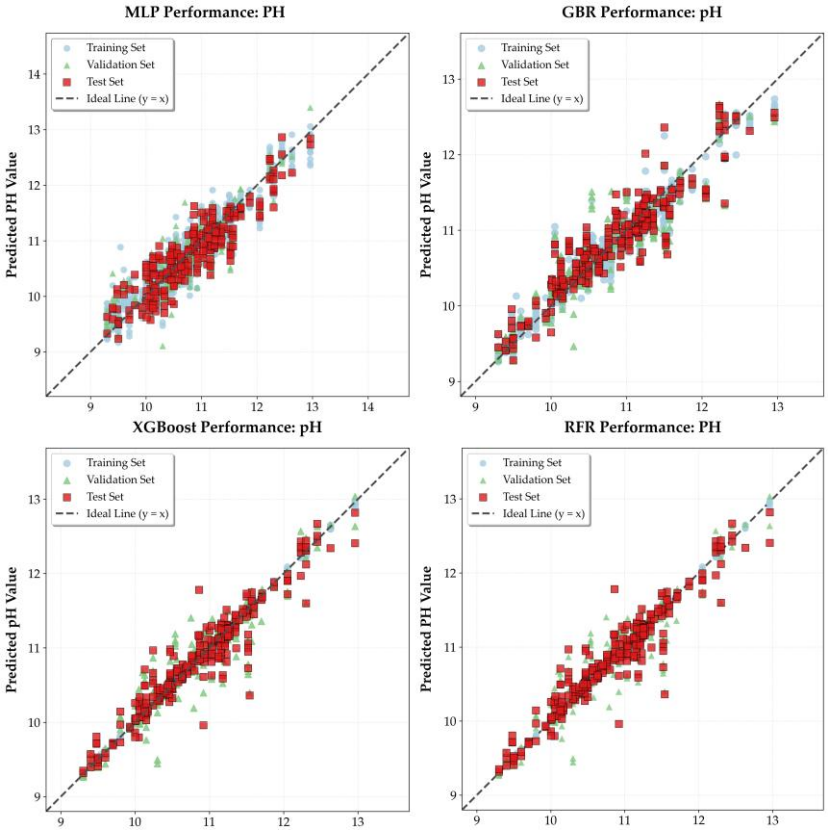


Figure 2. Correlation between actual and predicted pH of planted concrete for MLP, GBR, XGBoost, and RFR models.

In addition to analyzing the three subsets of the original database, it is also essential to perform statistical analysis for each data sample. The frequency distribution histogram of the percentage errors for various intervals across the entire database is shown in Figure 3. Evidently, an algorithm achieves higher fitting accuracy when the percentage errors of all samples are concentrated near zero. Although the XGBoost algorithm exhibits the highest frequency distribution of samples within the −5% to 5% error range, previous findings indicate that, while XGBoost achieves a high (R^2) on the training set, its performance declines significantly on the validation and test sets. Furthermore, the scatter plot reveals nearly perfect alignment between predicted and actual values on the training set but significant dispersion on the validation and test sets, highlighting pronounced overfitting. In contrast, the RFR method demonstrates more stable predictive performance, with superior generalization ability across all subsets.

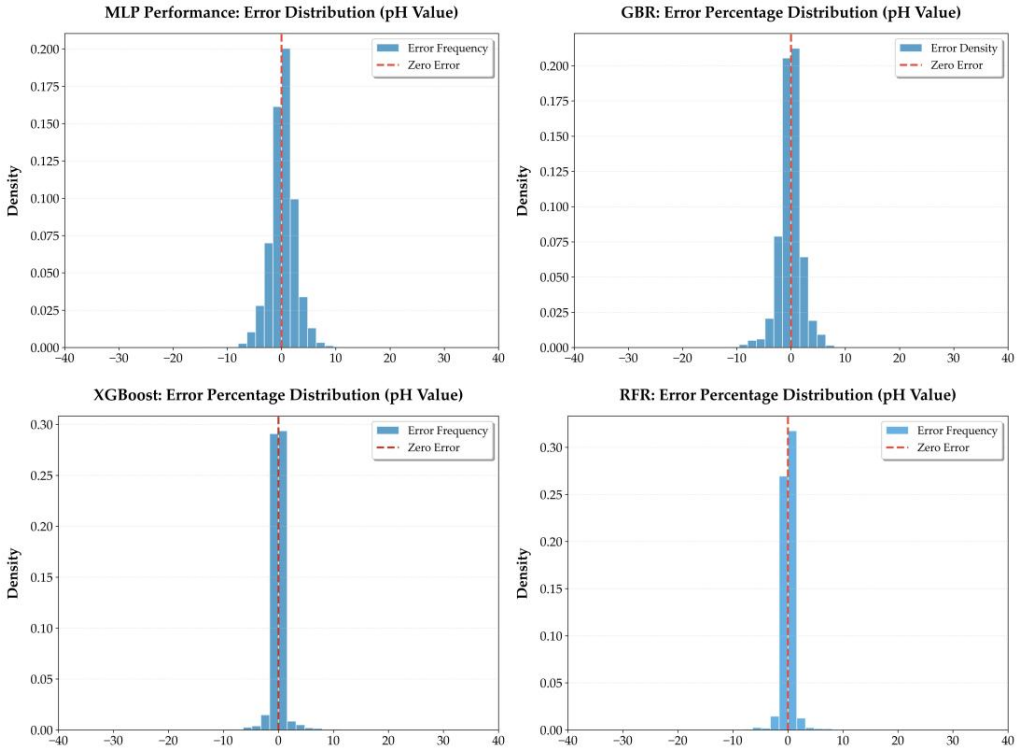


Figure 3. PH error percentage distribution histogram for MLP, GBR, XGBoost, and RFR models.

3.2. Predictive Models for Planted Concrete Effective Porosity

From the point of view of the porosity prediction performance metrics in Table 4, different models show varying results during training, validation, and testing phases. The MLP model demonstrates stable performance with R^2 values of 0.8293 and 0.8276 during the training and validation phases, respectively. However, its R^2 value drops to 0.7937 in the testing phase, indicating a slight lack of generalization ability. The Random Forest Regression (RFR) model achieves an exceptionally high R^2 value of 0.9965 in the training phase, and its R^2 values in the validation and testing phases are 0.9793 and 0.9747, respectively, showing strong fitting ability and good generalization performance. The GBR model performs excellently overall, with an R^2 value of 0.9655 during the training phase, and R^2 values above 0.93 in both the validation and testing phases, alongside low RMSE and MAE values, indicating high precision and stability in porosity prediction. The Extreme Gradient Boosting model achieves an R^2 value close to 1 (0.9991) during training, with R^2 values of 0.9627 and 0.9650 in the validation and testing phases, respectively. Its RMSE and MAE values are the lowest among all models, demonstrating extremely high predictive accuracy and strong generalization ability. In summary, the XGBoost and GBR models exhibit the best performance in porosity prediction.

Table 4. Results of different ML models on statistical metrics of effective porosity.

Effective Porosity		R^2	RMSE	MAE	MAPE	RSR
MLP	Training	0.8293	0.0164	0.0131	5.4090%	0.3649
	Validation	0.8276	0.0170	0.0134	5.4267%	0.3775
	Testing	0.7937	0.0182	0.0147	6.2250%	0.3969
	All	0.8240	0.0168	0.0134	5.5340%	0.3717

Table 4. *Cont.*

Effective Porosity		R^2	RMSE	MAE	MAPE	RSR
GBR	Training	0.9655	0.0074	0.0051	2.1325%	0.1858
	Validation	0.9379	0.0102	0.0071	2.9404%	0.2491
	Testing	0.9391	0.0099	0.0067	2.9029%	0.2469
	All	0.9572	0.0082	0.0056	2.3693%	0.2068
XGBoost	Training	0.9991	0.0012	0.0008	0.3331%	0.0294
	Validation	0.9627	0.0079	0.0050	2.0766%	0.1931
	Testing	0.9650	0.0075	0.0051	2.2344%	0.1872
	All	0.9883	0.0043	0.0021	0.8798%	0.1084
RFR	Training	0.9965	0.0024	0.0011	0.5044%	0.0595
	Validation	0.9793	0.0059	0.0029	1.2512%	0.1440
	Testing	0.9747	0.0064	0.0028	1.3385%	0.1589
	All	0.9905	0.0039	0.0016	0.7415%	0.0975

The comparative visual analysis of scatter plots depicting predicted versus actual porosity values across the MLP, GBR, XGBoost, and RFR models in Figure 4 revealed distinct differences in predictive performance. Among these, the scatter plot generated by the RFR model exhibited the closest alignment with the ideal 1:1 line, characterized by minimal dispersion and an absence of systematic bias across the porosity range. In contrast, the MLP model demonstrated substantial scatter and multiple outliers, particularly at higher porosity values, while the GBR and XGBoost models exhibited moderate dispersion, especially within the intermediate porosity range. The superior data clustering observed in the RFR model, coupled with its uniform distribution and the absence of localized overestimation or underestimation patterns, underscores its robustness in predicting the planted concrete effective porosity. These findings collectively position RFR as a reliable model for porosity prediction.

The visual analysis of frequency distribution histograms illustrating the percentage prediction errors for the MLP, GBR, XGR, and RFR models Figure 5 reveals distinct error patterns. The RFR model demonstrates a sharply peaked histogram, centered around 0% error, with over 75% of predictions falling within a narrow error range (±5%), indicating outstanding precision and consistency. In contrast, the MLP model exhibits a broader distribution, with significant error spread (±15%) and a secondary peak at higher errors (>10%), suggesting the presence of systematic inaccuracies. The GBR and XGR models display intermediate performance, with histograms exhibiting moderate dispersion (±10% error) and asymmetric tails indicative of overestimation. These visual patterns—tight clustering, minimal spread, and the absence of bias—collectively highlight the superior reliability of the RFR model in minimizing planted concrete effective porosity prediction errors.

3.3. Predictive Models for Planted Concrete Compressive Strength

A comprehensive evaluation of predictive performance across training, validation, and testing datasets, as assessed by R^2, RMSE, MAE, MAPE, and RSR metrics in Table 5, reveals that the RFR model consistently outperforms the MLP, GBR, and XGBoost models. RFR achieves the highest validation and testing R^2 values (0.9452 and 0.9439, respectively), accompanied by the lowest RMSE (0.4628–0.4690), MAE (0.2460–0.2492), MAPE (2.54–2.56%), and RSR (0.2341–0.2367), demonstrating superior accuracy and robustness. Notably, RFR maintains balanced performance across all datasets, with no evidence of overfitting, in contrast to XGBoost, which displays exceptional training metrics (R^2 = 0.9995) but suffers from degraded generalization in both validation and testing phases. These findings conclusively position RFR as the most reliable model for precise and consistent predictions of planted concrete compressive strength.

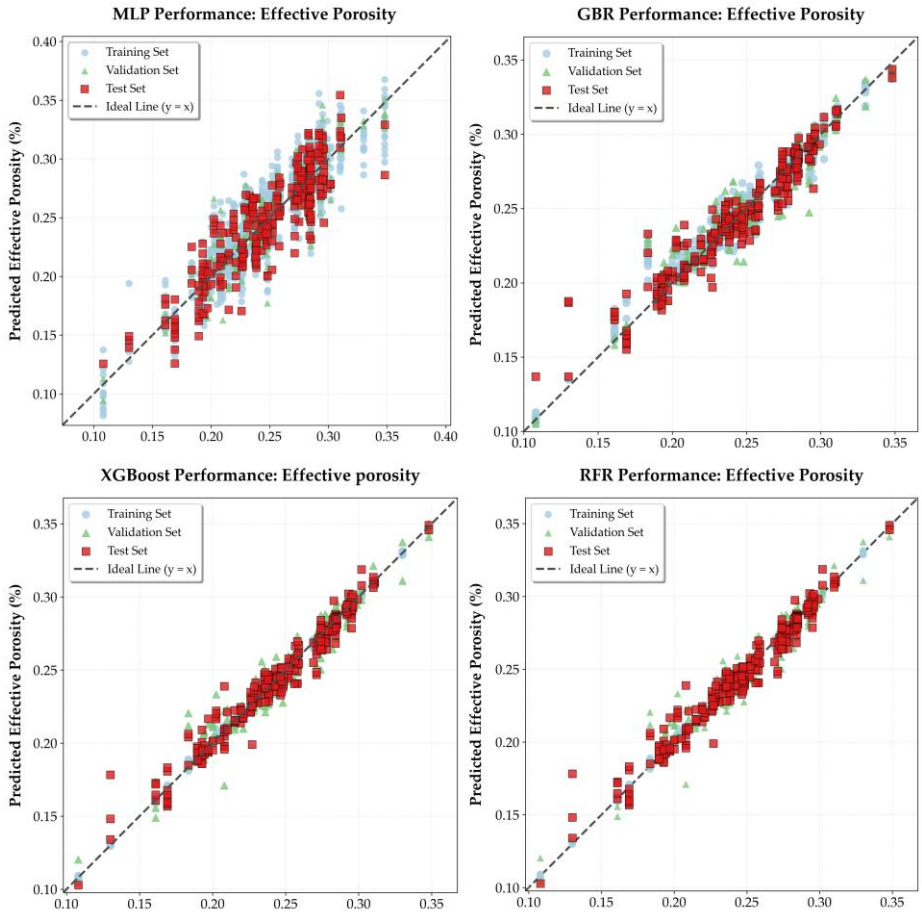


Figure 4. Correlation between actual and predicted effective porosity of planted concrete for MLP, GBR, XGBoost, and RFR models.

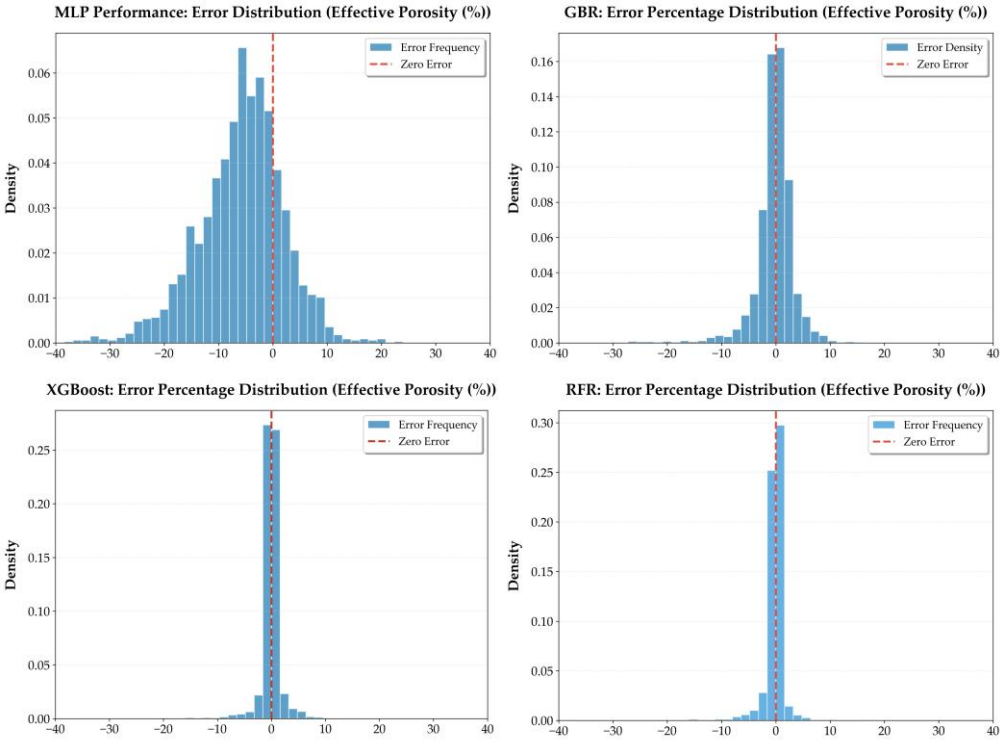


Figure 5. Effective porosity error percentage distribution histogram for MLP, GBR, XGBoost, and RFR models.

Table 5. Results of different ML models on intensity statistics metrics.

Compressive Strength		R^2	RMSE	MAE	MAPE	RSR
MLP	Training	0.9124	0.5832	0.4197	4.5112%	0.2912
	Validation	0.8757	0.6968	0.5031	5.4839%	0.3488
	Testing	0.8953	0.6413	0.4654	4.9115%	0.3219
	All	0.9043	0.6104	0.4391	4.7172%	0.3053
GBR	Training	0.9063	0.6030	0.4279	4.5576%	0.3061
	Validation	0.8608	0.7375	0.5235	5.5863%	0.3731
	Testing	0.8619	0.7363	0.5304	5.5659%	0.3716
	All	0.8927	0.6460	0.4576	4.8632%	0.3275
XGBoost	Training	0.9995	0.0437	0.0300	0.3264%	0.0222
	Validation	0.9187	0.5635	0.3464	3.6621%	0.2851
	Testing	0.9107	0.5923	0.3855	4.0102%	0.2989
	All	0.9739	0.3187	0.1308	1.3793%	0.1615
RFR	Training	0.9905	0.1913	0.0985	1.0244%	0.0971
	Validation	0.9452	0.4628	0.2460	2.5630%	0.2341
	Testing	0.9439	0.4690	0.2492	2.5426%	0.2367
	All	0.9767	0.3012	0.1433	1.4829%	0.1527

From Figures 6 and 7, the scatter distribution of the MLP model is more dispersed, indicating greater volatility and lower stability. In comparison, the GBR model exhibits some tendency towards aggregation, but its overall dispersion remains relatively high. While the XGBoost (Extreme Gradient Boosted Regression) model shows a higher degree of concentration and consistency, the scatter distribution of the RFR (Random Forest Regression) model is notably more centralized, with its prediction results closely aligning with the actual values, suggesting superior performance. A comprehensive analysis of the histogram of frequency distribution of the percentage of sample error shows that, although the XGBoost model demonstrates an excellent fit on the training set, it is prone to overfitting, as previously noted, which may hinder its ability to generalize to unseen data. In contrast, the scatter distribution of the RFR model is highly concentrated and consistent, reflecting strong agreement with actual values and greater stability across different datasets. These findings indicate that the RFR model outperforms others in terms of reliability and generalizability. Based on the preceding analysis, the MLP and XGBoost models demonstrated suboptimal predictive performance in terms of pH, effective porosity, and compressive strength of planted concrete, while the XGBoost and RFR models exhibited superior predictive accuracy. In the prediction of pH and compressive strength, the XGBoost model showed excellent training indices; however, a noticeable decline in generalization ability and the occurrence of overfitting were observed during both the validation and testing phases. These findings suggest that the RFR model delivers more robust prediction performance for both pH and compressive strength of planted concrete. In terms of effective porosity prediction, the RFR model demonstrated better predictive capacity, as evidenced by key performance metrics, scatter plots, and frequency distribution histograms of sample error percentages. Therefore, the RFR model emerges as the most stable and reliable model for simulating the effects of pH, effective porosity, and compressive strength in planted concrete.

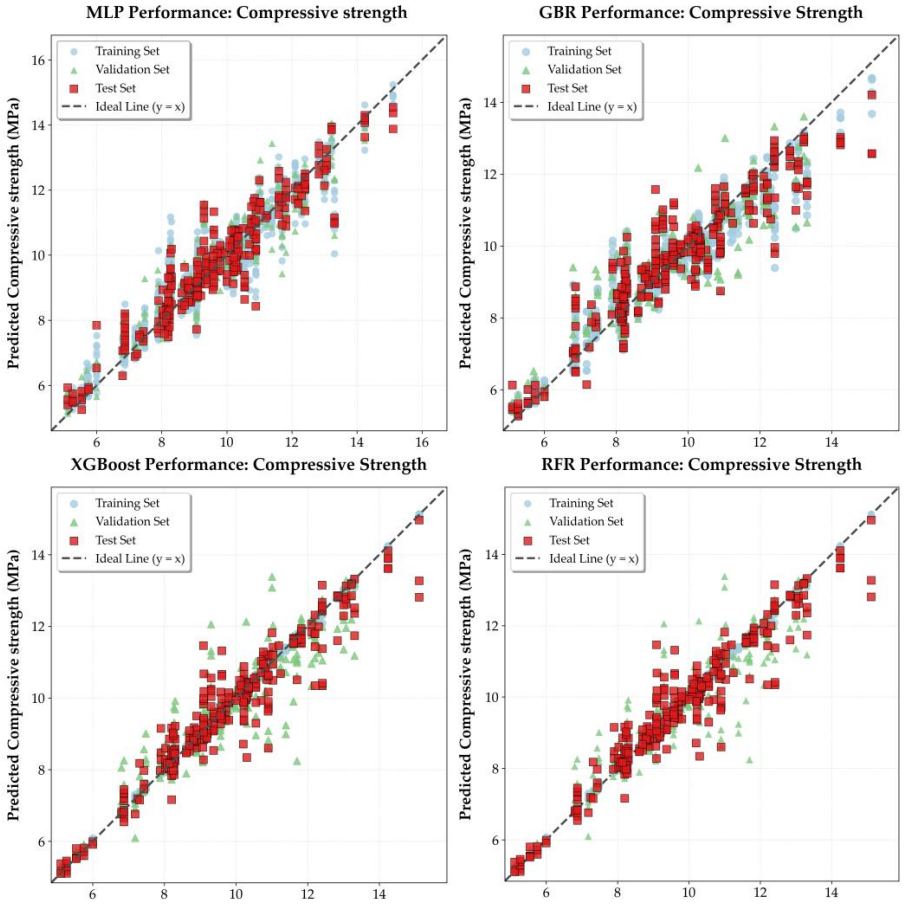


Figure 6. Correlation between actual and predicted compressive strength of planted concrete for MLP, GBR, XGBoost, and RFR models.

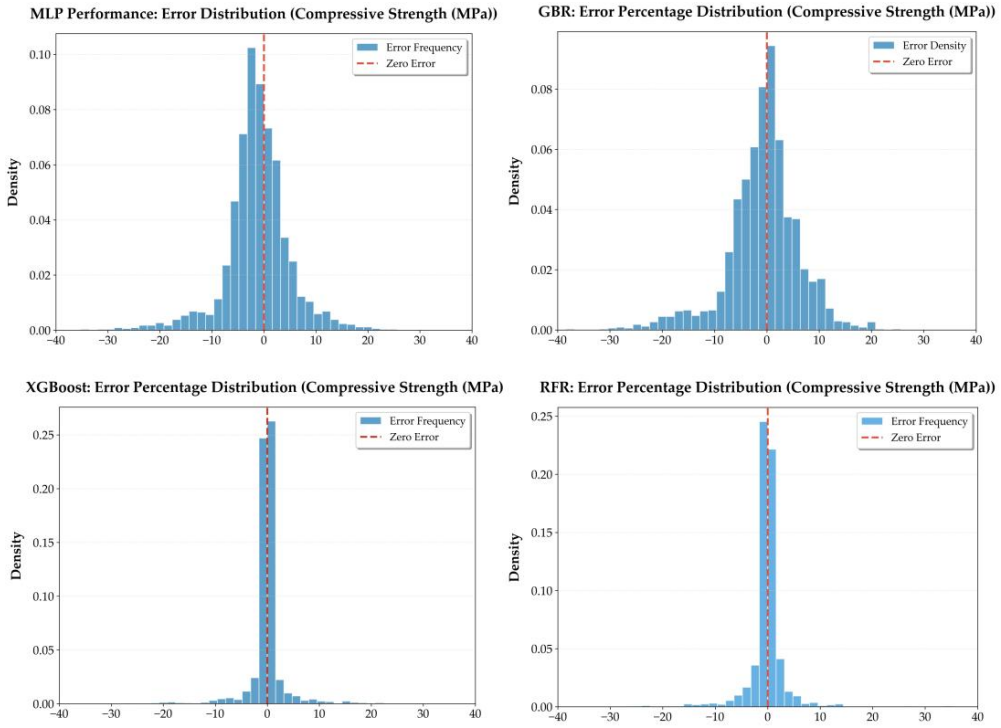


Figure 7. Compressive strength error percentage distribution histogram for MLP, GBR, XGBoost, and RFR models.

3.4. SHAP Global Explanation: Swarm Diagrams

Figure 8 presents a swarm plot of the Shapley Values for the ten input features. Each row corresponds to a specific feature, with the horizontal axis representing the Shapley Values, while feature importance decreases from top to bottom. Each point in the plot represents an individual sample; the redder the color, the smaller the Shapley Value will be.

Figure 8 presents the SHAP distribution results for three key properties. For the pore fluid pH, it is evident that cement dosing is the primary factor influencing the pore fluid pH, with significantly higher Shapley Values than the other features, and the high cement content (red dots) is concentrated in the region of positive Shapley Values (right), indicating that an increase in the cement content will significantly enhance the model-predicted pH values. This phenomenon is consistent with the chemical mechanism of cement hydration to produce $Ca(OH)_2$, which leads to enhanced alkalinity of the pore fluid (higher pH). Fly ash content and silica fume content follow closely as the second and third most important features, respectively: the Shapley Values of high fly ash content (red dots) are skewed towards the negative region (left), implying that it consumes $Ca(OH)_2$ through the volcanic ash reaction to neutralize alkaline substances to lower the pH value, while the distribution of the Shapley Values of high silica fume content (red dots) further reinforces this trend, with a nanoscale active ingredient that can accelerate the alkaline depletion to further reduce the pH value. It is worth noting that the Shapley Values of designed porosity have a weaker effect, but its high porosity (red dots) is slightly skewed towards the negative region, reflecting the potential dilution effect of increased water permeability on alkaline migration. In contrast, coarse aggregate content had low Shapley Value contributions, suggesting that their direct effects on pH values were limited. Lv Qifeng et al. [73] concluded that the cement type had the most significant effect on the pH of the planted concrete, which is similar to the findings of this paper. In summary, the pH regulation of planted concrete needs to be centered on reducing cement mixing, suppressing alkalinity by increasing the ratio of fly ash to silica fume, and balancing water permeability and plant compatibility by combining the pore design, and the trend of the pH value with the material proportion can be verified experimentally in the future.

Regarding the effective porosity of planted concrete shown in Figure 8, the SHAP values exhibit a broad distribution, with high eigenvalues (red dots) predominantly concentrated in the positive SHAP interval. This indicates a strong positive correlation between design porosity and the effective porosity of planted concrete. The primary reason for this relationship is that higher design porosity corresponds to a reduction in the cementitious slurry between aggregates, leading to the formation of more open pores. Furthermore, increased pore connectivity facilitates water infiltration and promotes plant root growth. Notably, the Shapley Values for the maximum coarse aggregate particle size show that high-value points (representing larger particle sizes) are concentrated on the right side of the plot (positive Shapley Value region), suggesting that larger aggregate particle sizes contribute to enhanced effective porosity. This phenomenon can be attributed to the formation of a tightly packed aggregate skeleton, where natural gaps between particles are not fully filled by cementitious slurry, resulting in unfilled voids that become interconnected pores. The simultaneous presence of red points for both large aggregate particle size and high design porosity in the upper-ranked features highlights the synergistic effect of these two factors in enhancing effective porosity. In addition, the results of the effective porosity study in this paper are similar to those of Peng Bo et al. [74]. Both of them concluded that the design porosity is the main influencing factor.

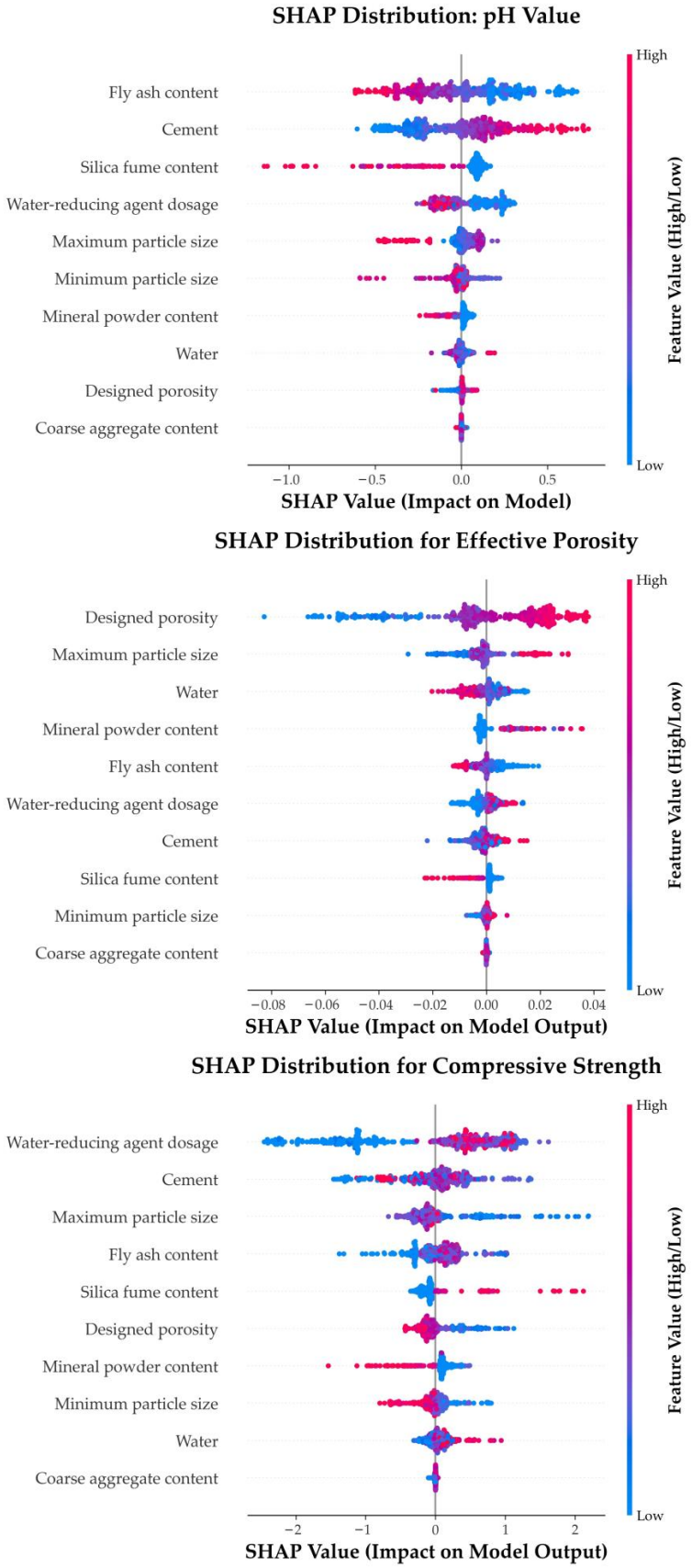


Figure 8. Summary of the SHAP analysis showing the influence of ten parameters for MLP, GBR, XGBoost, and RFR models.

For the compressive strength analysis in Figure 8, the distribution illustrates that the red high-value points (indicating high admixture levels) for cement are concentrated in the positive Shapley Value range, suggesting that an increase in cement admixture enhances the compressive strength of planted concrete. This can be attributed to cement's role as the core material in the hydration reaction, where higher cement content directly promotes the production of C-S-H gel, compressively strengthening the concrete matrix and increasing its density, thereby improving compressive strength. Additionally, the red high-value points for the water-reducing agent are also distributed within the positive Shapley Value region, indicating a positive effect on compressive strength. The water-reducing agent enhances slurry fluidity, ensuring better encapsulation of aggregates, which reduces the defects in the interfacial transition zone (ITZ), ultimately increasing compressive strength. Furthermore, the synergistic interaction between water reducers and cement amplifies hydration efficiency by reducing the water-to-binder (W/B) ratio, further contributing to compressive strength enhancement. Wanli Xing's [75] study came to a similar conclusion.

3.5. Model Generalizability and External Validation

To evaluate the model's practical robustness, an external validation was conducted using four independent experimental groups from Shen et al. [76], focusing on recycled aggregate planted concrete. As shown in Table 6, the RFR model demonstrated exceptional accuracy in predicting effective porosity, with the predicted value matching the experimental value exactly (0.240 vs. 0.240) for Case 1.

Table 6. Comparison of predicted and experimental performance for recycled aggregate planted concrete.

ID	Size	Exp. Strength (MPa)	Pred. Strength (MPa)	Exp. Porosity	Pred. Porosity
1	10–15 mm,	6.80	8.41	0.240	0.240
2	10–15 mm,	7.20	8.23	0.300	0.242
3	15–25 mm,	3.40	7.30	0.390	0.256
4	15–25 mm,	3.00	7.30	0.400	0.260

Regarding compressive strength, although a numerical offset was observed (predicted values being higher than experimental ones), the model successfully captured the correct descending trend as the aggregate-to-cement ratio increased. This numerical gap is scientifically consistent with the fact that the validation study utilized 100% recycled aggregates, which inherently yield significantly lower strength than the natural aggregates dominant in our training dataset. The model's ability to track physical trends and accurately anchor structural parameters (porosity) confirms its reliability for real-world engineering applications.

4. Conclusions

The objective of this study is to leverage ML techniques (MLP, GBR, XGBoost, and RFR) to accurately predict the pH, effective porosity, and compressive strength of planted concrete based on the provided input parameters. This study also aims to identify the model with the best predictive performance and, ultimately, utilize the interpretable SHAP algorithm to uncover the influence of variables on the prediction results. The performance of each model was thoroughly evaluated, and the following observations were made:

(1) A 5% Gaussian noise-based data enhancement method was used to expand the dataset and improve the prediction accuracy of the model.
(2) The ensemble models GBR, XGBoost, and RFR demonstrated superior predictive accuracy compared to the individual model MLP. A comprehensive evaluation using

ten statistical metrics revealed that RFR was the optimal model, with all test sets achieving an R^2 value greater than 0.93, making it the most accurate model of the four.

(3) SHAP analysis identified cement content as the primary driver for pH. For effective porosity, designed porosity and maximum aggregate size were the most significant factors. In terms of compressive strength, a strong synergy between the water reducer and cement was quantified: SHAP interaction plots indicate that increasing the water-reducer dosage yields a more pronounced strength enhancement at higher cement contents, as it optimizes paste workability and ensures a uniform mortar coating around aggregate particles.

(4) External validation using four groups of independent experimental data from Shen et al. (2025) [76] confirms the model's robust generalizability. The framework achieved high precision in structural prediction and successfully tracked mechanical trends across different material domains, effectively addressing concerns regarding data augmentation biases.

Author Contributions: X.D.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing—original draft, Visualization, Project administration. M.Z. and R.Z.: participated in Validation, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Open Research Fund Program of State Key Laboratory of Water Disaster Prevention(Grant No. 2024490411, 2023490211); National College Students' Innovation and Entrepreneurship Training Program (202310363003); The Key project of University Natural Science Research Project of Anhui Province (KJ2021A0505, 2024AH050111).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used for machine learning modeling in this study was extracted and compiled from relevant published literature. All literature providing the original data has been cited to support the findings. The compiled integrated dataset cannot be publicly shared due to copyright restrictions. However, the dataset can be reproduced by any researcher from the cited literature following the data collection and processing methods described in this paper. Upon reasonable request, the authors can provide detailed procedures for data extraction and processing.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rahayu, W.; Ramadhan, R.I.; Adinegara, A.W.; Adiguna, G.A.; Hamdany, A.H.; Wijaya, M.; Prakoso, W.A.; Sagitaningrum, F.H.; Satyanaga, A. Effect of slope protection using concrete waste on slope stability during rainfall. *Results Eng.* **2024**, *24*, 103244. [CrossRef]
2. Tang, W.; Mohseni, E.; Wang, Z. Development of vegetation concrete technology for slope protection and greening. *Constr. Build. Mater.* **2018**, *179*, 605–613. [CrossRef]
3. Ma, R.; Ma, D.; Long, E. Experimental study on the dynamic thermal performance of V-Ti black ceramic solar collector under multiple factors. *Sol. Energy* **2020**, *201*, 615–620. [CrossRef]
4. Kim, T.M.; Lee, S.-J.; Chang, D.-J.; Koo, J.; Kim, T.; Yoon, K.-H.; Choi, I.-Y. DynamiChain: Development of Medical Blockchain Ecosystem Based on Dynamic Consent System. *Appl. Sci.* **2021**, *11*, 1612. [CrossRef]
5. Salehi, P.; Dabbagh, H.; Ashengroph, M. Effects of microbial strains on the mechanical and durability properties of lightweight concrete reinforced with polypropylene fiber. *Constr. Build. Mater.* **2022**, *322*, 126519. [CrossRef]
6. Recking, A.; Piton, G.; Montabonnet, L.; Posi, S.; Evette, A. Design of fascines for riverbank protection in Alpine rivers: Insight from flume experiments. *Ecol. Eng.* **2019**, *138*, 323–333, Correction in *Ecol. Eng.* **2021**, *167*, 106254. https://doi.org/10.1016/j.ecoleng.2021.106254. [CrossRef]
7. McGinnis, M.J.; Davis, M.; de la Rosa, A.; Weldon, B.D.; Kurama, Y.C. Strength and stiffness of concrete with recycled concrete aggregates. *Constr. Build. Mater.* **2017**, *154*, 258–269. [CrossRef]

8. Zhuang, P.; Yan, X.; Wang, X.; Liu, J. Study on the Performance Optimization of Plant-Growing Ecological Concrete. *Sustainability* **2024**, *16*, 4575. [CrossRef]
9. Chen, J.; Du, W.; Zhao, G.; Shi, M.; Xue, B. Effect of Aggregate Size and Water/Cement on Compressive Strength and Physiological Performance of Planting Concrete. *Materials* **2022**, *15*, 6685. [CrossRef] [PubMed]
10. Yang, C.; Fang, Y.; Wei, P.; Zeng, H.; Shen, C.; Yang, J.; Zhang, H.; Chu, H. Optimization study on alkali-reduction techniques and vegetation performance of planting concrete incorporating phosphogypsum. *Constr. Build. Mater.* **2025**, *482*, 141663. [CrossRef]
11. Ganapathy, G.P.; Alagu, A.; Ramachandran, S.; Panneerselvam, A.S.; Arokiaraj, G.G.V.; Panneerselvam, M.; Panneerselvam, B.; Sivakumar, V.; Bidorn, B. Effects of fly ash and silica fume on alkalinity, strength and planting characteristics of vegetation porous concrete. *J. Mater. Res. Technol.* **2023**, *24*, 5347–5360. [CrossRef]
12. Kong, J.; Wang, Z.; Meng, X.; Zhao, Y.; Chen, M.; Quan, H. Study on Alkali Reduction Treatments and Plant Growth Properties of Planting Concrete. *Sustainability* **2022**, *14*, 12228. [CrossRef]
13. Cai, J.; Shen, C.; Ye, M.; Huang, S.; He, J.; Cui, D. Influencing Factors of Porosity and Strength of Plant-Growing Concrete. *Materials* **2023**, *17*, 31. [CrossRef] [PubMed]
14. Liu, P.; Xu, Y.; Liu, F.; Pan, J.; Fan, Z. Research on the performance of planted concrete based on orthogonal test methods. *Silic. Bull.* **2024**, *43*, 4398–4405. [CrossRef]
15. Sathiparan, N.; Heshan, W.S.; Rinduja, R.; Pratheeba, J.; Subramaniam, D.N. Response surface regression and machine learning models to predict the porosity and compressive strength of pervious concrete based on mix design parameters. *Road Mater. Pavement Des.* **2025**, *26*, 618–657. [CrossRef]
16. Liu, F.; Xu, J.; Tan, S.; Gong, A.; Li, H. Orthogonal Experiments and Neural Networks Analysis of Concrete Performance. *Water* **2022**, *14*, 2520. [CrossRef]
17. Cao, C. Prediction of concrete porosity using machine learning. *Results Eng.* **2023**, *17*, 100794. [CrossRef]
18. Mohtasham Moein, M.; Saradar, A.; Rahmati, K.; Ghasemzadeh Mousavinejad, S.H.; Bristow, J.; Aramali, V.; Karakouzian, M. Predictive models for concrete properties using machine learning and deep learning approaches: A review. *J. Build. Eng.* **2023**, *63*, 105444. [CrossRef]
19. Wu, Y.; Pieralisi, R.; Sandoval, F.G.B.; López-Carreño, R.D.; Pujadas, P. Optimizing pervious concrete with machine learning: Predicting permeability and compressive strength using artificial neural networks. *Constr. Build. Mater.* **2024**, *443*, 137619. [CrossRef]
20. Song, Y.; Zhao, J.; Ostrowski, K.A.; Javed, M.F.; Ahmad, A.; Khan, M.I.; Aslam, F.; Kinasz, R. Prediction of Compressive Strength of Fly-Ash-Based Concrete Using Ensemble and Non-Ensemble Supervised Machine-Learning Approaches. *Appl. Sci.* **2021**, *12*, 361. [CrossRef]
21. Asteris, P.G.; Skentou, A.D.; Bardhan, A.; Samui, P.; Pilakoutas, K. Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. *Cem. Concr. Res.* **2021**, *145*, 106449. [CrossRef]
22. Xie, H.-Z.; Li, L.G.; Ng, P.-L.; Liu, F. Effects of Solid Waste Reutilization on Performance of Pervious Concrete: A Review. *Sustainability* **2023**, *15*, 6105. [CrossRef]
23. Chou, J.-S.; Tsai, C.-F.; Pham, A.-D.; Lu, Y.-H. Machine learning in concrete strength simulations: Multi-nation data analytics. *Constr. Build. Mater.* **2014**, *73*, 771–780. [CrossRef]
24. Shemirani, A.B. Prediction of fracture toughness of concrete using the machine learning approach. *Theor. Appl. Fract. Mech.* **2024**, *134*, 104749. [CrossRef]
25. Liu, T.; Wang, Z.; Zeng, J.; Wang, J. Machine-learning-based models to predict shear transfer strength of concrete joints. *Eng. Struct.* **2021**, *249*, 113253. [CrossRef]
26. Wang, Z.; Liu, H.; Amin, M.N.; Khan, K.; Qadir, M.T.; Khan, S.A. Optimizing machine learning techniques and SHapley Additive exPlanations (SHAP) analysis for the compressive property of self-compacting concrete. *Mater. Today Commun.* **2024**, *39*, 108804. [CrossRef]
27. Wang, X.; Ke, Z.; Liu, W.; Zhang, P.; Cui, S.A.; Zhao, N.; He, W. Compressive Strength Prediction of Basalt Fiber Reinforced Concrete Based on Interpretive Machine Learning Using SHAP Analysis. *Iran. J. Sci. Technol. Trans. Civ. Eng.* **2024**, *49*, 2461–2480. [CrossRef]
28. Ly, H.-B. An interpretable machine learning-based measurement system for predicting the compressive strength of concrete at elevated temperatures. *Measurement* **2025**, *264*, 120273. [CrossRef]
29. Huang, Q.; Wang, W.-W.; Tang, Y.; Guo, K.; Zhou, C. Failure modes analysis of reinforced concrete beams under impact loads based on machine learning and SHAP approach. *Eng. Fail. Anal.* **2026**, *183*, 110281. [CrossRef]
30. Yadav, A.A.; Hirekhan, S.G.; Bhandari, P.S.; Bhagat, R.M.; Ranit, A.B.; Shelare, S.; Soudagar, M.E.M.; Sharma, S.; Raja, V.K.B.; Mahapatro, A.; et al. AI-driven sustainable concrete mix design: Hybrid deep Q-Learning and genetic algorithms-based multi-objective machine learning optimizations for high structural strength, low cost, and low carbon footprints. *Structures* **2025**, *82*, 110443. [CrossRef]

31. Ben Chaabene, W.; Flah, M.; Nehdi, M.L. Machine learning prediction of mechanical properties of concrete: Critical review. *Constr. Build. Mater.* **2020**, *260*, 119889. [CrossRef]
32. Salami, B.A.; Iqbal, M.; Abdulraheem, A.; Jalal, F.E.; Alimi, W.; Jamal, A.; Tafsirojjaman, T.; Liu, Y.; Bardhan, A. Estimating compressive strength of lightweight foamed concrete using neural, genetic and ensemble machine learning approaches. *Cem. Concr. Compos.* **2022**, *133*, 104721. [CrossRef]
33. Quan, Y.; Wang, F. Machine learning-based real-time tracking for concrete vibration. *Autom. Constr.* **2022**, *140*, 104343. [CrossRef]
34. Hoang, N.-D. Machine Learning-Based Estimation of the Compressive Strength of Self-Compacting Concrete: A Multi-Dataset Study. *Mathematics* **2022**, *10*, 3771. [CrossRef]
35. Sun, M. Research on the Application of Vegetative Eco-Concrete in Coastal Road Slopes. Master's Thesis, Zhejiang Ocean University, Zhoushan China, 2022. [CrossRef]
36. Huang, Q. Research on the Performance and Application Technology of Vegetated Eco-Concrete. Master's Thesis, Hebei University of Architecture, Zhangjiakou, China, 2020. [CrossRef]
37. Tan, S. Research on Alkali Reduction Technology of Vegetated Concrete and Plant Suitability. Master's Thesis, Guangzhou University, Guangzhou, China, 2020. [CrossRef]
38. Fang, J. Research on the Application of Vegetative Porous Concrete in Ecological Slope Protection. Master's Thesis, Chongqing Jiaotong University, Chongqing, China, 2016. [CrossRef]
39. Hribar-Lee, B. The influence of excipients on the viscosity of monoclonal antibody solutions. *J. Mol. Liq.* **2022**, *366*, 120349. [CrossRef] [PubMed]
40. Yufeng, Z.; Zheng, C.; Xiaoliang, Z.; Yuxuan, Z. Experimental study on the performance of vegetative eco-concrete based on orthogonal test method. *China Build. Mater. Technol.* **2022**, *31*, 22–25.
41. Li, S. Preparation of Low-Alkali Porous Concrete for Ecological Rehabilitation of Slopes and Research on Planting Technology. Master's Thesis, Chongqing Jiaotong University, Chongqing, China, 2024. [CrossRef]
42. Yuan, Y. Experimental Study on the Basic Properties and Ratio of Recycled Vegetated Concrete. 2023. Available online: https://kns.cnki.net/kcms2/article/abstract?v=weCMZZjTuL1VKOtNmzUG940OBmo3feIXorgTWm7wOSdBpbV0FpQQun9n3wkUB9vahAywWUfhvmkIHAxZF4hqMqRUgz37aKPtmEX0bzKoOirUYZK3gqpuqJBoMb91cBzUykECOviZaJcpezh7qvtYMFbWwgQjm8OVjRdpneJRS2Y4VkhN21VvDA==&uniplatform=NZKPT&language=CHS (accessed on 5 January 2026).
43. Wang, Z.-y.; Zhang, Q.-l.; Tao, Y.; Liu, J.; Qin, Y.-y.; Jin, X. Machine learning-based prediction of peak outflow of embankment dam break with data augmentation. *Phys. Fluids* **2025**, *37*, 086621. [CrossRef]
44. Rehman, S.U.; Riaz, R.D.; Usman, M.; Kim, I.-H. Augmented Data-Driven Approach towards 3D Printed Concrete Mix Prediction. *Appl. Sci.* **2024**, *14*, 7231. [CrossRef]
45. Kurhade, S.D.; Patankar, S. Predicting compressive strength of geopolymer concrete using machine learning models. *Innov. Infrastruct. Solut.* **2024**, *10*, 12. [CrossRef]
46. Li, Z.; Pei, T.; Ying, W.; Srubar, W.V.; Zhang, R.; Yoon, J.; Ye, H.; Dabo, I.; Radlińska, A. Can domain knowledge benefit machine learning for concrete property prediction? *J. Am. Ceram. Soc.* **2023**, *107*, 1582–1602. [CrossRef]
47. Jin, S.; Zhang, Y.; Yan, Y.; Xu, Z.; Li, A.; Wang, J.; Liu, P.; Li, L.; Wang, Z. Influence of aggregate characteristics on the plant growing environment of the planting concrete with SAC. *J. Clean. Prod.* **2024**, *445*, 141179. [CrossRef]
48. Liu, Y.; Wang, B.; Qian, Z.; Yu, J.; Shi, T.; Fan, Y.; Zhou, Y.; Ning, Y.; Zhou, X. State-of-the art on preparation, performance, and ecological applications of planting concrete. *Case Stud. Constr. Mater.* **2024**, *20*, e03131. [CrossRef]
49. Liu, C.; Xia, Y.; Chen, J.; Huang, K.; Wang, J.; Wang, C.; Huang, Z.; Wang, X.; Rao, C.; Shi, M. Research and Application Progress of Vegetation Porous Concrete. *Materials* **2023**, *16*, 7039. [CrossRef]
50. AlShareedah, O.; Nassiri, S. Pervious concrete mixture optimization, physical, and mechanical properties and pavement design: A review. *J. Clean. Prod.* **2021**, *288*, 125095. [CrossRef]
51. Ding, K.; Zeng, C. Study on the compressive strength, pore structure characteristics, and fractal dimension of the ecological porous concrete specimens based on ordinary Portland cement. *Constr. Build. Mater.* **2025**, *483*, 141795. [CrossRef]
52. Pallapothu, S.N.R.G.; Pancharathi, R.K.; Janib, R. Predicting concrete strength through packing density using machine learning models. *Eng. Appl. Artif. Intell.* **2023**, *126*, 107177. [CrossRef]
53. Quan Tran, V.; Quoc Dang, V.; Si Ho, L. Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach. *Constr. Build. Mater.* **2022**, *323*, 126578. [CrossRef]
54. Cao, Q.; Zhou, J.; Xu, W.; Yuan, X. Study on the Preparation and Properties of Vegetation Lightweight Porous Concrete. *Materials* **2024**, *17*, 251. [CrossRef]
55. Abellán-García, J. Four-layer perceptron approach for strength prediction of UHPC. *Constr. Build. Mater.* **2020**, *256*, 119465. [CrossRef]
56. Chou, J.-S.; Tsai, C.-F. Concrete compressive strength analysis using a combined classification and regression technique. *Autom. Constr.* **2012**, *24*, 52–60. [CrossRef]

57. Tsai, H.-C. Predicting strengths of concrete-type specimens using hybrid multilayer perceptrons with center-unified particle swarm optimization. *Expert Syst. Appl.* **2010**, *37*, 1104–1112. [CrossRef]
58. Bayram, S.; Ocal, M.E.; Laptali Oral, E.; Atis, C.D. Comparison of multi layer perceptron (MLP) and radial basis function (RBF) for construction cost estimation: The case of Turkey. *J. Civ. Eng. Manag.* **2016**, *22*, 480–490. [CrossRef]
59. Rathakrishnan, V.; Beddu, S.B.; Ahmed, A.N. Predicting compressive strength of high-performance concrete with high volume ground granulated blast-furnace slag replacement using boosting machine learning algorithms. *Sci. Rep.* **2022**, *12*, 9539. [CrossRef]
60. Kaloop, M.R.; Kumar, D.; Samui, P.; Hu, J.W.; Kim, D. Compressive strength prediction of high-performance concrete using gradient tree boosting machine. *Constr. Build. Mater.* **2020**, *264*, 120198. [CrossRef]
61. Xu, J.-G.; Chen, S.-Z.; Xu, W.-J.; Shen, Z.-S. Concrete-to-concrete interface shear strength prediction based on explainable extreme gradient boosting approach. *Constr. Build. Mater.* **2021**, *308*, 125088. [CrossRef]
62. Li, H.; Lin, J.; Lei, X.; Wei, T. Compressive strength prediction of basalt fiber reinforced concrete via random forest algorithm. *Mater. Today Commun.* **2022**, *30*, 103117. [CrossRef]
63. Iftikhar, B.; Alih, S.C.; Vafaei, M.; Elkotb, M.A.; Shutaywi, M.; Javed, M.F.; Deebani, W.; Khan, M.I.; Aslam, F. Predictive modeling of compressive strength of sustainable rice husk ash concrete: Ensemble learner optimization and comparison. *J. Clean. Prod.* **2022**, *348*, 131285. [CrossRef]
64. Zhang, D.; Chen, Y.; Zhang, C.; Xue, G.; Zhang, J.; Zhang, M.; Wang, L.; Li, N. Prediction of seismic acceleration response of precast segmental self-centering concrete filled steel tube single-span bridges based on machine learning method. *Eng. Struct.* **2023**, *279*, 115574. [CrossRef]
65. Yılmaz, Y.; Nayır, S. Machine learning based prediction of compressive and flexural strength of recycled plastic waste aggregate concrete. *Structures* **2024**, *69*, 107363. [CrossRef]
66. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. [CrossRef]
67. Bishop, C.M. Training with Noise is Equivalent to Tikhonov Regularization. *Neural Comput.* **1995**, *7*, 108–116. [CrossRef]
68. Yuan, X.; Tian, Y.; Ahmad, W.; Ahmad, A.; Usanova, K.I.; Mohamed, A.M.; Khallaf, R. Machine Learning Prediction Models to Evaluate the Strength of Recycled Aggregate Concrete. *Materials* **2022**, *15*, 2823. [CrossRef]
69. Chen, B.; Wang, L.; Feng, Z.; Liu, Y.; Wu, X.; Qin, Y.; Xia, L. Optimization of high-performance concrete mix ratio design using machine learning. *Eng. Appl. Artif. Intell.* **2023**, *122*, 106047. [CrossRef]
70. Kazemi, F.; Asgarkhani, N.; Shafighfard, T.; Jankowski, R.; Yoo, D.-Y. Machine-Learning Methods for Estimating Performance of Structural Concrete Members Reinforced with Fiber-Reinforced Polymers. *Arch. Comput. Methods Eng.* **2024**, *32*, 571–603. [CrossRef]
71. Abdulalim Alabdullah, A.; Iqbal, M.; Zahid, M.; Khan, K.; Nasir Amin, M.; Jalal, F.E. Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. *Constr. Build. Mater.* **2022**, *345*, 128296. [CrossRef]
72. Wu, Y.; Zhou, Y. Hybrid machine learning model and Shapley additive explanations for compressive strength of sustainable concrete. *Constr. Build. Mater.* **2022**, *330*, 127298. [CrossRef]
73. Lv, Q.; Chen, A.; Dai, P. Study on the Influencing Factors of pH of 3D Printed Vegetated Concrete. *Concr. Cem. Prod.* **2023**, 7–12. [CrossRef]
74. Bo, P.; Changbo, J.; Taishang, X.; Haowen, S. Experimental study on porosity of vegetated porous concrete. *China Water Transp.* **2013**, *13*, 281–283.
75. Xing, W. Design of concrete mix ratio and the influence of high water absorption resin and water reducer on strength and working performance. *Highl. Sci. Eng. Technol.* **2022**, *5*, 250–256. [CrossRef]
76. Shen, C.; Tang, L.; Wang, M.; Wang, X. Research on Physical Mechanics and Vegetation Characteristics of Vegetation Concrete with Different Particle sizes. *J. Build. Mater.* **2025**. Available online: https://link.cnki.net/urlid/31.1764.TU.20251209.1627.024 (accessed on 5 January 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.