Civil Engineering Beyond Limits 1 (2027) 11173
Civil Engineering Beyond Limits
journal home: www.acapublishing.com/journals/1/cebel
Research Article

Explainable Metaheuristic-Optimized Ensemble Machine Learning for Compressive Strength Prediction and Sustainable Mix Design of Recycled Powder Mortar

Sebghatullah Jueyendah[*1], Hedayatullah Hawadi[2], Khwaja Abdul Wahab Sediqi[3]
[1]Department of Civil Engineering, Sakarya University, Sakarya, 54050, Turkey
[2]Department of Geophysics, Faculty of Nano and Bio Science and Technology, Persian Gulf University, Bushehr, Iran
[3]Department of Civil Engineering, Faculty of Technology, Sakarya University of Applied Sciences, Sakarya, 54050, Turkey

Keywords

SHAP,
Recycled powder mortar,
Compressive strength,
Ensemble machine learning,
Metaheuristic optimization,
Sobol sensitivity analysis,
Sustainable construction,
Explainable artificial intelligence.

Abstract

Recycled powder mortar (RPM) represents a sustainable cementitious material whose compressive strength (CS) is governed by complex nonlinear interactions among material and mixture parameters. This study proposes an explainable, metaheuristic-optimized ensemble machine learning (ML) framework for accurate prediction and interpretation of CS of RPM. A dataset comprising 204 observations was employed, with recycled powder type (kind), particle size, water/binder (W/B), and mass replacement ratio (MRR) considered as input variables. Extra Trees (ET), Gradient Boosting (GB), and Histogram-Based Gradient Boosting (HGB) were integrated within bagging, voting, and stacking architectures, while hyperparameters were optimized using genetic algorithm (GA), particle swarm optimization (PSO), differential evolution (DE), and grey wolf optimizer (GWO). The dataset was partitioned into 80% training and 20% testing subsets, with model robustness further evaluated through shuffled 10-fold cross-validation. Predictive performance was quantified using R^2, RMSE, MAE, and MAPE. The GA-optimized ET–GB–HGB stacking model demonstrated superior predictive performance, achieving R^2 values of 0.949, 0.966, and 0.947, with corresponding RMSE values of 2.149, 1.931, and 2.129 MPa for the training, testing, and 10-fold cross validation datasets, respectively. SHAP and permutation importance consistently established the hierarchy MRR > particle size > kind > W/B, with MRR accounting for more than half of global SHAP importance. Sobol sensitivity and SHAP interaction analyses further confirmed MRR as the dominant factor and particle size–MRR as the strongest coupled interaction. Response-surface analysis additionally elucidated their nonlinear effects on CS. Overall, the proposed framework establishes a robust, interpretable, and data-driven methodology for accurate CS prediction and performance-informed design of sustainable RPM mixtures.

1. Introduction

The construction industry faces increasing pressure to mitigate environmental impacts while enhancing material efficiency [1]. Recycled powder mortar (RPM) has been developed as a sustainable cementitious material that valorizes fine construction and demolition waste, thereby reducing landfill disposal and dependence on conventional cement [2]. However, the mechanical performance of RPM is greatly affected by the mixture design parameters [3]. The compressive strength is a fundamental parameter for the structural performance and durability. Nevertheless, the experimental evaluation of this property is expensive and time consuming, especially for recycled materials with high variability [4]. Recycled powder (RP) is a fine particulate material created by crushing and processing construction and demolition waste such as concrete, mortar and brick [5]. RP is composed of fine, cementitious-rich fractions, which can be reused to reduce the disposal of low-value waste. The use of RP can partially replace traditional mortar constituents, which promotes waste valorization and material efficiency. However, variable physicochemical properties of RP may have a negative effect on mechanical performance [6].

The RPM strength is determined by the complex nonlinear interactions between RP type, particle size, W/B and MRR, which requires robust predictive approaches to capture the coupled effects of W/B and MRR [7]. In comparison with conventional models, the nonlinear interactions and uncertainties are often difficult to be captured, while the machine learning (ML) provides better prediction for cementitious materials [8-11]. Tree-based models are effective in modeling the nonlinear behavior while SHAP quantifies the effect of material parameters on CS [12-18]. ML is well suited for accurate and interpretable modeling of high-performance mixtures optimization [19-22]. In 1959, Arthur Samuel defined ML as the ability of systems to learn from experience without being explicitly programmed [23]. Artificial Intelligence (AI) models complex phenomena and predicts outcomes, while ML allows for pattern recognition and data-driven decision making [10, 24].

The following are recent developments in artificial intelligence (AI), ML, and deep learning (DL) that have improved the data-driven prediction and optimization of concrete and structural performance [25, 26]. Fei et al. [27] used ensemble ML for prediction of CS of RPM and optimization of mixture design. The CS prediction of waste glass powder mortar was investigated by Khan et al. [28] using experimental and ML techniques. Amin et al. [29] studied the flexural strength of waste glass powder mortar by experimental and ML methods. Yang et al. [30] developed AI-based models to predict mechanical properties of concrete containing ceramic waste powder, validating the potential of data-driven methods for sustainable concrete design. Bisciotti et al. [31] quantified attached mortar in recycled aggregates by DL and mineralogical modeling, which

*Corresponding Author: sebghatullahjueyendah@ogr.sakarya.edu.tr; sebghatullahjueyendah@gmail.com
Received 13 Aug 2026; Revised 23 Aug 2026; Accepted 28 Aug 2026
https://doi.org/10.36937/cebel.2027.11173

allows a more effective characterization of construction and demolition waste. Kellouche et al. [32] combined ML with metaheuristic optimization to predict and optimize CS of concrete with marble powder. Hasan et al. [33] developed an ML approach for predicting the CS of hybrid-fiber-reinforced recycled aggregate concrete, which is a strong tool for optimizing the mix design. Dahish and Alkharisi [34] used ML and response surface methodology to predict the properties of polypropylene fiber recycled aggregate concrete. Jueyendah et al.[35] demonstrated the efficiency of SVR with RBF kernel in cement mortar strength prediction. Jueyendah and Martins [36] developed hybrid SVR-RBF model for optimized design and performance assessment of welded structures. Jueyendah et al.[37] showed that nonlinear ML models have better accuracy to predict the strength of cement mortar than linear models. Hematibahar et al. [38] suggested a SADRA-based ML model for prediction of fiber reinforced concrete. Song et al. [39] developed a new self-compacting ultra-high-performance fiber-reinforced concrete with compounded high-activity powders. Ağcakoca et al. [40] proposed a hybrid ML–FE model using CDP to enhance prediction of cementitious composite behavior. Siddique et al. [41] modeled and predicted the CS of SCC using ANN. GA, PSO, DE, and GWO were selected for their complementary exploration–exploitation strategies within the ET–GB–HGB stacking framework. Bayesian and Optuna-based optimization are recommended for future comparative evaluation.

The principal novelty of this study lies in developing an integrated optimization–prediction–interpretation framework for the CS of RPM. Unlike conventional single-model approaches, the proposed framework integrates heterogeneous ensemble learners (ET, GB, and HGB) within bagging, voting, and stacking architectures, with hyperparameters optimized using GA, PSO, DE, and GWO. In particular, optimizing the ET–GB–HGB stacking architecture effectively leverages the complementary predictive strengths of randomized tree-based and boosting algorithms. Predictive reliability was evaluated using independent testing, shuffled 10-fold cross-validation, error-distribution and generalization analyzes, and integrated performance ranking. The GA-optimized ET–GB–HGB stacking model was the best performer overall. A multi-level explainability and sensitivity framework of SHAP, permutation importance, interaction analysis, Sobol sensitivity analysis and response-surface analysis elucidate feature importance, non-linear effects and interactions, more than prediction. The results consistently identify MRR as the dominant predictor, with the particle size–MRR interaction exhibiting the strongest pairwise effect. Collectively, the proposed framework advances RPM strength modeling beyond predictive estimation toward a robust, interpretable, and data-driven approach for sustainable mixture assessment and design. Despite recent advances in ML-based RPM strength prediction, the comparative integration of heterogeneous stacking ensembles with multiple metaheuristic optimization strategies remains insufficiently investigated.

2. Material and Methods

The dataset comprised 204 RPM samples from the literature [27], with kind, size (µm), W/B, and MRR (%) as predictors and CS (MPa) as the target variable. Prior to modeling, the data were screened for missing values, duplicates, and invalid entries, followed by Min–Max normalization of continuous predictors. The dataset was then split into 80% for training and 20% for independent testing sets. Pre-processing was applied exclusively to the training dataset to prevent information leakage. An optimization-assisted ensemble ML framework was developed using ET, GB, and HGB as the core learners. The ensemble was integrated thru bagging, voting, and stacking architectures to improve predictive accuracy and generalization. We systematically employed four metaheuristic algorithms, namely GA, PSO, DE, and GWO, for the optimization of hyperparameters with special focus on the joint optimization of the base and meta-learners of the ET-GB-HGB stacking architecture. To prevent information leakage, the model development and hyperparameter optimization were performed only on the training dataset. The predictive performance, generalization and robustness were evaluated by R^2, RMSE, MAE and MAPE via independent testing and shuffled 10-fold cross-validation. Moreover, generalization gaps, variability of cross-validation, error distributions, characteristics of regression errors, and normalized performance ranking were used to investigate the robustness and generalization of the models.

The best model was further interpreted by SHAP to quantify the contribution of the features globally and locally and to elucidate the nonlinear effects and interactions among the input variables. We applied permutation importance, Sobol sensitivity analysis, and response-surface analysis to enhance model interpretability and to gain an understanding of the relationships driving the CS of RPM. All the computational steps such as pre-processing, model development, optimization, validation, interpretability and visualization were performed in Python 3.12 using NumPy, Pandas and Scikit-learn [42, 43]. Table 1 shows the descriptive statistics of the input and output variables. The mean of the variable kind is 0.392 (range 0-1) and the particle size is 27.736 µm (range 0-103.6 µm). The dataset also contains W/B (mean = 0.492, range 0.35–0.55), MRR (mean = 23.816%, range 0–60%) and CS (mean = 36.309 MPa, range 5.4–56.8 MPa), with significant difference between parameters.

Table 1. Descriptive statistics of input and output variables

Variable	Mean	Std. Dev.	Min	25%	Median	75%	Max	Variance	Range
Kind	0.392	0.489	0.00	0.00	0.00	1.00	1.00	0.240	1.00
Size (µm)	27.736	23.722	0.00	12.663	20.83	35.00	103.60	562.726	103.60
W/B	0.492	0.041	0.35	0.50	0.50	0.50	0.55	0.0017	0.20
MRR (%)	23.816	14.141	0.00	10.00	30.00	30.00	60.00	199.962	60.00
CS (MPa)	36.309	9.743	5.40	30.45	37.00	43.40	56.80	94.922	51.40

The histograms with the superimposed kernel density estimation (KDE) of the input parameters (kind, size, W/B, MRR) and the output variable (CS) are shown in Figure 1. The distribution of the dataset indicates that kind is bimodal, size and MRR are positively skewed, W/B is tightly centered at .50, and CS is approximately normal. Such distributions highlight significant aspects of variability and central tendency that are helpful for preprocessing, feature selection and modeling. The Pearson correlation matrix is shown in Figure 2. The correlation between MRR and CS is strong and negative, r = −0.76. The correlations between size and CS, r = −0.01, and W/B and CS, r = −0.17, are weak and negligible in terms of linear correlation. To mitigate overfitting, optimization was restricted to the training data with cross-validation, while Kind was binary encoded as 0 = [Type 1] and 1 = [Type 2].

2.1. Ensemble Machine Learning Models

The ET, GB, and HGB were adopted as the principal ensemble learning algorithms [44, 45]. ET builds trees randomly and combines them to improve the predictive stability, while GB reduces prediction errors iteratively to find complex nonlinear relationships. HGB improves the computational efficiency by discretizing the features based on histograms. At the same time, it keeps the ability to model nonlinear relationships well. To take advantage of the complementary predictive capabilities of the core learners, voting and stacking architectures were used in combining them [46]. Bagging is based on bootstrapping samples and reduces predictive variance. Voting aggregates, the predictions from multiple models. Stacking uses a meta-learner to improve predictive accuracy, robustness, and generalization [47].

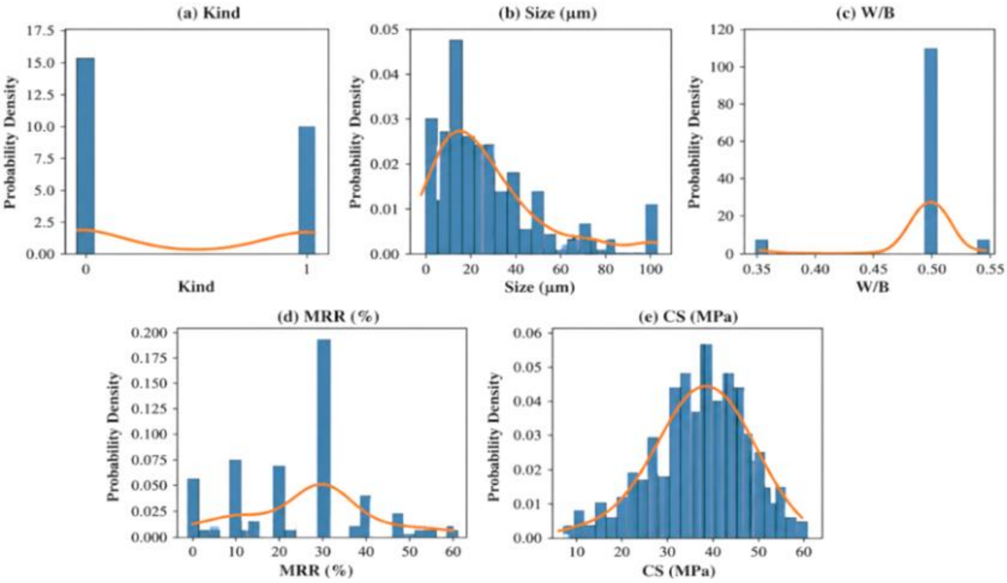

Figure 1. Histograms with KDE illustrating the statistical distributions of the input parameters and CS

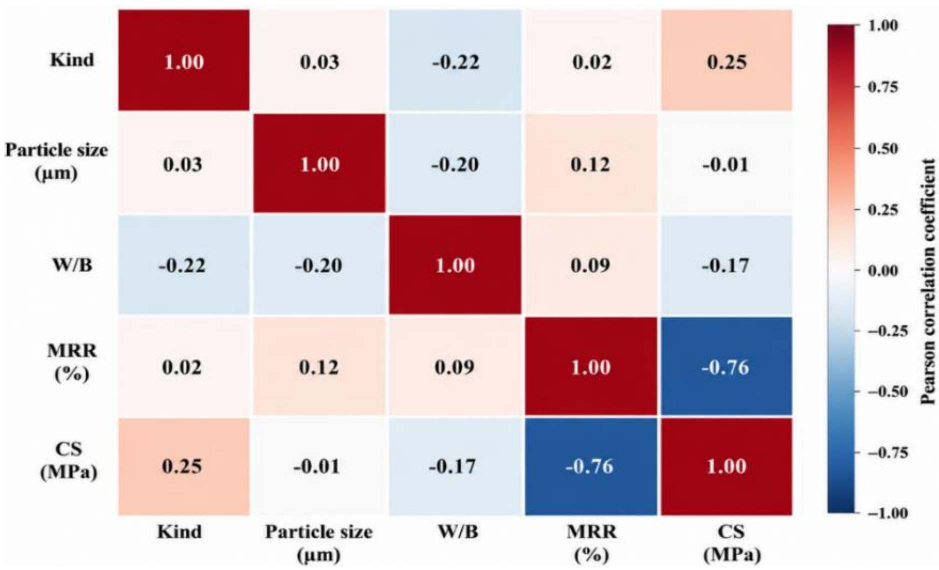

Figure 2. Correlation heatmap of input variables and CS

2.2. Metaheuristic Optimization Algorithms

Genetic algorithm (GA) is an evolutionary optimization technique based on the concept of natural selection. The GA iteratively improves candidate solutions through processes such as selection, crossover, mutation and elitism to find near-optimal hyperparameter configurations [47]. Particle swarm optimization (PSO) is a population-based metaheuristic, which iteratively updates candidate solutions based on individual and global optima, to efficiently explore the hyperparameter space [48]. Differential evolution (DE) is an evolutionary algorithm that is population-based and it has been efficiently utilized for optimizing continuous hyperparameter spaces by means of mutation, crossover and selection [49]. Gray wolf optimizer (GWO) is a swarm-intelligence algorithm that mimics gray wolf hierarchy and hunting behavior to balance exploration and exploitation in the optimization process [50]. These algorithms were employed to optimize ET, GB, HGB, and particularly the ET–GB–HGB stacking ensemble through population-based iterative search.

2.3. SHAP-Based Interpretability

SHapley additive explanations (SHAP) provide a rigorous and transparent interpretation of tree-based models by decomposing predictions into additive feature contributions based on cooperative game theory [48]. SHAP summary plots rank input variables by their global influence on CS while revealing the direction and magnitude of their effects. SHAP dependence plots depict the influence of individual features across their value ranges and reveal interaction effects, whereas SHAP dependency analysis quantifies higher-order interactions and evolving feature contributions. SHAP-based and traditional feature importance analyses improve model transparency by identifying the key factors controlling concrete strength and material behavior [49].

3. Results and Discussion

The hyperparameter search spaces in Table 2 were defined to optimize model complexity, regularization, and generalization. The key hyperparameters of ET, GB and HGB were optimized, and the ensemble architectures utilized their complementary learning characteristics to enhance predictive performance and generalization. Special attention was paid to the metaheuristic optimization of ET–GB–HGB stacking ensemble through GA, PSO, DE, and GWO, aiming at an efficient balance of exploration and exploitation in the hyperparameter search space. GA, PSO, DE and GWO use different evolutionary, swarm-intelligence, differential-evolution and hierarchical search strategies respectively to systematically tradeoff for global exploration and local exploitation. The optimized stacking models took advantage of the complementary benefits of ET, GB, and HGB, combining randomized learning, sequential error correction, and efficient boosting computation. The GA and PSO-optimized configurations exhibit a better performance, which confirms the effectiveness of the metaheuristic tuning in improving the complementary predictive power of the base learners. Collectively, metaheuristic-optimized ensembles enhanced the accuracy, robustness, and generalization of CS prediction for RPM.

Table 2. Hyperparameter search spaces, ensemble configurations, and optimization methods

Method	Hyperparameter search space / configuration	Optimization
Extra Trees (ET)	n_estimators: 100–1000; max_depth: 3–30/None; min_samples_split: 2–20; min_samples_leaf: 1–10; max_features: 0.5–1.0; bootstrap: T/F	GA, PSO, DE, GWO
Gradient Boosting (GB)	n_estimators: 50–1000; learning_rate: 0.01–0.30; max_depth: 2–10; min_samples_split: 2–20; min_samples_leaf: 1–10; subsample: 0.5–1.0	
Hist Gradient Boosting (HGB)	max_iter: 100–1000; learning_rate: 0.01–0.30; max_leaf_nodes: 5–50; max_depth: 2–20/None; min_samples_leaf: 5–30; L2: 0–10; max_bins: 32–255	
Bagging–ET/GB/HGB	n_estimators: 10–200; max_samples: 0.5–1.0; max_features: 0.5–1.0; bootstrap: T/F; base learner: ET/GB/HGB	–
Voting–ET/GB/HGB	Base learners: ET, GB and/or HGB; aggregation: mean/weighted prediction; weights: equal/optimized	–
Stacking–ET/GB/HGB	Base learners: ET, GB and/or HGB; meta-learner: Ridge; alpha: 10^{-4}–10^{2}; internal CV: 5-fold; passthrough: T/F	–
Stacking–ET–GB–HGB–GA	ET + GB + HGB stacking; base/meta-model parameters jointly searched	GA
Stacking–ET–GB–HGB–PSO	ET + GB + HGB stacking; base/meta-model parameters jointly searched	PSO
Stacking–ET–GB–HGB–DE	ET + GB + HGB stacking; base/meta-model parameters jointly searched	DE
Stacking–ET–GB–HGB–GWO	ET + GB + HGB stacking; base/meta-model parameters jointly searched	GWO

Table 3 demonstrates consistently high predictive performance across the developed models, with comparable training, testing, and 10-fold cross-validation results indicating robust generalization. Most R^2 values were above 0.94 with low RMSE and MAE values indicating high prediction accuracy of RPM compressive strength. The highest testing accuracy (R^2 = 0.965; RMSE ≈ 1.96 MPa) was obtained by the individually optimized GB-PSO and ET-DE, and ET-GWO also showed low prediction errors. The HGB–GWO showed good results for cross-validation (R^2 = 0.952; RMSE = 2.022 MPa), which confirm good generalization among the partitions. The conventional bagging, voting and stacking ensembles showed strong predictive performance. The highest validation accuracy was obtained from the voting–ET–GB–HGB (R^2 = 0.954; RMSE = 1.953 MPa). The metaheuristic optimization further improved the stacking performance, where the best testing results of the ET–GB–HGB optimized by the GA (R^2 = 0.966; RMSE = 1.931 MPa; MAE = 1.575 MPa; MAPE = 0.045) were obtained, while the PSO-optimized counterpart exhibited comparable accuracy and robust validation performance. The results verify the effectiveness of the optimized stacking ET-GB-HGB to balance the predictive accuracy and generalization. The superior performance of the GA-optimized configuration validates the efficacy of the evolutionary optimization in harnessing the complementary characteristics of the learners. It is the most appropriate model to predict the CS of RPM. The GA and PSO were found to be the best performing optimizers for the stacking architecture while DE and GWO provided moderate improvements. Metaheuristic optimization improved the predictive accuracy, reduced the errors, and improved the robustness and generalization of ensemble models for CS prediction.

Table 3. Predictive performance of optimized ensemble ML models

Model	Dataset	R^2	RMSE	MAE	MAPE
ExtraTrees-GA	Training	0.943	2.274	1.871	0.050
	Testing	0.955	2.229	1.842	0.052
	Validation	0.946	2.149	1.793	0.050
ExtraTrees-PSO	Training	0.948	2.168	1.778	0.049
	Testing	0.964	1.983	1.692	0.049
	Validation	0.946	2.117	1.788	0.050
ExtraTrees-DE	Training	0.952	2.080	1.726	0.047
	Testing	0.965	1.965	1.631	0.048
	Validation	0.947	2.110	1.831	0.051
ExtraTrees-GWO	Training	0.948	2.169	1.808	0.050
	Testing	0.964	1.985	1.522	0.043
	Validation	0.944	2.181	1.846	0.052
GradientBoosting-GA	Training	0.943	2.256	1.900	0.052
	Testing	0.959	2.135	1.727	0.049
	Validation	0.948	2.124	1.800	0.050
GradientBoosting-PSO	Training	0.951	2.094	1.796	0.050
	Testing	0.965	1.961	1.507	0.043
	Validation	0.946	2.129	1.803	0.050
GradientBoosting-DE	Training	0.938	2.361	2.016	0.054

	Testing	0.963	2.028	1.688	0.049
	Validation	0.946	2.176	1.835	0.051
GradientBoosting-GWO	Training	0.941	2.295	2.002	0.055
	Testing	0.959	2.122	1.825	0.053
	Validation	0.938	2.315	1.960	0.054
HistGradientBoosting-GA	Training	0.946	2.210	1.856	0.051
	Testing	0.956	2.198	1.898	0.054
	Validation	0.949	2.099	1.764	0.049
HistGradientBoosting-PSO	Training	0.940	2.322	1.964	0.053
	Testing	0.954	2.251	1.824	0.051
	Validation	0.948	2.093	1.747	0.049
HistGradientBoosting-DE	Training	0.948	2.170	1.832	0.051
	Testing	0.957	2.189	1.843	0.054
	Validation	0.946	2.129	1.776	0.049
HistGradientBoosting-GWO	Training	0.946	2.203	1.850	0.050
	Testing	0.956	2.206	1.753	0.050
	Validation	0.952	2.022	1.681	0.046
Bagging-ExtraTrees	Training	0.950	2.129	1.763	0.048
	Testing	0.957	2.181	1.838	0.052
	Validation	0.941	2.222	1.914	0.052
Bagging-GradientBoosting	Training	0.949	2.145	1.796	0.049
	Testing	0.964	1.999	1.621	0.046
	Validation	0.944	2.193	1.876	0.053
Bagging-HistGradientBoosting	Training	0.948	2.172	1.822	0.050
	Testing	0.961	2.079	1.681	0.048
	Validation	0.947	2.123	1.784	0.049
Voting-ET-GB	Training	0.950	2.129	1.748	0.047
	Testing	0.961	2.083	1.747	0.049
	Validation	0.936	2.329	1.982	0.055
Voting-ET-HGB	Training	0.949	2.139	1.799	0.050
	Testing	0.959	2.138	1.841	0.054
	Validation	0.940	2.263	1.932	0.053
Voting-GB-HGB	Training	0.943	2.257	1.899	0.052
	Testing	0.965	1.968	1.648	0.047
	Validation	0.944	2.185	1.831	0.051
Voting-ET-GB-HGB	Training	0.948	2.173	1.852	0.051
	Testing	0.959	2.119	1.754	0.050
	Validation	0.954	1.953	1.619	0.045
Stacking-ET-GB	Training	0.946	2.197	1.898	0.052
	Testing	0.962	2.057	1.716	0.050
	Validation	0.948	2.082	1.762	0.049
Stacking-ET-HGB	Training	0.950	2.116	1.770	0.049
	Testing	0.954	2.259	1.972	0.057
	Validation	0.945	2.185	1.844	0.051
Stacking-GB-HGB	Training	0.943	2.262	1.917	0.052
	Testing	0.963	2.029	1.632	0.045
	Validation	0.950	2.097	1.780	0.050
Stacking-ET-GB-HGB	Training	0.949	2.148	1.791	0.049
	Testing	0.956	2.206	1.874	0.052
	Validation	0.945	2.176	1.851	0.051
Stacking-ET-GB-HGB-GA	Training	0.949	2.149	1.804	0.050
	Testing	0.966	1.931	1.575	0.045
	Validation	0.947	2.129	1.792	0.049
Stacking-ET-GB-HGB-PSO	Training	0.947	2.181	1.827	0.050
	Testing	0.966	1.942	1.580	0.047
	Validation	0.951	2.047	1.761	0.049
Stacking-ET-GB-HGB-DE	Training	0.946	2.210	1.853	0.051
	Testing	0.955	2.226	1.880	0.053
	Validation	0.945	2.177	1.852	0.051
Stacking-ET-GB-HGB-GWO	Training	0.945	2.227	1.836	0.049
	Testing	0.958	2.159	1.760	0.050
	Validation	0.951	2.041	1.687	0.046
Ensemble-ET-GB	Training	0.950	2.128	1.740	0.047
	Testing	0.952	2.304	1.961	0.055
	Validation	0.942	2.199	1.862	0.051
Ensemble-ET-HGB	Training	0.944	2.244	1.889	0.052
	Testing	0.951	2.333	1.896	0.050
	Validation	0.948	2.136	1.815	0.051
Ensemble-GB-HGB	Training	0.942	2.280	1.872	0.050
	Testing	0.961	2.079	1.659	0.048
	Validation	0.948	2.102	1.756	0.049
Ensemble-ET-GB-HGB	Training	0.955	2.013	1.665	0.045
	Testing	0.956	2.213	1.835	0.053
	Validation	0.945	2.186	1.870	0.052

Although GA-stacking achieved the highest integrated ranking, its performance was closely comparable to PSO-stacking and other leading models. Therefore, the marginal differences in R^2 should not be interpreted as statistically significant superiority.

Figure 3 illustrates the agreement between observed and predicted CS values for the best-performing optimized ensemble models across the training, testing, and 10-fold cross-validation datasets, providing a visual assessment of predictive accuracy and model generalization. The close clustering of predicted values around the 1:1 reference line, with most observations falling within the ±10% error bounds, indicates high predictive accuracy, minimal systematic bias, and robust model generalization. The GA and PSO-optimized ET–GB–HGB stacking models demonstrate consistently strong performance across all data partitions, while ET–DE and GB–PSO exhibit comparable predictive accuracy and generalization. The consistency across training, testing, and cross-validation results demonstrates robust and stable model generalization across different data partitions. These results indicate that the metaheuristic optimization is effective in improving the predictive accuracy, robustness, and reliability of ensemble learning for the CS prediction of RPM.

The statistical analysis of prediction errors for the best ML models is shown in Figure 4. The distributions of absolute errors shown in Figure 4(a) are concentrated at low magnitudes, revealing the high predictive accuracy of all the models tested. The residual distributions in Figure 4(b) are compact and centered around zero, suggesting a small systematic bias and stable prediction performance. The residual concentrations around zero are evident in GB–PSO and stacking–ET–GB–HGB–PSO. It indicates that the prediction bias is low, and the prediction stability is high. The boxplots in Figure 4(c) confirm low error magnitudes and limited dispersion, with only a few isolated larger errors. The optimized ensemble models demonstrate low prediction errors, minimal bias, and robust RPM compressive-strength prediction.

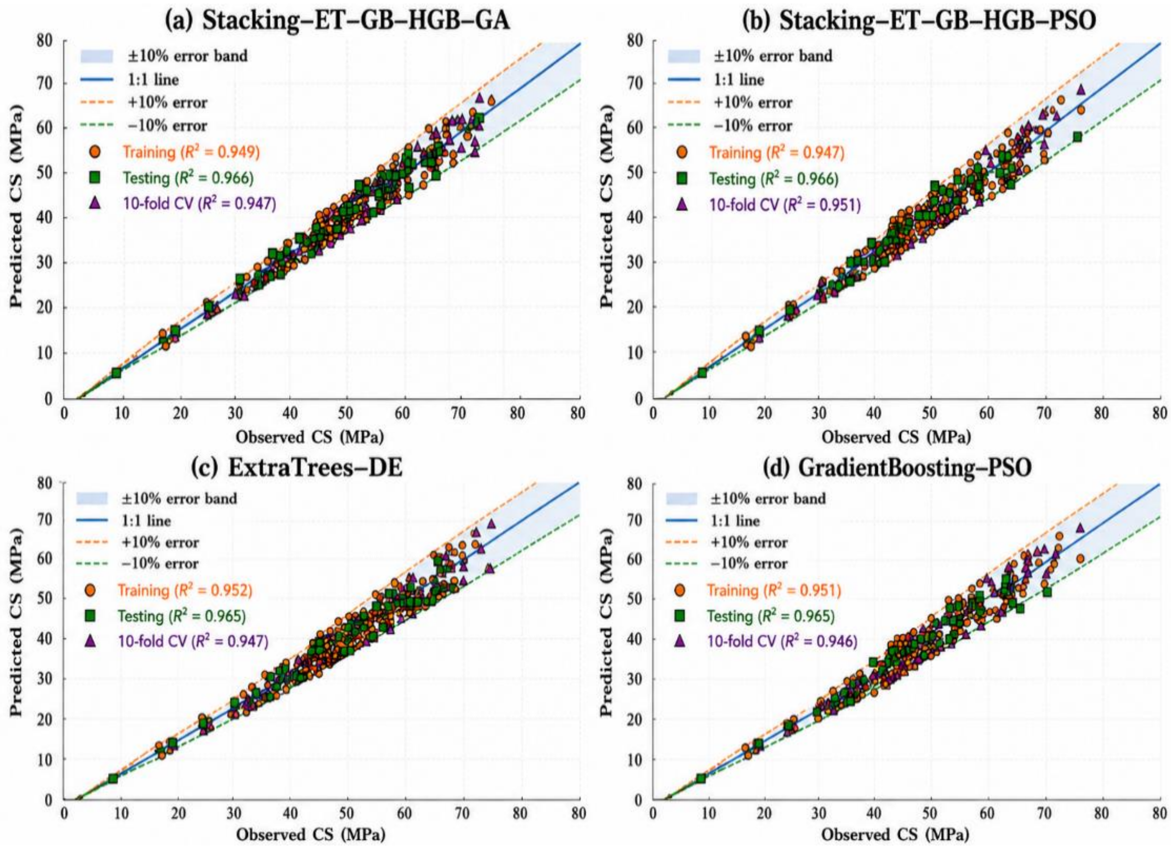

Figure 3. Observed versus predicted CS for the top-performing optimized ensemble models across training, testing, and 10-fold cross-validation

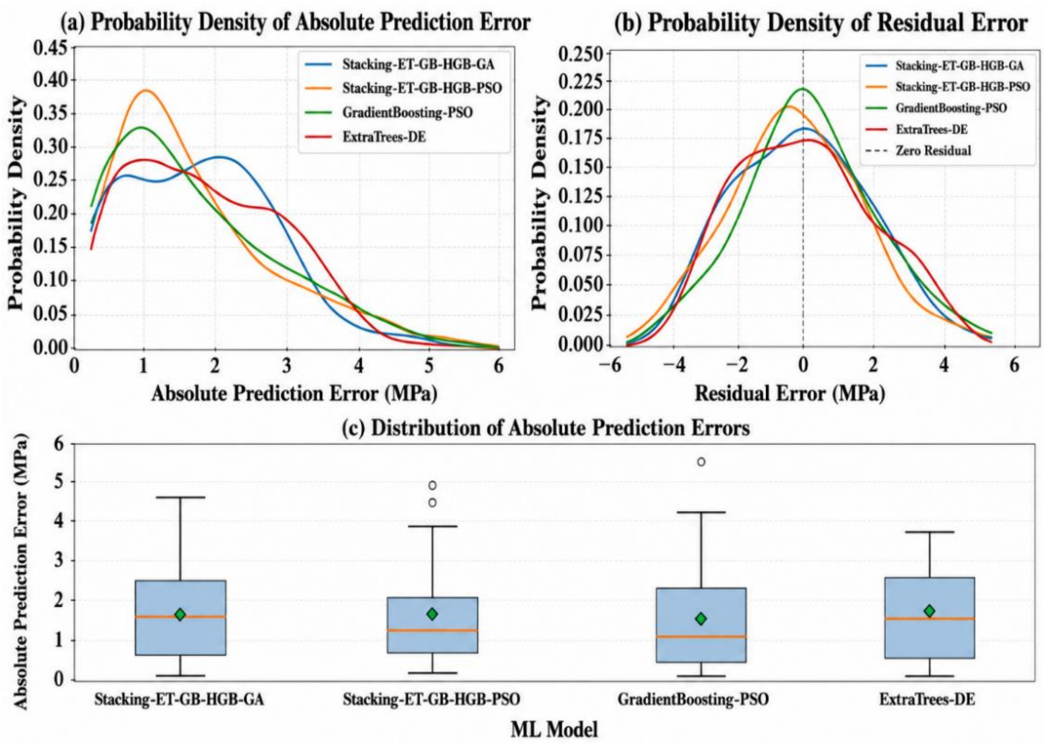

Figure 4. Statistical distributions of prediction errors for the top-performing ML models: (a) absolute-error density; (b) residual-error density; and (c) absolute-error distribution

The cross-validation stability of the tested ML models is shown in Figure 5, where the standard deviations of R^2, RMSE, and MAE among the 10 validation folds are shown. The smaller standard deviations mean that predictions are more consistent and robust across different splits of the data. GB–DE is the most stable in the cross-validation in general as it has the lowest standard deviations for all three metrics of performance, followed by stacking–ET–HGB–GA. Conversely, models with larger standard deviations in RMSE and MAE display higher fold-to-fold variation, i.e., they are less predictive stable across the cross-validation partitions. These findings identify GB–DE as the most stable model; however, comprehensive model superiority should be established by jointly considering predictive accuracy and cross-validation stability.

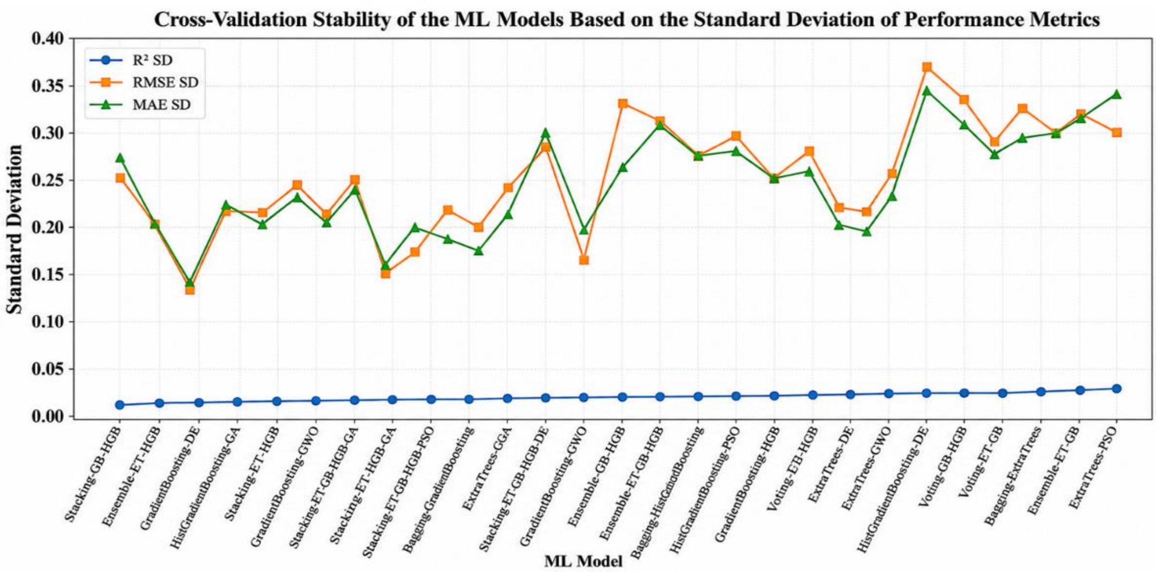

Figure 5. Cross-validation stability of the ML models based on the standard deviation of R^2, RMSE, and MAE

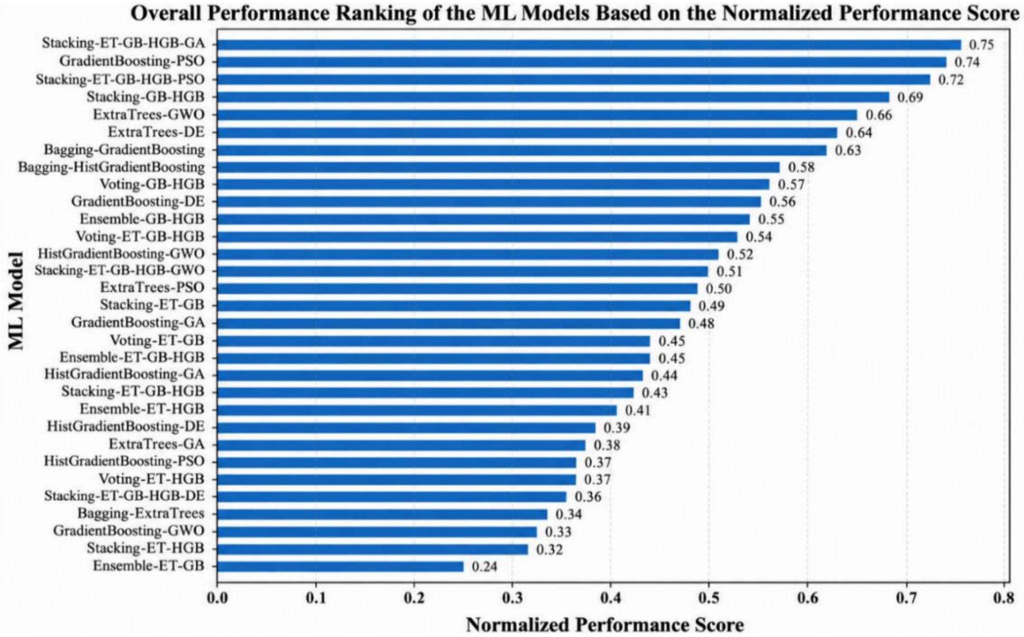

Figure 6. Overall performance ranking of the ML models based on the normalized performance score

Figure 6 presents the overall ranking of the evaluated ML models based on the normalized performance score, which integrates the considered evaluation criteria into a unified measure of model performance. The higher normalized score indicates the better overall predictive ability that is better tradeoff between predictive accuracy, error minimization and generalization performance. The stacking-ET-GB-HGB-GA model has the best normalized performance score of 0.75, followed by GB-PSO (0.74) and stacking-ET-GB-HGB-PSO (0.72). The results demonstrate that metaheuristic optimization is effective in improving the overall predictive performance and the generalization capability of the ensemble models. The non-optimized stacking-GB-HGB model is also competitive, achieving the fourth best normalized score of 0.69. In general, the dominance of optimized ensemble models over the top-ranked configurations shows that the integration of heterogeneous ensemble learners and metaheuristic optimization provides a good compromise between predictive accuracy and generalization ability. Therefore, the stacking-ET-GB-HGB-GA model is the best configuration since it has the highest integrated normalized performance score and the most favorable overall predictive performance among the evaluated models.

As indicated in Figure 6, the normalized performance score was obtained by transforming the considered metrics to a common dimensionless scale according to their optimization direction and combining them with equal weights. The score represents an integrated comparative measure, and the ranking of closely performing models may vary under alternative normalization or weighting schemes.

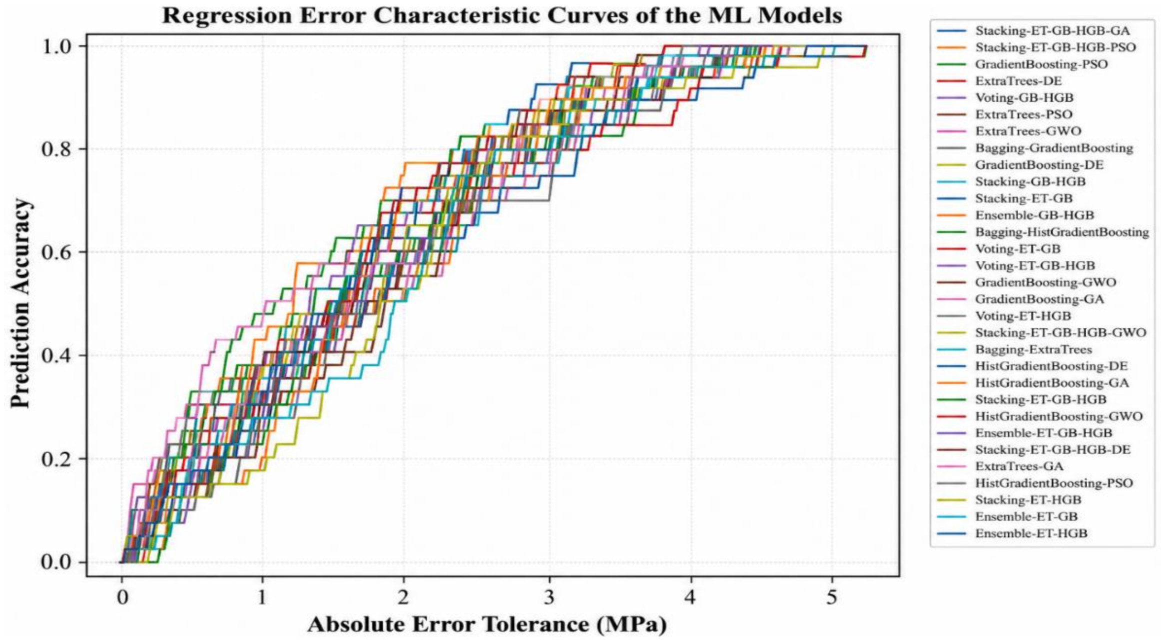

Figure 7. Regression error characteristic curves of the ML models

Figure 7 shows the regression error characteristic (REC) curves of the evaluated ML models, which give the cumulative proportion of predictions with absolute errors below a given tolerance threshold. Models with a steeper initial increase of the REC curve and a faster convergence toward the upper-left region show a better predictive accuracy, i.e., a higher proportion of the predictions within relatively small absolute-error tolerances. Among the evaluated models, stacking–ET–GB–HGB–GA exhibits the most favorable REC characteristics, with a rapid increase in cumulative accuracy at low error tolerances, confirming its superior predictive performance. This result is in line with the first ranking in the integrated normalized performance measure and also consolidates its overall better predictive performance. The remaining REC curves are closely grouped and exhibit roughly similar predictive performance, with more pronounced differences at lower error tolerances. As the error tolerance increases, the REC curves tend to 1.0, indicating that almost all the predictions are within the corresponding absolute-error tolerance. Thus, the REC analysis further corroborates the superior predictive capability of stacking–ET–GB–HGB–GA, while complementing conventional performance metrics by characterizing the cumulative distribution of absolute prediction errors.

Figure 8 presents an integrated interpretation of the optimized ML model using SHAP and permutation feature importance to quantify the magnitude and direction of input contributions to RPM compressive strength. The agreement of the SHAP and permutation analyzes provides complementary evidence of the relative importance of model inputs. The SHAP summary plot is shown in Figure 8(a). MRR is the most dominant predictor with the widest distribution of SHAP values and the most significant impact on model predictions. The higher values of MRR are mainly related to negative values of SHAP, which means that the increasing of the recycled powder mass replacement rate has generally decreasing effect on the predicted CS, while the lower values of MRR have positive influences. The second most influential predictor with nonlinear effect is particle size. Kind also has significant contribution, with distinct category-dependent SHAP effects on predicted CS. Conversely, W/B has the smallest SHAP range, i.e. the least relative influence on the predicted CS in the parameter range investigated. This hierarchy is confirmed by the global SHAP importance in Figure 8(b), where MRR has the highest contribution (4.693 MPa), followed by particle size (2.012 MPa) and kind (1.817 MPa), while W/B has a significantly lower influence (0.249 MPa).

The normalized SHAP contributions in Figure 8(c) also confirm the dominant effect of MRR (53.3%), followed by particle size (22.9%), kind (20.7%), and W/B (2.8%). So, MRR contributes more than 50% to the total feature importance of the model. It can be said that MRR is the main factor in CS prediction. The permutation importance results shown in Figure 8(d) agree with the SHAP-based ranking, with MRR leading to the largest reduction in R^2, and thus confirming its dominant contribution to the predictive performance. The influence of the particle size and kind on the predictive performance is relatively small but noticeable, whereas the effect of W/B is marginal. The error bars also provide a measure of the variability of permutation importance across repeated perturbations, and thus the stability of the estimated feature contributions. The combination of SHAP and permutation importance yielded a single order of feature importance: MRR > particle size > kind > W/B, indicating that MRR is the most important predictor for CS, followed by particle size and kind. The effect of W/B on the CS prediction is relatively small in the data range explored. Importantly, these importance measures describe predictive associations, not direct causal effects.

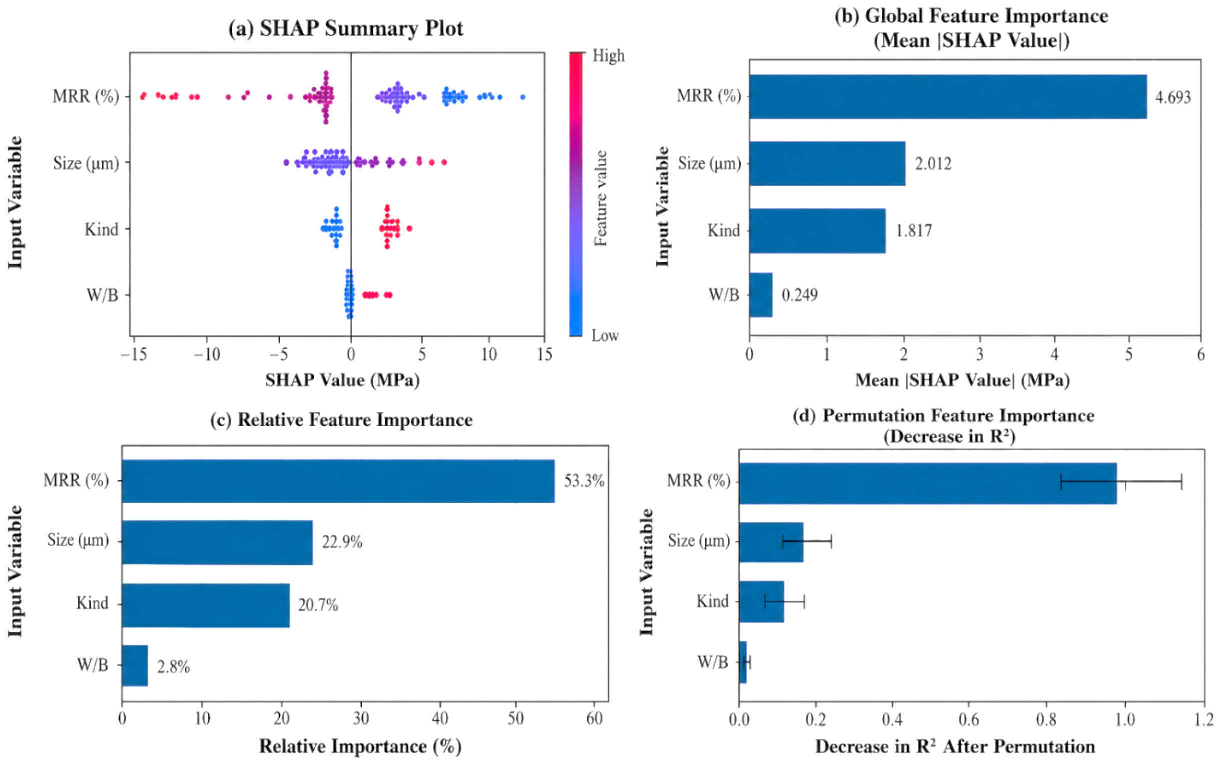

Figure 8. SHAP-based interpretation of the optimized ML model: (a) SHAP summary plot; (b) global feature importance; (c) relative feature importance; and (d) permutation feature importance

Figure 9 shows a complete SHAP analysis of the optimized ML model, combining local and global interpretability to characterize the influence of the input variables on the predicted CS. The individual waterfall plot is shown in Figure 9(a), where the predicted CS is reduced from the baseline of 36.415 MPa to 16.811 MPa. This decrease is mainly controlled by big negative contribution of MRR (−15.67 MPa), followed by particle size (−2.61 MPa) and kind (−1.17 MPa), while W/B has only a very small effect (−0.15 MPa). The results confirm that MRR is the main factor that drives the low-CS prediction for the instance studied. In contrast, the maximum predicted CS of 48.133 MPa is demonstrated in Figure 9(b) where all the input variables have a positive contribution in relation to the baseline prediction. The biggest positive contribution was provided by kind

(+3.61 MPa) followed by MRR (+3.35 MPa), particle size (+2.83 MPa) and W/B (+1.93 MPa). The opposite SHAP patterns in Figures 9(a) and (b) indicate that the magnitude and direction of feature contributions are specific to combination of mixture parameters. The SHAP decision plot is shown in Figure 9(c), which indicates the cumulative effect of the input variables in the shifting of the predictions from the baseline to the final CS values. The clear divergence of the prediction trajectories especially for MRR and particle size, shows their significant observation-dependent effects and implies the nonlinear relations with the predicted CS. Figures 9(d) and (e) also show the global SHAP distributions, which characterize the magnitude and direction of contributions of features. The SHAP distribution for MRR is the widest, indicating that it has the highest global impact on the predicted CS, followed by the particle size and kind, while the W/B has a relatively narrow global importance. The variation in SHAP values across feature magnitudes further demonstrates nonlinear and observation-dependent effects within the investigated parameter domain. Figure 9(f) quantifies the local feature contributions for the highest-CS prediction. Collectively, the SHAP analyses identify MRR as the dominant global predictor, followed by particle size and kind, while W/B has comparatively limited influence. The variation in local contributions further highlight the value of combining global and local SHAP interpretations.

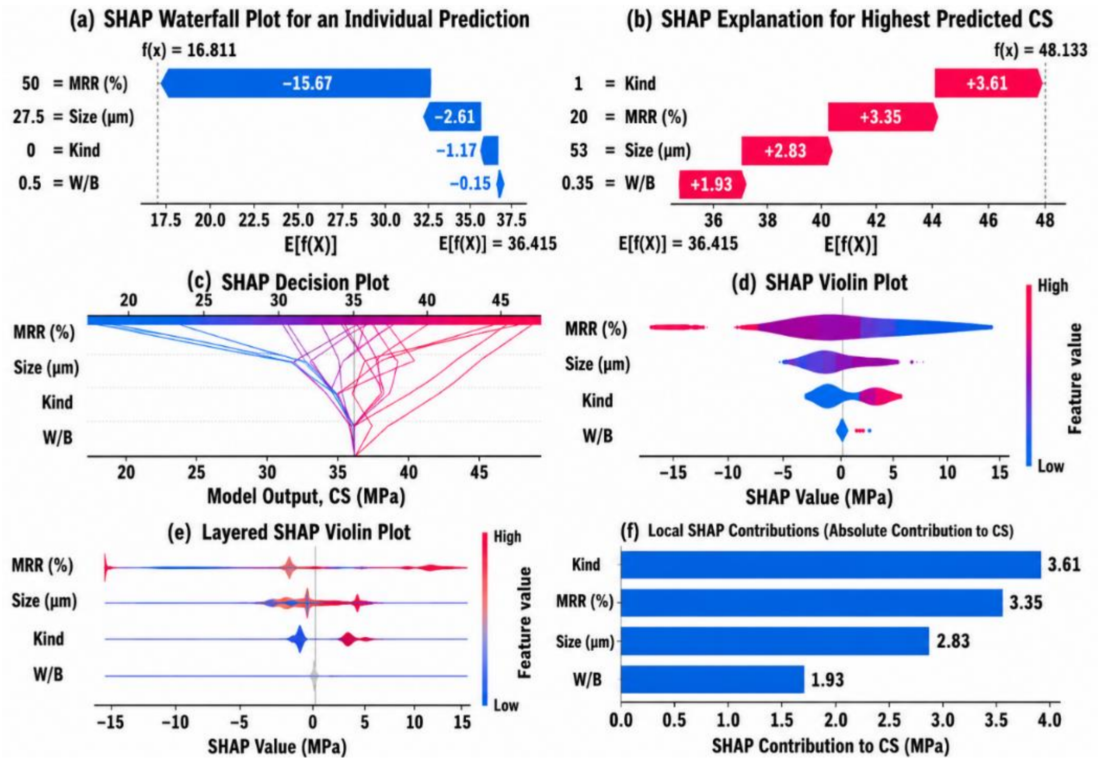

Figure 9. SHAP-based local and global interpretation of the optimized ML model: (a) individual prediction waterfall plot; (b) highest predicted CS explanation; (c) SHAP decision plot; (d) SHAP violin plot; (e) layered SHAP violin plot; and (f) local SHAP contributions

Figure 10 presents the SHAP heatmap, illustrating observation-wise feature contributions to predicted CS. The most significant positive and negative SHAP variations are shown by MRR, confirming its dominant predictive effect, followed by particle size and kind, while W/B shows relatively limited effects. Moreover, the heterogeneous SHAP patterns across testing observations further indicate nonlinear and sample-dependent feature contributions.

The SHAP dependence analysis is visualized in Figure 11, where we observe distinct non-linear and feature-specific impacts on the predicted CS. Kind has an obvious categorical influence, with kind = 1 contributing positively and kind = 0 negatively. Particle size has a nonlinear association with the mostly negative SHAP effects in the intermediate sizes and positive contributions in larger sizes followed by a decrease at the upper end. W/B has relatively small and discrete effects while MRR has the strongest dependence with SHAP values decreasing significantly as MRR increases. The results confirm MRR as the dominant factor governing CS predictions, followed by particle size and kind, while W/B exerts a comparatively limited influence.

Figure 12 illustrates the SHAP interaction effects among the principal input variables on predicted CS. Kind exhibits distinct size-dependent contributions, while the size–MRR interaction shows considerable variability in SHAP effects. The W/B–size interaction remains comparatively limited, whereas the MRR–size relationship exhibits the strongest interaction pattern, with increasing MRR generally shifting SHAP contributions from positive to strongly negative values. The results highlight MRR–size as the most influential interaction governing CS predictions.

Figure 13 presents the pairwise SHAP interaction strengths among the input variables. The size–MRR interaction exhibits the highest mean absolute SHAP interaction value (0.460 MPa), indicating the strongest coupled influence on predicted CS, followed by the kind–size interaction (0.214 MPa). In contrast, interactions with W/B are relatively weak. These results suggest that the interaction of the particle size and the MRR mainly affects the prediction response, which is further confirmed by the interaction trend shown in Figure 12.

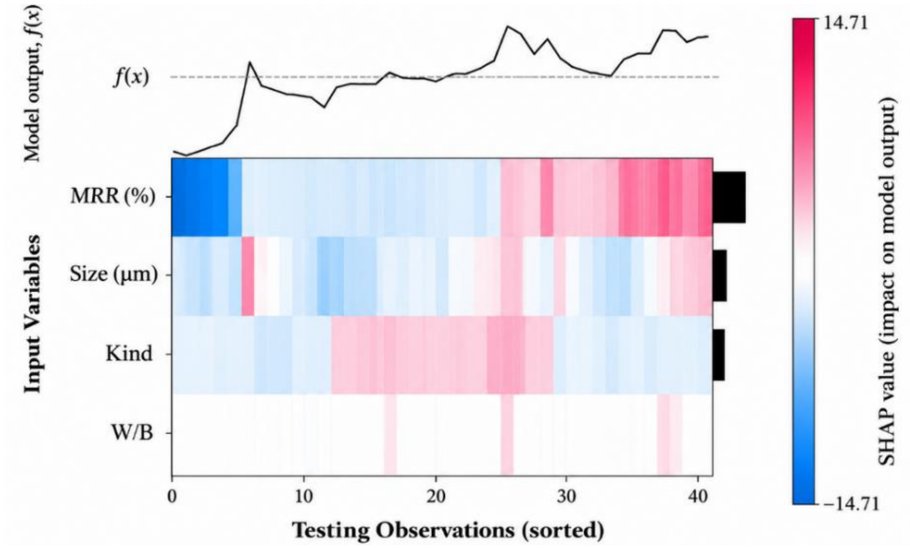

Figure 10. SHAP heatmap illustrating feature contributions to model predictions across the testing observations

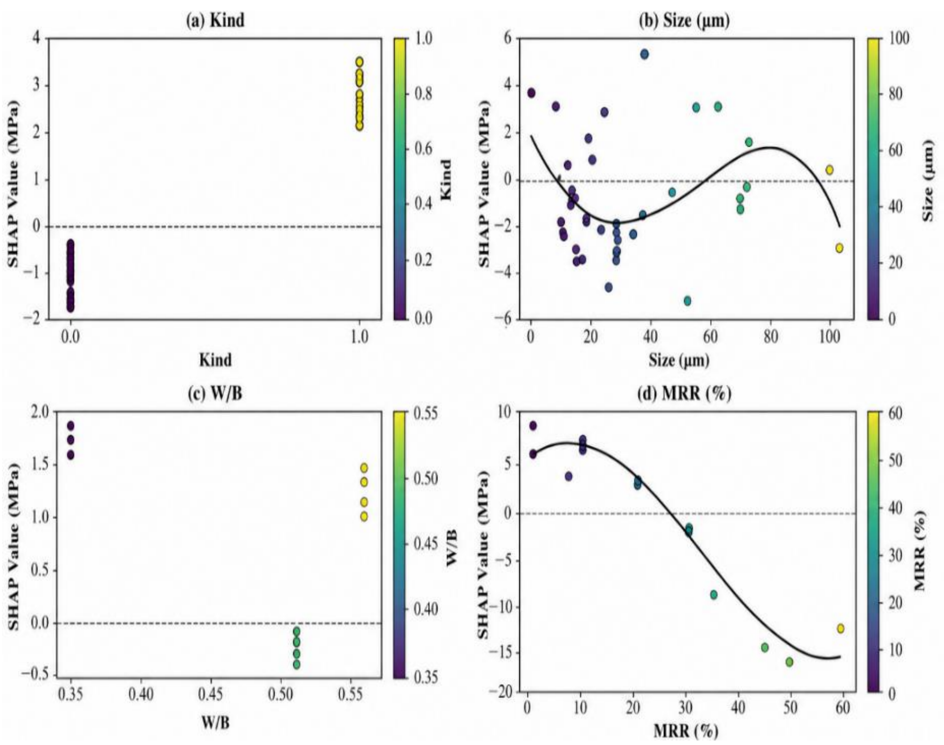

Figure 11. SHAP dependence analysis of the effects of (a) kind, (b) size, (c) W/B, and (d) MRR on predicted CS

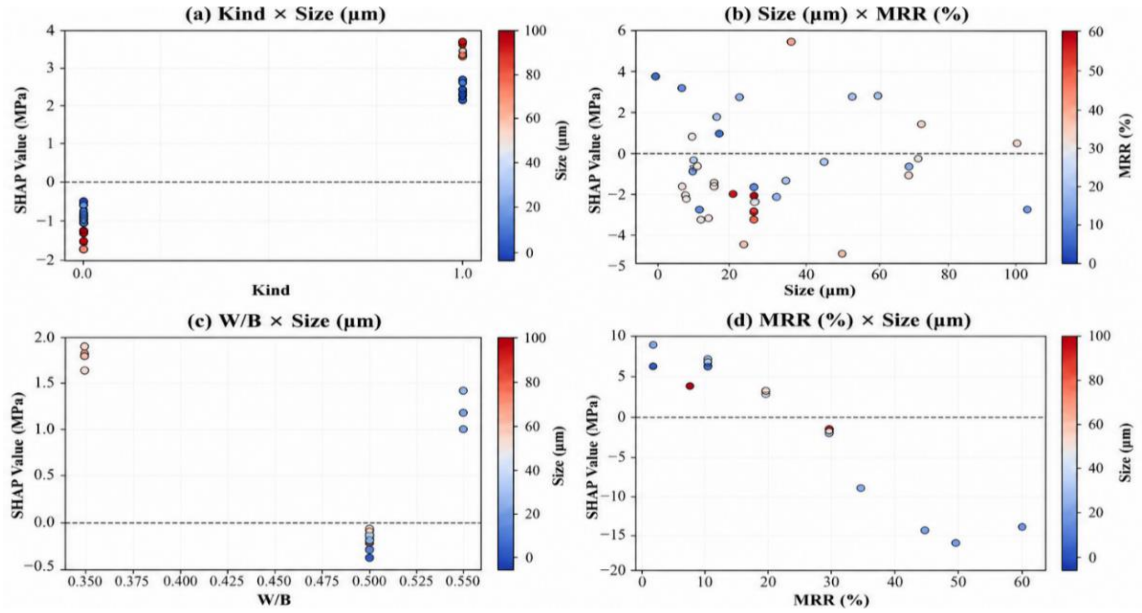

Figure 12. SHAP interaction analysis of the combined effects of kind, size, W/B, and MRR on predicted CS

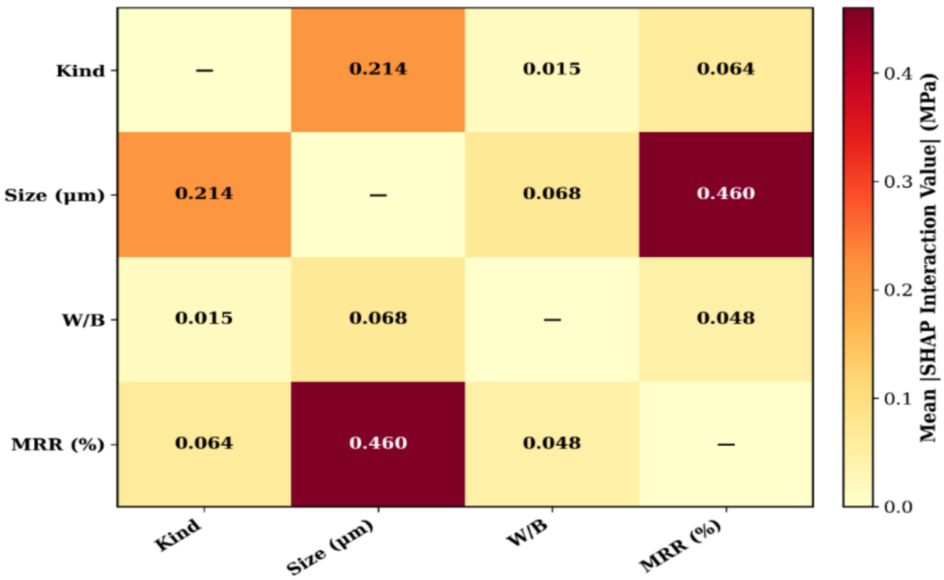

Figure 13. SHAP interaction heatmap illustrating pairwise interactions among the input variables

Figure 14 presents the directional distribution and magnitude of positive and negative SHAP contributions to predicted CS. As shown in Figure 14(a), kind and MRR exhibit relatively balanced positive and negative effects, whereas particle size and particularly W/B are more frequently associated with negative contributions. These effects are quantified further in Figure 14(b) which shows that the dominant contribution in both directions is from MRR and that the negative contribution from MRR is larger than the positive contribution. Particle size also has relatively stronger negative than positive effect. Kind has relatively stronger positive effect while W/B has marginal effect. These results together suggest that the model learns different directions of effects among the input variables and that MRR is the most influential factor overall for CS prediction. Figure 15 displays the Sobol global sensitivity analysis based on the first-order (S1) and total-order (ST) indices for kind = 0 and kind = 1. The highest sensitivity indices are observed for MRR for both the categories, showing the dominant contribution of MRR to the variability of predicted CS. The particle size has a significantly smaller but still measurable effect and W/B contributes only to a very small extent. The higher values of ST than S1 values indicate the presence of interaction effects, but small differences between S1 and ST values indicate that the main effects of individual variables mainly control the model response. The comparable sensitivity patterns for both kind categories further demonstrate the robustness of the identified feature hierarchy, corroborating the SHAP and permutation importance results.

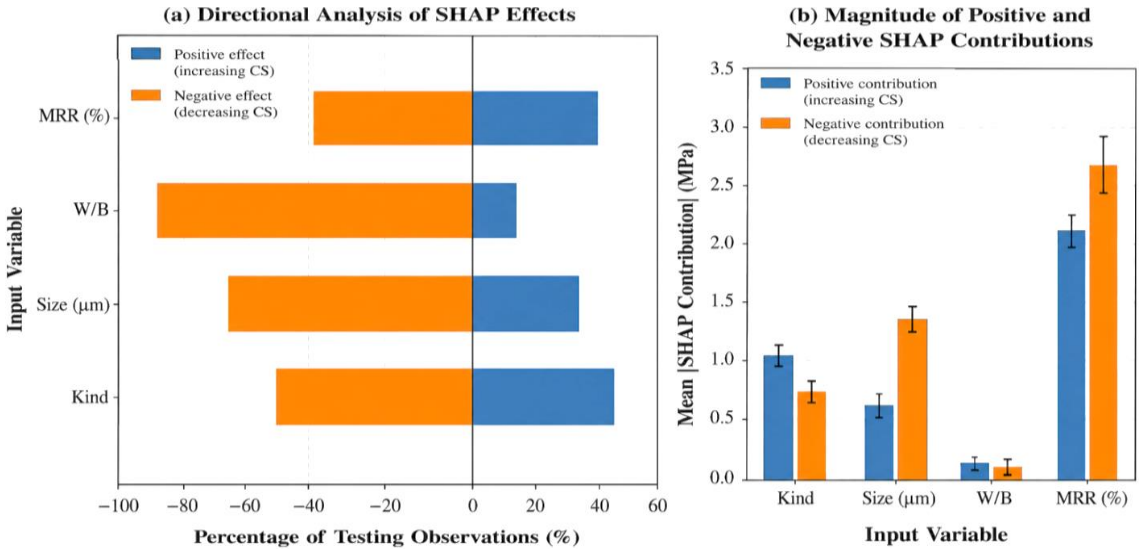

Figure 14. Directional and magnitude-based analysis of positive and negative SHAP contributions of the input variables to CS prediction

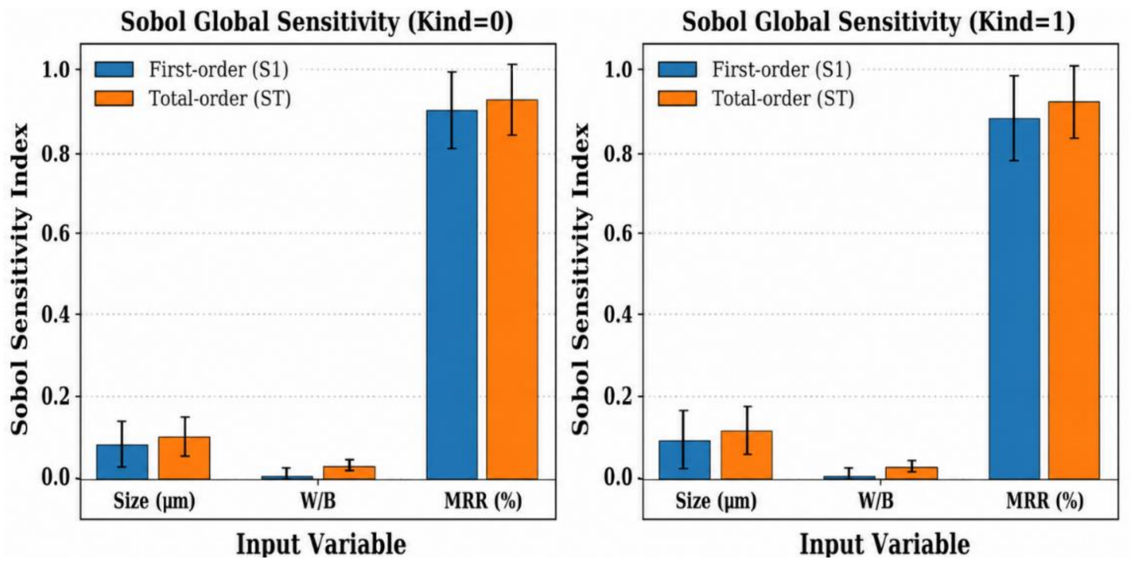

Figure 15. Sobol global sensitivity analysis of the input variables based on first-order (S1) and total-order (ST) sensitivity indices for (a) kind = 0 and (b) kind = 1

Figure 16 presents the predicted CS response surfaces, illustrating the coupled effects of particle size, W/B, and MRR for kind = 0 and kind = 1. In both kind categories, MRR is the most influential, with higher MRR resulting in a large decrease in predicted CS, consistent with the SHAP and Sobol sensitivity analyzes. The size–W/B surfaces indicate a more complex nonlinear response suggesting that the effects of particle size are different over the W/B range studied. In contrast, the W/B–MRR surfaces are predominantly governed by MRR, with comparatively moderate variations attributable to W/B. Differences between kind = 0 and kind = 1 further indicate that recycled powder type modifies the magnitude of the predicted response. These surfaces confirm the nonlinear and interactive dependence of CS on mixture parameters, while reinforcing MRR as the principal factor controlling model predictions.

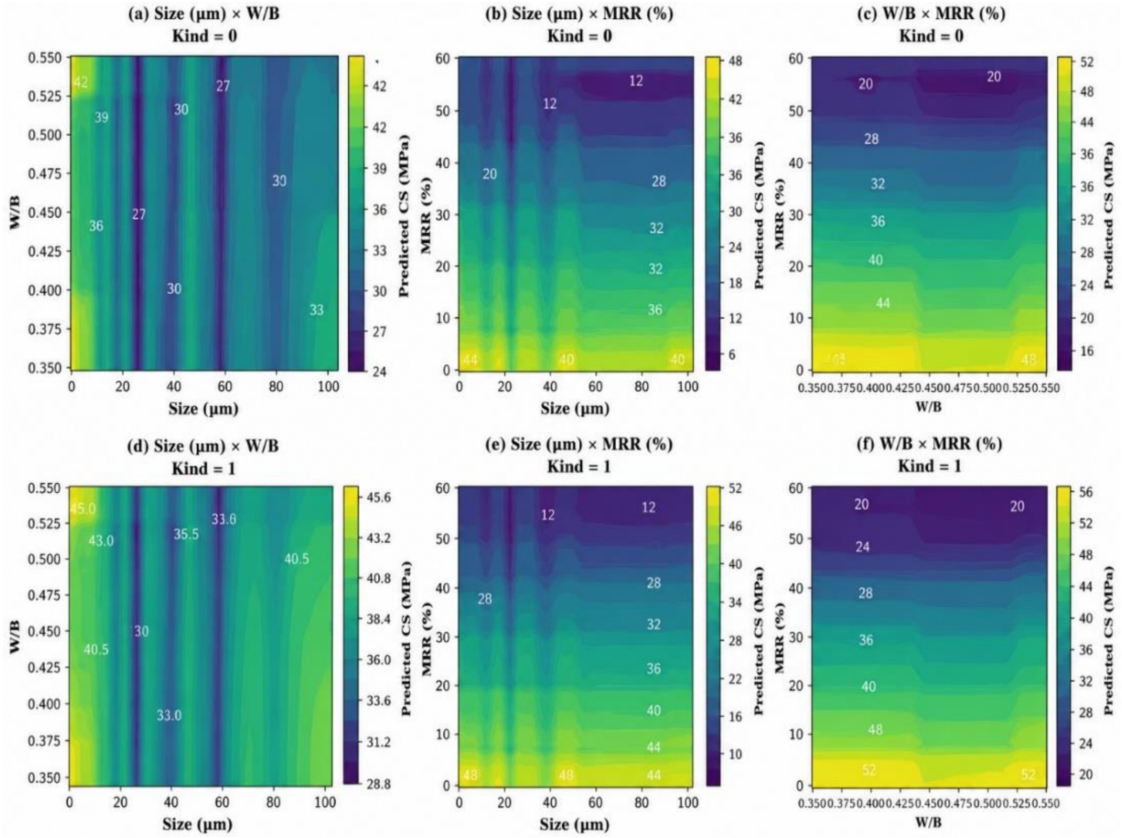

Figure 16. Predicted CS response surfaces illustrating the interactive effects of size (µm), W/B, and MRR (%)

Figure 17 shows the second order Sobol sensitivity indices (S2) of pairwise interaction between particle size, W/B and MRR for each kind category. The size-MRR interaction has the highest sensitivity for kind = 0 (S2 = 0.017) and kind = 1 (S2 = 0.023), followed by MRR-W/B. The size-W/B interaction has negligible effects. The relatively small S2 indices indicate that CS variability is governed predominantly by the main effects, particularly MRR, rather than pairwise interactions. Importantly, the dominant size–MRR interaction corroborates the SHAP interaction results, providing complementary evidence of its relevance to CS prediction. The overall workflow of the proposed explainable and metaheuristic-optimized ensemble ML framework for predicting and interpreting the CS of recycled powder mortar is summarized in Figure 18.

The pronounced MRR–particle size interaction can be physically associated with their coupled effects on particle packing, specific surface area, effective binder replacement, and matrix densification. Increasing MRR introduces a greater proportion of recycled powder, whereas particle size governs its filler efficiency and surface-area-related behavior. Finer particles may enhance void filling and matrix compactness but can also increase water demand because of their larger specific surface area. Consequently, the strength response to MRR depends on the particle-size characteristics of the recycled powder, providing a plausible materials-based explanation for the strong nonlinear interaction identified by SHAP, Sobol, and response-surface analyses. Direct confirmation of these mechanisms, however, requires experimental and microstructural validation.

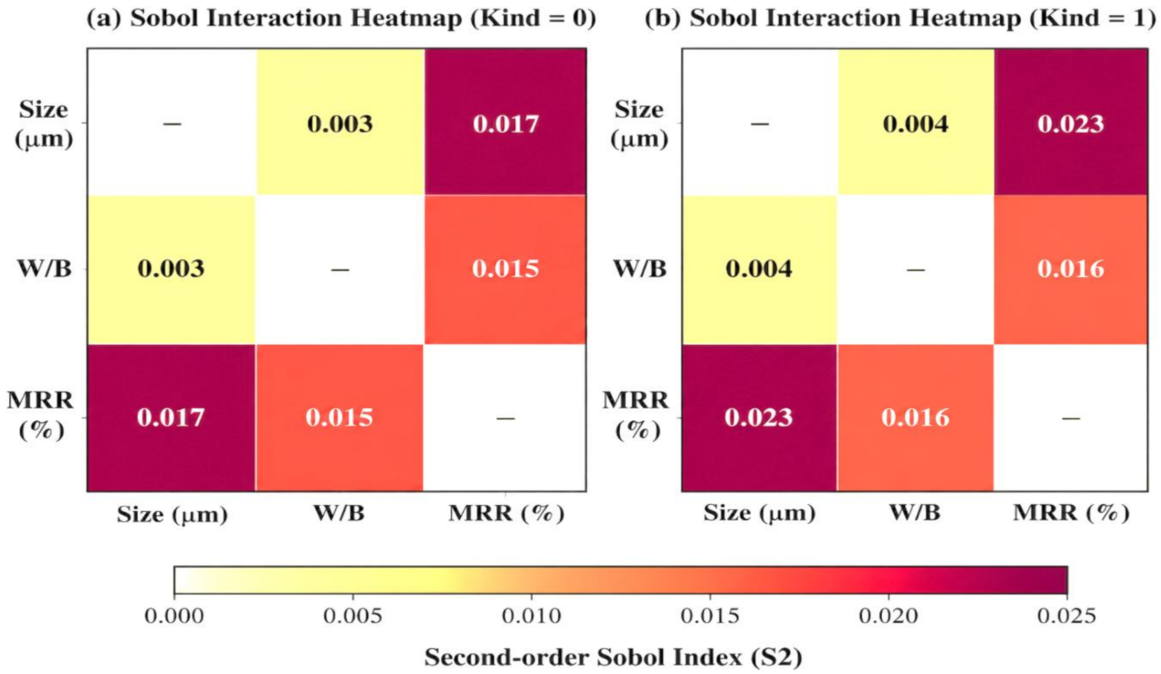

Figure 17. Second-order Sobol interaction heatmaps illustrating the pairwise effects of size (μm), W/B, and MRR (%) on predicted compressive strength for (a) kind = 0 and (b) kind = 1.

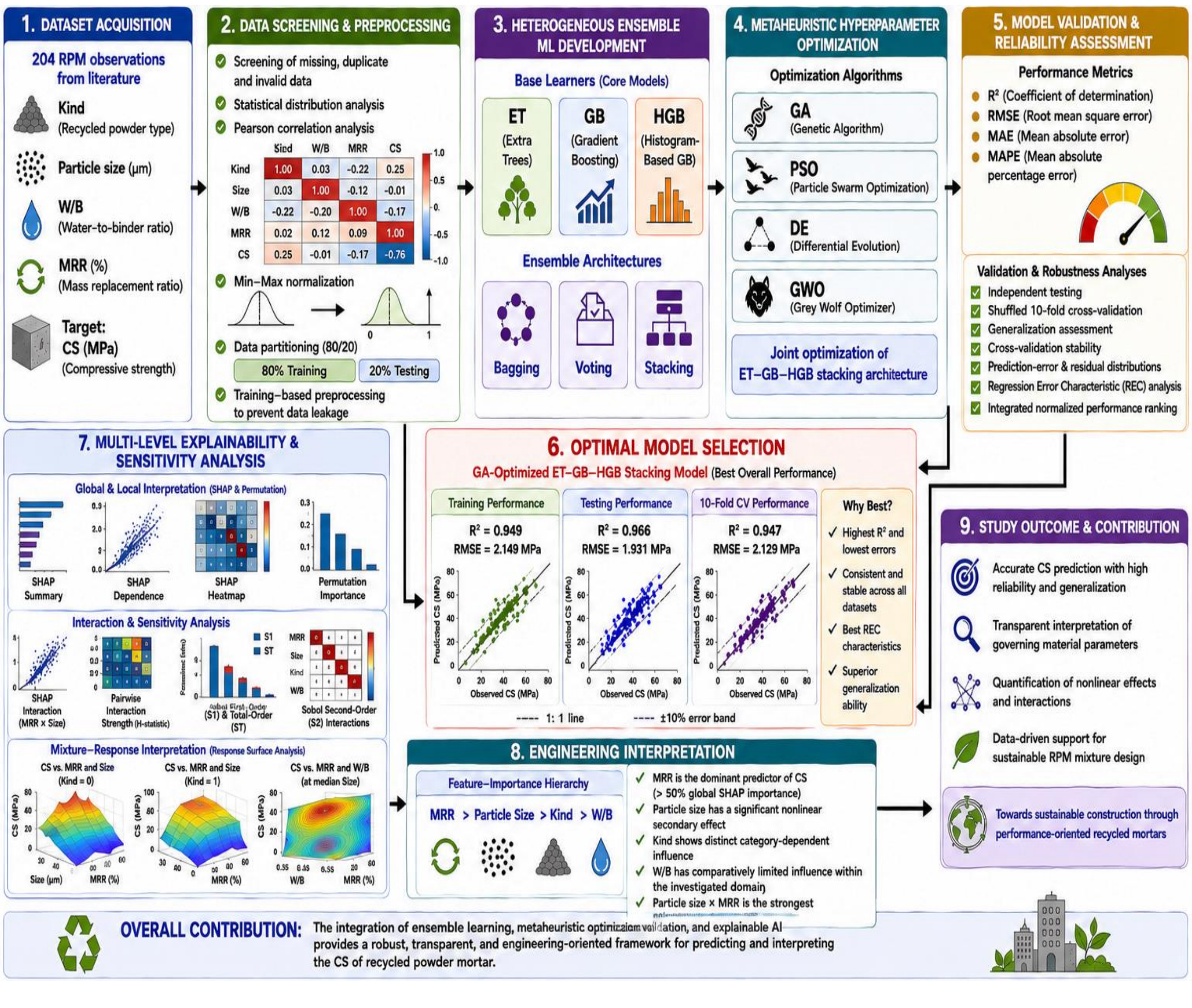

Figure 18. Graphical abstract of the proposed explainable and metaheuristic-optimized ensemble ML framework for predicting and interpreting the CS of recycled powder mortar

4. Engineering Implications and Interpretation of the Input Variables

Integrated SHAP, permutation, Sobol and response-surface analyze give engineering insight into the key factors driving CS of RPM. Although the mathematical formulations are different, all interpretability methods agree that MRR is the main CS predictor, followed by particle size and recycled powder type, with W/B being comparatively less influential. Global SHAP analysis reveals that MRR is the main predictor (53.3%) followed by particle size (22.9%), kind (20.7%) and W/B (2.8%). The integrated analyzes provide straightforward engineering insight into the variables that govern RPM compressive strength. MRR is the most significant factor with a strong negative correlation with CS ($r=-0.76$), negative SHAP contributions at higher levels of replacement and significant strength reductions across the response surfaces, highlighting the need to balance the incorporation of recycled powder with strength requirements. Particle size has a weak linear correlation with the target, but it has the second highest SHAP importance, indicating a nonlinear and mixture-dependent effect that is not well captured by conventional correlation analysis. The powder type (kind) used for recycling also induces unique category-dependent effects reflecting that different powder sources cannot be treated as mechanically equivalent. In contrast, W/B depicts the lowest global importance, but this should be referred to its relatively narrow distribution within the investigated dataset rather than to its intrinsic non-importance on the mortar strength. Interaction analyses consistently identify the particle size–MRR combination as the dominant pairwise effect in both SHAP and second-order Sobol analyses. However, the low Sobol S2 values indicate that CS variability is governed primarily by main effects, particularly MRR. Hence, MRR must be the primary consideration in RPM mix design, with particle size, type of recycled powder and their interactions being secondary indicators for performance optimization. These interpretations reflect model-learned relationships and should not be interpreted as direct causal mechanisms without independent experimental validation. The comparatively low SHAP importance of W/B should therefore be interpreted as a consequence of its restricted variability within the sampled range (0.35–0.55), rather than as evidence of intrinsically low physical importance.

5. Limitations and Future Research

The dataset contains 204 observations collected from the literature, which could limit the representativeness and diversity of mixture compositions and recycled powder properties. Independent testing and 10-fold cross-validation show that the model is robust in terms of internal generalizability, but the broader generalizability and practical utility of the model still need to be confirmed by external validation using completely independent experimental datasets that cover different material sources and laboratory conditions. The model is limited to kind, particle size, W/B and MRR, while RPM strength can be affected by other physicochemical, material and processing properties such as mineralogical and chemical composition, cement properties, curing condition, admixture content, particle morphology, specific surface area and processing history. Moreover, the limited distributions of some variables, especially W/B, can impact their apparent significance, suggesting that model interpretation and prediction should be confined to the parameter domain studied. Finally, SHAP and Sobol analyzes characterize sensitivity patterns and associations that are derived from the model, not direct physical causality. Future research should expand the database, include more physicochemical and processing variables and experimentally validate the identified effects and interactions. Additionally, incorporating multi-objective optimization, uncertainty quantification, and physics-informed learning can improve model generalizability and guide practical and sustainable design of RPM mixtures. The relatively small test subset limits the statistical resolution for distinguishing marginal differences among high-performing models. Future studies should employ repeated-split or bootstrap validation with confidence intervals to quantify predictive uncertainty. Future studies may also investigate feature-selection and data-simplification strategies to reduce effective dataset complexity, including approaches motivated by Kolmogorov complexity, which may facilitate more efficient learning and improve model generalization[50, 51].

6. Conclusion

This study developed an explainable, metaheuristic-optimized ensemble ML framework for predicting and interpreting the CS of RPM. ET, GB, and HGB were integrated through bagging, voting, and stacking and optimized using GA, PSO, DE, and GWO. Independent testing, 10-fold cross validation, error and stability analyzes; REC curves and normalized performance ranking were used to provide a complete assessment of model reliability.

The main results can be summarized as follows:

- The GA-optimized ET–GB–HGB stacking model achieved the highest integrated performance ranking, with R^2 values of 0.949, 0.966, and 0.947 and RMSE values of 2.149, 1.931, and 2.129 MPa for training, testing, and 10-fold cross-validation, respectively. However, its performance was closely comparable to the PSO-optimized and other leading configurations; therefore, the marginal numerical differences should not be interpreted as evidence of statistical superiority.
- SHAP and permutation importance consistently identified the global feature hierarchy as MRR > particle size > kind > W/B, with MRR emerging as the dominant predictor and contributing more than half of the global SHAP importance while exhibiting the strongest influence on individual predictions.
- SHAP dependence analysis revealed pronounced nonlinear and observation-dependent effects, particularly for MRR and particle size. The predominantly negative SHAP contribution at higher MRR levels identifies replacement rate as the principal factor governing the predicted compressive strength of RPM.
- Sobol global sensitivity analysis independently confirmed MRR as the dominant predictor, while second-order analysis identified particle size–MRR as the strongest pairwise interaction for both recycled powder categories. However, the relatively low S2 indices indicate that the predictive response is governed primarily by main effects rather than feature interactions.
- The response-surface analysis also indicated that CS was dependent on the mixture parameters non-linearly, with the magnitude and response patterns varying with the category of recycled powder.

Overall, the results indicate that the blend of heterogeneous ensemble learning, metaheuristic optimization, rigorous validation, SHAP explainability and Sobol sensitivity analysis provides an accurate and interpretable framework for the prediction of CS of RPM. MRR was always the most important design parameter and was emphasized as the main control parameter of the model. The presented framework offers a robust data-driven basis for sustainable RPM mixture design, but additional experimental and external validation is required for wider engineering application. In practice, the proposed framework can support batching plants and civil engineers by rapidly screening RPM mixture combinations, estimating compressive strength before experimental production, and identifying suitable ranges of MRR and particle size for more efficient and sustainable mix design.

Author Contributions

S.J.: Conceptualization, Methodology, Data Curation, Formal Analysis, Investigation, Visualization, Writing – Original Draft, Writing – Review & Editing.
H.H.: Investigation, Data Curation, Validation, Writing – Review & Editing.
K.A.W.S.: Investigation, Validation, Resources, Writing – Review & Editing.
All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data Availability Statement

The dataset used in this study is available from the corresponding author upon reasonable request.

Conflict of Interest Statement

The authors declare that there is no conflict of interest.

References

[1.] Spence, R. and H. Mulligan, Sustainable development and the construction industry. Habitat international, 1995. 19(3): p. 279–292.

[2.] Singh, A., et al., Use of recycled fine aggregates and recycled powders in sustainable recycled concrete. Journal of Building Engineering, 2023. 77: p. 107370.

[3.] De Luca, A., L. Chen, and K. Gharehbaghi, Sustainable utilization of recycled aggregates: Robust construction and demolition waste reduction strategies. International Journal of Building Pathology and Adaptation, 2021. 39(4): p. 666–682.

[4.] Nguyen, T.-D., Mechanical and durability performance of recycled aggregate concrete: Experimental and Predictive Studies. 2024.

[5.] Duan, Z., et al., Study on the essential properties of recycled powders from construction and demolition waste. Journal of Cleaner Production, 2020. 253: p. 119865.

[6.] Zhang, W., et al., The Utilization of Recycled Powder: A Critical Review. Buildings, 2026. 16(3): p. 649.

[7.] Tang, Q., et al., The utilization of eco-friendly recycled powder from concrete and brick waste in new concrete: A critical review. Cement and Concrete Composites, 2020. 114: p. 103807.

[8.] Bentegri, H., et al., Assessment of compressive strength of eco-concrete reinforced using machine learning tools. Scientific Reports, 2025. 15(1): p. 5017.

[9.] Dere, T., S. Jueyendah, and Z. Yaman, Machine Learning-Assisted Biomonitoring of Heavy Metal Accumulation in Pinus nigra Needles Across Urban, Industrial, and Pristine Sites in Adiyaman, Türkiye. Processes, 2026. 14(14): p. 2351.

[10.] Berti, F.G., et al. Advanced FEM-based assessment of rotational stiffness in HSS–UHPC composite cellular beams for lateral-distortional buckling. in Structures. 2026. Elsevier.

[11.] Jueyendah, S. and E. Ağcakoca, Explainable tree-based machine learning for sustainable mix optimization and compressive strength prediction of recycled aggregate fiber reinforced self-compacting concrete. Discover Materials, 2026.

[12.] Kim, B., et al., Ensemble machine learning-based approach for predicting of FRP–concrete interfacial bonding. Mathematics, 2022. 10(2): p. 231.

[13.] Mahmood, M.S., et al., Machine learning prediction and explainability analysis of high strength glass powder concrete using SHAP PDP and ICE. Scientific Reports, 2025. 15(1): p. 22089.

[14.] Jueyendah, S. and E. Ağcakoca, An Interpretable Ensemble Machine Learning Framework for Predicting the Ultimate Flexural Capacity of BFRP-Reinforced Concrete Beams. Polymers, 2026. 18(5): p. 601.

[15.] Jueyendah, S. and E. Ağcakoca, Explainable Ensemble Machine Learning for the Prediction and Optimization of Pozzolanic Concrete Compressive Strength. Polymers, 2026. 18(8): p. 933.

[16.] Jueyendah, S., E. Ağcakoca, and Z. Yaman, An interpretable hybrid machine learning and deep learning framework for predicting compressive strength of recycled powder mortar. Civ. Eng. Beyond Limits, 2026. 7: p. 11118.

[17.] Jueyendah, S. and I. Hollý, Integrating explainable hybrid machine learning and evolutionary optimization for high-fidelity prediction of ultra-high-performance concrete compressive strength. Array, 2026: p. 100774.

[18.] Ahmed Elsheikh, M.H., Sebghatullah Jueyendah , Abdelmalek H. Aljarah , Carlos Humberto Martins , Ghazi Shakah, Mohamed Ghalla , Viroon Kamchoom Real-time interpretable and cluster-stratified lightGBM framework for high-precision concrete strength prediction and instantaneous mixture optimization. Construction and Building Materials, 2026. 537: p. 146996.

[19.] Sobuz, M.H.R., et al., An explainable machine learning model for encompassing the mechanical strength of polymer-modified concrete. Asian Journal of Civil Engineering, 2024: p. 1–24.

[20.] Chaabene, W.B., et al., Machine learning prediction of mechanical properties of concrete: Critical review. 2020. 260: p. 119889.

[21.] Hosseinzadeh, M., M. Dehestani, and A.J.J.o.B.E. Hosseinzadeh, Prediction of mechanical properties of recycled aggregate fly ash concrete employing machine learning algorithms. 2023. 76: p. 107006.

[22.] Koya, B.P., et al., Comparative analysis of different machine learning algorithms to predict mechanical properties of concrete. 2022. 29(25): p. 4032–4043.

[23.] Samuel, A.L.J.I.J.o.r. and development, Some studies in machine learning using the game of checkers. 2000. 44(1.2): p. 206–226.

[24.] Xu, Y., et al., Artificial intelligence: A powerful paradigm for scientific research. 2021. 2(4).

[25.] Kongala, S.G.R., Explainable structural health monitoring (SHM) for damage classification based on vibration. 2024, University of Twente.

[26.] Javed, M.F., et al., Forecasting the strength of preplaced aggregate concrete using interpretable machine learning approaches. Scientific reports, 2024. 14(1): p. 8381.

[27.] Fei, Z., et al., Ensemble machine-learning-based prediction models for the compressive strength of recycled powder mortar. Materials, 2023. 16(2): p. 583.

[28.] Khan, K., et al., Evaluating the effectiveness of waste glass powder for the compressive strength improvement of cement mortar using experimental and machine learning methods. Heliyon, 2023. 9(5).

[29.] Amin, M.N., et al., Experimental and machine learning approaches to investigate the effect of waste glass powder on the flexural strength of cement mortar. PloS one, 2023. 18(1): p. e0280761.

[30.] Yang, J., et al., Experimental investigation and AI prediction modelling of ceramic waste powder concrete–An approach towards sustainable construction. Journal of Materials Research and Technology, 2023. 23: p. 3676–3696.

[31.] Bisciotti, A., et al., Estimating attached mortar paste on the surface of recycled aggregates based on deep learning and mineralogical models. Cleaner Materials, 2024. 11: p. 100215.

[32.] Kellouche, Y., et al., Prediction and optimization of the compressive strength of marble powder-based concrete using AI techniques: machine learning and metaheuristic approaches. Modeling Earth Systems and Environment, 2025. 11(6): p. 413.

[33.] Hasan, M.R., et al., Data-driven prediction of concrete strength by machine learning: hybrid-fiber-reinforced recycled aggregate concrete. World Journal of Engineering, 2025.

[34.] Dahish, H.A. and M.K. Alkharisi, Predicting the properties of polypropylene fiber recycled aggregate concrete using response surface methodology and machine learning. Buildings, 2025. 15(20): p. 3709.

[35.] Jueyendah, S., et al., Predicting the mechanical properties of cement mortar using the support vector machine approach. Construction and Building Materials, 2021. 291: p. 123396.

[36.] Jueyendah, S. and C.H. Martins, Optimal design of welded structure using SVM. Computational Engineering and Physical Modeling, 2024. 7(3): p. 84–107.

[37.] Jueyendah, S., et al., Comparative study of linear and non-linear ML algorithms for cement mortar strength estimation. Buildings, 2025. 15(16): p. 2932.

[38.] Hematibahar, M., C.H. Martins, and K. Momeni, Machine Learning Prediction of Fiber-Reinforced Concrete Using the SADRA Algorithm Based on Transcendental Philosophy. Journal of Soft Computing in Civil Engineering, 2026. 10(2): p.

[39.] Song, Q., et al., A novel self-compacting ultra-high performance fibre reinforced concrete (SCUHPFRC) derived from compounded high-active powders. Construction and Building Materials, 2018. 158: p. 883–893.

[40.] Ağcakoca, E., et al., Advanced Hybrid Modeling of Cementitious Composites Using Machine Learning and Finite Element Analysis Based on the CDP Model. Buildings, 2025. 15(17): p. 3026.

[41.] Siddique, R., P. Aggarwal, and Y.J.A.i.e.s. Aggarwal, Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks. 2011. 42(10): p. 780–786.

[42.] Guan, X., H. Burton, and T. Sabol, Python-based computational platform to automate seismic design, nonlinear structural model construction and analysis of steel moment resisting frames. Engineering Structures, 2020. 224: p. 111199.

[43.] Sarvade, S.M. and S.M. Pore. Use of python programming for interactive design of reinforced concrete structures. in National Conference on Exploring New Dimensions in Teaching Learning for Quality Education, Maharashtra, India. 2019.

[44.] Papadopoulos, S., et al., Evaluation of tree-based ensemble learning algorithms for building energy performance estimation. Journal of Building Performance Simulation, 2018. 11(3): p. 322–332.

[45.] Çiftçioğlu, A.Ö., et al., Machine learning based shear strength prediction in reinforced concrete beams using Levy flight enhanced decision trees. Scientific Reports, 2025. 15(1): p. 27488.

[46.] Zenko, B., L. Todorovski, and S. Dzeroski. A comparison of stacking with MDTs to bagging, boosting, and other stacking methods. in Proceedings of the First IEEE International Conference on Data Mining. 2001.

[47.] Erdal, H.I., Two-level and hybrid ensembles of decision trees for high performance concrete compressive strength prediction. Engineering Applications of Artificial Intelligence, 2013. 26(7): p. 1689–1697.

[48.] Wang, Z., et al., Optimizing machine learning techniques and SHapley Additive explanations (SHAP) analysis for the compressive property of self-compacting concrete. Materials Today Communications, 2024. 39: p. 108804.

[49.] Ngo, A.Q., L.Q. Nguyen, and V.Q. Tran, Developing interpretable machine learning-Shapley additive explanations model for unconfined compressive strength of cohesive soils stabilized with geopolymer. Plos one, 2023. 18(6): p. e0286950.

[50.] Bolón-Canedo, V. and B. Remeseiro, Feature selection in image analysis: a survey: V. Bolón-Canedo, B. Remeseiro. Artificial Intelligence Review, 2020. 53(4): p. 2905–2931.

[51.] Kabir, H. and N. Garg, Machine learning enabled orthogonal camera goniometry for accurate and robust contact angle measurements. Scientific reports, 2023. 13(1): p. 1497.

How to Cite This Article

Jueyendah, S., Hawadi, H., and Sediqi, K.A.W., "Explainable Metaheuristic-Optimized Ensemble Machine Learning for Compressive Strength Prediction and Sustainable Mix Design of Recycled Powder Mortar," Civil Engineering Beyond Limits, 1(2027),11173.
https://doi.org/10.36937/cebel.2027.11173