Civil Engineering Beyond Limits 1 (2026) 11118
Civil Engineering Beyond Limits
journal home: www.acapublishing.com/journals/1/cebel
Research Article

An Interpretable Hybrid Machine Learning and Deep Learning Framework for Predicting Compressive Strength of Recycled Powder Mortar

Sebghatullah Jueyendah[1], Elif Ağcakoca[1,*], Zeynep Yaman[1,2]
[1]Department of Civil Engineering, Sakarya University, Esentepe Campus, 6 Serdivan, Sakarya, 54050, Turkey
[2]A International Vision University Engineering Faculty Civil Engineering Department Gostivar North, Macedonia

Keywords

Mix design optimization, Recycled powder mortar, Interpretable machine learning, Compressive strength prediction, Sustainable cementitious materials.

Abstract

The accurate prediction of compressive strength (CS) is a fundamental requirement for structural safety, yet conventional empirical models remain limited in their ability to capture the complex nonlinear response of cementitious materials. Hybrid machine learning (ML) and deep learning (DL) models are investigated for predicting the CS of recycled powder mortar (RPM) using 10-fold cross-validation. Material type (classified as kind), particle size, water/binder (W/B), and mass replacement ratio (MRR) are used as input variables, with CS as the target response. Model performance was quantified using RMSE, R^2, MAE, and MAPE, and analysis of variance (ANOVA) was conducted to examine the statistical significance of the input variables. Based on 10-fold cross-validation, the SVR–PSO model exhibited the highest average predictive performance, attaining $R^2 = 0.950$ and RMSE = 2.042 MPa, thereby demonstrating superior accuracy and reliability relative to the other evaluated models. The interpretable hybrid ML–DL framework achieved high accuracy and robust generalization across folds, with error and stability analyses confirming unbiased predictions. MRR is the dominant factor controlling CS, with suggested optimal ranges; this influence is consistently supported by ALE, PDP, SHAP, and ICE analyses. Feature importance and PDP demonstrate Kind has a secondary yet meaningful effect on CS, reflecting differences in composition, morphology, and reactivity, with its subtler role in ALE and ICE analyses arising from its categorical and indirect mechanistic influence on hydration and pozzolanic activity. The framework offers a robust, interpretable tool for predicting CS, supporting data-driven mix design and sustainable mortar optimization.

1. Introduction

CS represents a key material parameter controlling the performance, durability, and serviceability of cementitious and composite materials [1]. Accurate estimation of CS is essential to ensure structural safety, durability, and cost-effectiveness in civil and structural engineering practice [2]. Although widely adopted, conventional empirical and regression approaches remain constrained by simplifying assumptions and a limited capacity to capture nonlinear material behavior [3]. Given the inherently nonlinear nature of material behavior, RPM produced from construction waste represents a sustainable alternative to conventional cement-based materials [4]. The incorporation of recycled powder into mortar enhances resource efficiency while reducing the environmental impact associated with cement production [5]. The inherent variability of recycled powders results in substantial uncertainty in CS of RPM, underscoring the need for accurate predictive approaches to ensure structural reliability and sustainable construction [6]. The physicochemical properties of recycled powders are intrinsically determined by their source materials, such as concrete, brick, or ceramic waste[7]. Such intrinsic variations govern particle packing, filler effectiveness, and potential pozzolanic activity, consequently influencing microstructural evolution of the cementitious matrix and the interfacial transition zone profiles. Thus, material type (classified as 'kind') fundamentally governs the mechanical behavior of RPM [8]. Drawing on a large body of experimental data, this study applies ML techniques to predict material strength by modeling complex nonlinear relationships [9]. Support vector regression (SVR) and artificial neural networks (ANNs) represent two of the most extensively used ML approaches for CS prediction [10]. The predictive performance of ML models is strongly influenced by hyperparameter selection, which motivates the use of particle swarm optimization (PSO) and genetic algorithms (GA) in hybrid frameworks [11]. Long short-term memory (LSTM) and convolutional neural network (CNN) architectures improve feature representation and predictive performance, with attention mechanisms highlighting key features [12]. Although physics-informed neural networks (PINNs) embed physical constraints within data-driven learning frameworks, consistent model comparison remains a challenge [13]. Many existing studies employ single train–test splits prone to sampling bias; in contrast, 10-fold cross-validation provides a more reliable assessment by mitigating bias and variance [14]. Statistical assessment and explainable artificial intelligence (AI) methodologies play a critical role in elucidating model behavior and strengthening confidence in machine learning–based predictions [15]. Feature importance, accumulated local effects (ALE), partial dependence plots (PDP), and individual conditional expectation (ICE) analyses offer complementary perspectives on the global and local effects of input variables [16]. Few studies have integrated hybrid optimization techniques, DL, explainable AI, and rigorous validation frameworks for CS prediction [17]. To address this gap, a unified framework is proposed for accurate, robust, and interpretable compressive strength prediction. Erdal [18] applied hybrid decision tree ensemble techniques for high-performance concrete CS prediction, highlighting the advantages of ensemble learning. Moges et al. [19] proposed SVR and ANN models for spray drift prediction, highlighting the enhanced capability of ML methods to represent nonlinear behavior relative to traditional regression techniques. Fei et al. [20] proposed ensemble machine learning models for recycled powder mortar CS prediction, demonstrating their potential for data-driven mix design optimization. Amin et al. [21] examined the influence of waste glass powder on cement mortar flexural strength through combined experimental and ML methodologies. Yang et al. [22] developed AI–based models to predict the mechanical properties of ceramic waste powder

*Corresponding Author: elifd@sakarya.edu.tr
Received 30 Jan 2026; Revised 01 Feb 2026; Accepted 01 Feb 2026
https://doi.org/10.36937/cebel.2026.11118

concrete, supporting sustainable construction practices. Açıkkar and Altunkol [23] introduced a hybrid particle swarm optimization–grid search (PSO–GS) strategy for SVR, demonstrating enhanced prediction accuracy via effective hyperparameter optimization. Parsaeimehr et al. [24] proposed a hybrid CNN–LSTM framework for improved feature extraction in entity identification tasks, highlighting the benefits of DL model integration. Martens and Dardenne [25] highlighted the necessity of rigorous validation and verification procedures for small datasets to achieve reliable and unbiased regression model evaluation. Bisciotti et al. [26] applied DL and mineralogical models to quantify attached mortar on recycled aggregates, supporting efficient evaluation of construction and demolition waste. Kellouche et al. [27] applied AI frameworks combining ML and metaheuristic methods for the prediction and optimization of marble powder concrete CS. Hasan et al. [28] applied ML techniques to predict the CS of hybrid fiber–reinforced recycled aggregate concrete, supporting robust mix design optimization. Jueyendah et al.[29] showed that SVM models employing a RBF kernel achieve effective prediction of cement mortar strength. Jueyendah and Martins [30] proposed a hybrid SVR–RBF model for optimized design and performance assessment of welded structure. Jueyendah et al.[31] demonstrated the superior predictive accuracy of nonlinear machine learning methods over linear models for cement mortar strength prediction. Hematibahar et al. [32] developed a ML framework based on SADRA for the prediction of fiber-reinforced concrete properties. Song et al. [33] developed self-compacting ultra-high-performance fiber-reinforced concrete mixtures incorporating high-activity powders. Ağcakoca et al. [34] developed a hybrid ML–finite element (FE) model based on the concrete damage plasticity (CDP) approach to improve the prediction of cementitious composite behavior. Siddique et al. [35] applied ANN models to predict the CS of self-compacting concrete. Minh et al. [36] provided a comprehensive review of explainable artificial intelligence (XAI) approaches, highlighting methods that improve interpretability and reliability in artificial intelligence systems. Although research on predicting RPM compressive strength is limited, classical ML approaches show poor predictive performance and fail to capture nonlinear interactions among key mix parameters. This study addresses these limitations by proposing a hybrid ML–DL framework with PSO and GA optimization, constituting the first such approach for RPM CS prediction. CNN–LSTM and LSTM–Attention sequence models extract latent features and represent complex hierarchical interactions among input variables. Ablation studies demonstrate improved accuracy and robustness of these architectures on tabular data. Ten-fold cross-validation combined with statistical analysis and explainable AI yields interpretable, generalizable models for optimized, sustainable mix design.

2. Methodology

A dataset of 204 recycled powder mortar (RPM) samples [20] is used, where material type (classified as 'kind'), mass replacement ratio (MRR), water/binder (W/B), and particle size serve as input variables, and CS is defined as the response. Input variables were min–max scaled to [0, 1], with parameters fitted separately within each training fold during ten-fold cross-validation to avoid data leakage and maintain unbiased model assessment. After cleaning and normalization, the dataset was evaluated using 10-fold cross-validation, and hybrid ML models combining GA and PSO with SVR, ANN, and RF algorithms were constructed. Advanced DL architectures, including LSTM–attention, CNN–LSTM, autoencoder-based feature extraction, and physics-informed neural networks (PINNs), were utilized to capture nonlinear relationships and enhance model performance and interpretability. Model performance was evaluated using root mean square error (RMSE), coefficient of determination (R^2), mean absolute error (MAE), and mean absolute percentage error (MAPE), while feature importance, accumulated local effects (ALE), partial dependence plots (PDP), and individual conditional expectation (ICE) plots were employed to interpret variable contributions and characterize nonlinear effects. By balancing the bias–variance tradeoff, the models capture complex nonlinear relationships while ensuring robustness. Categorical encoding of 'kind', 10-fold cross-validation, early stopping, dropout, and hyperparameter optimization were employed to mitigate overfitting. Within the PINNs framework, physical relationships linking recycled powder type, porosity, and compressive strength are embedded as constraints, enabling the model to integrate data-driven insights with fundamental material behavior. Data analysis and model development were conducted in Python 3.12 using an integrated development environment, enabling efficient preprocessing, training, and evaluation. Python, supported by scientific libraries such as NumPy and Pandas, provided a robust and efficient computational environment for data processing, model development, evaluation, and visualization in this study [37, 38]. Figures 1–3 illustrate the relationships between the input variables and compressive strength, their influence on CS, and the proposed workflow, while ANOVA results (Table 1) assess their statistical significance. Table 1 presents descriptive statistics for RPM input variables and CS, whereas analysis of variance (ANOVA) was employed to assess statistical significance. The results indicate that the MRR exerts the most pronounced effect on CS, followed by material type (classified as 'kind'), both showing significant influence ($p < 0.05$). ANOVA was used exclusively for preliminary feature screening to identify influential inputs; inference was not performed, since the ML and DL models inherently account for nonlinearities and complex interactions.

Table 1. Descriptive statistics and ANOVA results for the input variables and CS

Variable	Mean	Minimum	Median	Maximum	F-Value	P-Value
Kind	0.392	0.000	0.000	1.000	37.340	0.000
Size (µm)	27.74	0.000	20.830	103.600	2.860	0.092
W/B	0.492	0.350	0.500	0.550	0.360	0.550
MRR (%)	23.816	0.000	30.000	60.000	330.110	0.000
CS (MPa)	36.309	5.400	37.000	56.800	–	–

These patterns are in agreement with the ANOVA findings presented in Table 1, which indicate that MRR is the dominant factor influencing CS. Figure 1 presents the quantified relative contributions of the input variables to the CS of RPM. Material type (kind) and MRR display clear monotonic relationships, highlighting their dominant role in strength development. ANOVA results further demonstrate the dominant effects of MRR ($F = 330.11$) and kind ($F = 37.34$) on CS, with both factors exhibiting statistical significance. Figure 2 shows scatter relationships between the input variables and CS of RPM, with strong nonlinear trends for kind and MRR and weaker correlations for particle size and W/B. Figure 3 illustrates the structured workflow of the study, including data preprocessing, feature selection, and hybrid ML and DL model development. Model performance was assessed through resampling techniques and multiple evaluation metrics, with subsequent interpretability analysis and validation to ensure model reliability and practical applicability. The kind variable, representing material type, is binary (0 or 1) and was normalized to [0, 1] for model training. This encoding preserves its categorical meaning while enabling all models to capture nonlinear effects and interactions without implying ordinality. Sequence-based DL architectures, including CNN–LSTM and LSTM–Attention, were selected to investigate their capacity for modeling complex nonlinear relationships and to provide a comprehensive evaluation across diverse predictive paradigms.

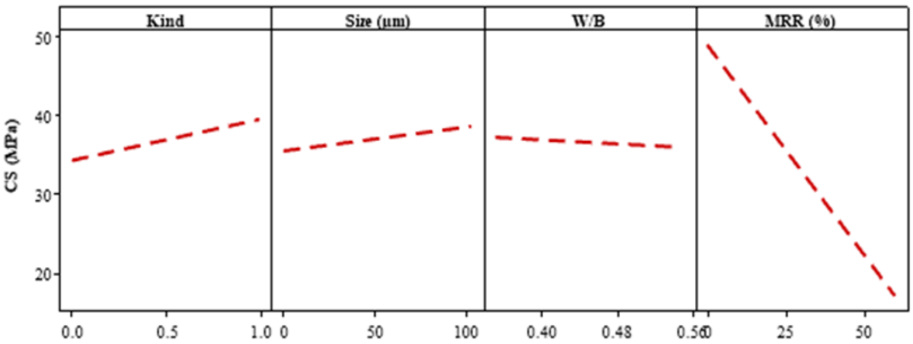

Figure 1. Contribution of input variables to the CS of RPM

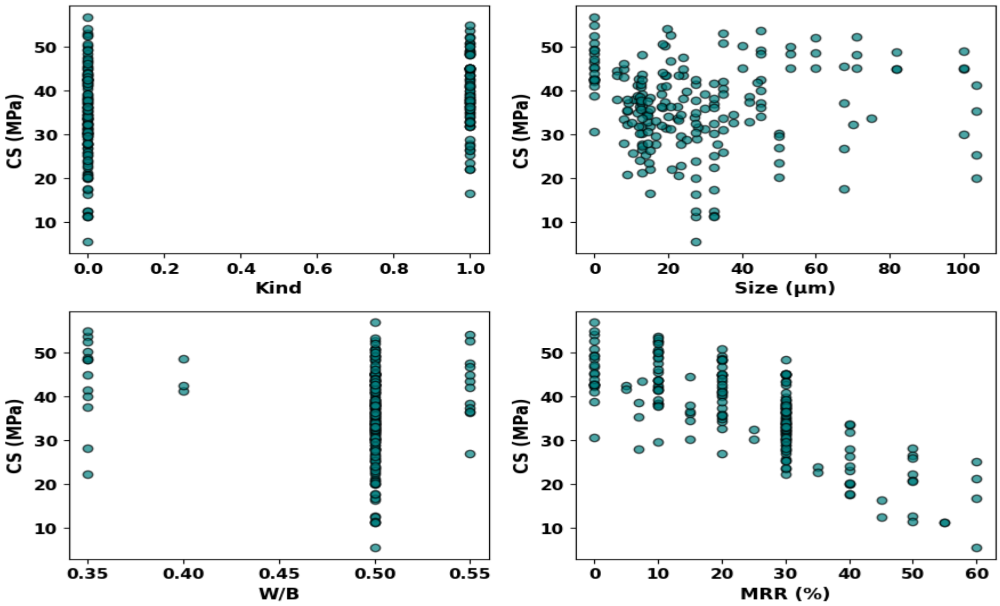

Figure 2. Input variable effects on RPM compressive strength

2.1. Hybrid machine learning models

By leveraging complementary algorithm strengths, hybrid ML models achieve superior predictive accuracy and generalization capability. Rather than relying on a single learning approach, these models integrate optimization or search-based techniques with learning algorithms to improve convergence and generalizability in complex engineering problems.

2.2. Deep learning approaches

Deep learning (DL) techniques have gained prominence for modeling highly nonlinear behaviors in engineering systems. Convolutional neural networks, recurrent neural networks, and physics-informed neural networks provide powerful feature-learning capabilities, especially in data-rich experimental and numerical settings.

2.3. Partial dependence analysis

Partial dependence analysis (PDA), illustrated through partial dependence plots (PDPs), evaluates the marginal effect of one or two input variables on model predictions by averaging over all other features [39]. PDPs capture nonlinear and threshold effects of input variables on predictions, providing robust global trend insights under the feature-independence assumption. In civil engineering, PDPs guide the selection of optimal material ranges, enabling performance-driven mix optimization [40].

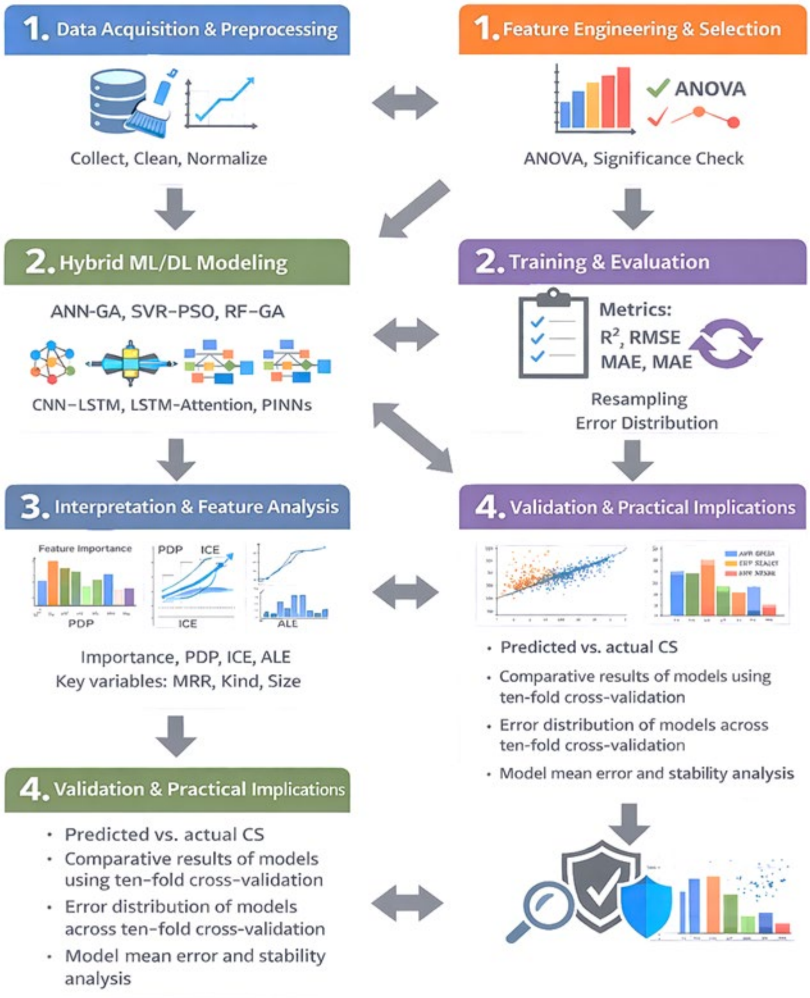

Figure 3. Proposed methodology workflow

2.4. Accumulated Local Effects Analysis

Accumulated local effects (ALE) plots quantify feature impacts by aggregating local variations in predictions, accounting for feature correlations and reducing extrapolation bias compared with PDPs [39]. ALE plots offer unbiased, interpretable assessments of feature effects for datasets with correlated variables. ALE analysis reveals how material variations impact structural performance, guiding robust mix design and material selection in civil engineering[40].

2.5. Individual Conditional Expectation Analysis

Individual conditional expectation (ICE) plots illustrate each sample's response to a specific feature, capturing heterogeneity in feature effects and identifying subgroups with distinct behavioral patterns [41]. ICE plots reveal local effects masked in PDPs, providing more detailed interpretability of predictions. ICE plots reveal how material variability affects individual mixtures, supporting reliable predictions for specific structural designs [42].

2.6. Statistical Parameters for Model Evaluation

Regression models were evaluated with multiple criteria to ensure robust and reliable predictions [43]. The coefficient of determination (R^2) quantifies the proportion of variance in the dependent variable explained by the independent variables, with higher values indicating better predictive performance. The root means square error (RMSE) and mean absolute error (MAE) measure the average deviation and absolute difference between predicted and measured values, with lower values indicating higher predictive accuracy. The mean absolute error (MAE) treats all errors equally, the root mean square error (RMSE) penalizes larger deviations more heavily, and the mean absolute percentage error (MAPE) assesses prediction accuracy relative to measured values, providing a scale-independent measure of model performance. Collectively, these evaluation criteria offer a comprehensive and reliable framework for assessing regression model performance in civil engineering applications [44]. The regression model was evaluated using R^2, RMSE, MAE, and MAPE (Eqs. 1–4).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{m,i} - y_{p,i})^2}{(y_{m,i} - y_a)^2} \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{m,i} - y_{p,i})^2} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{m,i} - y_{p,i}\right| \tag{3}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_{m,i} - y_{p,i}}{y_{m,i}}\right| \tag{4}$$

The parameters ym and yp represent the measured and predicted values, respectively, while ya denote the mean of observed values used as a reference for model accuracy. In these equations, n indicates the total dataset size, while p corresponds to the number of independent variables (predictors) included in the regression model. These metrics jointly assess the model's prediction quality and reliability.

3. Results and Discussion

Table 2 summarizes GA/PSO hyperparameters, learning rates, DL architectures, and PINNs physical constraints for the hybrid ML–DL models, supporting reproducibility, model comparability, and suitable complexity.

Table 2. Hybrid ML and DL model hyperparameters

Model	Key Hyperparameters	Typical Values / Notes
ANN-GA	Hidden layers, neurons, learning rate, GA population size, generations	1–3 layers, 10–64 neurons, LR=0.001, pop=50, gen=100
ANN-PSO	Hidden layers, neurons, learning rate, PSO particles, iterations	1–3 layers, 10–64 neurons, LR=0.001, particles=30, iter=100
SVR-GA	Kernel, C, epsilon, GA population size, generations	RBF kernel, C=1, ε=0.1, pop=50, gen=100
SVR-PSO	Kernel, C, epsilon, PSO particles, iterations	RBF kernel, C=1, ε=0.1, particles=30, iter=100
RF-GA	n_estimators, max_depth, GA population size, generations	100 trees, max_depth=None, pop=50, gen=100
CNN-LSTM	Conv layers, filters, kernel size, LSTM units, learning rate	2 conv layers, 32–64 filters, kernel=3, LSTM units=50, LR=0.001
LSTM-Attention	LSTM units, attention heads, learning rate, epochs, batch size	LSTM units=50–100, heads=4, LR=0.001, epochs=100, batch=32
Autoencoder-ML	Encoder/decoder layers, latent dim, learning rate, ML model parameters	2 layers, latent dim=16–32, LR=0.001, ML: RF or SVR
PINNs	Network layers, neurons, activation, learning rate, PDE parameters	3–5 layers, 50 neurons, tanh, LR=0.001, PDE-specific
ML-RSM	ML model parameters, RSM design levels	RF or ANN, design: factorial or CCD, tuning as nee

Table 3 shows 10-fold cross-validation results for all hybrid ML and DL models, including R², MAE, RMSE, and MAPE per fold. Optimal performance for the ANN–PSO model was observed in fold 8 (R^2 = 0.967, MAE = 1.820 MPa, RMSE = 2.094 MPa, MAPE = 0.051), while the SVR–GA model exhibited its highest predictive accuracy in fold 3 (R^2 = 0.968, MAE = 1.545 MPa, RMSE = 1.805 MPa, MAPE = 0.042). ANN–PSO achieved optimal performance in fold 8 (R^2 = 0.967, MAE = 1.820 MPa, RMSE = 2.094 MPa, MAPE = 0.051), whereas SVR–GA reached its highest accuracy in fold 3 (R^2 = 0.968, MAE = 1.545 MPa, RMSE = 1.805 MPa, MAPE = 0.042). For SVR–PSO, fold 8 achieved the highest predictive performance (R^2 = 0.977, RMSE = 1.720 MPa, MAE = 1.350 MPa, MAPE = 0.038), whereas RF–GA reached its maximum performance in fold 8 (R^2 = 0.963, RMSE = 2.211 MPa, MAE = 1.884 MPa, MAPE = 0.051). Among the DL architectures, CNN–LSTM achieved its best performance in fold 8 (R^2 = 0.963, RMSE = 2.191 MPa, MAE = 1.826 MPa, MAPE = 0.053), whereas LSTM–attention reached optimal performance in fold 2 (R^2 = 0.970, RMSE = 1.883 MPa, MAE = 1.581 MPa, MAPE = 0.047), demonstrating superior predictive accuracy.

Table 3. performance comparison of hybrid ML and DL models using 10-fold cross-validation

Model	Fold	R2	RMSE	MAE	MAPE
ANN-GA	1	0.965	1.851	1.399	0.040
ANN-GA	2	0.960	2.160	1.749	0.050
ANN-GA	3	0.948	2.298	1.966	0.053
ANN-GA	4	0.949	2.198	1.911	0.051
ANN-GA	5	0.927	2.027	1.631	0.046
ANN-GA	6	0.955	1.992	1.693	0.044
ANN-GA	7	0.942	2.032	1.713	0.050
ANN-GA	8	0.970	1.988	1.603	0.043
ANN-GA	9	0.925	2.397	2.081	0.055
ANN-GA	10	0.932	2.281	1.936	0.053

Average		0.947	2.122	1.768	0.048
ANN-PSO	1	0.959	2.001	1.643	0.046
ANN-PSO	2	0.959	2.204	1.941	0.055
ANN-PSO	3	0.956	2.109	1.735	0.049
ANN-PSO	4	0.946	2.252	1.851	0.048
ANN-PSO	5	0.916	2.167	1.737	0.047
ANN-PSO	6	0.952	2.059	1.582	0.039
ANN-PSO	7	0.948	1.935	1.599	0.050
ANN-PSO	8	0.967	2.094	1.820	0.051
ANN-PSO	9	0.952	1.923	1.689	0.046
ANN-PSO	10	0.928	2.351	1.909	0.051
Average		0.948	2.110	1.751	0.048
SVR-GA	1	0.957	2.066	1.725	0.047
SVR-GA	2	0.963	2.083	1.754	0.049
SVR-GA	3	0.968	1.805	1.545	0.042
SVR-GA	4	0.945	2.268	1.863	0.047
SVR-GA	5	0.922	2.088	1.725	0.048
SVR-GA	6	0.945	2.209	1.779	0.045
SVR-GA	7	0.937	2.123	1.786	0.054
SVR-GA	8	0.955	2.435	2.078	0.062
SVR-GA	9	0.945	2.045	1.801	0.050
SVR-GA	10	0.947	2.013	1.720	0.046
Average		0.948	2.114	1.777	0.049
SVR-PSO	1	0.964	1.877	1.594	0.044
SVR-PSO	2	0.966	1.995	1.649	0.046
SVR-PSO	3	0.958	2.073	1.631	0.044
SVR-PSO	4	0.943	2.324	1.857	0.046
SVR-PSO	5	0.926	2.032	1.724	0.047
SVR-PSO	6	0.952	2.064	1.606	0.040
SVR-PSO	7	0.949	1.917	1.687	0.056
SVR-PSO	8	0.977	1.720	1.350	0.038
SVR-PSO	9	0.907	2.661	2.365	0.063
SVR-PSO	10	0.959	1.760	1.430	0.039
Average		0.950	2.042	1.689	0.046
RF-GA	1	0.951	2.187	1.855	0.052
RF-GA	2	0.960	2.159	1.905	0.056
RF-GA	3	0.956	2.126	1.759	0.046
RF-GA	4	0.956	2.041	1.664	0.045
RF-GA	5	0.928	2.012	1.773	0.051
RF-GA	6	0.934	2.424	2.068	0.056
RF-GA	7	0.946	1.957	1.733	0.050
RF-GA	8	0.963	2.211	1.884	0.051
RF-GA	9	0.956	1.844	1.610	0.045
RF-GA	10	0.944	2.067	1.717	0.046
Average		0.949	2.103	1.797	0.050
CNN-LSTM	1	0.959	2.005	1.713	0.048
CNN-LSTM	2	0.936	2.743	2.442	0.068
CNN-LSTM	3	0.937	2.546	2.095	0.054
CNN-LSTM	4	0.946	2.258	1.896	0.051
CNN-LSTM	5	0.905	2.312	2.079	0.060
CNN-LSTM	6	0.933	2.445	1.936	0.050
CNN-LSTM	7	0.942	2.043	1.793	0.054
CNN-LSTM	8	0.963	2.191	1.826	0.053
CNN-LSTM	9	0.935	2.228	1.919	0.051
CNN-LSTM	10	0.925	2.400	2.141	0.056
Average		0.938	2.317	1.984	0.055
LSTM-Attention	1	0.958	2.035	1.665	0.047
LSTM-Attention	2	0.970	1.883	1.581	0.047
LSTM-Attention	3	0.945	2.364	2.035	0.053
LSTM-Attention	4	0.951	2.154	1.731	0.046
LSTM-Attention	5	0.889	2.496	2.174	0.061
LSTM-Attention	6	0.934	2.417	1.972	0.053
LSTM-Attention	7	0.952	1.851	1.491	0.045
LSTM-Attention	8	0.960	2.279	2.014	0.056
LSTM-Attention	9	0.933	2.266	1.953	0.051
LSTM-Attention	10	0.948	1.983	1.635	0.045
Average		0.944	2.173	1.825	0.050
Autoencoder-ML	1	0.934	2.554	2.175	0.059
Autoencoder-ML	2	0.954	2.332	2.013	0.056
Autoencoder-ML	3	0.948	2.309	1.912	0.050
Autoencoder-ML	4	0.954	2.072	1.596	0.041
Autoencoder-ML	5	0.924	2.061	1.658	0.045
Autoencoder-ML	6	0.939	2.316	2.037	0.052
Autoencoder-ML	7	0.952	1.854	1.580	0.050
Autoencoder-ML	8	0.978	1.696	1.365	0.045
Autoencoder-ML	9	0.934	2.237	1.888	0.052
Autoencoder-ML	10	0.936	2.202	1.639	0.042

Average		0.945	2.163	1.786	0.049
PINNs	1	0.951	2.184	1.939	0.055
PINNs	2	0.963	2.078	1.673	0.046
PINNs	3	0.957	2.084	1.773	0.050
PINNs	4	0.951	2.136	1.713	0.042
PINNs	5	0.923	2.082	1.780	0.049
PINNs	6	0.955	1.998	1.511	0.039
PINNs	7	0.940	2.077	1.752	0.054
PINNs	8	0.966	2.123	1.741	0.049
PINNs	9	0.935	2.221	1.805	0.047
PINNs	10	0.937	2.199	1.988	0.054
Average		0.948	2.118	1.767	0.048
ML-RSM	1	0.941	2.413	2.041	0.058
ML-RSM	2	0.949	2.439	2.101	0.058
ML-RSM	3	0.962	1.969	1.636	0.046
ML-RSM	4	0.955	2.065	1.846	0.048
ML-RSM	5	0.930	1.981	1.719	0.049
ML-RSM	6	0.950	2.115	1.726	0.045
ML-RSM	7	0.947	1.940	1.771	0.055
ML-RSM	8	0.954	2.454	2.063	0.053
ML-RSM	9	0.944	2.070	1.723	0.046
ML-RSM	10	0.946	2.025	1.698	0.045
Average		0.948	2.147	1.832	0.050

The autoencoder–ML model achieved its highest predictive performance in fold 8 (R^2 = 0.978, RMSE = 1.696 MPa, MAE = 1.365 MPa, MAPE = 0.045). The PINNs model achieved its optimal performance in fold 2 (R^2 = 0.963, RMSE = 2.078 MPa, MAE = 1.673 MPa, MAPE = 0.046), whereas the ML–RSM model attained its best results in fold 3 (R^2 = 0.962, RMSE = 1.969 MPa, MAE = 1.636 MPa, MAPE = 0.046). Overall, the autoencoder–ML model in fold 8 demonstrated the best performance across all models and folds, attaining the highest R^2 and the lowest error metrics. Overall, the autoencoder–ML model in fold 8 demonstrated the best performance across all models and folds, attaining the highest R^2 and the lowest error metrics. Figure 4 presents the comparison between predicted and measured CS values for the top-performing fold of each model. For a robust and unbiased assessment, performance metrics are reported as the mean values across all ten cross-validation folds. While the Autoencoder–ML model reached a maximum R^2 of 0.978 in fold 8, its fold-averaged metrics (R^2 = 0.945, RMSE = 2.163, MAE = 1.786) indicate consistent accuracy and reliable generalization. The PINNs were guided by fundamental material behavior, linking recycled powder type, porosity, and compressive strength, thereby combining mechanistic insight with data-driven learning to improve interpretability and predictive reliability. Future studies will strengthen model evaluation by reporting median values and interquartile ranges across cross-validation folds to better capture performance variability. Figure 4 illustrates the predicted versus measured CS values for the top-performing fold of each model, where the majority of points lie near the 1:1 line, reflecting high predictive fidelity. The autoencoder–ML model in fold 8 achieved the highest performance (R^2 = 0.978, RMSE = 1.696 MPa), followed closely by SVR–PSO (R^2 = 0.977, RMSE = 1.720 MPa), with both models demonstrating excellent agreement with the measured values. The autoencoder–ML model in fold 8 achieved the best performance (R^2 = 0.978, RMSE = 1.696 MPa), followed by SVR–PSO (R^2 = 0.977, RMSE = 1.720 MPa), with both models showing excellent agreement with the measured values. Table 3 summarizes 10-fold cross-validation results, allowing identification of each model's best-performing fold and elucidating variations due to algorithmic behavior, data partitioning, inherent data patterns, and hyperparameters, thereby supporting robust and transparent evaluation. Figure 6 presents a comparative evaluation of all models, illustrating variations in predictive performance across the 10 folds of cross-validation. Each model demonstrates fold-dependent behavior, showing sensitivity to data partitioning, yet overall performance remains clustered. This similarity in performance indicates that all models exhibit comparable predictive accuracy, with only minor variations in errors.

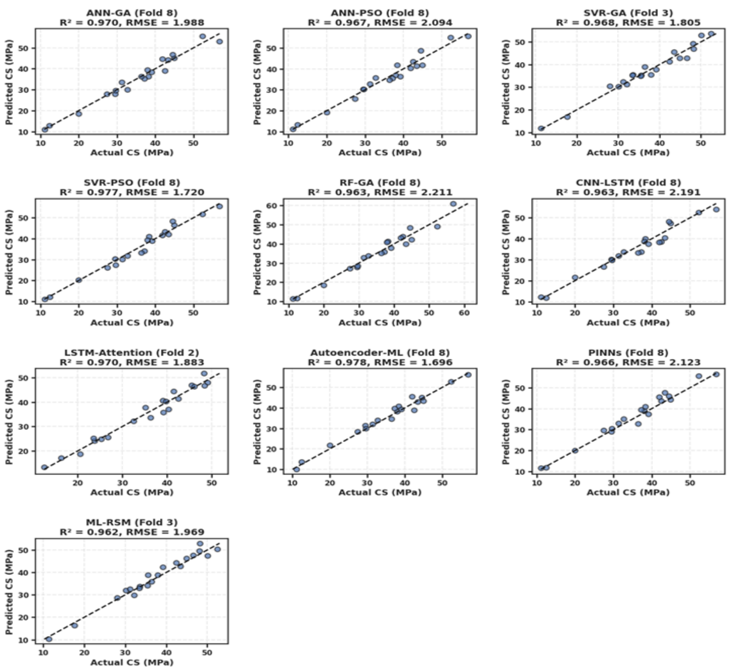

Figure 4. predicted and measured CS for the best-performing fold of each model

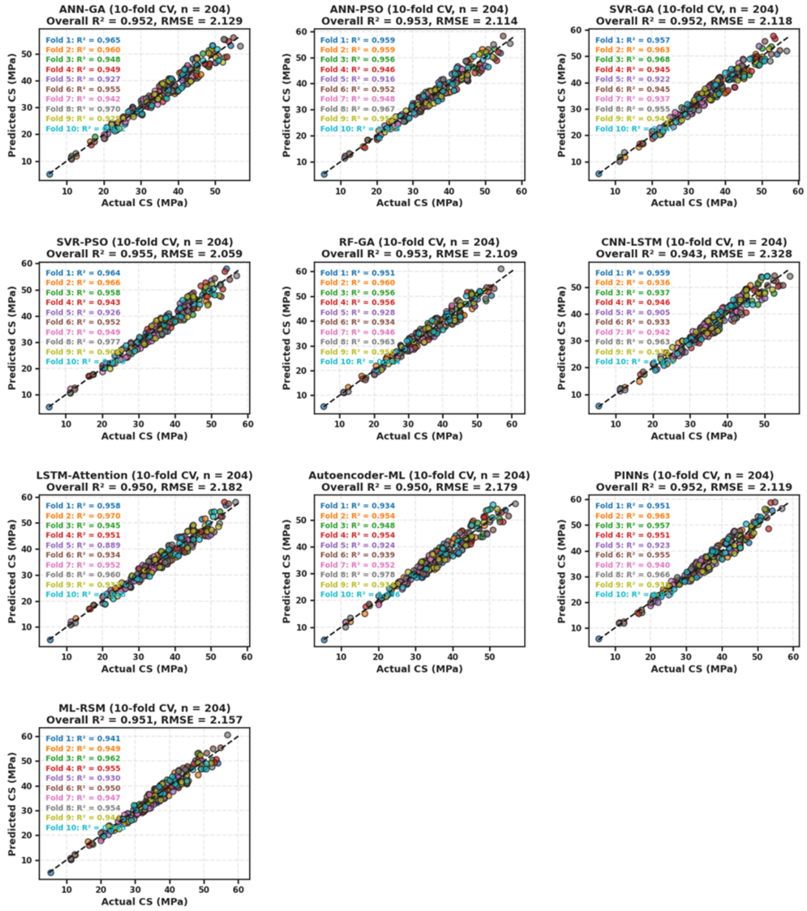

Figure 5. Predicted vs. measured CS across all folds for all models

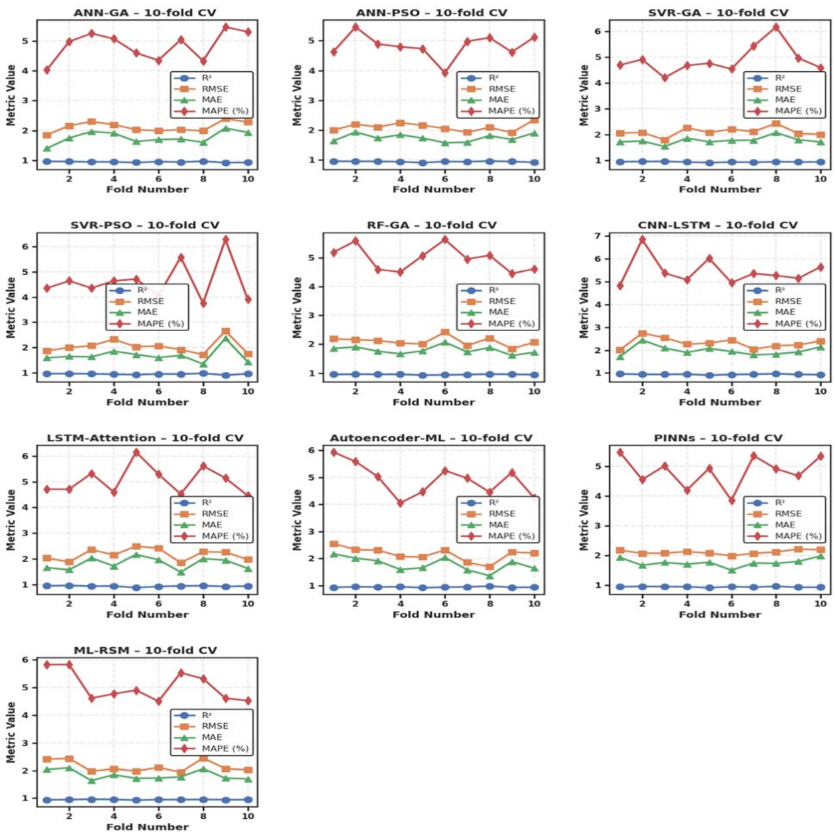

Figure 6. Comparative results of models using 10-fold cross-validation

Figure 7 shows that the autoencoder–ML model in fold 8 exhibits the lowest prediction error among all models, while the residuals remain symmetrically distributed around zero, confirming the absence of systematic bias. Furthermore, hybrid ML models display a relatively narrower error range compared with DL models, indicating greater robustness and reliability.

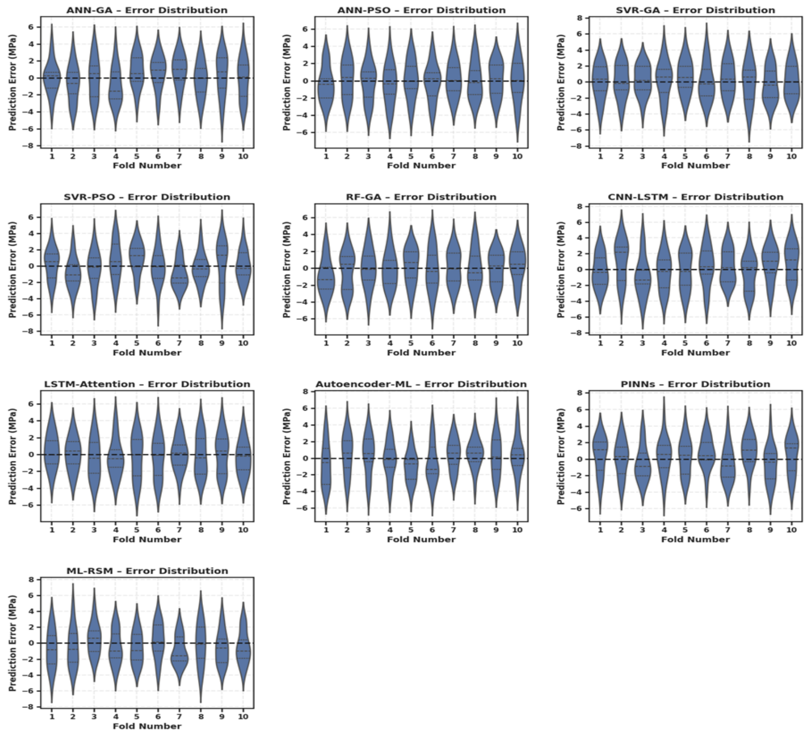

Figure 7. Error distribution of models across 10-fold cross-validation

Figure 8 illustrates the mean multivariate prediction error and its corresponding variability (± standard deviation) for all examined models. Mean errors closely centered around zero across folds indicate no systematic bias and highlight the reliability of the predictions. Hybrid ML models, especially ANN–PSO, RF–GA, and autoencoder–ML, demonstrate stable and consistent performance across folds, reflecting high robustness. DL models, especially CNN–LSTM, exhibit increased variability, indicating greater sensitivity to how the data are partitioned. To ensure robust and rigorous model evaluation, error distributions and fold-wise mean ± standard deviation was analyzed across all ten folds (Figures 7 and 8). Hybrid ML–DL models exhibited consistent predictions, characterized by near-zero median errors and low variability, thereby reducing potential reporting bias. With prediction errors confined to ±5 MPa and consistent fold-wise performance, the models demonstrate reliable and practically relevant guidance for recycled powder mortar mix design. Figure 9 illustrates the 10-fold cross-validation procedure, in which the dataset is divided into ten equal subsets. Each subset is used once for validation while the remaining nine subsets serve for training, ensuring full utilization of the data and reducing sampling bias.

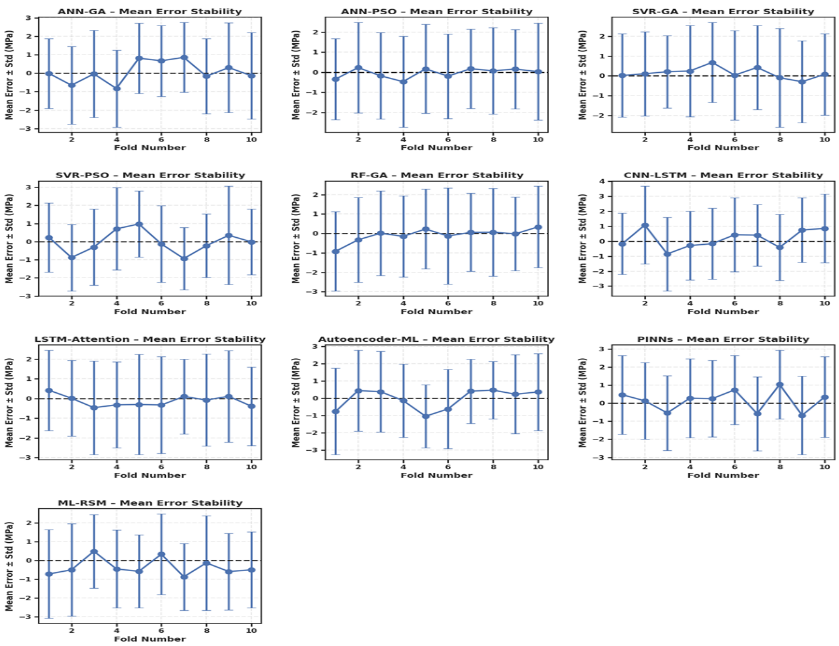

Figure 8. Model mean error and stability analysis

Figure 10 shows the relative importance of input variables for CS prediction, with MRR being the most influential, followed by particle size, kind, and W/B. This ranking is consistent with the ANOVA results, which indicate that MRR and particle size have statistically significant effects. SHAP summary plots were used to validate feature importance, confirming that the key variables identified by PDP and ALE consistently have the greatest impact on model predictions. Figure 11 shows PDPs illustrating the global marginal influence of each input variable on CS, revealing functional relationships between predictors and the model output. Kind has a positive effect on CS, while particle size shows a non-monotonic influence, suggesting an optimal range for improvement. W/B negatively correlates with CS, while MRR displays a nonlinear effect, increasing strength up to an intermediate level before decreasing, suggesting saturation at higher replacement rates

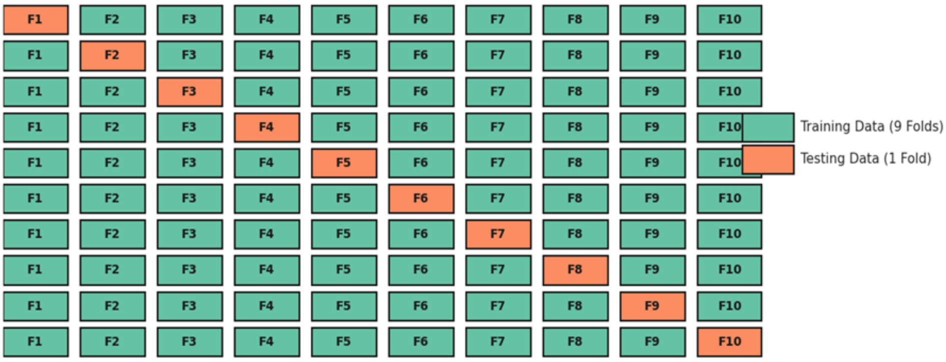

Figure 9. 10-fold cross validation scheme

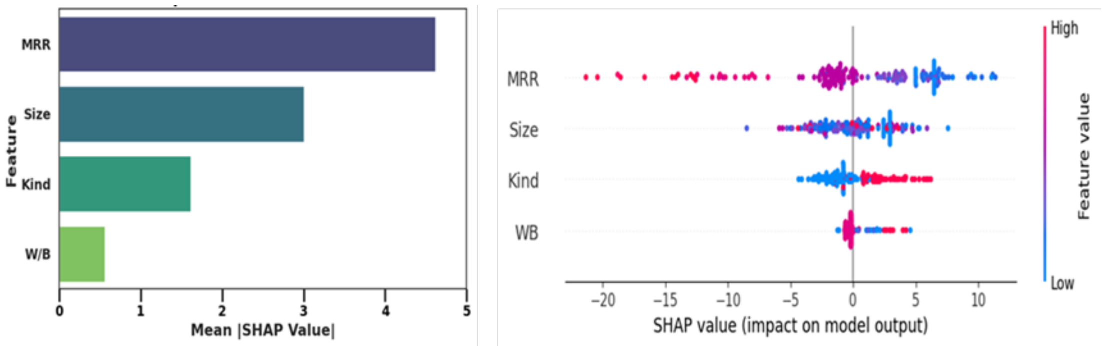

Figure 10. SHAP analysis of input variables influencing CS

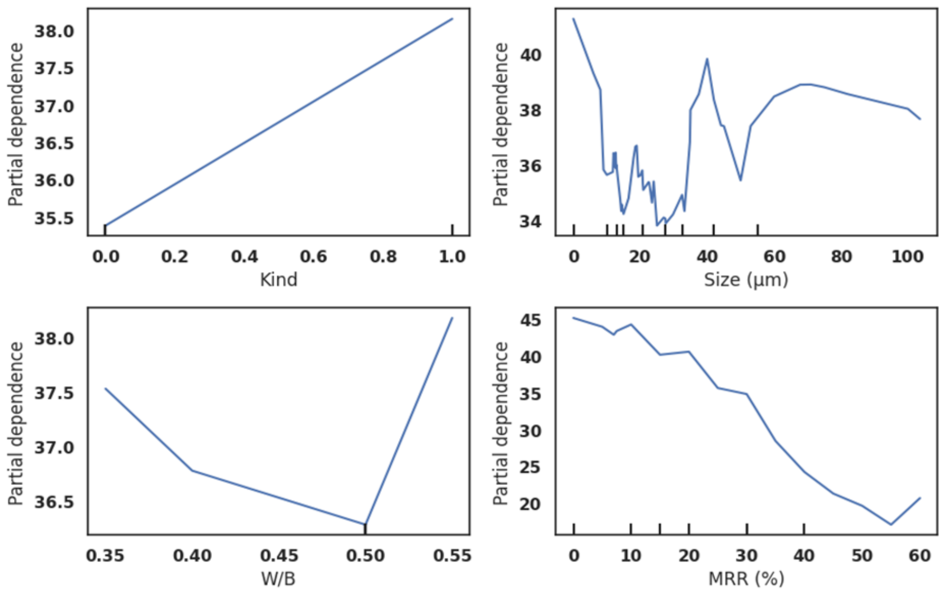

Figure 11. Partial dependence plots of input features

Figure 12 illustrates ALE plots, showing the contribution of each input variable to CS prediction while considering interactions with other features. The ALE plot for kind is essentially flat (mean |ALE| = 0.000), indicating a negligible effect on CS. In contrast, particle size exhibits a moderate, nonlinear effect (mean |ALE| = 2.156), with CS increasing significantly for sizes above approximately 30 μm, suggesting a size-dependent enhancement. The W/B ratio exhibits a linear negative effect on CS (mean |ALE| = 0.994), whereas MRR has the strongest influence (mean |ALE| = 7.086), with CS increasing up to a moderate level before declining, indicating a saturation effect at higher replacement rates.

Taken together, these plots provide an interpretable evaluation of the magnitude and functional effect of each input parameter on CS. The observed flatness of the ALE plot for 'kind' reflects the ALE methodology, which measures the average local effect of a feature while considering interactions with other variables. Limited variability within local intervals can cause ALE to appear nearly flat, despite a feature exerting significant overall influence. By comparison, SHAP and ANOVA quantify global contributions, reflecting the total effect of 'kind' on the model's predictions. Thus, despite the flat ALE, the importance of 'Kind' is consistently supported by SHAP, ANOVA, and ICE analyses, validating its key role in compressive strength prediction.

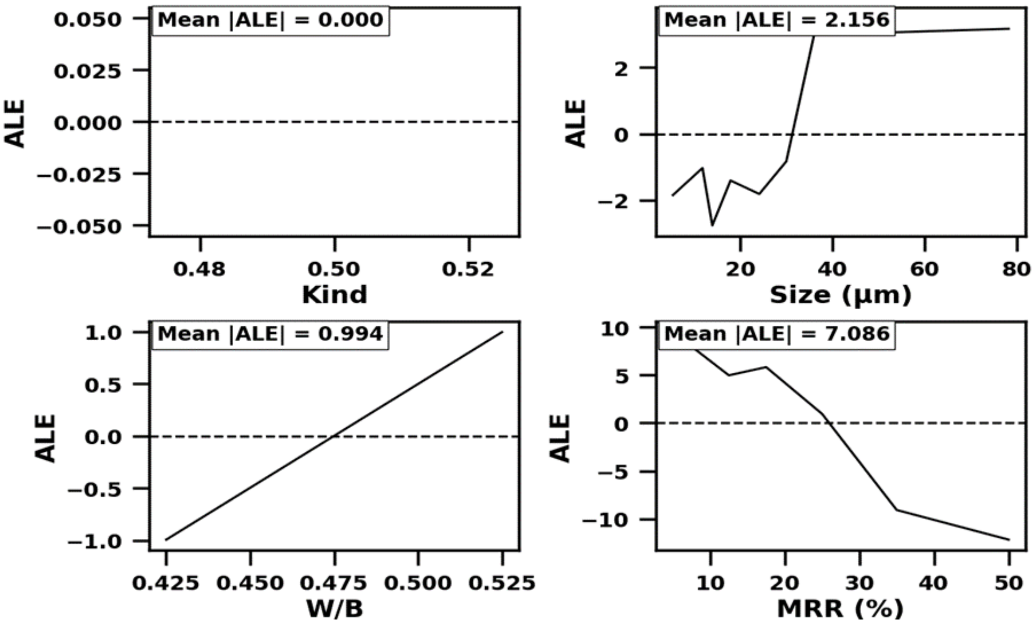

Figure 12. ALE plots showing the local effects of input variables on the CS

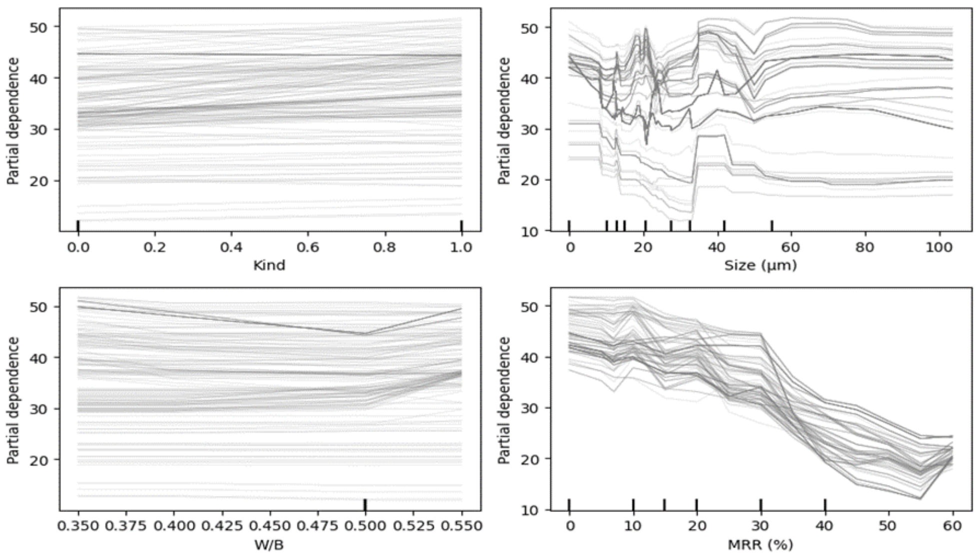

Figure 13. ICE plots illustrating the effect of each input feature on the predicted CS

Figure 13 illustrates ICE plots, showing the marginal impact of each input variable on predicted CS and revealing the model's response at the level of individual features. ICE curves indicate that kind has little effect on CS, while particle size exhibits moderate nonlinear behavior, with strength increasing above ~30 µm and notable variation across individual samples. W/B shows a clear negative effect, indicating that CS is highly sensitive to water content. MRR exhibits pronounced nonlinear effects, with CS increasing at intermediate levels and declining at higher levels, indicative of saturation behavior. ICE plots collectively illustrate how each feature influences CS predictions, offering a detailed and feature-specific view of model behavior. ALE and ICE analyses (Figures 12 and 13) indicate that compressive strength exhibits a nonlinear response to MRR (%), increasing up to an optimal range (~15–30%) before slightly decreasing. In contrast, the W/B ratio consistently reduces strength, while particle size exerts a moderate nonlinear effect, providing guidance for optimized mix design. Figure 14 illustrates the graphical user interface (GUI) developed for this study to predict the CS of mortar. The GUI allows users to enter material parameters and immediately receive predicted CS values, supporting practical use of the hybrid models.

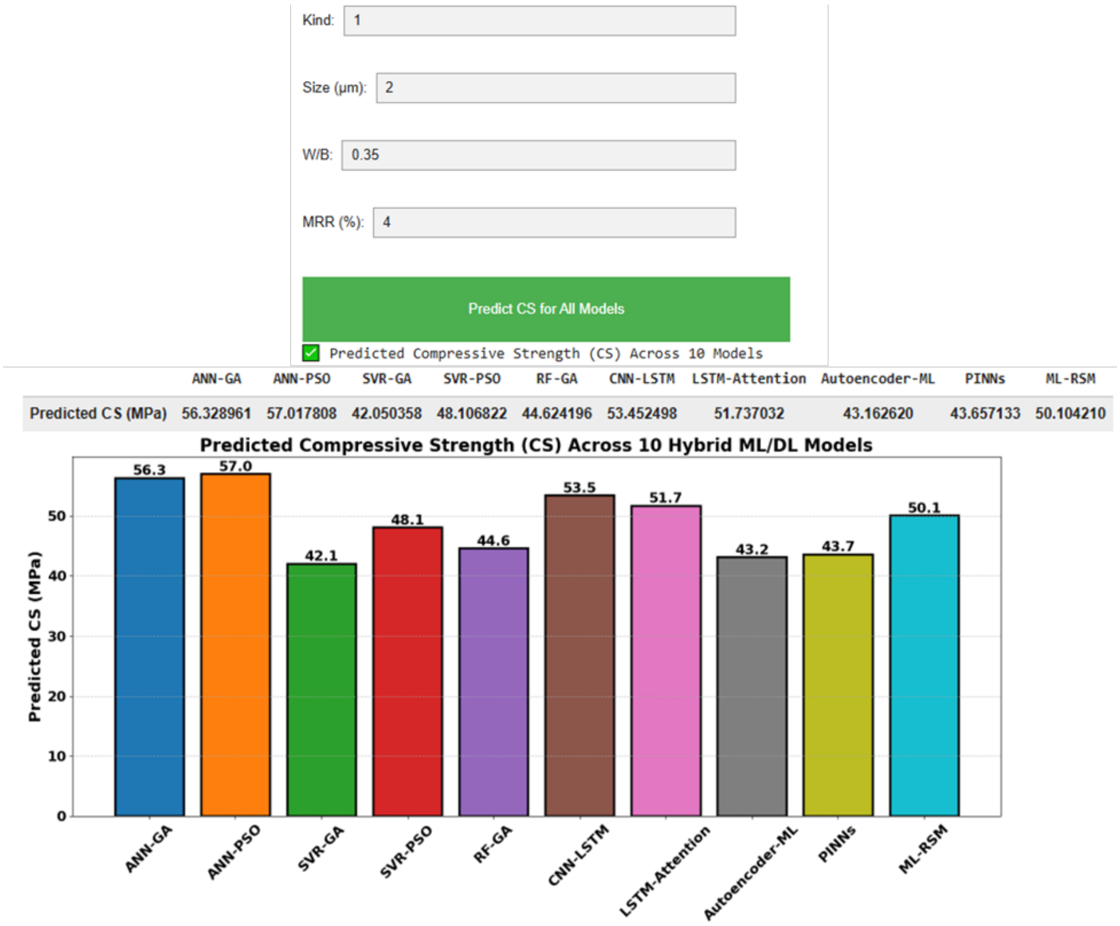

Figure 14. GUI of the models for CS prediction

- **Engineering Implications**

This study offers critical insights for the design and optimization of RPM in engineering applications. The hybrid ML–DL framework facilitates accurate prediction of CS, identifying key factors such as MRR, and delineating an optimal MRR range (~15–30%) to maximize performance. Implementing these guidelines enables engineers to optimize mix designs, enhance structural efficiency, and promote sustainability by utilizing industrial by-products, thereby reducing CO_2 emissions. The proposed approach underscores the potential of data-driven design for efficient, resilient, and sustainable construction.

4. Conclusion

This study investigated the CS of recycled powder mortar (RPM) through a hybrid ML–DL framework, demonstrating its potential for advanced material design and sustainable mix optimization. 10-fold cross-validation provided a rigorous and generalizable assessment of model performance, while the incorporation of hybrid optimization strategies substantially enhanced predictive accuracy. Across 10-fold cross-validation, the SVR–PSO model achieved the highest average performance, with R^2 = 0.950, RMSE = 2.042 MPa, MAE = 1.689 MPa, and MAPE = 0.046, slightly outperforming the Autoencoder–ML model (R^2 = 0.945, RMSE = 2.163 MPa, MAE = 1.786 MPa, MAPE = 0.049) and demonstrating superior predictive accuracy and robustness relative to all other evaluated models. ANN and RF–GA models showed high robustness, whereas DL approaches were competitive but exhibited slightly lower stability. Explainability analyses, including PDP, ALE, ICE, and ANOVA, identified mass replacement ratio (MRR) and material type as the most influential variables, thereby enhancing model interpretability. Notwithstanding these strengths, the study is constrained by reliance on a single dataset, simulated input parameters, and models that do not incorporate explicit physical constraints, potentially limiting generalizability. Practical deployment further entails risks such as extrapolation beyond the training domain, application to novel materials, and scale effects. While the graphical user interface (GUI) illustrates the predictive workflow, future investigations should incorporate experimental validation and uncertainty quantification to ensure reliable real-world application. Collectively, the study underscores the methodological novelty of a hybrid ML–DL framework for RPM strength prediction and provides actionable insights for sustainable, data-driven mix design.

Declaration of Conflict of Interests

The authors declare that there is no conflict of interest.

References

[1.] Fode, T.A., Y.A.C. Jande, and T. Kivevele, Effects of different supplementary cementitious materials on durability and mechanical properties of cement composite–Comprehensive review. Heliyon, 2023. 9(7).

[2.] Altuncı, Y.T., A Comprehensive Study on the Estimation of Concrete Compressive Strength Using Machine Learning Models. Buildings, 2024. 14(12): p. 3851.

[3.] Darlington, R.B. and A.F. Hayes, Regression analysis and linear models: Concepts, applications, and implementation. 2016: Guilford Publications.

[4.] Sun, C., et al., Low-carbon and fundamental properties of eco-efficient mortar with recycled powders. Materials, 2021. 14(24): p. 7503.

[5.] Yang, Y., et al., short review on the application of recycled powder in cement-based materials: Preparation, performance and activity excitation. Buildings, 2022. 12(10): p. 1568.

[6.] Ma, Z., et al., Using recycled aggregate and powder from high-strength mortar waste for durable cement-based materials: Microstructure and chloride transport. Journal of Cleaner Production, 2023. 417: p. 137998.

[7.] Hu, J., W. Ahmed, and D. Jiao, A critical review of the technical characteristics of recycled brick powder and its influence on concrete properties. Buildings, 2024. 14(11): p. 3691.

[8.] 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.

[9.] Hussain, A., A.H. Sakhaei, and M. Shafiee, Machine learning-based constitutive modelling for material non-linearity: A review. Mechanics of Advanced Materials and Structures, 2024: p. 1-19.

[10.] Yaswanth, K., et al., Compressive strength prediction of ternary blended geopolymer concrete using artificial neural networks and support vector regression. Innovative Infrastructure Solutions, 2024. 9(2): p. 32.

[11.] Narayanan, R. and N. Ganesh, A Comprehensive Review of Metaheuristics for Hyperparameter Optimization in Machine Learning. Metaheuristics for Machine Learning: Algorithms and Applications, 2024: p. 37-72.

[12.] Mienye, I.D. and T.G. Swart, A comprehensive review of deep learning: Architectures, recent advances, and applications. Information, 2024. 15(12): p. 755.

[13.] de la Mata, F.F., et al., Physics-informed neural networks for data-driven simulation: Advantages, limitations, and opportunities. Physica A: Statistical Mechanics and its Applications, 2023. 610: p. 128415.

[14.] Montesinos López, O.A., A. Montesinos López, and J. Crossa, Overfitting, model tuning, and evaluation of prediction performance, in Multivariate statistical machine learning methods for genomic prediction. 2022, Springer. p. 109-139.

[15.] Cifci, A., Interpretable prediction of a decentralized smart grid based on machine learning and explainable artificial intelligence. IEEE Access, 2025.

[16.] Wright, R., Interpreting black-box machine learning models using partial dependence and individual conditional expectation plots. Exploring SAS® Enterprise Miner Special Collection, 2018. 1950.

[17.] Mirzaei, A. and A. Aghsami, A Hybrid Deep Reinforcement Learning Architecture for Optimizing Concrete Mix Design Through Precision Strength Prediction. Mathematical and Computational Applications, 2025. 30(4): p. 83.

[18.] 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.

[19.] Moges, G., et al., Development and comparative analysis of ANN and SVR-based models with conventional regression models for predicting spray drift. Environmental Science and Pollution Research, 2023. 30(8): p. 21927-21944.

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

[21.] 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.

[22.] 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.

[23.] Açıkkar, M. and Y. Altunkol, A novel hybrid PSO-and GS-based hyperparameter optimization algorithm for support vector regression. Neural Computing and Applications, 2023. 35(27): p. 19961-19977.

[24.] Parsaeimehr, E., M. Fartash, and J. Akbari Torkestani, improving feature extraction using a hybrid of CNN and LSTM for entity identification. Neural Processing Letters, 2023. 55(5): p. 5979-5994.

[25.] Martens, H.A. and P. Dardenne, Validation and verification of regression in small data sets. Chemometrics and intelligent laboratory systems, 1998. 44(1-2): p. 99-121.

[26.] 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.

[27.] 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.

[28.] 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.

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

[30.] Jueyendah, S., C.H.J.C.E. Martins, and P. Modeling, Optimal Design of Welded Structure Using SVM. 2024. 7(3).

[31.] Jueyendah, S., et al., Comparative Study of Linear and Non-Linear ML Algorithms for Cement Mortar Strength Estimation. Buildings, 2025. 15(16): p. 2932.

[32.] 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. -.

[33.] 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.

[34.] 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.

[35.] 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.

[36.] Minh, D., et al., Explainable artificial intelligence: a comprehensive review. Artificial Intelligence Review, 2022. 55(5): p. 3503-3568.

[37.] 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.

[38.] 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.

[39.] Danesh, T., et al., Interpretability of neural networks predictions using Accumulated Local Effects as a model-agnostic method, in Computer aided chemical engineering. 2022, Elsevier. p. 1501-1506.

[40.] Lazaridis, P.C., et al., Interpretable machine learning for assessing the cumulative damage of a reinforced concrete frame induced by seismic sequences. Sustainability, 2023. 15(17): p. 12768.

[41.] Yeh, A. and A. Ngo. Bringing a ruler into the black box: uncovering feature impact from individual conditional expectation plots. in Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 2021. Springer.

[42.] Jia, J.-F., et al. An interpretable ensemble learning method to predict the compressive strength of concrete. in Structures. 2022. Elsevier.

[43.] Tatachar, A.V., Comparative assessment of regression models based on model evaluation metrics. International Research Journal of Engineering and Technology (IRJET), 2021. 8(09): p. 2395-0056.

[44.] Araba, A.M., et al., Estimation at completion in civil engineering projects: review of regression and soft computing models. Knowledge-Based Engineering and Sciences, 2021. 2(2): p. 1-12.

How to Cite This Article

Jueyendah, S., Ağcakoca, E., and Yaman, Z., An Interpretable Hybrid Machine Learning and Deep Learning Framework for Predicting Compressive Strength of Recycled Powder Mortar, Civil Engineering Beyond Limits, 1(2026), 11118
https://doi.org/10.36937/cebel.2026.11118