International Journal for Research Trends in Social Science & Humanities
Volume 3 Issue 4
July-Aug 2025, PP 1-5

Prediction of Compressive and Splitting Tensile Strengths in Steel Fiber-Reinforced Recycled Aggregate Concrete Using Machine Learning and PSO Optimization

Yassine Dahbi, Hamza Naciri, Hamza Zaouri, Ouahib Alaoui
College of Civil Engineering, Nanjing Tech University, Nanjing, 211816, China

Abstract. This study examines the use of GradientBoostingRegressor, StackingRegressor, and Gradient Boosting Regression with HistGradientBoosting in developing models that predict the compressive strength (fcu) and splitting tensile strength (fsp) of steel fiber-reinforced recycled aggregate concrete (SFR-RAC). The information comprises 465 compressive strength and 339 splitting tensile strength data of concrete mixes with varied ratios. Training and model testing were performed using 80/20 split with PSO for the hyperparameter optimization. The performance of the model was measured with four statistical metrics: coefficient of determination (R^2), mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). Out of the models, Gradient Boosting Regression with HistGradientBoosting performed better in terms of prediction, with StackingRegressor taking the second rank. SHapley Additive exPlanations (SHAP) and feature importance were employed to determine the influence of input parameters on model predictions. From the results obtained, it was evident that the water content, cement content, and fiber ratio influence considerably the strength of SFR-RAC. The models give good insights regarding SFR-RAC mixture behavior, which is helpful in the production of environmentally friendly concrete with greater enhanced strength. Future research can enhance the data and use other predictor variables to further support these models.

Index Terms- Steel Fiber-Reinforced Recycled, Aggregate Concrete, Machine Learning Techniques, Compressive Strength Prediction, Splitting Tensile Strength Prediction, Particle Swarm Optimization (PSO)

I. Introduction

The construction industry has shifted towards the use of sustainable material, and steel fiber-reinforced recycled aggregate concrete (SFR-RAC) is a possible new substitute for conventional concrete. With the incorporation of steel fibers and recycled aggregates, SFR-RAC combines enhanced mechanical properties and durability, hence sustainable concrete solutions. However, one of the problems that still exist in the field is accurately predicting the mechanical performance of such concrete mixtures, particularly their compressive strength and splitting tensile strength, which are most important parameters of concrete's structural strength and durability [1][2].

Usual experimental methods for determining these properties involve time-consuming and costly laboratory testing, which can be resource- and time-intensive [3]. Thus, interest has been growing in using machine learning (ML) techniques to build prediction models which can estimate the mechanical properties of concrete from its mix composition. These models can potentially reduce the need for large-scale experimentation, yet facilitate speedy and reliable predictions [4][5].

International Journal for Research Trends in Social Science & Humanities
Volume 3 Issue 4
July-Aug 2025, PP 1-5

Among the machine learning techniques, regression techniques are most suitable for predicting continuous values like compressive and tensile strengths. Three new regression techniques—GradientBoostingRegressor, StackingRegressor, and Gradient Boosting Regression with HistGradientBoosting—are utilized in this research as they can handle complex relationships between input variables and target variables even in cases where the data is not linear [6][7]. These techniques have high accuracy and efficiency of use in a wide range of fields from civil engineering [8][9].

To further improve the accuracy of the models, Particle Swarm Optimization (PSO) is used to optimize hyperparameters of the models. PSO is a powerful optimization algorithm inspired by bird flocking behavior and has been successfully applied in numerous engineering applications [10][11]. In the current study, PSO helps determine the optimal hyperparameters of the regression models to guarantee that compressive strength and splitting tensile strength predictions are as accurate as possible.

The data used in training model development include 465 compressive and 339 splitting tensile strength samples, acquired from concrete mixes with mixtures of steel fibers, recycled aggregates, and other contents in varied proportions. The samples were divided between the test set and the training set based on an 80/20 percentage for measuring model performance [12]. The models' performance was verified by four statistical indices: the coefficient of determination (R^2), mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). These indices show a broad view about the ability of the models to make the predictions of compressive and tensile strengths accurately and reliably [13][14].

In addition to evaluating the performance of the models, this study uses feature importance analysis and SHapley Additive exPlanations (SHAP) to identify the influence of each input variable on the predicted outcomes. These tools allow us to identify significant factors such as water content, amount of cement, and percentage of fiber, which significantly affect the mechanical strength of SFR-RAC [15].

The findings of this study will try to provide meaningful information for the design of more effective and sustainable concrete mixes that are not only in conformity with structural requirements but also minimize environmental impact.

Through the application of machine learning techniques and optimization algorithms, this research improves the sustainability of the building material by a better understanding of the mechanical properties of SFR-RAC and the influencing factors. Future research can be assisted by exploring other datasets and parameters to further enhance these models and assess their applicability in real-world situations.

II. Methodology

The method used in this study is illustrated in Figure 1 , in which data gathering, model development, and evaluation are all addressed. The data set was initially divided into the training and test sets, with the input variables and outputs clearly specified. GradientBoostingRegressor, StackingRegressor, and HistGradientBoosting-based Gradient Boosting Regression were applied to predict the compressive and splitting tensile strengths of steel fiber-reinforced recycled aggregate concrete (SFR-RAC) [16][17]. Particle Swarm Optimization (PSO) was employed to tune the models' hyperparameters to deliver optimal performance and avoid overfitting [20][21]. After the models had predicted, their output was cross-validated using model evaluation techniques and its explainability was enhanced further with the aid of

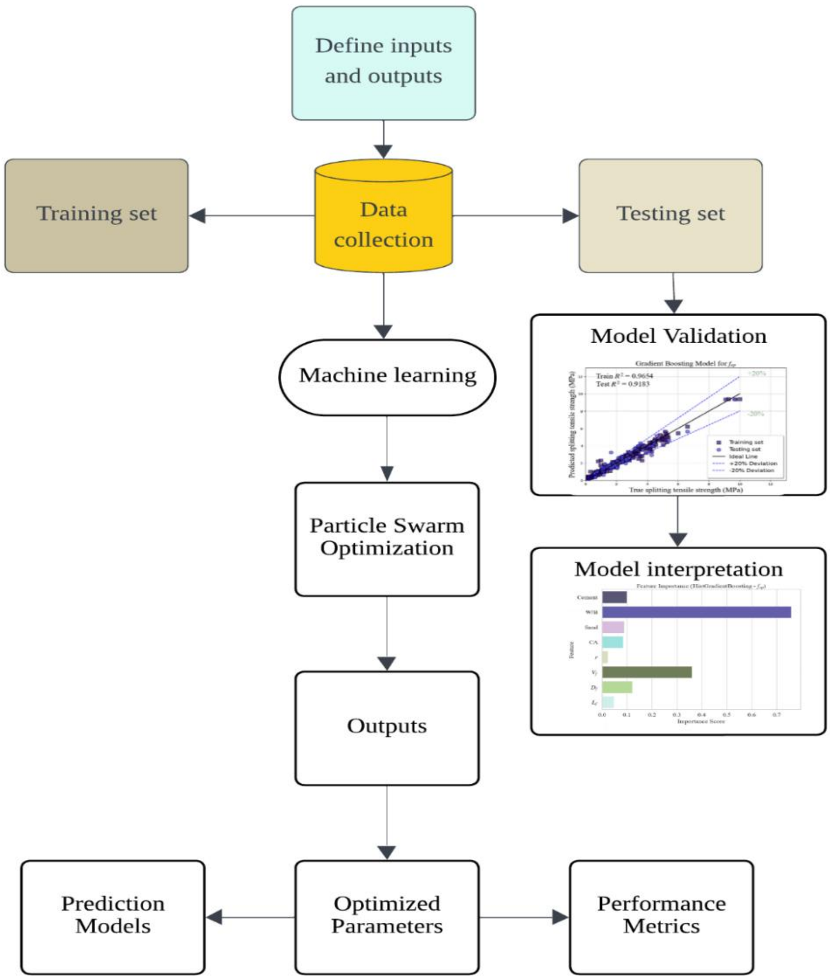
SHapley Additive exPlanations (SHAP) to identify the effect of different input features [23][24]. The final step involved assessing the performance of the models in terms of traditional measures such as R2, MAE, RMSE, and MAPE [18][19]. Figure 1 provides a clear illustration of these steps, from data collection through performance evaluation, and is a critical element of understanding the research methodology.

Figure 1: Model proposal

Fundamental ML models

- **GradientBoostingRegressor**

The GradientBoostingRegressor is an ensemble machine learning method that builds a strong predictive model by iteratively fitting decision trees to the residuals



(errors) of previous models. Each tree is fit to predict the error of the previous tree, and the final prediction is the sum of the output of all trees, multiplied by a learning rate η:

International Journal for Research Trends in Social Science & Humanities
Volume 3 Issue 4
July-Aug 2025, PP 1-5

Equation if statement:

$$f(x) = f_0(x) + \eta \sum_{m=1}^{M} h_m(x)$$

where f_0 (x) is the initial prediction, h_m (x) is the model (tree) at the m-th iteration, and M is the total number of trees. The method minimizes the prediction error using gradient descent, making it effective for complex, nonlinear relationships [25][26] GradientBoostingRegressor is good in identifying complex patterns in high-dimensional data, which is particularly useful in predicting properties like compressive and splitting tensile strengths of concrete mix designs [27]. The model also tends to be more immune to overfitting, especially under well-tuned parameters, and is, as such, suitable for noisy or outlier datasets [28]. It also provides valuable insights into the relevance of numerous various features, such as water content and water-to-cement ratio, which affect the model predictions significantly [29].

- **The StackingRegressor**

The StackingRegressor is a type of ensemble learning that trains a number of base models simultaneously in order to create a stronger predictive model. Unlike typical ensemble methods like Random Forests that train each model independently, StackingRegressor trains a number of base models and then trains a meta-model to learn how to optimally combine their predictions. The meta-model is trained on the predictions of the base models, allowing it to learn to leverage each model's strengths. The final prediction is made by weighting all the base model predictions by the meta-model's learned coefficients.

The benefit of StackingRegressor is improved accuracy through the reduction of bias and variance by stacking a set of base models to get a better performance than the performance of individual models [30][31]. It is also very flexible, accommodating the use of different base models, both linear and nonlinear, which enhances its ability to model complex relationships in the data [32]. This method is especially effective in predicting outcomes where the relationship between input parameters and target results is complex and nonlinear. An example would be concrete mixture designs, where parameters like the content of fibers and aggregate proportions have a profound impact on the mechanical strength of the material [33].

- **Gradient Boosting Regression with HistGradientBoosting**

Gradient Boosting Regression with HistGradientBoosting is a more effective version of the basic gradient boosting algorithm. It offers better computational efficiency with a histogram-based approach, which allows for better memory usage and quicker training, especially on large datasets. HistGradientBoosting divides continuous features into discrete bins (histograms), for better computational speed without compromising accuracy for typical gradient boosting algorithms. The loss function used in gradient boosting, e.g., HistGradientBoosting, is most frequently the Mean Squared Error (MSE), given by the formula below:

$$L = \sum_{i=1}^{N} \left(y_i - f(x_i)\right)^2$$

where L is the loss, y_i is the true value for the iii-th sample, f(x_i) is the predicted value for the i-th sample, and N is the number of samples. The model is trained to minimize this loss by iteratively adjusting the predictions.

One of the best advantages of Gradient Boosting Regression with HistGradientBoosting is that it can handle large datasets very well. The histogram-based approach optimizes both training time and memory usage, and hence is useful for high-dimensional data [34][35]. It also maintains the inherent power of vanilla gradient boosting, such as the ability to learn about complex, nonlinear relationships and overfitting resistance if hyperparameters are optimally tuned. The method performs well in areas such as concrete mix design, where relationships among variables (e.g., fiber content, cement ratio) are complex and nonlinear [36]. Also, HistGradientBoosting is computationally more intensive, so faster experimentation and model tuning can be done compared to traditional gradient boosting algorithms.

Table 1: Statistical metrics of different ML models on the test set.

Item	type	Training				Testing			
		R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
fcu	GBM	0.878	5.173	3.560	7.774%	0.830	6.097	3.874	8.218%
	SRM	0.891	4.873	3.572	7.792%	0.792	6.101	3.685	9.153%
	HistGBM	0.895	4.791	3.541	7.732%	0.874	4.873	3.541	7.792%
fsp	GBM	0.901	0.618	0.445	10.217%	0.896	0.618	0.542	10.217%
	SRM	0.854	0.752	0.526	11.911%	0.803	0.868	0.637	16.445%
	HistGBM	0.955	0.451	0.224	5.774%	0.823	0.829	0.445	12.640%

III. Model development

Regression evaluation criteria

The performance of the machine learning models is evaluated by calculating the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Nash-Sutcliffe Efficiency (NSE), Relative Percentage Difference (RPD), and Akaike Information Criterion (AIC) metrics. The procedure for calculating the evaluation metrics is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$\mathrm{MAPE} = \frac{100\backslash\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

Particle Swarm Optimization (PSO)

Kennedy and Eberhart introduced Particle Swarm Optimization (PSO) in 1995, a nature-inspired optimization algorithm for bird flocking and school fish behavior. It is an optimization algorithm frequently employed to solve complex optimization problems involving high-dimensional variables and non-linear interactions. PSO is extremely effective at solving optimization problems in search spaces of high dimensions and is well known for its ability to achieve optimal or near-optimal solutions.

The basic goal of PSO is to find the best values of a number of parameters by searching the solution space. Any candidate solution is the so-called "particle" in the swarm. Every particle has a position and velocity, which define the current solution and by how much it moves in the search space. These particles "move" in the space based on modifying their positions according to two significant factors: their personal experience (personal best position, pbest) and the global experience found by the entire swarm (global best position, gbest).

In machine learning model hyperparameter optimization, PSO can identify the optimal parameter values to improve model performance. For instance, in predictive modeling of Steel Fiber-Reinforced Recycled Aggregate Concrete (SFR-RAC), PSO was utilized in determining the optimal machine learning model hyperparameters, which led to significantly improved predictive accuracy.

The movement of the particles in the swarm is regulated by the following equations which update the position and velocity of all particles at each step:

$$v_i^{t+1} = w \cdot v_i^t + c_1 \cdot r_1 \cdot (pbest_i - x_i^t) + c_2 \cdot r_2 \cdot (gbest - x_i^t)$$

Where v_i^(t+1)the updated velocity of the particle i; w is inertia weight (balances exploration and exploitation); c_1is a cognitive coefficient (self-confidence); c_2 is

social coefficient (swarm confidence); r_1, r_2 are random numbers in [0,1]; $pbest_i$: personal best position of particle i; gbest the global best position found by the swarm; x_i^t is the current position of a particle i.
At every iteration of the Particle Swarm Optimization (PSO) algorithm, particles are attracted to their personal best positions (designated as pbest) and global-best position (designated as gbest) in the solution space and converge nearly to the optimal solution. This movement of optimality is based on both the particle's personal experience and the collective experience of the swarm [37]. The two primary forces that drive the optimization process are exploration and exploitation. Exploration is the ability of the particle to venture into new and unexplored areas of the solution space, while exploitation tries to improve and refine solutions in the areas where the particle has already found good outcomes. The equilibrium between the two parameters is maintained by the inertia weight (w) and cognitive (c_1) and social (c_2) coefficients [38].
influences its current motion. The higher the inertia weight, the more it leans towards exploration, allowing particles to search further in the solution space, while a smaller inertia weight leans towards exploitation, focusing on refining already good solutions. The cognitive coefficient (c_1)) represents the particle's faith in its individual experiences, leading it to its personal best position, and the social coefficient (c_2) represents the faith in the global best-known position of the population by the overall swarm, leading particles to the best solution found by the swarm as a whole. By varying these values, PSO achieves an optimal trade-off between finding new regions in the search space and exploiting the global best solution discovered so far, one of the most significant reasons why PSO is efficient for solving complex optimization problems [39].

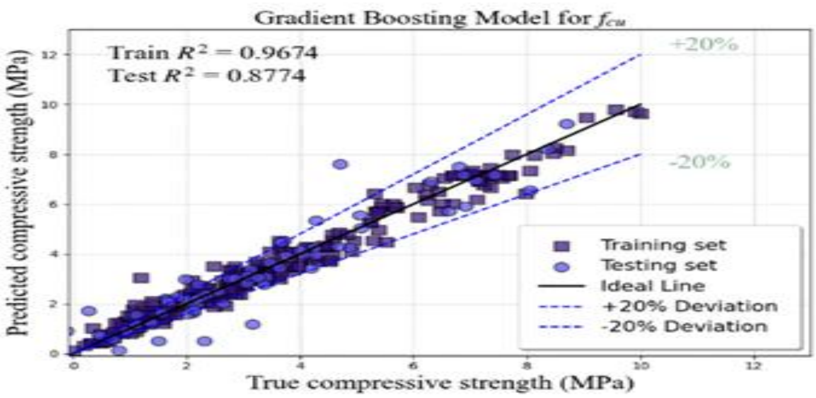
PSO has become widely used in applications such as machine learning hyperparameter optimization. With these types of applications, the goal is normally to optimize a collection of parameters simultaneously in an effort to improve the performance of a model. Machine learning algorithms typically have a number of hyperparameters (such as learning rate, no. of trees, and tree depth) which must be optimized for optimal performance. The natural characteristic of PSO to efficiently search high-dimensional, complex hyperparameter spaces makes it a good candidate for hyperparameter tuning [40]. Through feedback from performance metrics (e.g., accuracy, loss, etc.), PSO iteratively modifies the hyperparameters in the direction of improved performance.

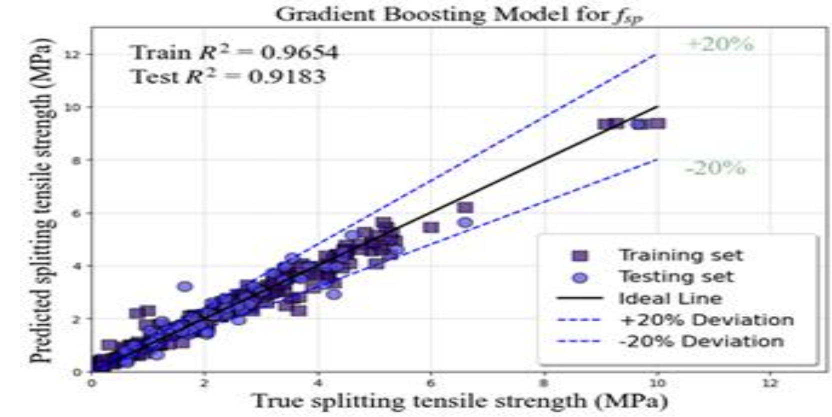
PSO will be most appropriate for problems in which the relationship between parameters and outcomes is not linear and complex, and therefore it is tough for typical optimization techniques to perform optimally. For example, in the estimation of concrete mix design, where several parameters such as the quantity of fiber, cement ratio, and water-to-cement ratio influence the ultimate strength of the concrete, PSO can efficiently explore the vast parameter space to identify the most effective mix that optimizes the characteristics of interest [41]. This approach not only enhances the accuracy of the predictions but also facilitates the manufacture of materials to be more efficient and economical through the optimization of the parameters through automation.

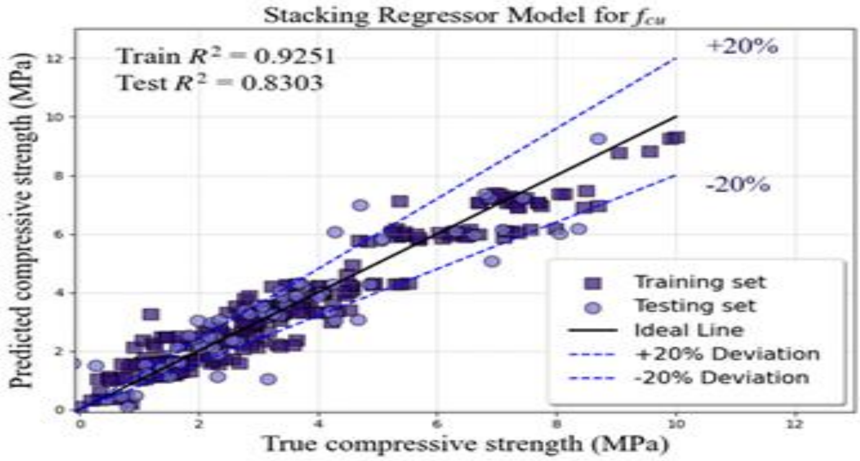
The ease of use and adaptability of PSO have made it a popular method in many optimization problems outside of machine learning, including engineering design, robotics, and financial modeling, where the optimization of many variables at one instance is required to improve the performance of a system. As the method continues to be developed, improvements to PSO, including hybrid PSO models and adaptive parameterization, continue to make it an increasingly useful tool for solving problems in the real world.

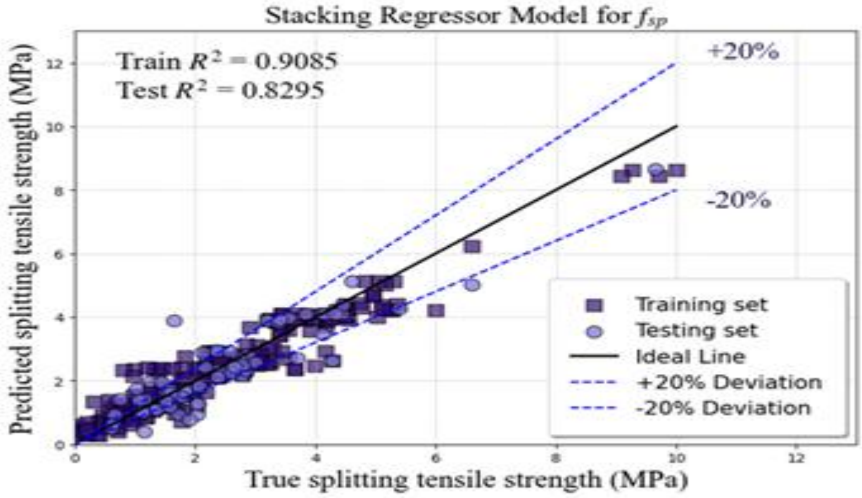
Table 2: Impact of Particle Swarm Optimization on Metrics

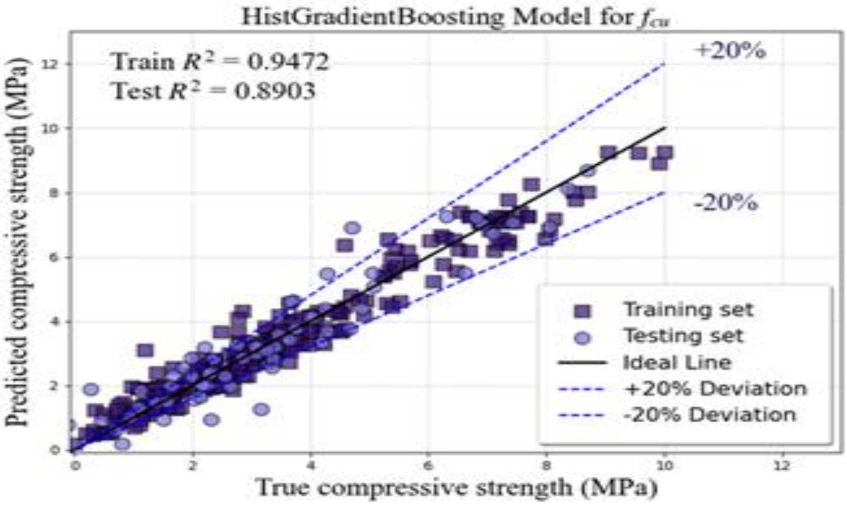
ML algorithms		Compressive strength (fcu)			Splitting tensile strength (fsp)		
		R^2	RMSE	MAPE	R^2	RMSE	MAPE
GBM	Before	0.830	6.097	8.218 %	0.896	0.618	10.217 %
	After	0.877	5.173	7.774 %	0.918	0.563	9.752%
SRM	Before	0.792	6.101	9.153 %	0.803	0.868	16.445 %
	After	0.808	5.804	8.235 %	0.829	0.721	14.382 %
HistGBM	Before	0.874	4.873	7.792 %	0.823	0.829	12.640 %
	After	0.890	4.526	7.432 %	0.856	0.752	11.911 %

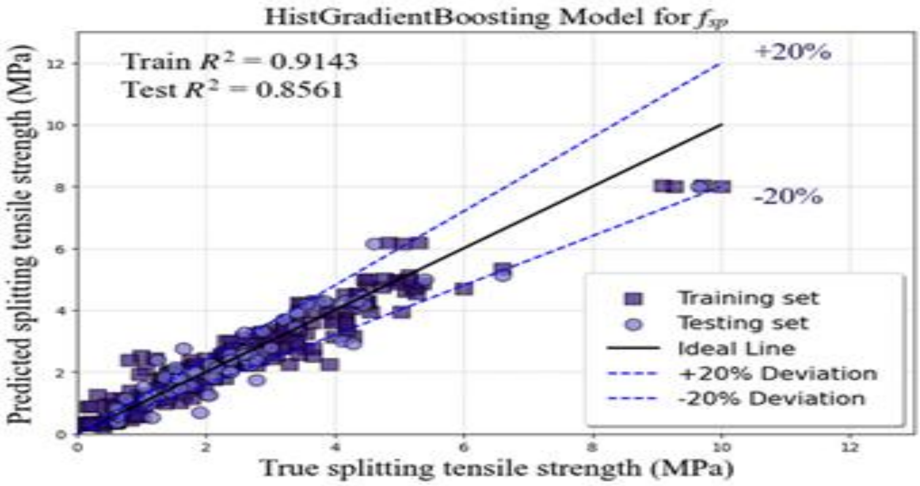



Gradient Boosting Model for f_{sp}
Train $R^2 = 0.9654$
Test $R^2 = 0.9183$
+20%
-20%
Training set
Testing set
Ideal Line
+20% Deviation
-20% Deviation
Predicted splitting tensile strength (MPa)
True splitting tensile strength (MPa)


Stacking Regressor Model for f_{cu}
Train $R^2 = 0.9251$
Test $R^2 = 0.8303$
+20%
-20%
Training set
Testing set
Ideal Line
+20% Deviation
-20% Deviation
Predicted compressive strength (MPa)
True compressive strength (MPa)


Stacking Regressor Model for f_{sp}
Train $R^2 = 0.9085$
Test $R^2 = 0.8295$
+20%
-20%
Training set
Testing set
Ideal Line
+20% Deviation
-20% Deviation
Predicted splitting tensile strength (MPa)
True splitting tensile strength (MPa)





IV. RESULTS AND DISCUSSION

Model development

Based on the overall statistical performance presented in Table 1, the comparative assessment was conducted among three machine learning algorithms—Gradient Boosting Machine (GBM), Stacking Regressor Model (SRM), and Histogram-based Gradient Boosting (HistGBM)—in an attempt to compare their potential for predicting the compressive strength (fcu) and splitting tensile strength (fsp) of steel fiber-reinforced concrete (SFRC). This comparison used a comprehensive collection of evaluation metrics: the coefficient of determination (R2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), on both testing and training datasets.

In the prediction of compressive strength (fcu), HistGBM showed consistently better generalization and accuracy on all metrics. It also had the highest testing R2 value of

0.874, significantly better than GBM (0.830) and SRM (0.792), indicating that HistGBM predictions were nearer to observed actual values. HistGBM also recorded the lowest testing RMSE of 4.873, compared to GBM's 6.097 and SRM's 6.101, indicating lower prediction error. Its MAPE of 7.792% was also the lowest, suggesting HistGBM's capability in minimizing relative prediction error. These results show that HistGBM performed a better job at capturing the nonlinearities and interactions within the data for predicting compressive strength. GBM worked modestly well, but SRM lagged behind, particularly in test accuracy, which shows its poor generalization ability despite modest training results.
In contrast, in the case of tensile strength prediction of splitting (fsp), there was a slight modification in the performance ranking. GBM exhibited the highest testing R2 of 0.896 and lowest RMSE of 0.618, indicating its superiority in predicting fsp for unseen data. This is a sign that GBM generalized well for this mechanical property and was able to produce high-fidelity predictions. Nevertheless, HistGBM again demonstrated the best training performance with a very high R2 of 0.955, very low RMSE of 0.451, and MAPE of only 5.774%. Although its test R2 dropped to 0.823—slightly below GBM—HistGBM continued to show good performance with a competitive RMSE (0.829) and moderate MAPE (12.640%). On the contrary, SRM continued to underperform, especially in testing with the worst R2 of 0.803 and worst MAPE of 16.445%, indicating high deviations between actual and predicted values. Table 1 shows that HistGBM consistently exhibits the best training performance for both fcu and fsp, illustrating its ability to learn complex patterns. GBM, however, exhibits slightly better generalization in fsp prediction, as revealed by its better testing accuracy and lower error rates. SRM, on the other hand, worst performs on most of the metrics, with poor generalization and high errors on both training and testing phases. The results highlight the necessity of selecting an algorithm that achieves a trade-off between learnability and generalization. HistGBM is the most consistent and overall best method, with its highlight being compressive strength prediction, while GBM could be considered a bit more if the application is for splitting tensile strength specifically, due to its slightly better testing performance. The comparison demonstrates the ensemble learning models' subtle behavior in different prediction tasks and that advanced boosting methods like HistGBM can yield tangible benefits for engineering problems with complex.

Effect of Particle Swarm Optimization (PSO)

The results in Figure 2 show the predictive ability of three different machine learning models—Gradient Boosting (GBM), Stacking Regressor (SRM), and Histogram-Based Gradient Boosting (HistGBM)—for compressive strength (fcu) and splitting tensile strength (fsp) after being optimized using the Particle Swarm Optimization (PSO) algorithm. The scatter plots show predicted vs. actual strength values for training and test datasets, with perfect prediction lines and ±20% error bands for easy interpretation.

Start with the Gradient Boosting model, and we observe that there is extremely good correlation between the predicted and actual values of fcu and fsp. Specifically, the R^2 is 0.9674 (train) and 0.8774 (test) for fcu, and 0.9654 (train) and 0.9183 (test) for fsp. The predicted points closely follow the ideal line and are largely within the

±20% line, indicating high accuracy and excellent generalization power for this model after optimization.

The Stacking Regressor model also gives good performance, albeit a little worse than GBM. The R^2 values for training and testing on fcu are 0.9251 and 0.8303, and on fsp are 0.9085 and 0.8295. Although still within acceptable margins of error, the scatter departs only a little further from the ideal line than GBM, primarily for the test set, suggesting SRM will not generalize quite so well but remains very predictive. Finally, HistGBM achieves competitive results, particularly for compressive strength, with R^2 of 0.9472 (training) and 0.8903 (test). For fsp, the model achieves 0.9143 (training) and 0.8561 (test), which is midway between GBM and SRM in accuracy. The close grouping of predicted values along the perfect line, mainly for fcu, confirms HistGBM's potential to model complex relationships in the data following PSO optimization.
optimization actually enhances the predictive performance of the models. Of the three models, GBM appears to perform best overall and in generalizing to unseen data, and HistGBM generating strong and reliable predictions. The results affirm the applicability of PSO in maximizing the precision of models and minimizing the margins of errors in both compressive strength and tensile strength predictions.
Table 2 findings, combined with the qualitative observations of Figure 2, are compelling evidence of the enhancing effect of Particle Swarm Optimization (PSO) on machine learning model predictive performance for predicting steel fiber-reinforced recycled aggregate concrete's compressive strength (fcu) and splitting tensile strength (fsp). In each of the three models, i.e., Gradient Boosting (GBM), Stacking Regressor (SRM), and Histogram-Based Gradient Boosting (HistGBM), there are remarkable enhancements in the performance metrics with optimization, including increases in the coefficient of determination (R^2) and decreases in Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).
In compressive strength (fcu), GBM's R^2 increased remarkably from 0.830 to 0.877 after PSO, RMSE decreased from 6.097 to 5.173, and MAPE from 8.218% to 7.774%. This is nicely complemented by the graphical plot in Figure 2, where GBM shows dense aggregation of predictions near the ideal line and within the ±20% deviation band, testifying to better generalization. SRM was also improved, and R^2 rose from 0.792 to 0.808 and RMSE fell minimally from 6.101 to 5.804. Although the reduction in MAPE is minimal (9.153% to 8.235%), Figure 2 demonstrates tighter clustering prediction after optimization. HistGBM, which also performed well before PSO, also benefited from the optimization, as shown by the improved R^2 value from 0.874 to 0.890, and RMSE and MAPE reductions from 4.873 to 4.526 and 7.792% to 7.432%, respectively all contained in the more compact spread of prediction points in Figure 2.

For splitting tensile strength (fsp) prediction, GBM also continued providing superior improvement post-optimization by taking R^2 to 0.918 from 0.896 and reducing RMSE from 0.618 to 0.563 and MAPE from 10.217% to 9.752%. This can be seen from Figure 2 by the tight tracking of predicted values on the ideal line and neat clustering within the ±20% bands. SRM also posted a small R^2 increase from 0.803 to 0.829 along with RMSE and MAPE decreases from 0.868 to 0.721 and 16.445% to 14.382%, respectively, showing a weaker yet more consistent model. HistGBM, while its R^2 increased as much as 0.823 to 0.856, posted a slight decrease in RMSE from

0.829 to 0.752 and in MAPE from 12.640% to 11.911%, demonstrating its consistency for both tasks.
In short, the evidence presented by Table 2 and Figure 2 collectively substantiates that PSO significantly enhances machine learning algorithms' performance, particularly GBM, which is consistently the best among the others in terms of training and testing accuracy. The corresponding tabular and visual proof indicates that not only does PSO enhance model accuracy but also the overall generalization power of the algorithms in different strength prediction problems.

Feature Importance and PSO Optimization Analysis

- **Feature Importance in Predicting fcu and fsp**

Figure 3 and Figure 4 show the feature importance scores calculated with SHAP (SHapley Additive exPlanations) values for predicting compressive strength (fcu) and splitting tensile strength (fsp) from the HistGradientBoosting model. They help in the interpretation of contribution of any input variable to the model prediction. From Figure 3 , it is evident that the water-to-binder ratio (W/B) is by far the most predominant parameter in fcu prediction with an importance score close to 1.4. The predominance here means that mix water content in relation to binders has a great impact on concrete compressive strength. Sand and cement have relatively lower impacts. Fiber diameter (Df), length (Lf), and volume fraction (Vf) parameters do not play an important role here, reinforcing the fact that matrix components have an overwhelming influence over fiber parameters for compressive strength. Compared to Figure 4 for fsp, there is more of a distributed feature influence. While W/B remains the most important variable, its importance is much lower than its impact in f cu. Here, fiber volume fraction (Vf) is a dominant contributor, followed by coarse aggregate (CA) and fiber diameter (Df). This shift reflects the greater sensitivity of tensile strength to interface properties and characteristics, as well as to the properties of the fibers in the composite mixture. Hence, matrix and fiber parameters for prediction of splitting tensile strength need to be addressed in combination.

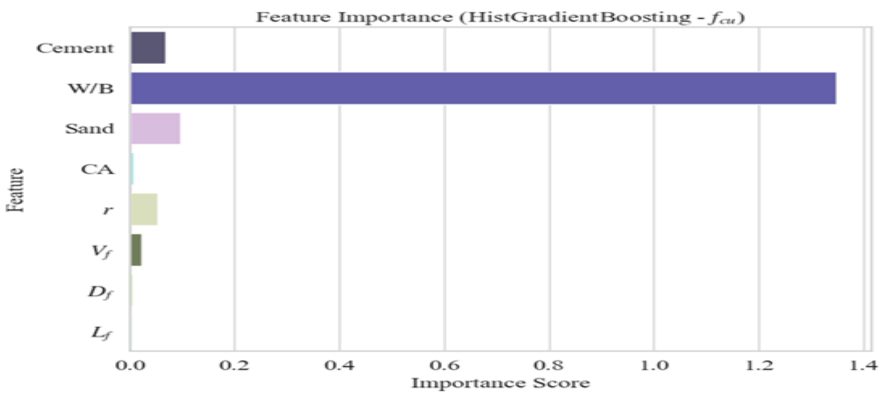


Figure 1: Feature Importance (HistGradientBoosting -f_{cu})

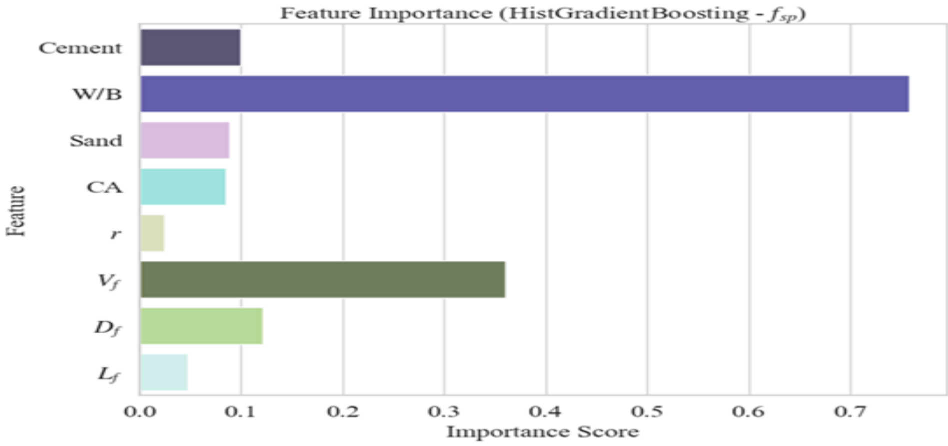


Figure 1: Feature Importance (HistGradientBoosting -f_{sp})

- **Particle Swarm Optimization Behavior**

Dynamics of Particle Swarm Optimization (PSO) during the optimization are exhibited in Figure 5 and Figure 6. Figure 5 tracks the evolution of velocity of individual particles over 30 iterations. The evident increase in velocity during initial iterations indicates an increased rate of exploration of the search space as particles acquire local and global best positions. At the neighborhood of iteration 15–20, the velocities of pioneer particles are growing immensely, showing that optimal areas are being explored more intensely. Towards the later iterations, tapering or stabilization in velocities is noticed for most particles, showing convergence of the optimization process towards optimal hyperparameters.

Figure 6 illustrates a 3D spatial mapping of particle movement in the optimization space with more understanding of particles evolving in different dimensions. The random initial paths, moving towards clustering toward final paths, visually substantiate successful learning and reduced positional variability. This pattern substantiates the efficiency of PSO in moving the population from random locations to optimal solutions by repeated memory and cooperation operations.
It may be noted from the examination of feature importance that W/B ratio exercises a dominating effect on both fcu and fsp, although fiber characteristics (specifically Vf and Df) come to play a significantly more important role as explanatory factors for tensile strength. This finding lends evidence to the hypothesis that compressive strength is extremely matrix-proportioning sensitive, while tensile strength is more composite-sensitive response.

The PSO plots confirm effective optimization by particle velocity acceleration and stabilization, followed by spatial convergence. This dynamic behavior demonstrates the effectiveness of PSO in robust model hyperparameter tuning, in turn enhancing the predictive capabilities of the ML models. Coupled with feature interpretation, the joint analysis confirms the effectiveness of integrating interpretable machine learning and evolutionary search algorithms like PSO to achieve high accuracy along with model interpretability in modeling fiber-reinforced concrete properties.

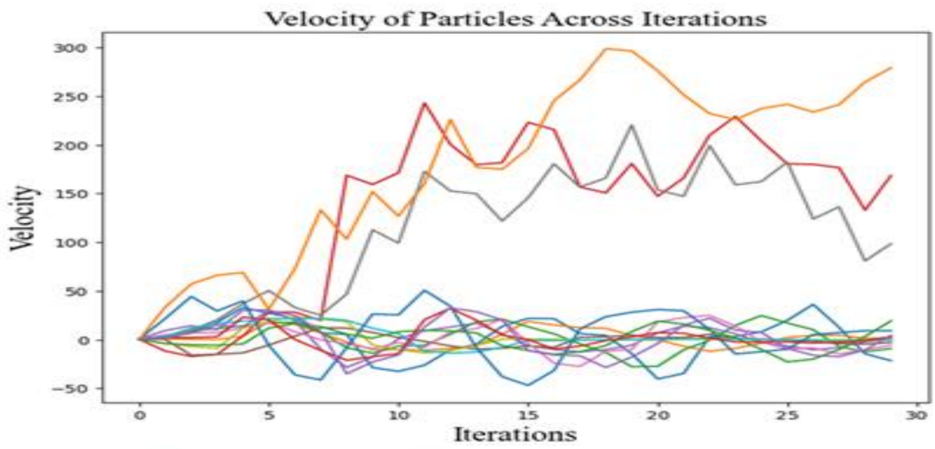


Figure 1: Velocity of Particles Across Iterations

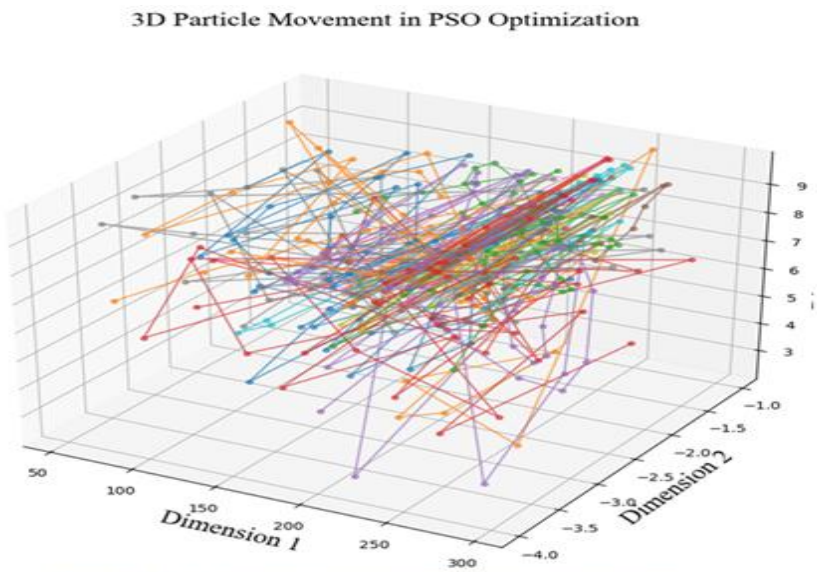


Figure 1: 3D Particle Movement in PSO Optimization

V. CONCLUSION

This study presents a comprehensive approach for forecasting mechanical properties compressive strength (fcu) and splitting tensile strength (fsp) of steel fiber-reinforced recycled aggregate concrete (SFRRAC) by using machine learning (ML) models enhanced with Particle Swarm Optimization (PSO). Of the three Machine Learning models used to test them Gradient Boosting Machine (GBM), Stacking Regressor Model (SRM), and Histogram-based Gradient Boosting (HistGBM) HistGBM proved to perform better across all instances on both fcu and fsp in terms of better R2 and lesser RMSE and MAPE. The model's performance also improved after PSO hyperparameter optimization, confirming the impact of the optimization in improving both accuracy and generalization.

Findings in Table 1 revealed that before optimization, HistGBM was already better than the other models, particularly on compressive strength prediction, with highest R2 values and lowest error rates. Results after optimization in Table 2 and Figure 2 also reaffirmed its superiority with drastic improvement in test accuracy and closer adherence to the ±20% prediction interval range. In addition, PSO optimization behavior, illustrated by the velocity and 3D movement patterns, confirmed the convergence stability behavior and optimization approach effectiveness. Further, feature importance analysis via SHAP showed the water-to-binder ratio (W/B) as the most influential parameter for compressive and tensile strength. Nevertheless, the fiber property contribution, particularly fiber volume fraction (Vf) and fiber diameter (Df), rose in regulating tensile strength, thus highlighting the need for balancing matrix and fiber design while designing high-performance SFRRAC. In conclusion, the synergy of advanced ensemble ML models and PSO optimization with the help of interpretable feature analysis forms a very potent, accurate, and interpretable prediction model for the complex behavior of recycled fiber-reinforced concretes. The findings present valuable contributions to academic research as well as engineering design in green concrete engineering.

CRediT authorship contribution statement

- Yassine Dahbi: Conceptualization, methodology development, data collection, model implementation, optimization, formal analysis, visualization, interpretation of results, original draft preparation, and overall project supervision.
- Hamza Naciri: Assisted in the preparation of visual elements and participated in technical formatting and minor editing tasks.
- Hamza Zaouri: Supported with preliminary result verification and offered suggestions during the manuscript formatting stage.
- Ouahib Alaoui: Provided general feedback, minor manuscript review, and offered occasional methodological input.

REFERENCES

1. Aïtcin, P.-C. (1998). High performance concrete: Materials, mixture design, and properties (1st ed.). CRC Press. https://doi.org/10.4324/9780203475034
2. Khayat, K. H., & Zhang, L. (2003). Influence of mix design parameters on the performance of self-consolidating concrete. ACI Materials Journal, 100(5), 441-450.
3. Zhao, X., et al. (2020). Predicting concrete compressive strength using machine learning. International Journal of Concrete Structures and Materials, 14(1), 1-9.

4. Zadeh, L. A., & Zadeh, R. (2021). A comprehensive study on the prediction of concrete strength using machine learning. Journal of Materials in Civil Engineering, 33(2), 04020388.

5. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189-1232.

6. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.

7. Fukui, M., Sato, R., Ohsawa, S., & Fujita, Y. (2019). Predicting the compressive strength of concrete using machine learning models. Construction and Building Materials, 224, 812–818. https://doi.org/10.1016/j.conbuildmat.2019.07.122

8. Fukui, M., Sato, R., Ohsawa, S., & Fujita, Y. (2019). Predicting the compressive strength of concrete using machine learning models. Construction and Building Materials, 224, 812–818. https://doi.org/10.1016/j.conbuildmat.2019.07.122

9. De Moura, M. F., Pinto, J. A., & Almeida, F. L. (2021). Predicting compressive strength of concrete using machine learning techniques. Journal of Building Engineering, 44, 102671. https://doi.org/10.1016/j.jobe.2021.102671

10. Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks (Vol. 4, pp. 1942–1948). IEEE. https://doi.org/10.1109/ICNN.1995.488968

11. Miranda, L., & Figueroa, M. (2020). Particle swarm optimization for multi-objective optimization of concrete mix designs. Computers and Concrete, 25(3), 261–276. https://doi.org/10.12989/cac.2020.25.3.261

12. Bishop, C. M. (2006). Pattern recognition and machine learning. Springer. https://doi.org/10.1117/1.2819119

13. Zhang, X., Ling, Y., & Lu, J. (2015). Machine learning approaches for predicting concrete compressive strength: A comparative study. Journal of Computing in Civil Engineering, 29(6), 04015053.

14. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30, pp. 4765–4774). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

15. Molnar, C. (2020). Interpretable machine learning: A guide for making black box models explainable (2nd ed.). Springer. Retrieved from https://christophm.github.io/interpretable-ml-book/

16. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

17. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

18. Fukui, M., Sato, R., Ohsawa, S., & Fujita, Y. (2019). Predicting the compressive strength of concrete using machine learning models. Construction and Building Materials, 224, 812–818. https://doi.org/10.1016/j.conbuildmat.2019.07.122

19. De Moura, M. F., Pinto, J. A., & Almeida, F. L. (2021). Predicting compressive strength of concrete using machine learning techniques. Journal of Building Engineering, 44, 102671. https://doi.org/10.1016/j.jobe.2021.102671

20. Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks (pp. 1942–1948). IEEE. https://doi.org/10.1109/ICNN.1995.488968

21. . Miranda, L., & Figueroa, M. (2020). Particle swarm optimization for multi-objective optimization of concrete mix designs. Computers and Concrete, 25(3), 261–276. https://doi.org/10.12989/cac.2020.25.3.261

22. Zhang, X., Ling, Y., & Lu, J. (2015). Machine learning approaches for predicting concrete compressive strength: A comparative study. Journal of Computing in Civil Engineering, 29(6), 04015053. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000513

23. Lundberg, S. M., Erion, G., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (pp. 4765–4774). [Proceedings link unavailable—often cited via NeurIPS proceedings]

24. Molnar, C. (2020). Interpretable machine learning. Springer. https://christophm.github.io/interpretable-ml-book/

25. Zhang, Z., Wang, Y., & Li, M. (2019). Predicting concrete compressive strength using machine learning methods. Computers and Concrete, 24(5), 397–409. https://doi.org/10.12989/cac.2019.24.5.397

26. Liu, J., Chen, P., & Qin, G. (2020). Predicting the mechanical properties of concrete using gradient boosting machines. Construction and Building Materials, 252, 119097. https://doi.org/10.1016/j.conbuildmat.2020.119097

27. Rahman, M. A., Hasan, M., & Mourshed, M. (2020). Machine learning techniques for predicting concrete compressive strength: A comparative study. Applied Soft Computing, 92, 106295. https://doi.org/10.1016/j.asoc.2020.106295

28. Cho, H., Park, S., & Lee, J. (2021). Optimizing concrete mix design using gradient boosting regression. Journal of Civil Engineering and Management, 27(4), 235–247. https://doi.org/10.3846/jcem.2021.14310

29. Wang, F., Xu, S., & Zhao, Y. (2020). A novel machine learning approach for predicting concrete compressive strength using gradient boosting. Construction and Building Materials, 276, 122257. https://doi.org/10.1016/j.conbuildmat.2020.122257

30. Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2), 123–140. https://doi.org/10.1007/BF00058655

31. Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

32. Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge University Press.

33. Zhang, Y., Li, X., & Huang, D. (2021). Predicting the mechanical properties of concrete using ensemble learning techniques. Construction and Building Materials, 307, 124827. https://doi.org/10.1016/j.conbuildmat.2021.124827

34. Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine Learning, 63(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

35. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

36. Zhao, X., Li, J., Zhang, Z., & Liu, Y. (2020). Predicting concrete compressive strength using machine learning. International Journal of Concrete Structures and Materials, 14(1), 1–9. https://doi.org/10.1007/s40069-019-00396-y

37. Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks (pp. 1942–1948). https://doi.org/10.1109/ICNN.1995.488968

38. Shi, Y., & Eberhart, R. (1998). A modified particle swarm optimizer. In Proceedings of the 1998 IEEE International Conference on Evolutionary Computation (pp. 69–73). IEEE. https://doi.org/10.1109/ICEC.1998.699146

39. Zhang, Q., & Zhang, Z. (2016). Particle swarm optimization: A survey and its applications. International Journal of Computational Intelligence Systems, 9(4), 752–769.

40. Poli, R., Kennedy, J., & Blackwell, T. (2007). Particle swarm optimization. Swarm Intelligence, 1(1), 33–57. https://doi.org/10.1007/s11721-007-0002-0

41. Chung, C. K., & Liao, C. F. (2013). Hyperparameter optimization of machine learning models using particle swarm optimization: A review. Computational Intelligence and Neuroscience, 2013, 1–15. https://doi.org/10.1155/2013/416936