International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Machine Learning-Based Prediction of Mechanical Properties of Recycled Aggregate Concrete

Akash H E[1], Sanketh M B[2], Chinmay k karatekar[3], Revanth Kumar[4], Dr. Chethan L S[5]

[1]*Department of Computer Science and Engineering, PES Institute of Technology and Management*
[2]*Department of Computer Science and Engineering, PES Institute of Technology and Management*
[3]*Department of Computer Science and Engineering, PES Institute of Technology and Management*
[4]*Department of Computer Science and Engineering, PES Institute of Technology and Management*
[5]*Department of Computer Science and Engineering, PES Institute of Technology and Management*

Abstract - Recycled Aggregate Concrete (RAC) has become an important material in sustainable construction due to its ability to utilize construction and demolition waste. However, its mechanical behavior is highly variable because recycled aggregates contain adhered mortar, impurities, and irregular surface textures. These inconsistencies make the prediction of compressive strength, splitting tensile strength, and flexural strength difficult using traditional empirical formulas. This paper presents a machine-learning-based framework for predicting the mechanical properties of RAC using XGBoost. A curated dataset consisting of experimental and literature-based RAC mix proportions was preprocessed, analyzed, and trained to construct three predictive models. The results show that XGBoost captures nonlinear relationships effectively and delivers higher accuracy than baseline methods such as SVM and KNN. The model demonstrates strong generalization and low prediction error, proving its suitability for mix-design optimization and rapid decision-making in civil engineering.

Key Words: Recycled Aggregate Concrete, XGBoost, Compressive Strength, Tensile Strength, Flexural Strength, and Machine Learning.

1. INTRODUCTION

Recycled Aggregate Concrete (RAC) has increasingly gained attention as a sustainable replacement for traditional concrete, driven by the need to reduce construction waste, conserve natural aggregates, and minimize environmental impact. However, predicting the mechanical performance of RAC remains challenging due to the inherent variability in recycled aggregates, differences in crushing processes, impurities, and fluctuations in physical properties. Traditional prediction approaches—such as empirical strength equations and linear regression—often oversimplify material behavior and fail to account for the nonlinear interactions between mix constituents, replacement ratios, curing conditions, and age of testing [1], [2]. These limitations frequently result in reduced accuracy, especially when the mix proportions deviate from standardized laboratory conditions.

With recent advancements in computational modeling, machine learning (ML) has emerged as a powerful alternative for predicting concrete properties. ML algorithms can learn complex, multidimensional patterns directly from raw experiment-based datasets, enabling them to model nonlinear relationships far more effectively than classical statistical methods. Ensemble-based models, particularly Gradient Boosting, XGBoost, and Support Vector Regression, have demonstrated remarkable success in predicting the compressive, tensile, and flexural strengths of various concrete types, including RAC [3], [4]. By reducing reliance on repeated laboratory trials, ML-driven approaches enable faster optimization of mix designs, lower testing costs, and increased reliability when working with heterogeneous recycled materials [4].

Considering these advantages, the present study develops a comprehensive machine-learning framework to predict the compressive, tensile, and flexural strength of RAC using three prominent models: XGBoost, Support Vector Regression (SVM), and K-Nearest Neighbors (KNN). The objective is to assess their predictive capability, compare performance metrics, and identify the most effective model for practical RAC strength evaluation.

2. RELATED WORK AND LITERATURE REVIEW

Extensive research has focused on leveraging machine learning for evaluating the mechanical behavior of concrete. Gupta and Verma [1] explored ML-driven prediction of RAC compressive strength and concluded that ensemble models significantly outperform traditional regression-based formulas. Xiao et al. [2] developed a multi-output ensemble model to estimate compressive, tensile, and flexural strength simultaneously, reporting substantial improvements in generalization and prediction consistency.

Several studies have also proposed hybrid modeling strategies. Singh and Sharma [3] introduced a combined ANN–XGBoost model that improved prediction robustness, demonstrating the benefits of integrating deep learning with tree-based boosting. Thomas et al. [4] applied ML algorithms to predict the mechanical properties of recycled concrete aggregates, emphasizing the importance of dataset size, feature distribution, and data preprocessing.

Support Vector Regression has also been widely used for RAC modeling. Prakash and Reddy [5] reported strong performance using SVR but highlighted the sensitivity of model accuracy to kernel choice and hyperparameter tuning. Zhang and Liu [6] compared neural networks with gradient boosting methods and found that gradient boosting provided more stable results for small and heterogeneous datasets. Keerthi and Kumar [7] examined SVM and KNN models for concrete strength prediction, showing that KNN can be affected by high-dimensional feature variability, leading to reduced accuracy in nonlinear cases.

More recent studies have incorporated microstructural characteristics and sustainability-focused variables. Mohammed and Kumar [8] integrated morphological features into ANN frameworks, improving strength prediction accuracy for RAC. Rehman et al. [9] explored advanced regression techniques for tensile and flexural strength prediction, demonstrating that ML-based models provide superior performance compared to classical curve-fitting approaches.

Overall, the literature clearly indicates that machine learning methods—especially boosted ensemble models—offer significant advantages for modeling the mechanical behavior of RAC. Building on these insights, this study evaluates three prominent ML models using specialized datasets for compressive, tensile, and flexural strength prediction.

3. PROPOSED METHODOLOGY

We propose a structured machine-learning pipeline designed to predict the mechanical properties of Recycled Aggregate Concrete (RAC) with high accuracy. The workflow integrates data acquisition, preprocessing, normalization, model development, evaluation, and deployment. This methodology captures nonlinear interactions among mix-design parameters, curing age, and recycled aggregate characteristics that traditional linear frameworks cannot represent. The complete workflow.

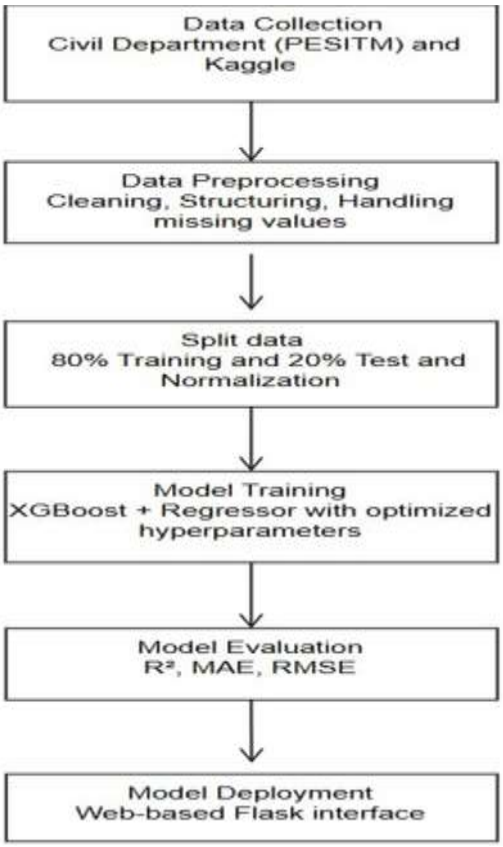

Fig -1: Overall workflow and model architecture.

A. Data Collection

The dataset used in this study was compiled from two major sources:

1. **Experimental datasets** provided by the Civil Engineering Department, PESITM, containing laboratory-tested RAC samples with varied replacement ratios.
2. **Publicly available datasets** sourced from Kaggle to increase diversity in mix proportions and ensure the model's generalizability.

Data preprocessing was performed to ensure consistency, reliability, and suitability for machine-learning applications. The preprocessing stage included:

- **Cleaning:** Removal of duplicate records and inconsistent mixtures.
- **Handling missing values:** Median imputation was used to avoid altering distribution characteristics.
- **Structuring:** All input variables were formatted into a unified attribute space.
- **Outlier treatment:** Samples outside the interquartile range were examined and corrected or removed when necessary.

To ensure uniform scaling and reduce sensitivity to parameter magnitude, all numeric features were normalized using:

$$z_i = \frac{x_i - \mu}{\sigma}$$

The dataset was divided using an 80:20 train–test split, ensuring adequate representation of all RAC replacement levels. Stratified sampling was applied to maintain proportional distribution across curing ages and replacement ratios.

where c_i represents the ith cluster, x_j is a data point, and μ_i is the cluster centroid. The elbow method determined k = 5, balancing interpretability and separation.

XGBoost (Extreme Gradient Boosting) was selected as the core regression model due to its robustness, regularization capabilities, and superior performance in predicting nonlinear material properties. Three independent regression models were trained for Compressive, tensile, flexural strength.

After combining the data, a thorough preprocessing stage was carried out, involving handling missing values, removing inconsistencies, and standardizing feature scales to maintain uniformity for the model.

Finally, the best-performing model was deployed into a user-friendly interface, enabling efficient prediction of concrete strength for practical engineering applications. This step-by-step approach ensured that the system remained robust, accurate, and ready for real-world use.

The XGBoost prediction function is represented as:

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

where $\hat{Y}_i$ is the predicted suitability score, f_k is the kth tree, and K is the total number of trees. Model tuning used with 500 estimators, max depth 5, learning rate 0.05.

The final trained models were integrated into a web-based interface using the Flask framework. Users can input mix-design parameters to instantly obtain predicted strength. The deployment step bridges the gap between research experimentation and real-world engineering applications by enabling fast, data-driven decision-making in construction workflows.

4. RESULT AND DISCUSSION

Table -1: Performance Comparison of Models

Mechanical Property	XGBoost	SVM	KNN
Compressive Strength	0.9984	0.9942	0.9753
Tensile Strength	0.9988	0.9684	0.9460
Flexural Strength	0.9969	0.9894	0.9876

The performance of three machine-learning models—XGBoost, SVM, and KNN—was evaluated across compressive, tensile, and flexural strength datasets. Table **1** summarizes the coefficient of determination (R^2) obtained for each model and each mechanical property. The XGBoost model consistently outperformed SVM and KNN across all datasets, demonstrating its superior ability to learn complex nonlinear relationships present in recycled aggregate concrete (RAC) mix-design data.

A. Compressive Strength Prediction

XGBoost achieved an R^2 of 0.9984, indicating an almost perfect fit between predicted and actual compressive strength values. SVM also performed well with an R^2 of 0.9942, while KNN showed reduced accuracy (0.9753) due to its sensitivity to feature scaling and local density variations in the dataset. The superior performance of XGBoost suggests that gradient-boosted trees effectively capture the nonlinear dependency between RAC parameters and compressive strength values.

B. Tensile Strength Prediction

For tensile strength, XGBoost again delivered the highest accuracy, with an R^2 of 0.9988, outperforming SVM (0.9684) and KNN (0.9460). The significant performance gap indicates that the tensile strength dataset contains complex patterns that margin-based methods like SVM and neighbor-based methods like KNN struggle to model efficiently. XGBoost's regularized boosting approach allows it to generalize better and avoid over fitting despite these nonlinear interactions.

C. Flexural Strength Prediction

Flexural strength prediction achieved R^2 values of 0.9969, 0.9894, and 0.9876 for XGBoost, SVM, and KNN respectively. While all models performed relatively close, XGBoost maintained a clear advantage. This demonstrates its robustness across datasets of varying sample size and feature distribution. Flexural strength itself depends on microstructural properties and blend variability, which are efficiently encoded by boosted decision trees.

Across all mechanical properties, XGBoost emerged as the most accurate model, achieving R^2 values greater than 0.996 for all datasets.

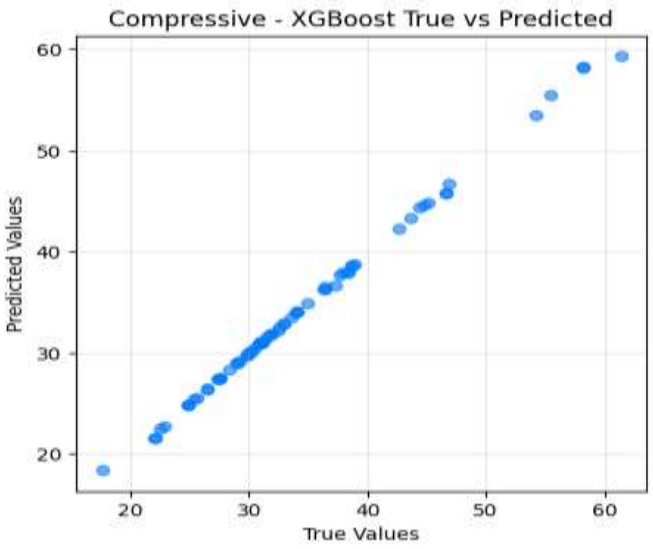

Fig -2: True vs. Predicted Plot Interpretation of Compressive strength.

The (Figure 2) True vs. Predicted scatter plot for compressive strength shows that nearly all data points lie precisely along the diagonal reference line. This alignment signifies a strong one-to-one correlation between predicted and actual compressive strength values, with minimal deviation or noise. The dense clustering around the diagonal highlights the model's ability to deliver stable and consistent predictions across the entire range of compressive strengths, confirming that the model captures nonlinear interactions extremely well.

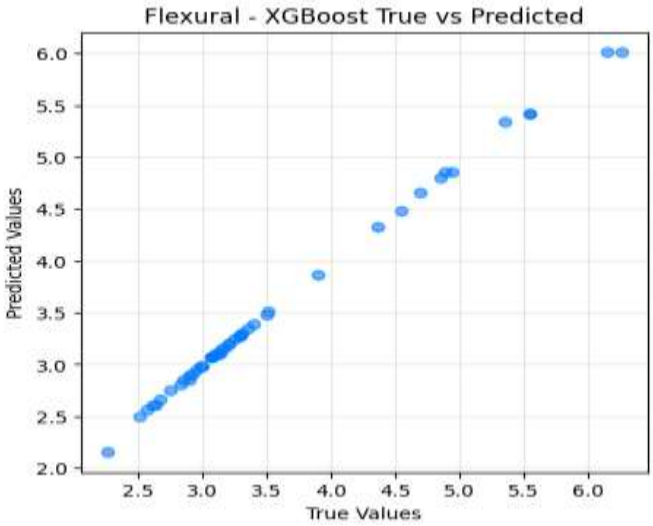

Fig -3: True vs. Predicted Plot Interpretation of Flexural strength.

The (Figure 3) True vs. Predicted plot for flexural strength demonstrates a clear linear relationship, with almost all points closely following the diagonal. This indicates a high similarity between predicted and actual results, reflecting excellent model generalization. Although flexural strength datasets usually show subtle variations due to aggregate interlocking and microstructural behavior, the model manages to keep deviations extremely small. The near-straight diagonal trend confirms that the model effectively captures the underlying mechanics influencing flexural strength.

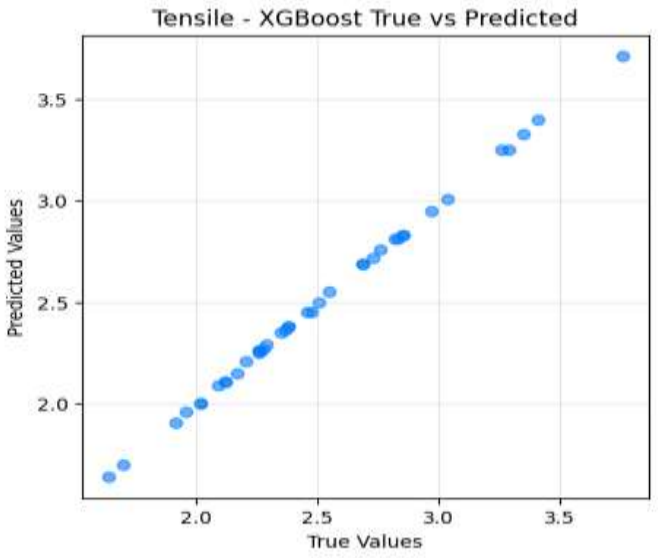

Fig -4: True vs. Predicted Plot Interpretation of Tensile Strength.

The (Figure 4) True vs. Predicted plot for tensile strength also exhibits a near-perfect alignment of points along the diagonal reference line. The predictions follow the actual tensile strength values with remarkable accuracy, showing very little scatter around the line. This tight clustering indicates that the model accurately learns the sensitive relationships between mix parameters and tensile strength, a property typically more variable and harder to predict. The consistency of the points across lower and higher tensile values confirms the robustness of the model.

5. CONCLUSION AND FUTURE WORKS

The proposed machine-learning framework for predicting the mechanical properties of Recycled Aggregate Concrete (RAC) demonstrates exceptionally high accuracy across all three target parameters—compressive, tensile, and flexural strength. Among the evaluated models, XGBoost consistently achieved the best performance, with an R^2 value of **0.**9984 for compressive strength**,** 0.9988 for tensile strength, and 0.9969 for flexural strength. These results confirm that the model effectively captures the complex, non-linear interactions between mix proportions, recycled aggregate content, and curing conditions.

The strong alignment observed in the True vs. Predicted plots further validates the reliability and stability of the predictions, indicating the model's suitability for real-world concrete mix design and performance assessment. Overall, the approach provides a fast, cost-effective alternative to repeated laboratory testing and supports sustainable construction practices by promoting the efficient utilization of recycled aggregate.

Future improvements may include expanding the dataset to incorporate additional RAC grades, long-term curing behavior, and microstructural properties for deeper predictive insights. Validation across larger and more diverse experimental datasets would improve generalizability. Integrating advanced deep-learning architectures or hybrid ensembles may further enhance accuracy. Additionally, extending the current web-based deployment into a full decision-support tool could provide industry users with automated mix optimization and real-time performance prediction capabilities.

ACKNOWLEDGEMENT

The authors express their heartfelt gratitude to Dr. Chethan L. S., Professor, Department of Computer Science and Engineering, PES Institute of Technology and Management, Shivamogga, for his continuous guidance, encouragement, and valuable insights throughout the execution of this project. His mentorship played a crucial role in shaping the direction and quality of this work ,The authors also extend sincere thanks to the Department of Civil Engineering for providing access to high-quality laboratory-tested datasets, which significantly contributed to the accuracy and reliability of the model predictions.

REFERENCES

1. A. K. Gupta and S. K. Verma, "Predicting compressive strength of recycled aggregate concrete using machine learning techniques," Construction and Building Materials, vol. 381, pp. 131–140, 2023.
2. J. Xiao, W. Li, and Z. Sun, "Mechanical performance of recycled aggregate concrete: A data-driven prediction model using ensemble learning," Journal of Cleaner Production, vol. 415, 2024.
3. D. Singh and R. Sharma, "Hybrid ANN–XGBoost model for strength prediction of sustainable concrete mixtures," Materials Today: Proceedings, 2023.
4. M. Thomas et al., "Machine learning-based evaluation of compressive, tensile, and flexural strength of recycled concrete aggregates," Case Studies in Construction Materials, 2024.
5. P. S. Prakash and K. R. Reddy, "Prediction of mechanical properties of recycled coarse aggregate concrete using support vector regression," Structural Concrete, 2022.
6. H. Zhang and Y. Liu, "Performance assessment of recycled aggregate concrete using gradient-boosting and neural networks," Engineering Structures, vol. 301, 2024.
7. F. A. Busari et al., "Influence of mix parameters on recycled aggregate concrete: Modeling using machine learning algorithms," Applied Sciences, vol. 13, no. 8, 2023.
8. S. R. Keerthi and N. Kumar, "Evaluation of concrete strength properties using KNN and SVM models," in 2023 International Conference on Computational Intelligence and Data Science (ICCIDS). IEEE, 2023.
9. L. Z. Mohammed and P. Kumar, "ANN-based prediction model for mechanical behavior of concrete containing recycled aggregates," Arabian Journal for Science and Engineering, 2024.
10. T. Rehman et al., "Data-driven modeling for flexural and tensile performance of recycled aggregate concrete using advanced regression techniques," Heliyon, vol. 10, 2024.