Revista ALCONPAT
www.revistaalconpat.org
eISSN 2007-6835
Revista de la Asociación Latinoamericana de Control de Calidad, Patología y Recuperación de la Construcción

Machine learning approaches for estimating the strength performance of recycled aggregate concrete in rigid pavement systems.

N. Kumar Gautam[1]*, A. Kumar[2]
*Contact author: navinphd2026@gmail.com
DOI: https://doi.org/10.21041/ra.v16i3.1025

Received: 20/02/2026 | Received in revised form: 22/08/2026 | Accepted: 31/08/2026 | Published: 01/09/2026

ABSTRACT
This study aims to develop a method for predicting the 28-day compressive strength of recycled aggregate concrete (RAC) for rigid pavement applications. A hybrid dataset of 385 observations, combining laboratory results and selected literature data, was used to develop and compare machine learning models. The models were assessed using five-fold cross-validation, error measures, bias analysis, multicollinearity assessment, and SHAP interpretation. Lasso Regression provided the best performance, with $R^2 = 0.7713$, MAE = 3.51 MPa, and RMSE = 4.47 MPa. The study is limited to five-fold cross-validation without independent external validation. Its originality lies in evaluating a pavement-oriented hybrid RAC dataset using predictive and interpretive analyses. The results show that regularized models can support preliminary RAC mix evaluation and material assessment.
Keywords: recycled aggregate concrete; rigid pavement; compressive strength prediction; machine learning; lasso regression; sustainable pavement materials.

Cite as: Kumar Gautam, N., Kumar, A. (2026), "*Machine learning approaches for estimating the strength performance of recycled aggregate concrete in rigid pavement systems.*", Revista ALCONPAT, 16 (3), pp. 493 – 520, DOI: https://doi.org/10.21041/ra.v16i3.1025.

[1] Research Scholar, Department of Civil Engineering, Sandip University, Sijoul, Madhubani-847235, Bihar, India.
[2] Assistant Professor, Department of Civil Engineering, Sandip University, Sijoul, Madhubani-847235, Bihar, India.

Contribution of each author
In this work, Navin Kumar Gautam contributed Conceptualization, experimental investigation, dataset development, machine learning modeling, formal analysis, visualization, manuscript preparation, and revision (75% contribution). Dr. Amrendra Kumar contributed to Supervision, methodological review, validation, interpretation of results, critical review, and manuscript editing (25% contribution). All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Discussions and subsequent corrections to the publication
Any dispute, including the replies of the authors, will be published in the second issue of 2027 provided that the information is received before the closing of the first issue of 2027.

Enfoques de aprendizaje automático para estimar el rendimiento de resistencia del hormigón agregado reciclado en sistemas de pavimento rígido.

RESUMEN

Este estudio tiene como objetivo desarrollar un método para predecir la resistencia a la compresión de 28 días del hormigón agregado reciclado (RAC) para aplicaciones de pavimento rígido. Se utilizó un conjunto de datos híbrido de 385 observaciones, combinando resultados de laboratorio y datos de literatura seleccionada, para desarrollar y comparar modelos de aprendizaje automático. Los modelos se evaluaron mediante validación cruzada de cinco vías, medidas de error, análisis de sesgos, evaluación multicolineal e interpretación SHAP. La regresión de lazo proporcionó el mejor rendimiento, con R^2 = 0,7713, MAE = 3,51 MPa y RMSE = 4,47 MPa. El estudio está limitado a una validación cruzada de cinco vías sin validación externa independiente. Su originalidad radica en evaluar un conjunto de datos híbrido de RAC orientado al pavimento mediante análisis predictivos e interpretativos. Los resultados muestran que los modelos regularizados pueden soportar la evaluación preliminar de la mezcla RAC y la evaluación de materiales.

Palabras clave: hormigón agregado reciclado; pavimento rígido; predicción de resistencia a la compresión; aprendizaje automático; regresión de lazo; materiales sostenibles para pavimentos.

Abordagens de aprendizado de máquina para estimar o desempenho de resistência do concreto agregado reciclado em sistemas de pavimentação rígida.

RESUMO

Este estudo tem como objetivo desenvolver um método para prever a resistência à compressão de 28 dias do concreto agregado reciclado (RAC) para aplicações em pavimentos rígidos. Um conjunto de dados híbrido de 385 observações, combinando resultados laboratoriais e dados selecionados da literatura, foi utilizado para desenvolver e comparar modelos de aprendizado de máquina. Os modelos foram avaliados usando validação cruzada de cinco vezes, medidas de erro, análise de viés, avaliação multicolinear e interpretação SHAP. A Regressão de Laço forneceu o melhor desempenho, com R^2 = 0,7713, MAE = 3,51 MPa e RMSE = 4,47 MPa. O estudo é limitado a validação cruzada de cinco vezes sem validação externa independente. Sua originalidade está na avaliação de um conjunto de dados híbrido RAC orientado para pavimento, utilizando análises preditivas e interpretativas. Os resultados mostram que modelos regularizados podem suportar a avaliação preliminar da mistura RAC e avaliação de materiais.

Palavras-chave: concreto agregado reciclado; pavimento rígido; previsão de resistência à compressão; aprendizado de máquina; regressão de laço; materiais sustentáveis para pavimento.

Legal Information

Revista ALCONPAT is a quarterly publication by the Asociación Latinoamericana de Control de Calidad, Patología y Recuperación de la Construcción, Internacional, A.C., Km. 6 antigua carretera a Progreso, Mérida, Yucatán, 97310, Tel. +52 1 983 419 8241, alconpat.int@gmail.com, Website: www.alconpat.org

Reservation of rights for exclusive use No.04-2013-011717330300-203, and ISSN 2007-6835, both granted by the Instituto Nacional de Derecho de Autor. Responsible editor: Pedro Castro Borges, Ph.D. Responsible for the last update of this issue, Informatics Unit ALCONPAT, Elizabeth Sabido Maldonado.

The views of the authors do not necessarily reflect the position of the editor.

1. INTRODUCTION

Rigid pavements are extensively used in highway and airport infrastructure because of their high load carrying capacity, structural rigidity, and long service life under repeated traffic loading. The structural performance and durability of rigid pavements are largely governed by the mechanical properties of concrete slabs, among which compressive strength is a fundamental parameter used in pavement thickness design, fatigue evaluation, and construction quality control (American Association of State Highway & Officials, 1993; Brandes et al., 2018; Huang, 2004). Reliable estimation of compressive strength is therefore essential to ensure safety, serviceability, and cost effectiveness in rigid pavement systems.

In recent years, the growing emphasis on sustainable infrastructure development and the depletion of natural aggregate resources have encouraged the use of recycled aggregate concrete (RAC) in pavement construction. Recycled aggregates obtained from construction and demolition waste offer significant environmental benefits by reducing landfill disposal and conserving natural resources (Behera et al., 2014; Silva et al., 2014; Tam et al., 2018). Several studies have reported that RAC can be effectively utilized in rigid pavement applications such as base layers, shoulders, and even structural concrete slabs, provided that appropriate mix design and quality control measures are adopted (Ismail et al., 2022; Xiao et al., 2018).

Despite these advantages, the mechanical performance of RAC is generally more variable than that of conventional concrete. Recycled aggregates often contain adhered old mortar, internal microcracks, and higher porosity, which increase material heterogeneity and lead to greater dispersion in compressive strength results (Etxeberria et al., 2007; Limbachiya et al., 2000; Thomas et al., 2013). The compressive strength of RAC is influenced by a complex interaction of factors, including aggregate replacement ratio, water-cement ratio, aggregate quality, and curing conditions (Butler et al., 2016; Pedro et al., 2017). These nonlinear and interdependent effects limit the applicability of traditional empirical or regression-based prediction models, especially when data are drawn from multiple sources (Asteris et al., 2020).

Conventional laboratory testing remains the most reliable approach for determining compressive strength; however, extensive experimental programs are time consuming, labour intensive, and costly. This challenge is particularly evident in rigid pavement projects, where multiple mix designs may need to be evaluated within limited timeframes during material selection and preliminary design stages (Chou et al., 2014; Neville, 2011). Consequently, there is increasing interest in complementary predictive methods that can reduce experimental effort while maintaining acceptable levels of accuracy.

Machine learning (ML) techniques have emerged as effective tools for modeling complex material behavior due to their ability to capture nonlinear relationships and interactions among multiple input variables without requiring predefined functional forms. In concrete engineering, ML methods have been successfully applied to predict compressive strength, elastic modulus, and durability-related properties using algorithms such as artificial neural networks, support vector machines, and ensemble learning techniques (Asteris et al., 2019; Chou & Pham, 2013; Duan et al., 2021; Yeh, 1998). Recent studies indicate that ML models can outperform conventional statistical approaches when dealing with heterogeneous materials such as RAC (Nguyen et al., 2021; Zhang et al., 2020).

Although machine learning has increasingly been applied to the prediction of RAC properties, its effectiveness depends strongly on the quality, diversity, and consistency of the data used for model development. Many existing investigations are based on datasets obtained from a single experimental programme or from narrowly defined literature sources. Such datasets are useful for establishing relationships between mixture parameters and concrete strength, but they may not adequately represent the variability associated with different recycled aggregate sources, parent concrete qualities, mixture proportions, and curing conditions encountered in practical applications. This issue is particularly relevant to rigid pavement concrete, where variations in material properties can influence strength-based design and quality assessment.

The present study addresses this limitation by developing a hybrid dataset that combines controlled experimental observations with systematically selected literature-derived data for recycled aggregate concrete. The purpose of this integration is not simply to increase the number of observations, but to broaden the range of material and mixture conditions represented in the modeling dataset while retaining experimentally grounded information. The resulting dataset incorporates variations in cement content, water-cement ratio, aggregate composition, recycled aggregate characteristics, parent concrete strength, admixture dosage, workability, and curing conditions. This provides a broader basis for evaluating compressive strength prediction for recycled aggregate concrete used in rigid pavement applications.

The main contribution of this study lies in the development and systematic evaluation of a pavement-oriented hybrid experimental–literature dataset for RAC strength prediction, rather than in the use of any individual machine learning algorithm. Established regression, kernel-based, and ensemble learning methods are compared under a common validation framework, with their performance further examined using error measures, bias analysis, multicollinearity assessment, and feature-importance interpretation. The objective is to determine which modeling approach provides stable and practically useful predictions for the heterogeneous RAC dataset considered in this study. The findings are intended to support preliminary material assessment, mix evaluation, and quality-control decisions in sustainable rigid pavement applications, while recognizing that laboratory verification remains necessary for final engineering decisions.

2. RESEARCH OBJECTIVES

The primary objective of this study is to develop a reliable and practical framework for predicting the compressive strength of recycled aggregate concrete (RAC) intended for rigid pavement applications by integrating experimental investigation with machine learning techniques.

To achieve this overall goal, the specific objectives of the study are as follows:

1. To experimentally evaluate the 28-day compressive strength of recycled aggregate concrete prepared with mix proportions relevant to rigid pavement construction.
2. To develop a hybrid dataset by combining experimentally generated data with carefully selected compressive strength data obtained from reviewed literature, ensuring broader representation of material variability.
3. To apply and compare different machine learning models, including regularized linear models, ensemble learning methods, and kernel-based approaches, for predicting the compressive strength of RAC.
4. To assess the predictive performance of the developed models using multiple statistical indicators, including error-based metrics, bias analysis, and hypothesis testing, to ensure robustness and reliability.
5. To identify suitable machine learning approaches that can support material level design, quality assessment, and decision-making for the sustainable use of recycled aggregate concrete in rigid pavement engineering.

3. EXPERIMENTAL PROGRAM

3.1 Materials.

The experimental program was designed to investigate the compressive strength behavior of recycled aggregate concrete (RAC) intended for rigid pavement applications. Ordinary Portland cement conforming to standard requirements was used as the primary binder. Natural aggregates and recycled concrete aggregates (RCA) were employed as coarse aggregate constituents, while natural river sand was used as fine aggregate. The use of RCA reflects practical pavement construction scenarios where recycled materials are incorporated to improve sustainability (Silva et al., 2014; Tam et al., 2018).

The recycled concrete aggregates were obtained from crushed parent concrete of varying strength levels, representing practical demolition waste conditions. Parent concrete strength was considered an important parameter due to its influence on the quality and mechanical performance of RCA (Katz, 2003; Pedro et al., 2017; Xiao et al., 2018). The maximum nominal size of RCA varied between 10 mm and 20 mm, which is consistent with aggregate gradations commonly adopted in rigid pavement concrete (Huang, 2004). The water absorption capacity of RCA, ranging approximately from 3.5% to 8.5%, was recorded to account for the higher porosity and moisture demand associated with recycled aggregates (Butler et al., 2016; Etxeberria et al., 2007). A polycarboxylate based superplasticizer was used to achieve the desired workability across different mix compositions. In some mixes, a small proportion of fibers or mineral admixtures was incorporated to reflect variations commonly reported in pavement related concrete studies (Li et al., 2022).

Table 1 presents the complete experimental dataset developed in this study for recycled aggregate concrete. The table summarizes mix proportions, properties of recycled concrete aggregates, curing conditions, and the corresponding 28-day compressive strength results. Cement content and water-cement ratio were varied over a wide range to capture differences in binder intensity and mix workability. The replacement of natural coarse aggregates with recycled concrete aggregates was systematically considered from 0 to 100%, allowing evaluation of the influence of recycled aggregate incorporation on strength performance.

Table 1. Summary of experimental dataset including mix composition, recycled concrete aggregate properties, curing regime, and 28-day compressive strength results.

Parameter	Range / Levels
Cement Content (kg/m³)	350 - 420
Water-Cement Ratio (w/c)	0.40, 0.45, 0.50, 0.55
Natural Aggregate (NA) Replacement (%)	100, 75, 50, 25, 0
Recycled Coarse Aggregate (RCA) Content (%)	0, 25, 50, 75, 100
Fine Aggregate (kg/m³)	650 - 780
Parent Concrete Strength (MPa)	25 - 45
RCA Water Absorption (%)	3.5 - 8.5
RCA Maximum Size (mm)	10, 12.5, 20
Superplasticizer Dosage (%)	0.40 - 1.20
Slump (mm)	60 - 140
Curing Method	Water, Moist, Ambient
Curing Temperature (°C)	26 - 34
Fiber/Mineral Admixture (%)	0 - 1.50
28-Day Compressive Strength (MPa)	11.29 - 35.26

Table 2. Classification of experimental concrete mixes according to recycled concrete aggregate (RCA) replacement level.

Group	Mix ID Range	Natural Aggregate (%)	RCA Replacement Level (%)
Group I	M1–M16	100	0
Group II	M17–M32	75	25
Group III	M33–M48	50	50
Group IV	M49–M64	25	75
Group V	M65–M80	0	100

Note: *RCA = recycled concrete aggregate; NA = natural coarse aggregate. Each group includes multiple mixes prepared with varying water–cement ratios, cement contents, curing regimes, and admixture dosages to represent a broad experimental domain.*

Table 2 summarizes the classification of the experimental concrete mixes based on the percentage replacement of natural coarse aggregate with recycled concrete aggregate. The mixes were organized into five groups ranging from conventional concrete with no recycled aggregate to fully recycled aggregate concrete. This grouping framework enables a systematic comparison of the influence of RCA content on the fresh and hardened properties of concrete while maintaining consistency across other mix design parameters.

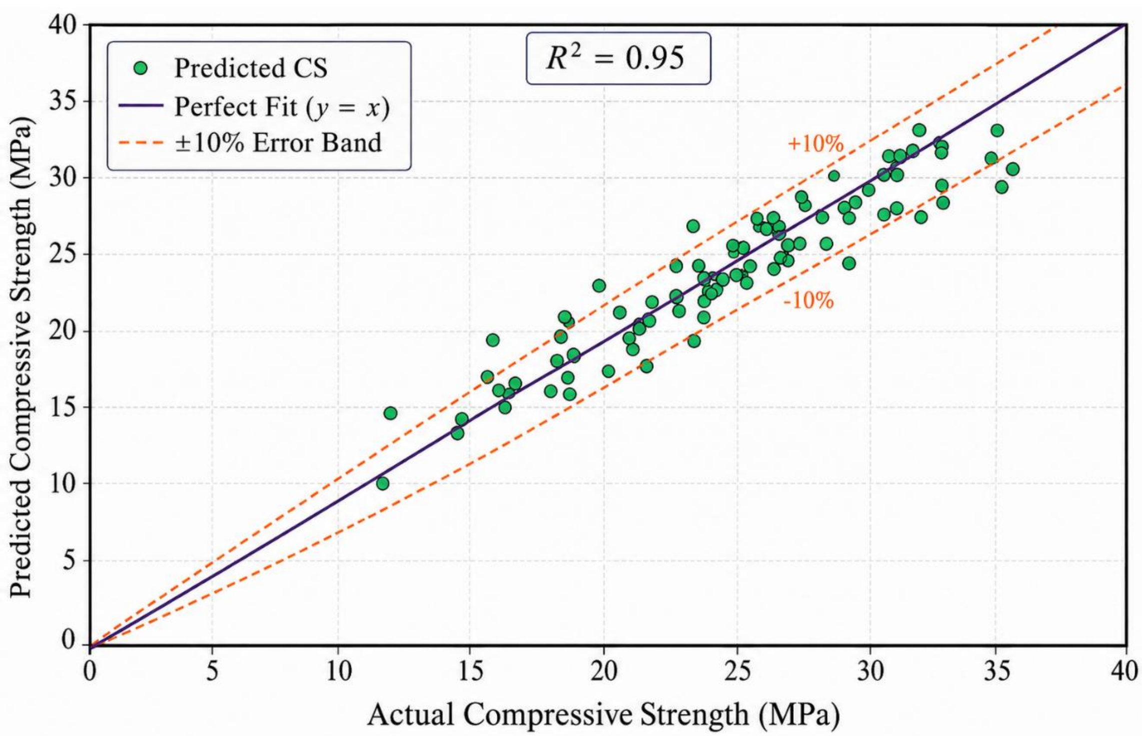

Figure 1. Actual versus predicted 28-day compressive strength obtained using the combined input parameters of the experimental recycled aggregate concrete dataset.

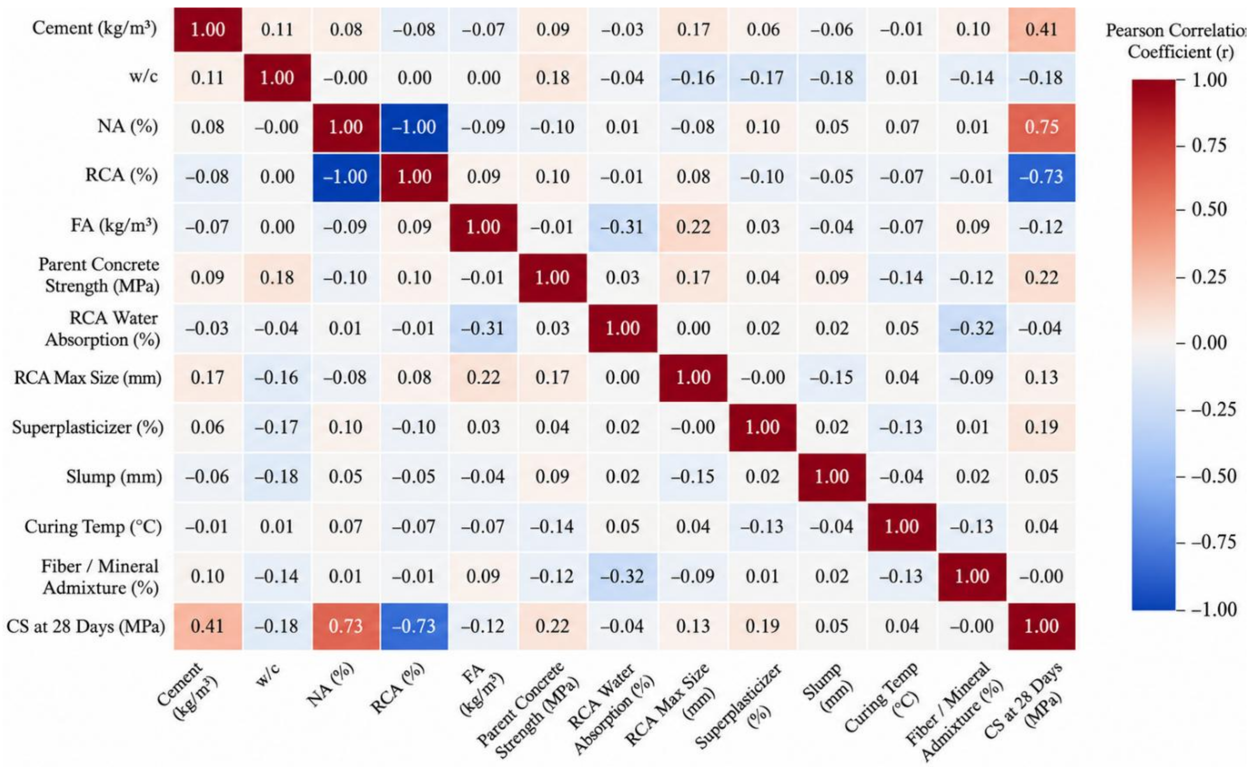

Figure 2. Correlation heatmap showing linear relationships between selected input parameters and 28-day compressive strength of recycled aggregate concrete.

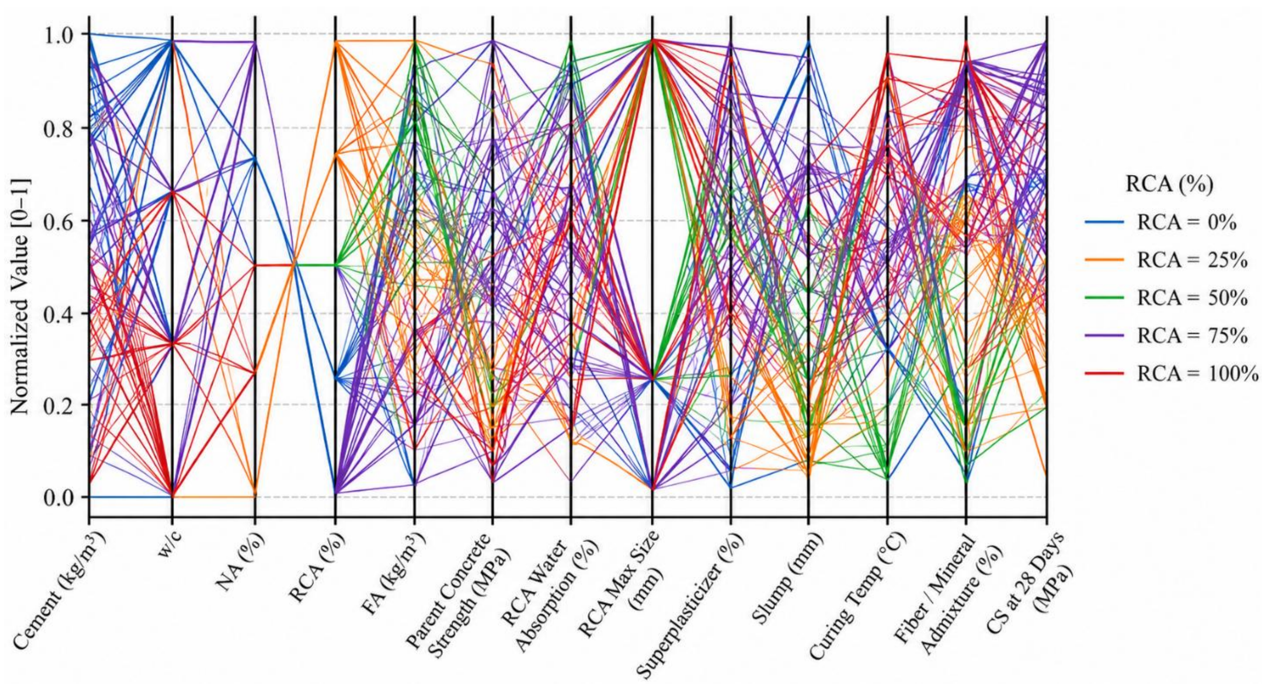

Figure 3. Parallel coordinates plot showing the normalized distribution of experimental input parameters for recycled aggregate concrete mixtures at different recycled coarse aggregate (RCA) replacement levels.

Figures 1-3 provide complementary views of the experimental dataset and its principal relationships with 28-day compressive strength. Figure 1 shows the agreement between measured and estimated strength values, with the data distributed close to the reference line and within the

±10% deviation limits, indicating a strong overall relationship ($R^2 = 0.95$). Figure 2 presents the correlation among the selected input variables and compressive strength, while Figure 3 illustrates the multidimensional distribution of the experimental mixtures across different RCA replacement levels. Together, these figures demonstrate the variability of the experimental dataset and the combined influence of mixture composition and recycled aggregate characteristics on strength development.

The R^2 value of 0.95 shown in Figure 1 should be interpreted as a descriptive relationship within the experimental dataset and not as the predictive accuracy of an individual machine learning model. The machine learning models were evaluated separately using five-fold cross-validation, and the corresponding validation results are presented in Section 5. Therefore, the cross-validated performance measures reported later in the manuscript provide the appropriate basis for comparing the predictive capability of the developed models.

3.2 Mix Proportions and experimental variables.

A comprehensive range of concrete mixes was prepared by systematically varying key mix design parameters relevant to rigid pavement performance. Cement content ranged approximately from 350 to 420 kg/m³, while the water-cement ratio varied between 0.40 and 0.55. These ranges are typical for pavement concrete designed to balance strength, durability, and workability (American Association of State Highway & Officials, 1993; Neville, 2011).

The replacement ratio of natural coarse aggregate with recycled concrete aggregate was varied from 0% to 100%, corresponding to natural aggregate contents of 100%, 75%, 50%, 25%, and 0%. This wide range enabled evaluation of the full transition from conventional concrete to fully recycled aggregate concrete, which is critical for assessing the feasibility of RAC in rigid pavement applications (Fanijo, 2023; Xiao et al., 2018).

Fine aggregate content was adjusted to maintain suitable grading and workability across all mixes. Slump values generally ranged from approximately 60 mm to 140 mm, ensuring workable mixes suitable for pavement concrete placement and compaction (Huang, 2004). The dosage of superplasticizer varied accordingly to compensate for increased water demand associated with higher RCA content.

3.3 Specimen preparation and curing conditions.

Concrete mixing was carried out following standard laboratory procedures to ensure uniformity and repeatability. Particular attention was given to moisture conditioning of RCA prior to mixing to minimize uncontrolled water absorption during batching, as recommended in previous studies on recycled aggregate concrete (Etxeberria et al., 2007; Pedro et al., 2017).

Fresh concrete was cast into standard cube or cylindrical molds for compressive strength testing. After casting, specimens were compacted adequately to remove entrapped air and ensure proper consolidation. Different curing regimes were adopted, including water curing, moist curing, and ambient curing, to reflect the range of curing conditions commonly encountered in pavement construction practice (Neville, 2011; Tam et al., 2005). Curing temperatures varied approximately between 26 °C and 34 °C, consistent with typical field conditions in pavement projects.

3.4 Compressive strength testing.

Compressive strength tests were conducted at 28-days, which is the standard reference age used for pavement concrete design and quality control. Testing was performed in accordance with established standards, and the average compressive strength was recorded for each mix (American Association of State Highway & Officials, 1993; Neville, 2011).

The experimentally measured 28-day compressive strength values exhibited a wide range due to variations in cement content, water-cement ratio, RCA replacement level, parent concrete strength, and curing conditions. This variability reflects realistic behavior of RAC used in rigid pavement applications and provides a robust foundation for subsequent data-driven modeling (Li et al., 2022; Pedro et al., 2017).

4. DATASET PREPARATION AND MACHINE LEARNING METHODOLOGY

4.1 Formation of hybrid dataset for pavement-oriented modelling.

The machine learning dataset was developed using a hybrid approach that integrates experimentally generated data with data collected from peer-reviewed literature. The experimental results obtained in the present study form the core of the dataset and provide a reliable baseline, as they were produced under controlled laboratory conditions with mix proportions relevant to rigid pavement concrete. To enhance the representativeness of the dataset and capture variability encountered in practical pavement construction, the experimental data were augmented with carefully selected literature data.

Only studies reporting complete information on mix composition, recycled aggregate characteristics, workability, curing conditions, and 28-day compressive strength were considered. Experimental investigations conducted in Indian and international contexts were prioritized, as they reflect material sources, curing environments, and construction practices similar to those encountered in rigid pavement projects (Ojha et al., 2024; Patil, 2025; Tank et al., 2014). Foundational studies addressing the influence of parent concrete strength, recycled aggregate quality, and replacement ratio were included to ensure adequate coverage across different performance levels (Limbachiya et al., 2000; Padmini et al., 2009).

Additional data were informed by studies examining recycled aggregate treatment methods, water absorption characteristics, and interfacial transition zone behavior, which are known to significantly affect strength development in recycled aggregate concrete (Chauhan & Singh, 2023; Ismail et al., 2022; Tam et al., 2005). Recent review and experimental studies were also considered to support inclusion of mineral admixtures and fiber content as secondary variables influencing pavement concrete performance (Jagadesh et al., 2024).

The final hybrid dataset comprised 385 samples and represents a wide range of mix compositions, recycled aggregate replacement levels (0-100%), curing regimes, and strength outcomes. This dataset structure is particularly suitable for rigid pavement material modeling, where variability in material properties must be explicitly accounted for during design and quality assessment.

4.2 Selection of input variables and pavement relevance.

The input variables selected for machine learning modeling were chosen based on their direct relevance to compressive strength development and rigid pavement performance. These include cement content, water-cement ratio, natural aggregate content, recycled aggregate replacement ratio, fine aggregate content, parent concrete strength, recycled aggregate water absorption, maximum aggregate size, superplasticizer dosage, slump, curing method, curing temperature, and the presence of fibers or mineral admixtures.

Several of these variables are particularly important in pavement applications. Parent concrete strength and recycled aggregate water absorption strongly influence recycled aggregate quality and strength variability, which directly affects slab stiffness and load-carrying capacity (Ismail et al., 2022; Padmini et al., 2009). Workability related parameters such as slump and admixture dosage indirectly affect compaction quality and air void structure, which are critical for pavement durability and long-term performance (Brandes et al., 2018). Curing method and temperature were included to reflect field curing conditions commonly observed in pavement construction, especially in warm climates (Chakradhara Rao et al., 2007; Patil, 2025).

The target variable for all models was the 28-day compressive strength, which remains the primary mechanical input for rigid pavement thickness design, fatigue assessment, and quality control.

Table 3. Statistical range of input variables used for model development

Input Parameter	Min	Max	Mean	Std. Dev.	Variance
Cement (kg/m³)	300.2	469.1	386.28	44.31	1963.56
Water-cement ratio (w/c)	0.35	0.65	0.49	0.08	0.0063
Natural aggregate content (%)	0	100	52.12	28.45	809.61
Recycled concrete aggregate (RCA) (%)	0	100	47.88	28.37	804.95
Fine aggregate (kg/m³)	550	819.6	696.9	66.59	4434.75
Parent concrete strength (MPa)	18.3	54.7	36.15	9.37	87.72
RCA water absorption (%)	1.24	11.99	6.64	2.79	7.76
RCA maximum size (mm)	10	20	15.29	4.02	16.13
Superplasticizer (%)	0	1.79	0.9	0.47	0.22
Slump (mm)	35	169	98.91	34.44	1186.36
Curing temperature (°C)	20	39.9	29.91	5.13	26.31
Fiber / mineral admixture (%)	0	2	0.49	0.55	0.31
Data source indicator	0	1	0.21	0.41	0.17

Note: *w/c = water-cement ratio; NA = natural coarse aggregate; RCA = recycled concrete aggregate. The reported minimum, maximum, mean, standard deviation, and variance were calculated from the combined hybrid dataset comprising both experimentally generated data and values collected from published literature. The data source indicator distinguishes between experimental and literature-based records. All parameters listed in the table were considered as input variables during model development.*

The statistical characteristics of the input variables employed for model development are summarized in Table 3. The table presents the range and distribution of key mix design parameters, recycled aggregate properties, and curing-related variables extracted from the combined experimental and literature dataset. The wide variation observed in cement content, water-cement ratio, aggregate composition, and recycled aggregate characteristics ensures adequate coverage of practical concrete mix designs. These statistical descriptors provide insight into the diversity of the dataset and support the robustness of the subsequent modeling and analysis.

4.3 Feature engineering for pavement-specific behavior.

To improve the physical interpretability and predictive capability of the models, additional derived features were introduced to represent pavement relevant material behavior. An effective water-cement ratio was computed by incorporating the influence of recycled aggregate water absorption, reflecting the additional water demand associated with porous recycled aggregates. This adjustment is particularly important for pavement concrete, where moisture variability can significantly influence early age strength development (Ismail et al., 2022; Tam et al., 2005).

A binder intensity parameter was introduced to represent the combined effect of cement content and water-cement ratio on strength development, which is consistent with strength-based pavement mix design concepts. In addition, a recycled aggregate severity index was formulated by combining recycled aggregate replacement ratio and water absorption to capture the combined effect of quantity and quality of recycled aggregates on compressive strength. Similar combined indicators have been recommended in previous recycled aggregate concrete studies to better represent material heterogeneity (Ojha et al., 2024; Padmini et al., 2009).

4.4 Data preprocessing.

Prior to model development, the hybrid dataset was subjected to preprocessing to ensure consistency and numerical stability. Missing values were handled using median imputation to preserve dataset integrity while minimizing distortion caused by outliers. Continuous variables were standardized to eliminate scale effects, which is essential for regularized linear models and kernel-based algorithms.

Categorical variables, such as curing method, were encoded numerically to enable their inclusion in machine learning models. The preprocessing strategy was designed to balance data normalization with preservation of physical meaning, an important consideration for engineering applications where interpretability is critical (Jagadesh et al., 2024).

4.5 Machine learning models and validation strategy.

The prediction of 28-day compressive strength of recycled aggregate concrete was formulated as a supervised regression problem, where the objective is to establish a functional relationship between multiple mix design parameters and a continuous output variable. Given the heterogeneous nature of recycled aggregate concrete and its relevance to rigid pavement design, a diverse set of machine learning models representing different modeling philosophies was employed.

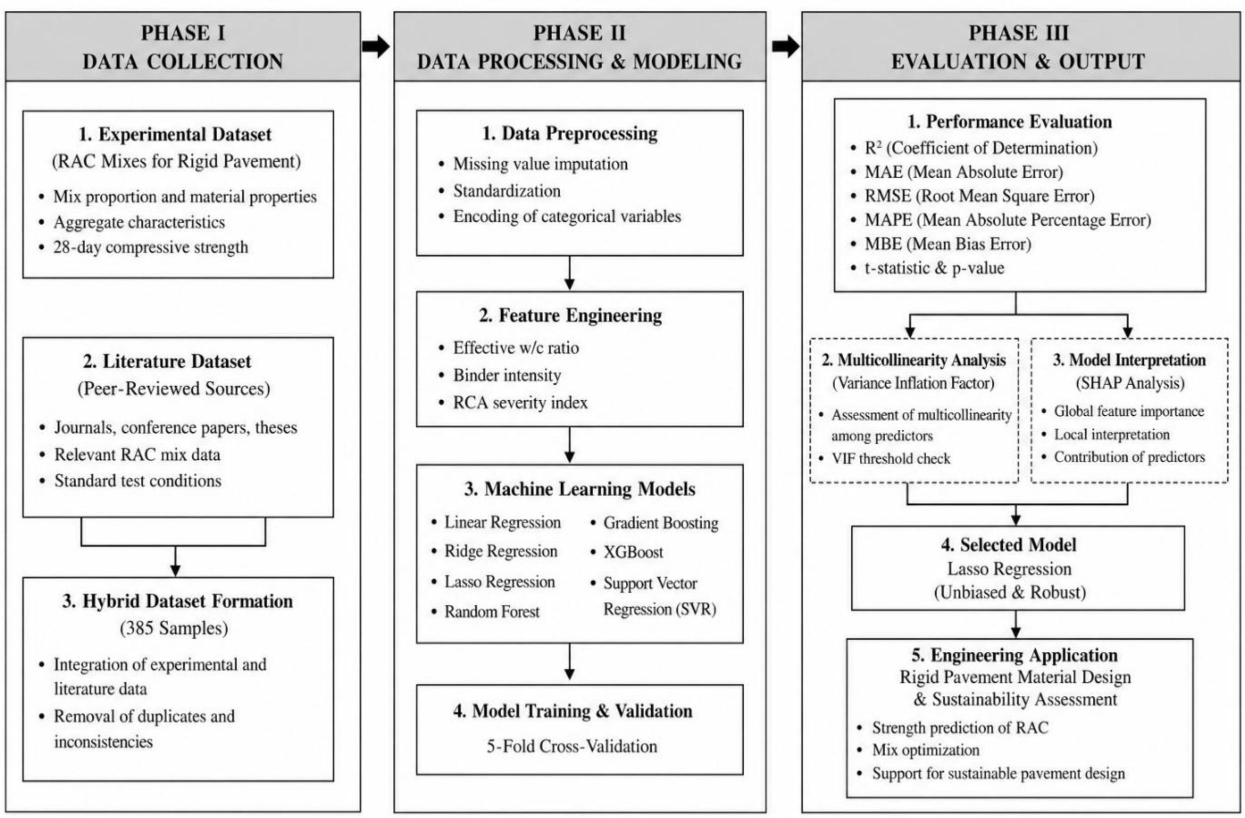

Figure 4: Machine learning workflow adopted for predicting the 28-day compressive strength of recycled aggregate concrete for rigid pavement applications.

Figure 4 illustrates the machine learning workflow adopted in this study for predicting the 28-day compressive strength of recycled aggregate concrete used in rigid pavement applications. Experimental and literature based data were combined to form a hybrid dataset, which was subjected to preprocessing and pavement-relevant feature engineering before model development. Multiple regression, ensemble, and kernel-based models were trained and evaluated using five-fold cross validation, followed by statistical performance, bias, and multicollinearity assessment. The workflow concludes with model interpretation and selection of the most reliable prediction model for engineering application.
The machine learning algorithms considered in this study were selected based on their established applicability in civil engineering material modeling and their ability to provide reliable predictions using datasets of moderate size. The selected models represent three major categories of supervised learning techniques, namely regularized regression models, kernel-based learning methods, and ensemble learning approaches. Artificial neural network models were not included in the present investigation because the primary objective was to compare interpretable machine learning techniques capable of providing both predictive accuracy and engineering insight. This approach also facilitates model transparency and practical implementation in recycled aggregate concrete applications, where understanding the influence of individual variables is often as important as prediction performance. Five-fold cross-validation was adopted to provide a robust assessment of model performance while maximizing the use of the available hybrid dataset. This validation strategy reduces the influence of a single train-test partition and enables performance evaluation across multiple data subsets. Although external validation using an independent dataset would provide additional evidence of model transferability, such datasets with consistent pavement-oriented recycled aggregate concrete variables were not readily available. Therefore, cross-validation was considered the most appropriate validation framework for the present investigation.

4.5.1 Model training and hyperparameter optimization.

Prior to model development, the hybrid dataset was subjected to preprocessing and quality verification to ensure consistency among the experimental and literature-derived records. Machine learning models were subsequently trained and evaluated using a five-fold cross-validation framework, whereby the dataset was divided into five subsets and each subset was used once as a validation set while the remaining subsets were used for model training. The final performance metrics reported in this study represent the average performance obtained across all validation folds.
To improve model reliability and predictive performance, hyperparameter tuning was performed for all machine learning algorithms before final model evaluation. For Ridge Regression and Lasso Regression, the regularization strength was adjusted to obtain an appropriate balance between model complexity and prediction accuracy. For Support Vector Regression, kernel-related parameters associated with the radial basis function kernel were optimized to improve nonlinear prediction capability. Ensemble learning models, including Random Forest, Gradient Boosting, and XGBoost, were tuned by adjusting model complexity and tree-learning parameters to achieve improved generalization performance.
The final configuration of each model was selected based on cross-validation performance to ensure a fair comparison among algorithms. This procedure reduced the likelihood of overfitting and enabled a more robust assessment of the suitability of different machine learning approaches for predicting the compressive strength of recycled aggregate concrete intended for rigid pavement applications.

4.5.2 Linear and regularized regression models.

Linear Regression was adopted as a baseline model to evaluate the linear relationship between input variables and compressive strength. The general form of the linear regression model is expressed as:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i \quad (1)$$

where y_i is the predicted compressive strength, x_{ij} are the input variables, β_j are regression coefficients, and ε_i represents random error. While linear regression provides transparency and interpretability, its performance may degrade in the presence of multicollinearity and nonlinear interactions, which are common in recycled aggregate concrete datasets.

To address these limitations, Ridge Regression and Lasso Regression were employed. Ridge Regression introduces an L2 regularization term that penalizes large coefficients and reduces model sensitivity to multicollinearity:

$$\min\left(\sum(y_i - y_i^0)^2 + \lambda\sum\beta_j^2\right) \quad (2)$$

Lasso Regression applies L1 regularization, which can shrink some coefficients to zero and effectively perform variable selection:

$$\min\left(\sum(y_i - y_i^0)^2 + \lambda\sum \mid \beta_j \mid\right) \quad (3)$$

Regularized regression models are particularly suitable for hybrid datasets that combine experimental and literature data, as they improve stability and generalization in the presence of correlated input variables and measurement variability (Asteris et al., 2021; Kumar et al., 2024). In rigid pavement applications, where conservative and reliable material predictions are preferred, such stability is essential.

4.5.3 Support vector regression.

Support Vector Regression (SVR) was employed to capture nonlinear relationships between mix parameters and compressive strength. SVR constructs a regression function by minimizing prediction error within an ε-insensitive tube, while controlling model complexity:

$$\min\left(\frac{1}{2} \mid\mid w \mid\mid^2 + C\sum(\xi_i + \xi_i^*)\right) \quad (4)$$

SVR is effective for modeling nonlinear material behavior and has been widely used in concrete strength prediction studies. However, its performance is sensitive to kernel selection and parameter choice, and it may struggle when dataset variability is high, as is often the case with hybrid pavement concrete datasets (Jagadesh et al., 2024; Nguyen et al., 2021). A radial basis function (RBF) kernel was adopted due to its widespread use in concrete strength prediction.

4.5.4 Ensemble learning models.

Ensemble learning models were employed to capture complex nonlinear interactions among mix design parameters. Random Forest is a bagging-based ensemble method that constructs multiple decision trees using bootstrap samples and averages their predictions:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t\,(x) \qquad (5)$$

where $f_t(x)$ represents the prediction of the t-th decision tree. Random Forest models are robust to noise and capable of handling high-dimensional data, making them suitable for concrete strength prediction.

Gradient Boosting and Extreme Gradient Boosting (XGBoost) are boosting-based ensemble methods that sequentially build models to minimize prediction error. These methods are particularly effective in capturing nonlinear dependencies and interaction effects among variables:

$$y_i^0 = \sum_{k=1}^{K} f_k\,(x_i) \qquad (6)$$

Ensemble methods have demonstrated strong predictive performance in concrete engineering studies; however, they are also prone to overfitting when dataset size is limited or variability is high (Ojha et al., 2024; Zhang et al., 2020). Their inclusion in this study allows a comprehensive comparison between simple and complex modeling approaches for rigid pavement material prediction.

4.5.5 Validation strategy.

To ensure robust evaluation of model generalization capability, a five-fold cross validation strategy was adopted. The dataset was randomly divided into five equal subsets; in each iteration, four subsets were used for training and one subset for validation. This process was repeated until each subset served as the validation set once, and the prediction results were aggregated.

Cross-validation is particularly important for hybrid datasets, where data originate from multiple experimental sources and literature studies. It minimizes bias associated with a single train test split and provides a reliable estimate of model performance under unseen conditions, which is critical for material level predictions used in rigid pavement design (Kumar et al., 2024; Patil, 2025).

4.6 Performance evaluation criteria.

The predictive performance of the machine learning models was evaluated using multiple statistical indicators to assess both accuracy and reliability. These indicators provide complementary perspectives on model behavior and are particularly relevant for engineering decision making in rigid pavement design.

The coefficient of determination (R^2) was used to measure the goodness of fit between predicted and observed compressive strength values:

$$R^2 = 1 - \frac{\sum(y_i - y_i^0)^2}{\sum(y_i - \bar{y})^2} \qquad (7)$$

Mean Squared Error (MSE) and Root Mean Square Error (RMSE) quantify the spread of prediction errors and penalize large deviations:

$$MSE = \frac{1}{N}\sum(y_i - y_i^0)^2 \qquad (8)$$

$$RMSE = \sqrt{\frac{1}{N}\sum(y_i - y_i^0)^2} \qquad (9)$$

Mean Absolute Error (MAE) measures the average magnitude of prediction errors without emphasizing extreme values:

$$MAE = \frac{1}{N}\sum \mid y_i - y_i^0 \mid \quad (10)$$

Mean Absolute Percentage Error (MAPE) provides a relative measure of prediction accuracy expressed as a percentage, facilitating comparison across different strength ranges:

$$MAPE = \frac{1}{N}\sum \mid \frac{y_i - y_i^0}{y_i} \mid \times 100 \quad (11)$$

To assess systematic prediction bias, Mean Bias Error (MBE) was calculated:

$$MBE = \frac{1}{N}\sum (y_i - y_i^0) \quad (12)$$

A positive MBE indicates overestimation, while a negative value indicates underestimation. Bias evaluation is particularly important for rigid pavement applications, where systematic over or under prediction of compressive strength can directly affect slab thickness design and safety margins (Brandes et al., 2018; Patil, 2025).
Finally, paired t-statistics were computed to evaluate the statistical significance of differences between predicted and observed values:

$$t = \sqrt{\frac{(N-1)\cdot MBE^2}{RMSE^2 - MBE^2}} \quad (13)$$

The corresponding p-values were used to assess whether prediction errors were statistically significant. Inclusion of hypothesis-testing metrics ensures that selected models are not only accurate but also statistically consistent and suitable for engineering applications involving rigid pavement materials.

5. RESULTS AND DISCUSSION

5.1 Overall model performance comparison.

The predictive performance of the developed machine learning models was evaluated using multiple statistical indicators, including accuracy and bias related metrics, to ensure their suitability for rigid pavement material assessment. Table 4 summarizes the performance of all models considered in this study.

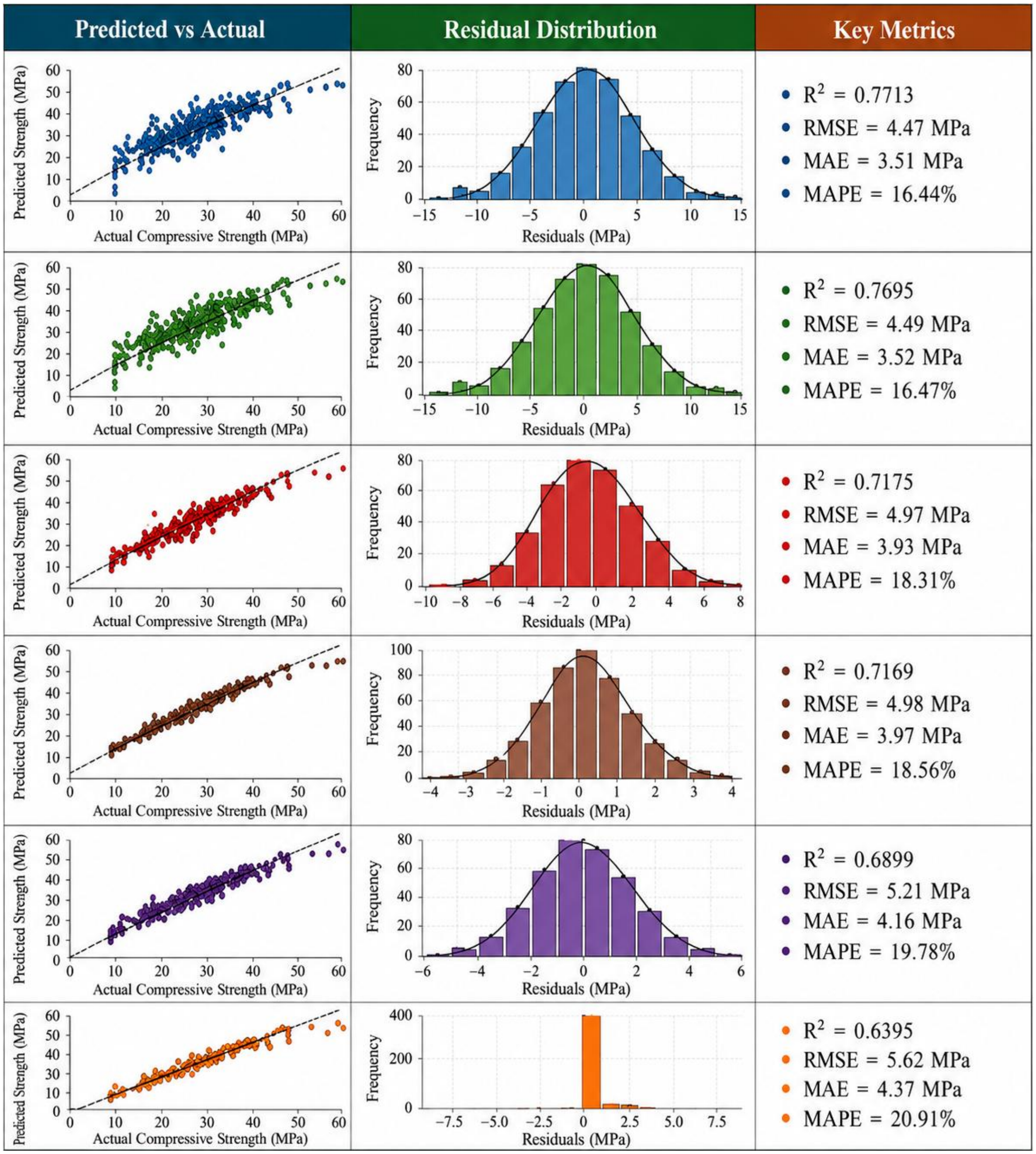

Figure 5. Performance comparison of six machine learning models for predicting 28-day compressive strength of recycled aggregate concrete. For each model, actual versus predicted compressive strength and the corresponding residual distribution are presented.

Figure 5 summarizes the predictive performance of six machine learning models developed for estimating the 28-day compressive strength of recycled aggregate concrete. The actual predicted plots show the level of agreement between experimental and model estimated values, while the residual distributions provide insight into the spread and consistency of prediction errors. The predictive performance reported in this section should not be directly compared with the correlation analysis presented in Figure 1. While Figure 1 illustrates the overall association between compressive strength and the combined input parameters within the experimental dataset, the machine learning results are derived from cross-validation and therefore represent model performance under independent validation conditions. The difference between these values reflects the distinction between descriptive correlation and predictive generalization, which is commonly observed in machine learning applications involving heterogeneous engineering datasets. The reported performance indicators correspond to the optimized versions of each machine learning model obtained through the hyperparameter tuning and cross-validation procedure described in Section 4.5.1, thereby providing a consistent basis for comparative evaluation.

Among the evaluated models, Lasso Regression demonstrated the highest predictive accuracy, achieving an R^2 value of 0.7713, along with the lowest MAE (3.51 MPa), RMSE (4.47 MPa), and MAPE (16.44%). Ridge Regression followed closely, with an R^2 value of 0.7695 and similar error levels. In contrast, ensemble learning models such as Gradient Boosting, XGBoost, and Random Forest exhibited comparatively lower R^2 values and higher prediction errors, while Support Vector Regression showed the weakest performance among all models. These results indicate that simpler, regularized regression models outperform more complex nonlinear models when applied to the present hybrid dataset (Zhuang et al., 2021). This outcome is consistent with recent findings that hybrid datasets combining experimental and literature data often exhibit high variability and multicollinearity, conditions under which regularized linear models tend to provide more stable and reliable predictions (Asteris et al., 2021; Kumar et al., 2024).

The superior performance of Lasso Regression can be attributed to the characteristics of the hybrid dataset used in this study. The dataset combines experimental and literature-derived observations and therefore contains considerable variability, correlated predictors, and potential measurement inconsistencies. Under such conditions, Lasso Regression benefits from L1 regularization, which suppresses the influence of redundant variables and reduces model complexity, thereby improving generalization performance. Ridge Regression also performed well because its regularization mechanism stabilizes coefficient estimation under multicollinearity; however, unlike Lasso Regression, it does not eliminate less informative predictors, which may explain its slightly lower performance.

In contrast, ensemble learning models such as Random Forest, Gradient Boosting, and XGBoost are designed to capture complex nonlinear interactions. While these methods often perform exceptionally well on large and highly structured datasets, their advantage may diminish when the available data exhibit substantial variability and moderate sample size. Under such conditions, the models may partially learn dataset-specific patterns that do not consistently generalize across validation folds. Similarly, Support Vector Regression demonstrated lower predictive capability, which may be attributed to the difficulty of representing heterogeneous recycled aggregate concrete behavior using a single kernel-based decision function. These observations indicate that the dominant relationships governing compressive strength in the present dataset are sufficiently captured through regularized linear modeling approaches.

Table 4. Performance comparison of machine learning and regression models for predicting 28-day compressive strength.

Model	R^2	MAE	MSE	RMSE	MAPE	MBE	t-statistic	p-value
Lasso Regression	0.7713	3.51	20.01	4.47	16.44	−0.03	0.139	0.89
Ridge Regression	0.7695	3.52	20.16	4.49	16.47	−0.04	0.163	0.871
Gradient Boosting	0.7175	3.93	24.71	4.97	18.31	0.2	−0.774	0.439
XGBoost	0.7169	3.97	24.77	4.98	18.56	0.27	−1.048	0.295
Random	0.6899	4.16	27.13	5.21	19.78	0.2	−0.751	0.453

Forest								
Support Vector Regression	0.6395	4.37	31.54	5.62	20.91	0.06	−0.200	0.842

Note: *R^2 = coefficient of determination; MAE = mean absolute error (MPa); MSE = mean squared error (MPa²); RMSE = root mean squared error (MPa); MAPE = mean absolute percentage error (%); MBE = mean bias error (MPa). The t-statistic and corresponding p-value were used to assess the statistical significance of model prediction bias. All performance indicators were evaluated using an identical data partitioning and validation strategy to ensure a fair comparison among the models.*

5.2 Interpretation from a rigid pavement perspective.

In rigid pavement design, compressive strength is a critical input parameter influencing slab thickness, load transfer efficiency, and fatigue performance. From an engineering standpoint, prediction reliability and bias control are as important as achieving high numerical accuracy. The superior performance of Lasso and Ridge Regression models suggests that the relationship between compressive strength and mix design parameters in recycled aggregate concrete for pavements is governed by dominant linear trends, despite the presence of some nonlinear interactions.

Recycled aggregate concrete used in pavements typically exhibits greater material heterogeneity due to variations in parent concrete strength, recycled aggregate water absorption, and curing conditions. In such cases, complex ensemble models may overfit localized patterns present in the training data, particularly when the dataset includes information from multiple experimental sources (Ojha et al., 2024; Patil, 2025). Regularized regression models mitigate this issue by constraining coefficient magnitudes, thereby improving generalization performance under variable conditions.

The relatively lower performance of Random Forest and boosting-based models observed in this study aligns with similar findings reported for recycled aggregate concrete, where ensemble models did not consistently outperform simpler approaches when data variability was high and sample size was moderate (Jagadesh et al., 2024; Zhang et al., 2020).

5.3 Error metrics and engineering significance.

From a pavement engineering perspective, absolute error measures such as MAE and RMSE are particularly important, as they reflect the expected deviation between predicted and actual compressive strength values. The MAE values obtained for Lasso and Ridge Regression models (approximately 3.5 MPa) fall within an acceptable range for preliminary pavement material design and quality control applications.

The MAPE values for the best-performing models remained below 17%, indicating reasonable prediction consistency across a wide strength range. This level of accuracy is comparable to, or better than, that reported in previous machine learning studies on recycled aggregate concrete strength prediction (Jagadesh et al., 2024; Nguyen et al., 2021). Support Vector Regression exhibited the highest MAE, RMSE, and MAPE values, suggesting limited robustness for hybrid pavement datasets.

Higher RMSE values observed for ensemble and kernel-based models indicate greater sensitivity to extreme prediction errors. In rigid pavement design, such variability may lead to unsafe or overly conservative slab thickness estimates, particularly when compressive strength predictions are used directly in mechanistic empirical design frameworks (Brandes et al., 2018).

5.4 Bias analysis and statistical consistency.

Bias evaluation plays a crucial role in assessing the suitability of predictive models for engineering applications. Mean Bias Error (MBE) values for Lasso and Ridge Regression models were close to zero (-0.0317 MPa and -0.0373 MPa, respectively), indicating negligible systematic over or under prediction. Furthermore, the associated t-statistics were low, and the p-values were significantly greater than 0.05, suggesting that the differences between predicted and observed values are statistically insignificant.

In contrast, ensemble models such as Gradient Boosting, XGBoost, and Random Forest exhibited higher positive MBE values, indicating a tendency toward overestimation of compressive strength. Although these biases were not statistically significant at the 95% confidence level, their magnitude is undesirable from a pavement design standpoint, as systematic overestimation can lead to unsafe structural designs and reduced service life (Patil, 2025).

Support Vector Regression showed a relatively small bias; however, its higher dispersion of errors, as indicated by RMSE and MAPE, reduces its practical applicability for rigid pavement material prediction. These findings reinforce the importance of combining bias-related metrics with conventional accuracy measures when evaluating machine learning models for pavement engineering applications (Brandes et al., 2018).

5.5 Comparison with previous studies.

The results obtained in this study are consistent with recent literature reporting that regularized regression models often provide competitive or superior performance compared to complex machine learning algorithms when predicting concrete strength using heterogeneous datasets (Asteris et al., 2021; Kumar et al., 2024). Studies focusing on recycled aggregate concrete have also highlighted the challenges associated with modeling strength variability arising from recycled aggregate quality and parent concrete characteristics (Ojha et al., 2024; Padmini et al., 2009).

While ensemble models have demonstrated strong performance in some concrete applications, their effectiveness appears to be context-dependent and influenced by dataset size, variability, and source consistency (Zhang et al., 2020). For rigid pavement applications, where prediction stability and bias control are critical, the present findings suggest that simpler, well-regularized models may offer a more reliable alternative. The predictive performance achieved in the present study should be interpreted within the context of cross-validation-based assessment. While cross-validation is widely accepted for evaluating machine learning models in engineering applications, external validation using completely independent datasets remains an important step for confirming model transferability under different material sources, construction practices, and environmental conditions. Consequently, the reported performance metrics should be regarded as indicators of internal model generalization rather than universal predictive capability.

Similar differences between correlation-based analyses and cross-validated predictive performance have been reported in previous machine learning studies involving recycled aggregate concrete and other construction materials. Correlation analyses are useful for identifying dominant trends and variable relationships, whereas cross-validation provides a more conservative and realistic estimate of model performance when applied to new datasets. Therefore, the predictive results presented in this study should be interpreted primarily on the basis of cross-validation metrics rather than simple correlation measures.

5.6 Implications for rigid pavement design and practice.

The findings of this study have practical implications for the use of recycled aggregate concrete in rigid pavement construction. The demonstrated reliability of Lasso and Ridge Regression models indicates that machine learning can be effectively used as a supporting tool for preliminary material assessment, mix optimization, and quality control when experimental testing resources

are limited.
Importantly, the ability of regularized models to deliver unbiased and statistically consistent predictions enhances confidence in their application for pavement thickness design and performance evaluation. By integrating experimentally grounded data with literature information, the proposed hybrid modeling framework provides a realistic representation of field conditions, supporting the sustainable use of recycled aggregate concrete in rigid pavement systems.
For practical implementation, the proposed framework may be used during preliminary mix design, material screening, and quality assessment stages of rigid pavement projects. Engineers can utilize the developed machine learning model to obtain rapid estimates of compressive strength based on readily available mix design parameters, thereby reducing dependence on extensive trial batching during early project planning. The framework is particularly useful when evaluating multiple recycled aggregate sources, alternative replacement levels, or different binder compositions.
The results further indicate that special attention should be given to binder intensity, cement content, parent concrete strength, and superplasticizer dosage, as these variables exert the greatest influence on compressive strength development. Consequently, optimization of these parameters should be prioritized during pavement concrete mix design. In contrast, recycled aggregate replacement ratio alone should not be considered the primary indicator of expected strength performance, as aggregate quality and source characteristics play a more significant role.
For engineering applications, the developed models should be viewed as decision-support tools rather than replacements for standard laboratory testing. Final pavement mix approval should continue to rely on experimental verification in accordance with applicable design specifications and quality control procedures. Nevertheless, the proposed framework can substantially reduce the number of trial mixtures required during material selection and preliminary pavement design.

5.7 Identification of the most suitable model.
The selection of the most suitable predictive model for rigid pavement applications must be based not only on numerical accuracy but also on robustness under multicollinearity, bias behavior, and engineering reliability. In rigid pavement design practice, conservative and statistically consistent predictions of compressive strength are essential, as this parameter directly governs slab thickness determination, fatigue performance, and long-term structural safety. Accordingly, model selection in this study was carried out using an integrated evaluation framework combining predictive accuracy metrics, bias analysis, multicollinearity assessment through Variance Inflation Factor (VIF), and feature contribution analysis using SHapley Additive exPlanations (SHAP).
Based on the comparative evaluation presented in Table 4, Lasso Regression emerged as the most suitable model for predicting the 28-day compressive strength of recycled aggregate concrete for rigid pavement applications. Lasso Regression achieved the highest coefficient of determination ($R^2 = 0.7713$), indicating the strongest agreement between predicted and measured compressive strength values among all evaluated models. It also recorded the lowest MAE (3.51 MPa), RMSE (4.47 MPa), and MAPE (16.44%), demonstrating superior accuracy and consistency across the full strength range. From a pavement engineering standpoint, bias behavior is particularly critical. The Mean Bias Error (MBE) obtained for the Lasso model was -0.0317 MPa, which is effectively negligible, indicating the absence of systematic over or under estimation. The associated t-statistic was very small, and the p-value (0.8896) was substantially higher than the 0.05 significance threshold, confirming that the prediction differences are statistically insignificant and reinforcing the reliability of the Lasso model for engineering decision making (Brandes et al., 2018; Hall & Hall, 2017; Patil, 2025).
The robustness of Lasso Regression is further justified by the VIF analysis. Extremely high VIF values were observed for several physically interrelated parameters, including water-cement ratio

(VIF ≈ 7836), effective water-cement ratio (VIF ≈ 7078), cement content (VIF ≈ 924), and binder intensity (VIF ≈ 847), indicating severe multicollinearity within the hybrid dataset. Such multicollinearity is inherent in concrete mix design variables and is particularly pronounced in hybrid datasets combining experimental and literature data. Under these conditions, conventional regression and complex nonlinear models are prone to instability. Lasso Regression effectively mitigates this issue through L1 regularization, which penalizes redundant predictors and stabilizes coefficient estimation, thereby enhancing generalization capability under highly correlated input conditions (Asteris et al., 2021; Kumar et al., 2024). The presence of severe multicollinearity in the hybrid dataset is not unexpected because many concrete mix design parameters are physically interconnected. For example, binder intensity, cement content, water-cement ratio, and effective water-cement ratio are derived from related mixture design principles and collectively influence strength development. Although these variables exhibit strong statistical correlation, they represent different engineering characteristics governing recycled aggregate concrete performance. Therefore, the objective of this study was not to eliminate correlated variables solely on statistical grounds but to preserve physically meaningful information relevant to pavement concrete behavior. The adoption of regularized machine learning models enabled the retention of these variables while minimizing the adverse effects of multicollinearity on prediction stability.

Further physical validation of model suitability was obtained through SHAP analysis, which revealed that binder intensity (mean |SHAP| ≈ 2.61) and cement content (≈ 2.34) are the most influential parameters governing compressive strength prediction, followed by parent concrete strength (≈ 2.05) and superplasticizer dosage (≈ 1.95). These findings are fully consistent with rigid pavement concrete behavior, where paste quality, recycled aggregate origin, and compaction efficiency dominate load carrying capacity and durability. The recycled aggregate severity index also exhibited a meaningful contribution (≈ 1.25), confirming that the combined effect of recycled aggregate quantity and quality is more influential than recycled aggregate replacement ratio alone, which showed a comparatively low SHAP contribution (≈ 0.13). Lower ranked variables such as curing method, aggregate size, and fiber or mineral admixture content exhibited secondary influence within the studied range, indicating that their effects are largely indirect or embedded within dominant mix parameters (Jagadesh et al., 2024; Ojha et al., 2024). The dominance of binder intensity and cement content observed in the SHAP analysis is consistent with the fundamental mechanisms governing strength development in concrete. Binder intensity reflects the amount of cementitious material available to generate hydration products and establish a dense microstructure, while cement content directly influences the formation of calcium silicate hydrate gel, which is primarily responsible for strength gain. Higher values of these parameters generally contribute to improved paste quality, reduced porosity, and enhanced load carrying capacity. Their strong contribution to model predictions therefore reflects well-established principles of concrete technology and rigid pavement material design.

Parent concrete strength emerged as another influential variable because recycled aggregates inherit physical characteristics from the original concrete from which they are produced. Recycled aggregates derived from stronger parent concrete generally possess lower levels of microcracking, reduced mortar deterioration, and improved aggregate integrity. Consequently, the quality of the parent concrete significantly affects the mechanical behavior of recycled aggregate concrete. The strong SHAP contribution associated with this variable confirms that recycled aggregate quality cannot be adequately represented by replacement ratio alone and must also account for the properties of the source material. The importance of superplasticizer dosage can be attributed to its influence on workability, particle dispersion, and compaction efficiency. Adequate dispersion of cement particles promotes more uniform hydration and allows lower effective water demand without compromising workability. This effect is particularly beneficial in recycled aggregate concrete, where the higher water absorption capacity of recycled aggregates can adversely influence fresh concrete properties. The SHAP results therefore indicate that chemical admixtures play a meaningful role in controlling both fresh and hardened concrete performance within the investigated range of mixture proportions.

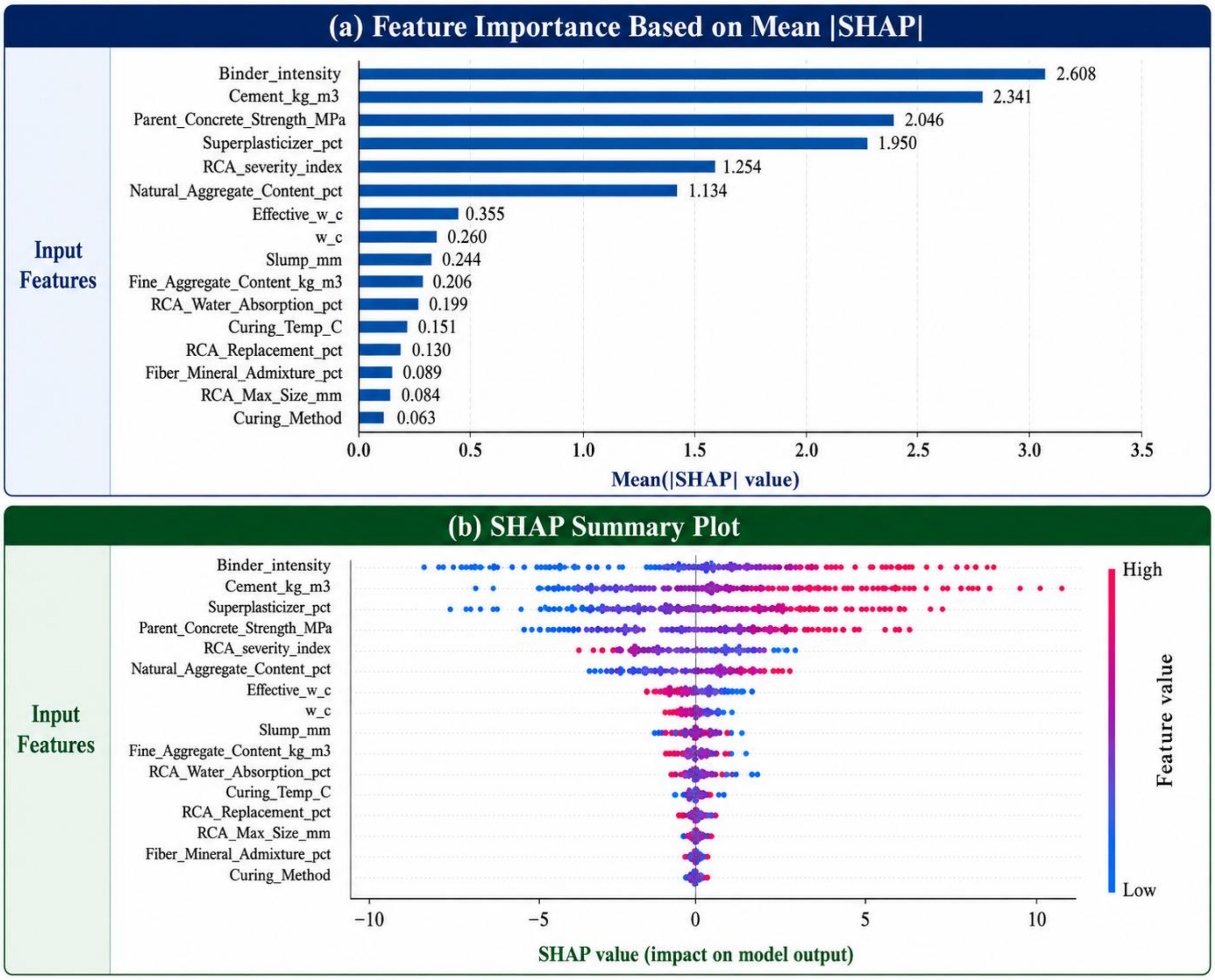

Figure 6. SHAP-based interpretation of the Lasso regression model for predicting 28-day compressive strength of recycled aggregate concrete: (a) mean absolute SHAP values indicating the relative contribution of each input parameter, and (b) SHAP summary-style plot illustrating the comparative influence of variables on model predictions.

Figure 6 illustrates the contribution of individual input variables to the prediction of 28-day compressive strength using the Lasso regression model, as interpreted through SHAP analysis. The mean absolute SHAP values presented in Figure 6(a) quantify the overall importance of each parameter in influencing model output, while Figure 6(b) provides a comparative visualization of feature effects across the dataset.

Although Ridge Regression demonstrated performance close to that of Lasso Regression, with similar R^2 and error values, its slightly higher prediction errors and marginally greater bias indicate reduced robustness under severe multicollinearity, as Ridge Regression performs coefficient shrinkage without eliminating redundant variables. Ensemble learning models, including Gradient Boosting, XGBoost, and Random Forest, exhibited lower predictive accuracy and a tendency toward positive bias, indicating systematic over estimation of compressive strength. Such behavior is undesirable in rigid pavement applications, as it may lead to unconservative slab thickness design and reduced safety margins, particularly under heavy traffic loading (Ojha et al., 2024;

Patil, 2025). Support Vector Regression showed the weakest overall performance, characterized by higher error dispersion and limited generalization capability for the present hybrid dataset.
The integrated evidence from predictive accuracy metrics, bias analysis, VIF assessment, and SHAP-based feature interpretation confirms that Lasso Regression offers the most reliable, stable, and physically meaningful framework for compressive strength prediction of recycled aggregate concrete in rigid pavement applications. Within the scope of this study, Lasso Regression is therefore identified as the most suitable model for supporting pavement material assessment and sustainable pavement design. More broadly, the study demonstrates that the value of machine learning for RAC strength prediction depends not only on algorithm selection but also on the quality and diversity of the underlying dataset. The hybrid experimental–literature dataset developed in this study provides the principal methodological contribution by combining experimentally grounded observations with a broader range of reported RAC conditions, while the comparative modeling analysis identifies Lasso Regression as the most suitable approach within the scope of the available data.

6. CONCLUSIONS AND PRACTICAL IMPLICATIONS

This study developed and systematically evaluated a hybrid experimental–literature dataset for predicting the 28-day compressive strength of recycled aggregate concrete (RAC) for rigid pavement applications. The dataset combines laboratory-based observations with carefully selected literature data to represent a wider range of RAC mixture conditions, recycled aggregate characteristics, and curing environments. The machine learning analysis was then used to evaluate how different modeling approaches perform within this heterogeneous dataset. The results demonstrate that the principal contribution of the study lies in establishing a broader and pavement-oriented data foundation for RAC strength prediction, while the modeling analysis provides evidence on the relative suitability of different prediction approaches.
The experimental investigation confirmed that the compressive strength of RAC used in rigid pavements is governed by a combined influence of binder characteristics, recycled aggregate quality, parent concrete strength, and curing conditions. The measured strength values exhibited noticeable variability, particularly at higher recycled aggregate replacement levels, highlighting the limitations of conventional empirical relationships when RAC is used in pavement concrete. This observation justified the development of a data driven prediction framework tailored to pavement relevant mix proportions.
The hybrid dataset successfully integrated experimental data with literature reported results, enabling the machine learning models to learn from a broader range of recycled aggregate qualities and mix designs. This approach ensured improved generalization and addressed the challenge of limited experimental data, which is common in pavement material studies. The comparative evaluation of machine learning models demonstrated that regularized linear regression techniques are more effective than ensemble and kernel-based models when applied to such hybrid datasets.
Among all evaluated models, Lasso Regression achieved the best predictive performance, with a coefficient of determination of 0.7713, a mean absolute error of 3.51 MPa, and a root mean square error of 4.47 MPa. These values indicate reliable strength prediction accuracy for rigid pavement material assessment. In contrast, ensemble learning models such as Gradient Boosting, XGBoost, and Random Forest yielded lower R^2 values in the range of 0.69-0.72 and higher RMSE values between 4.97 and 5.21 MPa, while Support Vector Regression showed the weakest performance with an RMSE of 5.62 MPa and a MAPE exceeding 20%. Bias analysis further reinforced the suitability of Lasso Regression, which exhibited a negligible mean bias error of -0.0317 MPa and a statistically insignificant difference between predicted and observed strengths, as indicated by a p-value of 0.8896. This unbiased and statistically consistent behavior is particularly important for

rigid pavement design, where conservative estimates of compressive strength are essential to ensure adequate slab thickness and fatigue performance.

The robustness of the selected model was further validated through Variance Inflation Factor analysis, which revealed severe multicollinearity among several physically interrelated mix design variables. Extremely high VIF values were observed for water-cement ratio and effective water-cement ratio (greater than 7000), as well as for cement content and binder intensity (greater than 800), reflecting their inherent physical dependency. Despite this, Lasso Regression remained stable due to its L1 regularization capability, confirming its suitability for pavement datasets where correlated variables are unavoidable. SHapley additive explanations (SHAP) analysis provided additional engineering insight into the mechanisms governing compressive strength development in recycled aggregate concrete. Binder intensity emerged as the most influential variable, highlighting the importance of cementitious material availability in hydration and microstructure formation. Cement content also exhibited a strong contribution, confirming its direct role in generating hydration products responsible for strength development. Parent concrete strength was identified as a major influencing parameter because the quality of recycled aggregates is strongly dependent on the characteristics of the source concrete. Recycled aggregates originating from stronger parent concrete generally possess lower levels of microcracking and improved aggregate integrity, which contribute positively to strength performance. Superplasticizer dosage also showed a meaningful influence by improving workability, particle dispersion, and compaction efficiency, particularly in mixtures containing recycled aggregates with higher water absorption. Collectively, these findings indicate that compressive strength development in recycled aggregate concrete is governed not only by recycled aggregate replacement levels but also by binder quality, source aggregate characteristics, and mixture optimization strategies. The recycled aggregate severity index further demonstrated that aggregate quality exerts a greater influence on strength performance than replacement ratio alone. The observed multicollinearity should be interpreted as an inherent characteristic of concrete mixture datasets rather than a modeling deficiency. Many concrete design variables are mathematically related and cannot be treated as completely independent parameters. Consequently, the retention of correlated variables was considered necessary to preserve engineering relevance and maintain consistency with practical concrete mix design procedures. The successful performance of the regularized models demonstrates that reliable strength prediction can still be achieved despite the presence of strong inter-variable dependency.

From a rigid pavement engineering perspective, these findings indicate that compressive strength prediction for RAC is primarily governed by dominant linear relationships associated with binder and aggregate quality parameters. The superior accuracy, negligible bias, robustness under extreme multicollinearity, and physically interpretable behavior of Lasso Regression make it particularly suitable for material-level decision-making in pavement slab design, fatigue life estimation, and sustainability-driven material selection. The tendency of ensemble models to overestimate strength further highlights the importance of bias evaluation when machine learning approaches are applied to safety critical pavement applications.

Despite the encouraging results, several limitations should be acknowledged. The hybrid dataset contains variables that exhibit substantial multicollinearity due to the inherent relationships among concrete mixture parameters. While regularization techniques effectively controlled its influence in the present study, the transferability of the developed models should be further examined using independent datasets from different geographical regions and material sources. Future investigations may also explore advanced feature selection and dimensionality reduction approaches to evaluate whether comparable prediction accuracy can be achieved using a reduced set of engineering variables.

The proposed framework enables reliable compressive strength estimation of recycled aggregate concrete with an average prediction error of approximately 3.5 MPa, supporting preliminary rigid pavement material design. The framework supports mix optimization and quality assessment while reducing dependence on extensive laboratory testing during early design stages. Moreover, the transparent interpretation provided by SHAP analysis allows engineers to prioritize binder quality and recycled aggregate characteristics during pavement mix design, supporting sustainable construction without compromising structural reliability. It should also be noted that model validation was performed using a five-fold cross-validation framework rather than an entirely independent external dataset. Although this approach provides a rigorous assessment of internal predictive capability, additional validation using datasets generated from different geographical regions, aggregate sources, and construction environments would further strengthen confidence in the broader applicability of the proposed framework. Accordingly, the present findings should be interpreted as a reliable comparative assessment of machine learning algorithms within the scope of the available hybrid dataset.

From a practical perspective, the proposed machine learning framework can assist engineers in rapidly evaluating recycled aggregate concrete mixtures during preliminary pavement design and material selection. The ability to identify influential strength-governing variables provides additional support for optimizing mix proportions while maintaining sustainable construction objectives.

Future research should focus on extending the proposed framework to predict flexural strength, elastic modulus, and fatigue performance, which are directly used in mechanistic empirical rigid pavement design. Incorporating long term durability indicators such as shrinkage, abrasion resistance, and freeze thaw performance would further enhance lifecycle-based pavement assessment. Additional work is also needed to develop field calibrated datasets that include traffic loading, environmental exposure, and construction variability, as well as to explore physics-informed machine learning models that integrate pavement mechanics with data driven prediction. Such efforts will further strengthen the applicability of machine learning in sustainable rigid pavement engineering.

Although the developed framework demonstrated satisfactory predictive performance, validation was limited to five-fold cross-validation of the available hybrid dataset. Future studies should incorporate independent external datasets and larger sample populations to further assess model transferability and robustness.

7. REFERENCE

Alyami, M., Alyousef, R., Alabduljabbar, H., Alrshoudi, F. (2024). *Machine learning-based prediction of concrete compressive strength using hybrid datasets.* Construction and Building Materials, 390, 131768. https://doi.org/10.1016/j.conbuildmat.2023.131768

American Association of State Highway, & Officials, T. (1993). *AASHTO Guide for Design of Pavement Structures.* AASHTO.

Asteris, P. G., Apostolopoulou, M., Skentou, A. D., Moropoulou, A. (2021). *Application of machine learning techniques for concrete strength prediction.* Materials, 14(6), 1434. https://doi.org/10.3390/ma14061434

Asteris, P. G., Kolovos, K. G., Douvika, M. G., Roussis, P. C. (2019). *Prediction of self-compacting concrete strength using artificial neural networks.* European Journal of Environmental and Civil Engineering, 23(3), 329–347. https://doi.org/10.1080/19648189.2017.1320233

Asteris, P. G., Skentou, A. D., Bardhan, A., Samui, P., Pilakoutas, K. (2020). *Predicting concrete strength using machine learning.* Construction and Building Materials, 252, 119134. https://doi.org/10.1016/j.conbuildmat.2020.119134

Behera, M., Bhattacharyya, S. K., Minocha, A. K., Deoliya, R., Maiti, S. (2014). *Recycled aggregate from C&D waste and its use in concrete.* Resources, Conservation and Recycling, 68, 30–43. https://doi.org/10.1016/j.resconrec.2012.10.007

Brandes, M. R., Butterfield, J., Roesler, J. R. (2018). *Effect of recycled concrete aggregates on strength and stiffness gain of concrete.* PCI Journal, 63(3), 52–65. https://doi.org/10.15554/pcij.06012018.52.65

Butler, L., West, J. S., Tighe, S. L. (2016). *Effect of recycled concrete aggregate properties on the bond strength between concrete and steel reinforcement.* Cement and Concrete Research, 41(10), 1037–1049. https://doi.org/10.1016/j.cemconres.2011.06.004

Chakradhara Rao, M., Bhattacharyya, S. K., Barai, S. V. (2007). *Assessment of durability of recycled aggregate concrete produced by two-stage mixing approach.* Journal of Materials in Civil Engineering, 19(5), 379–386. https://doi.org/10.1061/(ASCE)0899-1561(2007)19:5(379)

Chauhan, B. L., Singh, G. J. (2023). *Sustainable development of recycled concrete aggregate through optimized acid–mechanical treatment: A simplified approach.* Construction and Building Materials, 399, 132559. https://doi.org/10.1016/j.conbuildmat.2023.132559

Chou, J. S., Pham, A. D. (2013). *Hybrid computational model for predicting bridge construction cost.* Automation in Construction, 35, 331–340. https://doi.org/10.1016/j.autcon.2013.05.016

Chou, J. S., Tsai, C. F., Pham, A. D., Lu, Y. H. (2014). *Machine learning in concrete strength simulations: Multi-nation data analytics.* Construction and Building Materials, 73, 771–780. https://doi.org/10.1016/j.conbuildmat.2014.09.054

Duan, Z., Kou, S. C., Poon, C. S. (2021). *Using artificial neural networks to predict the compressive strength of recycled aggregate concrete.* Construction and Building Materials, 289, 123155. https://doi.org/10.1016/j.conbuildmat.2021.123155

Etxeberria, M., Vázquez, E., Marí, A., Barra, M. (2007). *Influence of amount of recycled coarse aggregates and production process on properties of recycled aggregate concrete.* Cement and Concrete Research, 37(5), 735–742. https://doi.org/10.1016/j.cemconres.2007.02.002

Fanijo, E. O. (2023). *A comprehensive review on the use of recycled concrete aggregates for pavement construction.* Cleaner Materials, 8, 100159. https://doi.org/10.1016/j.clema.2023.100159

Gogoi, S., Singh, J. P., Verma, A. (2023). *Multicollinearity effects in machine learning models for concrete.* Materials Today: Proceedings, 72, 2580–2587. https://doi.org/10.1016/j.matpr.2022.09.507

Hall, J. W., Hall, J. L. (2017). *Engineering Decision-Making under Uncertainty*. Springer. https://doi.org/10.1007/978-3-319-69953-0

Huang, Y. H. (2004). *Pavement Analysis and Design* (2nd ed.). Pearson Prentice Hall.

Ismail, S., Ramli, M., Karim, M. R. (2022). *Influence of recycled aggregate quality and proportioning criteria on recycled concrete properties: A review.* Construction and Building Materials, 340, 127746. https://doi.org/10.1016/j.conbuildmat.2022.127746

Jagadesh, P., Karthik, K., Kalaivani, P. (2024). *Examining the influence of recycled aggregates on the fresh and mechanical characteristics of high-strength concrete: A comprehensive review.* Sustainability, 16(20), 9052. https://doi.org/10.3390/su16209052

Katz, A. (2003). *Properties of concrete made with recycled aggregate from partially hydrated old concrete.* Cement and Concrete Research, 33(5), 703–711. https://doi.org/10.1016/S0008-8846(02)01033-5

Khan, M. A., Khan, M. I., Aslam, F. (2022). *Machine learning-based prediction of recycled aggregate concrete strength.* Journal of Building Engineering, 45, 103115. https://doi.org/10.1016/j.jobe.2021.103115

Kumar, M., Samui, P., Kumar, D. R. (2024). *State-of-the-art machine learning models for predicting compressive strength of concrete: A comparative study.* Construction and Building Materials, 375, 130890. https://doi.org/10.1016/j.conbuildmat.2023.130890
Li, L., Wang, Y., Zhang, J., Li, Z. (2022). *Effects of multicollinearity on machine learning prediction of concrete strength incorporating recycled aggregates.* Engineering Structures, 252, 113620. https://doi.org/10.1016/j.engstruct.2021.113620
Limbachiya, M. C., Leelawat, T., Dhir, R. K. (2000). *Use of recycled concrete aggregate in high-strength concrete. Materials* and Structures, 33(9), 574–580. https://doi.org/10.1007/BF02480538
Liu, H., Wu, J., Chen, B. (2021). *Prediction of compressive strength of recycled aggregate concrete using ensemble machine learning techniques.* Materials, 14(18), 5263. https://doi.org/10.3390/ma14185263
Lundberg, S. M., Lee, S. I. (2017). *A unified approach to interpreting model predictions.* Advances in Neural Information Processing Systems, 30, 4765–4774.
Neville, A. M. (2011). *Properties of Concrete* (5th ed.). Pearson Education.
Nguyen, T. T., Pham, T. M., Hao, H. (2021). *Prediction of concrete compressive strength using machine learning: A comparative study.* Structures, 33, 120–134. https://doi.org/10.1016/j.istruc.2021.04.056
Ojha, P. N., Kaura, P., Singh, B. (2024). *Studies on mechanical performance of treated and non-treated coarse recycled concrete aggregate and its performance in concrete: An Indian case study.* Research on Engineering Structures & Materials, 10(1), 341–362. https://doi.org/10.17515/resm2023.53me0728rs
Padmini, A. K., Ramamurthy, K., Mathews, M. S. (2009). *Influence of parent concrete on the properties of recycled aggregate concrete.* Construction and Building Materials, 23(2), 829–836. https://doi.org/10.1016/j.conbuildmat.2008.03.006
Park, S., Kim, J., Lee, H. (2022). *Machine learning-based prediction of mechanical properties of recycled aggregate concrete under controlled experimental conditions.* Sustainability, 14(9), 5407. https://doi.org/10.3390/su14095407
Patil, R. D. (2025). *The performance of concrete with various recycled aggregates: Indian case studies and comparative assessment.* European Transport Research Review, 17(1), 12. https://doi.org/10.1186/s12544-025-00632-1
Pedro, D., de Brito, J., Evangelista, L. (2017). *Structural concrete with incorporation of coarse recycled concrete aggregates: Mechanical, durability and long-term properties.* Construction and Building Materials, 154, 294–309. https://doi.org/10.1016/j.conbuildmat.2017.07.185
Shahmansouri, A. A., Bengar, H. A., Akbarzadeh Bengar, H. (2021). *Data-driven prediction of concrete properties.* Engineering Structures, 226, 111353. https://doi.org/10.1016/j.engstruct.2020.111353
Silva, R. V, de Brito, J., Dhir, R. K. (2014). *Properties and composition of recycled aggregates from construction and demolition waste suitable for concrete production.* Construction and Building Materials, 65, 201–217. https://doi.org/10.1016/j.conbuildmat.2014.04.117
Tam, V. W. Y., Soomro, M., Evangelista, A. C. J. (2018). *A review of recycled aggregate in concrete applications.* Journal of Cleaner Production, 172, 272–291. https://doi.org/10.1016/j.jclepro.2017.10.236
Tam, V. W. Y., Tam, C. M., Ng, W. C. Y. (2005). *Microstructural analysis of recycled aggregate concrete produced from two-stage mixing approach.* Cement and Concrete Research, 35(6), 1195–1203. https://doi.org/10.1016/j.cemconres.2004.10.025
Tamagusko, J., Ferreira, R. M., Jalali, S. (2024). *Practical limitations of laboratory-based datasets in machine learning applications for recycled aggregate concrete*. Automation in Construction, 156, 105116. https://doi.org/10.1016/j.autcon.2023.105116

Tank, Y. R., Parmar, J. P., Gadhiya, D. H., Goyani, J. S. (2014). *Experimental study of compressive strength of recycled aggregate concrete.* International Journal of Engineering Research & Technology, 3(4), 1380–1384.
Thomas, C., Setién, J., Polanco, J. A., Alaejos, P., de Juan, M. (2013). *Durability of recycled aggregate concrete.* Construction and Building Materials, 40, 1054–1065. https://doi.org/10.1016/j.conbuildmat.2012.11.106
Xiao, J., Li, W., Fan, Y., Huang, X. (2018). *An overview of study on recycled aggregate concrete in China.* Construction and Building Materials, 31, 364–383. https://doi.org/10.1016/j.conbuildmat.2011.12.074
Yeh, I. C. (1998). *Modeling of strength of high-performance concrete using artificial neural networks.* Cement and Concrete Research, 28(12), 1797–1808. https://doi.org/10.1016/S0008-8846(98)00165-3
Zhang, J., Zhao, Y., Chen, Y. (2020). *Prediction of recycled aggregate concrete strength using support vector machine and ensemble learning.* Materials & Design, 190, 108572. https://doi.org/10.1016/j.matdes.2020.108572
Zhuang, Y., Chen, Y., Li, Z. (2021). *Performance comparison of machine learning models for concrete strength prediction.* Construction and Building Materials, 276, 122132. https://doi.org/10.1016/j.conbuildmat.2020.122132