RECYCLING STEEL SCRAP IN CONCRETE: A MACHINE LEARNING APPROACH FOR PREDICTING THE COMPRESSIVE STRENGTH

Rhad – E - Parsa[1], Md. Asar Uddin[2] and MD. Mehedi Hassan Galib*[3]

[1] *Graduate Student, Department of Building Engineering & Construction Management,* Rajshahi University of Engineering & Technology*, Bangladesh, e-mail:* rhadeee484522@gmail.com

[2] *Graduate Student, Department of Building Engineering & Construction Management, Rajshahi University of Engineering & Technology, Bangladesh, e-mail:* asaruddinbecm@gmail.com

[3] *Lecturer, Department of Building Engineering & Construction Management, Rajshahi University of Engineering & Technology, Bangladesh, e-mail:* mehedihassangalib@gmail.com

***Corresponding Author**

ABSTRACT

This research investigates the potential of machine learning approach in predicting the compressive strength of concrete using recycled steel scrap. Steel scrap: the waste product of the steel plate manufacturing industry consumes space and contaminates soil through rust and corrosion. Disposal of these steel scraps is considered to be a major issue all over the world. This study proposes to utilize a significant amount of waste materials from the steel industry as the reinforcement of concrete mixtures and analyze their impact on the compressive strength of concrete. Industrially, it is known as Steel Fiber Reinforced Concrete (SFRC), which is composed of fine and coarse aggregates, cement, water, and discontinuous-discrete steel fibers. The dataset of SFRC has been collected from previous studies where the laboratory test rasults have been included. For laboratory tests, the estimation of recycled materials needed for different percentages of SFRC has been done. Cylindrical Molds of height 12 inches and diameter of 6 inches have been created with the wet curing process, and the compressive strength and peak load have been obtained using uniaxial compression tests of these cylindrical sections. These experimental findings have been then preprocessed for capturing the non-linear dependencies of compressive strength of SFRC with its constituents and building several machine learning algorithms for further prediction. The data has been splitted into two parts- train set and test set. The train set with 80% data has been used to train three Machine learning algorithms- Random Forest Regressor, Extreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost). The Extreme Gradient Boosting algorithm has demonstrated the highest predictive accuracy, succeeding CatBoost and Random Forest. The R^2 values have been found as 0.98, 0.95 and 0.91 from XGBoost, CatBoost and Random Forest respectively. These algorithms had effectively captured the non-linear dependencies between steel scrap, cement, sand, coarse aggregate and water-cement ratio with the compressive strength of SFRC. Thus, this study has revealed the potential of machine learning algorithms for predicting the compressive strength of SFRC. By integrating an experimental and data-driven approach, the research has established a scalable methodology for optimizing the use of recycled steel scrap in concrete, contributing to sustainable development and advancing environmentally responsible construction practices. The findings have underscored machine learning's role in enhancing efficiency, optimizing material usage, and supporting eco-friendly infrastructure development. This research opens doors for improving the prediction accuracy using larger datasets along with advanced hybrid machine learning algorithms.

Keywords: *Artificial intelligence, machine learning, hyperparameter tuning, SFRC, compressive strength*

1. INTRODUCTION

Growing environmental concerns and the depletion of natural resources have encouraged researchers to explore the use of recycled materials in construction. One promising approach is the incorporation of recycled steel scrap into concrete, which not only improves its compressive strength but also helps to address the problem of industrial waste management (Singh & Singh, 2018). Steel scrap, a by-product of steel manufacturing, can pose serious environmental issues such as soil contamination, iron leaching, and excessive use of landfill space. The global steel industry generates approximately 400 million tons of steel scrap each year, much of which remains unused and improperly disposed (Khan et al., 2022).

Steel Fiber Reinforced Concrete (SFRC) is a composite material that includes steel fibers within the concrete matrix to enhance its mechanical properties—particularly tensile and flexural strength (Mena-Alonso et al., 2024). The addition of steel fibers has also been found to improve the durability and cracking resistance of concrete. Furthermore, SFRC exhibits better resistance to water penetration and chloride migration, which contributes to its long-term performance in aggressive environments (Şanal, 2018). Alongside the inclusion of steel scrap, several factors such as the water-cement ratio, curing conditions, aggregate characteristics, admixtures, and specimen size significantly influence the compressive strength of concrete. The use of recycled steel fibers introduces a novel form of micro-reinforcement that bridges fine cracks and enhances post-cracking behavior, offering an alternative to conventional reinforcement methods (Barros et al., 2005). In recent years, the integration of Machine Learning (ML) techniques in construction research has opened new possibilities for material optimization and mix design analysis. Traditional methods of concrete strength prediction often rely on empirical equations and laboratory testing, which can be time-consuming, costly and limited in scope. In contrast, Machine Learning algorithms such as Random Forest, CatBoost and Support Vector Regression can analyze diverse experimental datasets, identify complex correlations among input parameters and generate accurate predictive models with minimal human intervention.
Machine Learning helps to overcome inefficiencies and inaccuracies in traditional testing by identifying complex and non-linear patterns within large experimental datasets (Sharma et al., 2021). These models can be used for predictive analysis, enabling better control over material properties and supporting sustainable design decisions (Adhikari et al., 2025). By analyzing extensive experimental data, Machine Learning algorithms can predict the performance of different steel scrap–based concrete mixes with high accuracy, reducing the dependency on laboratory tests.

The use of Machine Learning algorithms not only enhances the precision and efficiency of strength prediction but also allows for the development of adaptable models that can be continuously improved with new data (Abed et al., 2022). This adaptability supports sustainable construction practices by minimizing material waste, optimizing mix proportions and reducing the environmental impact associated with excessive laboratory experimentation. Furthermore, by integrating Machine Learning algorithms into material design processes, engineers can make data-driven decisions, forecast long-term performance and accelerate innovation in eco-friendly construction materials. Therefore, this study has focused on examining the effect of recycled steel scrap on the compressive strength of concrete and developed several reliable Machine Learning algorithms to predict the compressive strength of Steel Fiber Reinforced Concrete (SFRC). The outcome of this research has been expected to contribute to sustainable construction practices by promoting the use of industrial waste materials and demonstrating the potential of artificial intelligence in material design and performance prediction.

2. METHODOLOGY

2.1 Data Collection

Dataset for this study has been collected from previous 4 studies. 566 observations have been obtained from (Yeh, 1998), 146 observations from (Penido et al., 2022), 56 observations from (Raju et al., 2025) and 534 observations from (Nhat-Duc, 2023) resulting in a total of 1302 observations for building several machine learning algorithms. The observations from previous 4 studies have been combined into one dataset for further preparation. The summary of the dataset has been added to Table 1.

Table 1: Statistical summary of the dataset

Property	Mean	Median	Standard Dev.	Minimum	Maximum
Cement	281.17	272.90	104.51	102.00	540.00
Steel Scrap	73.90	22.00	86.28	0.00	359.40
Fine Aggregate	972.92	965.95	112.98	594.00	1215.00
Coarse Aggregate	773.58	797.80	84.97	594.00	1025.00
Water Content	181.57	185.00	21.35	121.80	247.00
W/C Ratio	0.79	0.74	0.26	0.23	2.50
Age (days)	45.66	28.00	63.17	1.00	365.00
Compressive Strength	35.82	34.45	16.71	2.33	82.60

2.2 Data Preparation

The explanatory variables in this study have been considered as cement content, steel scrap, fine aggregate, coarse aggregate, water content, water-cement ratio and the response variable has been considered as the corresponding compressive strength of concrete. For consistency, the quantities of cement, steel scrap, fine aggregate, coarse aggregate and water have been recorded in kilograms per cubic meter (kg/m³) for all samples. The water-cement ratio has been expressed as a dimensionless value. Compressive strength has been measured in MPa in accordance with standard testing procedures.

In data pre-processing phase, a correlation matrix has been plotted to observe the dependencies of concrete constituents with its compressive strength. The correlation matrix has been added in Figure 1. Then the dataset has been splitted into two sets: train set with 80% data for building machine learning algorithms and test set with 20% data for determining the prediction accuracy of those algorithms.

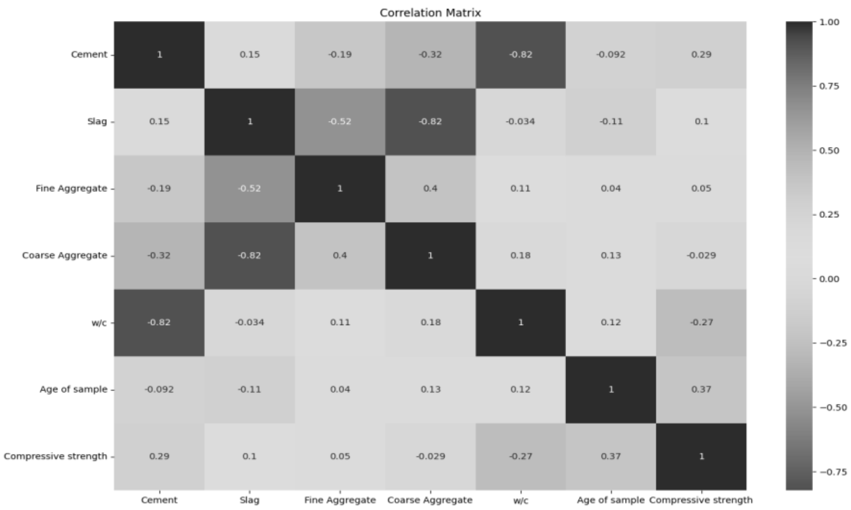

Figure 1: Correlation matrix

2.3 Machine Learning Algorithms

In this study, three ensemble machine learning algorithms: Random Forest (RF) Regressor, Extreme Gradient Boosting (XGBoost) and Categorical Boosting (CatBoost) have been used to observe the feasibility of using machine learning algorithms to predict the compressive strength of SFRC. An algorithm is said to be ensemble when it combines the predictions of multiple individual algorithms to improve the prediction accuracy. The typical flowchart of a machine learning algorithm has been added in Figure 2.

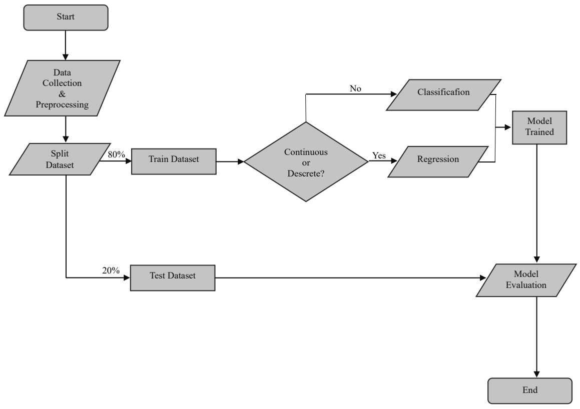

Figure 2: Typical flowchart of a machine learning algorithm

2.3.1 Random Forest Regressor

The Random Forest Regressor is a supervised machine learning algorithm that uses multiple decision trees as base learners. It operates on the principle of ensemble learning, where the combined predictions of several models lead to more reliable and accurate results. The number of decision trees used in a random forest is typically determined by the designer through a trial-and-error approach. A key feature of this algorithm is its use of resampling the data with replacement which is also known as the bootstrap method.

In this algorithm, different subsets of the training data have been generated and each subset has been used to train an individual decision tree. The final prediction of the random forest has been obtained by averaging the outputs of all the trees. This process, illustrated in Figure 3, has helped to improve prediction accuracy by creating a more generalized model and reducing the risk of overfitting. Additionally, hyperparameter tuning has been performed to optimize model performance and examine how different parameters affect prediction accuracy. The key hyperparameters include the number of trees in the forest and the maximum number of leaf nodes in each tree, both of which are typically adjusted through experimentation.

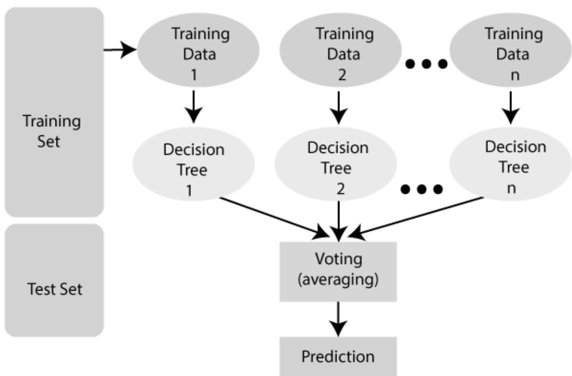

Figure 3: Typical structure of Random Forest algorithm

2.3.2 Extreme Gradient Boosting

The Extreme Gradient Boosting (XGBoost) Regression combines gradient boosting with decision trees as its base learners. Although it shares some conceptual similarities with the Random Forest algorithm, XGBoost introduces several enhancements that make it more powerful and efficient. Unlike Random Forest, which uses a fixed number of trees, XGBoost dynamically adjusts the number of trees during training based on performance. This adaptive process helps to prevent overfitting and improves the algorithm's ability to generalize to unseen data.

The core strength of XGBoost lies in its gradient boosting framework, where trees are built sequentially, each one correcting the errors made by its predecessors, as illustrated in Figure 4. Through this iterative learning process, the algorithm continuously refines its predictions by paying greater attention to the more difficult-to-predict samples.

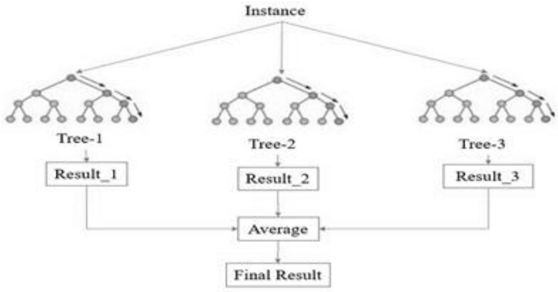

Figure 4: Typical structure of XGBoost regression

2.3.3 Categorical Boosting

The CatBoost Regression operates by combining multiple decision trees and improving predictive accuracy through gradient boosting. Gradient boosting works by sequentially training a series of relatively weak models, typically decision trees and each one learning from the errors of the previous model to form a stronger, more accurate overall predictor. This process is illustrated in Figure 5, where each subsequent decision tree learns from the mistakes of the earlier ones, thereby reducing the loss function step by step and refining the model's performance.

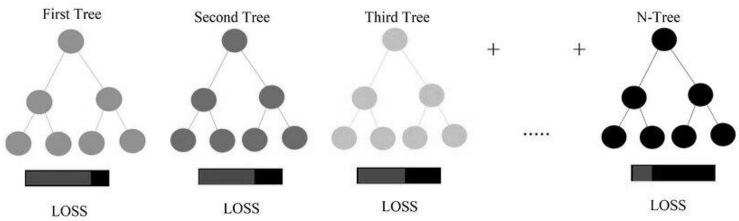

Figure 5: Sequential decision trees in CatBoost regression

A distinguishing feature of CatBoost Regression is its symmetric tree structure. Unlike XGBoost, where trees grow depth-wise and asymmetrically, CatBoost's trees grow in a balanced and symmetric manner. This means that all nodes at the same level split under the same conditions, resulting in more stable and efficient learning.

2.4 Evaluation of the Algorithm

Once the model has been trained, it has been required to evaluate the model using the test dataset. There are lots of techniques for carrying out performance measurement as well as error metrics to evaluate the models. In this study, coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) have been used as error metrics.

Formula of (R^2):

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{1}$$

In equation (1),
SS_{res} = Residuals sum of squares
SS_{tot} = Total sum of squares

Formula of MAE:

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|A_t - P_t\right| \tag{2}$$

Formula of MAPE:

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - P_t}{A_t}\right| \tag{3}$$

Formula of RMSE:

$$\mathrm{M} = \sqrt{\sum_{t=1}^{n}\frac{(A_t - P_t)^2}{n}} \tag{4}$$

In equation (2), Equation (3) and Equation (4),
M = Error metrics
A_t = Actual value
P_t = Predicted value
n = Number of observations be

3. RESULTS & DISCUSSION

3.1 Random Forest Regressor

The train dataset has been trained through Random Forest Regressor with three hyperparameter sets. The hyperparameters that have been considered are number of decision trees (n_estimators), maximum tree growth (max_depth) and minimum number of samples (min_samples_split). The R^2 values have been found as 0.91, 0.85 and 0.89 from RF-1, RF-2 and RF-3 respectively with three different hyperparameter sets. The best accuracy has been found from RF-1 with hyperparameter set: n_estimators as 1000, max_depth as 10 and min_samples_split as 3. The other error metrices have also been mentioned in Table 2 below.

Table 2: Hyperparameter sets of Random Forest Regressor

Set	Hyperparameters	R^2	RMSE	MAE	MAPE
RF-1	n_estimators:1000 max_depth: 10 min_samples_split: 3	0.91	5.90	4.08	16.02%
RF-2	n_estimators: 800 max_depth: 8 min_samples_split: 2	0.85	7.52	5.33	17.64%
RF-3	n_estimators: 500 max_depth: 10 min_samples_split: 3	0.89	6.56	4.36	14.64%

The difference between actual and predicted compressive strength obtained from RF-1 has been plotted in Figure 6.

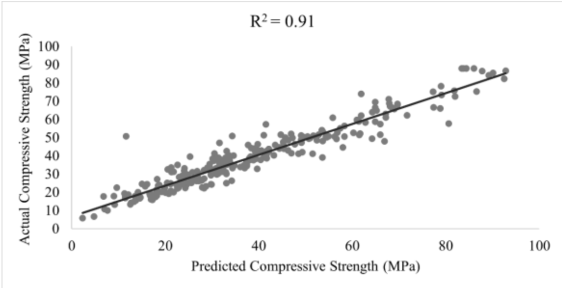

Figure 6: Actual vs predicted compressive strength (RF-1)

3.2 Extreme Gradient Boosting Regressor

Similarly, in XGBoost algortihm The train dataset has been trained with three hyperparameter sets. The hyperparameters that have been considered are number of decision trees (n_estimators), step size (learning_rate) and maximum tree growth (max_depth). The R^2 values have been found as 0.97, 0.98 and 0.97 from XGB-1, XGB -2 and XGB -3 respectively with three different hyperparameter sets. The best accuracy has been found from XGB -2 with hyperparameter set: n_estimators as 1000, learning_rate as 0.5 and max_depth as 5. The RMSE, MAE and MAPE have been found as 2.73, 1.77 and 5.92% respectively. The other error metrices have also been mentioned in Table 3 below.

Table 3: Hyperparameter sets of Extreme Gradient Boosting Regressor

Set	Hyperparameters	R^2	RMSE	MAE	MAPE
XGB-1	n_estimators: 500 learning_rate: 0.15 max_depth: 5	0.97	3.59	2.43	8.28%
XGB-2	n_estimators:1000 learning_rate: 0.5 max_depth: 5	0.98	2.73	1.77	5.92%
XGB-3	n_estimators: 1000 learning_rate: 0.05 max_depth: 10	0.97	3.52	2.13	7.52%

The difference between actual and predicted compressive strength obtained from XGB-2 has been plotted in Figure 7.

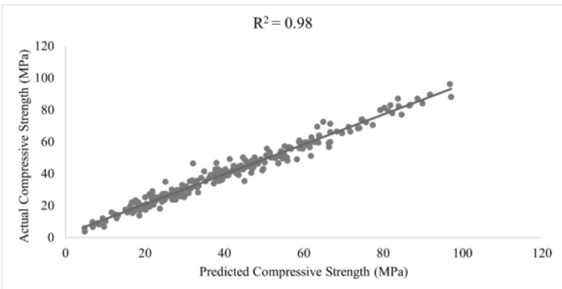

Figure 7: Actual vs predicted compressive strength (XGB-2)

3.3 CatBoost Regressor

In CatBoost Regressor, the hyperparameters that have been considered are number of boosting (iterations), step size (learning_rate) and tree growth (depth). The R^2 values have been found as 0.95, 0.94 and 0.95 from CTB-1, CTB -2 and CTB -3 respectively with three different hyperparameter sets. Same R^2 value has been found from CTB-1 and CTB-3. From the other error metrices mentioned in Table 4, it has been observed that the mean absolute percentage errors (mape) have been found as 9.03% and 9.50% from CTB-1 and CTB -3 respectively. It has indicated that CTB-1 has got the ability to predict the compressive strength of SFRC with more accuracy as compared to CTB-3.

Table 4: Hyperparameter sets of CatBoost Regressor

Set	Hyperparameters	R^2	RMSE	MAE	MAPE
CTB-1	iterations: 1000 learning_rate: 0.05 depth=7	0.95	4.49	2.90	9.03%
CTB-2	iterations: 1000 learning_rate: 0.5 depth=8	0.94	5.18	2.91	10.78%
CTB-3	iterations: 500 learning_rate: 0.5 depth=8	0.95	4.50	2.67	9.50%

The difference between actual and predicted compressive strength obtained from CTB-1 has been plotted in Figure 8.

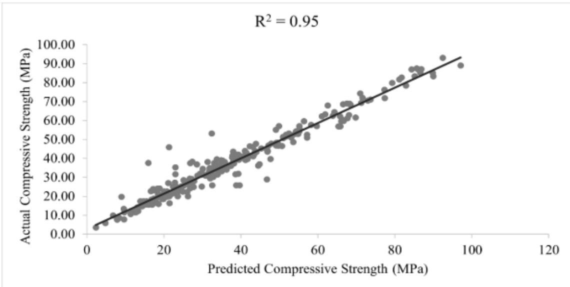

Figure 8: Actual vs predicted compressive strength (CTB-1)

3.4 Comparison of the Algorithms

It has been observed that the XGBoost Regressor has performed quite well in predicting the compressive strength of SFRC as compared to the other algorithms. In XGBoost Regressor, the decision trees have been built sequentially, with each new tree correcting the errors of the previous one, while in Random Forest, the decision trees have been built independently in parallel. Since XGBoost has considered the errors of the previous models, its accuracy has become better with each iteration and has got the ability to capture the complex non-linear dependencies between the explanatory and response variables.

The comparison of error metrices of the three algorithms has been mentioned in Table 5 below.

Table 5: Comparison of the algorithms

Set	R^2	RMSE	MAE	MAPE
RF-1	0.91	5.90	4.08	16.02%
XGB-2	0.98	2.73	1.77	5.92%
CTB-1	0.95	4.49	2.90	9.03%

4. CONCLUSIONS

This research has successfully demonstrated that incorporating Machine Learning algorithms for predicting compressive strength of SFRC has achieved significant accuracy. Machine learning algorithms such as Random Forest, XGBoost and CatBoost have been developed to predict the compressive strength and the prediction accuracies of those algorithms have been carried out. The hyperparameters of these algorithms have been tuned to achieve greater accuracy. As a result, models like XGBoost and CatBoost have shown significant predictive performance achieving higher R^2 values such as 0.98 and 0.95 respectively with lower error matrices values such as 5.92% and 9.03% respectively. These numbers have indicated their effectiveness in handling the non-linear behavior of

SFRC and its constituents. The integration of Machine Learning with concrete has not only improved compressive strength prediction possibilities for new features such as steel scrap but also contributed to waste reduction and sustainability in construction practices. The study has supported the idea that sustainable material usage and data driven approach can lead to smarter and greener infrastructure solutions. Future studies with hybrid algorithms and deep learning models can be carried out for improving further predictive accuracy.

DECLARATION OF USE OF AI

The authors declare that no artificial intelligence (AI) tools or technologies were used in the preparation, writing, analysis, or revision of this manuscript. All content, including the research design, data analysis, interpretation of results, and manuscript writing, was carried out only by the authors.

REFERENCES

Abed, Y. G., Hasan, T. M., Zehawi, R. N., & Maseer, Z. K. (2022). Machine learning algorithms for constructions cost prediction: A systematic review. International Journal of Nonlinear Analysis & Applications, 13(2), 2205–2218.

Adhikari, A., Karunaratne, T., & Sumanarathna, N. (2025). Machine learning techniques for predictive maintenance of building services: A comprehensive review and research outlook. Facilities. Facilities, 43(11–12), 803–817.

Barros, J. A. O., Cunha, V. M. C. F., Ribeiro, A. F., & Antunes, J. A. B. (2005). Post-cracking behaviour of steel fibre reinforced concrete. Materials and Structures, 38(1), 47–56.

Sharma, V., Zaki, M., Jha, K. N., & Krishnan, N. M. A. (2021). Machine learning-aided cost prediction and optimization in construction operations. Engineering, Construction and Architectural Management, 29(3), 1241–1257.

Khan, W. S., Asmatulu, E., Uddin, M. N., & Asmatulu, R. (2022). Recycling and reusing of metals and alloys. In Recycling and reusing of engineering materials (pp. 163–179).

Mena-Alonso, Á., González, D. C., Mínguez, J., & Vicente, M. A. (2024). Mechanical behavior of high-strength steel fiber-reinforced concrete. In Fiber-reinforced composites: Recent advances, new perspectives and applications (pp. 71–83).

Nhat-Duc, H. (2023). Estimation of the compressive strength of concretes containing ground granulated blast furnace slag using a novel regularized deep learning approach. Multiscale and Multidisciplinary Modeling, Experiments and Design, 6, 415–430.

Penido, R. E. K., da Paixão, R. C. F., Costa, L. C. B., Peixoto, R. A. F., Cury, A. A., & Mendes, J. C. (2022). Predicting the compressive strength of steelmaking slag concrete with machine learning – Considerations on developing a mix design tool. Construction and Building Materials, 341, 127896.

Raju, M. R., Ahmad, S. I., Hasan, M. M., Sadiqul Hasan, N. M., Islam, M. M., Basit, M. A., Hossain, I. T., Santo, S. A., Alam, M. S., & Rahman, M. (2025). Machine learning for predicting strength properties of waste iron slag concrete. Heliyon, 11, e42133.

Şanal, İ. (2018). Discussion on steel fiber reinforced concretes (SFRC). Trends in Civil Engineering and Its Architecture, 2(1), 1–2.

Singh, H., & Singh, Y. (2018). Applications of recycled and waste materials in infrastructure projects. Lecture Notes in Civil Engineering, 21, 569–585.

Yeh, I.-C. (1998). Modeling of strength of high-performance concrete using artificial neural networks. Cement and Concrete Research, 28(12), 1797–1808.