COMPRESSIVE STRENGTH PREDICTION OF HIGH-STRENGTH GRAPHENE REINFORCED CONCRETE USING MACHINE LEARNING

Rahat Aayaz*[1], Md. Habibur Rahman Sobuz[2], Sumaiya Mifra Akhter[3] and Jannat Ara Jabin[4]

[1] Graduate Student, Khulna University of Engineering & Technology, Bangladesh, e-mail: aayaz1823051@stud.kuet.ac.bd
[2] Associate Professor, Khulna University of Engineering & Technology, Bangladesh, e-mail: habib@becm.kuet.ac.bd
[3] Graduate Student, Khulna University of Engineering & Technology, Bangladesh, e-mail: sumayamifraakter@gmail.com
[4] Graduate Student, Khulna University of Engineering & Technology, Bangladesh, e-mail: jannatjabin6246@gmail.com

***Corresponding Author**

ABSTRACT

The cement industry is pivotal in the global construction sector, as cement is one of the most widely used building materials. The cement industry is of significant importance in the worldwide construction sector due to its status as one of the most prevalent building materials. On the contrary, the manufacturing process of cement is associated with substantial carbon emissions, thereby increasing environmental degradation. The potential environmental impact associated with cement production could be mitigated through the utilization of graphene in Graphene Reinforced Concrete (GRC), which improves the strength and durability of concrete. This could be achieved by decreasing the cement content necessary for construction projects. This study emphasizes the significance of incorporating sustainable alternatives into the building sector. The incorporation of graphene into concrete matrices has attracted considerable interest in light of the growing need for environmentally friendly and high-performing building materials. This is mainly due to graphene's remarkable mechanical properties and capacity to enhance concrete's durability and strength. The predictive models in this study are constructed utilizing various machine learning algorithms, such as support vector machines, random forests, and artificial neural networks. Feature engineering techniques are employed in order to extract crucial information from the dataset. In the meantime, the model's performance is thoroughly evaluated using cross-validation and evaluation metrics, including mean absolute error, mean squared error, root mean squared error, and coefficient of determination (R-squared). The predictive outcomes derived from this research undertaking revealed the compressive strength of complex concrete mixtures, including high-strength graphene-reinforced concrete, with an R^2 value of 0.8395 and an accuracy of 84%. The impacts of the results presented in this study for the construction industry are substantial, as they provide a data-driven methodology for optimizing GRC mix designs, which can lead to the creation of sustainable and robust building materials. This study enables the possibility of integration of machine learning and advanced materials science, thereby encouraging novel approaches in the domains of construction technology and civil engineering.

Keywords: *Graphene, Machine learning, Prediction, Compressive strength, Random Forest Regressors*

1. INTRODUCTION

Concrete is an essential building material made up of many ingredients that improve its binding and strength characteristics. It is necessary to comprehend its mechanical and physical characteristics, particularly its compressive strength. Temperature, curing period, and compaction level can all affect compressive strength. For this reason, estimating the compressive strength of concrete is crucial to producing durable, reasonably priced concrete mixes. The main component of concrete is cement, mixed with aggregates and chunky materials like sand and stones. Concrete is mixed well and then poured into forms to harden and set. The environmental impact of cement, especially its carbon footprint, is a significant problem. Approximately 7% of CO^2 emissions worldwide are attributed to the concrete industry's principal source, cement manufacture (Thomas Czigler, 2020). The graph displayed in Fig. 1 illustrates this effect (Thomas Czigler, 2020).

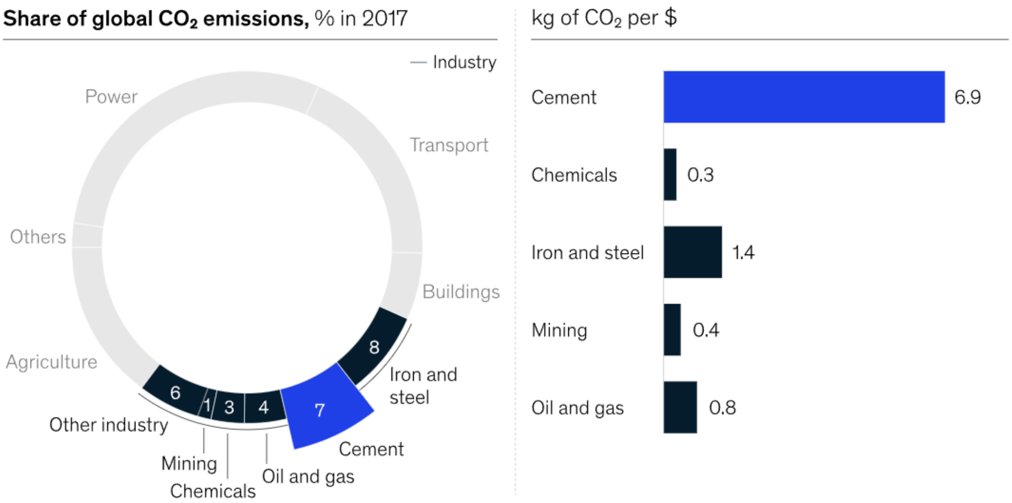

Figure 1: Global CO_2 emissions by category/cost of cement (Thomas Czigler, 2020).

The building construction sector encounters substantial obstacles in attaining swift innovation and sustained growth, heightening the significance of novel composite materials (Zaid et al., 2024). The development of advanced cement composites, such as multifunctional cementitious materials (MCMs), engineering cementitious composites (ECCs), high-strength concrete (HSC), and high-performance cement composites (HPCCs), is the main emphasis of current civil engineering (Monteiro et al., 2022; Mostafa et al., 2022). Because of its exceptional mechanical and electrical properties, graphene, a material discovered in 2004 by Andre Geim and Konstantin Novoselov, has recently garnered much attention in composite materials. (Geim & Novoselov, 2007). They were awarded the 2010 Nobel Prize in Physics for their revolutionary discovery of graphene, defined as "a thin sheet made of only one layer of carbon atoms." (Stojkovic & Louro, 2022). Graphene's unique chemical and physical characteristics have made it a focus of research (Bakhshandeh & Shafiekhani, 2018; Sadak et al., 2018). Because of its extraordinary endurance and strength, graphene is an achievable choice for increasing the mechanical properties of concrete. It can remarkably increase concrete mixtures' tensile and compressive strength, enhancing the material's resistance to structural failure and cracks (Zhao et al., 2020).
According to studies, adding graphene to concrete can increase its flexural strength by up to 95% and its compressive strength by up to 27% (Dalal & Dalal, 2021). A number of algorithms based on machine learning have been employed for predicting concrete's compressive strength (Liu, 2022). Several methods rely on basic mathematical models. For instance, Behnood et al. evaluated the concrete's compressive, split tensile, and flexural strengths using the M5P model. (Behnood & Golafshani, 2020) Yan and Shi discovered that when it came to predicting the elastic modulus of both regular and high-strength concrete, Support Vector Regression (SVR) performed more efficiently than alternative models. (Yan & Shi, 2010). However, it's common for basic mathematical models to struggle to generate precise equations. Using a hybrid GS-SVR model, Wu et al. (Wu & Zhou, 2022) found the compressive strength of sustainable concrete and achieved an R^2 score of 0.93. (Almahameed & Sobuz, 2023) observed that this model has an MSE value of 17.6, an R^2 = 0.83, and an R-squared value of

0.926858. It is challenging to employ machine learning to anticipate the behavior of graphene-enhanced concrete because of the random nature of concrete and the lack of data necessary to make precise predictions. Concrete is a complex and heterogeneous material used in building. It is made up of a combination of different ingredients, such as cement, water, aggregates, and maybe additives. These components demonstrate a great deal of variety. (Nguyen et al., 2022), Resulting in various qualities for the finished concrete. Furthermore, outside variables such as curing conditions—such as temperature and humidity—have an impact and further complicate the behavior of concrete over time (Conduit et al., 2023).

Concrete reinforced with graphene adds further complication (Cunningham et al.). The form and alignment of graphene sheets in the mixture can have a considerable impact on the material properties, which are challenging to manage and predict. The restricted availability of appropriate data for machine learning model training exacerbates the difficulty. There needs to be more historical data due to the relatively recent application of graphene in concrete, and gathering new data through experiments is expensive and time-consuming.

A multi-faceted approach is essential to address these formidable challenges (Chetty et al., 2022). This technique involves collecting certain experimental data, carefully designing the key attributes, and maybe utilizing data expansion techniques to duplicate the unpredictable behavior of the concrete. It is important to implement machine learning models capable of handling unpredictability and uncertainty, such as Bayesian models or ensemble techniques. Overfitting, a prevalent concern in small datasets, can be reduced by implementing regularization techniques.

1. METHODOLOGY

2.1 Dataset Preparation

The data utilized for this study was sourced from peer-reviewed published papers and laboratory experiments(Abiodun et al., 2023; Cao et al., 2016; Jiang et al., 2021; Liu et al., 2016; Lu et al., 2015; Rhee et al., 2016; Shahab et al., 2024; Zaid et al., 2022; Zhao et al., 2020). Furthermore, the lack of basic elements such as water, cement, coarse aggregates, and fine aggregates was a result of insufficient data presented in the articles. A total of 130 samples were collected to produce the final dataset for compressive strength analysis of high-performance concrete. The specimens underwent evaluation by various university research laboratories and were treated under controlled conditions.

- **w/c ratio:** The water-cement ratio, also known as the water-to-cement ratio (w/c ratio), represents the relationship between the weight of water and the weight of cement employed in a concrete mixture ("Water–cement ratio," 2023).

- **Graphene Content:** The column "Graphene" indicates the different levels of graphene present in each mixture, expressed as a percentage of the total volume of the mixture.

- **Age of Concrete (Days):** The column "Age (day)" in the dataset accounts for the element of time, reflecting measurements taken at various points during the curing process of concrete (e.g., 3 days, 28 days, 56 days, etc.). This information provides insights into how the strength of graphene-enhanced concrete changes as it matures over time.

- **Concrete Compressive Strength:** Compressive strength is a crucial attribute of concrete, signifying its capacity to withstand loads before succumbing to compression failure. A higher compressive strength signifies that the concrete is suitable for more demanding structural uses. This dataset evaluates the compressive strength of concrete with different levels of graphene content at various stages of maturity.

- **Percentage of Increase/Decrease:** This column provides valuable information by comparing the compressive strength of graphene-enhanced concrete to a reference scenario where no graphene is included (Graphene = 0%). This percentage illustrates the extent to which the concrete's strength is improved or diminished when graphene is introduced, relative to conventional concrete lacking graphene. To clarify, a percentage exceeding 100% indicates a strength enhancement compared to the reference, whereas a percentage below 100% suggests a reduction in strength.

Overall, this dataset enables an in-depth examination of the impact of graphene on the time-dependent compressive strength of concrete.

2.2 Data transformation:

Several data transformation strategies are used to optimize the model's performance and boost its effectiveness. Additionally, it facilitates faster computation times, which improves efficiency. Several techniques may be used to achieve this, such as data splitting, data normalization, and outlier removal (Davawala et al., 2023).

2.3 Data normalization

Data normalization is the process of standardizing the range of independent variables or characteristics in a dataset. (vpadmin, 2023). In machine learning, several normalization techniques are employed, such as:

- **Min-Max Normalization:** The feature is rescaled using this method to a defined range of [0,1].

- **Z-Score Normalization (Standardization):** With a mean of 0 and a standard deviation of 1, the feature is rescaled to resemble a conventional normal distribution using this procedure. (Orthey et al., 2019).

$$x' = (x - \mu) / \sigma \tag{1}$$

Where x = is the initial or supplied value.
μ = The average value of the feature sets.
σ = The feature values' standard deviation.

2.4Removing outliers:

Before removing outliers, it is essential to establish a method for identifying them. Common strategies include:

- **Standard Deviation:** Assuming a normal distribution, around 68% of data points will fall within one standard deviation of the mean, whereas approximately 95% will fall within two standard deviations. An outlier is a data point that deviates from the mean by a specific number of standard deviations.

- **Interquartile Range (IQR):** The interquartile range, as defined by Alshammari et al. (2023), refers to the difference between the 25th percentile (first quartile) and the 75th percentile (third quartile). In context, outliers are defined as observations that are below Q1 (the first quartile) minus 1.5 times the interquartile range (IQR) or above Q3 (the third quartile) plus 1.5 times the IQR (Alshammari et al., 2023).

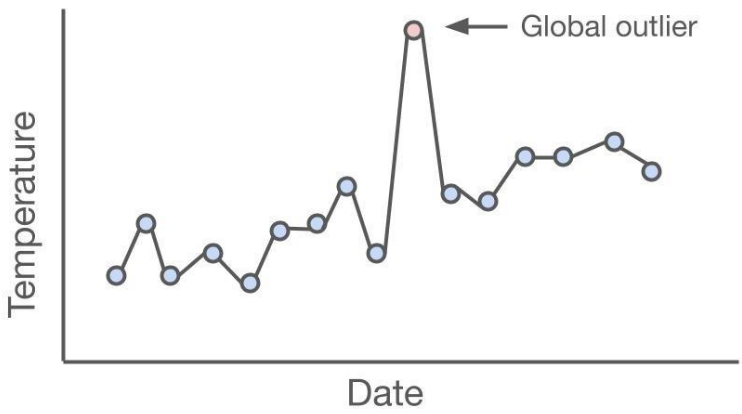

Figure 2: Line graph displaying temperature data with an outlier (M, 2021).

2.5Performance Indices

Table 1: Equation for statistical indicators and allowable ranges.

Statistical Indicator Equation	Acceptable range	References
$\mathrm{R}^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$	Close to 1	(Dong et al., 2022)
$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$	MAE < RMSE	(Iqbal et al., 2021)
$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$	Closer to zero (0≤MSE≤∞)	(M, 2021)
$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\lvert y_i - \hat{y}_i\rvert$	Greater than 0.65 for an excellent model	(N. Moriasi et al., 2007)

2.6Machine learning models

The predictions in this investigation were generated employing a modified version of machine learning [ML] algorithms, comprising gradient boosting regression, support vector regression, k-nearest neighbor (kNN) regression, Random Forest regression, Ridge regression, Lasso regression, and linear regression. This study utilized these advanced models because of their relevance, efficiency, simplicity, and expertise.

- **Linear regression:** Linear regression is a statistical technique employed in machine learning and artificial intelligence to forecast a dependent variable by considering the values of one or more independent variables ("Linear Regression Formula – Definition, Formula Plotting, Properties, Uses and Solved Questions,").

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ... + \theta_m x_m$$

$$J(\theta) = \sum_{i=1}^{n} (y - \hat{y})^2 / 2n$$

- **Ridge regression:** Ridge regression is a method used to estimate the coefficients of a multiple-regression model where the independent variables are strongly correlated (Hilt et al., 1977). Tikhonov regularization, coined by Andrey Tikhonov, is a method of regularization used to address ill-posed problems (Hilt et al., 1977). It is particularly efficient in mitigating the issue of multicollinearity in linear regression, a common occurrence in models with several parameters.

$$J(\theta) = \sum_{i=1}^{n} (y - \hat{y})^2 / 2n + \alpha \sum_{i=1}^{m} \theta_i^2$$

- **Lasso regression:** Regularization is a type of regression known as Lasso regression. Regression approaches are preferred over other methods for more precise prediction (Team, 2023). This model leverages the phenomenon of shrinking. Shrinkage refers to the procedure of reducing data values towards a central point called the mean (Team, 2023). The lasso method promotes the use of uncomplicated and sparse models, which have a reduced number of parameters. This regression method is well-suited for models that exhibit a significant degree of multicollinearity or require automated procedures for elements like variable selection or parameter removal (Team, 2023).

$$J(\theta) = \sum_{i=1}^{n} (y - \hat{y})^2 / 2n + \alpha \sum_{i=1}^{m} |\theta_i|$$

- **Decision tree:** A decision tree employs a hierarchical structure to construct models for regression or classification purposes. It iteratively partitions a dataset into increasingly smaller subsets while simultaneously constructing a corresponding decision tree. The final outcome is a tree consisting of leaf and decision nodes ("Decision Tree Regression,").

$$J(\theta) = m_L MSE_L + m_R MSE_R / m$$

$$MSE_{node} = \sum_{i \in node} (y - \hat{y})^2 / 2m_{node}$$

- **Gradient boosting:** Gradient boosting is a machine learning method employed in several applications, including regression and classification. The system provides a predictive model that consists of a collection of weak prediction models, specifically basic decision trees, which have minimal assumptions about the data (Piryonesi & El-Diraby, 2020).

- **k-nearest neighbor regression:** k-NN is a classification technique where the function is only estimated in a local manner, and all calculations are deferred until the function is assessed. (Candelaria et al., 2022). Normalizing the training data can greatly enhance the effectiveness of this method, as it relies on distance for classification. This is particularly beneficial when the characteristics exhibit diverse physical units or vary in magnitude (Candelaria et al., 2022).

$$d = \sqrt{(x - x_i)^2 + (y - y_i)^2 + ... + (z - z_i)^2}$$

- **Random forest regression:** Random Forest Regression is a supervised learning algorithm that employs ensemble learning techniques for regression analysis. Ensemble learning is a methodology that integrates predictions from numerous machine learning algorithms to generate a more precise forecast compared to using a single model (Chaya, 2022).

- **AdaBoost regressor:** The AdaBoost regressor method integrates multiple weak classifiers into a single robust and dependable classifier. This technique is distinctive in its allocation of more significance to cases that are challenging to foresee while assigning less importance to examples that are easily predictable.

- **Bagging regressor:** A Bagging regressor is a meta-estimator ensemble that trains base regressors on random subsets of the original dataset. It then combines their individual forecasts, either by voting or averaging, to provide a final prediction. This functionality can be found in the "sklearn.ensemble.BaggingRegressor" module. A meta-estimator, such as "sklearn.ensemble.BaggingRegressor," can be employed to decrease the variability of a black-box estimator, like a decision tree. This is achieved by including randomness into its creation process and creating an ensemble from it ("sklearn.ensemble.BaggingRegressor,").
- **Extra tree regressor:** A regressor utilizing the extra-trees algorithm. This class use a meta estimator to train several randomized decision trees (also known as extra-trees) on different subsets of the dataset. It employs averaging to enhance the accuracy of predictions and regulate over-fitting ("sklearn.ensemble.ExtraTreesRegressor,").
- **k-nearest neighbor (kNN):** KNN regression is a non-parametric technique that estimates the relationship between independent variables and a continuous outcome by calculating the average of observations in the nearby vicinity (Teixeira-Pinto).
- **Neural Network (MLP):** The term "Multilayer perceptron (MLP)" is a misnomer for a contemporary feedforward artificial neural network. This network comprises fully connected neurons that employ a non-linear activation function. It is structured into a minimum of three layers and is particularly noteworthy for its ability to differentiate input that cannot be separated linearly (Cybenko, 1989).

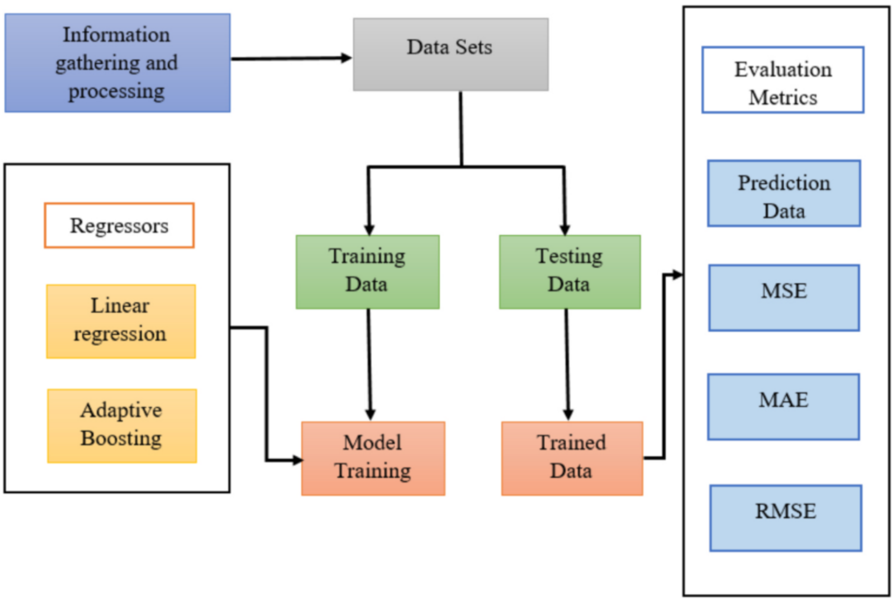

Figure 3: Visual depiction of the presented models (Habibur Rahman Sobuz et al., 2024).

2.7 Exploratory Data Analysis (EDA)

The process known as exploratory data analysis, or EDA, aids in analyzing and examining data sets to summarize their key characteristics. Techniques for data visualization are frequently used to do it. The present study examined the impact of varying concentrations of graphene on concrete by measuring its compressive strength at 3, 7, or 28-day intervals. Specifically, scatter plots were used to examine the link between graphene concentration and concrete strength, box plots were used to examine the distribution of strength at various ages, and histograms were made to show the distribution of each numerical variable.
To conduct an Exploratory Data Analysis (EDA), the following tasks were completed:

Basic Statistics: General statistics such as average, median, standard deviation, and so forth.

Distributions: The important numerical values' distribution.

Correlations: Relationships between several variables.

Graphical Analysis: Plots and graphs used to portray information visually are called graphic analyses.

Histogram Analysis: Regarding graphene, the graphene histogram has a bimodal distribution, indicating two comparable graphene contents across the samples. This suggests that in concrete compositions, specific concentrations are often chosen. Concrete sample distributions of age are typically homogeneous, with peaks occurring at specific ages. This implies that some healing times are studied more often than others. The histogram for the compressive strength of concrete displays a right-skewed distribution, indicating that while lower strength values are more common, there are large variations in the compressive strength of concrete samples. The distribution of the variable for Increase/Decrease in Percentage is similarly biased to the right. While some samples indicate a higher percentage, indicating an increase in compressive strength over a baseline, the majority of results are relatively close to 100%.

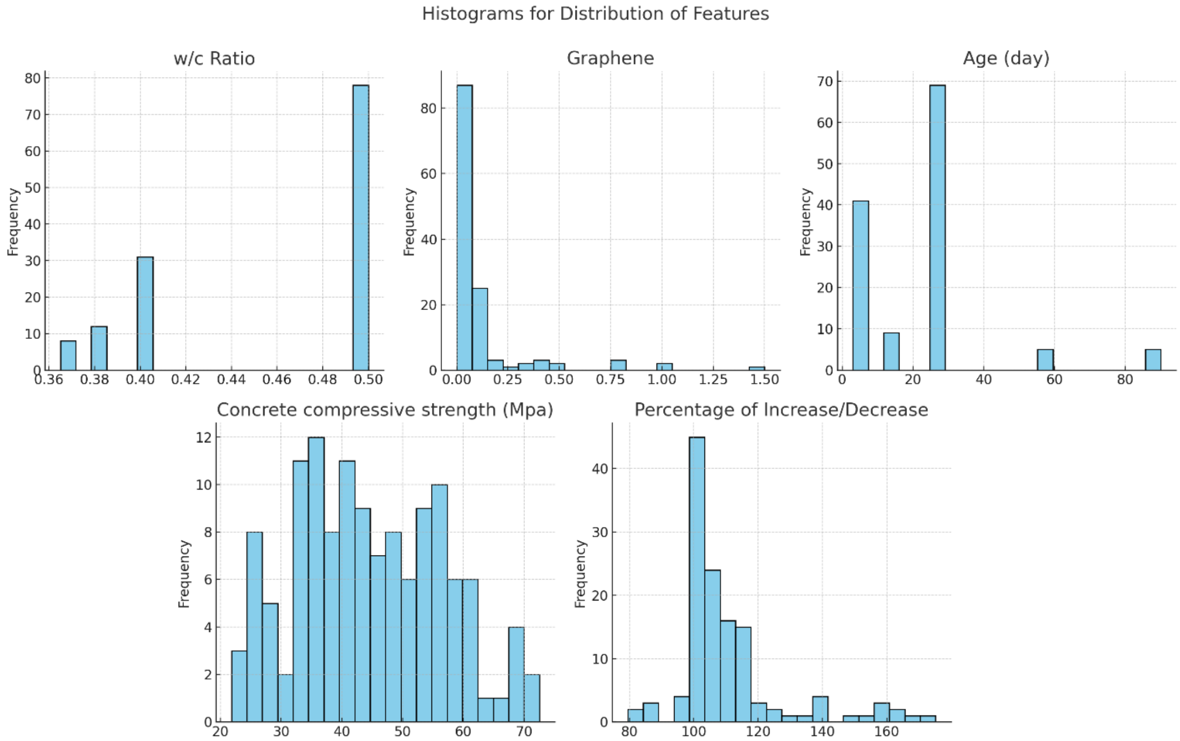

Figure 4: Histograms for distribution of the features.

Correlation Heatmap: The correlation coefficients between the various variables in your dataset are shown visually in the correlation heatmap. This heatmap displays the correlation between the variables on each axis as a square. The range of correlation is -1 to +1. Strong positive correlations are indicated by values closer to +1, strong negative correlations by values closer to -1, and no correlations are indicated by values around 0. The color scale makes these associations easier to see, going from blue (negative correlation) to red (positive correlation).

The main findings from the dataset are:

- **w/c Ratio vs. Concrete Compressive Strength (MPa):** This chart demonstrates a negative association, meaning that the concrete compressive strength tends to rise when the W/C ratio falls.

- **Graphene vs Concrete compressive strength (MPa):** This indicates a positive association, suggesting that increased concrete compressive strength is linked to a higher graphene concentration.

- **Age (day) vs. Concrete compressive strength (MPa):** Here, there is a significant positive association that suggests that the strength of concrete grows with age.

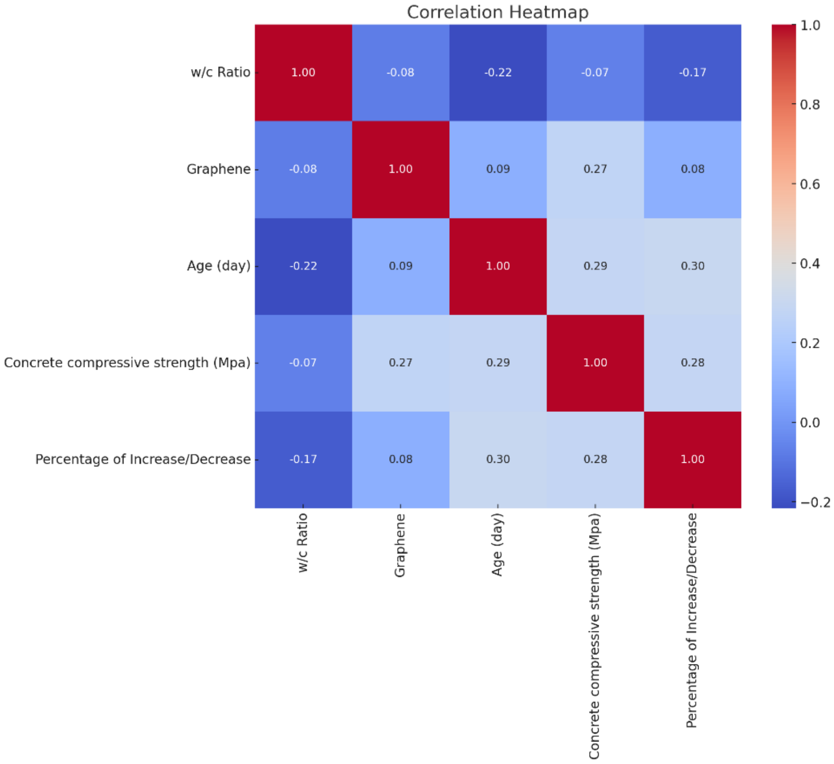

Figure 5: Correlation Heatmap

2. RESULTS AND DISCUSSIONS

Twelve different machine learning approaches were used to predict the compressive strength of concrete. The accuracy measures provided a comprehensive evaluation of the performance of each algorithm. The results suggest that the most effective strategy for this prediction has been identified. Both Table 2 and Figure 6 present the evaluation metrics for the different models used.

Table 2: Different model's accuracy scores and errors.

Model	R^2	MAE	MSE	RMSE
Bagging Regressor	**0.8395**	**1.9684**	**17.6655**	**4.2030**
Random Forest	**0.8268**	**2.1996**	**17.9423**	**4.2358**
Gradient Boosting	**0.8240**	**2.2498**	**18.2366**	**4.2704**
Extra Trees Regressor	**0.7901**	**2.1810**	**21.7474**	**4.6634**
AdaBoost	**0.7258**	**3.6146**	**28.4052**	**5.3297**
Decision Tree	**0.6581**	**2.6657**	**35.4165**	**5.9512**
KNeighbors Regressor	**0.5614**	**3.4303**	**45.4368**	**6.7407**
Neural Network (MLP)	**0.2532**	**6.6812**	**77.3583**	**8.7954**
Support Vector Regression	**0.1579**	**7.3562**	**87.2362**	**9.3400**
Linear Regression	**0.1448**	**7.1972**	**88.5939**	**9.4124**

Lasso Regression	**0.1440**	**7.5861**	**88.6721**	**9.4166**
Ridge Regression	**0.1263**	**7.4855**	**90.5027**	**9.5133**

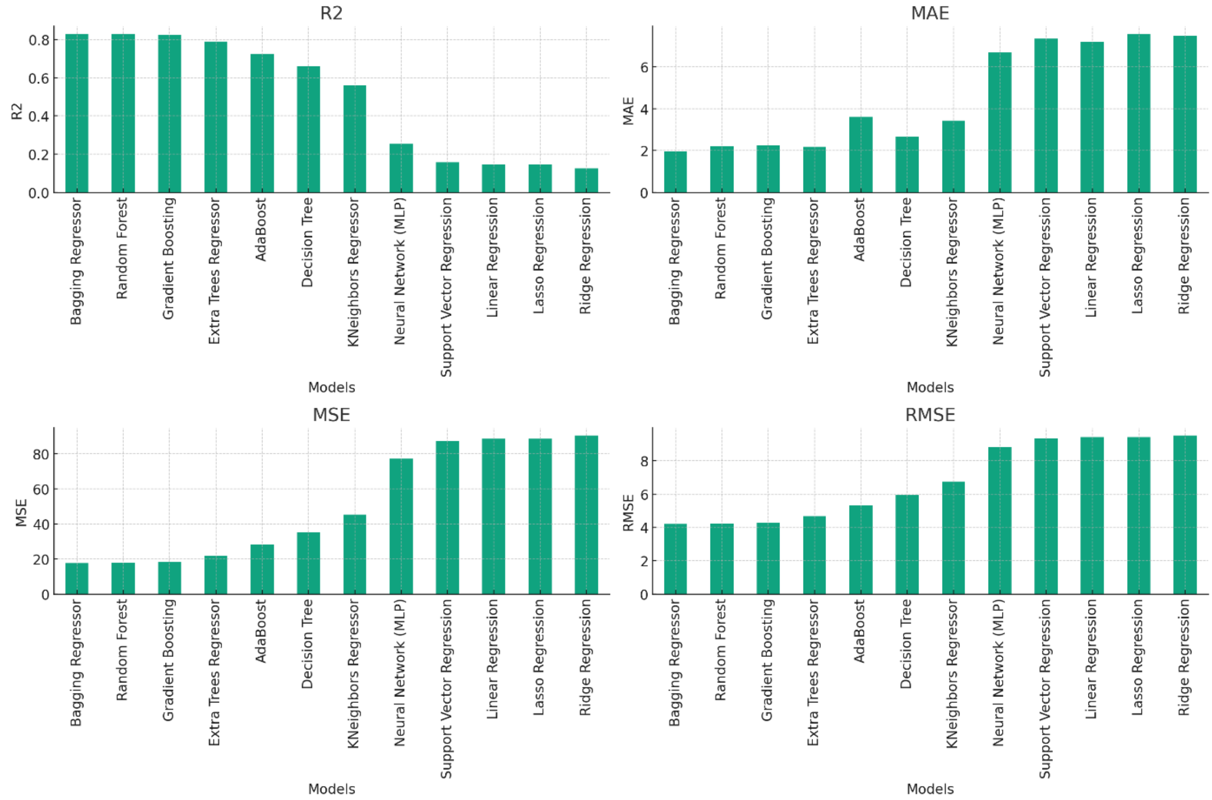

Figure 6: Bar plots for different models evaluating matrices.

Table 2 presents a hierarchy of regression model performance, with ensemble methods like Bagging Regressor and Random Forest at the top, indicating their superior ability to predict accurately in the given dataset. Simpler models and complex ones like Neural Networks and Support Vector Regression lag in effectiveness, as evidenced by lower R^2 and higher error values. The Bagging Regressor has the highest R^2 value of 0.8295, indicating it can explain approximately 82.95% of the dataset's variance. The Random Forest, an ensemble method known for constructing multiple decision trees, follows closely, with an R^2 of 0.8268. The Gradient Boosting model, which optimizes differentiable loss functions forward stage-wise, shows a marginal decline in performance compared to the Random Forest. The Extra Trees Regressor, which chooses random splits for decision trees, shows lower efficacy in the given metrics compared to the top three models. AdaBoost, a model combining multiple weak learners, shows a noticeable dip in performance. The Decision Tree, a simpler model relying on a tree-like decision schema, presents moderate**.**

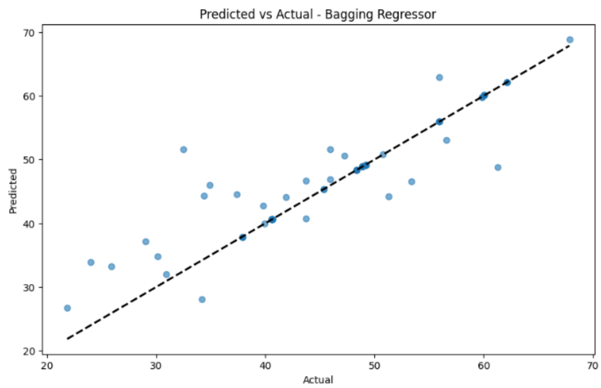

Figure 7: Plot for Bagging Regressor: predicted against actual.

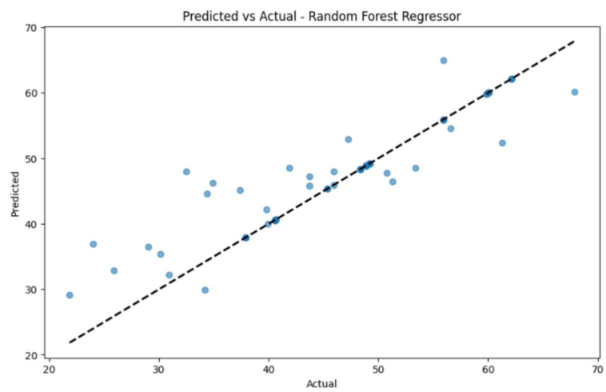

Figure 8: Plot for the Random Forest Regressor: predicted against actual.

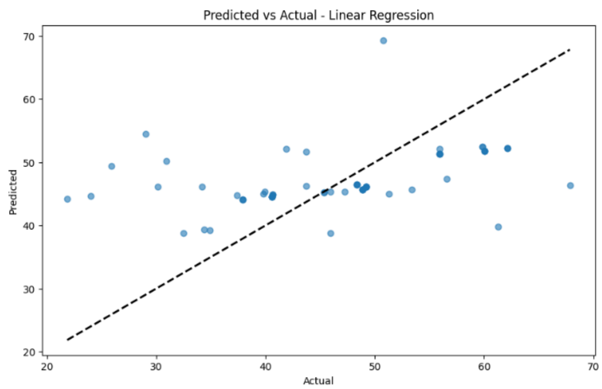

Figure 9: Plot of the linear regression: predicted against actual.

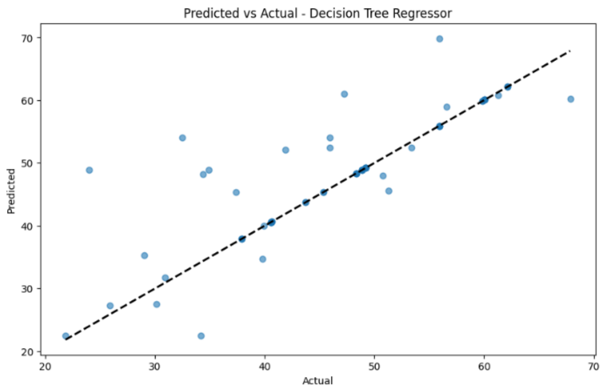

Figure 10: Decision Tree Regressor: Plot of predicted against actual data.

4. LIMITATIONS

This research dataset contains both concrete and other cementitious materials (such as mortar); adding cement and other related materials (such as fine/coarse aggregate) as variables could significantly affect the correlation among other features and potentially reduce the predictive model's accuracy. It is important to carefully consider the intricate interactions between the cement concentration and other compositional variables in concrete.

Moreover, most of the previous research includes a range of different cement kinds, which further complicates the ability to accurately anticipate outcomes. For example, the use of various types of cement, such as CEM II or CEM III, could affect the overall strength properties of the concrete, resulting in difficulties in accurately predicting outcomes from the model. The different features and performance traits linked to various types of cement may result in substantial disparities in the strength of concrete, which presents difficulty for the model to reliably forecast outcomes in such heterogeneous circumstances. Furthermore, this research offers significant insights into the forecast of compressive strength in graphene-reinforced concrete using machine learning. Additionally, it emphasizes the need for additional research in this important field.

5. CONCLUSIONS

This research project revealed valuable findings, aiming to predict the compressive strength of high-strength graphene-reinforced concrete by the application of several machine learning techniques. The study included many models, such as Decision Tree, Random Forest, Lasso, Ridge, and Linear regression. Their predictive performance was evaluated using criteria such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 Score. The Bagging Regressor model exhibited the highest performance in the experiment. It demonstrated superior accuracy in all statistical measures, showcasing its robustness in handling the intricacies of the dataset. The model's ensemble method, employing multiple decision trees, effectively captured the intricate interplay of variables such as concrete age, graphene quantity, and other elements that impact compressive strength. An important challenge in this research is the limited amount of data available for analysis. Graphene, being a novel material for reinforcing concrete, still lacks comprehensive research in many environments and compositions. This limitation impacts the scope and comprehensiveness of the training for machine learning models, thereby affecting the anticipated precision and applicability of the models.

Additionally, a comprehensive and diverse dataset is necessary to account for the intricate connections between graphene and concrete, which are influenced by multiple variables, including curing time, environmental factors, and graphene quality. Due to the complexity of these relationships, it is imperative to gather more data to accurately represent the wide range of possible outcomes and scenarios. As more comprehensive data is collected in the future, there will be a significant opportunity to enhance the predictive accuracy of models. Machine learning models can be improved and updated with the availability of more experimental data and the advancement of research on graphene-reinforced concrete. An improved dataset would facilitate a more thorough understanding of the material characteristics and lead to more accurate and reliable prediction models.

Further investigation could explore advanced machine learning techniques, such as deep learning, which have the potential to yield superior outcomes in handling complex, high-dimensional data. Partnerships among data scientists, material scientists, and industry practitioners are crucial to propel these significant advancements.

To summarize, this study establishes a fundamental structure for utilizing machine learning in forecasting the compressive strength of graphene-reinforced concrete. However, it also opens up avenues for additional, more comprehensive exploration. The predicted increase in data availability and developments in analytical methods make it feasible to develop precise and reliable prediction models. These advancements would have a tangible effect on enhancing construction methods and material innovation, as well as furthering scholarly understanding of graphene-reinforced concrete.

REFERENCES

Abiodun, O., Kabubo, C., Mutuku, R., & Ejohwomu, O. (2023). The Effect of Pristine Graphene on the Mechanical Properties of Geopolymer Mortar. *Sustainability*, *15*(2), 1706.

Almahameed, B. A. A., & Sobuz, H. R. (2023). The Role of Hybrid Machine Learning for Predicting Strength Behavior of Sustainable Concrete. *Civil Engineering and Architecture*, *11*(4), 2012-2032. https://doi.org/10.13189/cea.2023.110425

Alshammari, T., Ramadan, R. A., & Ahmad, A. (2023). Temporal Variations Dataset for Indoor Environmental Parameters in Northern Saudi Arabia. *Applied Sciences*, *13*(12), 7326.

Bakhshandeh, R., & Shafiekhani, A. (2018). Ultrasonic waves and temperature effects on graphene structure fabricated by electrochemical exfoliation method. *Materials Chemistry and Physics*, *212*, 95-102. https://doi.org/https://doi.org/10.1016/j.matchemphys.2018.03.004

Behnood, A., & Golafshani, E. M. (2020). Machine learning study of the mechanical properties of concretes containing waste foundry sand. *Construction and Building Materials*, *243*, 118152. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2020.118152

Candelaria, M. D. E., Chua, N. M. M., & Kee, S.-H. (2022). Evaluation of Heat-Induced Damage in Concrete Using Machine Learning of Ultrasonic Pulse Waves. *Materials*, *15*(22), 7914.

Cao, M.-l., Zhang, H.-x., & Zhang, C. (2016). Effect of graphene on mechanical properties of cement mortars. *Journal of Central South University*, *23*(4), 919-925.

Chaya. (2022). Random Forest Regression. *Medium*. https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

Chetty, K., Watson, M., Raine, T., McGurgan, T., Ladislaus, P., Chen, J., Zhang, S., Lin, L., & Jiang, G. (2022). Enhancing Concrete and Mortar Properties and Durability Using Pristine Graphene Particles. *Coatings*, *12*(11), 1703. https://www.mdpi.com/2079-6412/12/11/1703

Conduit, G. J., Forsdyke, J. C., Lees, J. M., & Zviazhynski, B. (2023). Probabilistic selection and design of concrete using machine learning. *Data-Centric Engineering*, *4*, e9, Article e9. https://doi.org/10.1017/dce.2023.5

Cunningham, L. S., Gallardo, A. F. S., Foster, A. S. J., McDermott, A., Hibberd, R., Scullion, L., Dawson, C., Ijije, H., & Withers, P. Practical application of graphene-enhanced concrete. *Proceedings of the Institution of Civil Engineers - Construction Materials*, *0*(0), 1-5. https://doi.org/10.1680/jcoma.22.00068

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*(4), 303-314. https://doi.org/10.1007/BF02551274

Dalal, S. P., & Dalal, P. (2021). Experimental Investigation on Strength and Durability of Graphene Nanoengineered Concrete. *Construction and Building Materials*, *276*, 122236. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2020.122236

Davawala, M., Joshi, T., & Shah, M. (2023). Compressive strength prediction of high-strength concrete using machine learning. *Emergent Materials*, *6*(1), 321-335. https://doi.org/10.1007/s42247-022-00409-4

Decision Tree Regression. https://saedsayad.com/decision_tree_reg.htm

Dong, W., Huang, Y., Lehane, B., & Ma, G. (2022). Multi-objective design optimization for graphite-based nanomaterials reinforced cementitious composites: A data-driven method with machine learning and NSGA-II. *Construction and Building Materials*, *331*, 127198. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2022.127198

Geim, A. K., & Novoselov, K. S. (2007). The rise of graphene. *Nature Materials*, *6*(3), 183-191. https://doi.org/10.1038/nmat1849

Habibur Rahman Sobuz, M., Khan, M. H., Kawsarul Islam Kabbo, M., Alhamami, A. H., Aditto, F. S., Saziduzzaman Sajib, M., Johnson Alengaram, U., Mansour, W., Hasan, N. M. S., Datta, S. D., & Alam, A. (2024). Assessment of mechanical properties with machine learning modeling and durability, and microstructural characteristics of a biochar-cement mortar composite. *Construction and Building Materials*, *411*, 134281. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2023.134281

Hilt, D. E., Northeastern Forest Experiment, S., Seegrist, D. W., & United, S. (1977). *Ridge, a computer program for calculating ridge regression estimates*. Dept. of Agriculture, Forest Service, Northeastern Forest Experiment Station. http://www.biodiversitylibrary.org/bibliography/68934

Iqbal, M. F., Javed, M. F., Rauf, M., Azim, I., Ashraf, M., Yang, J., & Liu, Q.-f. (2021). Sustainable utilization of foundry waste: Forecasting mechanical properties of foundry sand based concrete using multi-expression programming. *Science of The Total Environment*, *780*, 146524. https://doi.org/https://doi.org/10.1016/j.scitotenv.2021.146524

Jiang, Z., Sevim, O., & Ozbulut, O. E. (2021). Mechanical properties of graphene nanoplatelets-reinforced concrete prepared with different dispersion techniques. *Construction and Building Materials*, *303*, 124472.

Linear Regression Formula – Definition, Formula Plotting, Properties, Uses and Solved Questions. *VEDANTU*. https://www.vedantu.com/formula/linear-regression-formula

Liu, Q., Xu, Q., Yu, Q., Gao, R., & Tong, T. (2016). Experimental investigation on mechanical and piezoresistive properties of cementitious materials containing graphene and graphene oxide nanoplatelets. *Construction and Building Materials*, *127*, 565-576.

Liu, Y. (2022). High-Performance Concrete Strength Prediction Based on Machine Learning. *Computational Intelligence and Neuroscience*, *2022*, 5802217. https://doi.org/10.1155/2022/5802217

Lu, Z., Hou, D., Meng, L., Sun, G., Lu, C., & Li, Z. (2015). Mechanism of cement paste reinforced by graphene oxide/carbon nanotubes composites with enhanced mechanical properties. *RSC advances*, *5*(122), 100598-100605.

M, P. (2021). A Comprehensive Introduction to Evaluating Regression Models. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2021/10/evaluation-metric-for-regression-models/

Monteiro, H., Moura, B., & Soares, N. (2022). Advancements in nano-enabled cement and concrete: Innovative properties and environmental implications. *Journal of Building Engineering*, *56*, 104736. https://doi.org/https://doi.org/10.1016/j.jobe.2022.104736

Mostafa, S. A., Tayeh, B. A., & Almeshal, I. (2022). Investigation the properties of sustainable ultra-high-performance basalt fibre self-compacting concrete incorporating nano agricultural waste under normal and elevated temperatures. *Case Studies in Construction Materials*, *17*, e01453. https://doi.org/https://doi.org/10.1016/j.cscm.2022.e01453

N. Moriasi, D., G. Arnold, J., W. Van Liew, M., L. Bingner, R., D. Harmel, R., & L. Veith, T. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, *50*(3), 885-900. https://doi.org/https://doi.org/10.13031/2013.23153

Nguyen, N. T., Nguyen, T. T. T., & Tran, A. D. (2022). Assessing the Spatial Variability of Compressive Strength of In-Situ Concrete: From Laboratory to Construction Site. *Strength of Materials*, *54*(5), 918-928. https://doi.org/10.1007/s11223-022-00468-2

Orthey, R., Vrij, A., Meijer, E., Leal, S., & Blank, H. (2019). Eliciting response bias within forced choice tests to detect random responders. *Scientific Reports*, *9*(1), 8724.

Piryonesi, S. M., & El-Diraby, T. E. (2020). Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index. *Journal of Infrastructure Systems*, *26*(1), 04019036. https://doi.org/doi:10.1061/(ASCE)IS.1943-555X.0000512

Rhee, I., Lee, J. S., Kim, Y. A., Kim, J. H., & Kim, J. H. (2016). Electrically conductive cement mortar: Incorporating rice husk-derived high-surface-area graphene. *Construction and Building Materials*, *125*, 632-642.

Sadak, O., Sundramoorthy, A. K., & Gunasekaran, S. (2018). Facile and green synthesis of highly conducting graphene paper. *Carbon*, *138*, 108-117. https://doi.org/https://doi.org/10.1016/j.carbon.2018.05.076

Shahab, Z., Anwar, W., Alyami, M., Hammad, A. W., Alabduljabbar, H., Nawaz, R., & Javed, M. F. (2024). Experimental investigation and predictive modeling of compressive strength and electrical resistivity of graphene nanoplatelets modified concrete. *Materials Today Communications*, *38*, 107639.

sklearn.ensemble.BaggingRegressor. *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.ensemble.BaggingRegressor.html

sklearn.ensemble.ExtraTreesRegressor. *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.ensemble.ExtraTreesRegressor.html

Stojkovic, V., & Louro, P. (2022). Simulation of plasmonic effects of Metal (Au, Ag and Al) NPs and rGO embedded in aqueous solutions. *i-ETC: ISEL Academic Journal of Electronics Telecommunications and Computers*, *8*(1).

Team, G. L. (2023). A Complete understanding of LASSO Regression. *Great Learning Blog: Free Resources what Matters to shape your Career!* https://www.mygreatlearning.com/blog/understanding-of-lasso-regression/

Teixeira-Pinto, A. *2 K-nearest Neighbours Regression | Machine Learning for Biostatistics*. https://bookdown.org/tpinto_home/Regression-and-Classification/k-nearest-neighbours-regression.html

Thomas Czigler, S. R., Patrick Schulze and Ken Somers. (2020). *Laying the foundation for Zero-Carbon Cement*. McKinsey & Company. https://www.mckinsey.com/industries/chemicals/our-insights/laying-the-foundation-for-zero-carbon-cement

vpadmin. (2023). A Comprehensive Guide to Database Normalization with Examples. *Visual Paradigm Guides*. https://guides.visual-paradigm.com/a-comprehensive-guide-to-database-normalization-with-examples/

Water–cement ratio. (2023). In *Wikipedia*.

Yan, K., & Shi, C. (2010). Prediction of elastic modulus of normal and high strength concrete by support vector machine. *Construction and Building Materials*, *24*(8), 1479-1485. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2010.01.006

Zaid, O., Abdulwahid Hamah Sor, N., Martínez-García, R., de Prado-Gil, J., Mohamed Elhadi, K., & Yosri, A. M. (2024). Sustainability evaluation, engineering properties and challenges relevant to geopolymer concrete modified with different nanomaterials: A systematic review. *Ain Shams Engineering Journal*, *15*(2), 102373. https://doi.org/https://doi.org/10.1016/j.asej.2023.102373

Zaid, O., Hashmi, S. R. Z., Aslam, F., Abedin, Z. U., & Ullah, A. (2022). Experimental study on the properties improvement of hybrid graphene oxide fiber-reinforced composite concrete. *Diamond and Related Materials*, *124*, 108883.

Zhao, Y., Liu, Y., Shi, T., Gu, Y., Zheng, B., Zhang, K., Xu, J., Fu, Y., & Shi, S. (2020). Study of mechanical properties and early-stage deformation properties of graphene-modified cement-based materials. *Construction and Building Materials*, *257*, 119498.