Predicting Concrete Compressive Strength using Machine Learning

Osama Ibrahim
Department of Civil and Environmental Engineering
University of Illinois Urbana-Champaign
Urbana, IL, USA
osamani2@illinois.edu

Praneeth Shivashankarappa
Department of Civil and Environmental Engineering
University of Illinois Urbana-Champaign
Urbana, IL, USA
ps104@illinois.edu

Kazi Ishat Mushfiq
Department of Civil and Environmental Engineering
University of Illinois Urbana-Champaign
Urbana, IL, USA
mushfiq2@illinois.edu

Georg Bauer
Department of Civil and Environmental Engineering
University of Illinois Urbana-Champaign
Urbana, IL, USA
georgb2@illinois.edu

***Abstract*—This project investigates the prediction of concrete compressive strength using machine learning models. We use the Concrete Compressive Strength dataset from the UCI Machine Learning Repository, which contains 1030 experimental data points with eight input variables (cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate, and curing age) and one output variable (compressive strength in MPa). Our study compares linear regression, decision tree regression, and neural networks to determine the most effective approach for modeling the nonlinear relationships that influence the development of concrete strength.**

***Index Terms*—Concrete Compressive Strength, Machine Learning, Civil Engineering, Regression Models**

I. Introduction

Concrete compressive strength is one of the most important performance indicators in civil engineering, because it determines whether a mix design will satisfy project requirements for safety and durability. Strength testing is widely performed using laboratory cylinders, yet experimental testing requires time, materials, and controlled curing conditions, and this creates a practical need for fast predictive tools that can estimate strength based on mix proportions. Traditional empirical relationships capture only part of the strength behavior, because concrete is influenced by nonlinear interactions among cementitious materials, water content, and curing age. Machine learning offers a flexible framework to explore these nonlinearities and has been applied in several studies to predict strength more accurately across a wide range of mix designs.

The objective of this project is to evaluate whether machine learning models can reliably predict concrete compressive strength when given common mix design variables. The research question guiding the analysis is the following: Can modern predictive models, particularly nonlinear learners, capture the complex effects of mixture ingredients and curing age better than a linear regression benchmark? Our hypothesis is that nonlinear models, including Decision Trees (DT), Random Forests (RF), and Neural Networks (NN), will outperform linear regression because the dataset exhibits strong nonlinear patterns.

To address this question, we use the Concrete Compressive Strength dataset from the UCI Machine Learning Repository [1], [2]. The dataset contains 1030 samples, each representing a concrete mix tested under laboratory conditions. The input variables include cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate, and testing age. All quantities are expressed per cubic meter. The output variable is the measured compressive strength in megapascals (MPa). A summary of the variables is provided below.

Variables included in the dataset:

- Cement (kg/m^3)
- Blast Furnace Slag (kg/m^3)
- Fly Ash (kg/m^3)
- Water (kg/m^3)
- Superplasticizer (kg/m^3)
- Coarse Aggregate (kg/m^3)
- Fine Aggregate (kg/m^3)
- Age (days, from 1–365)

Target variable:

- Concrete Compressive Strength (MPa)

This dataset is widely used as a benchmark in the concrete materials community because it contains substantial variability in mix proportions and curing conditions. The absence of missing values and the broad distribution of ingredients make it suitable for both exploratory analysis and predictive modeling. The goal of this report is to connect statistical observations from the exploratory data analysis to the modeling decisions made in later sections, creating a complete narrative from data characteristics to final predictive performance.

The remainder of this report is organized as follows. The next section presents an exploratory analysis that describes the structure and patterns in the dataset. The predictive modeling section then evaluates a range of linear and nonlinear models and compares their performance. The discussion section interprets the modeling results in the context of the original research question, and the report concludes with a summary of findings and potential next steps for future work.

II. Exploratory Data Analysis

Understanding the characteristics of the dataset is an essential step before building predictive models. The goal of this analysis is to examine the distributions, correlations, and patterns present in the Concrete Compressive Strength dataset, and to identify the variables that most strongly influence compressive strength. These observations guide the modeling decisions in the next section and help explain why certain algorithms perform better than others.

A. Dataset Overview

The dataset contains 1030 samples with eight input features and one continuous output variable. The source and structure of the data were introduced in the Introduction, and Table I shows representative rows that illustrate the format of the mix design variables and strength measurements. The dataset includes a wide range of ingredient quantities and curing ages, which supports the development of predictive models that generalize beyond narrow mix proportions.

TABLE I
Representative Entries from the Concrete Compressive Strength Dataset

Row	Cement	Slag	Fly-Ash	Water	SuperP.	CA	FA	Age	Strength
1	540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28	79.99
2	540.0	0.0	0.0	162.0	2.5	1055.0	676.0	28	61.89
3	332.5	142.5	0.0	228.0	0.0	932.0	594.0	270	40.27
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
1028	148.5	139.4	108.6	192.7	6.1	892.4	780.0	28	23.70
1029	159.1	186.7	0.0	175.6	11.3	989.6	788.9	28	32.77
1030	260.9	100.5	78.3	200.6	8.6	864.5	761.5	28	32.40

B. Summary Statistics

Descriptive statistics for all variables are presented in Table II. These values summarize the minimum, maximum, mean, median, and standard deviation for each ingredient and for curing age. The data display substantial variability, for example, cement content ranges from 102 to 540 kg/m³, and age ranges from 1 to 365 days. This variation indicates that the dataset is sufficiently broad for identifying nonlinear patterns that relate mixture proportions to compressive strength.

The presence of wide standard deviations in several variables, such as slag and fly ash, reflects the heterogeneity in the mix designs. This variety strengthens the dataset for model training because it avoids over-representation of narrow patterns and reduces the risk of overfitting to limited scenarios. The summary statistics form the basis for diagnosing potential issues in predictive modeling such as skewed distributions, collinearity, and zero-inflated features.

TABLE II
Summary statistics of concrete mix features

Feature	Min	Max	Mean	Median	Std-Dev
Cement (kg/m³)	102.0	540.0	281.17	272.9	104.51
Blast Furnace Slag (kg/m³)	0.0	359.4	73.9	22.0	86.28
Fly Ash (kg/m³)	0.0	200.1	54.19	0.0	64.0
Water (kg/m³)	121.75	247.0	181.57	185.0	21.36
Superplasticizer (kg/m³)	0.0	32.2	6.2	6.35	5.97
Coarse Aggregate (kg/m³)	801.0	1145.0	972.92	968.0	77.75
Fine Aggregate (kg/m³)	594.0	992.6	773.58	779.51	80.18
Age (day)	1	365	45.66	28.0	63.17

C. Data Visualizations

Visual analysis helps reveal patterns that guide feature selection and model development. The following subsections describe histograms, scatter plots, correlation analysis, boxplots, and a Principal Component Analysis (PCA) that summarizes the main sources of variance in the dataset.

Histograms of the input features are shown in Fig. 1. They display the distribution of ingredient quantities and age, and several characteristic patterns emerge. Cement content is concentrated between 150 and 350 kg/m³ with fewer mixes above 400 kg/m³. Blast furnace slag, fly ash, and superplasticizer show strong spikes at zero, which indicates zero-inflated distributions. Most mixes therefore omit these supplementary materials. Water content follows a roughly symmetric distribution centered between 170 and 200 kg/m³. Coarse and fine aggregates show multimodal distributions that likely reflect different design strategies and target strengths. The testing age histogram has a strong peak at early ages, especially at 1–7 days and at 28 days, which is the standard reporting age.

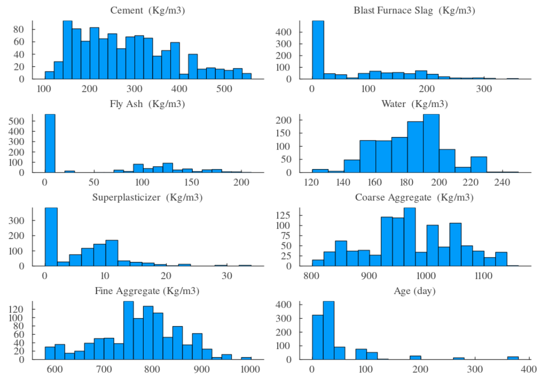

Fig. 1. Histograms of input features

Scatter plots showing the relationship between each ingredient and compressive strength are presented in Fig. 2. These plots provide intuitive evidence of nonlinear effects. Cement content shows a generally increasing trend, although strength may decrease at very high cement dosages due to shrinkage and poor workability. Slag shows no consistent pattern because its effect depends on curing age and replacement percentage. Fly ash is absent in many samples, but mixes containing 50–150 kg/m³ show variable performance.

A clear negative trend appears between water content and strength. Increasing water improves workability but increases porosity in the hardened matrix, which reduces strength. Superplasticizer content is mostly low, and when used in moderate amounts, it improves flowability and supports higher strengths through reduced water requirements. Aggregate contents show weak correlation with compressive strength, which reflects their more indirect role in strength formation. Age displays a strong positive relationship, especially up to about 90 days, after which strength gains become slower.

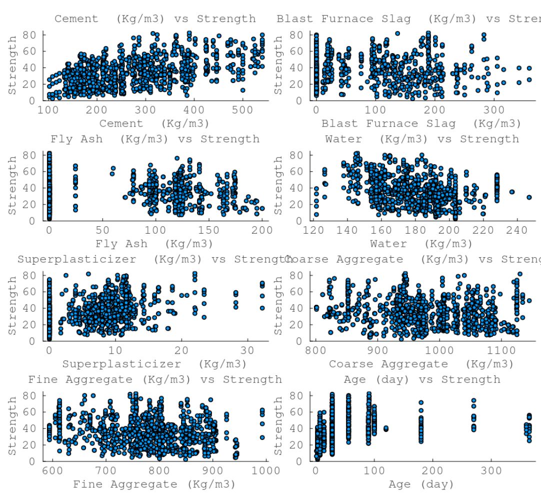

Fig. 2. Scatter plots of input features vs compressive strength

The correlation heatmap in Fig. 3 summarizes linear relationships between variables. Compressive strength is positively correlated with both cement content and age, and negatively correlated with water content [3]. Superplasticizer shows a small positive correlation with strength. Aggregate variables show weaker correlations, while coarse and fine aggregates are negatively correlated with each other due to volumetric balance. Slag and fly ash are negatively correlated with cement, which reflects their role as partial cement replacements in sustainable concrete designs.

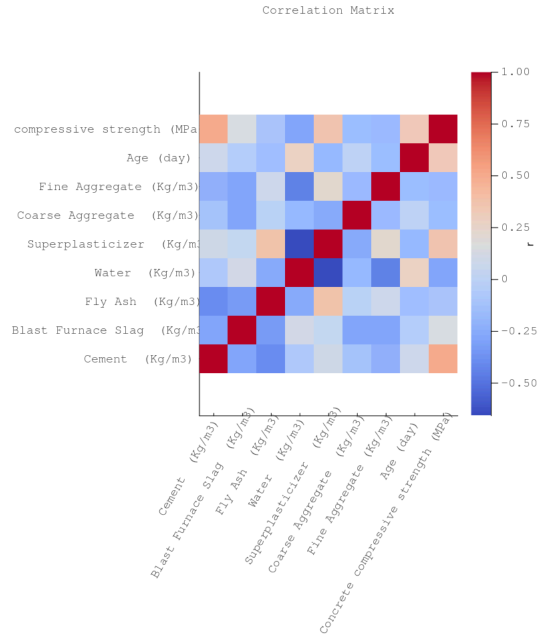

Fig. 3. Correlation heatmap

Boxplots of the input features are shown in Fig. 4. Most variables show moderate variation with a few outliers, which is expected for experimental materials data. Age has notable long-term values up to 365 days, but most data points cluster around 28 days. These visualizations highlight where extreme values may influence predictive models, especially tree-based methods that can partition extreme values more easily than linear models.

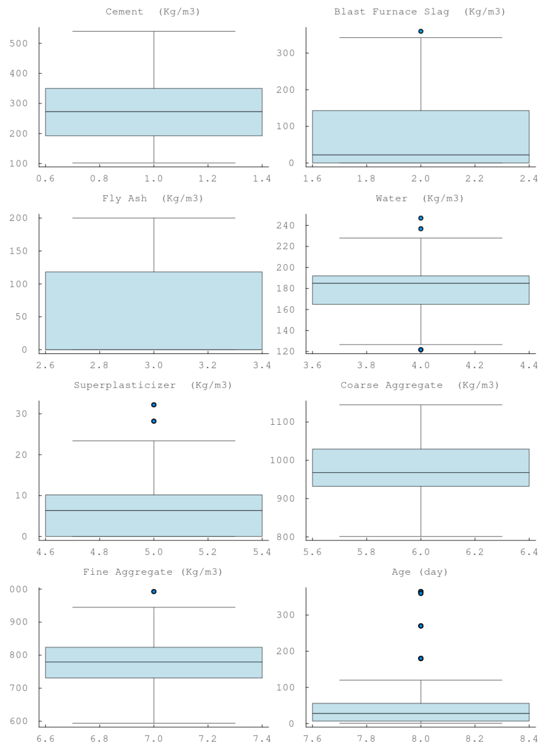

Fig. 4. Boxplots of input features

Finally, Principal Component Analysis (PCA) summarizes how groups of variables jointly contribute to variance. Fig. 5 shows that the first principal component (PC1) captures 28.5 percent of the total variance. The loading plot identifies which variables influence each principal component. PC2 is primarily shaped by blast furnace slag, water, coarse aggregate, and age, all with negative contributions, indicating that increases in these features reduce the value of PC2. PCA supports the observations from earlier plots by showing that ingredient proportions interact in complex ways.

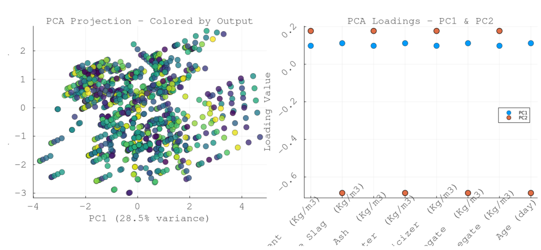

Fig. 5. PCA plot of input features

This exploratory analysis highlights several important findings that directly motivate the predictive modeling approaches used in the next section. The strong nonlinear relationships between cement, water, age, and compressive strength suggest that linear regression alone may not capture the full behavior of the dataset. Zero-inflated variables, such as slag and fly ash, may reduce performance for models sensitive to sparse distributions. These insights influenced the modeling strategy and explain why nonlinear models such as Decision Trees, Random Forests, and Neural Networks were included in the analysis that follows.

III. Predictive Modeling

The exploratory analysis demonstrated that concrete compressive strength depends on nonlinear interactions between mixture ingredients and curing age. These observations motivate the use of both linear and nonlinear models to determine how well each can capture the patterns in the data. The predictive modeling phase builds on the findings from the previous section by evaluating multiple approaches in parallel. Linear Regression establishes a baseline, while Decision Tree (DT), Random Forest (RF), and Neural Network (NN) models are used to capture nonlinear effects.

This section summarizes the modeling process, hyperparameter exploration, and performance evaluation for each model. The goal is to determine whether nonlinear methods provide a meaningful improvement over linear regression and, ultimately, whether machine learning can predict compressive strength with accuracy suitable for engineering applications.

A. *Research Question and Hypothesis*

Based on the exploratory analysis, we hypothesize that machine learning models can accurately predict concrete compressive strength and that nonlinear models will outperform linear regression. This is expected because the dataset includes several nonlinear relationships, such as the effect of water content on porosity and the asymptotic increase in strength with curing age.

B. *Methods*

a) *Data Preprocessing:*

The dataset contains 1030 samples with eight primary input features. Several engineered features were added to enhance model performance. The preprocessing steps were: , Dataset: 1030 samples with 11 features (8 original + 3 engineered) , Train-Test Split: 80 percent training (824 samples), 20 percent testing (206 samples) , Standardization: Applied to models sensitive to feature scaling (linear regression variants and neural network) , Duplicate Removal: All repeated rows removed , Feature Engineering: Created Water-Cement Ratio, Total Binder, and Aggregate-Binder Ratio

These engineered features incorporate domain knowledge from concrete materials science and help the models capture mix design interactions more effectively.

b) *Model Specifications:*

Four models were trained and evaluated:

1) **Linear Regression (LR)**: Serves as a baseline for comparison.
2) **Decision Tree (DT)**: Captures nonlinear relationships through hierarchical splits.
3) **Random Forest (RF)**: Uses an ensemble of decision trees to improve stability.
4) **Neural Network (NN)**: A two-layer feed-forward model using ReLU activation.

Performance was assessed using R^2, RMSE, and MAE. Cross-validation was used for LR, DT, and RF [4]. The NN was trained with full-batch gradient descent, so cross-validation was not applied.

IV. Model 1 – Linear Regression

This section presents the development and evaluation of the Linear Regression models implemented in Julia using gradient descent optimization. Several preprocessing strategies and feature-engineering techniques were compared.

A. Basic Linear Model

The first model used raw, unstandardized features.
Equation:

$$\hat{y} = x\beta \tag{1}$$

Performance metrics are summarized in Table III, and Fig. 6 shows the relationship between predicted and actual strengths.

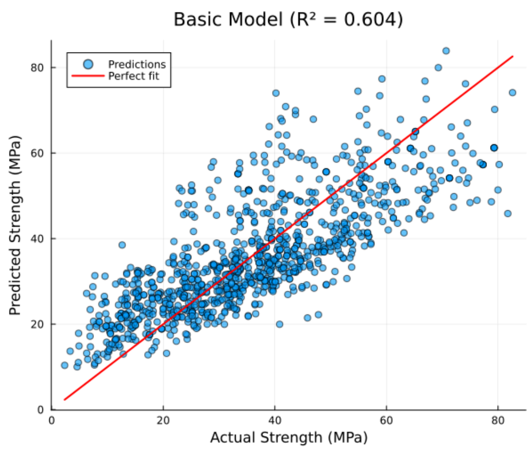

Fig. 6. Basic linear regression model performance: predicted vs actual strengths.

TABLE III
Performance metrics of the Basic Linear Model.

Metric	Value
R^2	0.6108
MSE	108.5376

B. Standardized Linear Model

The input and output variables were standardized before training, which improved model stability and convergence. The resulting performance is shown in Fig. 7 and Table IV.

Fig. 7. Standardized linear regression model performance.

TABLE IV
Performance metrics of the Standardized Model.

Metric	Value
R^2	0.6144
MSE	0.3852

C. K-fold Cross Validation

Model robustness was evaluated with 4-fold blocked cross validation. Each fold was used once as a test set, and R^2 values were averaged. Table V summarizes the fold-wise results.

TABLE V
Performance metrics of the Standardized Model under K-Fold Cross Validation.

Fold	R^2 Score
1	0.4587
2	0.4432
3	0.4791
4	0.4085

The mean and standard deviation across the folds are summarized in Table VI.

TABLE VI
Mean and Standard Deviation of 4-Fold Cross Validation.

Metric	Value
Mean R^2	0.4474
Std R^2	0.0298

D. Regularized Model (L1 – LASSO)

To improve sparsity and avoid overfitting, an L1 penalty was added:

$J(\beta) = \text{MSE}(\hat{y}, y) + \lambda\|\beta\|_1$

The regularization parameter lambda was tuned using grid search. The resulting coefficient magnitudes are shown in Fig. 8, and performance metrics are summarized in Table VII.

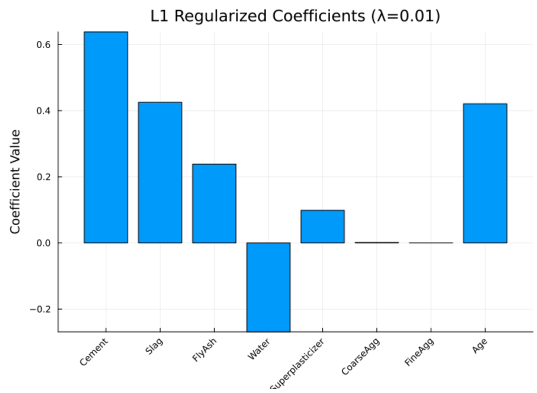

Fig. 8. L1-regularized coefficient magnitudes.

TABLE VII
Performance metrics of the L1 Regularized Model.

Metric	Value
R^2	0.6136
MSE	0.3863

E. *Feature Correlation and Collinearity*

The correlation matrix between input variables and the target strength was computed to assess collinearity effects. The heatmap in Fig. 9 shows the pairwise correlations.

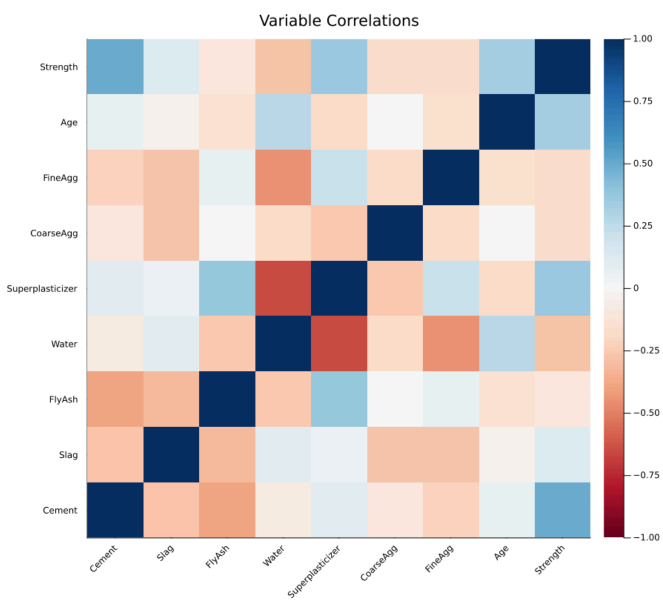

Fig. 9. Correlation heatmap showing relationships among all variables.

Highly collinear features (|r| > 0.8) were removed to create a simplified model with improved interpretability. The performance of this model is shown in Fig. 10.

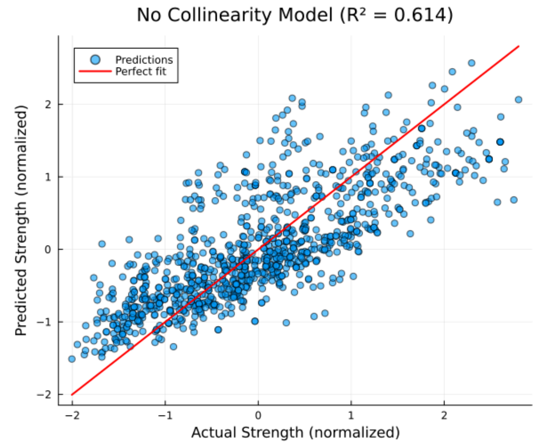

Fig. 10. Linear regression model after removing collinear features.

F. *PCA-Based Model*

Principal Component Analysis (PCA) was applied to reduce dimensionality. The model was trained on the first eight principal components, explaining over 95% of the total variance. Fig. 11 shows the performance of this PCA-based regression model.

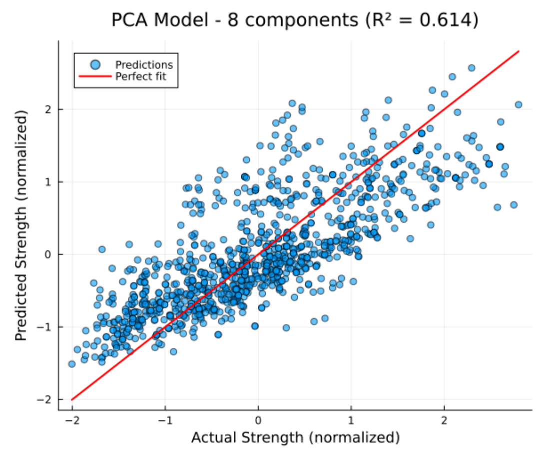

Fig. 11. PCA-based regression model (8 components).

G. *Engineered Feature Model*

Feature engineering was performed to capture meaningful ratios and combined terms such as Water/Cement ratio, Total Binder, and Binder/Aggregate ratio. This model achieved the highest performance as shown in Fig. 12.

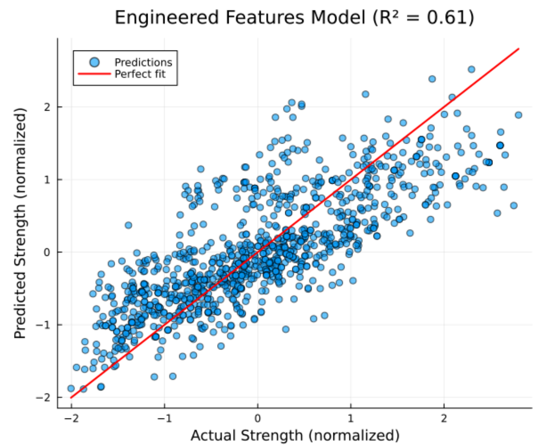

Fig. 12. Engineered feature regression model performance.

H. Comparative Performance

A comparison of the R^2 values for all models is shown in Fig. 13. The engineered-feature model yielded the best predictive accuracy, indicating that domain-informed features significantly improved generalization.

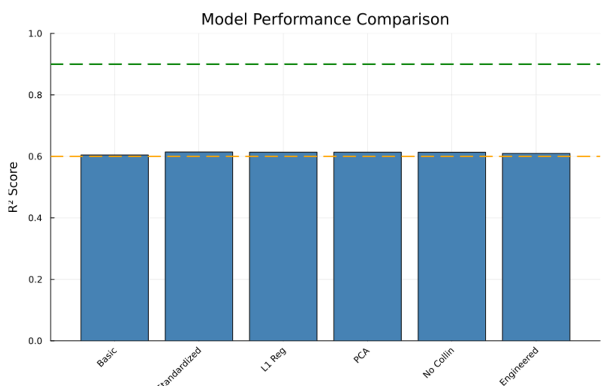

Fig. 13. Comparison of R^2 scores for all linear regression model variants.

V. Model 2 - Decision Tree

In this section, we developed a decision tree regression model to predict concrete compressive strength using eight mix design varibles and several engineered features. The dataset was standardized and randomly divided into 80% training set and 20% test set.

A. Untuned Model

At first we used un-tuned decision tree model as a baseline model to predict the compressive strength. This model was trained pruned to prevent overfitting. The untuned model provided R^2 score of 0.59 and 0.99 for train and test data respectively. Fig. 14 and Table VIII provide a comparison of the test and train R^2 score.

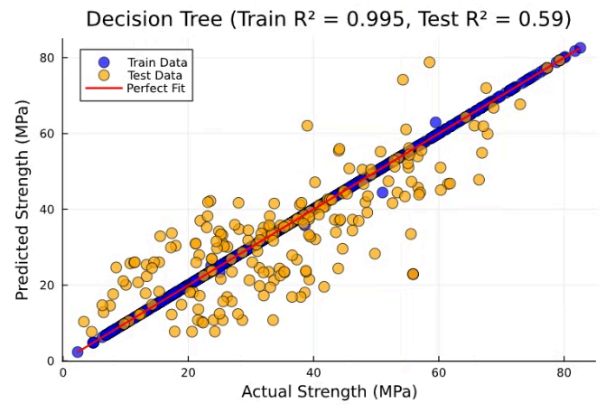

Fig. 14. Comparison of R^2 scores for test and train model using decision tree.

TABLE VIII
Performance metrics of the Baseline Decision Tree Model.

Metric	Value
Model (Train)	"Decision Tree"
R^2 (Train)	0.995071
RMSE (Train)	1.18517
MAE (Train)	0.0595223
Model (Test)	"Decision Tree"
R^2 (Test)	0.590449
RMSE (Test)	10.194
MAE (Test)	7.75932

B. Tuned Model

To improve the model performance,we performed comprehensive hyperparameter tuning. It was performed by systematically searching across multiple configurations of maximum tree depth, minimum samples per leaf, minimum samples required for a split, impurity reduction threshold, and pruning. Then the R^2 score was evaluated for the model to identify the best confituration.

Fig. 15 shows the R^2 score for the tuned model. The R^2 score improved from 0.59 to 0.68 after tuning the model. Table IX provides a comparison of the untuned and tuned model performance.

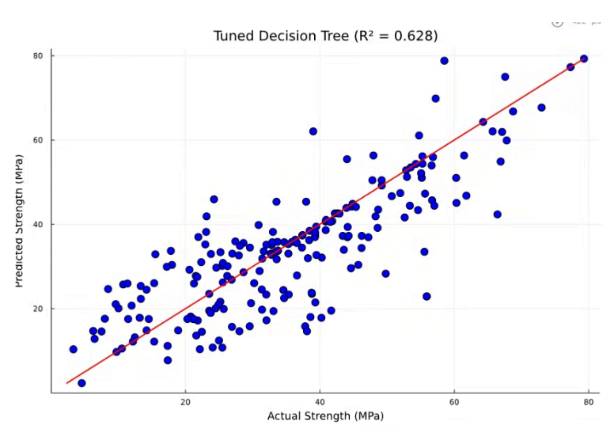

Fig. 15. R^2 score for the tuned model using decision tree.

TABLE IX
Performance comparison of the Untuned and Tuned Decision Tree Models.

Model	R^2	RMSE	MAE
"Untuned Tree"	0.590449	10.194	7.75932
"Tuned Tree"	0.628332	9.71112	7.17482

C. 5 Fold Cross Validation for Decision Tree Models

Both of the tuned and untuned models were validated throught 5 fold cross validation to measure its robustness across different data partitions. The performance of the both tuned and untuned model are shown in Table X and Table XI respectively.

TABLE X
Performance metrics of the Tuned Decision Tree Model under 5-Fold Cross Validation.

Fold (Tuned)	R^2 Score
1	0.631071
2	0.598450
3	0.671774
4	0.390783
5	0.581910
Mean	0.5748

TABLE XI
Performance metrics of the Untuned Decision Tree Model under 5-Fold Cross Validation.

Fold (Untuned)	R^2 Score
1	0.660438
2	0.554768
3	0.337916
4	0.611585
5	0.544826
Mean	0.5419

VI. Model 3 – Random Forest

In this section, we developed a random forest regression model to predict concrete compressive strength using eight mix-design variables and several engineered features. The dataset was standardized and randomly divided into an 80% training set and a 20% test set. It follows the same pathway of optimization as the decision tree model

A. Untuned Model

We first used an untuned random forest model as a baseline to evaluate ensemble performance. This initial model was trained using default hyperparameters without limiting tree depth or adjusting sampling parameters. The untuned model provided an R^2 score of 0.99 for the train set and 0.63 for the test set. Fig. 16 and Table XII show the comparison of the train and test R^2 scores.

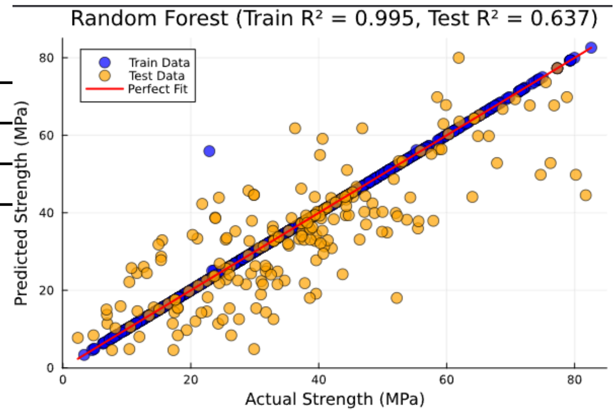

Fig. 16. Comparison of R^2 scores for test and train model using random forest.

TABLE XII
Performance metrics of the Baseline Random Forest Model.

Metric	Value
Model (Train)	"Random Forest"
R^2 (Train)	0.995178
RMSE (Train)	1.1526
MAE (Train)	0.0449522
Model (Test)	"Random Forest"
R^2 (Test)	0.637315
RMSE (Test)	10.2703
MAE (Test)	7.25876

B. Tuned Model

To improve predictive accuracy, we performed hyperparameter tuning on the random forest model. The tuning process explored multiple configurations of the number of estimators, maximum tree depth, bootstrap sampling, minimum samples per leaf, and feature selection strategy. Each configuration was evaluated using its R^2 score to identify the most effective setup.

Fig. 17 illustrates the improvement in performance. After tuning, the test R^2 increased from 0.63 to 0.652. Table XIII provides a comparison of the untuned and tuned model performances.

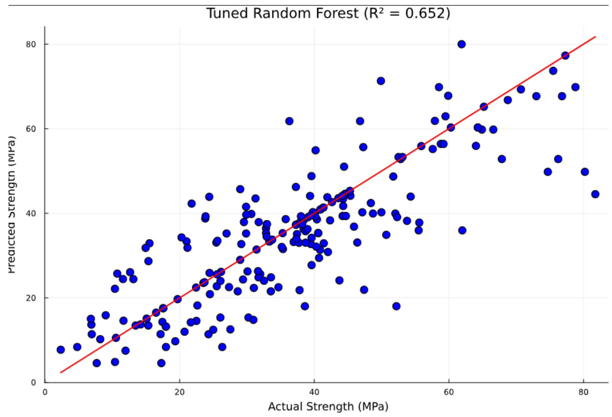

Fig. 17. R^2 score for the tuned model using random forest.

TABLE XIII
PERFORMANCE COMPARISON OF THE UNTUNED AND TUNED RANDOM FOREST MODELS.

Model	R^2	RMSE	MAE
"Untuned RF"	0.637315	10.2703	7.25876
"Tuned RF"	0.652318	10.0557	7.19412

C. 5 Fold Cross Validation for Random Forest Models

Both tuned and untuned random forest models were validated using 5-fold cross-validation to evaluate their robustness across different data splits. The R^2 scores for each fold are shown below in Table XIV and Table XV respectively.

TABLE XIV
PERFORMANCE METRICS OF THE TUNED RANDOM FOREST MODEL UNDER 5-FOLD CROSS VALIDATION.

Fold (Tuned)	R^2 Score
1	0.514840
2	0.505447
3	0.613936
4	0.661547
5	0.434754
Mean	0.5461

TABLE XV
PERFORMANCE METRICS OF THE UNTUNED RANDOM FOREST MODEL UNDER 5-FOLD CROSS VALIDATION.

Fold (Untuned)	R^2 Score
1	0.538777
2	0.535721
3	0.564592
4	0.602024
5	0.430688
Mean	0.5344

VII. MODEL 4 – NEURAL NETWORK

To capture the non-linear relationships present in the concrete mixture data, we developed a feed-forward neural network trained using a custom gradient descent optimizer. The model uses eight standardized input features and produces a single continuous prediction of compressive strength.

The architecture consists of an input layer with eight features, followed by one hidden layer with ten neurons using the ReLU activation function, and finally an output neuron representing the predicted compressive strength. The forward propagation can be described as follows. First, the model computes the linear transformation $z1 = W1 \cdot x + b1$, where W1 and b1 denote the weights and biases of the hidden layer. The activation of this hidden layer is then obtained through a ReLU operation, expressed as $a1 = \max(0, z1)$. The output prediction is calculated through a second linear transformation, $y\_pred = W2 \cdot a1 + b2$.

Training was performed using full-batch gradient descent. We optimized the parameters with a learning rate of $\eta = 0.001$ over a total of 5000 training steps. The loss function used during training was the Mean Squared Error (MSE). Because gradient-based optimization is sensitive to input scaling, all features were standardized prior to training.

After optimization, the neural network achieved strong predictive performance on the held-out test set. The final metrics were: MSE = 43.64, RMSE = 6.61, and MAE = 5.07. The coefficient of determination reached $R^2 = 0.837$, indicating that the model explains approximately 84% of the variance in compressive strength. This performance surpasses the linear model and confirms that non-linear approaches such as neural networks are capable of capturing complex interactions between concrete components that linear regression cannot represent.

To visually assess the predictive quality of the neural network, the true compressive strengths from the test set were plotted against the model's predicted values in Fig. 18.

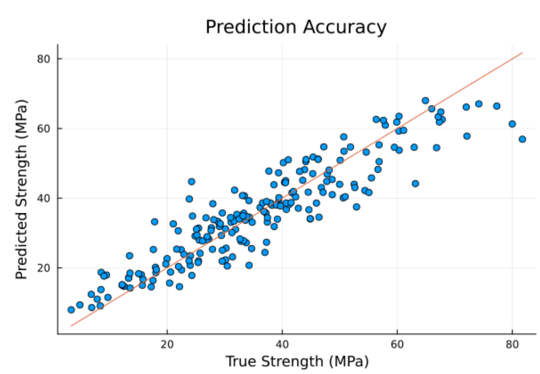

Fig. 18. Model accuracy plot comparing predicted vs. true concrete strengths. Points close to the diagonal indicate high predictive performance.

VIII. Summarized Results

A. Model Performance Comparison

Table XVI summarizes the performance metrics for all four models evaluated in this study. The Neural Network achieved the highest test R^2 of 0.837, followed by the Random Forest at 0.652, Decision Tree at 0.628, and Linear Regression at 0.614. The RMSE and MAE values also reflect this ranking, with the Neural Network showing the lowest error metrics.

TABLE XVI
Comparison of model performance metrics across all evaluated algorithms.

Model	Dataset	R^2	RMSE	MAE
Linear Regression	Train	0.614	—	—
Linear Regression	Test	0.614	—	—
Decision Tree	Train	0.995	1.185	0.060
Decision Tree	Test	0.628	9.711	7.175
Random Forest	Train	0.995	1.153	0.045
Random Forest	Test	0.652	10.056	7.194
Neural Network	Train	0.854	6.210	4.780
Neural Network	Test	0.837	6.610	5.070

B. Cross-Validation Results

a) *Linear Regression (4-Fold Blocked):*
- Fold Scores (R^2): 0.4587, 0.4432, 0.4791, 0.4085
- Mean R^2: 0.447
- Standard Deviation: 0.030

b) *Decision Tree – Tuned (5-Fold):*
- Fold Scores (R^2): 0.6311, 0.5985, 0.6718, 0.3908, 0.5819
- Mean R^2: 0.575
- Standard Deviation: 0.101

c) *Decision Tree – Untuned (5-Fold):*
- Fold Scores (R^2): 0.6604, 0.5548, 0.3379, 0.6116, 0.5448
- Mean R^2: 0.542
- Standard Deviation: 0.109

d) *Random Forest – Tuned (5-Fold):*
- Fold Scores (R^2): 0.5148, 0.5054, 0.6139, 0.6615, 0.4348
- Mean R^2: 0.546
- Standard Deviation: 0.085

e) *Random Forest – Untuned (5-Fold):*
- Fold Scores (R^2): 0.5388, 0.5357, 0.5646, 0.6020, 0.4307
- Mean R^2: 0.534
- Standard Deviation: 0.067

C. Key Findings

1) **Best Performing Model**: Neural Network achieved the highest test accuracy with R^2 = 0.837
2) **Most Influential Features**:
   - Age (days)
   - Aggregate/Binder Ratio
   - Total Binder Content
3) **Engineered Features**: Water-Cement Ratio ranked 5th in importance based on Random Forest feature ranking

D. Visualizations

Fig. 19 and Fig. 20 illustrate the comparative model performances and feature importance rankings, respectively.

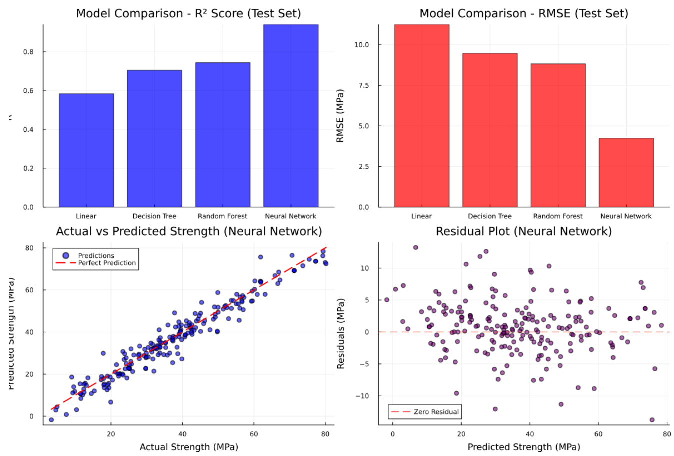

Fig. 19. Model performance comparison and diagnostic plots

Fig. 20. Feature importance analysis from tree-based models

IX. Discussion

The results of this project provide clear evidence that machine learning models can predict concrete compressive strength with meaningful accuracy. The exploratory analysis demonstrated strong nonlinear relationships between cement content, water content, curing age, and the resulting compressive strength. These patterns motivated the use of nonlinear models, and the predictive modeling results validate this decision. The comparison across models confirms the initial hypothesis that nonlinear approaches outperform linear regression when modeling complex material behavior.

The Neural Network achieved the highest predictive performance with an R^2 value of 0.837 on the test set. This indicates that the model explains approximately eighty four percent of the variance in compressive strength and demonstrates that nonlinear models can successfully capture the interaction effects among concrete ingredients. While this level of accuracy is not intended to replace standardized laboratory testing, it is sufficiently strong for early stage mix design screening or for evaluating the expected trends when adjusting mixture proportions.

The Random Forest model performed moderately well, with an R^2 of 0.652 on the test set. This suggests that ensemble tree based models can extract part of the nonlinear structure but may require larger datasets to achieve their full potential. The Decision Tree model showed higher variance and lower predictive stability due to its sensitivity to the specific data splits. Linear regression performed the weakest among the models. Although engineered features improved its accuracy, the model was still unable to fully capture nonlinear behavior in the data.

Several factors help explain the remaining prediction error across the models. The dataset contains zero inflated variables such as fly ash and slag, which make it challenging for some algorithms to learn consistent patterns. The wide range of curing ages also introduces nonlinear strength development behavior that is difficult to approximate using simple functions. Experimental variability in materials and testing conditions likely contributes additional noise that the models cannot capture. These limitations suggest that improvement is possible with larger datasets, more detailed mix design descriptors, or more advanced modeling techniques.

If this project were to continue beyond this semester, several next steps would be valuable. Collecting or incorporating external datasets would allow more robust training and better generalization. Additional engineered features could be included, such as water binder ratios accounting for chemical admixtures or terms representing aggregate gradation. Exploring deeper neural network architectures or advanced ensemble techniques may also increase predictive accuracy. Finally, evaluating model performance using uncertainty quantification would provide more insight into the reliability of predictions in practical engineering settings.

Overall, the findings support the hypothesis that nonlinear machine learning models improve the prediction of concrete compressive strength and demonstrate that these techniques can offer meaningful insight into how mix proportions influence material performance.

X. Conclusion

This project applied a range of machine learning techniques to predict concrete compressive strength using a widely studied dataset of mix design variables. The analysis showed that concrete strength is influenced by several interacting factors, including cement content, water content, and curing age, and these interactions introduce nonlinear behavior that simple linear models cannot fully capture. By examining the dataset through exploratory visualizations and summary statistics, we identified patterns that directly informed the modeling choices.

The predictive modeling results show that nonlinear models provide clear improvements in accuracy. The Neural Network achieved the highest performance, explaining more than eighty percent of the variance in the test data, and demonstrated the capability to model the complex relationships within the dataset. Random Forest models also performed reasonably well, although with higher variance across folds. Linear regression served as a useful baseline but was limited by its linear structure even when enhanced with engineered features.

Taken together, these results indicate that machine learning can play an effective role in preliminary concrete strength estimation. While these tools do not replace laboratory testing, they can support early stage decision making in mix proportioning and can help engineers identify promising designs before committing to physical trials. The findings also highlight the importance of domain informed feature engineering, careful preprocessing, and model selection when applying machine learning to civil engineering materials data.

Future work could focus on expanding the dataset, improving engineered features, exploring deeper neural architectures, or integrating uncertainty quantification to assess prediction confidence. These extensions would help further strengthen the applicability of data driven methods in concrete materials research and engineering practice.

References

[1] D. Dua and C. Graff, "UCI Machine Learning Repository: Concrete Compressive Strength Data Set." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength

[2] I.-C. Yeh, "Modeling of strength of high-performance concrete using artificial neural networks," *Cement and Concrete Research*, vol. 28, no. 12, pp. 1797–1808, 2007, doi: 10.1016/S0008-8846(98)00165-3.

[3] S. Popovics and J. Ujhelyi, "Contribution to the concrete strength versus water-cement ratio relationship," *Journal of Materials in Civil Engineering*, vol. 20, no. 7, pp. 459–463, 2008.

[4] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *Peerj computer science*, vol. 7, p. e623, 2021.