Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Investigation of Multi-Output Regression Modeling in Predicting Concrete Mix Design

AvaniDedhia[1*], Dr. N.K. Arora[2†], Hepil Italiya[3†] and Anuj Shukla[4†]

[1]Research Scholar, Gujarat Technological University, Ahmedabad, 382424, Gujarat, India.
[2] Applied Mechanics Department, GEC Gandhinagar, 382010, Gujarat, India.
[3] Computer EngineeringDepartment, Gujarat Technological University, Gujarat, India.
[4] Information Technology Department, Gujarat Technological University,Ahmedabad,382424, Gujarat, India.

ARTICLE INFO

Received: 22 Dec 2024

Revised: 14 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Current research is a piece of an innovative approach to concrete mix-design prediction by implementing advanced regression techniques, that addressed the limitations of traditional methods in IS 10262 and ACI 318 standards which rely heavily on empirical relationships and require extensive trial batching. The study investigates eight important mix-design parameters namely water-cementratio, cement content, flyash content, fine aggregate content, 10 mm and 20 mm aggregate content, water content, and superplasticizer content. The methodology utilizes comprehensive Multioutput Regression with gradient boosting and decision tree regressors, chosen, for their ability to capture complex non-linear relationships between material properties and mix-design proportions that traditional methods often oversimplify. Through k-cross validation using a 70-30 train-test split on a dataset of 180 actual laboratory samples, the Multioutput Regression achieved a coefficient of determination (R-squared) of 0.99, significantly outperforming both the Decision Tree Regressor (R-squared: 0.89), and traditional design methods. While traditional methods exhibit 15-20% prediction errors, the current model reduced this error margin to 3-5%, leading to potential material cost savings of 8-12% and reducing trial batching by up to 30%. Furthermore, Mean Squared Error was calculated across all predicted parameters which helped in verifying the model's robustness and good performance in aspects like water-cement ratio (MSE: 0.051) and cement content (MSE: 1:52). Unlike conventionally used methods which require multiple trial and errors of mix-design, the current method simultaneously optimizes all mix-designingredients while taking into account 27 distinct material properties offering a more efficient and accurate design process.

Keywords - Construction materials,Concrete mix design optimization, Multioutput Regression, Gradient Boosting, Decision Tree Regressor

1. Introduction

Concrete is the world's most widely used man-made material, second only to water in overall usage. Its versatility allows for applications ranging from intricate architectural designs to massive infrastructure projects like dams and bridges. A thoroughly designed concrete optimizes the balance between strength, workability, and durability, leading to structures requiring less frequent repairs and

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

significant economic benefits.The designing method derived from the Indian standard code IS 10262 (2019)and ACI 318-25standard (2019) for concrete mix design rely on a lot of trial and errors for predicting concrete mix design and faces several other challenges like limited ability to handle material variability, difficulty in capturing non-linear relationships between ingredients, inconsistent results across different batches, and inability to simultaneously optimize multiple parameters.These limitations have significant implications for construction efficiency and concrete quality. Also, in spite of having done numerous mix-designs, every single time the ingredients change in terms of the source or properties, it becomes essential to carry out the entire trial and error process again to come up with an optimum concrete mix design satisfying the needs of the compressive strength and slump. While IS 10262 and ACI standards provide valuable guidelines, this process when done manually, requires extensive laboratory testing and expert interpretation, leading to increased cost and time efforts, potential inconsistencies and wastage of material.

With the advent of powerful tools like Fuzzy Logic, Support Vector Machines (SVM), Artificial Intelligence, and Machine Learning it has become very easy for analyzing complex mathematical relationships in concrete mix design. Several researchers have explored these techniques to develop more optimum and efficient prediction models. The following studies demonstrate the evolution and application of these computational methods in concrete mix design optimization.

Many researchers have focused on predicting concrete compressive strength using various machine learning approaches. Feng et al. (2020) employed adaptive boosting, while Silva and Moita (2019) conducted a comparative study of multiple machine learning methods. Young et al. (2019) combined statistical analysis with machine learning techniques, and Al-Salloum et al. (2012)specifically utilized neural networks for strength prediction. These studies demonstrated that machine learning approaches consistently achieved prediction accuracies above 85% for compressive strength, highlighting their potential as reliable alternatives to traditional prediction methods.

Artificial Neural Networks (ANN) have been extensively applied in concrete research not just for traditional but even advanced concretes, with Hodhod et al. (2019) developing models for nano silica/silica fume concrete strength prediction, Asteriset al. (2016) investigating self-compacting concrete strength, and PinanditaFaiz (1998) focusing on high-performance concrete strength prediction. In the realm of geopolymer concrete, Van Dao, Trinh, S.H. et al. (2019) and Van Dao, Ly, H.B. (2019) proposed hybrid artificial intelligence approaches for strength prediction of mixes using steel slag aggregates, while Lahoti et al. (2017) studied mix design factors and strength prediction of metakaolin-based geopolymer. The widespread adoption of ANN across various types of concrete research indicates its versatility and effectiveness in modeling complex concrete properties, particularly for specialized concrete mixes incorporating supplementary cementitious materials.

Workability and slump characteristics have been analyzed using various computational methods. Cihan(2019)and Moayedi et al. (2019) investigated concrete slump prediction using machine learning and metaheuristic techniques respectively. Hoang and Pham (2016)employed least squares support vector regression for workability estimation, while Aydogmus et al. (2015) conducted a comparative assessment of bagging ensemble models for slump flow modeling. Jain et al. (2008) and Bai et al. (2003) utilized artificial neural networks for analyzing concrete slump and workability, with the latter specifically incorporating metakaolin and fly ash. These studies collectively demonstrated that computational methods can effectively predict concrete workability parameters with average prediction errors typically below 10%, making them valuable tools for quality control in concrete production.

Research on self-compacting concrete properties has been conducted by Siddique et al. (2011, 2008), who applied both artificial neural networks and support vector machines in their studies. Ziolkowski and Niedostatkiewicz(2019) explored broader applications of machine learning in concrete

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

mix design, while Khashman and Akpinar (2017) focused on non-destructive strength prediction using neural networks.

Wu Zheng et al. (2023) have also explored sophisticated approaches combining multiple techniques like K-fold cross-validation, Bayesian hyperparameter optimization, and regression feature elimination to achieve high prediction accuracies of 0.993 for compressive strength. In foam concrete applications, Mohamed Abdellatief et al. (2024) performed comparative studies of various algorithms including Linear Regression (LR), Support Vector Regression (SVR), Multilayer-Perceptron Artificial Neural Network (MLP-ANN), and Gaussian Process Regression (GPR) which demonstrated GPR's superior performance with R^2 values of 0.98, significantly outperforming traditional methods. These studies highlight how advanced computational approaches can achieve remarkably high prediction accuracies while offering insights into the complex relationships between mix design parameters and concrete performance.

The majority of this work aims at optimizing mix design by predicting only the individual properties of fresh and hardened concrete, like workability and compressive strength. However, researchers have also attempted to optimize mix design by predicting the ratio of the ingredients of concrete, i.e., predicting the actual optimum mix design using mathematical models. Fan et al. (2020) developed a fuzzy weighted relative error support vector machine for reverse prediction of concrete components, integrating fuzzy logic and support vector machines for more accurate predictions. Aaron (2015)predicted the mix design for Geopolymer concrete using Class F flyash and GGBFSusing ANN.

While these studies have made valuable contributions to the field and has successfully investigates the use of AI and other mathematical models in the field of concrete technology, they have done so without exhaustively addressing the comprehensive set of 27 parameters for predicting concrete mix design specified by IS 10262 code (2019) and the ACI 318-19 (2019), the guidelines for concrete mix-design. The present research addresses this gap by developing a novel model that simultaneously predicts seven crucial mix design ingredients: water content, cement content, fine aggregate content, 10mm and 20mm aggregate contents, flyash content, and superplasticizer dosage, also making it one of the very first models to predict 7 mix-design outputs simultaneously.

The research involved collecting extensive data from consulting laboratories working on various government and private infrastructure projects, analyzing material properties as per Indian Standard codes, and creating a unique dataset containing diverse mix designs for different concrete grades, different workability and concrete strength tested at different ages; making this study a far more advanced and accurate version of the similar technique used by concrete mix design consultancies. Concrete mix-design consultancies are hired by government and private agencies to conduct a thorough investigation of several raw material properties, and come up with an optimum mix-design. The consultancies do so by computing formulae laid by thecode guidelines using excel sheets and relying heavily on trial-error methods as well as expert intervention to optimize the mix-design as per site requirements. Nonetheless, these traditional methods often fall short in capturing the detailed interdependencies among the concrete ingredients and their properties which this research has taken into account.

The proposed holistic mathematicalstudy that showcases the effectiveness of the Multioutput Regression frameworkfor concrete mix design predictionincorporates important ingredient properties while maintaining high accuracy. It represents a significant advancement over existing techniques and offers a more practical, time-saving, cost-effective, and reliable solution for industry as well as research applications. This technique can also be generalized across different concrete grades and workability, as well as variability of material properties, since it has been trained and validated on diverse real-world data. The overarching goal is to identify the most suitable approach for predicting

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

concrete mix proportions with high precision, thereby enhancing the efficiency and sustainability of construction practices.

2. Data Collection& Analysis

2.1 Data Collection

The data collection phase involved obtaining 27 distinct concrete material properties and characteristics from a concrete mix design consultancy, where standardized testing procedures were conducted in alignment with Indian Standard codes as listed in Table 1. These parameters were systematically documented along with the mix proportions of concrete samples prepared using traditional methods. As part of the consultancy's process, key measurements of fresh concrete like slump; and hardened concrete like compressive strength at various ages, were recorded using standard-sized cubes standard testing procedures. This data collection effort resulted in a dataset comprising 180 rows and 35 columns, with 27 columns being input values and 8 columns being the predicted output values.

The 27 inputs as mentioned in Table 1 resulted in the prediction of the mix design of concrete which contained the following 8 outputs:

1. Water content (liter per meter cube)
2. Cement Content (kg per meter cube)
3. Water/Cement ratio
4. Fine aggregate content (kg per meter cube)
5. 10mm coarse aggregate content (kg per meter cube)
6. 20mm coarse aggregate content (kg per meter cube)
7. Flyash (kg per meter cube)
8. Superplasticizer (kg per meter cube)

2.2 Data Analysis

The input parameters were thoroughly analyzed to understand their mathematical correlations, interactions, and impacts on the concrete mix design. Below are the graphs showcasing the results of the exploratory data analysis.

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Table 127 Concrete Parameters and Material Characteristics for Predicting Mix-Design

Sr. No.	Material	Property/Characteristic	Range	Reference/Standard
1	Fine aggregate	Zone of Sand	1-2	IS 383 (2016)
2	Fine aggregate	Fineness Modulus	2.18-3.25	IS 383 (2016)
3	Fine aggregate	Silt Content	0.2-2.8	IS 2386-2 (1963)
4	10mm coarse aggregate	Flakiness %	9.54-34.16	IS 2386-1 (1963)
5	20mm coarse aggregate	Flakiness %	0-25.67	IS 2386-1 (1963)
6	10mm coarse aggregate	Elongation %	4.4-20.3	IS 2386-1 (1963)
7	20mm coarse aggregate	Elongation %	0-25.78	IS 2386-1 (1963)
8	10mm coarse aggregate	Impact Value %	11.19-14.95	IS 2386-4 (1963)
9	20mm coarse aggregate	Impact Value %	0-15.11	IS 2386-4 (1963)
10	Fine aggregate	Bulk Density (g/cc)	1.46-1.64	IS 2386-3 (1963)
11	10mm coarse aggregate	Bulk Density (g/cc)	1.459-1.546	IS 2386-3 (1963)
12	20mm coarse aggregate	Bulk Density (g/cc)	0-1.629	IS 2386-3 (1963)
13	Cement	Grade	43, 53	IS 269 (2015)
14	Cement	Type of Cement	OPC, PPC, SRC, Slag	IS 269 (2015)
15	Cement	Cement brand	Ultratech, Ambuja, Kamal SRC, Hi Bound, Sanghi, Platinum, JK Laxmi, JK Super, Sidhee, Nuvoco,	Manufacturer
16	Cement	Specific Gravity	2.96-3.15	IS 4031-11 (1988)
17	Fine aggregate	Specific Gravity	2.58-2.72	IS 2386-3 (1963)
18	10mm coarse aggregate	Specific Gravity	2.77-2.84	IS 2386-3 (1963)
19	20mm coarse aggregate	Specific Gravity	2.85-2.89	IS 2386-3 (1963)
20	Fine aggregate	Water absorption	0.6-4.45	IS 2386-3 (1963)
21	10mm coarse aggregate	Water absorption	0.65-1.08	IS 2386-3 (1963)
22	20mm coarse aggregate	Water absorption	0.5-0.96	IS 2386-3 (1963)
23	Concrete	Grade	7.5-50	IS 456 (2000)

e-ISSN: 2468-4376
https://www.jisem-journal.com/

24	Superplasticizer	% Dosage	0-3.96	IS 9103 (2021)
25	Fresh concrete	Slump (mm)	10-170	IS 1199-2 (2018)
26	Hardened concrete	Compressive strength (N/mm^2)	6.84-59.19	IS 516 (1959)
27	Hardened concrete	Age (days)	3-28	IS 516 (1959)

The correlations illustrating the interrelationships between various concrete parameters and material properties in the dataset are presented in the heat map shown in figure 1. The heatmap employs a color gradient scale ranging from dark purple (indicating strong negative correlations of -1.0) through red (neutral correlations around 0) to light pink (strong positive correlations of 1.0). The diagonal elements display perfect positive correlations (1.0) as expected, represented by white squares, since each correlates perfectly with itself.

Several notable correlation patterns emerge from the analysis. The physical properties of aggregates, including flakiness and elongation percentages for both 10mm and 20mm aggregates, demonstrate moderate to strong correlations with each other, suggesting inherent relationships in aggregate shape characteristics. The bulk density measurements across different aggregate sizes (sand, 10mm, and 20mm) also exhibit significant correlations, indicating consistency in material density properties. Water absorption characteristics across different aggregate sizes show similar patterns of correlation, highlighting the interconnected nature of material porosity properties. The mixture proportion parameters, particularly cement content, water content, and superplasticizer dosage, display distinct correlation patterns with the final concrete properties such as slump and average strength. The water-cement ratio demonstrates expected correlations with strength parameters, aligning with fundamental concrete technology principles. The superplasticizer dosage shows notable correlations with workability parameters, reflected in the slump measurements.

The correlation heatmap analysis provided foundational insights that enhanced the model's predictive capabilities. By revealing strong correlations between aggregate properties (flakiness, elongation, and water absorption across different sizes) and identifying significant relationships between mixture proportions and final concrete properties, the heatmap validated the feature selection approach. These correlations aligned with established concrete technology principles, demonstrating that the model's predictions were grounded in theoretical understanding. The visualization of parameter interdependencies, particularly in water absorption characteristics and strength parameters, guided the feature engineering decisions and ultimately contributed to the model's ability to accurately predict optimal concrete mix designs.

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

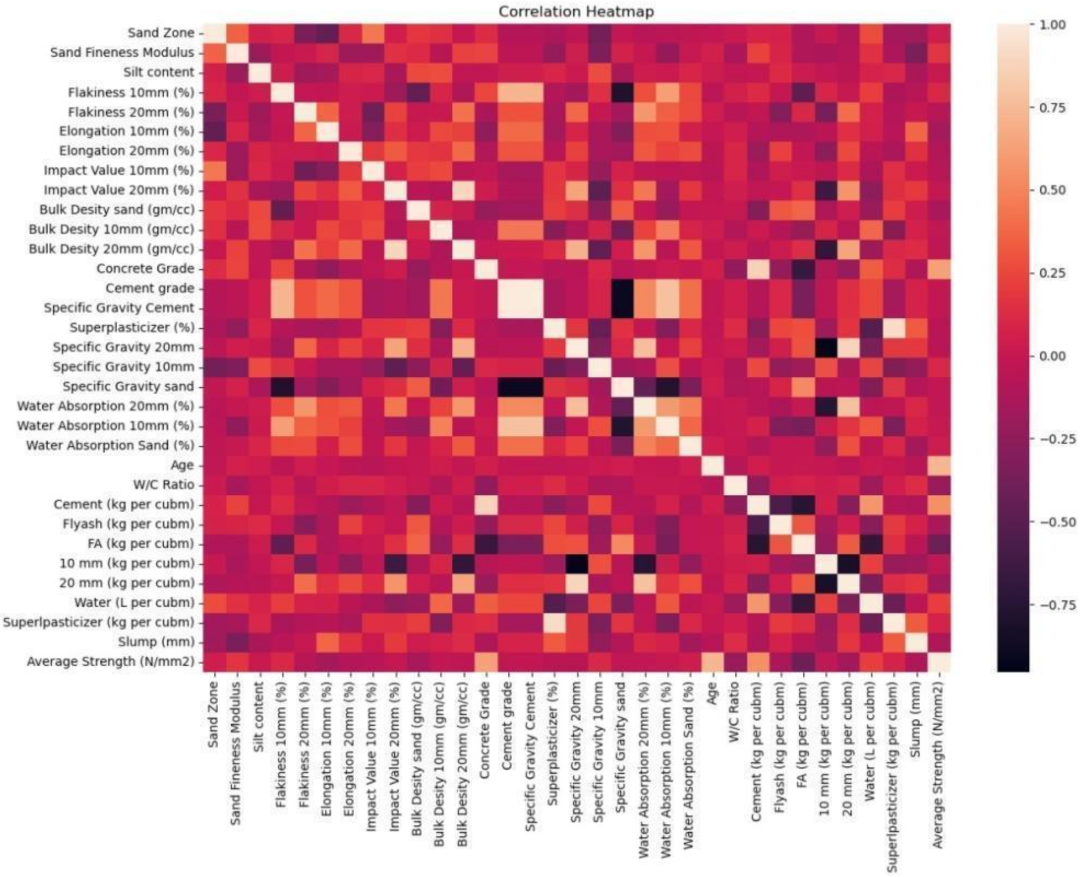

Fig. 1 Heat Map showing correlations among all parameters

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

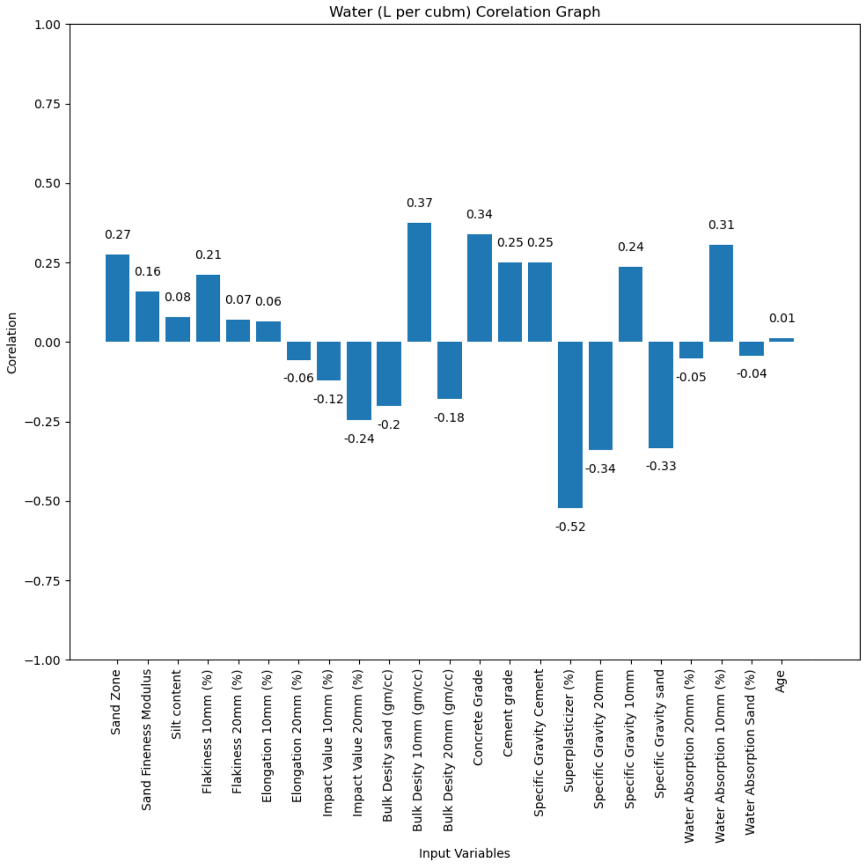

Fig. 2 Correlation graph of Water (output) with other input parameters

The given correlation graph in fig. 2 illustrates how various parameters influence the water content in the given concrete mixes. The height of each bar indicates the strength and direction (positive or negative) of the correlation. Key conclusions are that superplasticizer percentage (-0.52) has negative and the strongest correlation.

The individual correlations of all the output mix-design parameters with all the input properties were also studied through a correlation graph. The correlation analysis presented in figure 2 illustrates the relationships between water content (L per cu.m.) and various input parameters in concrete mix design. The strong negative correlation (-0.52) between water content and superplasticizer dosage substantiates the fundamental role of superplasticizers in concrete technology, as they effectively reduce water demand while maintaining workability, thereby enabling lower water-cement ratios for enhanced concrete performance.

Significant correlations are also noted with aggregate physical properties, including flakiness (0.21) and water absorption characteristics (-0.33 to 0.31), validating their inclusion as predictive features in the machine learning model. These relationships align with established concrete technology principles, where aggregate characteristics substantially influence water demand in concrete mixtures. The correlation patterns identified through this analysis served to validate the feature selection approach further and enhanced the gradient boosting regressor's predictive accuracy. The model's architecture benefited from these insights, particularly in capturing the complex interactions between material properties and their collective impact on water requirements in concrete mix design.

3. Methodology

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The computational analysis began with thorough data preprocessing to ensure optimal model training conditions. The framework, as depicted in Figure 3, shows the systematic flow from input data processing through model training to final testing and validation, representing a comprehensive approach to concrete mix design optimization through machine learning techniques.

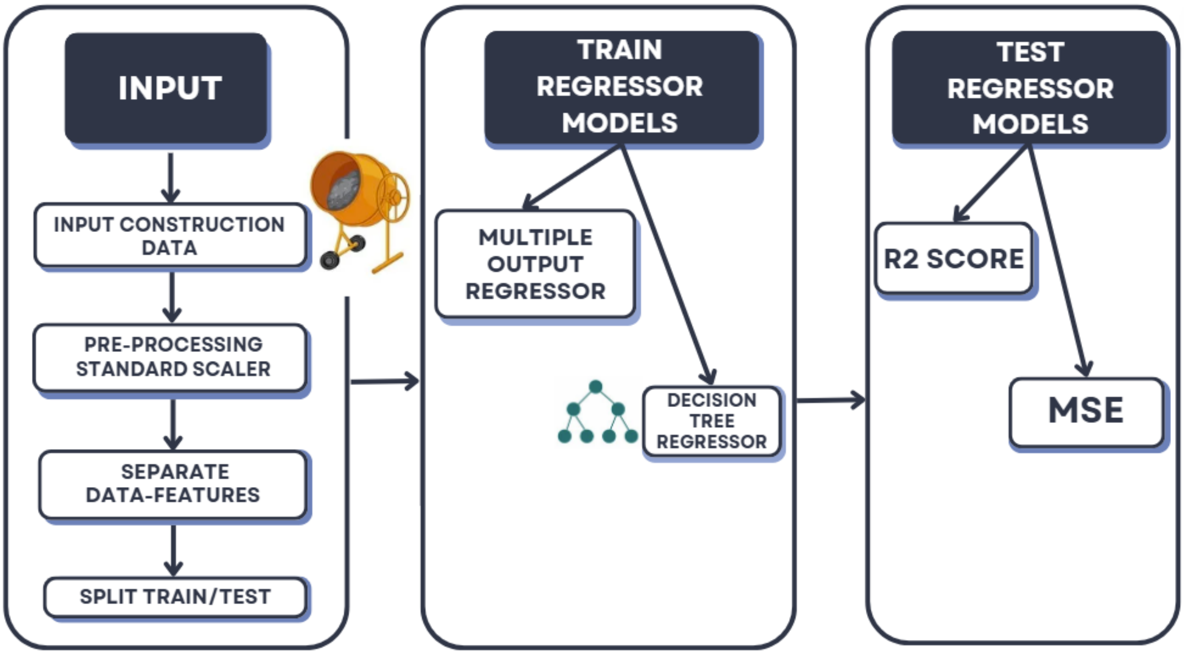

Fig. 3 Proposed Regression Framework

The dataset, consisting of 180 data points with 27 input parameters and 8 output variables representing final concrete mix proportions, was standardized using the StandardScaler technique. This crucial step removed mean values and scaled to unit variance, ensuring uniform contribution of all features to the model's training process, particularly important for algorithms utilizing distance measures or gradient descent optimization. Further data preparation phase involved systematically dividing the preprocessed data into input features and target variables as listed in chapter 2. The dataset was then split into training and testing sets using a 70:30 ratio enabling robust model validation. Feature engineering techniques were implemented to enhance the model's predictive capabilities, creating composite indices that better captured the relationships between related parameters such as aggregate gradation and specific gravity.
The modeling phase employed two distinct regression approaches: Multioutput Regression and Decision Tree Regressor. The Multioutput Regression framework, an ensemble learning method, was particularly effective in handling the simultaneous prediction of multiple dependent variables. The Decision Tree Regressor complemented this approach by capturing non-linear relationships and providing interpretable insights into the impact of material properties on mix design outcomes. Both models underwent rigorous hyperparameter tuning using Grid Search and Random Search techniques to optimize their performance.Model evaluation incorporated comprehensive cross-validation techniques, particularly k-fold cross-validation, to ensure robust performance assessment. Sensitivity analysis was conducted to identify the most influential input parameters affecting concrete mix design, providing valuable insights for practical applications. The models' performance was evaluated using two key metrics: the R^2 score, indicating the model's effectiveness on unseen data, and Mean

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Squared Error (MSE), measuring the average squared differences between predicted and actual values.

$$R^2 Score = \frac{Correct\ Prediction}{All\ Predictions} \tag{1}$$

$$MSE = \frac{1}{No.of\ Predictions} \sum_1^n (Actual\ Value - Predicted\ Value)^2 \tag{2}$$

The methodology's scalability represents a significant advantage, as it can effectively handle datasets containing thousands of data points without excessive computational demands, owing to the text-based nature of the data. Furthermore, the trained model's learning can be preserved and applied to evaluate individual parameter sets, enabling on-demand mix design optimization. This comprehensive approach ensures precise predictions for optimizing concrete compositions based on site-specific workability requirements while maintaining long-term structural durability.

4. Results and Discussion

The Decision Tree Regressor model initially showed signs of overfitting, indicated by a perfect training score of 1.0 and a lower testing score of 0.88 as shown in figure 4. This prompted the implementation of strategies to enhance the model's ability to generalize, including limiting the tree's maximum depth and employing cost-complexity pruning techniques. Limiting the 'max_depth'hyperparameter to 5 resulted in training and testing scores of 0.92 and 0.86, respectively, while a depth of 7 yielded scores of 0.95 and 0.89, demonstrating improved generalization while maintaining model complexity.

```
from sklearn.tree import DecisionTreeRegressor
model = DecisionTreeRegressor()
model.fit(xtrain, ytrain)
```

```
▾ DecisionTreeRegressor
DecisionTreeRegressor()
```

```
score = model.score(xtrain, ytrain)
print("Training score:", score)
```

```
Training score: 1.0
```

```
result = model.score(xtest, ytest)
print("Testing score:", result)
```

```
Testing score: 0.8820760164874455
```

Fig. 4 Training and Testing Score of Decision Tree Regressor

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

To get insights into the decision tree model's poor prediction behavior, a residual plot in figure 5 was plotted. The residual plot clearly illustrates why the model struggles with generalization. The plot displays residuals against predicted values, with most points clustering around the zero line, indicating reasonable predictions. However, there is notable heteroscedasticity, with larger residual spread at lower predicted values (0-200 range) and decreasing variability as predicted values increase. Several outliers are observed, with residuals ranging from +30 to -30, particularly in the lower prediction ranges. This pattern suggests that the model's predictions are more volatile and less reliable for lower values, while showing more stability in higher value ranges. This observation aligns with the discussion on overfitting, highlighting the need for more robust regression techniques.

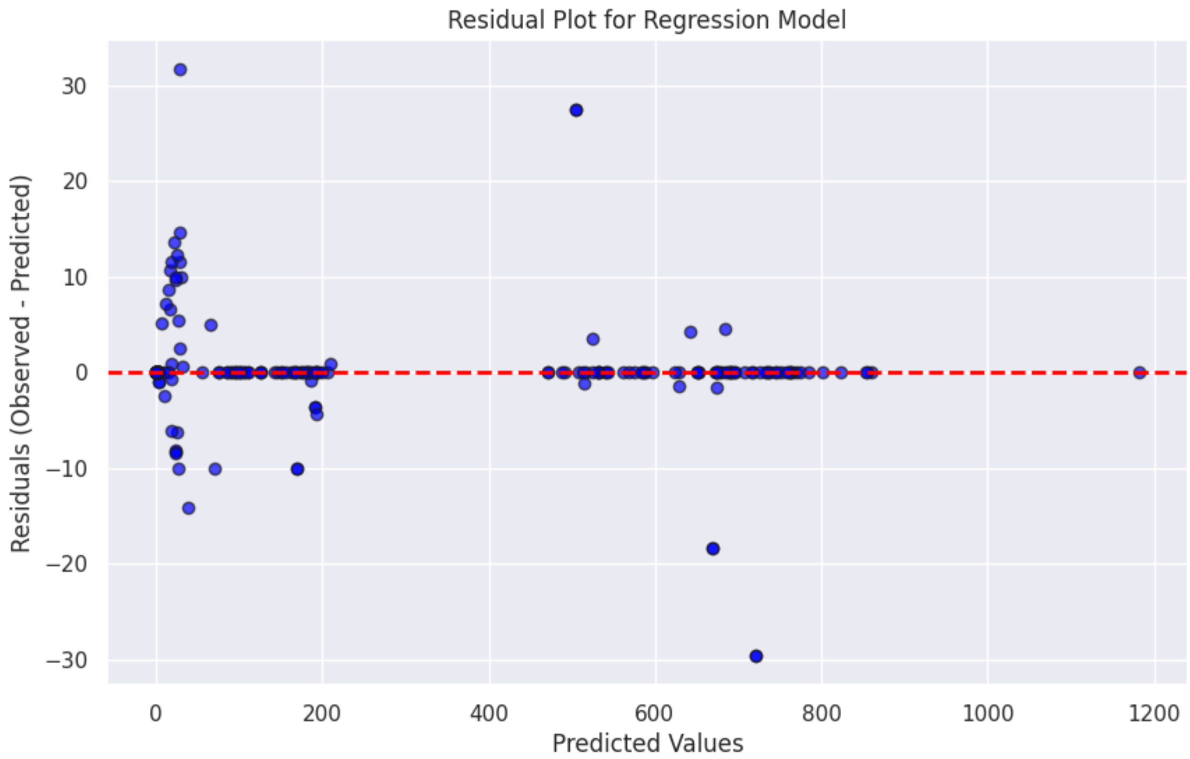

Fig. 5 Residual Analysis for Decision Tree Regressor

The Multioutput Regression model, utilizing a Gradient Boosting estimator, achieved high training and testing scores of 0.9982 and 0.9901, respectively, as shown in figure 6, suggesting excellent generalization and robustness. Advanced validation techniques, including the generation of learning curves, calculation of bootstrapped confidence intervals, and implementation of stratified k-fold cross-validation, were employed to critically evaluate the model's performance and address potential overfitting concerns. Learning curves demonstrated performance stabilization around 120-150 data points, with a narrowing gap between training and validation scores, indicating potential data sufficiency. Bootstrapped 95% confidence intervals were calculated for key performance metrics, including an R-squared value of 0.99, and Mean Squared Error (MSE) ranges for individual parameters such as W/C Ratio (0.051), Cement Content (1.52), and Superplasticizer (1.12). Stratified k-fold cross-validation (k=10) revealed consistent performance across different data splits, reducing the risk of dataset-specific bias and confirming the model's generalizability.

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

```
gbr = GradientBoostingRegressor()
model = MultiOutputRegressor(estimator=gbr)
print(model)
```

```
MultiOutputRegressor(estimator=GradientBoostingRegressor())
```

```
model.fit(xtrain, ytrain)
score = model.score(xtrain, ytrain)
print("Training score:", score)
result = model.score(xtest, ytest)
print("Testing score:", result)
```

```
Training score: 0.9982026768239773
Testing score: 0.9901485418206937
```

Fig. 6 Training and Testing Score of Multioutput Regression

A comparison of the two ML techniques as presented in table 2, revealed that Multioutput Regression consistently outperforms the Decision Tree Regressor, indicating robust generalization capabilities. The Multi-Output Regression model consistently yielded lower MSE values across all components, particularly for cement (1.52), fly ash (7.49), and superplasticizer (1.12), compared to the Decision Tree Regressor's values of 6.004, 56.10, and 12.98. The Multioutput Regression model achieved a prediction error of 3-5% with consistent performance across concrete grades, while the optimized Decision Tree Regressor showed a prediction error of 7-10%, improved from its initial unconstrained version. These results highlight the superior accuracy and reliability of the Multi-Output Regression model in predicting various concrete mix design components. Benchmarking against the traditional IS 10262 standard method, which exhibits an average prediction error of 15-20% and high variability across different concrete types, the machine learning models demonstrated superior performance. Hence these findings translate to potential material cost savings of 8-12%, a reduction in trial batching of up to 30%, and improved mix design consistency. The ability of the model to capture complex material interactions, exceeding traditional empirical methods, is crucial for optimizing concrete compositions, ensuring desired workability and compressive strength, and ultimately enhancing the overall quality and sustainability of construction projects.

Table 2 Comparison of Performance of the Models

Model	**Train Score**	**Test Score**	**MSE of Actual and Predicted Value**							
			W/C Ratio	**Cement**	**Flyash**	**FA**	**CA 10 mm**	**CA 20 mm**	**Water**	**Superplasticizer**
Multi output Regression	0.99	0.99	0.051	1.52	7.49	8.088	5.54	1.48	0.003	1.12
Decision Tree Regressor	0.95	0.89	0.00	6.004	56.10	0.662	1.74	0.06	12.03	12.98

5. Conclusion

In conclusion, this research critically examined the application of machine learning techniques in concrete mix design, revealing both significant potential and inherent limitations. The Multioutput Regression demonstrated superior generalization with an R-squared value of 0.99, while the Decision

Tree Regressor showed marked improvement through targeted optimization techniques. Traditional mix design methods exhibit 15-20% prediction errors, whereas the implemented machine learning models successfully reduced these error margins to 3-10%, indicating significant potential for industry optimization. The practical implications of these findings are substantial, with estimated material cost savings of 8-12%, reduced trial batching by up to 30%, and enhanced mix design consistency and predictability. Sensitivity analysis highlighted the critical parameters influencing concrete mix design, including the Water-Cement Ratio, Aggregate Specific Gravity, and Superplasticizer Percentage. This research represents a crucial initial step towards data-driven, precision-engineered concrete mix design. While promising, it underscores the need for continued research, focusing on expanding datasets and fostering interdisciplinary collaboration across materials science, civil engineering, and machine learning to realize the full potential of these techniques.

The technique presented in this study opens several such avenues for future research. The prediction based models of machine learning can help discover a lot more connections among several different ingredients that are used in concrete nowadays [39]. Moreover it can also help us predict concrete mix designs, various properties of fresh and hardened concrete for concrete other than the routine applications such as 3D printed concrete [40] and various other types of advanced concretes [41]. With future advancements and exploration in regression techniques, we could effectively conduct a wide variety of research and accurately predict their behavior which can help us cut costs and save valuable time.

Data availability statement

The datasets supporting the findings of this study are available as supplementary material with this article.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Abdellatief M, Wong LS, Din NM, Mo KH, Ahmed AN, El-Shafie A (2024) Evaluating enhanced predictive modeling of foam concrete compressive strength using artificial intelligence algorithms. Materials Today Communications 40. https://doi.org/10.1016/j.mtcomm.2024.110022

[2] ACI Committee 318 (2025) Building Code Requirements for Structural Concrete (ACI 318-25)

[3] Al-Salloum YA, Shah AA, Abbas H, Alsayed SH, Almusallam TH, Al-Haddad MS (2012) Prediction of compressive strength of concrete using neural networks. Computers and Concrete 10:197-217. https://doi.org/10.12989/cac.2012.10.2.197

[4] Asteris PG, Kolovos KG, Douvika MG, Roinos K (2016) Prediction of self-compacting concrete strength using artificial neural networks. European Journal of Environmental and Civil Engineering 20:s102-s122. https://doi.org/10.1080/19648189.2016.1246693

[5] Aydogmus HY, Erdal HI, Karakurt O, Namli E, Turkan YS, Erdal H (2015) A comparative assessment of bagging ensemble models for modeling concrete slump flow. Computers and Concrete 16:741-757. https://doi.org/10.12989/cac.2015.16.5.741

2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

[6] Bai J, Wild S, Ware JA, Sabir BB (2003) Using neural networks to predict workability of concrete incorporating metakaolin and fly ash. Advances in Engineering Software 34:663-669. https://doi.org/10.1016/S0965-9978(03)00102-9

[7] Bureau of Indian Standards (1963) IS 2386 (Part 1):1963 -- Methods of Test for Aggregates for Concrete -- Part 1: Particle Size and Shape. BIS, New Delhi

[8] Bureau of Indian Standards (1963) IS 2386 (Part 2):1963 -- Methods of Test for Aggregates for Concrete -- Part 2: Estimation of Deleterious Materials and Organic Impurities. BIS, New Delhi

[9] Bureau of Indian Standards (1963) IS 2386 (Part 3):1963 -- Methods of Test for Aggregates for Concrete -- Part 3: Specific Gravity, Density, Voids, Absorption, and Bulking. BIS, New Delhi

[10] Bureau of Indian Standards (1963) IS 2386 (Part 4):1963 -- Methods of Test for Aggregates for Concrete -- Part 4: Mechanical Properties. BIS, New Delhi

[11] Bureau of Indian Standards (1959) IS 516:1959 -- Method of Tests for Strength of Concrete. BIS, New Delhi

[12] Bureau of Indian Standards (1980) IS 2720 (Part 3/Sec 1):1980 -- Methods of Test for Soils -- Part 3: Determination of Specific Gravity -- Section 1: Fine-Grained Soils. BIS, New Delhi

[13] Bureau of Indian Standards (1988) IS 4031 (Part 11):1988 -- Methods of Physical Tests for Hydraulic Cement -- Part 11: Determination of Specific Gravity of Cement. BIS, New Delhi

[14] Bureau of Indian Standards (1999) IS 9103:1999 -- Specification for Concrete Admixtures. BIS, New Delhi

[15] Bureau of Indian Standards (2000) IS 456:2000 -- Code of Practice for Plain and Reinforced Concrete. BIS, New Delhi

[16] Bureau of Indian Standards (2015) IS 269:2015 -- Ordinary Portland Cement -- Specification. BIS, New Delhi

[17] Bureau of Indian Standards (2016) IS 383:2016 -- Coarse and Fine Aggregate for Concrete -- Specification. BIS, New Delhi

[18] Bureau of Indian Standards (2018) IS 1199 (Part 2):2018 -- Fresh Concrete -- Part 2: Determination of Consistency Using Compacting Factor Test. BIS, New Delhi

[19] Bureau of Indian Standards (2019) IS 10262:2019 -- Guidelines for Concrete Mix Design Proportioning (Second Revision). BIS, New Delhi

[20] Cihan MT (2019) Prediction of Concrete Compressive Strength and Slump by Machine Learning Methods.Advances in Civil Engineering. https://doi.org/10.1155/2019/3069046

[21] Fan Z, Chiong R, Hu Z, Lin Y (2020) An unbiased fuzzy weighted relative error support vector machine for reverse prediction of concrete components. Computers and Structures 230:106171. https://doi.org/10.1016/j.compstruc.2019.106171

[22] Feng DC, Liu ZT, Wang XD, Chen Y, Chang JQ, Wei DF, Jiang ZM (2020) Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. Construction and Building Materials 230:117000. https://doi.org/10.1016/j.conbuildmat.2019.117000

[23] Hoang ND, Pham AD (2016) Estimating Concrete Workability Based on Slump Test with Least Squares Support Vector Regression. Journal of Construction Engineering 2016:1-8. https://doi.org/10.1155/2016/5089683

[24] Hodhod OA, Khalafalla MS, Osman MSM (2019) ANN models for nano silica/silica fume concrete strength prediction. Water Science 33:118-127. https://doi.org/10.1080/11104929.2019.1669005

[25] Jain A, Jha SK, Misra S (2008) Modeling and Analysis of Concrete Slump Using Artificial Neural Networks. Journal of Materials in Civil Engineering 20:628-633. https://doi.org/10.1061/(asce)0899-1561(2008)20:9(628)

[26] Khashman A, Akpinar P (2017) Non-Destructive Prediction of Concrete Compressive Strength Using Neural Networks. Procedia Computer Science 108:2358-2362. https://doi.org/10.1016/j.procs.2017.05.039

[27] Lahoti M, Narang P, Tan KH, Yang EH (2017) Mix design factors and strength prediction of metakaolin-based geopolymer. Ceramics International 43:11433-11441. https://doi.org/10.1016/j.ceramint.2017.06.006

[28] Malik UJ, Riaz RD, Rehman SU, Usman M, Riaz RE, Hamza R (2024) Advancing mix design prediction in 3D printed concrete: Predicting anisotropic compressive strength and slump flow. Case Studies in Construction Materials 21. https://doi.org/10.1016/j.cscm.2024.e03510

[29] Moayedi H, Kalantar B, Foong LK, Bui DT, Motevalli A (2019) Application of three metaheuristic techniques in simulation of concrete slump. Applied Sciences 9. https://doi.org/10.3390/app9204340

[30] Nwankwo CO, Mahachi J, Olukanni DO, Musonda I (2023) Natural fibres and biopolymers in FRP composites for strengthening concrete structures: A mixed review. Construction and Building Materials 363. https://doi.org/10.1016/j.conbuildmat.2022.129661

[31] PinanditaFaiz R (1998) Modeling of strength of high-performance concrete using artificial neural networks. Cement and Concrete Research 43:1-9

[32] Siddique R, Aggarwal P, Aggarwal Y (2011) Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks. Advances in Engineering Software 42:780-786. https://doi.org/10.1016/j.advengsoft.2011.05.016

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

[33] Siddique R, Aggarwal P, Aggarwal Y, Gupta SM (2008) Modeling properties of self-compacting concrete: Support vector machines approach. Computers and Concrete 5:461-473. https://doi.org/10.12989/cac.2008.5.5.461

[34] Silva PFS, Moita GF (2019) A comparative study of machine learning methods for compressive strength of concrete. World Congress on Civil, Structural, and Environmental Engineering:0-6. https://doi.org/10.11159/icsect19.136

[35] Tam VWY, Butera A, Le KN, Da Silva LCF, Evangelista ACJ (2022) A prediction model for compressive strength of CO2 concrete using regression analysis and artificial neural networks. Construction and Building Materials 324. https://doi.org/10.1016/j.conbuildmat.2022.126689

[36] Van Dao D, Ly HB, Trinh SH, Le TT, Pham BT (2019) Artificial intelligence approaches for prediction of compressive strength of geopolymer concrete.Materials 12. https://doi.org/10.3390/ma12060983

[37] Van Dao D, Trinh SH, Ly HB, Pham BT (2019) Prediction of compressive strength of geopolymer concrete using entirely steel slag aggregates: Novel hybrid artificial intelligence approaches. Applied Sciences 9:1-16. https://doi.org/10.3390/app9061113

[38] Wilson AP (2015) Establishing a Mix Design Procedure for Geopolymer Concrete. Faculty of Health, Engineering and Sciences, University of Southern Queensland, Australia

[39] Wu Z, Shui Z, Xu Z, Gai X, Zhang S (2023) Multi-objective optimization of concrete mix design based on machine learning. Journal of Building Engineering 76. https://doi.org/10.1016/j.jobe.2023.107396

[40] Young BA, Hall A, Pilon L, Gupta P, Sant G (2019) Can the compressive strength of concrete be estimated from knowledge of the mixture proportions: New insights from statistical analysis and machine learning methods. Cement and Concrete Research 115:379-388. https://doi.org/10.1016/j.cemconres.2018.09.006

[41] Ziolkowski P, Niedostatkiewicz M (2019) Machine learning techniques in concrete mix design. Materials 12. https://doi.org/10.3390/ma12081256