Prediction and Optimization of Cementitious Composites Using Machine Learning Algorithms

B J S Varaprasad[1] [1*], V Rangaveni[2] and B Yashwanth[3]

[1] Professor, Civil Engineering Department, G Pulla Reddy Engineering College (A), Kurnool, Andhra Pradesh-518007, India.
[2] Assistant Professor, Humanities and Basic Sciences Department, G Pulla Reddy Engineering College (A), Kurnool, Andhra Pradesh-518007, India.
[3] Post Graduate Student, Civil Engineering Department, National Institute of Technology Warangal, Warangal, Telangana-506004, India.

drbjsvp.ce@gprec.ac.in

Abstract. Advanced machine learning (ML) techniques are employed in this study to enhance the design and performance of cementitious composites, with a focus on optimizing concrete's compressive strength—a key attribute for resilient structures. By leveraging algorithms such as Genetic Algorithms (GA), Artificial Neural Networks (ANN), and Support Vector Machines (SVM), the research refines concrete mix compositions, carefully adjusting parameters like water-cement ratio, aggregate proportions, and admixture levels. Rigorous data preprocessing ensures high-quality input, allowing the ML models to capture intricate, nonlinear relationships between these variables and compressive strength, significantly reducing the need for traditional, time-intensive testing. Through GA-based optimization, the study iteratively improves mix designs to maximize both durability and cost-efficiency, achieving outcomes superior to conventional empirical methods. This integration of ML and optimization techniques provides a scalable, adaptive framework for concrete design, promoting material efficiency and structural resilience. By setting a new standard for performance-based mix design, the study highlights ML's transformative potential in civil engineering, advancing sustainable practices while meeting rigorous performance requirements.

Keywords: Machine Learning, Genetic Algorithm, Optimization.

1 Introduction

Cementitious composites, particularly concrete, are fundamental materials in modern construction due to their inherent strength, durability, and versatility. As a widely used material, concrete consists of a combination of cement, aggregates, water, and various supplementary components such as fly ash, silica fume, and slag, which contribute to its performance in different environmental conditions. Optimizing the mix

B. S. Babu et al. (eds.), *Proceedings of International Conference on Advanced Materials, Manufacturing and Sustainable Development (ICAMMSD-2024)*, Advances in Engineering Research 257,
https://doi.org/10.2991/978-94-6463-662-8_31

design of concrete to enhance its mechanical and durability properties is a critical, ongoing challenge in the field of civil engineering.

In recent years, advancements in machine learning (ML) have shown great potential in predicting critical material properties like compressive strength (CS) and providing efficient methods for optimizing concrete mix designs. By leveraging large datasets from experimental studies and applying ML algorithms, engineers are now able to achieve precise predictions and enhance concrete performance more effectively than traditional methods. This project focuses on the implementation of advanced ML techniques, including Genetic Algorithms (GA), Artificial Neural Networks (ANN), and Gene Expression Programming (GEP), to predict and optimize the compressive strength of cementitious composites. These algorithms not only offer powerful tools for analysing vast amounts of data but also enable the fine-tuning of mix designs to meet specific performance criteria. This approach aims to improve both the accuracy and efficiency of the design process, contributing to the development of more resilient and sustainable concrete structures.

2 Literature Review

Machine learning (ML) is rapidly advancing the field of concrete technology by enabling precise predictions and optimized designs for high-performance concrete materials. Recent studies reveal the effectiveness of ML models in enhancing critical properties like compressive strength, durability, and microstructural integrity in diverse and challenging conditions.

One major application is the use of ML models—such as Random Forest (RF), Adaboost, CNN, K-Nearest Neighbour (KNN), and Bidirectional LSTM (BI-LSTM)—to predict compressive strength in self-compacting concretes (SCC) with high volumes of fly ash and silica fume. Of these models, RF achieved the highest accuracy based on R^2 and RMSE scores. Leveraging this model, researchers developed an open-source graphical interface for practical SCC mix optimization. Complementary microstructural analyses via SEM and XRD confirmed that these optimized SCC mixes promote denser structures, improving strength while minimizing bleeding and segregation. The use of ML in this domain provides a more efficient alternative to traditional strength-testing methods, which are labor-intensive and slow.

Furthermore, ML-driven optimization algorithms are being applied to concrete mixes designed for extreme environments, such as cold or marine conditions. By integrating models like RF with feature selection and genetic algorithms, researchers have developed hybrid methods that optimize properties such as frost resistance, chloride ion penetration resistance, and compressive strength. These enhanced mixtures demonstrated significant improvements in durability metrics, showcasing the ability of ML-enhanced designs to perform under environmental stresses.

Incorporating novel additives like graphene oxide (GO) and graphite-based nanomaterials (GN) is also being explored through ML frameworks. GO, at precise dosages, has shown substantial improvements in both static and dynamic properties of standard and high-strength concrete, enhancing compressive strength, elasticity, and

dynamic properties such as natural frequency. However, excessive amounts of GO can lead to particle agglomeration, underscoring the need for precise dosage control, which ML models can support. Meanwhile, GNs are proving beneficial in electrically conductive concrete composites, which facilitate structural health monitoring through non-destructive means. With Bayesian-optimized ML models, researchers have been able to fine-tune the balance between mechanical strength and electrical conductivity, creating optimized, multi-functional materials for advanced structural applications.
Altogether, ML-driven approaches in concrete design not only streamline the prediction and optimization of concrete properties but also enable highly tailored, efficient solutions for varied applications. These advancements mark a significant shift towards data-driven, intelligent materials engineering, paving the way for stronger, more durable, and sustainable concrete structures.

3 Data Collection

Data was systematically gathered from various literature sources, focusing on key parameters influencing concrete mix designs. This dataset includes detailed information on cement content, fly ash content, water content, aggregate content, admixture content, water-cement ratio, and curing time.
After the data collection, preprocessing steps were applied to ensure the data is clean, properly scaled, and ready for machine learning analysis. Missing data points were either imputed, using techniques such as mean or median substitution, or removed if their absence significantly skewed the dataset. Outliers, or extreme values that could distort model accuracy, were also detected and managed. Depending on their impact, outliers were either adjusted or removed to maintain data consistency. Feature scaling was a crucial next step. Since concrete mix parameters vary in scale, normalization and standardization were applied to ensure uniformity across features. Normalization was particularly useful for rescaling values to a fixed range, while standardization transformed features to have a mean of zero and a standard deviation of one. This scaling was essential to ensure that all features were comparable and to improve the convergence speed of the machine learning algorithms. Categorical variables, if any, were converted into numerical values using encoding techniques like one-hot encoding or label encoding, making them interpretable by machine learning algorithms. Finally, the pre-processed dataset was split into training and testing subsets. This split allowed for model validation and helped in assessing the model's performance on unseen data.

Fig. 1. Key Steps in Data Preprocessing

With these preprocessing steps completed, the data, including the selected parameters—cement content, fly ash content, water content, aggregate content, admixture content, water-cement ratio, and curing time—is now prepared for accurate and reliable training of machine learning models.

4 Machine learning Algorithms

Machine learning (ML) algorithms are computational methods that enable systems to learn from data and improve their performance without being explicitly programmed. By identifying patterns and relationships within large datasets, ML algorithms can make predictions, classify information, and uncover insights that were previously unknown. These algorithms are the foundation of artificial intelligence (AI) and are used in a wide range of applications, from recommendation systems and speech recognition to predictive maintenance and medical diagnostics.

Machine learning are algorithms can be broadly classified into three categories: Supervised Learning, Unsupervised Learning, and Reinforcement Learning.

1. **Supervised Learning:** In supervised learning, the model is trained using labelled data, meaning that both the input data and the corresponding correct output are provided. The goal is to learn a mapping from inputs to outputs, enabling the model to make predictions or decisions when given new, unseen data. Supervised learning algorithms can be further divided into regression (predicting continuous values) and classification (predicting discrete categories).
2. **Unsupervised Learning**: Unlike supervised learning, unsupervised learning deals with unlabelled data. The algorithm must find patterns, relationships, or clusters in the data without any prior knowledge of what the correct outputs should be. Common tasks include clustering, where the algorithm groups similar data points, and dimensionality reduction, which simplifies large datasets while retaining essential information.
3. **Reinforcement Learning**: In reinforcement learning, the algorithm learns by interacting with an environment and receiving feedback in the form of rewards or penalties. This type of learning is used in applications where an agent must make a series of decisions to achieve a goal, such as robotics, game playing, and autonomous driving.

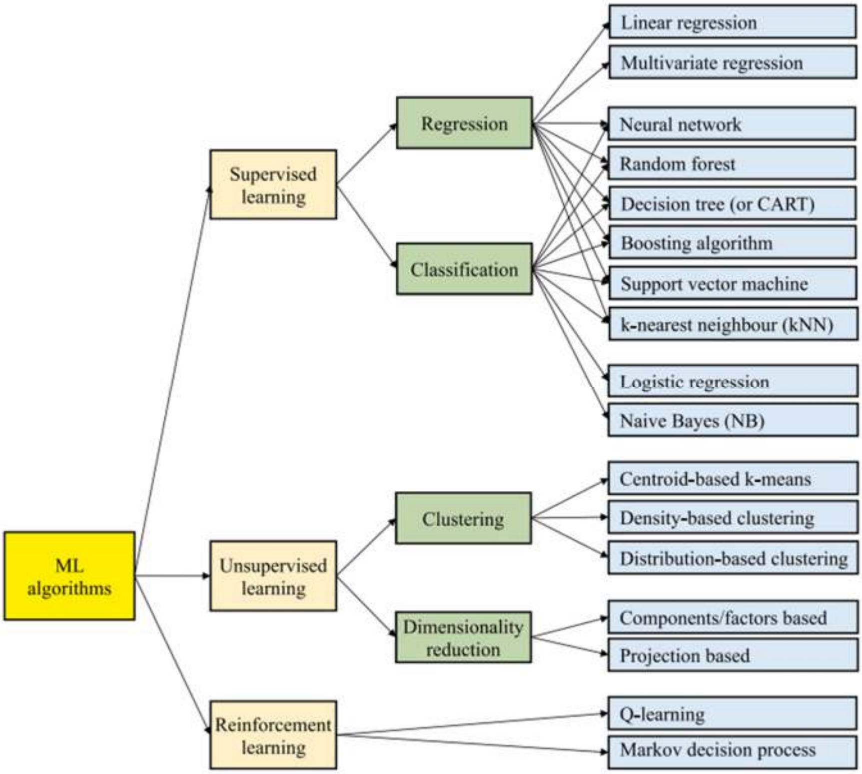


Fig. 2. Machine Learning Algorithms

[https://www.researchgate.net/publication/371444263/figure/fig1/AS:11431281166674540@1686330614255/Classification-of-ML-and-corresponding-algorithms-19-149.png]

5 Methodology

The methodology for this study follows a structured, multi-phase approach aimed at predicting and optimizing the compressive strength of cementitious composites containing fly ash. This approach begins with data extraction and data collection from various literature sources, where relevant experimental data on concrete mix proportions and associated compressive strengths are gathered. This collected data undergoes initial analysis to understand the distribution and range of parameters, which include factors like water-to-cement ratio, fly ash content, curing time, and aggregate ratios. The processed dataset is then split into training and testing subsets to facilitate model development and validation.

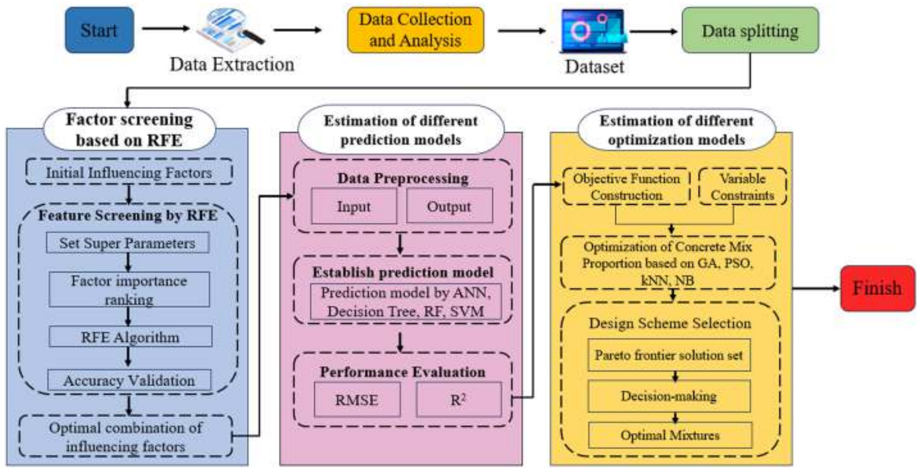


Fig. 3. Workflow for Predicting and Optimizing Concrete Mix Design Using Machine Learning and Optimization Algorithms

Factor screening is the first major step in refining the dataset to ensure that only the most influential features are considered in the prediction and optimization processes. This is achieved through Recursive Feature Elimination (RFE), a technique used to rank and select the most relevant factors that influence compressive strength. The RFE process starts with setting super parameters that control the elimination process, allowing the model to rank factors based on their relative importance. By systematically removing the least significant features, RFE isolates a subset of the most critical influencing factors. This optimized feature set undergoes accuracy validation to confirm that the selected factors contribute effectively to predictive accuracy, resulting in an optimal combination of influencing factors for subsequent modeling.

With the influential factors identified, the study proceeds to prediction model estimation. Data preprocessing is conducted to structure the dataset into distinct input and output variables, where input variables consist of the selected features from RFE, and the output variable is the compressive strength. Various machine learning models are then employed to predict compressive strength, including Artificial Neural Networks (ANN), Decision Trees (DT), Random Forest (RF), and Support Vector Machines (SVM). Each of these models has unique strengths: ANNs capture complex, non-linear relationships; Decision Trees are interpretable and handle categorical data well; Random Forests are robust due to ensemble learning; and SVMs are effective in high-dimensional spaces. The models are trained on the training dataset, and their performance is evaluated on the test dataset using metrics like Root Mean Square Error (RMSE), which measures prediction accuracy by penalizing large errors, and R-squared (R^2), which indicates the proportion of variance in compressive strength explained by the model. These metrics help identify the model with the best predictive performance for the data.

The final phase involves optimization model estimation to find the best mix proportions that maximize compressive strength. This is achieved by defining an objective function that aims to optimize compressive strength, with variable constraints based on practical limitations (e.g., minimum and maximum allowable values for water-to-cement ratio, fly ash content, etc.). Optimization techniques such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), k-Nearest Neighbors (kNN), and Naive Bayes (NB) are utilized to navigate the solution space. GA and PSO are particularly suited for this task due to their ability to explore a broad range of possible solutions and converge on optimal or near-optimal solutions by mimicking natural selection and swarm behavior, respectively. The optimization process iteratively adjusts the mix proportions, guided by the objective function, to produce a set of potential solutions.

After the optimization process, the study employs a Pareto frontier analysis to create a solution set that includes the most efficient trade-offs between compressive strength and other desired properties. The Pareto frontier provides a visual representation of solutions that are non-dominated, meaning that no other solution in the set is strictly better in all criteria. This frontier aids in design scheme selection, allowing decision-makers to choose from among the most balanced solutions based on project requirements. The final output of this process is a selection of optimal concrete mixtures that achieve high compressive strength while adhering to practical constraints, supporting the development of high-performance cementitious composites.

6 Prediction of Compressive Strength

Predicting the compressive strength of cementitious composites, such as concrete, is a crucial and complex task in the field of civil engineering. This process is essential for ensuring the structural integrity and safety of various civil structures, from buildings to bridges. Traditionally, empirical methods or experimental testing have been used to estimate compressive strength, but these methods often require extensive physical testing and can be time-consuming and resource intensive. Machine learning (ML) algorithms, however, present a powerful and efficient alternative, providing a means to forecast compressive strength with greater accuracy by analyzing relationships among various input parameters, including cement content, water-cement ratio, fly ash percentage, curing time, and other mix design variables.

In ML-based approaches, models are developed by training them on large datasets that contain known outcomes, such as specific compressive strength values corresponding to different combinations of input parameters. Through this training process, machine learning models learn complex relationships between these input factors and compressive strength. The resulting models can predict compressive strength for new mix designs with high accuracy, even in cases where interactions between variables are highly non-linear and complex. This capability makes ML models far more adaptable and precise than conventional empirical methods, which might overlook these intricate interactions.

One of the primary benefits of using machine learning techniques in predicting compressive strength is their ability to significantly reduce the reliance on extensive physical testing. Once the models have been adequately trained, they can offer fast and reliable predictions, allowing engineers to evaluate the strength potential of various mix designs quickly. This saves both time and resources during the concrete design process, as designers can test multiple combinations in a virtual environment before committing to actual physical trials. Furthermore, ML models provide the flexibility to evaluate compressive strength in relation to other critical parameters, such as cost efficiency and workability, enabling a more balanced approach to mix design optimization.

7 Optimization of Concrete Mix

Optimizing concrete mix proportions is a vital aspect of civil engineering, especially in ensuring that structures meet both performance and durability standards. Concrete, as a composite material, consists of cement, water, aggregates, and various supplementary materials like fly ash and silica fume. The proportions of these materials greatly influence the mechanical properties, including compressive strength (CS), durability, and resistance to environmental factors like frost or chloride ion penetration. In complex environments such as cold climates, oceanic settings, or underground infrastructures, designing optimal concrete mixes becomes more challenging, requiring precision and adaptability in the mix design process.

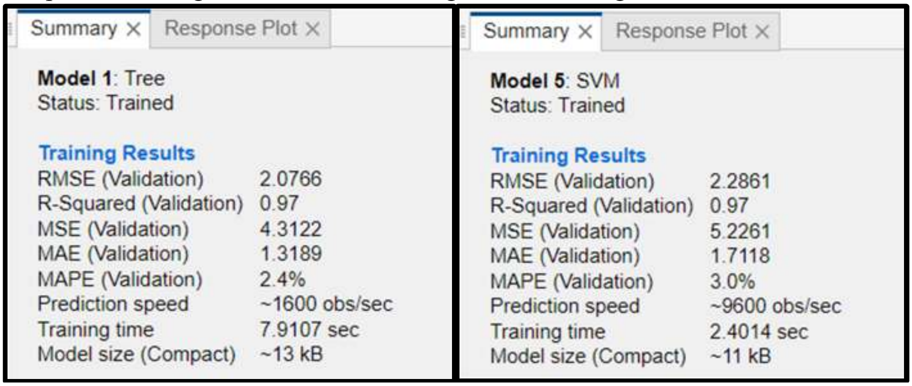
Recent advancements in optimization techniques, particularly the use of genetic algorithms (GA) and other machine learning (ML) methods, have revolutionized how concrete mix designs are optimized. Traditional trial-and-error approaches are often time-consuming and resource-intensive, whereas optimization algorithms enable the precise tuning of multiple variables to achieve the desired performance characteristics. Genetic algorithms, inspired by the principles of natural selection and evolution, offer a robust solution to the multi-objective optimization of concrete mixes. These algorithms work by iteratively selecting, combining, and mutating input parameters such as water-cement ratio, aggregate content, and supplementary material dosages, to find the most optimal mix that maximizes compressive strength, durability, and workability while minimizing costs and material use.

This project explores the application of genetic algorithms (GA) alongside other ML techniques such as Artificial Neural Networks (ANN) and Recursive Feature Elimination (RFE) for optimizing the mix design of concrete. By integrating GA into the design process, it becomes possible to navigate the complex, multi-variable landscape of concrete proportions, providing a more efficient and effective solution for maximizing performance metrics. Through the use of genetic algorithms, this study aims to develop an intelligent, data-driven approach for optimizing concrete mixes, improving compressive strength. This approach can greatly enhance the reliability and sustainability of concrete in demanding environments, while also contributing to advancements in non-destructive structural health monitoring (SHM) techniques.

8 Results

In this study, various machine learning models were employed to predict the compressive strength of cementitious composites, including the **Fine Regression Tree** and the **Support Vector Machine (SVM)**. These models were chosen for their ability to handle complex, non-linear relationships between the mix design parameters and the resultant compressive strength. The goal of using these models was to forecast the compressive strength based on key input factors such as cement, fly ash, water content, fine and coarse aggregates, and curing days.

The **Fine Regression Tree** is a decision tree algorithm that splits the data based on key features and learns relationships from the dataset to predict the compressive strength. It is particularly useful for interpreting which parameters have the most significant impact on strength. The **Support Vector Machine (SVM)**, on the other hand, is a robust algorithm known for its capacity to handle non-linear data by transforming the input data space to a higher dimension. This makes it effective in predicting the compressive strength based on a wide range of influencing factors.



(a) **(b)**

Fig. 4. (a). Summary and **(b).** Response Plot

Both models were trained on pre-processed data, which underwent steps such as cleaning, scaling, and feature selection to ensure the highest accuracy possible. The model results were evaluated using performance metrics like **R²** and **RMSE** (Root Mean Square Error), which indicate how well the models could generalize to new, unseen data. The accuracy of these models confirms their suitability for predicting the compressive strength in this study, enabling researchers to make quick and reliable predictions without relying solely on extensive physical testing.

In addition to prediction models, the study also implemented an optimization strategy using the **Genetic Algorithm (GA)**. GA is an evolutionary algorithm inspired by the principles of natural selection and genetics. It works by iteratively improving a population of solutions (in this case, different concrete mix designs) to achieve the optimal combination of factors that yield the maximum compressive strength.

The optimization process began with defining a population of mix designs, each representing a possible solution to the mix proportioning problem. Through selection, crossover, and mutation, GA searched for the optimal mix design that maximized compressive strength while considering material constraints and mix proportion limits. The fitness function used in this process evaluated each mix design's compressive strength, guiding the evolution toward better-performing combinations.

Editor - C:\Users\benya\OneDrive\Documents\MATLAB\Randomforest.m

Randomforest.m

```matlab
% Load Data
data = readtable('Data_220824.xlsx'); % Load your dataset
X = data(:, 1:7); % Features: cement, fine agg., coarse agg., water, SCMs
y = data(:, 8); % Target: compressive strength

% Split Data into Training and Testing Sets
cv = cvpartition(size(X, 1), 'HoldOut', 0.3);
X_train = X(training(cv), :);
y_train = y(training(cv), :);
X_test = X(test(cv), :);
y_test = y(test(cv), :);

% Train Random Forest Model
rng('default');
rf_model = fitrensemble(X_train, y_train, 'Method', 'Bag', 'NumLearningCycles', 100, 'Learners', 'tree');

% Define Objective Function for GA
objectiveFunction = @(x) predict(rf_model, x);

% Define Constraints (Bounds for Mix Proportions)
lb = [180.3, 0, 0, 185.45, 617.5, 686.67, 14]; % Lower bounds for each constituent
ub = [515.14, 100, 10, 185.45, 799, 888.25, 90]; % Upper bounds for each constituent

% Run Genetic Algorithm
options = optimoptions('ga', 'Display', 'iter', 'PopulationSize', 240, 'MaxGenerations', 480);
[x_optimal, fval] = ga(objectiveFunction, 7, [], [], [], [], lb, ub, [], options);

% Predict Optimal Mix Strength
optimal_strength = predict(rf_model, x_optimal);

% Display Results
fprintf('Optimal Mix Proportions:\n Cement: %.2f\n Fly ash: %.2f\n SF: %.2f\n Water: %.2f\n Coarse Aggregate: %.
fprintf('Predicted Compressive Strength: %.2f MPa\n', optimal_strength);
```

Fig. 5. MATLAB Code for Optimization of Concrete Mix using GA

The results from GA optimization revealed significant improvements in compressive strength compared to standard mix designs. By fine-tuning the cement content, fly ash proportion, water-cement ratio, and other key parameters, the optimized mix design not only maximized strength but also balanced cost-effectiveness and material performance. The implementation of GA in this project highlights the potential of advanced optimization techniques in refining concrete mix designs beyond conventional empirical methods.

```
Optimal Mix Proportions:
 Cement: 310.92
 Fly ash: 89.18
 SF: 0.44
 Water: 185.45
 Coarse Aggregate: 660.93
 Fine Aggregate: 736.46
 Curing Days: 14.00
Predicted Compressive Strength: 40.25 MPa
fx >>
```

Fig.6. Optimal Mix Proportions

9 Conclusions

The study emphasizes the potential of machine learning (ML) techniques in advancing the field of civil engineering, particularly in predicting and optimizing the compressive strength of cementitious composites such as concrete. By employing ML algorithms, namely the Fine Regression Tree and Support Vector Machine (SVM), the research successfully demonstrates that these models can handle complex, non-linear relationships between various input parameters, including cement content, fly ash ratio, water content, aggregate proportions, and curing time. These models outperform traditional empirical methods by offering more precise and faster predictions, which significantly reduce the reliance on extensive and resource-intensive physical testing. This not only saves time but also makes it feasible to evaluate a wide range of mix designs in the virtual space before committing to physical trials.

The study further highlights the importance of data preprocessing, including cleaning, scaling, and feature selection, to enhance the models' predictive accuracy. These preparatory steps ensure that the input data aligns well with the model requirements, reducing errors and allowing the ML models to generalize effectively to new, unseen data. Through rigorous evaluation using metrics such as R^2 and RMSE (Root Mean Square Error), the research confirms that the Fine Regression Tree and SVM are well-suited for compressive strength prediction, offering reliable outputs that civil engineers can use with confidence in practical applications.

In addition to prediction, the study introduces an optimization strategy for concrete mix design, using the Genetic Algorithm (GA). GA, inspired by the natural principles of evolution and genetics, plays a crucial role in identifying the optimal mix design that maximizes compressive strength while meeting additional requirements such as cost-effectiveness and durability. GA operates by generating an initial population of mix designs and iteratively refining this population through processes akin to biological evolution, including selection, crossover, and mutation. Each mix design, or "individual," is evaluated for its performance based on a fitness function that considers compressive strength and other key attributes. Through multiple iterations, GA converges towards an optimized solution that balances various design objectives.

The optimization results from the GA revealed significant improvements in compressive strength compared to standard, non-optimized mix designs. By fine-tuning key parameters such as cement content, fly ash proportion, and water-cement ratio, the GA not only increased the compressive strength of the concrete but also provided a balanced mix design that aligns with project-specific requirements, such as cost constraints and desired durability. This marks an important advancement over conventional trial-and-error approaches to mix design, as GA enables a more systematic and targeted search for the optimal mix composition, considering multiple objectives and constraints simultaneously.

This research thus validates the integration of ML models and optimization algorithms as effective tools for modern concrete design. ML techniques provide reliable predictions of compressive strength, enabling engineers to quickly assess various mix combinations without the need for exhaustive physical testing. Meanwhile, optimization algorithms like GA help refine these mix designs, ensuring that they not only meet compressive strength requirements but also align with broader project goals. This combined approach offers a more data-driven, efficient, and adaptive solution to concrete design challenges, paving the way for more resilient and cost-effective concrete structures.

Additionally, the study highlights the adaptability of ML models, which can be retrained and updated as more data becomes available. This means that with continuous data collection from ongoing projects, the models will become increasingly robust and versatile, making them suitable for a wider range of concrete types and environmental conditions. As such, the research establishes a foundation for future work in the field, with potential extensions into areas such as sustainability, where ML and optimization can help develop concrete mixes that balance environmental impact with performance. The approach also supports advancements in non-destructive testing techniques, contributing to enhanced structural health monitoring capabilities that ensure long-term durability and safety.

In conclusion, this research successfully demonstrates that integrating ML and optimization algorithms into concrete mix design and strength prediction can significantly improve predictive accuracy, reduce resource consumption, and facilitate a more flexible, data-driven approach to civil engineering. The findings underscore the potential for these techniques to transform traditional methods, making concrete design more efficient, adaptable, and aligned with modern engineering demands. These advancements offer promising implications for the future of sustainable and resilient concrete structures.

Disclosure of Interests. Authors do not have any conflicts among each other.

References

1. S. Kumar, R. Kumar, B. Rai, and P. Samui, Prediction of compressive strength of high-volume fly ash self-compacting concrete with silica fume using machine learning technique, Elsevier, Construction and Building Materials,438(9),(2024).
2. H. Song, A. Ahmad, F.Farooq, K.A. Ostrowski, M.Maslak, S.Czarnecki, and F. Aslam, Predicting the compressive strength of concrete with fly ash admixture using machine learning algorithms, Elsevier, Construction and Building Materials,308(15), (2021).
3. S. Yang, H. Chen, Z. Feng, Y. Qin, J. Zhang, Y. Cao, and Y. Liu, Intelligent Mult objective optimization for high-performance concrete mix proportion design: A hybrid machine learning approach, Engineering Applications of Artificial Intelligence, 126, (2023).
4. P.V.R.K. Reddy, and D. Ravi Prasad, The role of graphene oxide in the strength and vibration characteristics of standard and high-grade cement concrete, Journal of Building Engineering, 63 (A), (2023).
5. W. Dong, Y. Huang, B.Lehane, and G. Ma, Multi-objective design optimization for graphite-based nanomaterials reinforced cementitious composites: A data-driven method with machine learning and NSGA-II, Elsevier, Construction and Building Material, 331, (2022)