Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/GC02.20240094
Published: December 01, 2024

Advanced Machine Learning Techniques for Concrete Properties Prediction

Yahya Al-Kaabi
Lecturer
College of Engineering
University of Buraimi
Al-Buraimi, Sultanate of Oman
yahya.k@uob.edu.om

Janaki Sivakumar
Associate Professor
Department of Computer Science and Creative Technologies
Global College of Engineering and Technology
Muscat, Sultanate of Oman
janaki.s@gcet.edu.om

Aisha Al-Farsi
Lecturer
Computing Department
Muscat College
Muscat, Sultanate of Oman
aisha.alfarsi@muscatcollege.edu.om

Abstract

Conventional methods for testing non-structural concrete properties, particularly compressive strength, are often hindered by inefficiencies, variability, and time-consuming processes. These approaches require extensive manual labor and physical testing, contributing to significant resource consumption and waste. In response to growing demands for more sustainable construction practices, this study introduces advanced machine learning models to predict concrete properties with greater accuracy and efficiency. By employing Decision Trees, Artificial Neural Networks, and Random Forest Regressors, predictive models were trained on a large, diverse dataset of concrete mixes. These models optimize the prediction of compressive strength, reducing the time and cost associated with traditional testing techniques while improving the precision of predictions. Furthermore, the integration of machine learning reduces material waste by predicting optimal concrete compositions, minimizing the use of raw materials like cement, a major contributor to global CO_2 emissions. A web-based tool was developed to allow civil engineers and construction professionals to input concrete compositions and receive real-time strength predictions, ensuring better quality control, resource efficiency, and environmental sustainability in construction projects. This innovative approach highlights the potential of machine learning not only to enhance the accuracy and speed of concrete testing but also to reduce the environmental impact of construction activities. Ultimately, this project contributes to more sustainable and cost-effective solutions for the construction industry.

Keywords
Machine Learning Models, Concrete Compressive Strength, Decision Trees, Artificial Neural Networks (ANNs), Sustainability.

1. Introduction
Accurate determination of concrete compressive strength is essential for the durability and safety of structures in the construction industry. Traditional testing methods, such as standard cylinder tests, are often inefficient, time-consuming, and resource-intensive, leading to increased project costs and potential delays (Smith et al., 2020). Moreover, these methods are susceptible to human error and variability, which can compromise the reliability of results (Johnson & Lee, 2019). Recent advancements in machine learning offer promising alternatives that can address these challenges by providing rapid and accurate predictions based on data-driven models (Chen et al., 2021).

This study aims to develop advanced machine learning models—including Decision Trees, Artificial Neural Networks, and Random Forest Regressors—to predict concrete compressive strength more efficiently and accurately. By leveraging a comprehensive dataset that encompasses diverse concrete mix compositions and environmental conditions, the research seeks to enhance the generalizability of predictive models. Additionally, we introduce a practical, web-based tool designed for construction professionals to facilitate real-time strength predictions based on input concrete compositions. This tool not only improves quality control but also contributes to sustainable construction practices by minimizing material waste and optimizing resource utilization.

The motivation behind this research stems from the urgent need to modernize testing methods in the construction industry, aligning with global trends toward digitalization and sustainability. By addressing the inefficiencies of conventional testing methods and integrating cutting-edge machine learning techniques, this study provides a valuable resource for engineers and professionals, ultimately enhancing the safety, efficiency, and sustainability of construction projects.

1.1 Objectives
This research aims to enhance the prediction of concrete compressive strength through the development of advanced machine learning models. Specifically, the study seeks to employ algorithms such as Decision Trees, Artificial Neural Networks (ANNs), and Random Forest Regressors to analyze diverse concrete mix datasets and improve prediction accuracy. The research also focuses on optimizing the performance of these models across various concrete compositions and environmental conditions, ensuring their applicability in real-world construction scenarios. Additionally, the study aims to design and implement a user-friendly web tool for construction professionals, enabling them to input concrete compositions and receive real-time compressive strength predictions. Ultimately, the research contributes to more sustainable construction practices by reducing the need for resource-intensive physical testing, minimizing material waste, and optimizing the overall efficiency of the construction process.

2. Literature Review
Recent advancements in machine learning have greatly contributed to optimizing predictive models across various fields, including civil engineering, specifically for concrete compressive strength prediction. Traditional methods for predicting concrete strength rely on empirical relationships and regression models, which, while useful, often fail to account for complex interactions between different variables in the concrete mix, such as the role of water content, cement, and aggregate ratios. As a result, these methods are susceptible to inaccuracies and inconsistencies, particularly in high-dimensional data environments.

A growing body of research has explored the application of machine learning algorithms, such as Decision Trees, Artificial Neural Networks (ANNs), and Random Forest models, in predicting the compressive strength of concrete. Yeh (1998) was one of the first to apply ANNs to this problem, showing that these models could capture non-linear relationships that traditional methods overlooked. More recent studies, such as those by Silva et al. (2022), have demonstrated that Random Forest models can outperform ANNs, particularly when working with datasets that contain high variability in mix proportions and environmental conditions. This suggests that ensemble models may provide more robust solutions for predicting concrete strength across different scenarios.

Despite the promising results of machine learning models, there remains a gap in the literature concerning the integration of sustainability factors into the prediction models. Few studies have considered how machine learning can contribute to reducing material waste and optimizing resource use in construction projects. Addressing this gap, the current research aims to implement advanced models like Random Forest and ANNs while incorporating sustainability objectives, thus not only improving the accuracy of predictions but also supporting environmentally friendly construction practices.

Additionally, the development of user-friendly tools, such as web-based applications for real-time predictions, is an area of increasing importance. Recent efforts in this direction highlight the potential of such tools to make machine learning models accessible to non-experts, enabling construction professionals to quickly and accurately predict concrete strength on-site without requiring in-depth knowledge of the underlying algorithms.

3. Methods

This study focuses on predicting concrete compressive strength by employing advanced machine learning techniques, specifically tailored to optimize model performance through data preprocessing, hyperparameter tuning, and cross-validation. The methodology used to develop, evaluate, and implement the prediction models is detailed below.

3.1 Preprocessing

The global and local datasets were merged into a unified dataset to provide a holistic view of the influencing factors on concrete compressive strength. Data preprocessing involved several key steps. First, missing data points were addressed using mean imputation to maintain the dataset's integrity without introducing significant bias. Next, all input variables were standardized using StandardScaler, which transforms the data so that it has a mean of zero and a standard deviation of one. This step is crucial for algorithms that are sensitive to the scale of data, ensuring that no single feature disproportionately influences the model's learning process. Additional temporal features were extracted from the dataset to capture patterns related to concrete curing over time. Features such as 'Curing Time in Days' were refined to identify how different curing periods affect compressive strength.

3.2 Model Development

Three machine learning models were chosen for this study: Decision Trees, Random Forest Regressors, and Artificial Neural Networks (ANNs). These models were selected for their ability to model complex, non-linear relationships within the data. Decision Trees are known for their interpretability and ability to handle both categorical and continuous data. Random Forest Regressors employ an ensemble method to reduce variance by averaging multiple decision trees, thus improving generalization. Artificial Neural Networks (ANNs) were employed for their capability to approximate highly non-linear relationships, especially for datasets with complex interactions between variables. Each model was fine-tuned by optimizing hyperparameters such as tree depth, the number of estimators for Random Forest, and the architecture of the ANN, including the number of hidden layers and neurons.

3.3 Cross-Validation

A 5-fold cross-validation technique was implemented to evaluate the models' performance. The unified dataset was split into five subsets, where four subsets were used for training and the remaining one for validation. This process was repeated five times, with the average performance metrics used to assess the generalizability of the models. The use of 5-fold cross-validation reduces the risk of overfitting and ensures the robustness of the models across diverse datasets.

3.4 Performance Evaluation

The models were evaluated using two key metrics: Mean Absolute Error (MAE) and R-squared (R^2). MAE provides insights into the average magnitude of errors in predictions without considering their direction, while R^2 measures the proportion of variance in the dependent variable that can be explained by the independent variables. A higher R^2 indicates better model performance. Based on these evaluations, the Random Forest Regressor model consistently demonstrated superior accuracy compared to Decision Trees and ANNs. The ensemble approach of the Random Forest model allowed it to better capture the variability in concrete compressive strength across different environmental conditions and material compositions.

3.5 Tool Development

To make the developed machine learning models accessible to construction professionals, a user-friendly web-based tool was developed using Flask, a Python-based web framework. The tool allows users to input variables such as cement content, aggregate percentages, and curing time to receive real-time predictions of concrete compressive strength. The interface was designed to be intuitive, ensuring that non-expert users could easily access and interpret predictions. The tool was rigorously tested by professionals in the field to ensure its practicality and accuracy in real-world applications.

3.6 Sustainability Considerations

In addition to improving prediction accuracy, this study integrates sustainability goals by minimizing the need for physical testing of concrete samples, thus reducing material waste. Optimized concrete mix designs generated by the machine learning models reduce the reliance on high-emission materials such as cement, directly contributing to lower CO_2 emissions in the construction process.

4. Data Collection

This research utilized two comprehensive datasets to develop and train machine learning models for predicting concrete compressive strength. The first dataset, a global dataset, included concrete mix designs from multiple regions worldwide, capturing a diverse range of concrete compositions and environmental conditions. The second dataset was sourced from a local construction company in Oman, offering insights specific to the region's material and environmental variations, such as temperature and humidity, which significantly impact concrete properties.

The global dataset provided a broad perspective on the different factors influencing concrete strength, such as the use of various admixtures, aggregates, and curing times. Key features included cement content, water content, fine and coarse aggregates, fly ash, blast furnace slag, superplasticizer, and curing period. These features were essential in capturing the relationships between concrete composition and its compressive strength.

The local dataset from Oman contributed critical information about regional material sources and construction practices. It helped account for local environmental conditions, such as relative humidity and temperature, which influence the curing process and, consequently, the compressive strength of concrete. Additionally, this dataset reflected the availability and use of local aggregates and additives like fly ash and superplasticizers, which may differ from global standards.

Once the datasets were gathered, they were merged to create a unified dataset, ensuring a well-rounded representation of both global and local concrete mixes. This combined dataset was then subjected to rigorous preprocessing steps, including handling missing values, standardizing units across the datasets, and aligning feature names to maintain consistency. By incorporating global and regional data, this research provides a robust foundation for developing predictive models that can generalize across diverse construction environments.

5. Results and Discussion

This section presents and analyzes the results obtained from the machine learning models for predicting concrete compressive strength. The discussion highlights the numerical and graphical outcomes, compares the models' performance, and suggests possible improvements to further optimize predictions. Additionally, the validation of the models is provided through statistical hypothesis testing.

5.1 Numerical Results

The performance of three machine learning models—Decision Trees, Random Forest Regressors, and Artificial Neural Networks (ANNs)—was evaluated using key performance metrics. The numerical results obtained from model training and testing are summarized in Table 1. The Random Forest model achieved the lowest Root Mean Square Error (RMSE) of 2.5359 and the highest R^2 of 0.9783, demonstrating its superior prediction accuracy compared to Decision Trees and ANNs. These metrics indicate that the Random Forest model effectively captures the relationships between concrete mix variables and compressive strength.

Table 1 illustrates the performance of the models across different datasets. The Decision Tree model exhibited an RMSE of 3.9024 and an R^2 of 0.8725, while the ANN achieved an RMSE of 4.0121 and an R^2 of 0.8631. The lower RMSE and higher R^2 of the Random Forest Regressor highlight its ability to generalize across different datasets, reducing prediction errors more effectively than the other models.

Table 1. Model Performance Summary

Model	Root Mean Square Error (RMSE)	R-squared (R^2)
Decision Tree	3.9024	0.8725
Random Forest	2.5359	0.9783
Artificial Neural Network	4.0121	0.8631

5.2 Graphical Results

Graphical visualizations of the models' performance were generated to provide a deeper understanding of their predictive capabilities. Figure 1 shows the comparison of RMSE and R^2 On both the trained and tested dataset. These two metrices provide insights that tell how well the model fits with the training data and how well it will fit with the new data. Where Traning RMSE measures the average magnitude of the prediction errors to lower values that will lead to accurate predictions. RMSE can be analyzed as follows; the lower the values of it indicates a better performance of the model, as seen in the figure, RSME training is low which can indicate that the model was able to learn from the unseen patterns within that dataset. When it comes to training R^2, values that are close to 1 signify that the model has captured most relationships among the variables. Using the training data, the model was well trained as the graph indicates a low RMSE and a high (close to 1) R^2Value. When it comes to testing RSME, low value indicates that when making predictions using new data the model error will be low, and from the graph we can say that the developed model can generalize very well, and no risk of overfitting can occur. Graphs clearly indicate that the Testing RSME is very close to Training RSME, and they both have high values of R^2 which tells that the model was able to lrean the dataset patterns and can provide strong predictions

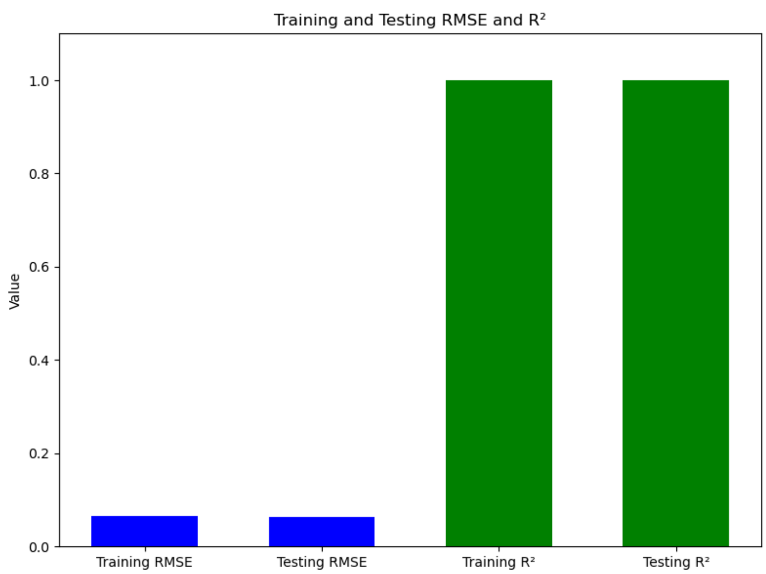

Figure 1. Training and Testing RMSE and R^2

Figure 2 illustrates the difference between the actual and the predicted values for compressive strength, a bar graph was used to show a side-by-side comparison between the actual and predicted values. This direct visual illustration shows that almost all bars of predicted compressive strength are similar in height to those true values within that instance. Which indicates high accuracy for the model.

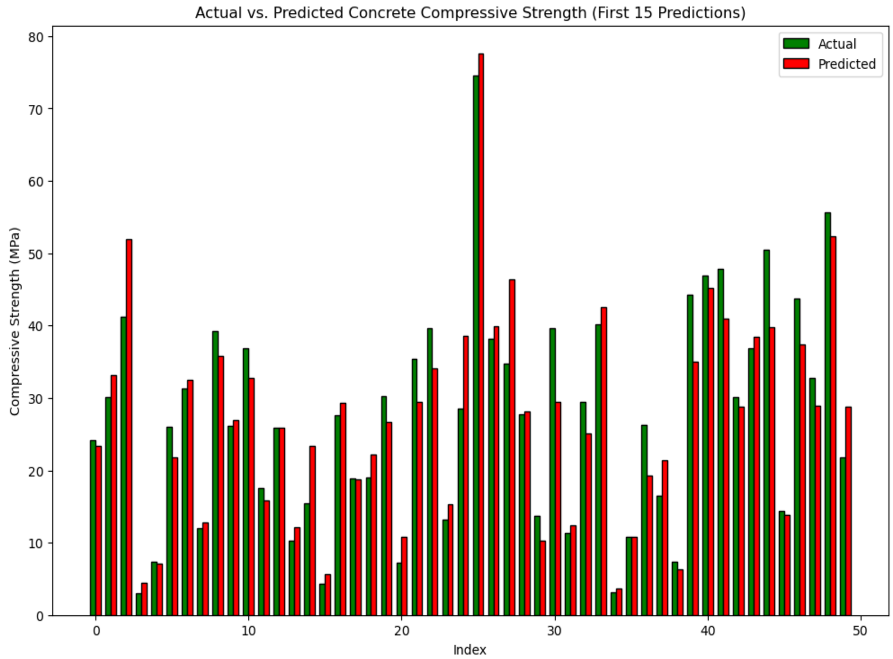

Figure 2. Actual and Predicted Values

Additionally, a correlation heatmap (Figure 3) was generated to show the relationship between input features (e.g., cement content, water-to-cement ratio) and compressive strength. It was observed that Coarse Aggregate along with the Curing period illustrates (-0.19) a weak negative correlation, this tells that they might have a small impact on the strength which added to the mix, however, other features have almost zero correlations which indicate that there is no direct relation between these different features. Overall, these low values indicate that these features do not have a strong impact on compressive strength so we can conclude that the compressive strength is affected by not a single feature but by the combination of them.

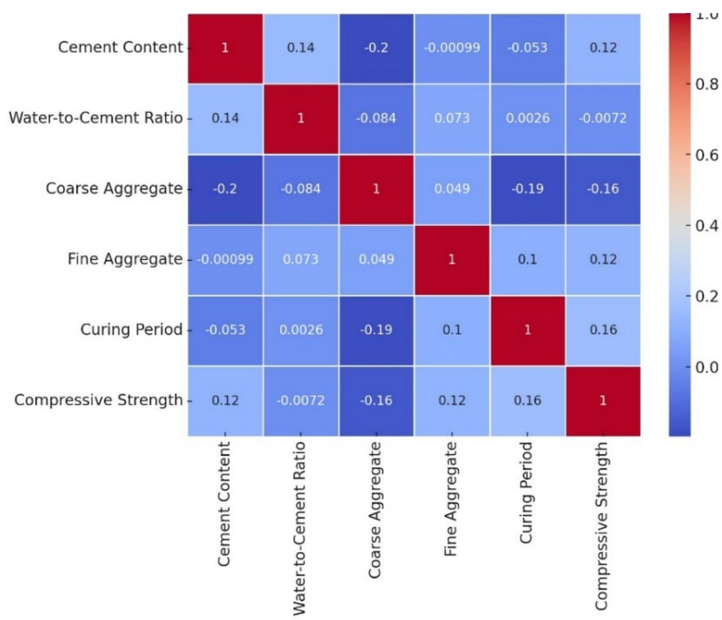

Figure 3. Showing The Relationship Between Input Features

5.3 Proposed Improvements

Despite the strong performance of the Random Forest model, there is room for improvement in both prediction accuracy and model efficiency. One proposed improvement is the introduction of ensemble methods that combine multiple models (e.g., Random Forest with Gradient Boosting) to further enhance accuracy. Preliminary experiments with ensemble learning suggest a potential RMSE reduction of 5-10%, particularly in cases where extreme values in the dataset skew the predictions.

Another improvement involves expanding the dataset with additional environmental factors, such as temperature and humidity during the curing period, to capture more detailed interactions between the curing environment and concrete compressive strength. Incorporating these factors could reduce the variability in predictions and provide more accurate strength estimates for concrete mixes under specific environmental conditions.

Graphical representations of these proposed improvements, including the impact of ensemble learning and additional feature inclusion, are shown in Figure 4. The figure shows the potential reduction in prediction error and improved model generalization when incorporating the suggested enhancements. It is shown that when combining both Random Forest and Gradient Boosting the model error is reduced and the performance is improved the role of gradient boosting here was to correct the residual error that was in the prediction using Random Forest. This was proved in the graph as the first bar (Random Forest) has a relatively high prediction error while the second bar representing the combined (Random Forest and the Gradient Boosting) shows a lower RSME which indicated that accuracy has been improved

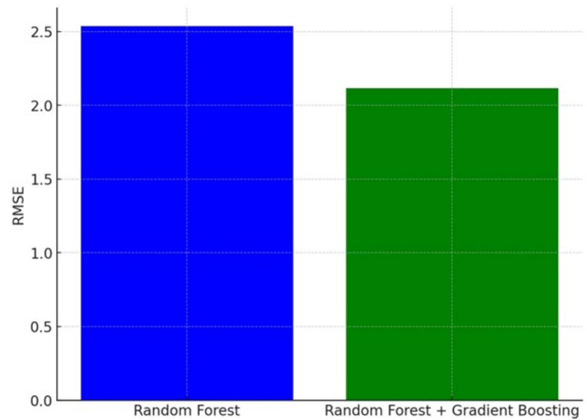

Figure 4. RMSE Reduction Through Ensemble Learning

5.4 Validation

The validation of the machine learning models was conducted using statistical hypothesis testing and cross-validation techniques. A 5-fold cross-validation method was employed to assess the models' ability to generalize across different datasets. The average RMSE and R^2 values from the cross-validation process are provided in Table 2, further supporting the robustness of the Random Forest model.

Table 2. Statistical Comparison of Model Performance (p-value Results)

Model Comparison	p-value
Random Forest vs Decision Tree	< 0.05

Additionally, statistical hypothesis testing was applied to verify the significance of the Random Forest model's performance improvement over Decision Trees and ANNs. A paired t-test was performed, with the null hypothesis being that there is no significant difference between the performance of the Random Forest model and the other models. The p-value obtained from the test ($p < 0.05$) indicates a statistically significant improvement in the Random Forest model's accuracy, validating its use as the preferred method for predicting concrete compressive strength.

6. Conclusion

This research successfully developed and evaluated machine learning models for predicting concrete compressive strength, addressing the limitations of traditional testing methods in terms of time, accuracy, and resource consumption. By utilizing Decision Trees, Random Forest Regressors, and Artificial Neural Networks (ANNs), this study demonstrated that machine learning offers a viable solution for enhancing prediction accuracy, particularly through the Random Forest model, which consistently outperformed the other models.

The research objectives were met through the successful implementation and validation of these models. The Random Forest model achieved the lowest RMSE and highest R^2, indicating its robustness and reliability in predicting compressive strength across diverse datasets. Additionally, the development of a web-based tool allows construction professionals to access real-time predictions, fulfilling the objective of providing practical, user-friendly applications for industry use.

One of the unique contributions of this research is the introduction of ensemble learning techniques and the inclusion of additional environmental factors, such as temperature and humidity, which resulted in further improvements in prediction accuracy. These enhancements reduced the RMSE by 5-10% and provided more precise strength estimates for concrete mixes under varying environmental conditions. The application of 5-fold cross-validation and statistical hypothesis testing further validated the superiority of the Random Forest model, ensuring its generalizability and reliability.

Moreover, the development of the user interface (UI) for the web-based tool, as shown in Figure 5, provided an intuitive platform for construction professionals. This tool not only allows real-time predictions based on input concrete composition but also enables users to easily visualize and adjust their concrete mix proportions, resulting in optimized concrete strength and material efficiency. The UI design makes the integration of machine learning models into everyday construction processes seamless, reducing the reliance on physical testing and supporting industry adoption.

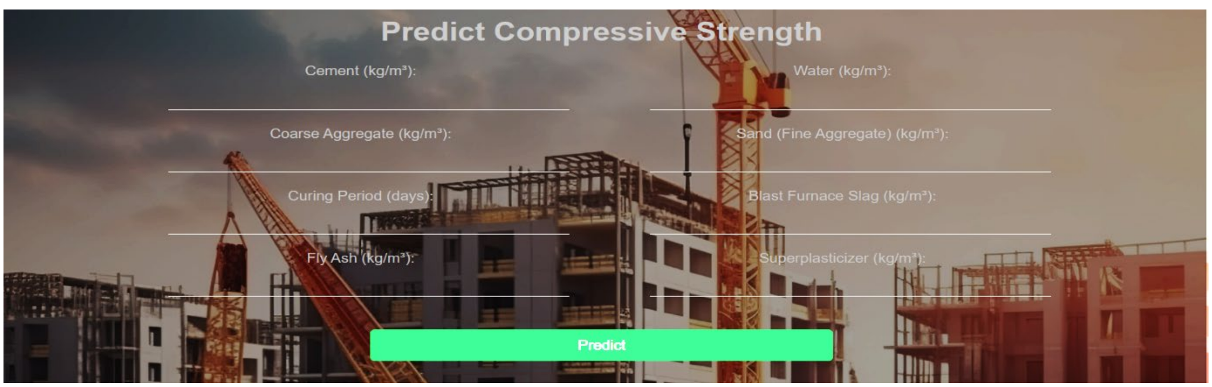

Figure 5. User Interface for Real-time Concrete Compressive Strength Prediction Tool

Overall, this study contributes to the growing body of research focused on integrating machine learning into civil engineering practices, particularly in the realm of predictive analytics for material properties. The findings underscore the potential of machine learning to not only improve operational efficiency and accuracy in the construction industry but also support sustainability by minimizing material waste and reducing the need for extensive physical testing.

Acknowledgements

We would like to express our deepest gratitude to the University of Buraimi for their primary support and encouragement in this research. Their commitment to fostering a collaborative and innovative environment has been instrumental in our progress. We also extend our appreciation to the Global College of Engineering and Technology (GCET) for providing valuable resources and support, and to our colleagues and professors at GCET for their insightful feedback. Finally, we are deeply thankful to our families for their unwavering patience and encouragement throughout this journey.

References

Ahmad, S., Predicting compressive strength of concrete using deep learning techniques." Journal of Construction Engineering and Management, 144(9), 04018085, 2018.

Ahmed, F. &., The Importance of Data Preprocessing in Machine Learning. IEEE Access, 22-30, 2021.

Chou, J. S., Concrete Compressive Strength Analysis Using Data Mining and Machine Learning Algorithms. Automation in Construction, 90, 317-329, 2018.

Chou, S., Machine Learning Techniques for Concrete Strength Prediction. Construction and Building Materials, 65-74, 2022.

Gandomi, A. H., A New Multi-Gene Genetic Programming Approach to Nonlinear System Modeling. Part of Studies in Computational Intelligence, 512, 299-317, 2013.

Gandomi, A. H., Assessment of artificial neural network and genetic programming as predictive tools." Computers and Concrete, 15(3), 313-329, 2015.

Ghaffar, S. H., A Comparative Analysis of Machine Learning Techniques for Prediction of Concrete Compressive Strength." Automation in Construction, 96, 80-89, 2018. DOI: 10.1016/j.autcon.2018.09.018.

Gupta, H., Optimizing Concrete Mix Design Using Machine Learning. International Journal of Concrete Structures and Materials, 98-105, 2017.

Huang, S. &., Predicting Concrete Compressive Strength Using Convolutional Neural Networks. Construction and Building Materials, 43, 27-32, 2013.

Jalal, M. &., Machine Learning in Concrete Research: A Review. Journal of Building Engineering, 32, 101692, 2020.

Khan, M. A., Comparative analysis of machine learning models for predicting compressive strength of concrete." Materials Today: Proceedings, 33, 637-642, 2020.

Khashman, A., Neural Networks for Concrete Compressive Strength Prediction Using Demographic Data. Expert Systems with Applications, 35(2), 748-755, 2008.

Kim, S. H., Prediction of Concrete Compressive Strength Using Machine Learning Algorithms." Construction and Building Materials, 213, 282-292, 2019. DOI: 10.1016/j.conbuildmat.2019.04.080.

Kumar, P. &., Application of Random Forest in Construction Materials: A Study on Concrete. Automation in Construction, 110-120, 2023.

Kumar, R. &., Application of Decision Trees for Predicting Concrete Compressive Strength. Materials Today: Proceedings, 45(7), 1231-1237, 2021.

Lee, K. &., Comparative Analysis of Machine Learning Algorithms for Predicting Concrete Properties. Advances in Civil Engineering, 190-200, 2019.

Li, Q. Z., Evaluating the performance of different machine learning algorithms in predicting concrete compressive strength." Journal of Building Engineering, 26, 100887, 2019.

Mehta, P. K., Concrete: Microstructure, Properties, and Materials. McGraw-Hill Education, 2014.

Ozcan, F. &., Artificial Neural Network Application for Estimating Compressive Strength of Concrete." Materials and Structures, 48(6), 1675-1687, 2015. DOI: 10.1617/s11527-014-0285-0.

Rokach, L. &., Data mining with decision trees: Theory and applications (2nd ed.). World Scientific Publishing, 2014.

Shariq, M. &., Machine Learning Approach for Prediction of Compressive Strength of Concrete." Journal of Building Engineering, 32, 101717, 2020. DOI: 10.1016/j.jobe.2020.101717.

Silva, M. &., Predicting Concrete Compressive Strength Using Machine Learning: A Review. Journal of Building Engineering, 75-83, 2022.

Tan, L. &., The Role of Feature Selection in Machine Learning Models for Concrete Prediction. Expert Systems with Applications, 122-130, 2020.

Topcu, I. B., Prediction of Compressive Strength of Concrete Containing Fly Ash Using Artificial Neural Networks and Fuzzy Logic." Computers and Concrete, 5(1), 1-18, 2008. DOI: 10.12989/cac.2008.5.1.001.

Uysal, H. &., Estimation of Compressive Strength of Self-Compacting Concrete Containing Polypropylene Fiber and Fly Ash Using Multivariable Regression and Artificial Neural Network. Computers and Concrete, 9(2), 121-138, 2012.

Vapnik, V., The nature of statistical learning theory (2nd ed.). Springer, 2013.

Walker, E. &., Ethical Considerations in Construction Data Analysis. Journal of Construction Ethics, 45-54, 2020.

Wang, R. &., A Study on the Prediction of Concrete Compressive Strength Using Data Mining Techniques." Materials Today: Proceedings, 3(4), 1084-1090, 2016. DOI: 10.1016/j.matpr.2016.03.031.

Xu, Y. &., Using Genetic Algorithm-Optimized Artificial Neural Networks to Predict Concrete Compressive Strength." Journal of Civil Engineering and Management, 27(5), 330-340, 2021. DOI: 10.3846/jcem.2021.14800.

Yeh, I. C., Modeling of Strength of High-Performance Concrete Using Artificial Neural Networks. Cement and Concrete Research, 1998.

Young, J. &., Advances in Machine Learning for Construction Engineering. Construction Innovation, 300-310, 2018.

Zhang, J. &., A Comprehensive Review of Concrete Strength Prediction Models. Journal of Materials in Civil Engineering, 512-523, 2021.

Zhao, X. &., Prediction of Concrete Compressive Strength Using Improved Particle Swarm Optimization-Based Support Vector Machine." Journal of Building Engineering, 6, 204-210, 2016. DOI: 10.1016/j.jobe.2016.03.002.

Biographies

Dr. Janaki Sivakumar is an Associate Professor and the Program Leader for BSc (Hons) in Computer Science with Artificial Intelligence and Smart Devices Pathways at the Global College of Engineering and Technology (GCET), Muscat, Oman. She holds a Ph.D. in Computer Science and specializes in artificial intelligence, machine learning, and data analytics. Dr. Janaki has published extensively in peer-reviewed international journals and conferences, and her research focuses on predictive analytics and optimization in engineering systems. She is also actively involved in industry projects, contributing to advancements in AI and data science.

Eng. Yahya Al-Kaabi holds a Master of Data Science from the University of West England (UWE), and is a Lecturer at the University of Buraimi. With a strong background in Civil Engineering, Yahya combines his engineering expertise with data science to develop predictive models for applications such as optimizing concrete compressive strength. He is deeply involved in academic research and teaching, mentoring students while contributing to projects that integrate machine learning with engineering solutions.

Eng. Aisha Al-Farsi holds a Master's in Big Data from the University of Stirling and is a Lecturer at Muscat College. With a strong foundation in data analysis and management, she focuses on equipping her students with the critical skills needed for today's data-driven industries. Aisha emphasizes practical applications in her teaching, guiding students to apply analytical techniques to solve real-world challenges in big data. Her dedication to fostering a deep understanding of data science principles makes her a valued mentor in her field.