

A NEURAL NETWORK APPROACH FOR COMPRESSIVE STRENGTH PREDICTION OF ULTRA-HIGH-PERFORMANCE CONCRETE: DEVELOPMENT, ANALYSIS, AND ROBUSTNESS TESTING

Dr. M. Adil Khan^{*1}, Wasim Khan², Jamal khan Lodhi³, Shah Khalid⁴,
Muhammad khubaib⁵, Fawad Iqbal⁶

^{*1}Resident Engineer, NESPAK

^{2,4,5}Department of Civil Engineering, University of Engineering and technology Peshawar Pakistan

⁵Department of Civil Engineering, Department of Civil Engineering, Cecos University of IT and Emerging Science Peshawar.

⁶School of Civil and environmental engineering (SCEE), National university of science and technology Islamabad (NUST)Pakistan

¹adee.uol@gmail.com, ²wasim.khan.utmanzai@gmail.com, ³22pwciv5898@uetpeshawar.edu.pk, ⁴sk7190518@gmail.com, ⁵22pwciv5809@uetpeshawar.edu.pk, ⁶18pwciv5041@uetpeshawar.edu.pk

DOI: <https://doi.org/10.5281/zenodo.17919023>

Keywords

Ultra-High-Performance Concrete (UHPC), Artificial Neural Network (ANN), Compressive Strength Prediction, Machine Learning, Model Interpretability (SHAP), Feature Importance Analysis

Article History

Received: 11 October 2025

Accepted: 21 November 2025

Published: 13 December 2025

Copyright @Author

Corresponding Author: *

Dr. M. Adil Khan

Abstract

The development of Ultra-High-Performance Concrete (UHPC) is constrained by the empirical and costly nature of mixture optimization, necessitating robust predictive models that capture the complex, non-linear interactions between its constituent materials. While Artificial Neural Networks (ANNs) have been widely adopted for this purpose, their application is frequently criticized for lack of transparency, inconsistent interpretability, and uncertain generalization beyond training datasets. This study addresses these critical gaps by developing, analysing, and rigorously validating a fully connected ANN model to predict UHPC compressive strength using a substantial and curated dataset of 810 mixture designs. The model, featuring a single hidden layer with 100 neurons and L2 regularization, demonstrated high predictive accuracy on an independent test set, achieving an R^2 of 0.922, RMSE of 11.267 MPa, and a mean absolute percentage error (MAPE) of 8.305%. Twenty-fold cross-validation confirmed robustness, revealing consistent performance across most folds while identifying specific regions in the mixture design space where predictive accuracy diminished, highlighting data limitations. A critical, multi-method feature importance analysis—employing Univariate Regression, RRelieff, and SHapley Additive exPlanations (SHAP)—provided a nuanced hierarchy of influential variables. While filter methods ranked silica fume and fiber content highest, SHAP analysis revealed the ANN's internal logic prioritizes curing age, cement content, and fiber dosage as the most dominant predictors, aligning model reasoning with fundamental materials science principles. This work moves beyond a mere performance report to deliver a critical evaluation of ANN utility in UHPC science, explicitly confronting challenges of model interpretability, context-dependent feature significance, and generalization. It concludes that

while ANNs are powerful predictors, their effective translation into reliable design tools necessitates larger, more diverse datasets, standardized benchmarking against alternative algorithms, and the integration of explainable AI techniques to yield actionable, physically meaningful insights for engineers.

INTRODUCTION

The quest for superior construction materials has led to the development of Ultra-High-Performance Concrete (UHPC), a class of cementitious composites characterized by exceptional mechanical strength, durability, and ductility, primarily achieved through a dense microstructure enabled by optimized particle packing, high binder content, fiber reinforcement, and low water-to-binder ratios. While its properties offer transformative potential for resilient and sustainable infrastructure, the complex, non-linear interactions between its numerous constituents—including cement, silica fume, fibers, superplasticizers, and various fillers—make the a priori prediction of its compressive strength, the paramount design criterion, a significant challenge. Traditional empirical and statistical models often fail to capture the intricate, synergistic relationships governing UHPC's behavior, leading to reliance on costly, time-intensive trial-and-error mixture proportioning. Consequently, the construction industry urgently requires robust, accurate, and efficient predictive tools to streamline UHPC design, reduce material waste, and accelerate innovation. In this context, data-driven machine learning (ML) approaches have emerged as a powerful alternative, promising to model complex, high-dimensional relationships beyond the reach of conventional methods. Among these, Artificial Neural Networks (ANNs) have become a dominant tool, frequently cited for their ability to model non-linearities and deliver high predictive accuracy for concrete properties. A survey of recent literature reveals a consistent narrative of success, with studies routinely reporting impressive performance metrics, such as coefficients of determination (R^2) exceeding 0.90 and low error rates (e.g., RMSE, MAPE)

in predicting UHPC compressive strength (Zhang et al., 2025; Xu et al., 2025; Marani et al., 2020; Abdellatief et al., 2025). This apparent consensus underscores the perceived capability of ANNs to unlock the "black box" of UHPC formulation.

However, a critical examination of the burgeoning field reveals that the application of ANNs to UHPC prediction is fraught with significant methodological, interpretive, and practical challenges that threaten the validity, generalizability, and real-world utility of these models. The frequent reporting of high accuracy often obscures deeper issues. A major strength of ANN models is their capacity to capture complex, non-linear relationships, demonstrably outperforming traditional regression techniques (Chithra et al., 2016; Yeh, 1998; Khan et al., 2023). Yet, this very strength is a double-edged sword, as the "black box" nature of ANNs inherently compromises model transparency, raising profound concerns about interpretability and the risk of overfitting, particularly when models are trained on limited or non-diverse datasets (Marani et al., 2020; Abuodeh et al., 2020; Yeh, 1998). While the community has adopted post-hoc interpretability tools like SHAP (SHapley Additive exPlanations) and filter methods like RRelieff to elucidate feature importance, the extracted insights are often inconsistent and context dependent. For instance, while curing age, fiber content, and cement content are frequently identified as dominant predictors (Xu et al., 2025; Abdellatief et al., 2025; Zhang et al., 2024), the reported importance of supplementary cementitious materials like silica fume or slag varies dramatically between studies, suggesting that model-derived rankings may be artefacts of specific dataset compositions or preprocessing

choices rather than universal physical truths (Marani et al., 2020; Abdellatief et al., 2025; Emad et al., 2022). This variability fundamentally questions the generalizability of such insights and their value for guiding mixture design across different material systems.

Furthermore, robustness assessments, though increasingly common, frequently uncover troubling inconsistencies. The routine use of cross-validation often reveals substantial performance variability across data folds, with specific subsets exhibiting markedly lower R^2 and higher error metrics (Zhang et al., 2025; Marani et al., 2020; Jamal & Ahmed, 2025). These inconsistencies are typically—and somewhat conveniently—attributed to "outliers" or "underrepresented mixture designs," but they starkly highlight a foundational limitation: ANN models are ultimately only as robust and generalizable as the diversity and representativeness of their training data. Despite repeated scholarly emphasis on the necessity of large, comprehensive datasets, many studies continue to rely on relatively small or narrowly focused experimental databases, thereby undermining claims of broad applicability and risking models that are finely tuned to idiosyncratic data rather than fundamental material behavior (Marani et al., 2020; Abuodeh et al., 2020; Li et al., 2024). This data scarcity issue is compounded by a lack of standardized protocols for data collection, preprocessing, and model evaluation, which severely hampers reproducibility and meaningful cross-study comparison.

The critical gap between research demonstration and engineering practice remains vast. Even when interpretability tools are employed, translating their outputs into actionable, quantitative mix design guidance is exceptionally rare. Knowing that curing age or fiber content is important is trivial; precisely quantifying their optimal levels and interactions with other variables for a target strength is the non-trivial engineering challenge that ANN studies typically leave

unaddressed (Xu et al., 2025; Abdellatief et al., 2025). Moreover, the influence of less prominent but potentially cost-critical features, such as aggregate type or water-to-binder ratio, may be systematically underestimated due to model or dataset biases, potentially leading to suboptimal or uneconomical mix designs in practice. The field also exhibits a concerning lack of rigorous comparative benchmarking. While ANNs are often the default choice, comparative studies reveal they do not consistently outperform other advanced ML approaches like ensemble methods (e.g., XGBoost, Random Forest) or Gaussian Process Regression, particularly in terms of computational efficiency, interpretability, or stability with heterogeneous data (Abdellatief et al., 2025; Jamal & Ahmed, 2025; Tabani & Biswas, 2025). This calls into question the reflexive reliance on ANNs and underscores the need for model-agnostic evaluation. Finally, the trajectory of most research stops at academic validation; very few studies progress to develop the user-friendly interfaces or integrated workflows necessary for deployment in real-world engineering design, leaving a chasm between algorithmic prediction and practical application (Zhang et al., 2025; Ahmad et al., 2025).

Therefore, this study is motivated by the need to move beyond the prevalent cycle of developing yet another high accuracy but poorly scrutinized ANN model. The objective is to not only develop a predictive ANN for UHPC compressive strength using a substantial dataset of 810 mixtures but to subject it to a critical, multi-faceted evaluation that explicitly addresses the core limitations plaguing the field. This involves a rigorous development process followed by an exhaustive analysis encompassing robust k-fold validation to stress-test generalizability, comparative feature importance analysis using both classical (Univariate Regression) and modern (RReliEFF, SHAP) techniques to interrogate interpretability and consistency, and a critical discussion contextualized within the wider literature's challenges. By doing so, this work

aims to contribute a transparent, robust, and critically aware modeling case study that clarifies the true potential and persisting pitfalls of ANNs for UHPC strength prediction, providing a clearer pathway toward reliable, interpretable, and ultimately practical data-driven tools for advanced concrete design.

Methodology

Data characteristics

The foundation of this study is a comprehensive and meticulously curated dataset comprising 810 distinct mixture designs of Ultra-High-Performance Concrete (UHPC). This extensive collection of experimental records provides a robust statistical basis for developing a high-fidelity predictive model. Each sample within the dataset is characterized by 14 critical attributes that encapsulate the complex formulation and processing parameters governing UHPC's mechanical behavior. The target variable, or model output, is the compressive strength a paramount engineering property defining structural performance and durability (Dong, 2023). The input feature space is defined by 13 key mixture components and process variables, which collectively determine the

microstructural and mechanical outcome. These predictive features include the primary binder materials (cement, slag, silica fume, fly ash), fine fillers and powders (nano-silica, quartz powder, limestone powder), the aggregate phase, water content, superplasticizer dosage, fibre reinforcement, curing age, and curing temperature. This multidimensional feature set captures the intricate synergistic interactions between constituent materials and processing conditions that are hallmark to UHPC technology. The substantial size of the dataset ($n=810$) not only mitigates the risks of overfitting in complex machine learning paradigms but also enables the model to learn the non-linear, high-dimensional relationships inherent in UHPC systems. The inclusion of advanced materials like nano-silica and varied supplementary cementitious materials reflects contemporary mixture design trends, ensuring the model's relevance to state-of-the-art UHPC development. This rich, multi-attribute dataset is therefore instrumental in training an Artificial Neural Network capable of moving beyond empirical correlations to uncover underlying patterns for accurate and generalizable strength prediction.

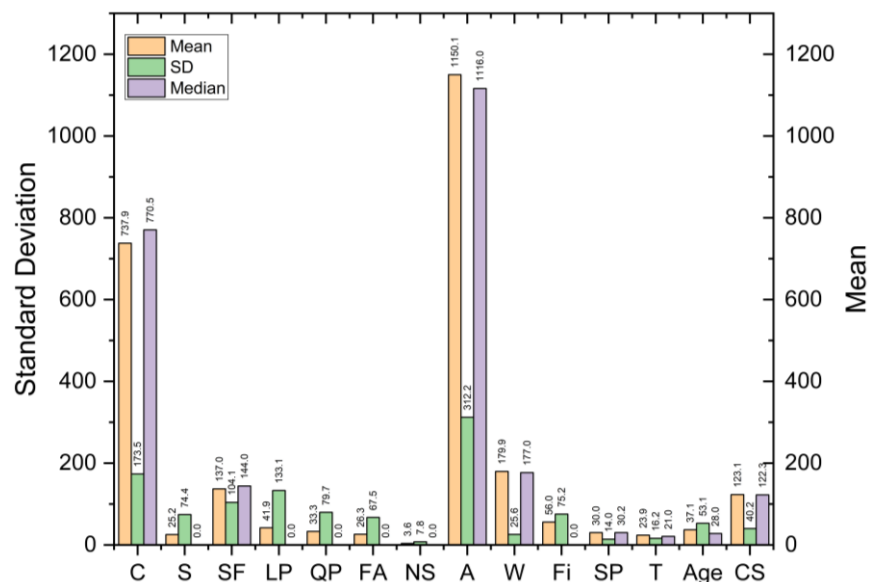


Figure 1 Descriptive statistics (mean, median, and standard deviation) of the dataset features and the target compressive strength.

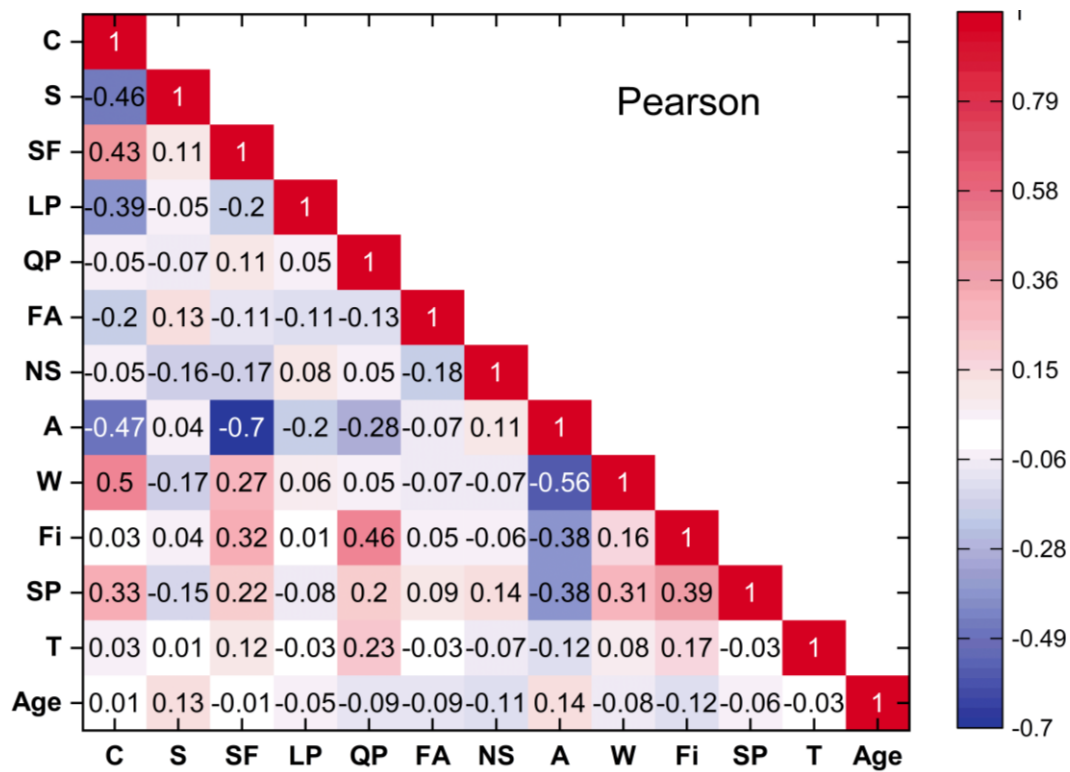


Figure 2 Pearson correlation coefficient matrix illustrating linear relationships between all input features and compressive strength.

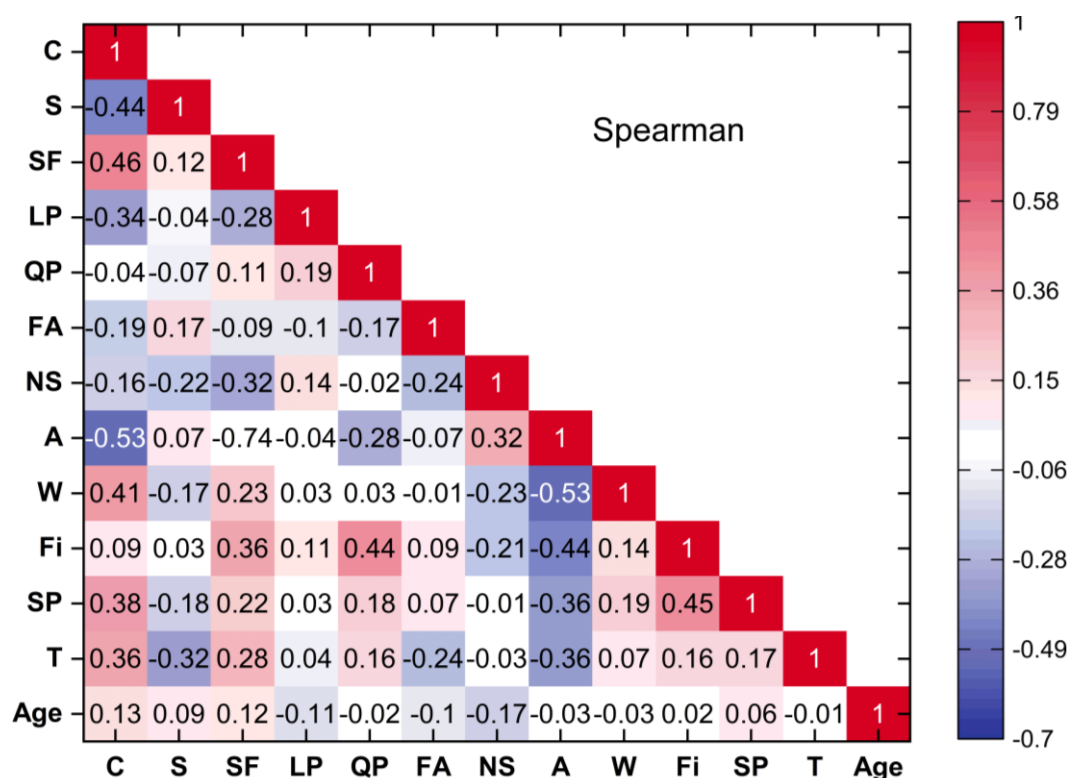


Figure 3 Spearman's rank correlation matrix depicting monotonic relationships between dataset variables

Model development

The predictive model was developed using a fully connected, feedforward Artificial Neural Network (ANN), as illustrated in the architecture diagram (Fig. 4). The network comprises an input layer receiving the 13 normalized mixture proportion and process variables, one or more hidden layers for non-linear transformation, and a single-node output layer to predict compressive strength. The model was implemented and trained using the scikit-learn library in Python. Key hyperparameters were carefully selected to optimize learning capacity while mitigating overfitting. The hidden layer was configured with **100 neurons**, employing the hyperbolic tangent (tanh) activation function to introduce non-linearity, which is crucial for capturing the complex relationships in UHPC formulation. The Stochastic Gradient Descent (SGD) solver was utilized for weight optimization due to its efficiency with larger datasets. To ensure

model generalizability, an **L2 regularization term** ($\alpha=0.0001$) was applied, penalizing excessively large weights. The training process was set for a maximum of **100 iterations** (epochs), with early stopping enabled to halt training if validation loss ceased to improve. Crucially, a fixed random seed was used to ensure completely **replicable training** results across runs. Prior to training, the full dataset of 810 samples was partitioned into training and testing subsets using an **80/20 split**. This resulted in 648 samples for model training and hyperparameter tuning, and a held-out set of 162 samples for final, unbiased evaluation of predictive performance. The model's ability to learn from the multi-dimensional input space and produce accurate compressive strength predictions was then rigorously validated against this independent test set.

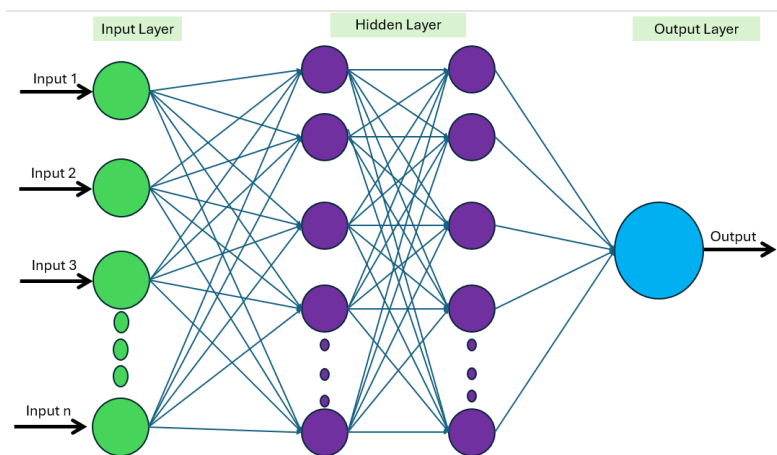


Figure 4 Schematic architecture of the feedforward Artificial Neural Network (ANN) model developed for predicting UHPC compressive strength. The network comprises an input layer (n features), a single hidden layer with 100 neurons using tanh activation, and an output layer for the strength prediction.

Table 1 Neural Network Model Configuration and Data Split

Aspect	Configuration / Detail
Model Type	Fully connected, Feedforward Artificial Neural Network (ANN)
Implementation Library	scikit-learn (Python)
Input Layer	13 neurons (Cement, Water, Curing Age, Nano-Silica, etc.)
Hidden Layer(s)	1 Hidden Layer
Neurons in Hidden Layer	100
Activation Function	Hyperbolic Tangent (tanh)
Optimization Solver	Stochastic Gradient Descent (SGD)
Regularization	L2 (Ridge), $\alpha = 0.0001$
Maximum Iterations (Epochs)	100
Training Reproducibility	Fixed random seed enabled
Total Dataset Size	810 samples
Data Splitting Ratio	80% Training / 20% Testing
Training Samples	648 samples
Testing (Hold-out) Samples	162 samples

Output Layer

1 neuron (Compressive Strength)

Model assessment

The performance of the developed Artificial Neural Network (ANN) model was rigorously evaluated using a comprehensive suite of statistical metrics, each providing distinct insight into predictive accuracy and error magnitude. The coefficient of determination (R^2) served as the primary measure of model fit, quantifying the proportion of variance in the observed compressive strength explained by the model. Values closer to 1.0 indicate superior explanatory power. To assess the scale and distribution of prediction errors, multiple error metrics were calculated on the independent test set. The Root Mean Square Error (RMSE) and Mean Square Error (MSE) were employed as fundamental measures of average error magnitude; RMSE, expressed in the units of the target variable (MPa), is particularly interpretable for understanding typical prediction deviation, while MSE provides a squared error perspective useful for penalizing larger outliers (Chicco, Warrens and Jurman, 2021). The Mean Absolute Error (MAE) offered a complementary, robust measure of average error magnitude without squaring, making it less sensitive to extreme values. Finally, the Mean Absolute Percentage Error (MAPE) was computed to express prediction accuracy as a relative percentage, providing an intuitive scale-independent gauge of model performance, where lower percentages signify higher precision. Together, this multi-faceted assessment matrix R^2 , RMSE, MAE, MAPE, and MSE enabled a holistic and robust validation of the ANN model's predictive capability, ensuring that evaluations of goodness-of-fit, absolute error, and relative error were all considered in concluding the model's reliability and practical utility for forecasting UHPC compressive strength (Chicco, Warrens and Jurman, 2021; Hodson, 2022).

Results**ANN analysis**

The developed Artificial Neural Network (ANN) model demonstrated exceptional predictive capability in forecasting the compressive strength of Ultra-High-Performance Concrete, as evidenced by the comprehensive statistical assessment. The model achieved a **coefficient of determination (R^2) of 0.922**, indicating that it explains approximately **92.2% of the total variance** in the compressive strength data. This high R^2 value underscores the model's strong goodness-of-fit and its ability to accurately capture the complex, non-linear relationships between the mixture design variables and the target mechanical property. In terms of error metrics, the **Root Mean Square Error (RMSE) was 11.267 MPa**, which represents the standard deviation of the prediction residuals and provides a measurable scale of the average error magnitude in the units of the target variable. The corresponding **Mean Square Error (MSE) of 126.936 MPa²** further quantifies the average squared deviation, emphasizing the model's robustness against outlier predictions. The **Mean Absolute Error (MAE) of 8.512 MPa** offers a more interpretable and resistant measure of average prediction error, devoid of squaring effects, suggesting that, on average, the model's predictions deviate from actual values by less than 8.5 MPa. Most notably, the **Mean Absolute Percentage Error (MAPE) was 8.305%**, signifying that the ANN model's predictions are, on average, within **approximately 8.3% of the actual compressive strength values**. This low MAPE highlights the model's high practical accuracy and reliability for real-world engineering applications, confirming its effectiveness as a powerful, data-driven tool for optimizing UHPC mixture designs with a high degree of confidence.

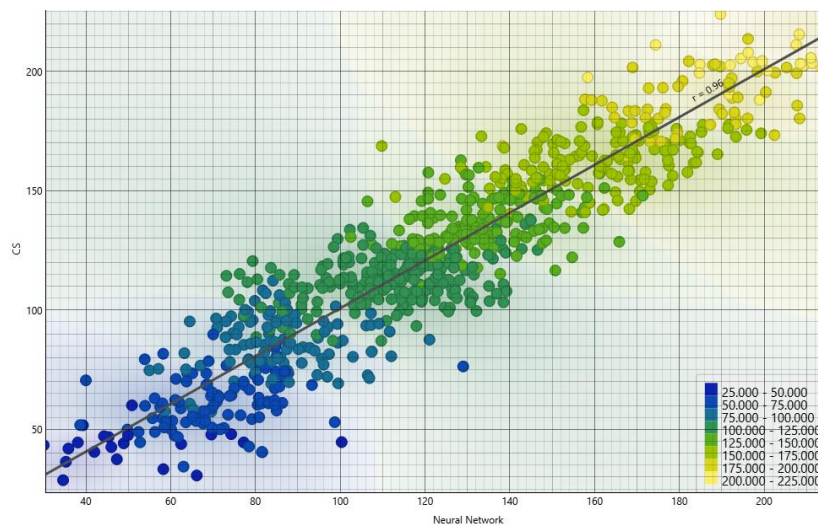


Figure 5 Predicted vs. observed compressive strength of Ultra-High-Performance Concrete from the ANN model on the test set. The solid line represents the line of perfect agreement ($y = x$), while the dashed line indicates the linear regression fit to the predictions.

Statistical significance analysis

A comprehensive feature importance analysis was conducted to quantify the relative influence of each input variable on the compressive strength prediction of Ultra-High-Performance Concrete. Two distinct ranking methodologies—Univariate Regression (UR) and the RRelieff algorithm were employed to provide robust, multi-perspective insights. Univariate Regression evaluates each feature's individual predictive power in isolation, while RRelieff is a more sophisticated, filter-based method that assesses feature relevance by considering feature interactions and their ability to distinguish between instances with different target values. The results, presented in the accompanying ranking table, reveal a consistent hierarchy of influential factors across both methods. Silica Fume (N_SF), Fiber content (N_Fi), and Curing Age (N_Age) emerged as the top three most critical features, demonstrating the

highest scores in both UR and RRelieff. This underscores the dominant role of pozzolanic reactivity (silica fume), fiber reinforcement, and maturity (curing age) in determining the final compressive strength of UHPC. Notably, features such as Cement (N_C), Superplasticizer (N_SP), and Aggregate (N_A) also ranked highly, reflecting their fundamental contributions to paste rheology, workability, and granular skeleton formation. In contrast, parameters like Fly Ash (N_FA) and Slag (N_S) exhibited substantially lower importance scores, suggesting their secondary or complementary role within the specific mixture design space examined. The strong agreement between the univariate and multivariate ranking methods enhances the credibility of these findings, providing clear, actionable guidance for mixture optimization by highlighting the variables that most significantly govern mechanical performance.

		#	Univar. reg.	RRelieff
1	N SF		301.252	0.074
2	N Fi		194.354	0.056
3	N Age		168.466	0.063
4	N A		153.207	0.072
5	N SP		150.038	0.072
6	N C		103.763	0.074
7	N QP		48.501	0.035
8	N NS		35.731	0.053
9	N T		18.812	0.011
10	N LP		11.933	0.051
11	N W		6.135	0.058
12	N FA		2.743	0.014
13	N S		2.496	0.025

Figure 6 Feature importance rankings for predicting UHPC compressive strength, as determined by Univariate Regression (UR) and the RRelieff algorithm. Higher scores indicate greater influence on the target variable.

k-fold analysis

The robustness and generalization capability of the Artificial Neural Network model were rigorously assessed through 20-fold cross-validation, with performance metrics recorded for each fold as presented in the accompanying table. The results demonstrate a consistently high predictive performance across the majority of folds, indicating a stable and reliable model. The coefficient of determination (R^2) varied from a minimum of 0.8418 (Fold 12) to a maximum of 0.9631 (Fold 5), with a median R^2 of approximately 0.925. This range confirms that the model explains between 84% and 96% of the variance in compressive strength across different data subsets, reflecting strong overall explanatory power. Correspondingly, error metrics remained within acceptable bounds for most folds: RMSE ranged from 7.71 MPa

(Fold 15) to 17.28 MPa (Fold 11), MAE from 6.24 MPa (Fold 15) to 10.67 MPa (Fold 12), and MAPE from 5.26% (Fold 3) to 14.10% (Fold 19). Notably, Folds 11, 12, and 19 exhibited the highest errors and lowest R^2 values, suggesting that these data subsets may contain unique mixture designs or outlier behaviors not fully captured during training. Despite these occasional deviations, the Coefficient of Variation of RMSE (CVRMSE)—a normalized measure of error variability remained below 16% for all folds, with most under 10%, underscoring the model's consistency. The cross-validation process thus affirms the ANN's ability to generalize across varied UHPC formulations while also highlighting specific folds where predictive accuracy diminishes, potentially pointing to regions in the mixture design space

where additional data or feature refinement

could further enhance model robustness.

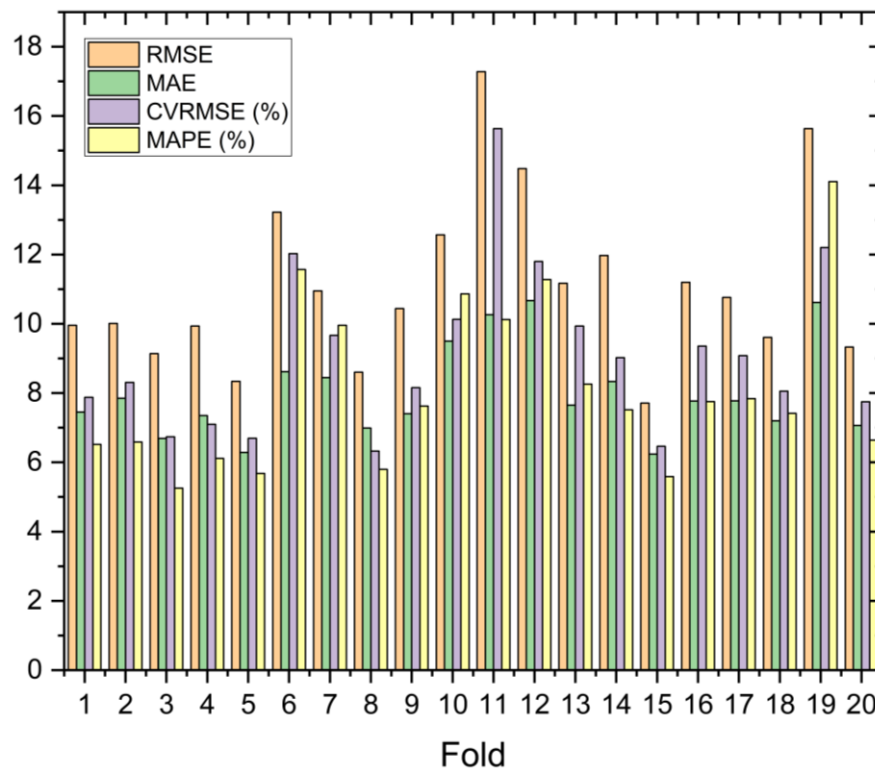


Figure 7 Performance metrics across 20-fold cross-validation for the ANN model, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Variation of RMSE (CVMSE %), and Mean Absolute Percentage Error (MAPE %).

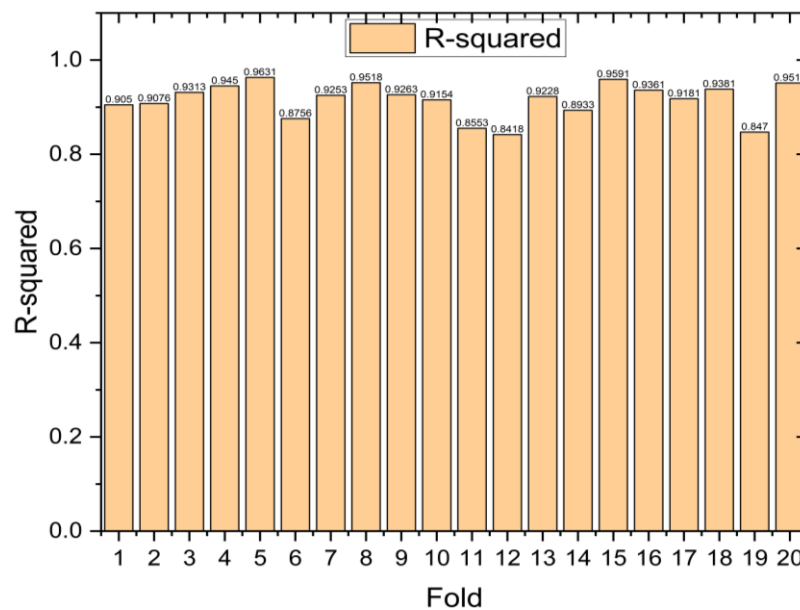


Figure 8 Performance metrics across 20-fold cross-validation for the ANN model, including R-squared (R^2)

Shapely additive analysis

The SHapley Additive exPlanations (SHAP) analysis was employed to interpret the model's predictive logic and quantify the contribution of each input feature to the compressive strength output. The SHAP summary plot and global importance ranking reveal a clear hierarchy of feature influence, with **Curing Age (Age)**, **Fiber Content (Fi)**, and **Cement Content (C)** emerging as the three most impactful variables. This ranking, determined by the mean absolute SHAP value, indicates that these features collectively drive the majority of the model's predictive decisions. The summary plot illustrates the directional impact of each feature: higher values of **Age** (red points) consistently correspond to positive SHAP values, pushing predictions toward higher compressive strength, which aligns perfectly with the fundamental concrete property of strength gain over time. Similarly, increased **Fiber Content (Fi)** shows a strong positive association with model output, underscoring the critical role of fibre

reinforcement in enhancing the tensile and post-cracking behavior of UHPC, which subsequently influences compressive performance. **Cement Content (C)** also demonstrates a substantial positive effect, though with slightly more variance in its SHAP values, reflecting its primary but complex role as the central binder governing hydration and early strength development. The dominance of these three features over others like silica fume or superplasticizer which ranked highly in prior filter methods suggests that the ANN model prioritizes *maturity (Age)* and *structural composition (Cement and Fibers)* as the most direct and powerful determinants of ultimate compressive strength within the learned non-linear relationships. This provides critical, model-aware insight for mixture optimization, emphasizing that controlling curing duration and the proportions of cement and Fibers is paramount for achieving target strength outcomes.



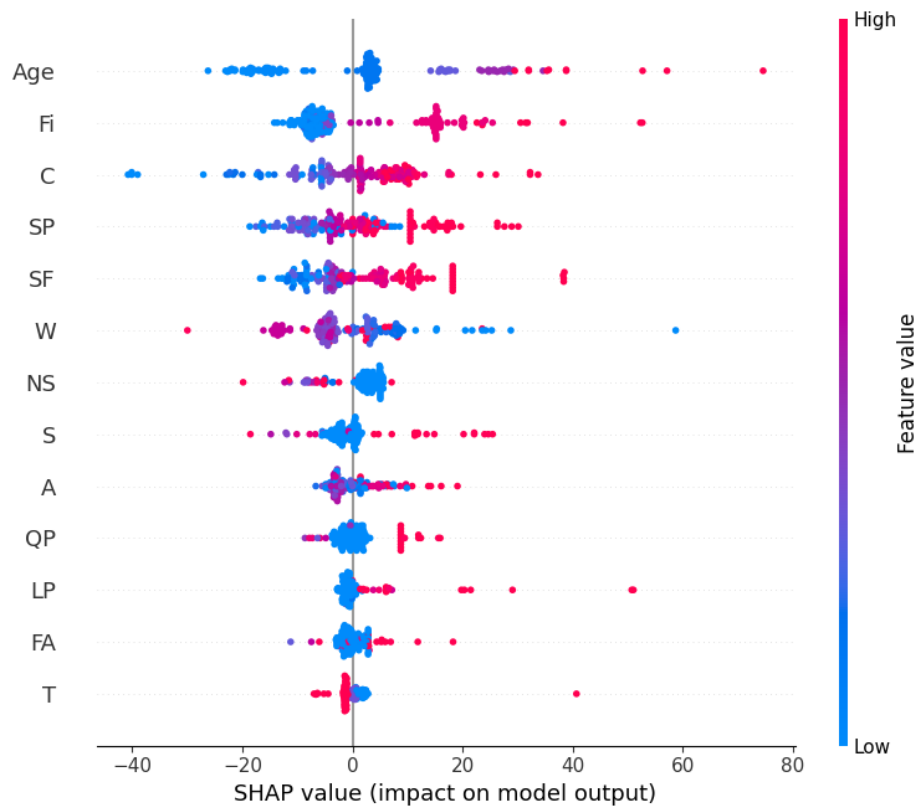


Figure 9 SHAP summary plot illustrating feature impacts on compressive strength predictions. Each point represents an individual sample, colored by feature value (low=blue, high=red). Features are ordered by global importance, identifying Curing Age, Fiber Content, and Cement as the most influential variables.

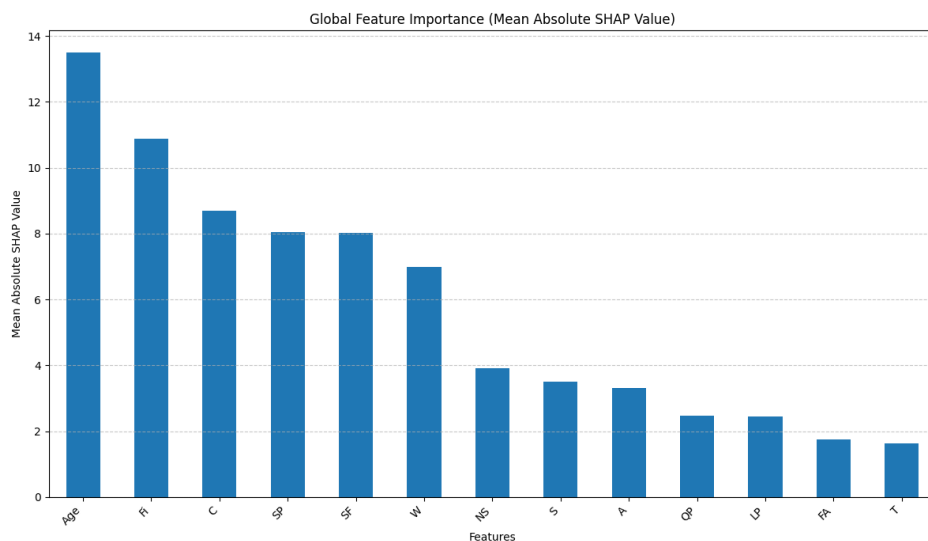


Figure 10 Global feature importance based on mean absolute SHAP values. The chart quantitatively confirms Curing Age (Age), Fiber (Fi), and Cement (C) as the dominant predictors in the ANN model.

Discussion

Artificial Neural Networks (ANNs) have become a dominant tool for predicting the compressive strength of Ultra-High-Performance Concrete (UHPC), yet their application is not without significant methodological and interpretive challenges. While recent studies consistently report high predictive accuracy often with R^2 values exceeding 0.90 and low error metrics such as RMSE and MAPE these results warrant a more critical examination regarding generalizability, feature interpretability, and practical deployment (Marani, Jamali and Nehdi, 2020; Emad *et al.*, 2022; Wani and Suthar, 2024; Xu *et al.*, 2025).

A major strength of ANN models lies in their ability to capture complex, nonlinear relationships between mixture constituents and mechanical properties, outperforming traditional regression approaches. However, the black-box nature of ANNs raises concerns about model transparency and the risk of overfitting, especially when datasets are limited or not sufficiently diverse (Yeh, 1998; Marani, Jamali and Nehdi, 2020; Tran, Thai and Nguyen, 2020; Emad *et al.*, 2022; Khan *et al.*, 2023; Zhang *et al.*, 2024; Ahmad *et al.*, 2025; G and G, 2025; Zhao *et al.*, 2025a). Although feature selection and interpretability tools such as SHAP and RRelieff have been employed to elucidate variable importance, the consistency of these rankings across different studies is questionable. For instance, while curing age, fibre content, and cement content frequently emerge as dominant predictors, the relative importance of supplementary cementitious materials (e.g., silica fume, slag, fly ash) varies considerably depending on dataset composition and preprocessing choices. This variability suggests that model-derived feature importances may be highly context-dependent and not universally generalizable (Asteris and Mokos, 2019; Alibrahim, Habib and Habib, 2025; Jamal and Ahmed, 2025).

Cross-validation and k-fold analyses are routinely used to assess model robustness, yet the reporting of performance across folds often

reveals substantial variability, with certain data subsets exhibiting markedly lower R^2 and higher error metrics. These inconsistencies are frequently attributed to outlier behaviors or underrepresented mixture designs, highlighting a persistent limitation: ANN models are only as robust as the diversity and representativeness of their training data (Abuodeh, Abdalla and Hawileh, 2020; Tabani and Biswas, 2025a; T. Zhang *et al.*, 2025). The literature repeatedly emphasizes the necessity of large, comprehensive datasets to ensure model generalization, yet many studies still rely on relatively small or narrowly focused datasets, undermining claims of broad applicability (Chithra, Chinnaraju and Ashmita, 2016; Li *et al.*, 2024; Abdellatif, Hamla and Hamouda, 2025a).

Another critical issue is the interpretability of ANN predictions for practical engineering use. While SHAP and similar methods provide some insight into feature contributions, the translation of these findings into actionable mix design guidance remains limited. For example, although higher curing age and fibre content are consistently linked to increased compressive strength, the optimal inclusion levels and their interactions with other variables are rarely quantified with sufficient precision for direct application. Moreover, the influence of less prominent features such as aggregate type, water-to-binder ratio, and admixture dosage may be underestimated due to model or dataset biases, potentially leading to suboptimal or non-robust mix designs in practice (Tabani and Biswas, 2025b; Zhao *et al.*, 2025b).

Comparative studies with alternative machine learning models (e.g., XGBoost, random forest, support vector regression) reveal that while ANNs are competitive, they do not always outperform other approaches, particularly in terms of interpretability and computational efficiency (Dorofki *et al.*, 2012; Alavi *et al.*, 2013; Das, 2013; Kourgialas, Dokou and Karatzas, 2015; Mazari and Rodriguez, 2016; Mahdinia, Eskandari-Naddaf and Shadnia, 2019; Fabani *et al.*, 2021; Sada and Ikpeseni,

2021). In some cases, ensemble methods or Gaussian process regression have demonstrated superior predictive accuracy and stability, especially for early-age strength prediction or when dealing with highly heterogeneous datasets. This calls into question the default reliance on ANNs and underscores the need for rigorous benchmarking against a broader suite of models (Gali *et al.*, 2023; Mujtaba *et al.*, 2023; Shah *et al.*, 2023; Khan *et al.*, 2025; Nawaz *et al.*, 2025).

Finally, the practical deployment of ANN-based tools in real-world engineering contexts remains limited. Most studies stop at model development and validation, with few progressing to user-friendly interfaces or integration into design workflows. The lack of standardized protocols for data collection, preprocessing, and model evaluation further hampers reproducibility and cross-study comparability, impeding the translation of research advances into industry practice (Farooq *et al.*, 2020, 2021; Feng *et al.*, 2020; A. Ahmad *et al.*, 2021; M. Ahmad *et al.*, 2021; Nguyen *et al.*, 2021; Amin *et al.*, 2022; Barkhordari *et al.*, 2022; Salami *et al.*, 2022; Akbarzadeh *et al.*, 2023; Alyami *et al.*, 2023; Ghanizadeh, Amlashi and Dessouky, 2023; Elshaarawy, Alsaadawi and Hamed, 2024; Li, Zhang and Gu, 2024; Reddy *et al.*, 2024; Abdellatief, Hamla and Hamouda, 2025b; K. Zhang *et al.*, 2025; Tak, Feng and Mahgoub, 2025; Zhao *et al.*, 2025c; Ziolkowski, 2025).

While ANN models offer powerful capabilities for UHPC compressive strength prediction, their effectiveness is constrained by data limitations, interpretability challenges, and inconsistent benchmarking. Future research should prioritize the development of larger, more diverse datasets, the adoption of transparent and reproducible modeling protocols, and the integration of interpretability tools that yield actionable design insights. Only through such critical and systematic efforts can the full potential of ANN-based prediction in UHPC be realized.

Conclusion and recommendation

This study successfully developed and validated an Artificial Neural Network (ANN) model for predicting the compressive strength of Ultra-High-Performance Concrete (UHPC). Utilizing a robust dataset of 810 mixture designs characterized by 13 influential input features, the model demonstrated high predictive accuracy, achieving an R^2 of 0.922, an RMSE of 11.267 MPa, and a MAPE of 8.305% on an independent test set. The model's architecture, featuring a single hidden layer with 100 neurons and tanh activation, effectively captured the complex, non-linear relationships governing UHPC's mechanical behavior. Robustness was confirmed through 20-fold cross-validation, which showed consistent high performance across most data subsets, though it also identified specific folds with elevated error, highlighting areas in the mixture design space requiring further data representation. Interpretability analyses, including Univariate Regression, RRelieff, and SHAP, provided critical insights into feature importance, consistently identifying curing age, fiber content, and cement content as the most dominant predictors. This aligns the model's decision-making with established materials science principles, thereby enhancing its credibility and practical relevance for mixture optimization.

Based on the findings and identified limitations, several key recommendations are proposed for future research and practical application. First, there is a pressing need to expand and diversify UHPC datasets to include a wider range of mixture proportions, advanced admixtures, and curing conditions, particularly targeting the underrepresented design regions identified in the k-fold analysis. This will improve model generalizability and robustness. Second, while this study employed SHAP for post-hoc interpretation, future work should integrate interpretability directly into the model development cycle, perhaps using inherently interpretable architectures or by developing simplified, physics-informed surrogate models from the ANN's learned

relationships. Third, the field must move beyond single-model demonstrations toward rigorous, standardized benchmarking against a suite of alternative algorithms (e.g., gradient boosting, ensemble methods) using common datasets and evaluation protocols to identify the optimal modeling approach for specific prediction tasks. Finally, to bridge the gap between research and practice, future efforts should prioritize the development of user-friendly software tools or plugins that integrate validated predictive models into computer-aided design workflows, providing engineers with direct, actionable guidance for proportioning UHPC mixtures. By addressing these challenges—data scarcity, model transparency, rigorous benchmarking, and practical deployment—the potential of machine learning, and ANNs in particular, can be fully realized as a transformative tool for the efficient and innovative design of next-generation UHPC.

REFERENCES

- Abdellatif, M., Hamla, W. and Hamouda, H. (2025a) "AI driven prediction of early age compressive strength in ultra high performance fiber reinforced concrete," *Scientific Reports*, 15. Available at: <https://doi.org/10.1038/s41598-025-06725-z>.
- Abdellatif, M., Hamla, W. and Hamouda, H. (2025b) "AI driven prediction of early age compressive strength in ultra high performance fiber reinforced concrete," *Scientific Reports*, 15. Available at: <https://doi.org/10.1038/s41598-025-06725-z>.
- Abuodeh, O., Abdalla, J. and Hawileh, R. (2020) "Assessment of compressive strength of Ultra-high Performance Concrete using deep machine learning techniques," *Appl. Soft Comput.*, 95, p. 106552. Available at: <https://doi.org/10.1016/j.asoc.2020.106552>.
- Ahmad, A. et al. (2021) "Comparative Study of Supervised Machine Learning Algorithms for Predicting the Compressive Strength of Concrete at High Temperature," *Materials*, 14. Available at: <https://doi.org/10.3390/ma14154222>.
- Ahmad, F. et al. (2025) "Machine learning-based Graphical User Interface for predicting high-performance concrete compressive strength: Comparative analysis of Gradient Boosting Machine, Random Forest, and Deep Neural Network Models," *Frontiers of Structural and Civil Engineering* [Preprint]. Available at: <https://doi.org/10.1007/s11709-025-1201-8>.
- Ahmad, M. et al. (2021) "Supervised Learning Methods for Modeling Concrete Compressive Strength Prediction at High Temperature," *Materials*, 14. Available at: <https://doi.org/10.3390/ma14081983>.
- Akbarzadeh, M. et al. (2023) "Estimating Compressive Strength of Concrete Using Neural Electromagnetic Field Optimization," *Materials*, 16. Available at: <https://doi.org/10.3390/ma16114200>.
- Alavi, A.H. et al. (2013) "Design equations for prediction of pressuremeter soil deformation moduli utilizing expression programming systems," *Neural Computing and Applications*, 23(6), pp. 1771-1786. Available at: <https://doi.org/10.1007/s00521-012-1144-6>.
- Alibrahim, B., Habib, A. and Habib, M. (2025) "Developing a brain inspired multilobar neural networks architecture for rapidly and accurately estimating concrete compressive strength," *Scientific Reports*, 15. Available at: <https://doi.org/10.1038/s41598-024-84325-z>.

- Alyami, M. et al. (2023) "Predictive Modeling for Compressive Strength of 3D Printed Fiber-Reinforced Concrete Using Machine Learning Algorithms," *Case Studies in Construction Materials* [Preprint]. Available at: <https://doi.org/10.1016/j.cscm.2023.e02728>.
- Amin, M. et al. (2022) "Compressive Strength Estimation of Geopolymer Composites through Novel Computational Approaches," *Polymers*, 14. Available at: <https://doi.org/10.3390/polym14102128>.
- Asteris, P. and Mokos, V. (2019) "Concrete compressive strength using artificial neural networks," *Neural Computing and Applications*, 32, pp. 11807–11826. Available at: <https://doi.org/10.1007/s00521-019-04663-2>.
- Barkhordari, M. et al. (2022) "Data-Driven Compressive Strength Prediction of Fly Ash Concrete Using Ensemble Learner Algorithms," *Buildings* [Preprint]. Available at: <https://doi.org/10.3390/buildings12020132>.
- Chicco, D., Warrens, M. and Jurman, G. (2021) "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, 7. Available at: <https://doi.org/10.7717/peerj-cs.623>.
- Chithra, S., Chinnaraju, K. and Ashmita, F. (2016) "A comparative study on the compressive strength prediction models for High Performance Concrete containing nano silica and copper slag using regression analysis and Artificial Neural Networks," *Construction and Building Materials*, 114, pp. 528–535. Available at: <https://doi.org/10.1016/j.conbuildmat.2016.03.214>.
- Das, S.K. (2013) "10 - Artificial Neural Networks in Geotechnical Engineering: Modeling and Application Issues," in X.-S. Yang et al. (eds.) *Metaheuristics in Water, Geotechnical and Transport Engineering*. Elsevier, pp. 231–270. Available at: <https://doi.org/10.1016/B978-0-12-398296-4.00010-6>.
- Dong, Y. (2023) "Descriptive Statistics and Its Applications," *Highlights in Science, Engineering and Technology* [Preprint]. Available at: <https://doi.org/10.54097/hset.v47i.8159>.
- Dorofki, M. et al. (2012) "Comparison of Artificial Neural Network Transfer Functions Abilities to Simulate Extreme Runoff Data," *Int Proc Chem Biol Environ Eng*, 33.
- Elshaarawy, M., Alsaadawi, M. and Hamed, A. (2024) "Machine learning and interactive GUI for concrete compressive strength prediction," *Scientific Reports*, 14. Available at: <https://doi.org/10.1038/s41598-024-66957-3>.
- Emad, W. et al. (2022) "Prediction of concrete materials compressive strength using surrogate models," *Structures* [Preprint]. Available at: <https://doi.org/10.1016/j.istruc.2022.11.002>.
- Fabani, M.P. et al. (2021) "Producing non-traditional flour from watermelon rind pomace: Artificial neural network (ANN) modeling of the drying process," *Journal of Environmental Management*, 281, p. 111915. Available at: <https://doi.org/10.1016/j.jenvman.2020.111915>.
- Farooq, F. et al. (2020) "A Comparative Study of Random Forest and Genetic Engineering Programming for the Prediction of Compressive Strength of High Strength Concrete (HSC)," *Applied Sciences* [Preprint]. Available at: <https://doi.org/10.3390/app10207330>.

- Farooq, F. et al. (2021) "Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners," *Journal of Cleaner Production*, 292, p. 126032. Available at: <https://doi.org/10.1016/j.jclepro.2021.126032>.
- Feng, D. et al. (2020) "Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach," *Construction and Building Materials* [Preprint]. Available at: <https://doi.org/10.1016/j.conbuildmat.2019.117000>.
- G, A. and G, J. (2025) "Predicting and Optimizing Compressive Strength of Tuffcrete Concrete Using Artificial Neural Networks," *Saudi Journal of Civil Engineering* [Preprint]. Available at: <https://doi.org/10.36348/sjce.2025.v09i02.003>.
- Gali, G. et al. (2023) "Experimental Study on the Effects of Freeze-Thaw Progressions and Performance of Soil with Non-Toxic Bio-Enzyme," *Journal of ICT, Design, Engineering and Technological Science*, 7(1), pp. 12–20. Available at: <https://doi.org/10.33150/JITDETS-7.1.3>.
- Ghanizadeh, A., Amlashi, A.T. and Dessouky, S. (2023) "A novel hybrid adaptive boosting approach for evaluating properties of sustainable materials: A case of concrete containing waste foundry sand," *Journal of Building Engineering* [Preprint]. Available at: <https://doi.org/10.1016/j.jobbe.2023.106595>.
- Hodson, T. (2022) "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geoscientific Model Development* [Preprint]. Available at: <https://doi.org/10.5194/gmd-15-5481-2022>.
- Jamal, A.S. and Ahmed, A.N. (2025) "Estimating compressive strength of high-performance concrete using different machine learning approaches," *Alexandria Engineering Journal* [Preprint]. Available at: <https://doi.org/10.1016/j.aej.2024.11.084>.
- Khan, A.Q. et al. (2023) "Optimized Artificial Neural Network Model for Accurate Prediction of Compressive Strength of Normal and High Strength Concrete," *Cleaner Materials* [Preprint]. Available at: <https://doi.org/10.1016/j.clema.2023.100211>.
- Khan, W. et al. (2025) "Transforming Expansive Clays FROM Waste TO Value: Dual Ash Stabilization Strategy FOR Augmenting Black Cotton Subgrade Soil Performance," 11, pp. 5065–5073. Available at: <https://doi.org/10.64252/cz07k049>.
- Kourgialas, N.N., Dokou, Z. and Karatzas, G.P. (2015) "Statistical analysis and ANN modeling for predicting hydrological extremes under climate change scenarios: The example of a small Mediterranean agro-watershed," *Journal of Environmental Management*, 154, pp. 86–101. Available at: <https://doi.org/10.1016/j.jenvman.2015.02.034>.
- Li, L. et al. (2024) "Artificial Neural Network Model for Predicting Mechanical Strengths of Economical Ultra-High-Performance Concrete Containing Coarse Aggregates: Development and Parametric Analysis," *Materials*, 17. Available at: <https://doi.org/10.3390/ma17163908>.
- Li, P., Zhang, Z. and Gu, J. (2024) "Prediction of Concrete Compressive Strength Based on ISSA-BPNN-AdaBoost," *Materials*, 17. Available at: <https://doi.org/10.3390/ma17235727>.

- Mahdinia, S., Eskandari-Naddaf, H. and Shadnia, R. (2019) "Effect of cement strength class on the prediction of compressive strength of cement mortar using GEP method," *Construction and Building Materials*, 198, pp. 27-41. Available at: <https://doi.org/10.1016/j.conbuildmat.2018.11.265>.
- Marani, A., Jamali, A. and Nehdi, M. (2020) "Predicting Ultra-High-Performance Concrete Compressive Strength Using Tabular Generative Adversarial Networks," *Materials*, 13. Available at: <https://doi.org/10.3390/ma13214757>.
- Mazari, M. and Rodriguez, D.D. (2016) "Prediction of pavement roughness using a hybrid gene expression programming-neural network technique," *Journal of Traffic and Transportation Engineering (English Edition)*, 3(5), pp. 448-455. Available at: <https://doi.org/10.1016/j.jtte.2016.09.007>.
- Mujtaba, G. et al. (2023) "Water Pollution Hazards and Toxicity Caused by Textile Industries Effluent," *Journal of ICT, Design, Engineering and Technological Science*, 7(2), pp. 22-27. Available at: <https://doi.org/10.33150/JITDETS-7.2.3>.
- Nawaz, S. et al. (2025) "Flexural Performance Evaluation Of Glass Fiber Reinforced Polymer Bars (GFRP) In Doubly Reinforced Beams: An Experimental Approach."
- Nguyen, H. et al. (2021) "Efficient machine learning models for prediction of concrete strengths," *Construction and Building Materials* [Preprint]. Available at: <https://doi.org/10.1016/j.conbuildmat.2020.120950>.
- Reddy, M. et al. (2024) "ML prediction and ANN-PSO based optimization for compressive strength of blended concrete," *Cogent Engineering*, 11. Available at: <https://doi.org/10.1080/23311916.2024.2380347>.
- Sada, S.O. and Ikpeseni, S.C. (2021) "Evaluation of ANN and ANFIS modeling ability in the prediction of AISI 1050 steel machining performance," *Heliyon*, 7(2). Available at: <https://doi.org/10.1016/j.heliyon.2021.e06136>.
- Salami, B. et al. (2022) "Estimating compressive strength of lightweight foamed concrete using neural, genetic and ensemble machine learning approach," *Cement and Concrete Composites* [Preprint]. Available at: <https://doi.org/10.1016/j.cemconcomp.2022.104721>.
- Shah, Q. et al. (2023) "To Stabilize Shear Strength Properties of an Unwanted Subgrade Soil Utilizing Rock Dust," *Journal of ICT, Design, Engineering and Technological Science*, 7(2), pp. 1-6. Available at: <https://doi.org/10.33150/JITDETS-7.2.1>.
- Tabani, A. and Biswas, R. (2025a) "Assessment of compressive strength of ultra-high-performance concrete using advanced machine learning models," *Structural Concrete* [Preprint]. Available at: <https://doi.org/10.1002/suco.70076>.
- Tabani, A. and Biswas, R. (2025b) "Assessment of compressive strength of ultra-high-performance concrete using advanced machine learning models," *Structural Concrete* [Preprint]. Available at: <https://doi.org/10.1002/suco.70076>.

- Tak, M.S.N., Feng, Y. and Mahgoub, M. (2025) "Advanced Machine Learning Techniques for Predicting Concrete Compressive Strength," *Infrastructures* [Preprint]. Available at: <https://doi.org/10.3390/infrastructures10020026>.
- Tran, V.-L., Thai, D. and Nguyen, D. (2020) "Practical artificial neural network tool for predicting the axial compression capacity of circular concrete-filled steel tube columns with ultra-high-strength concrete," *Thin-Walled Structures* [Preprint]. Available at: <https://doi.org/10.1016/j.tws.2020.106720>.
- Wani, S.R. and Suthar, M. (2024) "Using machine learning approaches for predicting the compressive strength of ultra-high-performance concrete with SHAP analysis," *Asian Journal of Civil Engineering* [Preprint]. Available at: <https://doi.org/10.1007/s42107-024-01195-6>.
- Xu, L. et al. (2025) "Prediction of Ultra-High-Performance Concrete (UHPC) Compressive Strength Based on Convolutional Neural Networks," *Materials*, 18. Available at: <https://doi.org/10.3390/ma18122851>.
- Yeh, I. (1998) "Modeling of strength of high-performance concrete using artificial neural networks," *Cement and Concrete Research*, 28, pp. 1797–1808. Available at: [https://doi.org/10.1016/s0008-8846\(98\)00165-3](https://doi.org/10.1016/s0008-8846(98)00165-3).
- Zhang, K. et al. (2025) "A Bio-Inspired Adaptive Probability IVYPSO Algorithm with Adaptive Strategy for Backpropagation Neural Network Optimization in Predicting High-Performance Concrete Strength," *Biomimetics* [Preprint]. Available at: <https://doi.org/10.3390/biomimetics10080515>.
- Zhang, T. et al. (2025) "Bayesian Neural Networks With Robust Feature Interpretation for Enhanced Compressive Strength Prediction of Ultra-High-Performance Concrete," *Engineering Reports*, 7. Available at: <https://doi.org/10.1002/eng2.70170>.
- Zhang, Y. et al. (2024) "Predicting the compressive strength of high-performance concrete using an interpretable machine learning model," *Scientific Reports*, 14. Available at: <https://doi.org/10.1038/s41598-024-79502-z>.
- Zhao, Y. et al. (2025a) "Estimation of compressive strength of ultra-high performance lightweight concrete (UHPLC) using neural network," *PLOS One*, 20. Available at: <https://doi.org/10.1371/journal.pone.0326652>.
- Zhao, Y. et al. (2025b) "Estimation of compressive strength of ultra-high performance lightweight concrete (UHPLC) using neural network," *PLOS One*, 20. Available at: <https://doi.org/10.1371/journal.pone.0326652>.
- Zhao, Y. et al. (2025c) "Estimation of compressive strength of ultra-high performance lightweight concrete (UHPLC) using neural network," *PLOS One*, 20. Available at: <https://doi.org/10.1371/journal.pone.0326652>.
- Ziółkowski, P. (2025) "Influence of Optimization Algorithms and Computational Complexity on Concrete Compressive Strength Prediction Machine Learning Models for Concrete Mix Design," *Materials*, 18. Available at: <https://doi.org/10.3390/ma18061386>.