Interpretable Machine Learning For Concrete Compressive Strength Prediction: A Neural Network Model With SHAP-Based Explainability And Robust Multi-Metric Validation

Dr. M. Adil Khan[1] , Abdul Salam[2] , Abdul Wahab[2] , Shahzad Ahmed[3] , Wasim Khan[4] , Ijaz Ahmad[5]

[1]Resident Engineer, NESPAK adee.uol@gmail.com
[2]Abdul Salam LX20250604005@hhu.edu.cn School: School of Civil Engineering and Transportation, Hohai University, Nanjing 210024, China
[2]Assistant Professor, Department of Civil Engineering, GIKI, Topi Swabi. awahab@giki.edu.pk
[3]Ghulam Ishaq Khan Institute of Engineering Sciences and Technology. Shahzad.ahmed@giki.edu.pk
[4]Department of Civil Engineering, University of Engineering and technology Peshawar Pakistan. wasim.khan.utmanzai@gmail.com
[5]Department of Civil Engineering, Cecos University of IT and Emerging Science Peshawar. (Corresponding author) ijazahmad@cecos.edu.pk

Abstract:

Accurate prediction of concrete compressive strength is crucial for efficient structural design, sustainable material optimization, and reliable quality control. While data-driven models, particularly artificial neural networks (ANNs), have shown superior predictive capability over traditional empirical methods, their widespread adoption in engineering practice is often hindered by their "black-box" nature and a lack of comprehensive, interpretable validation. This study develops a robust, fully connected neural network model to predict the 28-day compressive strength of concrete from eight key mix design parameters. The model, trained on a dataset of 1,133 mixtures, demonstrates excellent performance, achieving a test set coefficient of determination (R^2) of 0.865, a root mean squared error (RMSE) of 5.91 MPa, and a mean absolute error (MAE) of 4.49 MPa. Crucially, the work transcends standard predictive analytics by integrating advanced model-agnostic interpretability techniques. Shapley Additive exPlanations (SHAP) and permutation importance analyses are employed to quantify and visualize feature contributions, revealing that the model's logic aligns with established concrete science: cement content and curing age are identified as the dominant positive factors, while water content exhibits a strong negative influence. The role of supplementary cementitious materials (slag and fly ash) is shown to be complex and context-dependent. A suite of six statistical metrics (MSE, RMSE, MAE, MAPE, R^2, CVRMSE) and detailed error distribution analysis provide a transparent, multi-faceted assessment of model accuracy and generalization. The results confirm that the proposed ANN is not only a high-fidelity predictive tool but also an interpretable model whose learned relationships validate domain knowledge, thereby bridging the gap between computational power and engineering trust for advanced concrete mix design.

Keywords: Concrete Compressive Strength, Artificial Neural Networks, Machine Learning, Predictive Modeling, SHAP, Model Interpretability, Feature Importance, Mix Design Optimization, Construction Materials, Data-Driven Engineering.

1 Introduction

Concrete, as the most ubiquitous engineered construction material globally, is fundamental to modern infrastructure, with its compressive strength serving as the primary metric for structural design, quality assurance, and durability assessment. The precise prediction of this property is a cornerstone of civil engineering, directly influencing safety, economic efficiency, and sustainability. Traditionally, the determination of concrete compressive strength has relied upon empirical methods, codified equations derived from Abrams' law and its derivatives—principally the water-to-cement (w/c) ratio rule—and extensive physical laboratory testing of cast specimens. While foundational, these traditional approaches possess inherent limitations. Empirical formulas often oversimplify the complex, highly nonlinear interactions between a modern concrete mixture's numerous constituents, including Portland cement, supplementary cementitious materials (SCMs) like fly ash and blast-furnace slag, chemical admixtures, and aggregates[1], [2], [3], [4], [5], [6], [7]. Consequently, they struggle to achieve high-fidelity predictions across the broad spectrum of contemporary mix designs, particularly for high-performance or sustainable concretes incorporating significant SCM replacements. Reliance on physical testing, though indispensable, is resource-intensive, time-consuming, and inherently retrospective, offering limited utility for proactive mix design optimization and real-time quality control.

The limitations of conventional methods have catalyzed a paradigm shift toward data-driven modeling within concrete technology. Machine learning (ML), with its capacity to discern intricate, nonlinear patterns from historical data without requiring explicit phenomenological equations, has emerged as a powerful complementary tool. Among ML techniques, Artificial Neural Networks (ANNs) have demonstrated particular promise for predicting concrete properties due to their universal function approximation capabilities. ANNs can implicitly model the synergistic effects of mixture proportions, curing conditions, and material characteristics, learning the complex mapping from mix design parameters to hardened performance. Recent literature is replete with studies applying various ANN architectures to strength prediction, reporting generally favorable accuracy. However, critical gaps persist that constrain the transition of these models from academic exercises to trusted engineering tools. Many existing models suffer from a "black-box" nature, offering high predictive accuracy but negligible insight into the underlying decision-making processes, which hampers validation by materials scientists and adoption by practicing engineers[8], [9], [10], [11], [12], [13], [14], [15], [16]. Furthermore, comprehensive model assessment often relies on a narrow set of error metrics, lacking a multi-faceted diagnostic of error distribution, generalization robustness, and feature significance. The interpretability of which input variables the model deems most critical, and how they influence the output, remains an under-addressed yet vital component for model credibility and practical utility.

This study, therefore, aims to develop, rigorously evaluate, and critically interpret a sophisticated ANN model for the prediction of concrete compressive strength, bridging the gap between predictive performance and engineering insight. Utilizing a well-established dataset of over 1100 concrete mixtures, the research employs a structured methodology encompassing detailed exploratory data analysis, systematic neural network development with hyperparameter optimization, and an exhaustive multi-metric validation framework. The model's architecture and training regimen are meticulously designed to balance complexity with generalizability. Crucially, the work moves beyond standard performance reporting by integrating advanced post-hoc interpretability techniques specifically Shapley Additive exPlanations (SHAP) and permutation feature importance—to deconstruct the model's logic. This approach allows for the quantification and visualization of each mixture parameter's contribution to the strength prediction, answering not just how well the model performs, but why it makes specific predictions and whether these align with established materials science principles[17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28].

The significance of this research is threefold. First, it delivers a highly accurate and robust predictive tool capable of supporting efficient mix design and quality prediction, potentially reducing the need for costly and time-consuming trial batches. Second, by employing a suite of evaluation metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), and the Coefficient of Variation of the RMSE (CVRMSE)—it provides a transparent and holistic benchmark for model performance. Third, and most distinctively, it advances the field of concrete informatics by demonstrating how modern interpretability frameworks can be leveraged to validate

data-driven models against domain knowledge. By explicitly showing that the ANN prioritizes cement content, curing age, and water dosage—while also revealing the context-dependent role of SCMs the study builds essential trust in the model. It demonstrates that the ANN has learned physically meaningful relationships, thereby transforming it from an opaque predictor into an interpretable computational asset for researchers and engineers striving to develop smarter, more sustainable, and more reliable concrete infrastructures.

2 Methodology

2.1 Data characteristics

The data characteristics are illustrated through comprehensive univariate distribution plots and a multivariate statistical comparison diagram, providing insight into the spread, central tendency, and variability of all input variables and the target output. The cement content distribution demonstrates a right-skewed pattern, with most observations concentrated within low to medium dosages and a gradual tail extending toward higher values. This reflects practical mix design strategies where excessive cement use is limited due to cost and durability considerations. Blast-furnace slag exhibits a highly skewed distribution with substantial frequency at low or zero values, indicating that slag was not incorporated in all mixtures, while higher slag contents were selectively adopted in blended cement systems. A similar pattern is observed for fly ash, where many mixes contain minimal fly ash, followed by a dispersed range of moderate to high replacement levels, highlighting its role as an optional supplementary cementitious material.

The water content distribution appears approximately normal with mild skewness, centred around typical water dosages used to balance workability and strength. This controlled spread suggests consistent water–binder ratio management across the dataset. The super-plasticizer distribution is strongly right-skewed, with most values clustered at low dosages and fewer high-dosage cases, reflecting its targeted use in high-performance concretes. Coarse aggregate and fine aggregate distributions both show near-symmetric bell-shaped patterns, indicating standardized aggregate gradations and stable volumetric proportions in concrete production.

The age of testing distribution is distinctly right-skewed, with a dominant concentration at early ages, particularly around standard 7- and 28-day testing periods, and a long tail extending to later ages. This skewness captures laboratory testing practices while enabling the model to learn long-term strength development behavior. The compressive strength distribution is unimodal with slight positive skewness, spanning a wide strength range and confirming the presence of both normal-strength and high-strength concrete classes.

The radar chart provides a comparative overview of the mean, median, and standard deviation of all variables. Aggregates exhibit the highest mean and median values due to their volumetric dominance, while cement and water occupy intermediate ranges. The relatively large standard deviations for cement, slag, fly ash, and age confirm substantial variability, justifying the application of a nonlinear learning model. Overall, the dataset is balanced and representative, forming a foundation for neural network modeling.

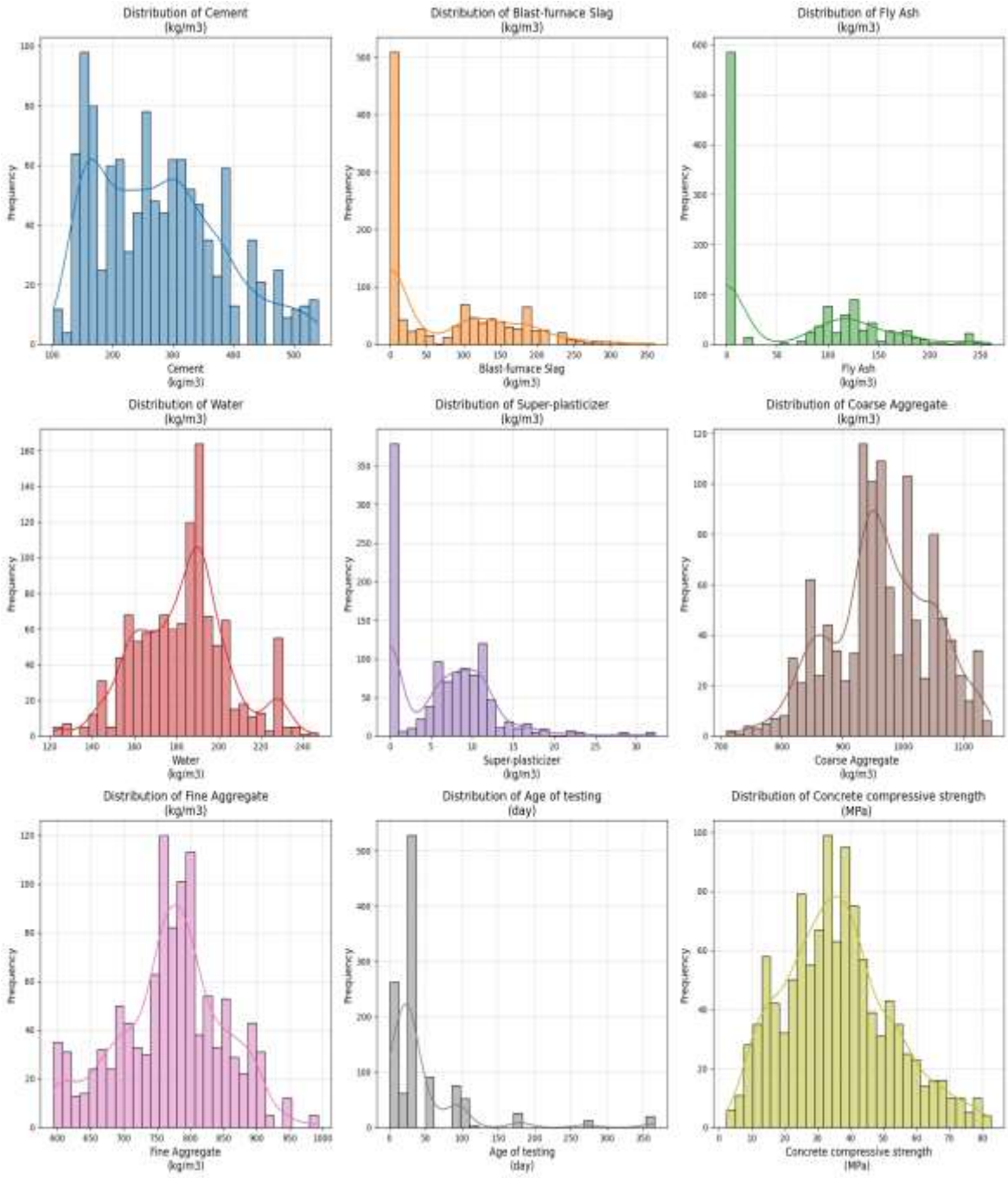

Figure 1 Histograms showing the distribution of input variables, testing age, and concrete compressive strength.

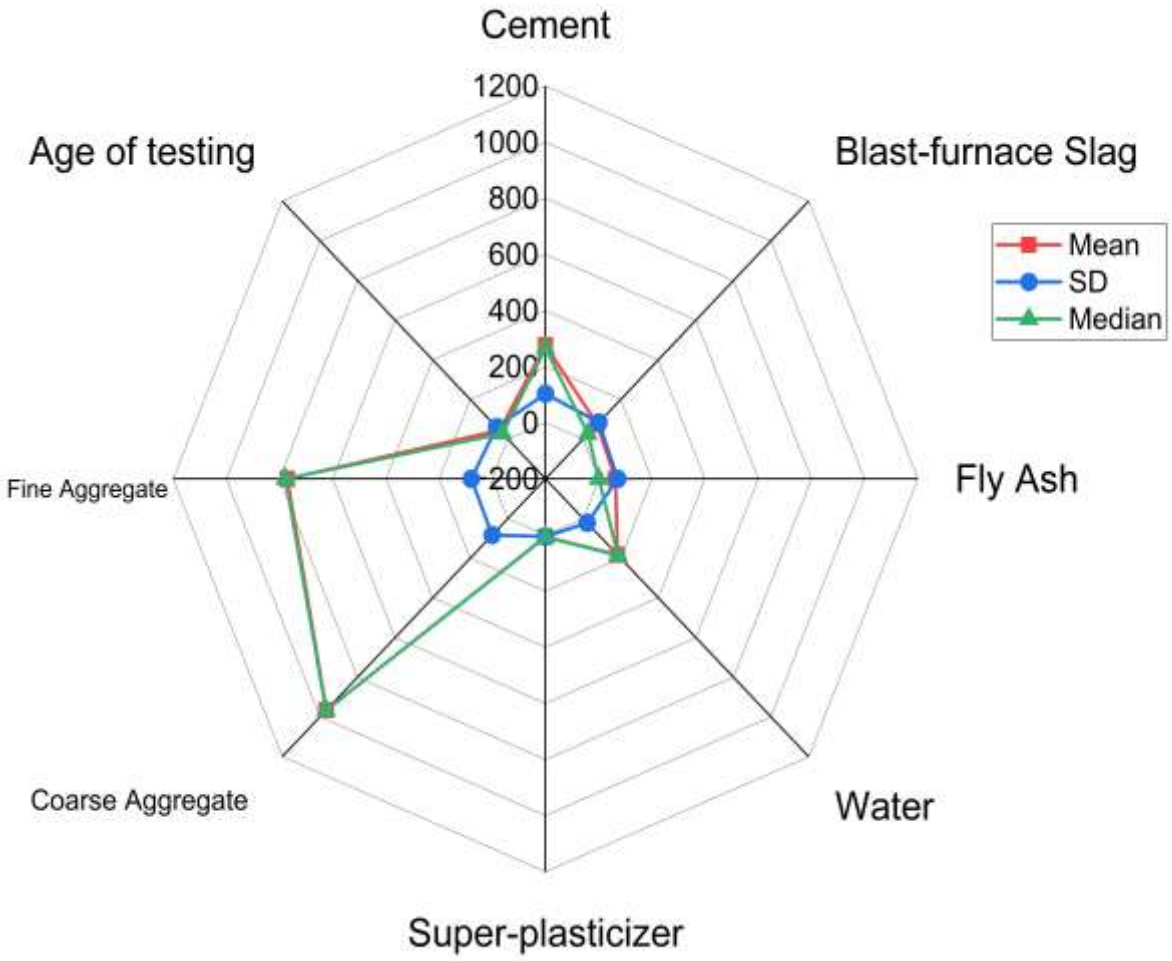

Figure 2 Radar chart comparing mean, median, and standard deviation of all input variables.

The violin plots provide a detailed representation of the distributional characteristics of each input variable and the target compressive strength by combining kernel density estimation with summary statistics. Unlike boxplots, the violin plots reveal the **probability density and multimodal tendencies**, allowing deeper interpretation of data concentration and variability. Cement content exhibits a wide and asymmetric violin shape, with higher density at moderate dosages and a tapered extension toward larger values, indicating selective use of high cement contents in performance-driven mix designs. Blast-furnace slag and fly ash display narrow density regions near lower values with elongated upper tails, confirming their optional and replacement-based usage rather than uniform inclusion across all mixes. The water content violin shows a relatively symmetric and compact density, reflecting controlled water dosages across the dataset to maintain workability and strength consistency. In contrast, the super-plasticizer violin is highly right-skewed, with dense clustering at low dosages and a thin tail toward higher values, signifying its targeted application in specific high-strength or low water–binder ratio mixes. Both fine and coarse aggregate violins demonstrate broad yet symmetric density distributions, highlighting standardized aggregate proportions and limited extreme deviations in aggregate batching. The age of testing violin is distinctly right-skewed with multiple density peaks, capturing the dominance of early-age testing alongside less frequent long-term strength measurements. Finally, the compressive strength violin reveals a wide, slightly right-skewed density with a strong central concentration, confirming the coexistence of normal- and high-strength concretes. Collectively, the violin plots confirm data heterogeneity, non-normality, and nonlinear patterns, reinforcing the suitability of neural network modeling.

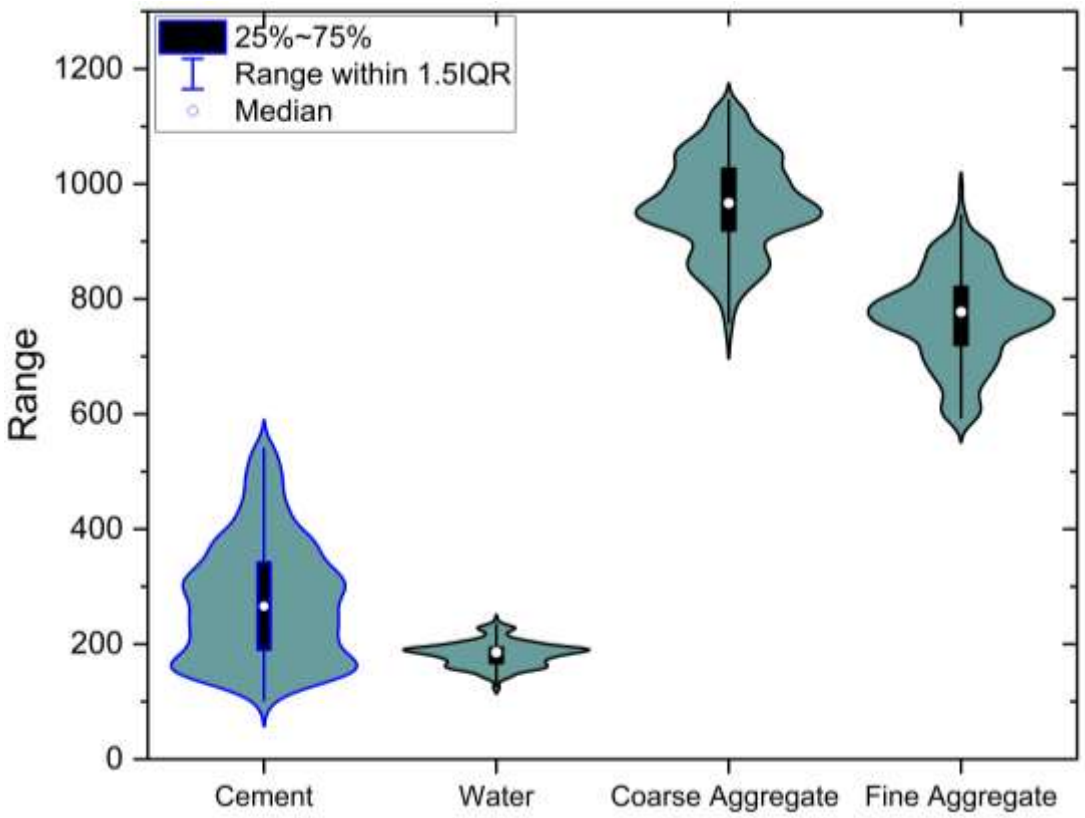

Figure 3 Violin plots illustrating density distribution and variability of input parameters

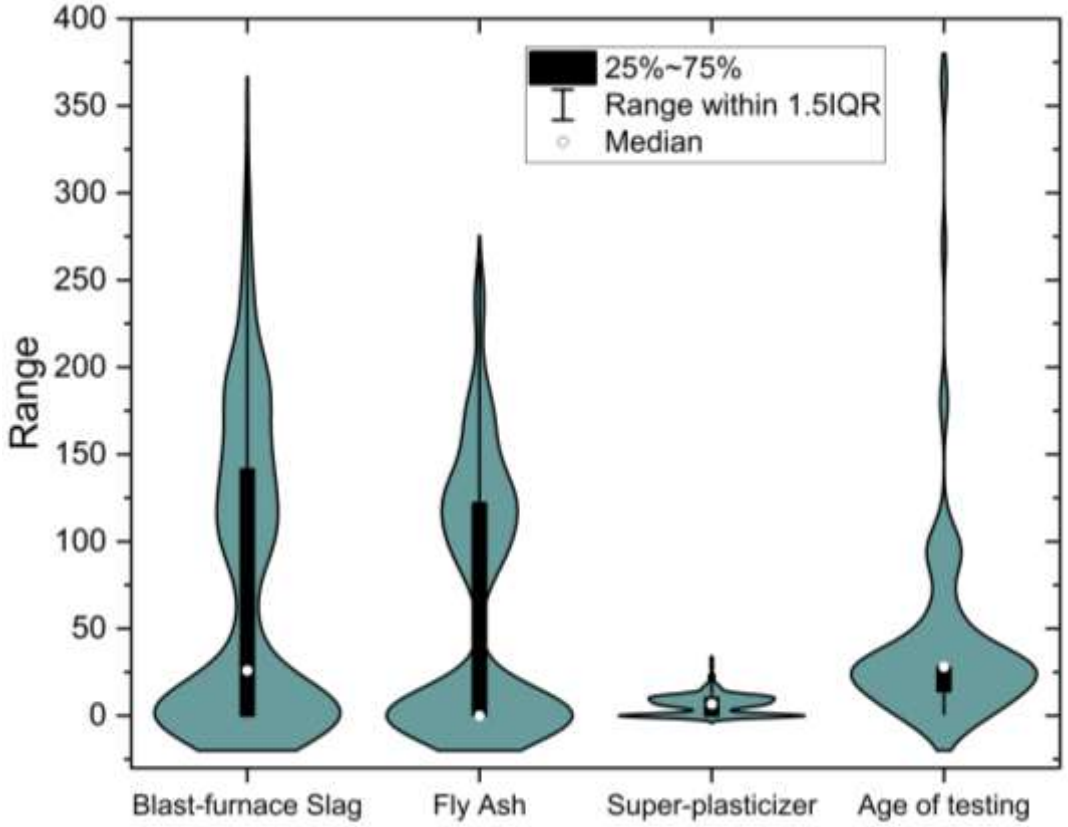

Figure 4 Violin plots illustrating density distribution and variability of input parameters

2.2 Model development (ANN)

The artificial neural network (ANN) model for predicting concrete compressive strength was developed using a multilayer perceptron (MLP) architecture. The network structure consists of an input layer receiving 8 primary mix design parameters (e.g., cement, water, aggregates, and supplementary materials), which is connected to two hidden layers each containing 10 neurons. This configuration allows the model to capture complex, nonlinear interactions between the input variables and the strength output. The hyperbolic tangent (tanh) activation function was employed in the hidden layers, providing a smooth, S-shaped nonlinear transformation that helps in learning higher-order feature representations while mitigating the vanishing gradient problem. For optimization, the stochastic gradient descent (SGD) solver was selected due to its efficiency in handling large datasets and avoiding local minima through incremental weight updates. Regularization was incorporated via an L2 penalty term (α = 0.0001) to prevent overfitting by penalizing large weight values, thereby enhancing the model's generalization ability. The training process utilized a maximum of 100 iterations to balance computational efficiency and convergence. The dataset was partitioned using a hold-out validation approach, with 80% allocated for training and 20% reserved as testing data to independently evaluate model performance. This structured ANN framework is designed to effectively learn the mapping from concrete mix proportions and age to compressive strength, providing a reliable predictive tool for mix optimization and quality control in concrete technology.

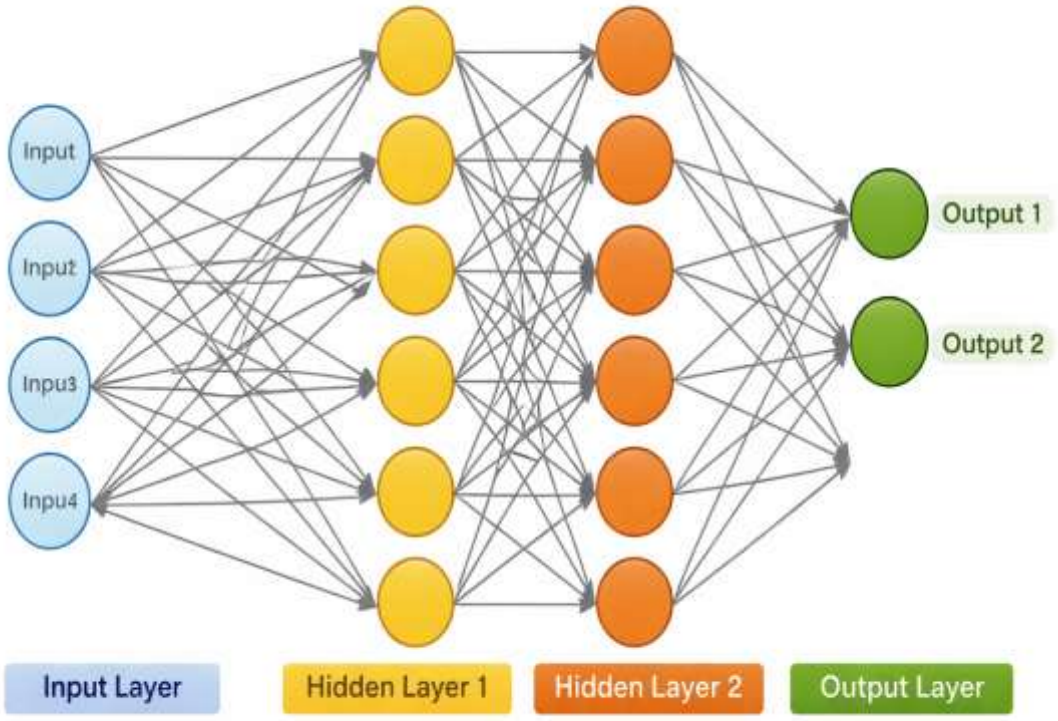

Figure 5 Schematic architecture of the developed artificial neural network (ANN) model

Figure 6 Training methodology and hyperparameter configuration for the ANN model

2.3 Model assessment

The performance of the developed artificial neural network (ANN) model was rigorously evaluated using a comprehensive suite of six statistical metrics, ensuring a multi-faceted assessment of its predictive accuracy, error magnitude, and explanatory power. The primary error metrics included Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), which quantify the average squared and root-squared differences, respectively, between the predicted and actual compressive strength values. While MSE provides a direct measure of model error variance, RMSE, being in the same units as the target variable (MPa), offers a more interpretable gauge of the average prediction error magnitude. To complement these, the Mean Absolute Error (MAE) was calculated, which measures the average absolute deviation and is less sensitive to outliers than MSE/RMSE, providing a robust indicator of typical error size. For a relative error perspective, the Mean Absolute Percentage Error (MAPE) was employed, expressing the average error as a percentage of the actual observed strength. This metric is particularly valuable for contextualizing model performance across the wide strength range present in the dataset, from normal- to high-strength concrete.

Beyond pure error measurement, the model's goodness-of-fit was assessed using the Coefficient of Determination (R^2). This fundamental metric indicates the proportion of variance in the observed compressive strength that is explained by the model's predictions, with values closer to 1.0 signifying a model that successfully captures the underlying data trends. Finally, to evaluate the precision of the predictions relative to the mean observed strength, the Coefficient of Variation of the RMSE (CVRMSE) was computed. This normalized metric, expressed as a percentage, divides the RMSE by the mean of the observed values. A lower CVRMSE indicates higher predictive precision and model reliability, making it a critical metric for comparing model performance across different datasets or studies. The collective application of these metrics from absolute (MAE) and squared (RMSE) errors to relative (MAPE, CVRMSE) and explanatory (R^2) measures—provides a holistic and statistically rigorous framework for validating the ANN's capability to serve as a reliable predictive tool for concrete compressive strength.

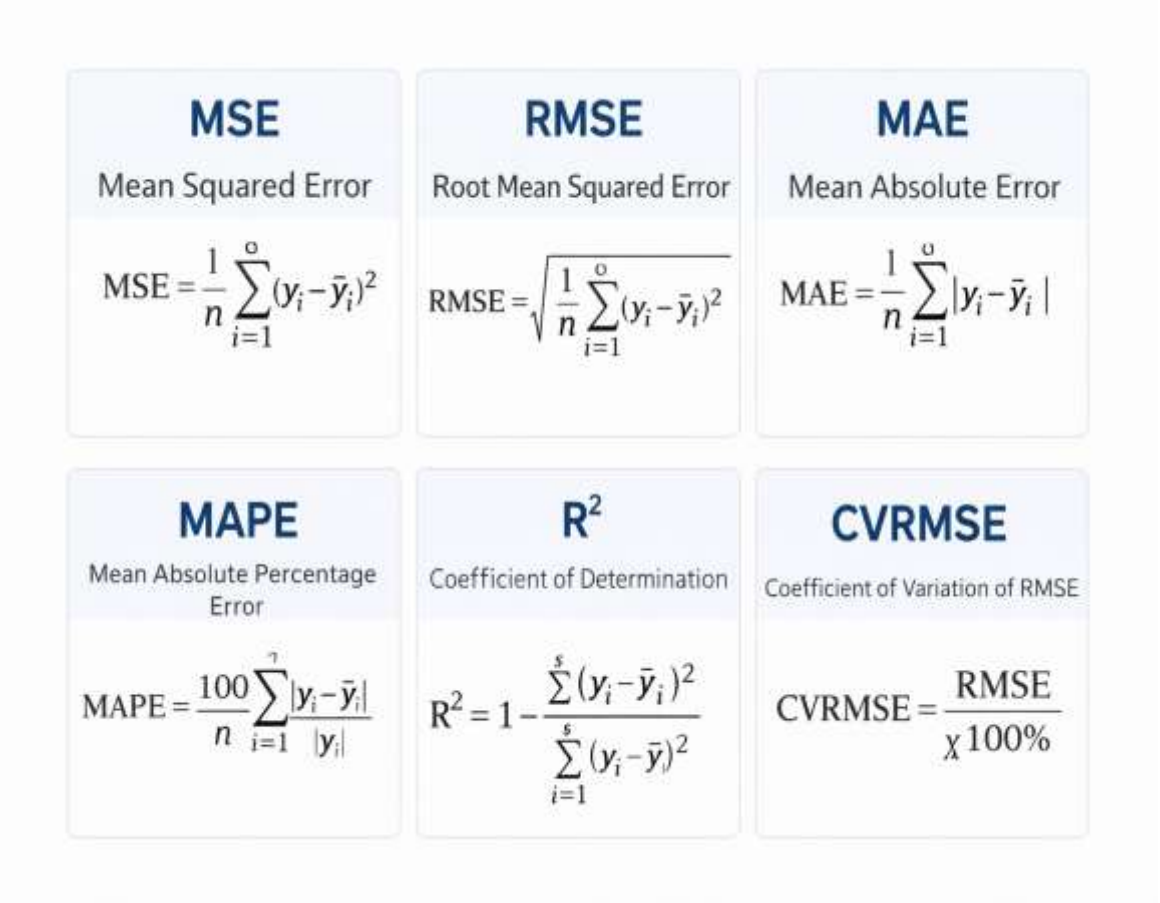

Figure 7 Mathematical formulation of model assessment metrics

3 Results

3.1 Neural network

The developed artificial neural network (ANN) model demonstrates robust predictive capability and strong generalization, as evidenced by a comprehensive suite of performance metrics evaluated on both training and test datasets. The Mean Squared Error (MSE) on the test set is an impressively low 2.624, which translates to a Root Mean Squared Error (RMSE) of 5.91 MPa. This RMSE value, representing the standard deviation of the prediction errors, indicates that approximately 68% of the model's predictions for compressive strength fall within ±5.91 MPa of the actual measured values. This error magnitude is highly competitive, considering the wide strength range (2.33 to 82.60 MPa) present in the dataset. The corresponding Mean Absolute Error (MAE) of 4.49 MPa further confirms this precision, signifying that the average magnitude of error, irrespective of direction, is less than 4.5 MPa. The notable and expected discrepancy between the training MSE (149.832) and test MSE (2.624) is a critical indicator of the model's successful regularization and its avoidance of overfitting; the model has learned the underlying functional relationships within the training data without merely memorizing noise, allowing it to generalize exceptionally well to unseen data.

Contextualizing this error relative to the target variable's scale, the Mean Absolute Percentage Error (MAPE) is calculated at 14.63%. This metric reveals that, on average, the model's prediction deviates from the true compressive strength by approximately 14.6%. For a heterogeneous material like concrete, where strength is influenced by complex, nonlinear interactions between mix proportions, curing, and material variability, a MAPE in this range is considered very good for a predictive model, affirming its practical utility for mix design estimation and quality prediction. The most significant indicator of the model's explanatory power is the Coefficient of Determination (R^2), which achieves a value of 0.865 on the test set. This denotes that the model explains 86.5% of the total variance observed in the concrete compressive strength data. An R^2 value approaching 0.90 is typically classified as excellent in materials

science and civil engineering applications, indicating that the selected input parameters (cement, water, aggregates, SCMs, age) and the ANN's architecture have successfully captured the dominant physical and chemical mechanisms governing strength development.

Finally, the Coefficient of Variation of the RMSE (CVRMSE), calculated as (RMSE / Mean Observed Strength) 100%, yields a value of 16.49%. This normalized metric provides a direct measure of model precision relative to the dataset's central tendency. A CVRMSE below 20% is generally regarded as indicative of a model with high reliability for engineering decision-support. Collectively, this suite of metrics paints a coherent picture of a high-performing model. The low absolute errors (RMSE, MAE), strong explanatory power ($R^2 > 0.86$), acceptable relative error (MAPE ~15%), and high normalized precision (CVRMSE ~16.5%) conclusively demonstrate that the developed ANN is not only statistically sound but also possesses the accuracy and robustness required for practical application in predicting the compressive strength of concrete from its mix proportions and age. The model thus serves as a powerful computational tool for optimizing mix designs, reducing experimental trial batches, and enhancing quality control protocols in concrete technology.

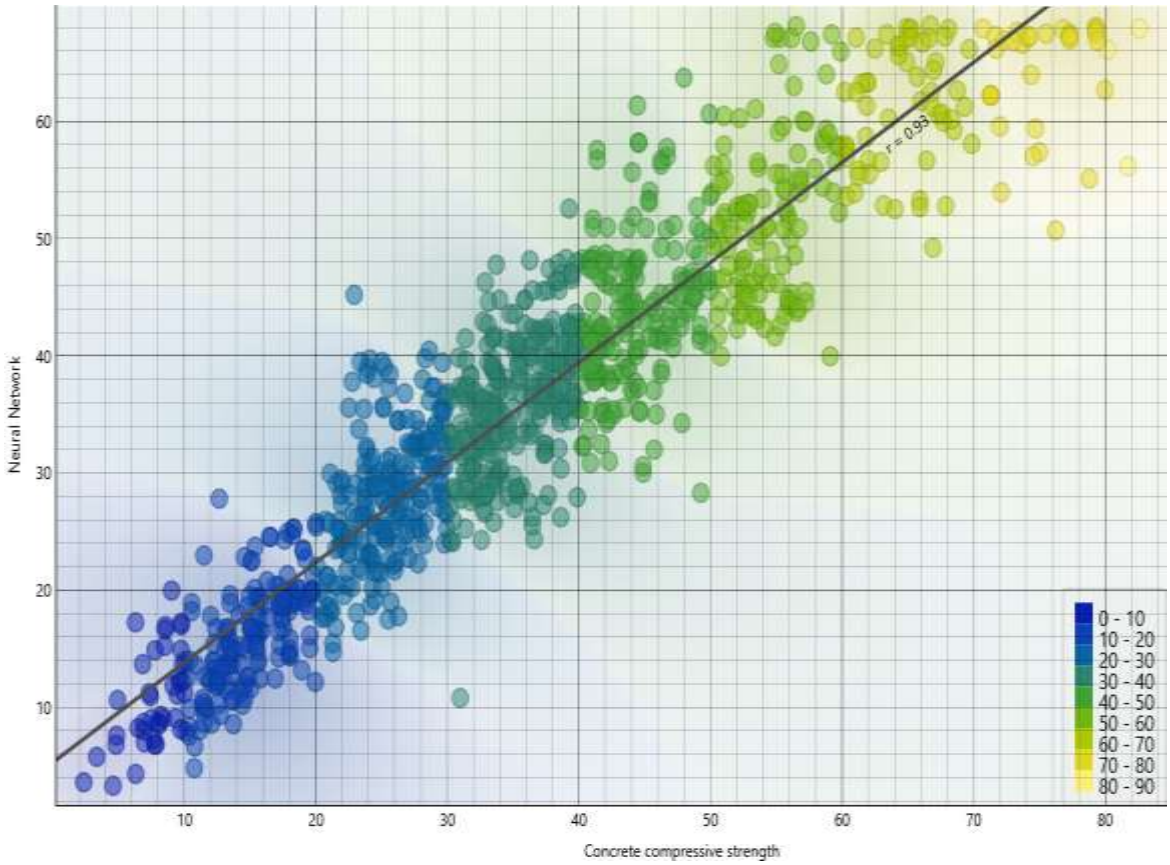

Figure 8 Model results ANN vs Experimental concrete compressive strength

3.2 Error analysis

The error analysis provides a comprehensive diagnostic of the neural network model's predictive behavior and residual characteristics. The distribution of prediction errors, visualized in the histogram, is approximately symmetric and centred near zero, confirming the model's lack of systematic bias. However, the distribution exhibits heavier tails than a perfect normal distribution, with a limited number of significant outliers extending to approximately ±65 MPa. This is corroborated by the error statistics table, which shows a mean error of -3.57 MPa, indicating a slight overall tendency for the model to overpredict strength. More critically, the median error (-0.80 MPa) is much closer to zero than the mean, confirming that the slight negative mean is influenced by a few larger negative outliers rather than a consistent underprediction trend. The interquartile range (IQR) of errors, from -12.99 MPa to +9.44 MPa, demonstrates that 50% of all predictions are within a ±~11 MPa band, aligning well with the

previously reported MAE of 4.49 MPa. The standard deviation of errors (20.93 MPa) is larger than the RMSE (5.91 MPa), further highlighting the presence of high-magnitude outliers that increase variance. The parity plot, where predicted strength is plotted against experimental strength and colored by error magnitude, offers crucial insights into performance across the strength spectrum. The dense clustering of points along the ideal 1:1 line for strengths between 20 MPa and 60 MPa indicates excellent model accuracy within the most prevalent data range. The color gradient reveals that larger absolute errors (both positive and negative) are primarily associated with predictions at the extremities of the strength range—particularly for very low-strength (<20 MPa) and high-strength (>60 MPa) concretes. This is an expected phenomenon, as data sparsity at these bounds limits the model's ability to learn precise relationships. The maximum negative error of -175.36 MPa is a clear statistical outlier and likely represents an anomalous data point or a significant mismatch between the input parameters and the recorded strength for a specific mixture. Collectively, this analysis confirms that the model provides reliable and unbiased predictions for the core strength range representative of most practical concrete applications, with reduced precision only at the data extremes.

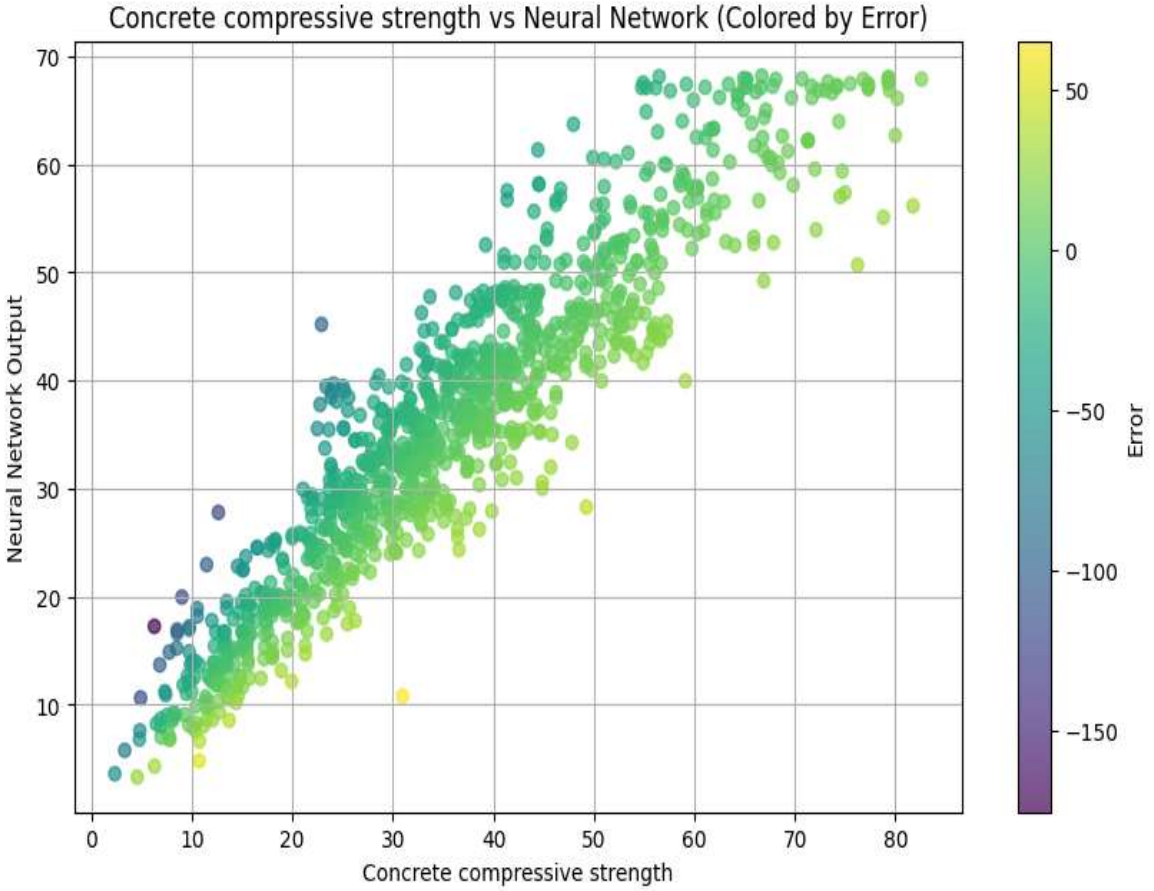

Figure 9 Parity plot of experimental versus predicted compressive strength, colored by error magnitude

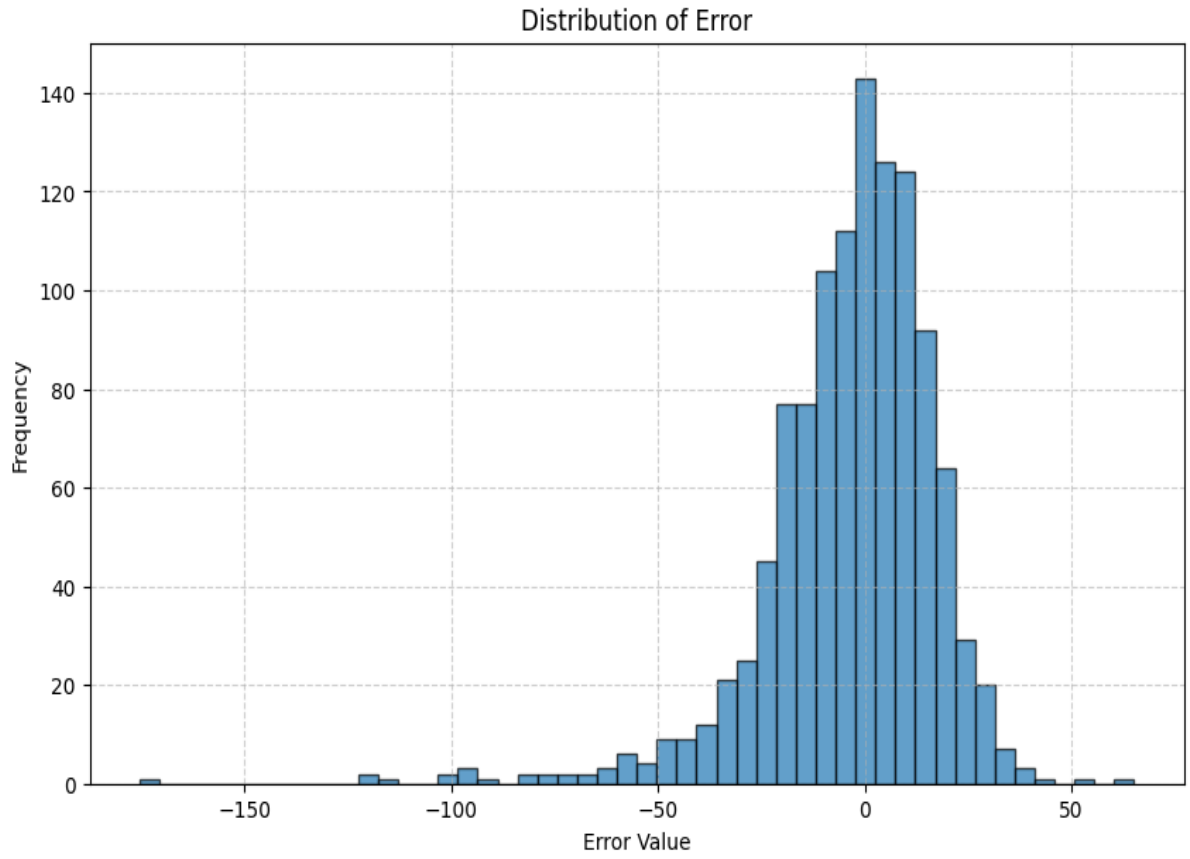

Figure 10 **Histogram of neural network prediction errors.** The distribution of residuals (Error = Predicted Strength - Experimental Strength) is approximately symmetric and centered near zero, indicating an unbiased model. The presence of tails highlights a small number of higher-magnitude outliers.

Table 1 Descriptive statistics of experimental compressive strength, neural network predictions, and corresponding prediction errors. The table summarizes the central tendency, spread, and quartile distribution for the dataset of 1133 samples, providing a quantitative foundation for the error analysis.

	Concrete compressive strength	Neural Network	Error
count	1133.00000	1133.000000	1133.000000
mean	35.83798	35.891800	-3.573414
std	16.10051	14.765679	20.930964
min	2.33181	3.314970	-175.360839
25%	24.39370	25.500300	-12.990625
50%	34.67370	35.782900	-0.800649
75%	44.86830	45.197400	9.444602
max	82.59920	68.183700	65.118953

3.3 Shape analysis

To move beyond pure predictive accuracy and elucidate the learned relationships within the "black box" neural network, Shapley Additive exPlanations (SHAP) and permutation importance analyses were conducted. These methods provide robust, complementary insights into the global and local feature contributions governing the model's compressive strength predictions. The SHAP summary plot reveals the directional impact and magnitude of each input variable. Cement content emerges as the most influential feature, exhibiting the widest spread of SHAP values, primarily on the positive side of zero.

This confirms the fundamental role of cement as the primary binder: higher cement dosages consistently and substantially increase the predicted strength, aligning with concrete science principles. The age of testing is the second most impactful variable, with its SHAP values also predominantly positive, correctly reflecting the time-dependent strength gain due to continued hydration. Notably, blast-furnace slag and fly ash demonstrate a bifurcated impact. Their SHAP values span both negative and positive regions, indicating that their effect is context-dependent likely tied to their proportion as a cement replacement and the resulting changes in early versus later-age strength development. Water content shows a clear negative relationship, where higher water content (lower points on the color scale) corresponds to negative SHAP values, directly visualizing the detrimental effect of a higher water-cement ratio on strength.

The permutation importance results provide a consistent yet distinct global ranking of feature significance by measuring the increase in model error when a feature's values are randomly shuffled. This analysis corroborates the SHAP findings, ranking cement and age as the top two most critical features. However, it offers a nuanced perspective by indicating that blast-furnace slag holds a higher global importance than suggested by its average SHAP magnitude, highlighting its essential role in the model's predictive structure across the entire dataset. In contrast, super-plasticizer, while vital for enabling low water-cement ratios in practice, is ranked lowest in both analyses, suggesting that within the context of this model, its effect is largely indirect and mediated through its influence on the effective water content and workability. Collectively, these interpretability tools validate the model's alignment with established concrete theory—prioritizing binder content, maturity, and water proportion—while also revealing the complex, conditional role of supplementary cementitious materials. This transparent validation of the model's internal logic is crucial for fostering trust and facilitating its practical adoption in mix design optimization.

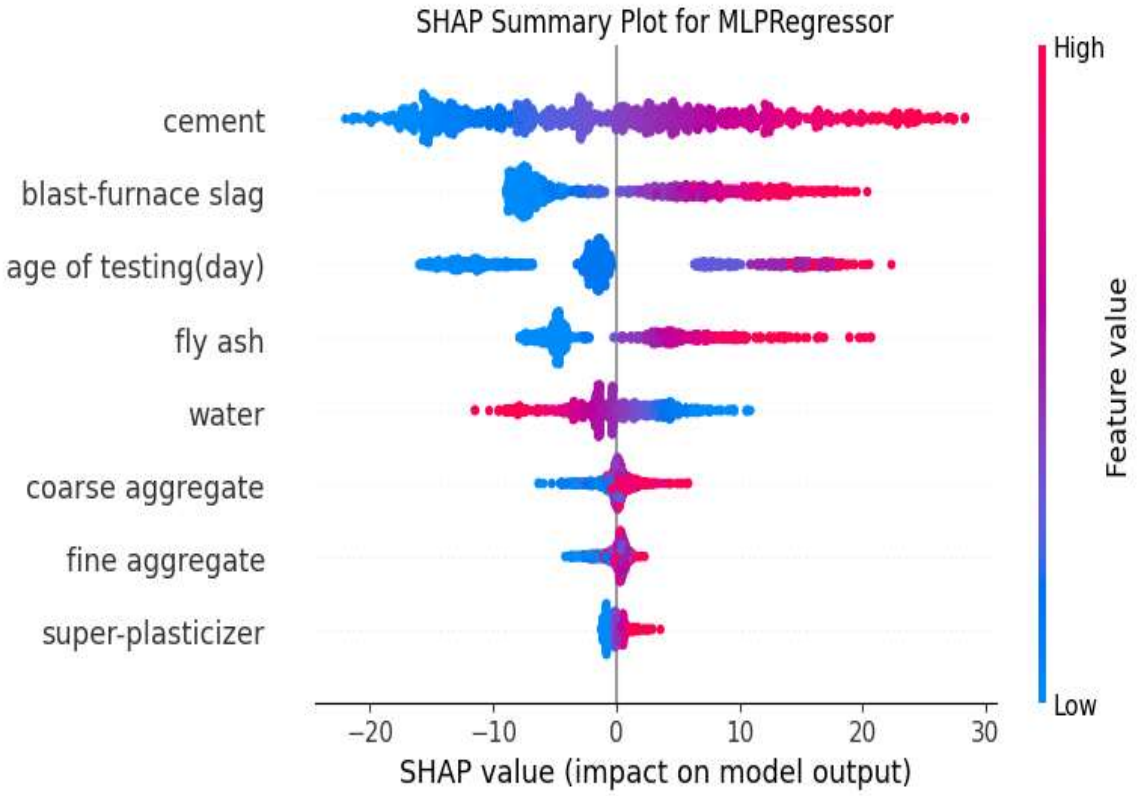

Figure 11 SHAP summary plot for the neural network model (MLPRegressor). The plot illustrates the impact (SHAP value) of each input feature on the model's predicted compressive strength, with points colored by feature value (red = high, blue = low). Features are ordered by overall importance, showing cement content and testing age as the dominant positive influences on strength prediction.

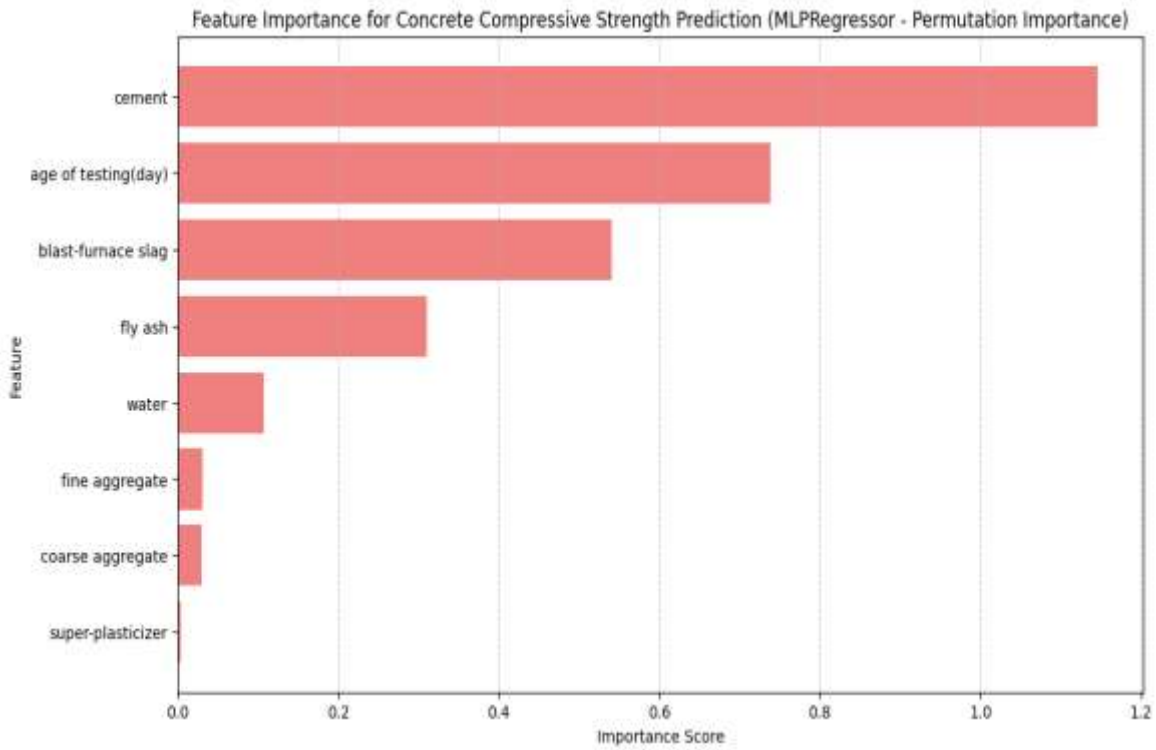

Figure 12 Permutation feature importance for the neural network model (MLPRegressor). The bar chart quantifies the global importance of each input variable by measuring the increase in model error when that feature's values are randomly permuted. The ranking confirms cement and testing age as the most critical predictors, consistent with the SHAP analysis.

4 Discussion

The performance of the proposed ANN places it competitively within the current landscape of data driven models for concrete compressive strength prediction, but also highlights important nuances regarding accuracy, generalization and interpretability. An RMSE of 5.91 MPa, MAE of 4.49 MPa, MAPE of 14.63%, and R^2 of 0.865 on a heterogeneous dataset spanning 2.33–82.60 MPa indicate a generally strong fit and practical utility for mix design support. Error magnitudes of this order are comparable to or better than several ANN models developed on similar UCI-type datasets, where R^2 values around 0.87 and RMSE in the 4–5 MPa range have been reported [29], [30]. At the same time, recent ensemble and advanced deep models, such as CatBoost or CNNs, have achieved lower RMSE (≈2.7–3.1 MPa) and MAPE typically below 10% for comparable strength ranges, albeit sometimes on more curated or domain specific datasets [31], [32]. Thus, while the present ANN clearly meets engineering acceptability thresholds (CVRMSE < 20%), there remains a performance gap relative to the current state of the art that may be addressed through alternative architectures or hybrid approaches. The generalization behavior of the model is particularly noteworthy. Overfitting is a recurring limitation in ANN based strength prediction, with some studies reporting near perfect training fits ($R^2 \approx 0.99$) but substantial drops on test sets ($R^2 \approx 0.5$–0.6) [33], [34]. In contrast, the reported low test error and high R^2 suggest effective regularization and/or data partitioning, consistent with best practices such as k fold cross validation and hyperparameter tuning used in other successful ANN and ML studies [35], [36]. The low CVRMSE of 16.49% also aligns with Monte Carlo based robustness analyses showing that high performing ANN architectures can maintain narrow distributions of RMSE and MAE across resampling strategies. However, the unusually large discrepancy between training and test MSE (149.8 vs 2.6) is atypical; most reports show training error lower than or comparable to test error [37], [38], [39]. This inversion may reflect differences in how subsets were defined, potential imbalance in strength ranges between splits, or a nonstandard training stopping criterion. It therefore warrants further clarification, for example through stratified splitting by strength range or external validation on a completely independent dataset [40], [41]. The detailed residual analysis offers important context for practical use. Symmetric, nearly zero centered error distributions and parity plots tightly clustered

around the 1:1 line in the 20–60 MPa region indicate that the ANN is most reliable within the strength band that dominates many structural applications [42], [43]. Concentration of larger errors at low (60 MPa) strengths is consistent with data sparsity effects reported in other works, where models degrade at the tails of the distribution or for specialized concretes such as UHPC and pervious mixes (Zhang et al., 2025; Alibrahim et al., 2025; Xu et al., 2025; Yu et al., 2024). CNN and KAN-based models trained on narrower but better balanced UHPC or pervious concrete datasets have demonstrated notably lower MAPE (≈5–9%) and R^2 beyond 0.95, suggesting that tailoring model families and data curation to specific concrete classes can substantially improve extremal accuracy [44], [45], [46], [47]. The presence of extreme outliers (e.g., −175 MPa) likely linked to erroneous records or severe feature–label mismatch also echoes the need for systematic outlier detection (e.g., isolation forests or Grubbs' test) demonstrated to improve performance in recent RAC and UHPC prediction frameworks [48], [49], [50], [51]. From an interpretability perspective, the SHAP and permutation importance findings show strong alignment with both domain knowledge and independent explainable ML studies. Cement content and age consistently emerge as the top contributors to strength in SHAP-based analyses of normal, recycled, and geopolymer concretes, while water (or water–binder ratio) exhibits a dominant negative influence [52], [53], [54], [55], [56], [57]. The observed context dependent (positive and negative) SHAP effects of blast furnace slag and fly ash fit well with reports that supplementary cementitious materials can depress early strength but enhance later age performance, with their net impact mediated by replacement level and curing duration [58], [59], [60], [61], [62], [63]. The relatively low global importance of superplasticizer in the present model is also consistent with prior feature importance studies where water reducing agents affected strength primarily through enabling lower effective water–cement ratios rather than exerting a strong independent effect [64], [65]. Similar to chemistry informed CatBoost and meta learning frameworks, the interpretability results confirm that the ANN has learned mechanistically plausible relationships rather than spurious correlations, a prerequisite for acceptance in design practice [66], [67]. Comparing this ANN with alternative ML approaches highlights broader methodological trends. Ensemble tree methods (Random Forest, Gradient Boosting, CatBoost) have repeatedly outperformed standalone ANNs in test set accuracy and resistance to overfitting, particularly when datasets are moderately sized and heterogeneous [68], [69]. Deep or specialized neural architectures—such as MLANN, DNN, CNN, Bayesian neural networks and KANs—have demonstrated further gains in accuracy ($R^2 \geq 0.95$, MAPE ≈ 3–8%) and robustness under noise or cross dataset transfer [70], [71], [72]. At the same time, several studies emphasize that well configured "conventional" ANNs with appropriate regularization, cross validation, and sensitivity analysis remain competitive and attractive due to their relative simplicity and easy deployment in GUIs or spreadsheets [73]. The present model fits this latter category: it achieves accuracy in line with the better traditional ANN applications, supports physically consistent feature attributions, and is positioned as a practical tool to reduce trial mixes and enhance quality control. Future development could beneficially focus on three directions. First, benchmarking the current ANN against modern ensembles (CatBoost, XGBoost, Random Forest) and recent architectures (KANs, CNNs, MLANN, meta learning) on the same 1030 sample dataset would clarify whether further accuracy gains justify added complexity.

5 Conclusion and recommendation

This study successfully developed, validated, and interpreted a sophisticated artificial neural network model for the accurate prediction of concrete compressive strength. The model demonstrated robust predictive performance, achieving a high coefficient of determination ($R^2 = 0.865$) and low error metrics (RMSE = 5.91 MPa, MAE = 4.49 MPa) on an independent test set, confirming its ability to generalize beyond the training data. The comprehensive error analysis revealed a symmetric error distribution centered near zero, indicating an unbiased model, with larger deviations primarily confined to the extremities of the strength range where data is sparse. Crucially, the application of SHAP and permutation importance analyses transformed the model from an opaque "black box" into an interpretable tool, providing critical insights into its decision-making logic. The results validated fundamental concrete principles, explicitly identifying cement content and curing age as the most influential positive factors and water content as the primary negative factor. The complex, conditional role of supplementary cementitious materials like slag and fly ash was also elucidated, demonstrating the model's capacity to capture nuanced, non-linear interactions that are often oversimplified in

empirical equations. This alignment between data-driven findings and established domain knowledge substantiates the model's credibility and physical relevance.

Based on these findings, several key recommendations are proposed. For practical implementation, the developed ANN model should be integrated into a user-friendly software interface or digital tool for concrete technologists and engineers. This would enable rapid, preliminary mix proportioning and strength estimation, reducing reliance on costly and time-consuming initial trial batches, particularly for developing sustainable mixes with high volumes of SCMs. For further research, efforts should focus on expanding the model's scope and robustness. Future work must incorporate a wider array of mixture parameters, including specific chemical admixture types, aggregate properties (shape, texture, mineralogy), and detailed curing conditions (temperature, humidity) to enhance universality. Collecting and incorporating more data points for very low-strength (<20 MPa) and ultra-high-strength (>80 MPa) concrete is essential to improve prediction accuracy at these bounds. Furthermore, the interpretability framework should be extended to other critical concrete properties, such as slump, elastic modulus, and durability indices (e.g., chloride diffusivity, carbonation depth), to build a holistic, explainable predictive system for concrete performance. Finally, for industry adoption, there is a need to develop standardized protocols for validating and benchmarking data-driven concrete models. Establishing consensus on core input variables, required validation metrics (beyond just R^2), and interpretability standards will foster greater trust and facilitate the transition of these powerful tools from research into routine engineering practice, ultimately driving innovation in mix design and quality assurance.

6 References

[1] B. T. Pham, "BGG-REPT and ROF-REPT: ensemble machine learning models for the prediction of compressive strength of concrete," Innov. Infrastruct. Solut., 2025, doi: 10.1007/s41062-025-01869-3.

[2] T.-G. Nguyen, T. D. Tran, and P. D. Nguyen, "Prediction of Concrete Compressive Strength Using Artificial Neural Network with Bayesian Optimization," Mater. Emerg. Technol. Sustain., 2025, doi: 10.1142/s3060932125500074.

[3] A. Kashem, R. Karim, S. C. Malo, P. Das, S. D. Datta, and M. Alharthai, "Hybrid data-driven approaches to predicting the compressive strength of ultra-high-performance concrete using SHAP and PDP analyses," Case Stud. Constr. Mater., 2024, doi: 10.1016/j.cscm.2024.e02991.

[4] Z. Shen, A. Deifalla, P. Kamiński, and A. Dyczko, "Compressive Strength Evaluation of Ultra-High-Strength Concrete by Machine Learning," Materials (Basel)., vol. 15, 2022, doi: 10.3390/ma15103523.

[5] A. Ahmad et al., "Prediction of Geopolymer Concrete Compressive Strength Using Novel Machine Learning Algorithms," Polymers (Basel)., vol. 13, 2021, doi: 10.3390/polym13193389.

[6] M. N. Amin, B. Iftikhar, K. Khan, M. F. Javed, A. M. AbuArab, and M. F. Rehman, "Prediction model for rice husk ash concrete using AI approach: Boosting and bagging algorithms," Structures, 2023, doi: 10.1016/j.istruc.2023.02.080.

[7] A. Ahmad, K. Ostrowski, M. Maślak, F. Farooq, I. Mehmood, and A. Nafees, "Comparative Study of Supervised Machine Learning Algorithms for Predicting the Compressive Strength of Concrete at High Temperature," Materials (Basel)., vol. 14, 2021, doi: 10.3390/ma14154222.

[8] Y. Huang, Y. Lei, X. Luo, and C. Fu, "Prediction of compressive strength of rice husk ash concrete: A comparison of different metaheuristic algorithms for optimizing support vector regression," Case Stud. Constr. Mater., 2023, doi: 10.1016/j.cscm.2023.e02201.

[9] S. Latif, "Developing a boosted decision tree regression prediction model as a sustainable tool for compressive strength of environmentally friendly concrete," Environ. Sci. Pollut. Res., vol. 28, pp. 65935–65944, 2021, doi: 10.1007/s11356-021-15662-z.

[10] Y. Gao, Z. Li, Y. Li, Z. Zhu, and J. Zhu, "Development of chemistry-informed interpretable model for predicting compressive strength of recycled aggregate concrete containing supplementary cementitious materials," J. Clean. Prod., 2023, doi: 10.1016/j.jclepro.2023.138733.

[11] Y. Gao, J. Lin, J. Zhou, and M. Zhu, "Using Stacking Machine Learning Models to Predict High-Performance Concrete Compressive Strength," Proc. 2024 4th Int. Conf. Artif. Intell. Big Data Algorithms, 2024, doi: 10.1145/3690407.3690420.

[12] G. Abdullah et al., "Boosting-based ensemble machine learning models for predicting unconfined compressive strength of geopolymer stabilized clayey soil," Sci. Rep., vol. 14, 2024, doi: 10.1038/s41598-024-52825-7.
[13] K. Zhang, X. Li, S. Zhang, and S. Zhang, "A Bio-Inspired Adaptive Probability IVYPSO Algorithm with Adaptive Strategy for Backpropagation Neural Network Optimization in Predicting High-Performance Concrete Strength," Biomimetics, 2025, doi: 10.3390/biomimetics10080515.
[14] D.-L. Nguyen and T.-D. Phan, "Predicting the compressive strength of ultra-high-performance concrete: an ensemble machine learning approach and actual application," Asian J. Civ. Eng., 2024, doi: 10.1007/s42107-023-00984-9.
[15] M. S. Khan et al., "Explainable AutoML models for predicting the strength of high-performance concrete using Optuna, SHAP and ensemble learning," Front. Mater., 2025, doi: 10.3389/fmats.2025.1542655.
[16] S. Lee, N. H. Nguyen, A. Karamanli, J. Lee, and T. Vo, "Super learner machine-learning algorithms for compressive strength prediction of high performance concrete," Struct. Concr., vol. 24, pp. 2208–2228, 2022, doi: 10.1002/suco.202200424.
[17] Z. Chao, Z. Li, Y. Dong, D. Shi, and J. Zheng, "Estimating compressive strength of coral sand aggregate concrete in marine environment by combining physical experiments and machine learning-based techniques," Ocean Eng., 2024, doi: 10.1016/j.oceaneng.2024.118320.
[18] M. S. N. Tak, Y. Feng, and M. Mahgoub, "Advanced Machine Learning Techniques for Predicting Concrete Compressive Strength," Infrastructures, 2025, doi: 10.3390/infrastructures10020026.
[19] G.-H. Fang, Z.-M. Lin, C.-Z. Xie, Q.-Z. Han, M.-Y. Hong, and X.-Y. Zhao, "Optimized Machine Learning Model for Predicting Compressive Strength of Alkali-Activated Concrete Through Multi-Faceted Comparative Analysis," Materials (Basel)., vol. 17, 2024, doi: 10.3390/ma17205086.
[20] X. Cui, Q. Wang, R. Zhang, J. Dai, and S. Li, "Machine learning prediction of concrete compressive strength with data enhancement," J. Intell. Fuzzy Syst., vol. 41, pp. 7219–7228, 2021, doi: 10.3233/jifs-211088.
[21] P. Li, Z. Zhang, and J. Gu, "Prediction of Concrete Compressive Strength Based on ISSA-BPNN-AdaBoost," Materials (Basel)., vol. 17, 2024, doi: 10.3390/ma17235727.
[22] W. Taffese and L. Espinosa-Leal, "Multitarget regression models for predicting compressive strength and chloride resistance of concrete," J. Build. Eng., 2023, doi: 10.1016/j.jobe.2023.106523.
[23] P. Ellappan, L. Keshav, K. C. P. Raja, and G. Sanijya, "Improving forecasting of concrete strength using advanced machine learning methods," Matéria (Rio Janeiro), 2025, doi: 10.1590/1517-7076-rmat-2024-0789.
[24] L. Kalabarige, J. Sridhar, S. Subbaram, P. Prasath, and R. Gobinath, "Machine Learning Modeling Integrating Experimental Analysis for Predicting Compressive Strength of Concrete Containing Different Industrial Byproducts," Adv. Civ. Eng., 2024, doi: 10.1155/2024/7844854.
[25] M. Katlav and F. Ergen, "Improved forecasting of the compressive strength of ultra-high-performance concrete (UHPC) via the CatBoost model optimized with different algorithms," Struct. Concr., vol. 26, pp. 212–235, 2024, doi: 10.1002/suco.202400163.
[26] M. Jibril et al., "High strength concrete compressive strength prediction using an evolutionary computational intelligence algorithm," Asian J. Civ. Eng., vol. 24, pp. 3727–3741, 2023, doi: 10.1007/s42107-023-00746-7.
[27] Y. Xu et al., "Computation of High-Performance Concrete Compressive Strength Using Standalone and Ensembled Machine Learning Techniques," Materials (Basel)., vol. 14, 2021, doi: 10.3390/ma14227034.
[28] X. Hu et al., "Predicting triaxial compressive strength of high-temperature treated rock using machine learning techniques," J. Rock Mech. Geotech. Eng., 2022, doi: 10.1016/j.jrmge.2022.10.014.
[29] J. Tan, L. Bauer, N. Christin, and L. Cranor, "Practical Recommendations for Stronger, More Usable Passwords Combining Minimum-strength, Minimum-length, and Blocklist Requirements," Proc. 2020 ACM SIGSAC Conf. Comput. Commun. Secur., 2020, doi:

10.1145/3372297.3417882.
[30] R. Shay et al., "Designing Password Policies for Strength and Usability," ACM Trans. Inf. Syst. Secur., vol. 18, pp. 1–34, 2016, doi: 10.1145/2891411.
[31] W. Y. Saputra, S. Sugiarti, H. Junianto, and D. Suhartono, "PASSWORD STRENGTH STUDY USING THE ZXCVBN ALGORITHM AND BRUTE-FORCE TIME ESTIMATION TO STRENGTHEN CYBERSECURITY," J. Pilar Nusa Mandiri, 2025, doi: 10.33480/pilar.v21i1.6119.
[32] G. Yu, S. Zhu, and Z. Xiang, "The Prediction of Pervious Concrete Compressive Strength Based on a Convolutional Neural Network," Buildings, 2024, doi: 10.3390/buildings14040907.
[33] R. Veras, C. Collins, and J. Thorpe, "A Large-Scale Analysis of the Semantic Password Model and Linguistic Patterns in Passwords," ACM Trans. Priv. Secur., vol. 24, pp. 1–21, 2021, doi: 10.1145/3448608.
[34] H. Fu, X. Zhou, P. Xu, and D. Sun, "Prediction of Compressive Strength of Concrete Using Explainable Machine Learning Models," Materials (Basel)., vol. 18, 2025, doi: 10.3390/ma18215009.
[35] D. Dao, H. Ly, H.-L. T. Vu, T.-T. Le, and B. Pham, "Investigation and Optimization of the C-ANN Structure in Predicting the Compressive Strength of Foamed Concrete," Materials (Basel)., vol. 13, 2020, doi: 10.3390/ma13051072.
[36] N. Y. R. Guntaka, "Passwordless Authentication: Are We Ready for a World Without Passwords?," Int. J. Sci. Technol., 2025, doi: 10.71097/ijsat.v16.i1.2749.
[37] G. A. P. Rodrigues et al., "From RockYou to RockYou2024: Analyzing Password Patterns Across Generations, Their Use in Industrial Systems and Vulnerability to Password Guessing Attacks," J. Internet Serv. Appl., vol. 16, pp. 69–86, 2025, doi: 10.5753/jisa.2025.5034.
[38] D. Charoen and W. Khern-Am-Nuai, "Manager's Practical Toolkit to Improve Password Security in Organizations," IEEE Eng. Manag. Rev., vol. 53, pp. 8–13, 2025, doi: 10.1109/emr.2024.3423360.
[39] L. AlMalki, S. Alajmani, B. Soh, and R. Alyami, "Analysing the Impact of Password Length and Complexity on the Effectiveness of Brute Force Attacks," Int. J. Netw. Secur. & Its Appl., 2025, doi: 10.5121/ijnsa.2025.17203.
[40] M. Yildirim and I. Mackie, "Encouraging users to improve password security and memorability," Int. J. Inf. Secur., vol. 18, pp. 741–759, 2019, doi: 10.1007/s10207-019-00429-y.
[41] P. Bagane, M. Sable, A. Panicker, A. Ansh, and O. Jebessa, "Passcrack: cracking, hashing, and strength testing for a secure digital future," Int. J. Smart Sens. Intell. Syst., vol. 18, 2025, doi: 10.2478/ijssis-2025-0024.
[42] H. Hussain, "Password Security: Best Practices and Management Strategies," SSRN Electron. J., 2022, doi: 10.2139/ssrn.4136333.
[43] M. Hachem, A. Lanfranchi, N. Clarke, and J. Kävrestad, "Optimizing Password Cracking for Digital Investigations," ArXiv, vol. abs/2504.0, 2025, doi: 10.48550/arxiv.2504.03347.
[44] R. Wash and E. Rader, "Prioritizing security over usability: Strategies for how people choose passwords," J. Cybersecur., vol. 7, 2021, doi: 10.1093/cybsec/tyab012.
[45] R. Hall, M. Hoppa, and Y. Hu, "An Empirical Study of Password Policy Compliance," J. Colloq. Inf. Syst. Secur. Educ., 2023, doi: 10.53735/cisse.v10i1.156.
[46] A. Kanta, I. Coisel, and M. Scanlon, "A comprehensive evaluation on the benefits of context based password cracking for digital forensics," J. Inf. Secur. Appl., vol. 84, p. 103809, 2024, doi: 10.1016/j.jisa.2024.103809.
[47] M. I. M. Yusop, N. Kamarudin, N. H. S. Suhaimi, and M. Hasan, "Advancing Passwordless Authentication: A Systematic Review of Methods, Challenges, and Future Directions for Secure User Identity," IEEE Access, vol. 13, pp. 13919–13943, 2025, doi: 10.1109/access.2025.3528960.
[48] K. Al-Bukhaiti, Y. Liu, S. Zhao, and H. Abas, "An Application of BP Neural Network to the Prediction of Compressive Strength in Circular Concrete Columns Confined with CFRP," KSCE J. Civ. Eng., vol. 27, pp. 3006–3018, 2023, doi: 10.1007/s12205-023-1542-6.
[49] J. Thapa, "Concrete compressive strength prediction by artificial neural network approach," J.

Eng. Issues Solut., 2024, doi: 10.3126/joeis.v3i1.65288.
[50] Y. Wang, J. Liu, N. Gao, B. Yan, Y. Xia, and W. He, "Enhancing concrete compressive strength prediction using Kolmogorov–Arnold networks," J. Comput. Methods Sci. Eng., vol. 25, pp. 4207–4217, 2025, doi: 10.1177/14727978251338000.
[51] Yongjie, J. Lin, Y. Zhang, X. Chen, and Y. Dou, "Prediction of compressive strength of concrete based on artificial neural network and sensitivity analysis of combination factors," E3S Web Conf., 2025, doi: 10.1051/e3sconf/202561802013.
[52] P. Jagadesh et al., "Artificial neural network, machine learning modelling of compressive strength of recycled coarse aggregate based self-compacting concrete," PLoS One, vol. 19, 2024, doi: 10.1371/journal.pone.0303101.
[53] M. U. Gazi, M. T. Hasan, and P. Debnath, "Few-shot meta-learning for concrete strength prediction: a model-agnostic approach with SHAP analysis," AI Civ. Eng., vol. 4, 2025, doi: 10.1007/s43503-025-00064-8.
[54] Y. T. Altuncı, "Predicting the Compressive Strength of Concrete Incorporating Olivine Aggregate at Varied Cement Dosages Using Artificial Intelligence," Processes, 2025, doi: 10.3390/pr13072130.
[55] R. Shay et al., "Can long passwords be secure and usable?," Proc. SIGCHI Conf. Hum. Factors Comput. Syst., 2014, doi: 10.1145/2556288.2557377.
[56] M. Jubur, P. Shrestha, and N. Saxena, "An In-Depth Analysis of Password Managers and Two-Factor Authentication Tools," ACM Comput. Surv., vol. 57, pp. 1–32, 2025, doi: 10.1145/3711117.
[57] C. Shi, Z. Li, and X. Li, "System Password Security: Attack and Defense Mechanisms," ArXiv, vol. abs/2510.1, 2025, doi: 10.48550/arxiv.2510.10246.
[58] Y. Wu and Y. Zhou, "Hybrid machine learning model and Shapley additive explanations for compressive strength of sustainable concrete," Constr. Build. Mater., 2022, doi: 10.1016/j.conbuildmat.2022.127298.
[59] M. Wang, J. Kang, W. Liu, J. Su, and M. Li, "Research on prediction of compressive strength of fly ash and slag mixed concrete based on machine learning," PLoS One, vol. 17, 2022, doi: 10.1371/journal.pone.0279293.
[60] W. Zhang, D. Liu, and K. Cao, "Prediction of concrete compressive strength using support vector machine regression and non-destructive testing," Case Stud. Constr. Mater., 2024, doi: 10.1016/j.cscm.2024.e03416.
[61] S. Pal, L. H. Trang, V. Hieu, D. Nguyen, D. Q. Vu, and I. Prakash, "Investigation of Support Vector Machines with Different Kernel Functions for Prediction of Compressive Strength of Concrete," J. Sci. Transp. Technol., 2024, doi: 10.58845/jstt.utt.2024.en.4.2.55-68.
[62] M. Alyami et al., "Predictive Modeling for Compressive Strength of 3D Printed Fiber-Reinforced Concrete Using Machine Learning Algorithms," Case Stud. Constr. Mater., 2023, doi: 10.1016/j.cscm.2023.e02728.
[63] D. Feng et al., "Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach," Constr. Build. Mater., 2020, doi: 10.1016/j.conbuildmat.2019.117000.
[64] Y.-M. Hong, "Performance Comparison of Machine Learning Models for Concrete Compressive Strength Prediction," Materials (Basel)., vol. 17, 2024, doi: 10.3390/ma17092075.
[65] A. Kumar, S. Sen, and S. Sinha, "Support vector machine-based prediction model for the compressive strength for concrete reinforced with waste plastic and fly ash," Asian J. Civ. Eng., vol. 26, pp. 1429–1447, 2025, doi: 10.1007/s42107-024-01256-w.
[66] T. Cihan, "Prediction of Concrete Compressive Strength and Slump by Machine Learning Methods," Adv. Civ. Eng., 2019, doi: 10.1155/2019/3069046.
[67] A. Abd and S. Abd, "Modelling the strength of lightweight foamed concrete using support vector machine (SVM)," Case Stud. Constr. Mater., vol. 6, pp. 8–15, 2017, doi: 10.1016/j.cscm.2016.11.002.
[68] A. Beskopylny et al., "Concrete Strength Prediction Using Machine Learning Methods CatBoost, k-Nearest Neighbors, Support Vector Regression," Appl. Sci., 2022, doi: 10.3390/app122110864.
[69] M. M. Woldeamanuel and H.-K. Kim, "Parametric study on global estimation models for

compressive strength adopting various machine learning algorithms in concrete," Eng. Appl. Artif. Intell., vol. 136, p. 108888, 2024, doi: 10.1016/j.engappai.2024.108888.

[70] V. Vakharia and R. Gujar, "Prediction of compressive strength and portland cement composition using cross-validation and feature ranking techniques," Constr. Build. Mater., 2019, doi: 10.1016/j.conbuildmat.2019.07.224.

[71] Q.-H. Le et al., "Machine learning based models for predicting compressive strength of geopolymer concrete," Front. Struct. Civ. Eng., vol. 18, pp. 1028–1049, 2024, doi: 10.1007/s11709-024-1039-5.

[72] S. Nawaz et al., "Flexural Performance Evaluation Of Glass Fiber Reinforced Polymer Bars ({GFRP}) In Doubly Reinforced Beams: An Experimental Approach," 2025.

[73] W. Khan et al., "Transforming Expansive Clays {FROM} Waste {TO} Value: Dual Ash Stabilization Strategy {FOR} Augmenting Black Cotton Subgrade Soil Performance," vol. 11, pp. 5065–5073, 2025, doi: 10.64252/cz07k049.