Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44) 3** : 1–16.
ISSN-0037-8712
doi:10.5269/bspm.82004

# Creating a Mathematical Model to Predict Concrete's Compressive Strength and Applying Data-Driven Inverse Techniques

Ibrahim Namiq Faiaq, Shilan Othman Hussein and Taysir Emhemed Dyhoum

ABSTRACT: This study introduces a mathematical model for predicting concrete compressive strength and employs data-driven inverse techniques. Six models are tested: Linear Regression, Pure Quadratic, Interaction, Full Quadratic, Artificial Neural Network (ANN), and Adaptive Neuro-Fuzzy Inference System (ANFIS). To evaluate inverse sensitivity, noise is introduced into the ANN models to assess how small errors in the data affect the results for concrete compressive strength. Tikhonov regularization is applied to mitigate these errors, ensuring reliable outcomes. The dependent variable, compressive strength (CS), is categorized into low-strength, normal-strength, and high-strength concrete, with values ranging from 7.98 to 92.93 MPa. Independent variables include coarse aggregate (CA), sand (S), cement (C), fly ash (FA), cement replacement (CR), water-to-binder ratio (w/b), calcium oxide ratio (CaO), silicon dioxide ($SiO_2$), and curing time (t). Model performance is evaluated using multiple metrics, including the correlation coefficient ($R^2$), objective function (OBJ), scatter index (SI), mean absolute error (MAE), and root mean squared error (RMSE). The results indicate that ANFIS outperforms the other models in both accuracy and efficiency, establishing it as the most reliable approach for predicting concrete compressive strength.

Keywords: Mathematical modeling, concrete compressive strength, statistical metrics, inverse techniques, regularization method.

## Contents



## 1. Introduction

Compressive strength is a crucial property of concrete and is vital for structural design and safety. Since concrete is a heterogeneous mixture of cement, aggregates, water, and admixtures, its strength can vary significantly. The physical, chemical, and mechanical properties of these components, along with their ratios, greatly influence concrete's ability to withstand compressive forces [19,2]. Compressive strength is usually measured in laboratory tests using crushed concrete cubes or standard-sized cylinders [17]. Although this method is accepted worldwide, it can be expensive and time-consuming due to the use of costly materials and lengthy procedures. This research explores various modeling techniques, including linear regression, pure quadratic, interaction, and full quadratic models, as well as artificial neural networks (ANN) and adaptive neuro-fuzzy inference systems (ANFIS), to forecast the compressive strength. These models have also seen extensive use in the literature [13,16,6,7,5,23,22] across different datasets. For comparison, another study [14] trained and tested an ANN model on 236 datasets and achieved the highest accuracy and reliability among the evaluated models. A sensitivity analysis indicated that curing time is the most critical factor in predicting concrete's compressive strength. Furthermore, a separate study [18] employed 450 datasets for model development.

This study gathered 477 data points on compressive strength from existing research. The analyzed concrete mixes included fly ash (FA) at 71-316 kg/m$^3$, calcium oxide (CaO) at 0.31-32%, and silica ($SiO_2$) at 30.5-62.54%. Cement (C) content ranged from 67 to 356 kg/m$^3$, with cement replacement (CR) varying

2020 *Mathematics Subject Classification:* 74G15, 65D10.
Submitted February 15, 2026. Published April 13, 2026

between 18% and 100%. Coarse aggregate (CA) was present in amounts of 801 to 1246 kg/m$^3$, while fine aggregate (S) ranged from 522 to 905 kg/m$^3$. The water-to-binder ratio (w/b) spanned 0.28 to 0.60, and curing periods ranged from 3 to 365 days. The resulting compressive strengths (CS) varied from 7.98 to 92.93 MPa. To evaluate model reliability, compressive strength data are categorized into 5-34 MPa, 35-64 MPa, and 65-95 MPa groups, and the best-performing model is selected for each category. The research examines various models that predict compressive strength through different approaches and explores the effect of changing calcium oxide (CaO) and silicon dioxide ($SiO_2$) ratios on the strength of fly ash-enhanced cement concrete. Additionally, model reliability and accuracy are assessed using several statistical metrics, including the correlation coefficient ($R^2$), objective function (OBJ), scatter index (SI), mean absolute error (MAE), and root mean squared error (RMSE).

Section 4 uses inverse problem methods to analyze how small errors impact the dataset. These errors can be minimized using the Tikhonov regularization method; more details can be found in [20,9,8,12,11, 10]. The paper is structured as follows: Section 2 introduces the mathematical formulation, Section 3 discusses the results, and Section 5 presents the conclusions.

## 2. Mathematical Formulation

This study aims to develop models for estimating the compressive strength of concrete. In mathematical modeling, it is essential to define input variables $\mathbf{x}$ and an output variable $\mathbf{y}$. This research considers nine input variables $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, ..., \mathbf{x}_9\}$, each with 477 observations collected from existing literature. In engineering contexts, these nine variables represent coarse aggregate (CA), sand (S), cement content (C), fly ash content (FA), cement replacement (CR), water-to-binder ratio (w/b), calcium oxide (CaO), silicon dioxide ($SiO_2$), and curing time (t). These variables are treated as independent variables. The output variable $\mathbf{y}$ corresponds to the compressive strength of concrete.

The models [14,16,6,7,5,23,22,1,21,24,3,15,4] and their formulations used in this study are listed below, along with a flowchart explaining their operation. The dependent variable is $\mathbf{y} = [y_1, y_2, ..., y_n]^T$, and the independent variables are $\mathbf{x}_i = [x_{1i}, x_{2i}, ..., x_{ni}]^T$, $i = \overline{1,9}$ and $n = 477$:

**Linear Regression Model (LR):**

$$\mathbf{y} = \beta_0 + \sum_{i=1}^{9} \beta_i \mathbf{x}_i. \tag{2.1}$$

such that the intercept is represented by $\beta_0$, while the coefficients for the variables $\mathbf{x_1}$ to $\mathbf{x_9}$ are represented by $\beta_1$ to $\beta_9$, respectively.

**Pure Quadratic Model (PQ):**

$$\mathbf{y} = \beta_0 + \sum_{i=1}^{9} \beta_i \mathbf{x}_i + \sum_{i=1}^{9} \beta_{i+9} \mathbf{x}_i^2, \tag{2.2}$$

where $\beta_0$ is the constant term, and $\sum_{i=1}^{9} \beta_i x_i$, $\sum_{i=1}^{9} \beta_{i+9} x_i^2$ represent the linear and quadratic terms, respectively.

**Interaction Regression (IN):**

$$\mathbf{y} = \beta_0 + \sum_{i=1}^{9} \beta_i \mathbf{x}_i + \sum_{1 \le i < j \le 9} \beta_{ij} \mathbf{x}_i \mathbf{x}_j. \tag{2.3}$$

$\beta_0$ stands for the intercept term, $\sum_{i=1}^{9} \beta_i \mathbf{x}_i$ for the linear terms, and $\sum_{1 \le i < j \le 9} \beta_{ij} \mathbf{x}_i \mathbf{x}_j$ for the interaction (product) terms.

**Full Quadratic Model (FQ):**

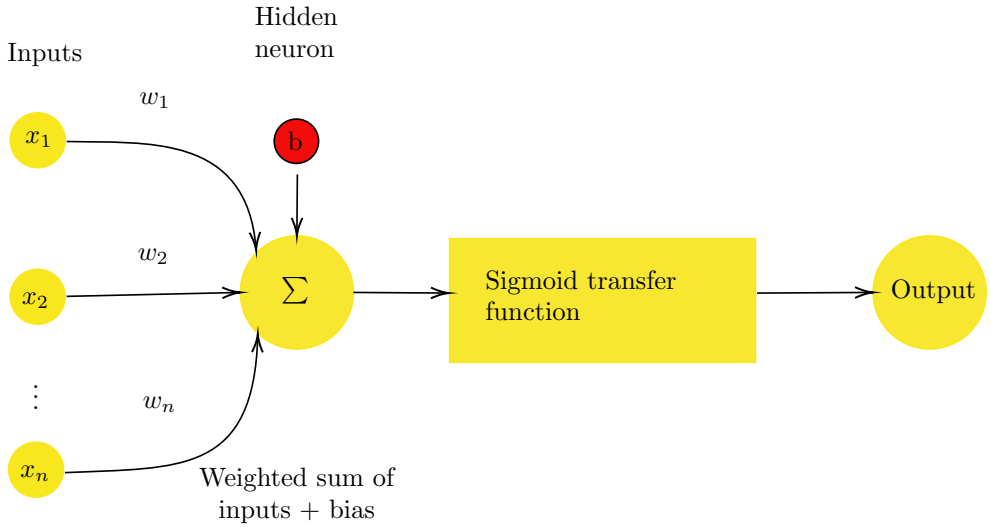
$$\mathbf{y} = \beta_0 + \sum_{i=1}^{9} \beta_i \mathbf{x}_i + \sum_{1 \le i < j \le 9} \beta_{ij} \mathbf{x}_i \mathbf{x}_j + \sum_{i=1}^{9} \beta_{ii} \mathbf{x}_i^2, \tag{2.4}$$

where $\beta_0$ indicates the constant term; $\sum_{i=1}^{9} \beta_i \mathbf{x}_i$ stands for the linear terms; $\sum_{1 \le i < j \le 9} \beta_{ij} \mathbf{x}_i \mathbf{x}_j$ shows the interaction terms; and $\sum_{i=1}^{9} \beta_{ii} \mathbf{x}_i^2$ identifies the quadratic terms. The parameters of the model are symbolized by the coefficients $\beta_i$, $\beta_{ij}$, and $\beta_{ii}$.

**Artificial Neural Network (ANN):**

$$\mathbf{y} = \sum_{j=1}^{m} w_j f \left( \sum_{i=1}^{9} v_{ji} \mathbf{x}_i + b_j \right) + b_0, \tag{2.5}$$

where $v_{ji}$ and $w_j$ are connection weights, $b_j$ and $b_0$ are biases, and $f(\cdot)$ is an activation function. The flowchart clearly illustrates inputting an ANN in MATLAB, helping you understand and implement it.




This model scales data to the [0, 1] range using the sigmoid function $\delta(x) = \frac{1}{1+e^{-x}}$, which transforms real-valued inputs into this interval.

**Adaptive-Network-Based Fuzzy Inference System (ANFIS):**

$$\mathbf{y} = \sum_{k=1}^{N} \bar{w}_k \left( \beta_{k0} + \sum_{i=1}^{9} \beta_{ki} \mathbf{x}_i \right). \tag{2.6}$$

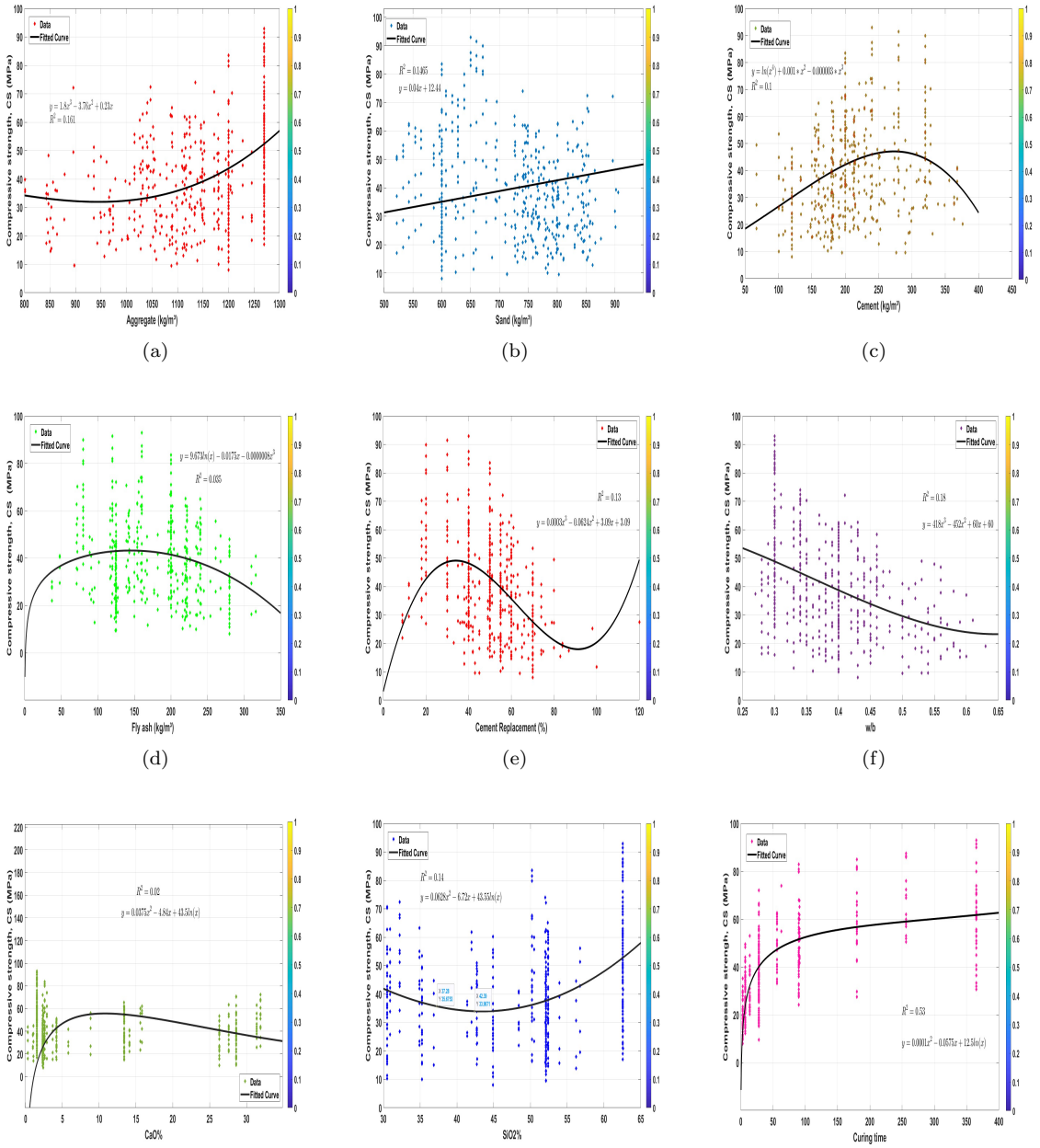
where: $N$ is the total number of fuzzy rules, $\beta_{k0}$ and $\beta_{ki}$ are the consequent parameters of the $k$-th rule, $\bar{w}_k$ is the normalized firing strength of the $k$-th rule, and Normalized Firing Strength The normalized firing strength is defined as:

$$\bar{w}_k = \frac{w_k}{\sum_{m=1}^{N} w_m}, \quad w_k = \prod_{i=1}^{9} \mu_{A_i^{(k)}} (\mathbf{x}_i),$$

where $\mu_{A_i^{(k)}}(\cdot)$ denotes the membership function associated with the $i$-th input in the $k$-th fuzzy rule.

## 3. Result Discussion

The dataset contains 477 samples divided into three subsets. The first subset, consisting of 333 samples (70% of the total), is used for training. The remaining 144 samples, which make up 30%, are split evenly into validation and test sets, each containing 72 samples. All models (2.1)-(2.6) examined in this study can be described as a linear algebraic system in the form of $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$. The unknown parameter vector $\boldsymbol{\beta}$ is estimated using the least-squares method, given by $\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}\mathbf{y}$, based on the provided values of $\mathbf{y}$. After calculating $\boldsymbol{\beta}$, the approximate $\mathbf{y}$ is obtained by substituting $\boldsymbol{\beta}$ into the equation $\mathbf{y}_{\text{approximate}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}\mathbf{y}$. Each model produces matrices of different sizes, resulting in different parameter vectors $\boldsymbol{\beta}$ for the training, validation, and test datasets. With 477 observations, MATLAB code is used to compute the results and generate the corresponding figures.




(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 1: Relation between the independent {CA,S,C,FA,CR,w/b,CaO,$SiO_2$,t} and dependent variables (Compressive strength of concrete)

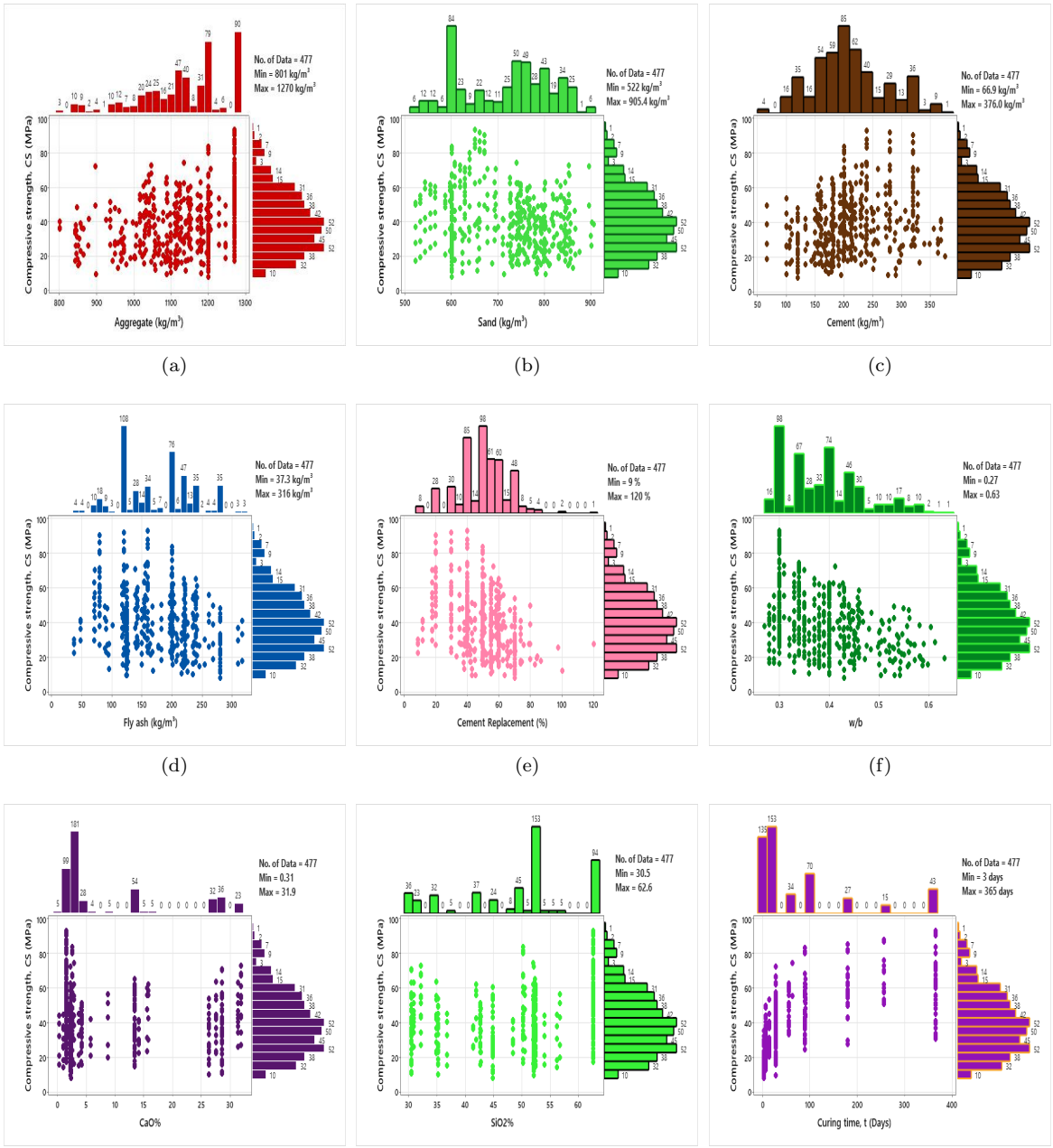



Figure 2: Marginal plots showing the distribution and effects of input variables on concrete's compressive strength.

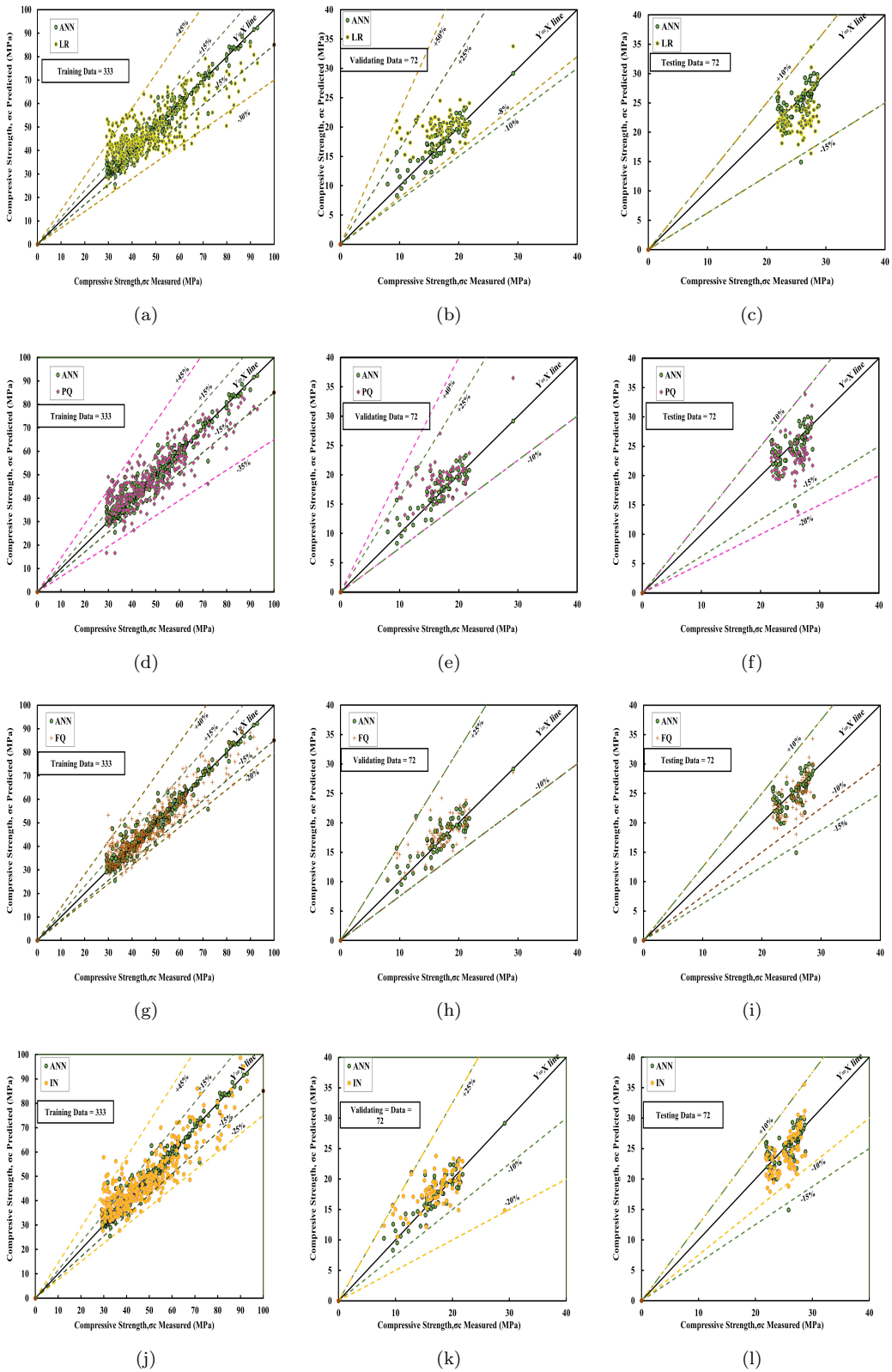



Figure 3: Comparison of the ANN model with (a)–(c) LR; (d)–(f) PQ; (g)–(i) FQ; and (j)–(l) IN across training, validation, and testing datasets.

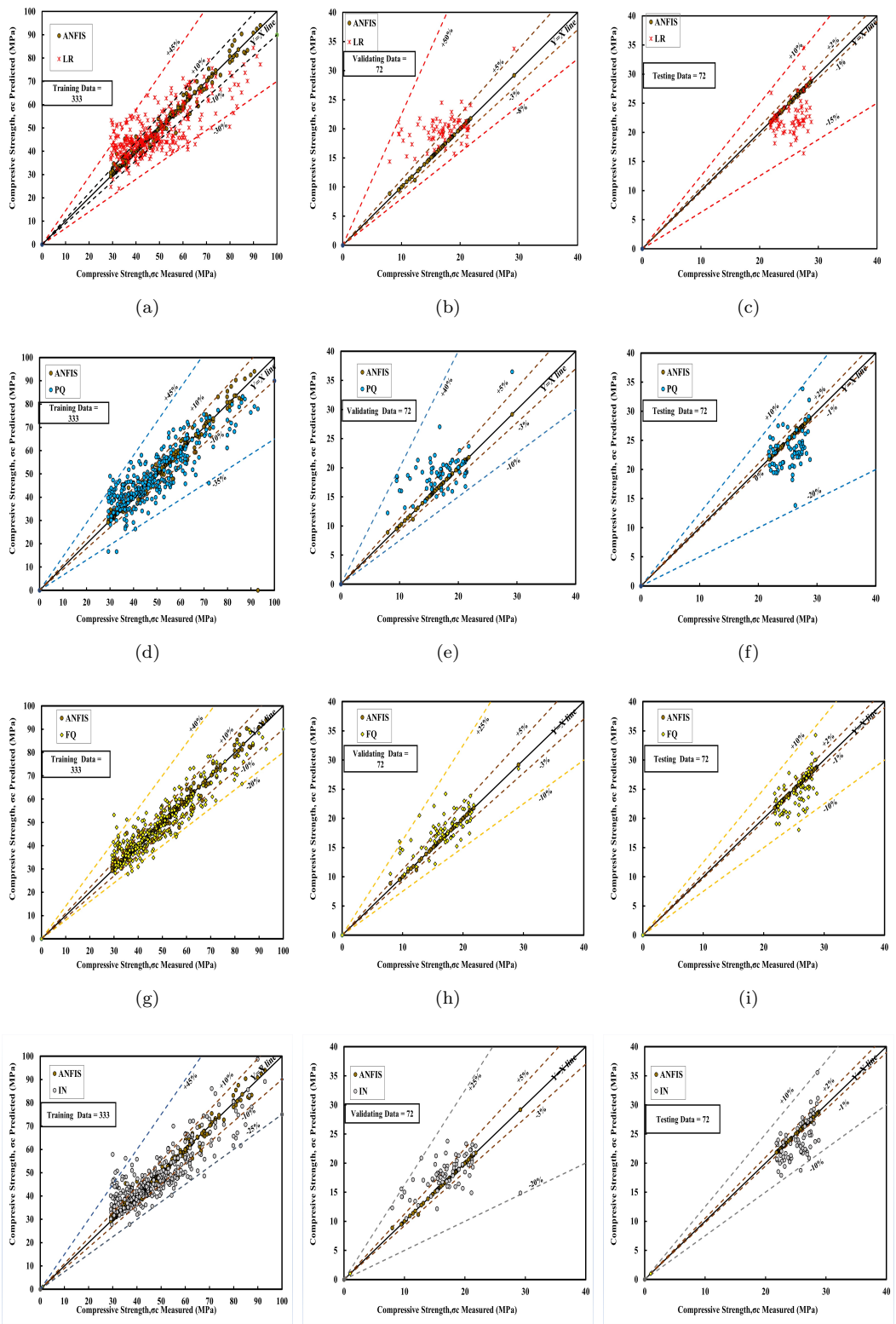



Figure 4: Comparison of the ANFIS model with (a)–(c) LR; (d)–(f) PQ; (g)–(i) FQ; and (j)–(l) IN across training, validation, and testing datasets.

Figures 1 and 2 illustrate how each independent variable individually impacts the concrete's compressive strength. Figure 1 features trend curves, whereas Figure 2 displays marginal plots generated through various plotting techniques. Both figures show that the relationships between variables are unclear and inconsistent. The data points are scattered, and most variables do not exhibit a clear trend when looked at separately. Despite being created with different software, both figures arrive at the same conclusion: each variable's effect on compressive strength is minor and unstable. This indicates that compressive strength depends on the combined influence of multiple variables rather than any single one. The next figures, Figures 3 and 4, show that integrating all nine independent variables yields more accurate results. These variables are included in the models analyzed in this study, indicating that the combined models perform better than those based on each variable alone.

Furthermore, Figure 3 shows that the ANN provides more accurate, stable, and reliable predictions than traditional regression models across all datasets. Subfigures (a)–(c) compare ANN with LR, (d)–(f) with PQ, (g)–(i) with FQ, and (j)–(l) with IN models. Meanwhile, the conventional models (LR, PQ, FQ, and IN) display greater dispersion and larger errors, especially during validation and testing. Additionally, Figure 4 presents a comparison between the predicted and measured concrete compressive strength from the ANFIS model and several benchmark models for three datasets: training, validation, and testing. Subfigures (a)–(c) compare ANFIS with LR, (d)–(f) with PQ, (g)–(i) with FQ, and (j)–(l) with IN models. In all cases, the ANFIS predictions closely follow the 45° reference line and remain within the ±10% and ±20% error bounds, indicating that ANFIS achieves higher accuracy and better generalization than the other models.

To accurately evaluate the proposed models, their performance is measured using three key metrics: Objective function (OBJ), Mean absolute error (MAE), and Scatter index (SI) [12, 26, 33]. These metrics are used together because relying on only one can be misleading; a model might perform poorly on one but well on others. OBJ, MAE, and SI are the most reliable for overall assessment. Figure 5 compares five models—FQ, Interaction, PQ, LR, ANN, and ANFIS—across various metrics. In Figures 5(a-c), the ANN model outperforms the regression-based models by having the lowest Objective (OBJ), Mean Absolute Error (MAE), and Stability Index (SI), indicating it offers high accuracy and consistency. Conversely, the FQ and LR models show higher errors and greater variability. The Interaction and PQ models perform at a moderate level. Furthermore, Figures 5(d-f) indicate that ANFIS surpasses the other models, achieving the lowest OBJ of 0.98 MPa, MAE of 0.23 MPa, and SI of 0.004, which reflects its superior accuracy and stability. The FQ and LR models show higher error metrics, with LR especially displaying more variability. Table 1 and Figure 6 illustrates the residuals of different models on the training, testing, and validation datasets. The ANFIS-based models produce smaller, more consistent residuals nearer to zero compared to other models. This indicates better accuracy and stability, confirming that ANFIS outperforms the other models. Overall, ANFIS delivers the best performance throughout the training, testing, and validation phases.

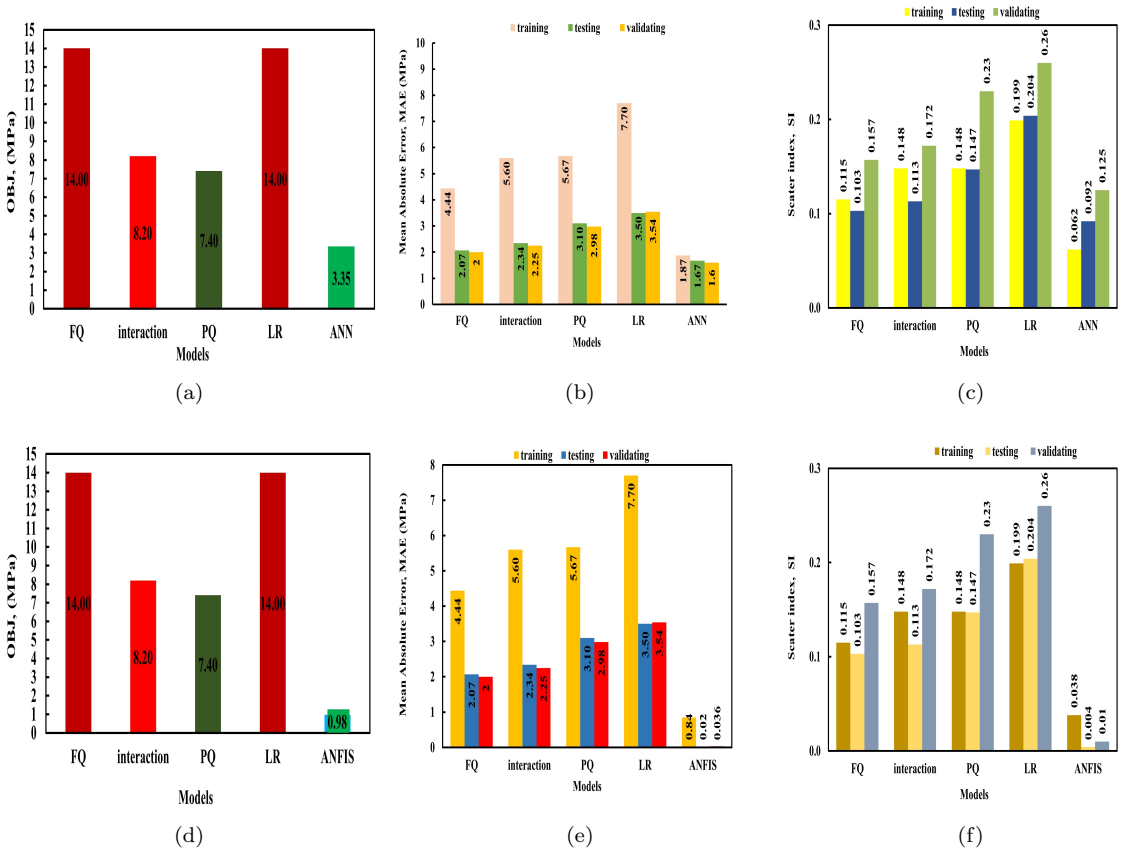



Figure 5: Measured models using three key metrics: (a) and (d) OBJ; (b) and (e) MAE; (c) and (f) SI, across training, validation, and testing datasets.

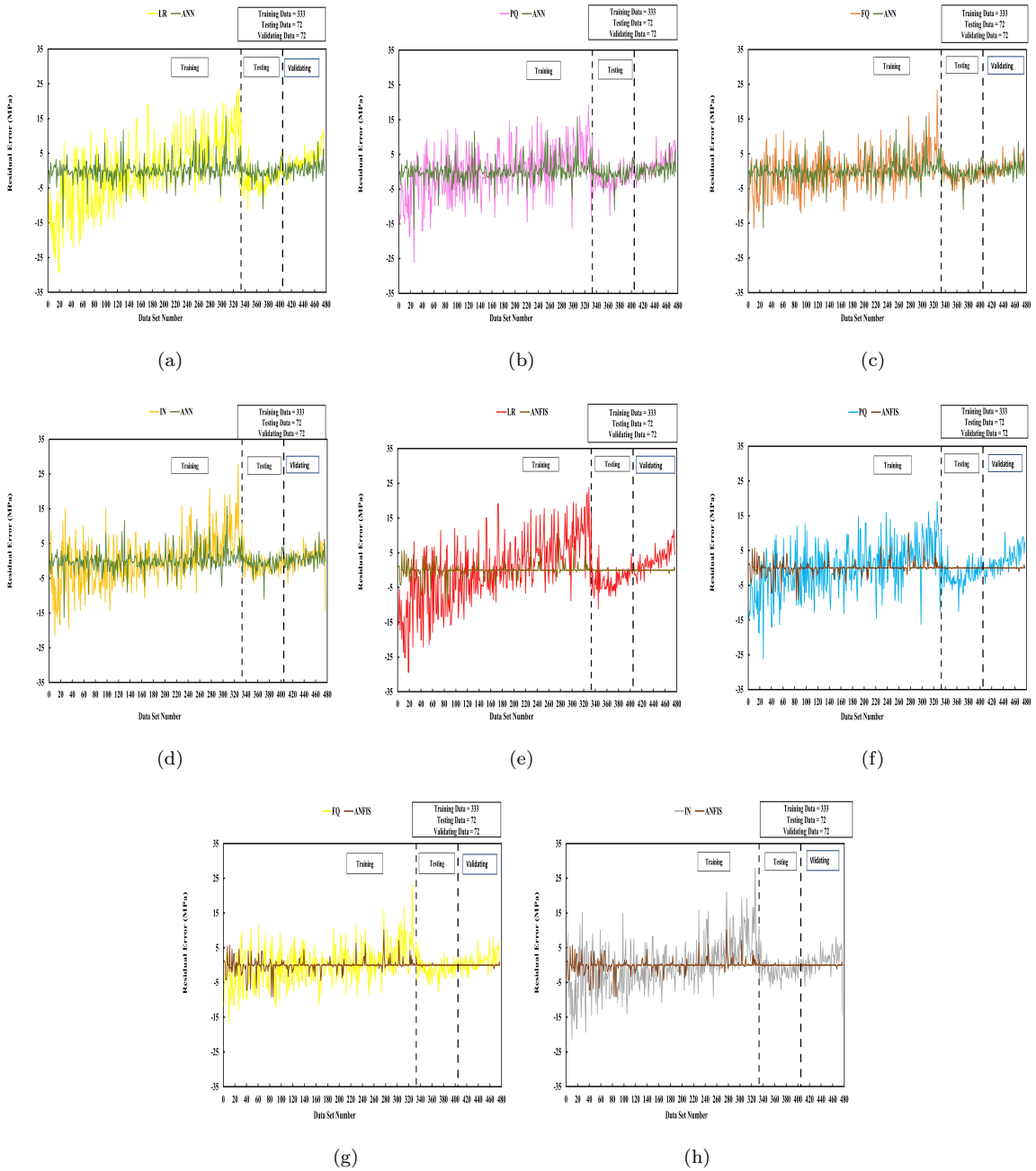
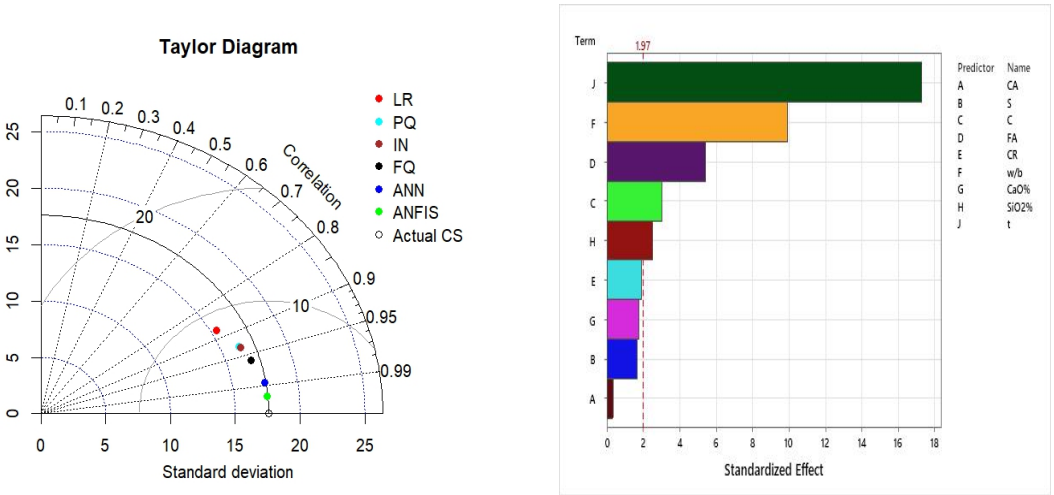



Figure 6: Comparison of residuals across different models, emphasizing the better performance of the ANFIS model.

Figure 7(a) shows the Taylor diagram evaluates the performance of different models—LR, PQ, IN, FQ, ANN, and ANFIS—against the actual reference data (Actual CS). It shows that ANFIS (shown in green) has the highest correlation and the closest standard deviation to the reference, making it the most effective among the tested models. Conversely, models like LR (red) and PQ (cyan) exhibit lower correlations and larger standard deviations, indicating weaker performance. As a result, ANFIS is identified as the most accurate and reliable model for this data set. Figure 7(b) shows the effects of input variables on output. (t) has the strongest influence, then (w/b) and (FA). The other variables have smaller effects, indicating less impact.



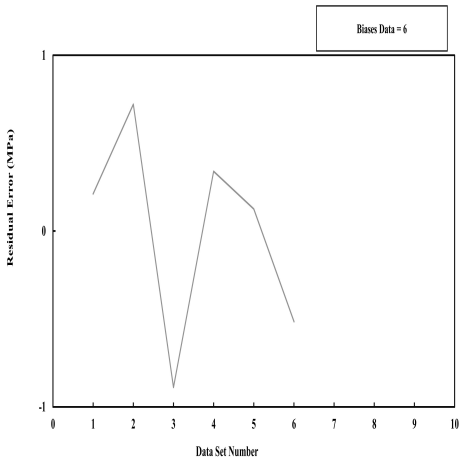

(a) (b)

Figure 7: (a) Comparison between developed models based on standard deviation and the correlation coefficient between measured and predicted CS using all datasat. (b) Sensitivity analysis examines how changes in an independent variable influence a dependent variable.

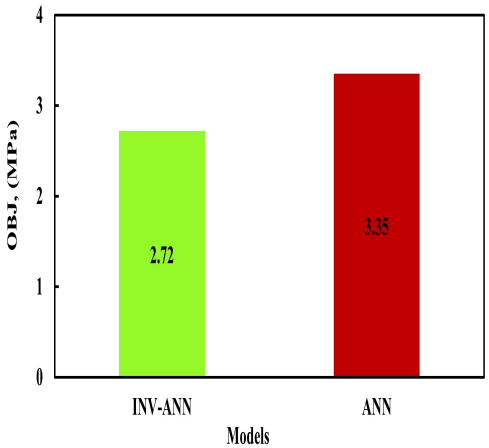
## 4. Inverse Data Analysis

This section examines the impact of small errors on accuracy. Such errors may arise from laboratory processes, during recording, from approximate data, or within the models themselves, which often normalize data to the [0, 1] range rather than using raw input. For this reason, multiple models are tested to reduce errors. The study highlights mathematical techniques for handling data uncertainty, mainly Tikhonov regularization, which helps stabilize solutions and lessen the influence of data errors.

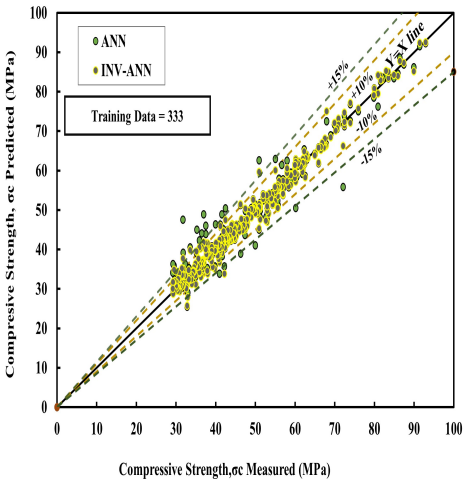
The figures and Table 1 clearly show that ANFIS and ANN perform significantly better than the other models (LR, IN, PQ, and FQ). While ANFIS surpasses ANN, it has difficulty handling noise effectively due to its high computational complexity. ANFIS involves roughly 512 parameters per input, making it computationally demanding. As a result, noise is introduced only into the ANN model in this analysis. To introduce noise into Equation (2.5), a Gaussian distribution is added to the bias term [20,9,8,12,11,10]. As shown in Figure 8(a), the residual error nears zero, indicating that noise impacts the data and causes overestimation and underestimation. A negative residual near dataset 3 indicates higher noise sensitivity. Overall, noise causes residual instability. To address this issue, Tikhonov regularization $\mathbf{y}_{\text{approximate}} = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda I)^{-1}\mathbf{X}\mathbf{y}$ is used to stabilize the results, as illustrated in Figures 8(b) and 8(c), where this research uses an optimal regularization parameter, $\lambda = 10^{-2}$. Exploring different $\lambda$ values helps identify the best option. While methods such as the L-curve criterion or $L_2$-norm minimization [15] can be used to select $\lambda$, this approach empirically optimizes $\mathbf{y}_{\text{approximate}}$. Figures 8(b) and 8(c) show that regularization improves accuracy and reduces noise.




(a)




(b)




(c)

Figure 8: (a) Adding noise to biased data (b) Comparison between ANN and inverse of ANN based on objective function (c) Line error between ANN and inverse ANN.

Table 1: Compares models across three datasets: training, testing, and validation, with $R^2$, RMSE, MAE, SI, and OBJ

| Training "333 data" | | | | | | | |
|---|---|---|---|---|---|---|---|
| Measure Metrics | INV-ANN | ANFIS | ANN | FQ | IN | PQ | LR |
| $R^2$ | 0.98 | 0.98 | 0.95 | 0.84 | 0.739 | 0.739 | 0.532 |
| RMSE | 2.19 | 1.87 | 3 | 5.624 | 7.225 | 7.221 | 9.699 |
| OBJ | 2.72 | 0.98 | 3.35 | 14.325 | 8.179 | 7.366 | 14.035 |
| SI | 0.04 | 0.04 | 0.06 | 0.115 | 0.148 | 0.148 | 0.198 |
| MAE | 1.56 | 0.84 | 1.87 | 4.435 | 5.6 | 5.674 | 7.697 |
| Testing "144 data" | | | | | | | |
| $R^2$ | 0.49 | 0.99 | 0.35 | 0.414 | 0.425 | 0.131 | 0.089 |
| RMSE | 2.2 | 0.09 | 2.33 | 2.624 | 2.878 | 3.737 | 4.23 |
| OBJ | 1.1 | 0.01 | 1.25 | 10.709 | 2.417 | 0.537 | 0.384 |
| SI | 0.06 | 0.01 | 0.09 | 0.103 | 0.113 | 0.147 | 0.166 |
| MAE | 1.3 | 0.02 | 1.67 | 2.077 | 2.336 | 3.101 | 3.459 |
| Validating "144 data" | | | | | | | |
| $R^2$ | 0.81 | 0.99 | 0.71 | 0.586 | 0.316 | 0.341 | 0.279 |
| RMSE | 1.91 | 0.17 | 2.11 | 2.65 | 3.35 | 3.871 | 4.385 |
| OBJ | 0.31 | 0.02 | 0.35 | 0.517 | 0.897 | 1.639 | 4.911 |
| SI | 0.1 | 0.01 | 0.13 | 0.157 | 0.199 | 0.23 | 0.26 |
| MAE | 1.3 | 0.04 | 1.6 | 2.006 | 2.447 | 2.984 | 3.537 |

## 5. Conclusion

This research developed a mathematical model to predict concrete compressive strength and applied data-driven inverse methods. The inverse analysis showed that even small measurement errors can significantly affect the results; therefore, Tikhonov regularization was applied to improve stability and robustness. Six predictive models were tested: Linear Regression, Pure Quadratic, Interaction, Full Quadratic, Artificial Neural Network, and Adaptive Neuro-Fuzzy Inference System. These models incorporated nine independent variables: coarse aggregate, sand, cement, fly ash, cement replacement, water-to-binder ratio, calcium oxide ratio, silicon dioxide, and curing time. Notably, curing time was identified as the most influential variable. Model performance was evaluated using several statistical metrics, particularly the coefficient of determination ($R^2$). The results demonstrate that the Adaptive Neuro-Fuzzy Inference System consistently outperforms the other models in terms of accuracy and reliability, confirming its effectiveness in predicting concrete compressive strength.

## References

1. Hamza Aamir, Kinza Aamir, and Muhammad Faisal Javed. Linear and non-linear regression analysis on the prediction of compressive strength of sodium hydroxide pre-treated crumb rubber concrete. *Engineering Proceedings*, 44(1):5, 2023.

2. Ali S Al-Harthy, Ramzi Taha, and Faisal Al-Maamary. Effect of cement kiln dust (ckd) on mortar and concrete mixtures. *Construction and Building Materials*, 17(5):353–360, 2003.

3. Turki S Alahmari, Irfan Ullah, and Furqan Farooq. Machine learning models for estimating the compressive strength of rubberized concrete subjected to elevated temperature: Optimization and hyper-tuning. *Sustainable Chemistry and Pharmacy*, 42:101763, 2024.

4. Hussein YH Alnajjar and Osman Üçüncü. Enhance and improve modelling prediction by using an adaptive neuro-fuzzy inference system-based model to predict pollution removal efficacy in wastewater treatment plants. *Desalination and Water Treatment*, 286:52–63, 2023.

5. Carl De Boor and Carl De Boor. *A practical guide to splines*, volume 27. springer-verlag New York, 1978.

6. Jerome H Friedman. Multivariate adaptive regression splines. *The annals of statistics*, 19(1):1–67, 1991.

7. Jerome H Friedman and Bernard W Silverman. Flexible parsimonious smoothing and additive modeling. *Technometrics*, 31(1):3–21, 1989.

8. MS Hussein, Taysir E Dyhoum, SO Hussein, and Mohammed Qassim. Identification of time-wise thermal diffusivity, advection velocity on the free-boundary inverse coefficient problem. *Mathematics*, 12(17), 2024.
9. Shilan Othman Hussein. *Inverse force problems for the wave equation*. PhD thesis, University of Leeds, 2016.
10. Shilan Othman Hussein. Placement inverse source problem under partition hyperbolic equation. *Iraqi Journal of Science*, pages 1994–1999, 2021.
11. SO Hussein. Inverse one-dimensional wave equation problem under upper-base as additional information. *Italian Journal of Pure and Applied Mathematics*, 47:596–608, 2022.
12. SO Hussein and Taysir E Dyhoum. Solutions for non-homogeneous wave equations subject to unusual and neumann boundary conditions. *Applied Mathematics and Computation*, 430:127285, 2022.
13. Dilshad Kakasor Ismael Jaf, Payam Ismael Abdulrahman, Ahmed Salih Mohammed, Rawaz Kurda, Shaker MA Qaidi, and Panagiotis G Asteris. Machine learning techniques and multi-scale models to evaluate the impact of silicon dioxide (sio2) and calcium oxide (cao) in fly ash on the compressive strength of green concrete. *Construction and Building Materials*, 400:132604, 2023.
14. Dilshad Kakasor Ismael Jaf, Payam Ismael Abdulrahman, Ahmed Salih Mohammed, Rawaz Kurda, Shaker M.A. Qaidi, and Panagiotis G. Asteris. Machine learning techniques and multi-scale models to evaluate the impact of silicon dioxide (sio2) and calcium oxide (cao) in fly ash on the compressive strength of green concrete. *Construction and Building Materials*, 400:132604, 2023.
15. Raphael Langbauer, Georg Nunner, Thomas Zmek, Jürgen Klarner, René Prieler, and Christoph Hochenauer. Development of an artificial neural network (ann) model to predict the temperature of hot-rolled steel pipes. *Advances in Industrial and Manufacturing Engineering*, 5:100090, 2022.
16. Ahmad Khalil Mohammed, AMT Hassan, and Ahmed Salih Mohammed. Predicting the compressive strength of green concrete at various temperature ranges using different soft computing techniques. *Sustainability*, 15(15):11907, 2023.
17. Ahmed Mohammed, Serwan Rafiq, Parveen Sihag, Rawaz Kurda, and Wael Mahmood. Soft computing techniques: systematic multiscale models to predict the compressive strength of hvfa concrete based on mix proportions and curing times. *Journal of Building Engineering*, 33:101851, 2021.
18. Ahmed Mohammed, Serwan Rafiq, Parveen Sihag, Rawaz Kurda, Wael Mahmood, Kawan Ghafor, and Warzer Sarwar. Ann, m5p-tree and nonlinear regression approaches with statistical evaluations to predict the compressive strength of cement-based mortar modified with fly ash. *Journal of Materials Research and Technology*, 9(6):12416–12427, 2020.
19. Nzar Shakr Piro, Ahmed Salih Mohammed, and Samir M Hamad. Retracted: The impact of ggbs and ferrous on the flow of electrical current and compressive strength of concrete, 2022.
20. Shnyar Ali Rashid and Shilan O Hussein. The technique of analyzing a non-homogeneous partial differential equation by splitting the problem into the inverse and direct problems. *Journal of University of Babylon for Pure and Applied Sciences*, 33(4):339–362, 2025.
21. Nasser A Saeed, Hend A Saleh, Wedad A El-Ganaini, Jan Awrejcewicz, and Haitham A Mahmoud. An unusual chaotic system with pure quadratic nonlinearities: Analysis, control, and synchronization. *Chinese Journal of Physics*, 88:311–331, 2024.
22. L Schumaker. Fitting surfaces to scattered data. approximation theory ii, c. chui, l. schumaker and g. lorenz, 1976.
23. Larry L Schumaker. On spaces of piecewise polynomials in two variables. In *Approximation theory and spline functions*, pages 151–197. Springer, 1984.
24. Thomas Welchowski and Dominic Edelmann. Interaction difference hypothesis test for prediction models. *Machine Learning and Knowledge Extraction*, 6(2):1298–1322, 2024.


## Appendix

$$R^2 = \left( \frac{\sum_i (xi - \bar{x}) \times (yi - \bar{y})}{\sqrt{\sum_i (xi - \bar{x})^2} \times \sqrt{\sum_i (yi - \bar{y})^2}} \right)^2; \qquad MAE = \frac{\sum_{i=1}^{n} (|yi - xi|)}{n};$$

$$OBJ = \left( \frac{n_{tr}}{n_{\text{all}}} \times \frac{RMSE_{tr} + MAE_{tr}}{R_{tr}^2 + 1} \right) + \left( \frac{n_{tst}}{n_{\text{all}}} \times \frac{RMSE_{tst} + MAE_{tst}}{R_{tst}^2 + 1} \right);$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (yi - xi)^2}{N}}; \qquad a20 - index = \frac{N20}{N}; \qquad SI = \frac{RMSE}{yi},$$

where; $xi$= predicted value, $yi$ = tested value, $\bar{x}$ = average of predicted value, $\bar{y}$ = average of tested data, $n_{\text{all}}$ = total number of training and testing datasets, $n_{tr}$ = number of the training dataset, $n_{tst}$ = a a number of the testing dataset, $N$ = total data, and $N20$ is the total number of predicted to the measured data of compressive strength ratio ranging from 0.8 to 1.2. $R^2$ and a-20 index values typically range from 0 to 1, with 1 indicating optimal performance. The RMSE, MAE, and OBJ values vary from 0 to

$\infty$. The ideal value is zero. Furthermore, if the SI value is smaller than 0.1, the model is deemed to be high-performing. In contrast, the SI value ranges from 0.1 to 0.2.

*Ibrahim N. Faiaq and Shilan O. Hussein,*
*Department of Mathematics,*
*College of Science, University of Sulaimani,*
*Sulaymaniyah, Kurdistan Region, Iraq.*
*and*
*Taysir E. Dyhoum,*
*Department of Computing and Mathematics, Faculty of Science and Engineering,*
*Manchester Metropolitan University, Manchester,*
*Manchester, UK.*
*E-mail address:* ibrahim.07000839@univsul.edu.iq*, shilan.husen@univsul.edu.iq and t.dyhoum@mmu.ac.uk