RESEARCH

Developing machine learning models to predict the fly ash concrete compressive strength

Abhinav Kapil[1] · Koteswaraarao Jadda[1] · Arya Anuj Jee[1]

Abstract

The advent and progress of machine learning (ML) have profoundly influenced civil engineering, especially in forecasting concrete's mechanical properties. This research focuses on predicting the fly ash (FA) concrete compressive strength (CS) using six different ML models: linear regression (LR), decision tree (DT), random forest (RF), extreme Ggradient boosting (XGB), support vector regression (SVR), and artificial neural network (ANN). A dataset comprising 1089 records, each with 12 input features, including the chemical compositions of FA, was used to train these models. The models' performance was assessed and compared using mean square error (MSE), mean absolute error (MAE), and the coefficient of determination (R^2), with validation achieved through the K-fold cross-validation method. Among all the models evaluated, XGB was the most accurate, attaining an R^2 value of 0.95. To interpret and understand the ML model predictions, Shapley Additive Explanations (SHAP) analysis was employed. It revealed that curing days, water-binder ratio, cement content, and superplasticizer are the most critical factors in predicting the FA concrete CS. These results indicate the potential of ML models, especially extreme gradient boosting, in accurately predicting concrete strength, promoting more efficient and effective use of FA in construction. Additionally, a graphical user interface (GUI) was created to enhance user interaction with the prediction models, improving the utility and accessibility of ML applications.

Keywords Compressive strength · Fly ash · Machine learning · Prediction · Graphical user interface

Introduction

Carbon dioxide (CO_2) emission is identified as a primary environmental concern, with cement production contributing approximately 8 to 10 percent of the total CO_2 emissions (Suhendro, 2014). This process plays a substantial role in greenhouse gas emissions and global warming (Bildirici, 2019). Currently, tackling climate change is of paramount importance worldwide. Concrete is highly valued in construction due to its mechanical strength and cost-effectiveness (Andrew, 2019). However, the construction industry, including its factories, has the largest environmental footprint among human activities. Integrating supplementary cementitious materials (SCMs) into concrete is a viable method for reducing CO_2 emissions (Scrivener et al., 2018). Hence, utilizing SCMs in concrete is an effective and environmentally responsible approach. Among these SCMs, FA is considered the predominant substitute for replacing cement in concrete mixtures (Li et al., 2022).

FA is a pozzolanic material abundant in silica and alumina, recognized for its fine powder consistency, even finer than cement. FA is a by-product which comes from the coal combustion process. According to ASTM C618 standards, FA is categorized into Class F and Class C based on its chemical composition. In the past, various researchers have extensively investigated the effect of FA on concrete performance, considering factors such as its type, chemical composition, quantity, and the extent of its replacement (Tkaczewska, 2021). Beyond its role in reducing carbon emissions, FA concrete offers several benefits. It enhances concrete's flow, binding, and water retention properties, improving its workability and performance during application (Nayak et al., 2022). FA inclusion also helps mitigate

✉ Abhinav Kapil
abhnvkapil542@gmail.com

Koteswaraarao Jadda
jaddak@nitj.ac.in

Arya Anuj Jee
jeeaa@nitj.ac.in

1 Department of Civil Engineering, Dr B R Ambedkar National Institute of Technology Jalandhar, Punjab, India

heat release during concrete hydration, reducing the risk of temperature-related cracks. Furthermore, through secondary hydration effects, FA increases compactness and improves the interface structure in concrete, resulting in enhanced impermeability and resistance against sulfate corrosion. Additionally, the prolonged reaction of volcanic ash in FA concrete improves its durability compared to conventional cement concrete.

Achieving the desired CS of FA concrete usually requires numerous adjustments to the concrete mix ratio using conventional methods. This involves casting laboratory concrete specimens and performing compression tests to evaluate CS. If the measured strength does not meet the desired standard, new specimens must be prepared, which is time-consuming and increases labor expenses. Therefore, developing an effective alternative approach that could predict the CS from a particular mix before performing compression tests would be highly beneficial. This could provide valuable insights in advance, enabling more efficient adjustments to the mix ratio and reducing the need for repeated specimen creation and testing.

The emergence and development of ML have significantly impacted civil engineering (Kaveh, 2024; Manzoor et al., 2021). Various ML models have been successfully applied to predict the compressive strength of concrete, yielding promising results (Al-Gburi & Yusuf, 2022; Sathiparan, 2024; Sathiparan et al., 2023). These techniques rely on extensive datasets to build precise models. The accuracy of their predictions primarily depends on the quality and completeness of the data samples collected from experimental procedures during specimen casting or from literature studies. Researchers utilize these algorithms to predict the mechanical properties of concrete with improved reliability and efficiency.

Kaveh et al. (1999) developed a hybrid method integrating graph theory with neural networks for domain decomposition, enhancing accuracy and efficiency in structured finite element meshes. Iranmanesh and Kaveh (1999) introduced a neurocomputing strategy combining neural networks with structural optimization techniques. Singh et al. (2023) used ML models with 14 input parameters on a dataset of 400 points to predict the CS of red mud (RM)-based concrete. DT and extra tree regressor (ET) models provided the best fit. Microstructural analysis and leaching tests confirmed the safety and compliance of RM concrete, making it suitable for eco-friendly construction, especially for low-traffic or rural roads. Albostami et al. (2023) applied data-driven approaches to predict the CS of self-compacting concrete (SCC) with recycled plastic aggregates (RPA). Using 400 experimental datasets, they employed multi-objective genetic algorithm evolutionary polynomial regression (MOGA-EPR) and gene expression programming (GEP). These models outperformed the traditional LR model. Kaveh et al. (2021) applied ML to relate

fiber angle and buckling capacity under bending-induced loads. Their deep learning model, trained on a dataset of 11,000 cases, outperformed RF, DT, and LR models, demonstrating superior accuracy and generalization. Kaveh et al. (2023) developed metaheuristic-trained ANNs to predict the ultimate buckling load of high-strength steel columns. Using particle swarm optimization and genetic algorithms to optimize ANN weights and biases, their models achieved up to 99.8% accuracy.

In the context of FA-based concrete, Ahmad et al. (2021) conducted a study on the utilisation of ML techniques to predict the CS of concrete incorporating SCMs. They employed bagging, DT, adaptive boosting, and GEP models. Among these, the bagging regressor provided the best prediction results. In their study, coarse aggregate, fine aggregate, and cement contributed 24.6%, 18.4%, and 16.3%, respectively to the prediction outcomes. Jiang et al. (2022) used ML algorithms to predict the CS of concrete made with FA. They employed four ML models: RF, extreme learning machine, SVR, and support vector regression with grid search (SVR-GS). The SVR-GS model produced the most accurate predictions, with age and water-cement ratio being the most influential features affecting CS. Mahajan and Bhagat (2022) investigated ANN, DT, GEP, and bagging regressor to predict the CS of concrete with FA admixture. Their prediction model used seven input elements (cement content, fine aggregate, coarse aggregate, fly ash, superplasticizer, water content, and curing days) to predict the output parameter. The bagging algorithm outperformed ANN, DT, and GEP, achieving an R^2 value of 0.97, compared to 0.81, 0.78, and 0.82, respectively. Chopra et al. (2016) utilized genetic programming and ANN to forecast the concrete CS, both with and without FA. They collected the relevant data from controlled laboratory experiments at various curing periods. The prediction results indicated that the ANN model, using the Levenberg–Marquardt (LM) training function, was the most effective tool for predicting concrete CS.

Research significance

Several experimental studies have investigated the impact of adding FA on concrete CS. However, only a few have focused on predicting FA concrete CS using ML models. Moreover, many of these studies have relied on a limited number of data-set points and input parameters. Notably, the use of chemical composition (silica content, lime content, iron oxide content, aluminum oxide content, and loss on ignition) of FA as input parameters for predicting concrete CS has rarely been reported in the literature. This inclusion of chemical composition addresses the variability in FA properties, which significantly influence concrete's mechanical properties.

Addressing these gaps, the current research aims to employ six distinct ML models with 1,089 dataset points and 12 input parameters to predict the FA concrete CS. The research objectives are:

- To develop ML models that can accurately predict the FA concrete CS.
- To compare the performance of the models using metrices: MSE, MAE, and R^2.
- To examine the relative significance and impact of each input feature on the CS.
- To develop a comprehensive graphical user interface (GUI) to facilitate user interaction with the prediction models.

Methodology

The flowchart for the methodology used in the current study is shown in Fig. 1.

Data collection

A total of 1,089 dataset points based on the use of FA in concrete were collected from existing literature (Alaka & Oyedele, 2016; Balakrishnan & Awal, 2014; Barbhuiya et al., 2009; Chen et al., 2019; Chindaprasirt et al., 2007; Atis, 2003; Durán-Herrera et al., 2011; Felekoglu, 2006; Hashmi et al., 2020; Golewski, 2018; Hansen, 1990; Huang et al., 2013; Kumar et al., 2007; Kumar et al., 2021; McCarthy & Dhir, 2005; Mehta & Gjorv, 1982; Mukherjee et al., 2013; Nochaiya et al., 2010; Oner et al., 2005; Reiner & Rens, 2006; Saha, 2018; Shaikh & Supit, 2015; Siddique, 2004; Siddique & Khatib, 2010; Sun et al., 2019; Woyciechowski et al., 2019; Yazici et al., 2012) in terms of twelve input parameters: water-binder (w/b) ratio, cement content (kg/

m^3), coarse aggregate (kg/m^3), fine aggregate (kg/m^3), silica dioxide (%), calcium oxide (%), ferric oxide (%), aluminum oxide (%), loss on ignition (%), superplasticizer (kg/m^3), curing days, and replacement percentage, and one output parameter: compressive strength (Mpa). The dataset included 35 different types of fly ashe, each characterized by diverse chemical and physical properties. The database incorporated data from concrete specimens of varying shapes and sizes, with four distinct configurations utilized. Relevant shape factors were employed in analyzing these specimens. Out of the 1,089 datapoints, 872 (80%) were allocated for training the models, while 217 (20%) were designated for testing the models. The various input parameters are depicted in Fig. 2.

Pre- processing

In the pre-processing phase, the dataset was subjected to standard scaling to ensure all numeric features were on a comparable scale. This involved centering the data around zero and rescaling it to unit variance using Python's standard scaling functionality. By standardizing the features in this manner, potential issues stemming from varying scales were mitigated, ensuring that each feature contributed equally to the model's learning process.

Statistical analysis

Descriptive statistical analysis of the input and output variable (CS) is summarized in Table 1, where 'mean' represents average value, 'std' represents standard deviation, 'min' and 'max' signify minimum and maximum values, '25%', '50%', and '75%' represent first, second, and third quartile, and 'skew' and 'kurt' signify skewness and kurtosis, respectively.
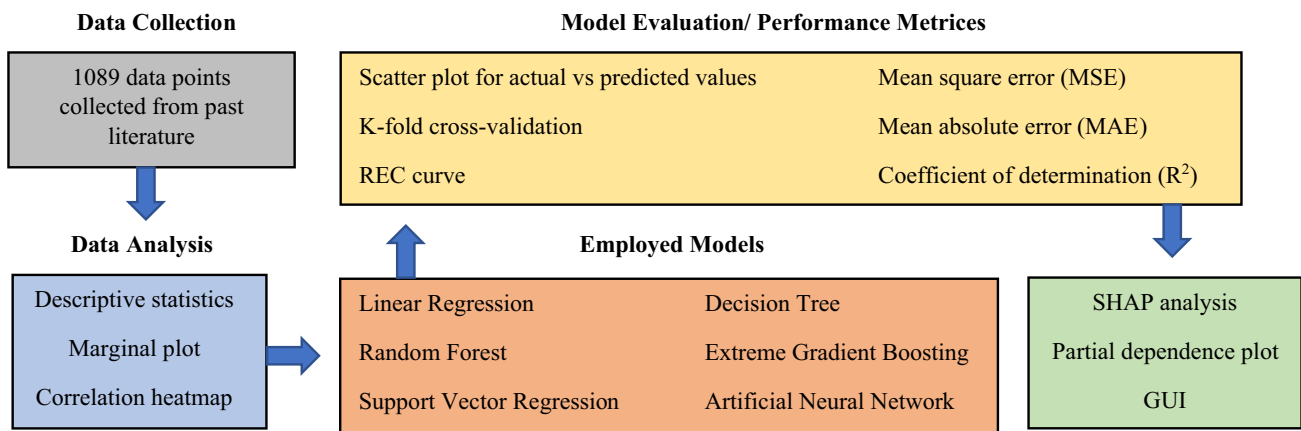

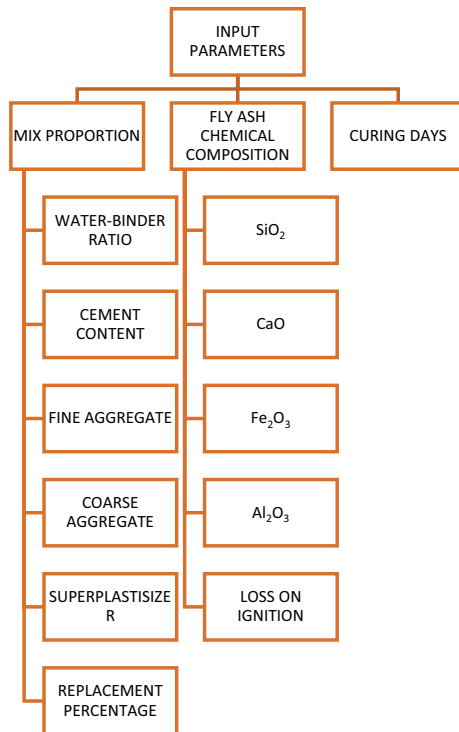**Fig. 1** Flowchart of the methodology adopted in the currrent study

Fig. 2 Input parameters used in the current study

Machine learning models employed

Linear regression (LR)

LR is a foundational statistical technique used to model the relationship between an outcome variable and one or more predictor variables. This is achieved by fitting a linear equation to the observed data to capture underlying patterns and trends (Su et al., 2012). The model

determines the coefficients for each input feature by minimizing the sum of squared errors between the predicted and actual values. The predicted value is calculated as a linear combination of the input features, where each feature is multiplied by its respective coefficient and summed together. For this study, an LR model was initialized and trained using the LinearRegression class from the scikit-learn library in Python. The mathematical equation for the trained LR model is represented in Eq. (1) below.

$$
\begin{aligned}
y =\ & 104.295 + (-76.280 * \text{W/b ratio}) \\
& + (-0.004 * \text{Cement Content}) \\
& + (0.011 * \text{Fine Aggregate}) \\
& + (0.001 * \text{Coarse Aggregate}) \\
& + (-0.478 * \text{SiO2}) + (-0.512 * \text{CaO}) \\
& + (0.005 * \text{Fe2O3}) + (-0.116 * \text{Al2O3}) \\
& + (-0.233 * \text{Loss on ignition}) \\
& + (0.397 * \text{Superplastisizer}) \\
& + (0.109 * \text{Curing Days}) \\
& + (-0.483 * \text{Replacement Percentage})
\end{aligned}
\tag{1}
$$

Decision tree (DT)

DT is a supervised learning algorithm employed for predictive modeling. The model functions by recursively dividing the feature space into regions (Myles et al., 2004). Each internal node represents a decision based on a particular attribute, while each leaf node represents a predicted value. This approach allows the model to capture nonlinear relationships between input features and the target variable. In this study, a DT model was initialized using the DecisionTreeRegressor class from the

Table 1 Descriptive statistical analysis of variables

Parameter	Mean	Std	Min	Max	25%	50%	75%	Skew	Kurt
W/b ratio	0.47	0.13	0.22	0.94	0.36	0.45	0.55	0.61	0.14
Cement content	224.88	89.24	83	486	140	205	290	0.52	− 0.57
Fine aggregate	757.14	210.81	0	1293	684	783.20	895	− 1.19	3.11
Coarse aggregate	1047.59	219.87	436	1970	975	1062	1180	0.30	3.33
Silica oxide (SiO_2)	53.39	8	33.14	76.34	49.70	50.90	60	0.58	0.52
Calcium oxide (CaO)	6.93	6.90	0.59	35.18	2.35	4.40	6	1.93	3.67
Ferric oxide (Fe_2O_3)	8.60	6.17	3.30	26.87	5.27	6.98	8	2.31	4.16
Aluminum oxide (Al_2O_3)	22.71	7.24	4.27	36.60	21.63	23.55	28	− 1.32	1.33
Loss on ignition	2.49	2	0.05	7.80	1	1.66	3.60	1.16	0.82
Superplastisizer	2.89	3.69	0	18.20	0	1.70	4.25	1.63	2.30
Curing days	67.88	91.15	1	365	7	28	90	2.10	3.90
Replacement percentage	41.58	18.70	5	80	30	40	60	0.16	− 0.70
Compressive strength	33.42	18.85	0	95.58	18.52	30.98	47.02	0.46	− 0.44

Table 2 Hyperparameters for the DT model

Hyperparameter	Value
Maximum tree depth	15
Maximum features per split	'sqrt''
Minimum samples per leaf	1
Minimum samples per split	2

Table 3 Hyperparameters for the RF model

Hyperparameter	Value
Maximum tree depth	15
Maximum features per split	'sqrt'
Minimum samples per leaf	1
Minimum samples per split	2
Number of estimators	100

scikit-learn library in Python. Unlike ensemble methods such as random forests, decision tree regression involves constructing a single decision tree trained on the entire dataset. The hyperparameters chosen for the model are shown in Table 2. These hyperparameters were carefully selected to balance the model's complexity and predictive accuracy. Decision tree up to the depth of three is shown in Fig. 3.

Random forest (RF)

RF is an ensemble learning technique that builds multiple decision trees during training and averages their predictions to produce a final output (Biau & Scornet, 2016). This method enhances predictive performance and reduces overfitting by training each tree on a random subset of features and data samples. In this research, an RF model was initialized using the RandomForestRegressor class from the scikit-learn library in Python. To determine the best hyperparameters, a grid search was conducted, wherein the mean square error was computed for different leaf sizes and plotted against the number of estimators, allowing us to visualize the relationship between the number of estimators and model performance. By analyzing the graph, hyperparameters were tuned to minimize MSE and improve model accuracy. The chosen hyperparameters and the aforementioned graph are shown in Table 3 and Fig. 4, respectively.
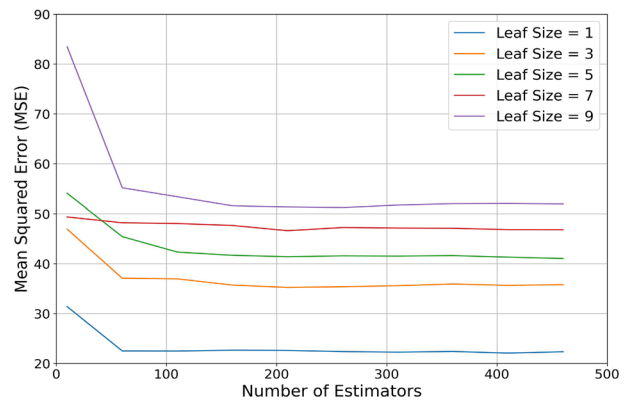
Fig. 4 RF MSE vs. number of estimators for different leaf sizes

Extreme gradient boosting (XGB)

XGB is a highly optimized and scalable implementation of gradient-boosting machines. It is renowned for its exceptional performance in various ML tasks, particularly in regression and classification problems. XGB operates by iteratively incorporating decision trees into an ensemble, wherein each tree is trained to rectify the errors made by the preceding ones (Chen & Guestrin, 2016). This boosting process focuses on minimizing a loss function by optimizing the predictions of the ensemble. An XGB model was
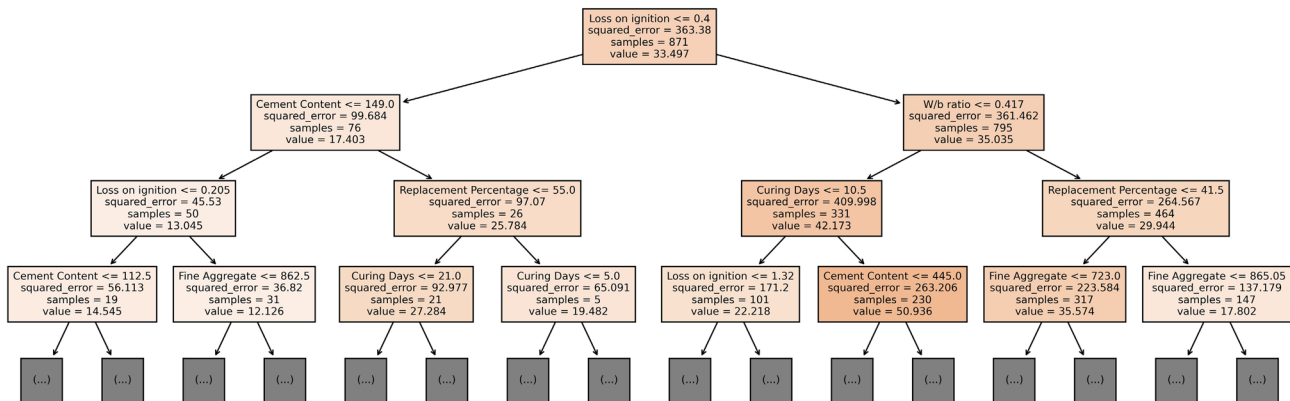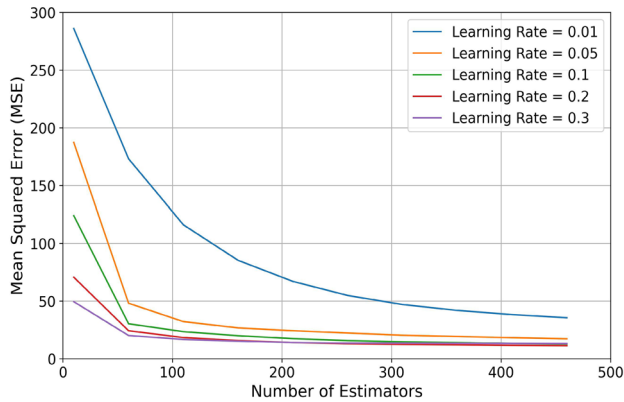
Fig. 3 DT regressor up to tree depth three

Table 4 Hyperparameters for the XGB model

Hyperparameter	Value
Maximum depth	3
Learning rate	0.2
Number of estimators	200

Table 5 Hyperparameters for the SVR model

Hyperparameter	Value
Kernel	'rbf'
C	400
Epsilon	1
Gamma	0.1

Fig. 5 XGB MSE vs. number of estimators for different learning rates
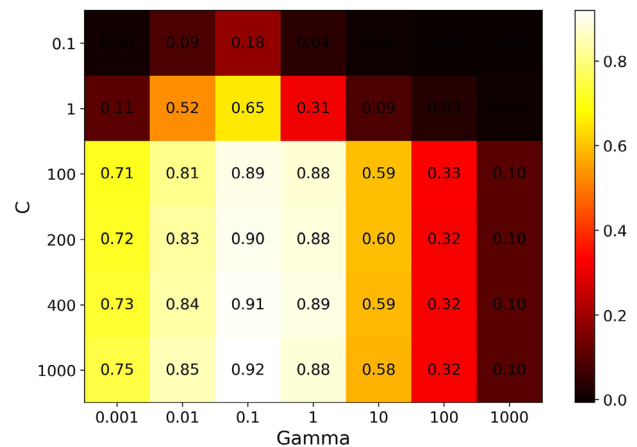
Fig. 6 Accuracy heatmap for SVR model for different combinations of C and Gamma

initialized and trained using the XGBRegressor class from the xgboost library in Python. To determine the optimal hyperparameters, a grid search was conducted, wherein the mean square error was computed for different learning rates and plotted against the number of estimators. This allowed us to visualize the relationship between the number of estimators and model performance, aiding in the decision about hyperparameters. The chosen hyperparameters and the relevant graph are shown in Table 4 and Fig. 5, respectively.

Support vector regression (SVR)

SVR leverages the principles of support vector machines for regression analysis, offering a robust technique. Its objective is to identify the optimal hyperplane that maximizes the margin while minimizing the error between the predicted and observed values (Pisner & Schnyer, 2019). In this study, an SVR model was trained using the SVR class from the scikit-learn library in Python. Prior to training, the features were standardized with the StandardScaler from the same library to ensure consistent scaling across different features, thereby enhancing model performance. To fine-tune the hyperparameters and find the best combination of C and Gamma, a (r-squared) value heatmap was plotted for various combinations. The selected hyperparameters and the accuracy heatmap are shown in Table 5 and Fig. 6, respectively.

Artificial neural network (ANN)

ANNs are computational models consisting of interconnected nodes arranged in layers: an input layer, one or more hidden layers, and an output layer. Each node performs a transformation on its input and forwards the outcome to the nodes in the subsequent layer. Through a process termed training, ANNs adjust the weights of connections between nodes to minimize a loss function and enhance predictive accuracy (Khan, 2018). In the current research, the model architecture was defined using the Keras library, with a sequential model featuring an input layer, a dense hidden layer with a variable number of neurons, and an output layer. The hidden layer utilized the rectified linear activation function (ReLU), while the output layer used a linear activation function, suitable for regression tasks. To determine the optimal number of neurons in the hidden layer, a graph was plotted showing the mean squared error versus the number of neurons. This visualization helped select the model complexity that best balanced underfitting and overfitting. The chosen hyperparameters and the above-mentioned graph are shown in Table 6 and Fig. 7, respectively.

Table 6 Hyperparameters for the ANN model

Hyperparameter	Value
Hidden layer neurons	50
Activation function	'relu'
Output layer activation	'linear'
Optimizer	'adam'
Loss function	MSE
Epochs	1000
Batch size	10
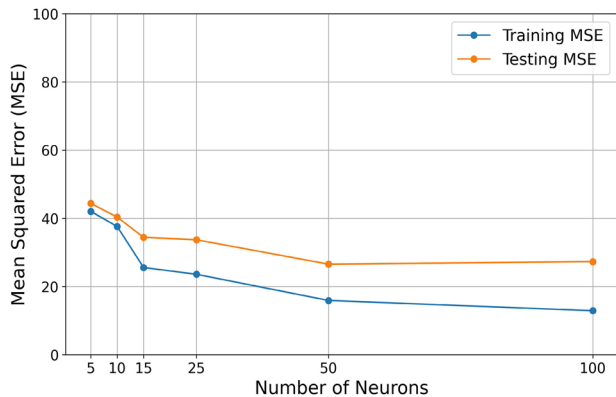
Fig. 7 ANN MSE vs. number of neurons

Results and discussion

Data visualization plots

Marginal plot

A marginal plot combines a scatter plot of input variables against the output variable with histograms or density plots of each variable along the axes. This plot allows for a simultaneous examination of the relationship between predictor variables and the output variable, while also visualizing the distribution of each variable. It facilitates understanding of how the input variables collectively affect the output variable and provides insights into their individual distributions. The marginal plot for all input variables with respect to the output variable is shown in Fig. 8.

Correlation heatmap

A correlation heatmap visually represents a correlation matrix, using colors to indicate the magnitude and direction of correlations between variables. Typically, warmer colors denote positive correlations, cooler colors represent negative correlations, and neutral colors signify no correlation. These heatmaps illustrate linear correlations between all possible combinations of variables in a dataset, offering insights into relationships and patterns that may exist among them. The heatmap for the employed dataset is presented in Fig. 9.

Curing days (0.56) and cement content (0.34) exhibit the highest positive correlation with the output CS, while the water-binder ratio (-0.39) and replacement percentage (-0.29) show the highest negative correlation coefficients. Furthermore, since no features are uncorrelated, all twelve input parameters can be utilized for predicting the CS.

Performance metrices

The comparison of the six regression models revealed distinct performance differences, highlighted through three key metrics: MSE, MAE, and R^2. These metrics are essential for understanding how well our models predict outcomes. MSE acts as a ruler, emphasizing larger errors by squaring the differences between forecasted and observed values (Allen, 1971). MAE focuses on the average size of errors without considering their direction (Willmott & Matsuura, 2005). R^2 indicates how effectively the model explains the variability of the dependent variable using the independent variables. Higher R^2 values suggest that the model better captures the patterns and relationships in the data.

The mathematical expressions for MSE, MAE, and R^2 are given in Eqs. (2), (3) and (4) (Chicco et al., 2021).

$$\text{MSE} = (1/n) * \Sigma(y_i - \hat{y}_i)^2 \tag{2}$$

$$\text{MAE} = (1/n) * \Sigma|y_i - \hat{y}_i| \tag{3}$$

$$R^2 = 1 - (\text{SSres} / \text{SStot}) \tag{4}$$

where 'n' represents the number of samples, 'y_i' denotes the actual value, '$\hat{y}_i$' denotes the predicted value, 'SSres' represents the sum of squared residuals (errors), and 'SStot' stands for the total sum of squares.

The MSE, MAE, and R² values for the employed models are presented in Table 7.

Ensemble models, namely XGB and RF outperform other models exhibiting low MSE and high R^2 values on both training and testing datasets. DT also performs well on the training data but shows moderate generalization to the testing data. While SVR and ANN show moderate performance on both training and testing datasets.
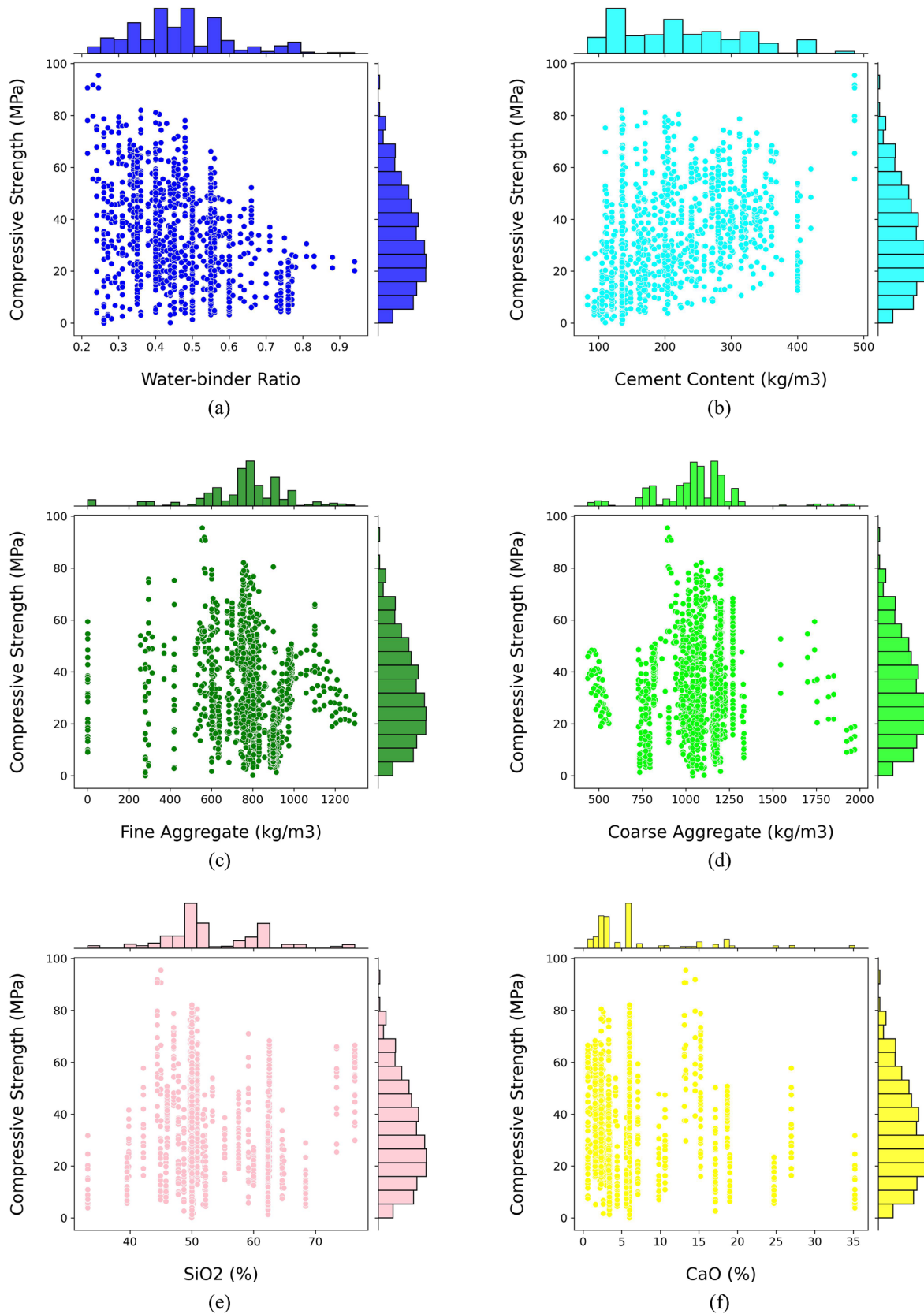
Fig. 8 Marginal plot between FA concrete CS and **a** water-binder ratio, **b** cement content, **c** fine aggregate, **d** coarse aggregate, **e** SiO$_2$, **f** CaO, **g** Fe$_2$O$_3$, **h** Al$_2$O$_3$, **i** loss on ignition, **j** superplastisizer, **k** curing days, and **l** replacement percentage
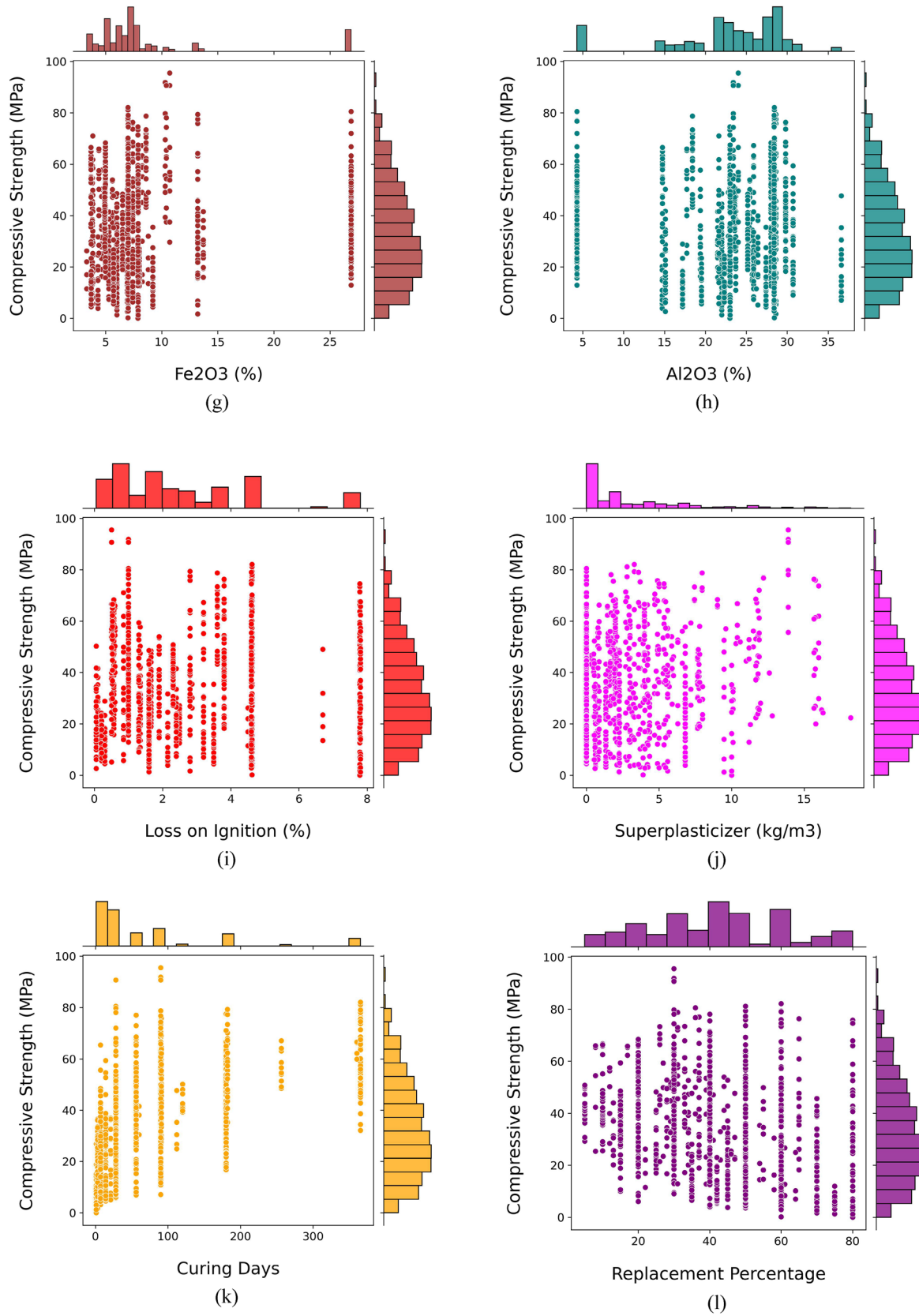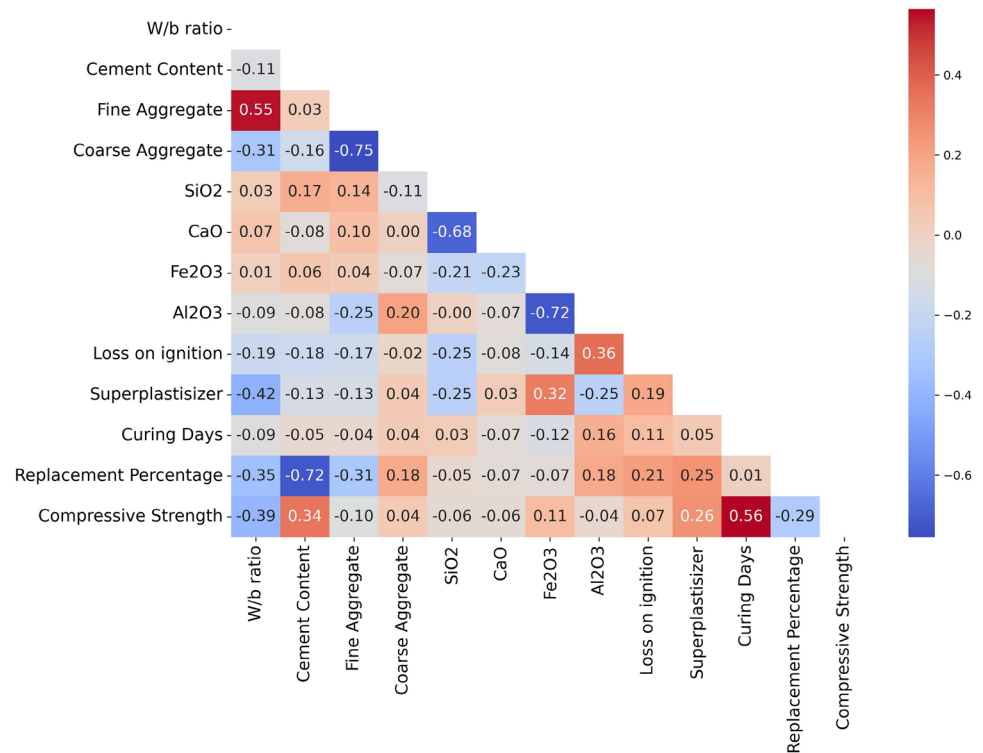
Fig. 8 (continued)

Fig. 9 Correlation matrix heatmap for the employed data-set

Table 7 Performance metrics for employed models for training and testing data-sets set

	Training			Testing		
	MSE	MAE	R^2	MSE	MAE	R^2
Linear regression	122.19	8.59	0.66	95.42	7.65	0.7
Decision tree	8.12	0.93	0.97	63.96	5.93	0.80
Random forest	4.19	1.47	0.98	22.37	3.57	0.93
Extreme gradient boosting	4.59	1.59	0.98	14.17	2.67	0.95
Support vector regression	16.5	2.48	0.95	28.71	3.69	0.91
Artificial neural network	15.47	2.78	0.95	23.48	3.65	0.92

Prediction plot

A scatter plot of actual versus predicted values visually assesses how well the model predictions align with the true values. Each point on the plot represents a data instance, where the x-coordinate denotes the actual value, and the y-coordinate represents the predicted value. Ideally, all points would lie on the diagonal line (the identity line), indicating perfect alignment between forecasted and actual values. Figure 10 displays the scatter plot of actual vs. predicted values for the FA concrete CS for all the models used in this study.

Residual plot and distribution of residuals

In regression analysis, residual plots and the distribution of residuals play pivotal roles in assessing the adequacy and

validity of the regression model (Suleiman et al., 2015). A residual plot visually depicts the differences between observed and predicted values, typically plotted against the independent variable(s) or the predicted values themselves. This graphical representation enables researchers to scrutinize key aspects of the model's performance: linearity and homoscedasticity. Specifically, a horizontal pattern in the residual plot suggests a linear relationship between the independent and dependent variables, while a consistent spread of residuals across all levels of the independent variable(s) or predicted values indicates homoscedasticity.

The distribution of residuals provides insights into normality, skewness, and kurtosis, aiding in the assessment of the assumptions underlying the regression model. Deviations from normality or symmetry in the residual distribution may signal issues with the model's validity and highlight areas for refinement or further investigation. The percentage of
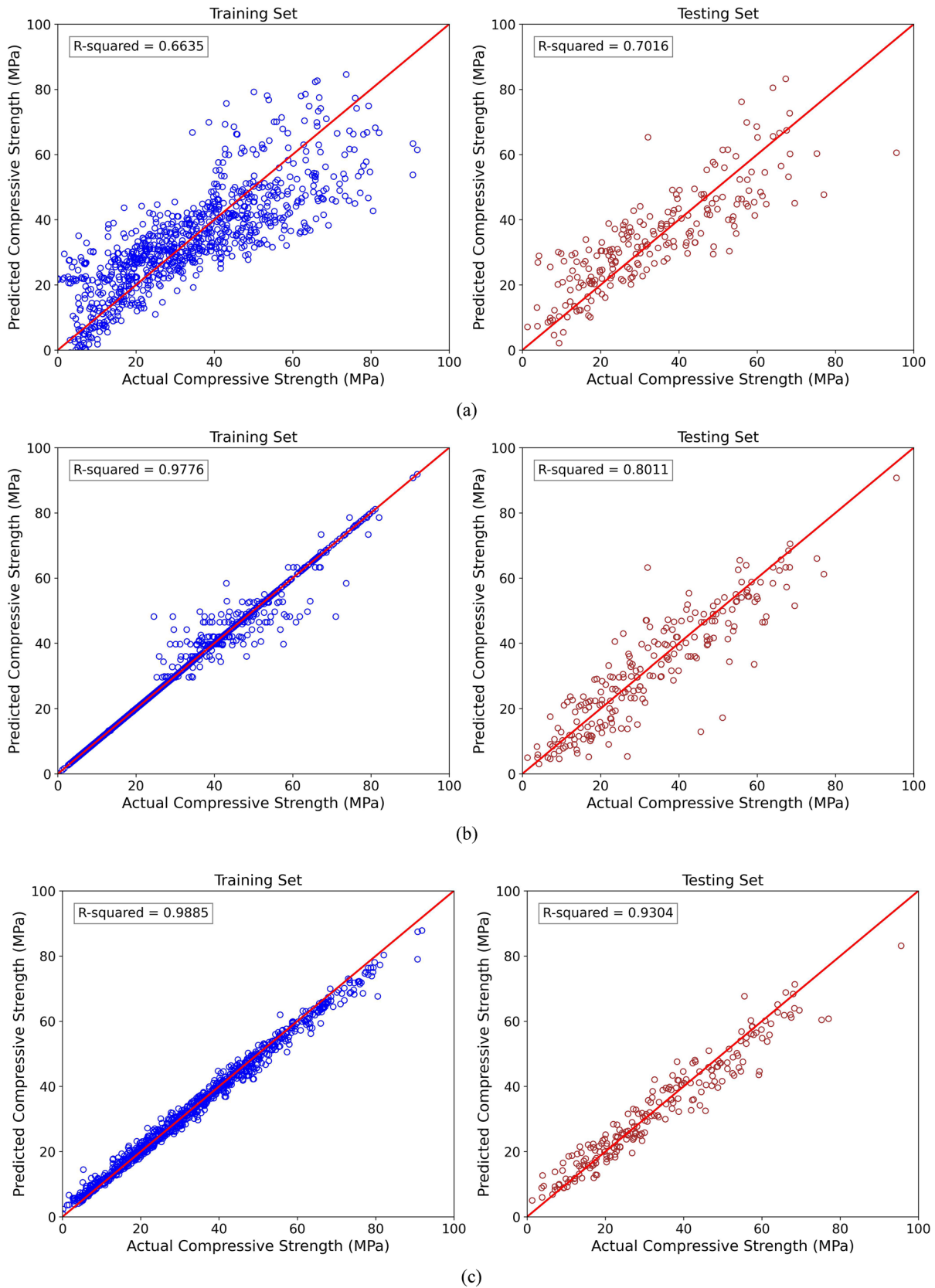
Fig. 10 Scatter plot of actual vs. predicted values of FA concrete CS for **a** LR, **b** DT, **c** RF, **d** XGB, **e** SVR, and **f** ANN model for training and testing data-sets
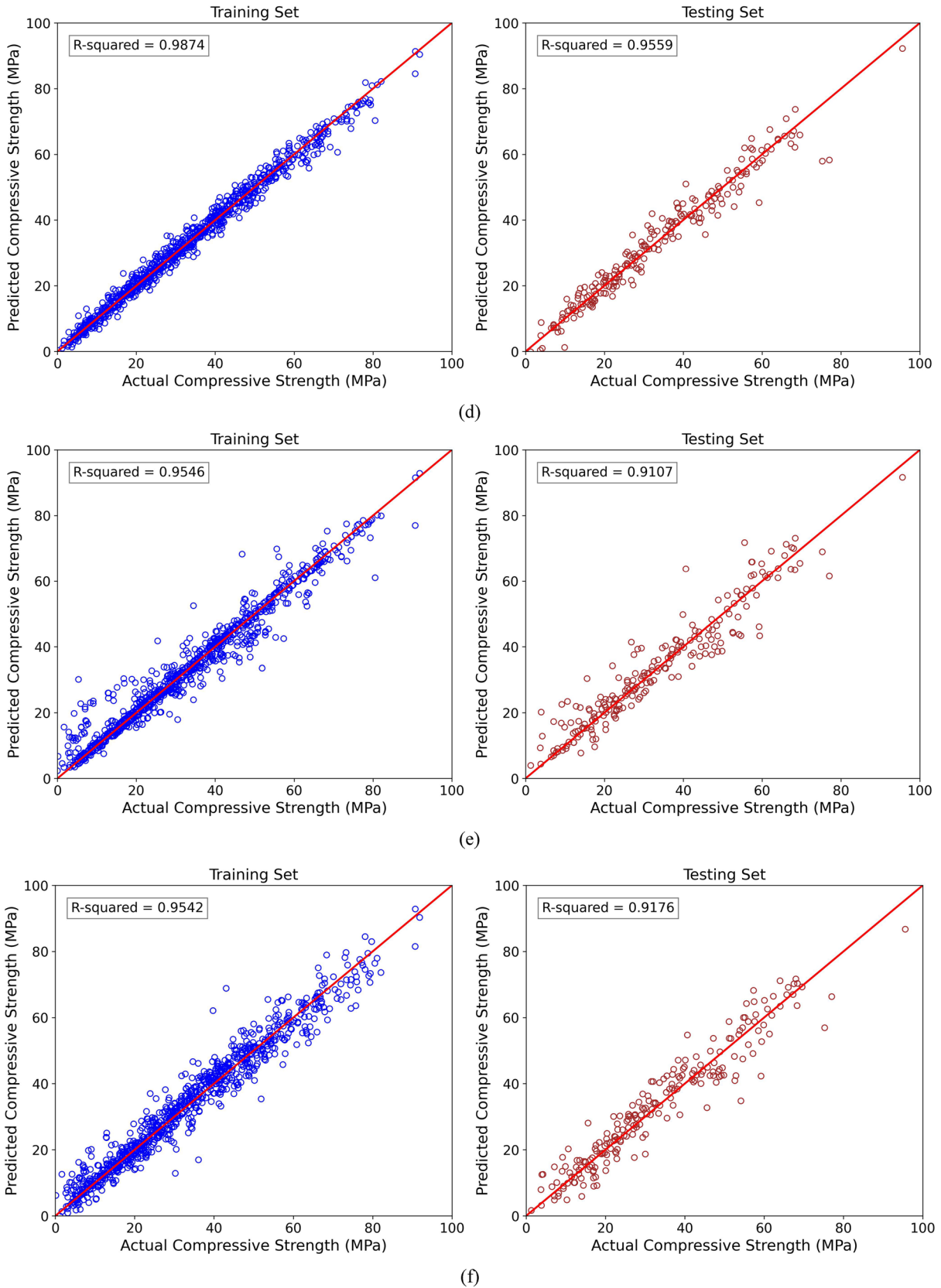
(d)

(e)

(f)

Fig. 10 (continued)

Table 8 Percentage predictions within ± 5 for employed models

Model	Percentage predictions within ± 5
LR	40.83
DT	50.46
RF	73.39
XGB	87.16
SVR	75.23
ANN	75.69

predictions within ± 5 for the employed models is shown in Table 8.

In examining all six models, the consistent presence of random scatter in the residual plots indicates that our modeling techniques effectively capture the diverse relationships within the data, without exhibiting systematic patterns. Furthermore, the residuals' normal distribution, centered around zero for most models, reinforces the reliability and robustness of our approach. In summary, the collective analysis of these models provides strong evidence supporting the validity of our statistical modeling framework in comprehensively explaining the inherent variability within the dataset.

The residual plot and distribution of residuals for employed models are shown in Fig. 11.

K-fold cross validation

K-fold cross-validation is a method employed to evaluate the performance of ML models accurately. In this method, the dataset is randomly divided into k equally sized subsets. One subset is set aside for validation, while the remaining k-1 subsets are utilized for training. This procedure is repeated k times, with each subset serving as the validation set exactly once. By averaging the results of these iterations, a more reliable assessment of the model's performance is achieved, reducing potential biases. In the described study, a 10-fold cross-validation approach was utilized. The outcomes were evaluated using MSE, MAE, and R^2, as shown in Fig. 12.

Across the folds, RF and XGB consistently demonstrate the lowest MSE and highest R^2 values, indicating robust performance and strong predictive accuracy. SVR also maintained competitive performance with moderate MSE and high R^2 values. Conversely, DT exhibits higher variability and generally higher MSE, while LR consistently displays the highest MSE and lowest R^2 values, suggesting less reliable predictive capability.

Regression error characteristics (REC)

The regression error characteristic (REC) curve is a graphical tool for evaluating regression models. It plots absolute error values on the x-axis and cumulative distribution function (CDF) values on the y-axis (Bennett et al., 2003). This curve illustrates how prediction error varies across different levels of accuracy. The CDF, representing the cumulative proportion of data points with absolute errors less than or equal to a certain threshold, provides valuable information about the distribution of errors in the predictions made by the regression model. REC curves are instrumental for comparing models and understanding how the dataset size affects prediction accuracy. The REC curve for the employed models is depicted in Fig. 13.

XGB, RF, ANN, and SVR show strong performance, as evidenced by their close alignment with the x-axis, indicating lower error rates across various thresholds. In contrast, LR and DT perform poorly and moderately, respectively.

Shapley additive explanation (SHAP) analysis

SHAP, or Shapley Additive Explanations, is a mathematical method used to interpret the predictions of ML models. A SHAP summary plot provides a comprehensive overview of feature imnportance and their influence on model predictions. It displays features along the y-axis, ranked by their importance, while the x-axis represents the average magnitude of SHAP values, indicating the direction and magnitude of each feature's impact on predictions across all data points. For this study, SHAP analysis was performed using the XGB model due to its superior performance, as shown in Fig. 14.

Curing days, water-binder ratio, cement content, and replacement percentage are the most impactful parameters in the prediction of FA concrete CS for the given data set.

Partial dependence plot

Partial dependence plots (PDPs) are visual tools that show the relationship between a subset of input features and the predicted outcome of a model. PDPs display how changes in specific features affect the predicted response while averaging out the effects of all other features. This helps in understanding the influence of each feature on the model's predictions, providing insights into the model's behavior and feature importance. PDPs can also serve as a validation tool, ensuring that the model's predictions are consistent with domain knowledge or expectations. In this study, PDPs were constructed for the most influential parameters— curing days, water-binder ratio, and cement content— using the XGB model while keeping the values of other features constant (at their mean), as shown in Fig. 15.

Graphical user interface (GUI)

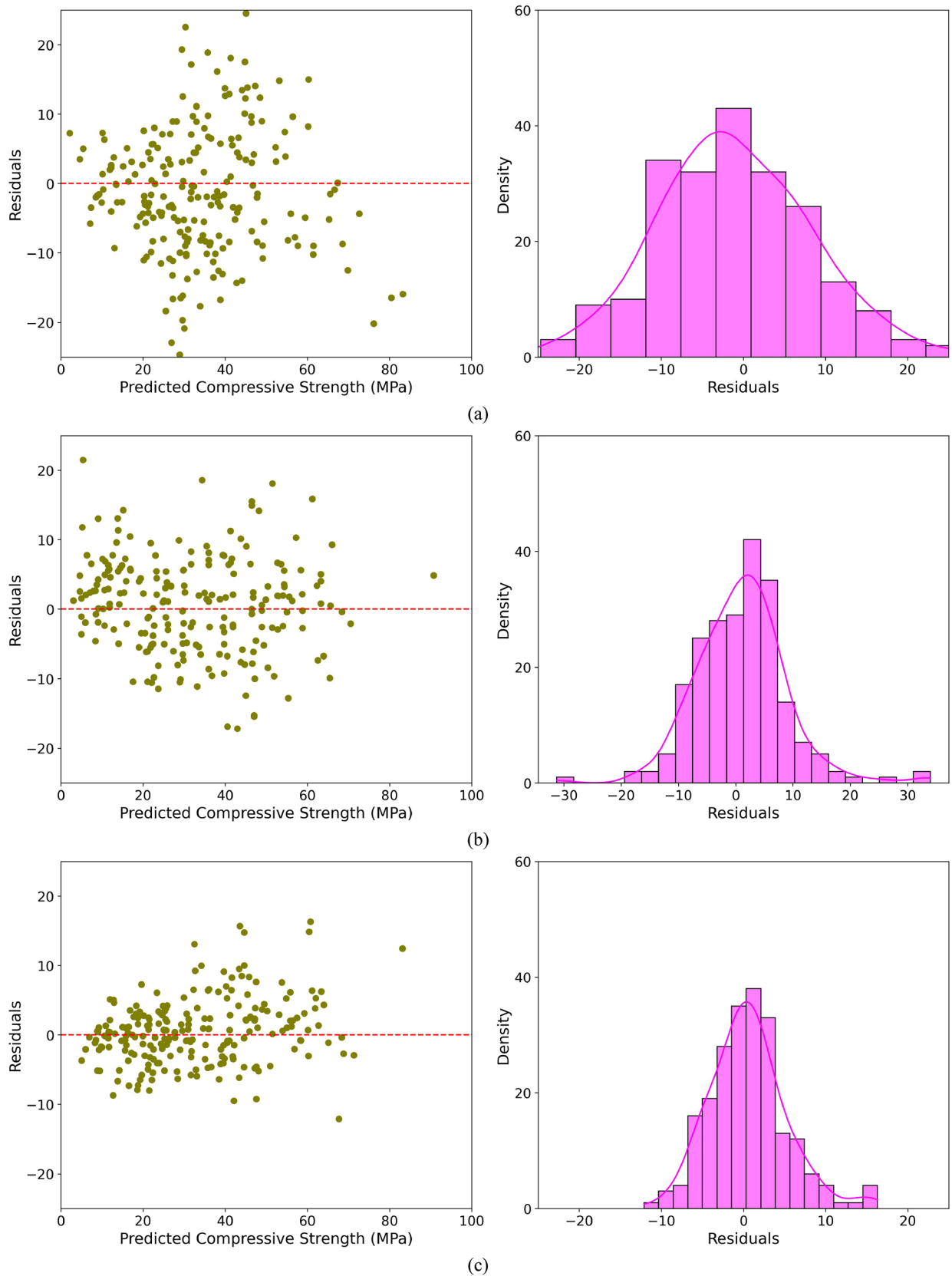The development of a graphical user interface (GUI) for the prediction models marks a major step forward in enhancing

Fig. 11 Residual plot and distribution of residuals for **a** LR, **b** DT, **c** RF, **d** XGB, **e** SVR, and **f** ANN model
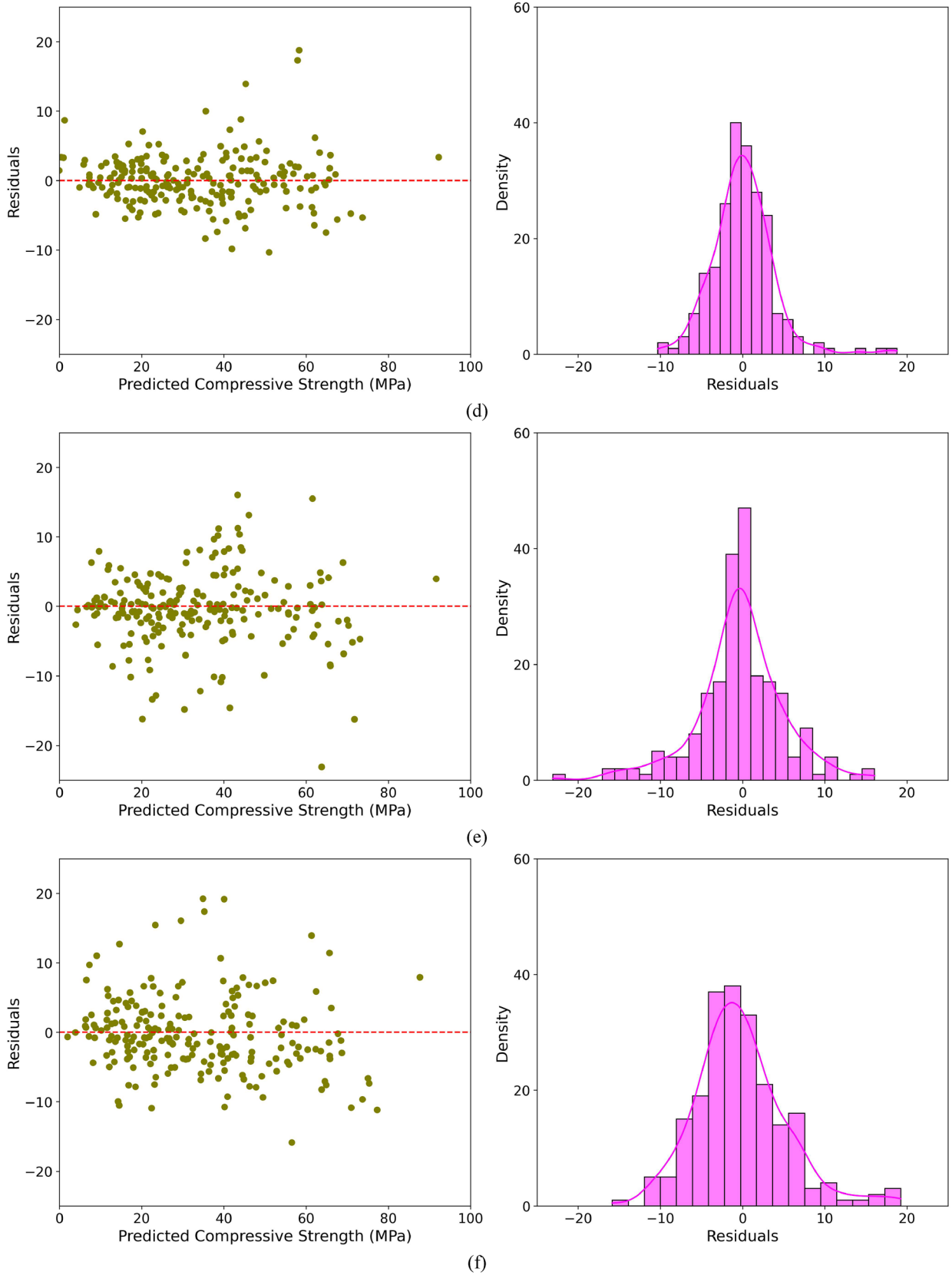
(d)

(e)

(f)
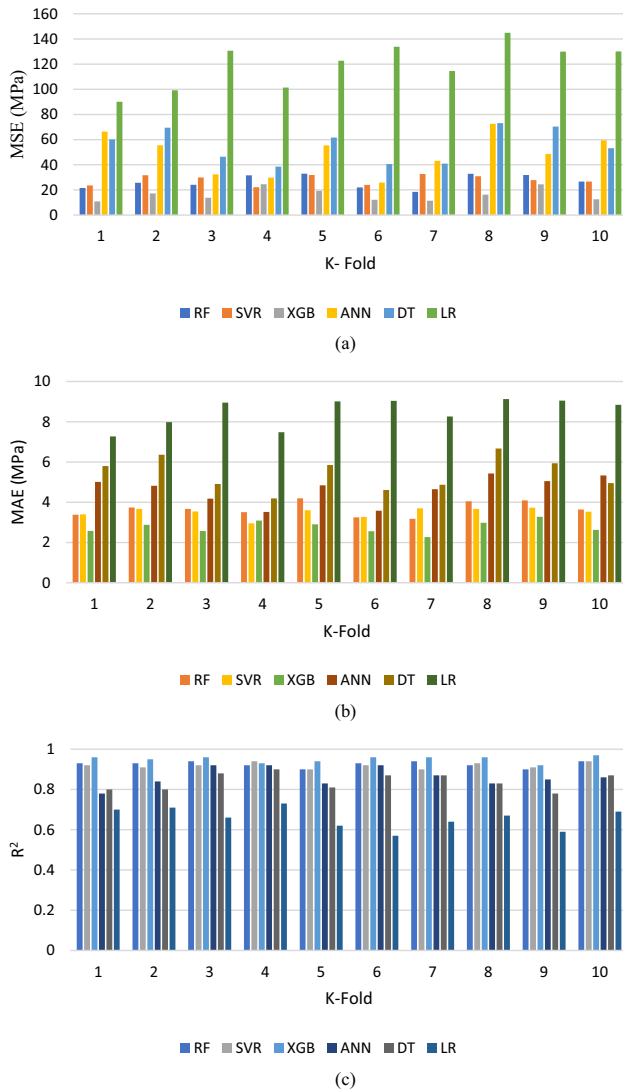
Fig. 11 (continued)

Fig. 12 K-fold cross validation results using **a** MAE, **b** MSE, and **c** R^2 for employed models
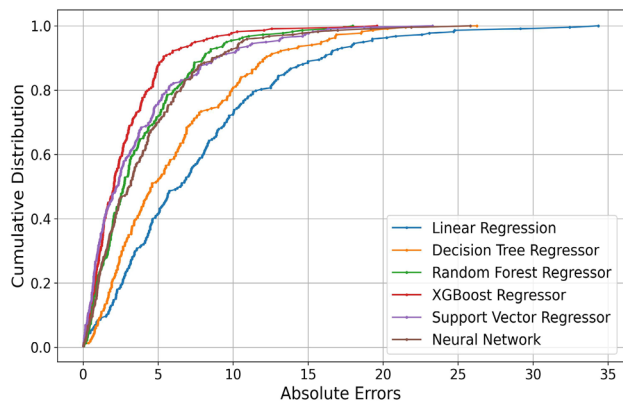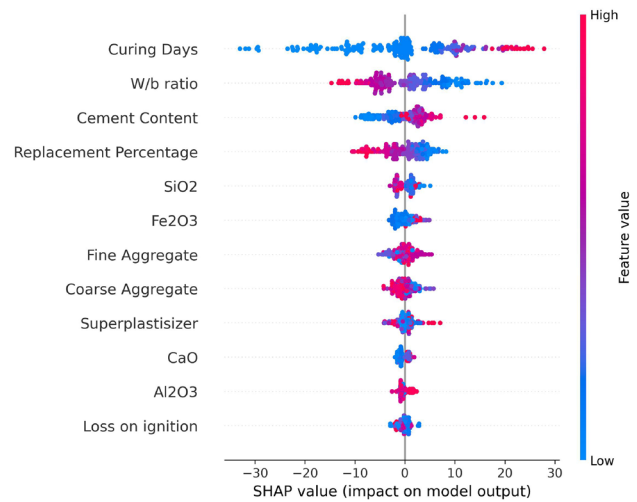

Fig. 13 REC analysis of employed models

Fig. 14 SHAP analysis values for the XGB model

the practicality and availability of ML applications. A GUI was built using the Flask framework and subsequently deployed on Render. The interface features a dedicated space for users to input values for all relevant features, ensuring comprehensive data entry. Additionally, a drop-down menu is incorporated, allowing users to select the type of model they wish to employ for predictions, thus providing flexibility and adaptability to varying model architectures and algorithms. This interface enables users to trigger predictions with a single click and receive immediate feedback on the predicted outcomes. The GUI and related code files are available at https://fa-cs-pred-ekc0.onrender.com/, and https://github.com/abhinavkapil/FA_CS_PRED, respectively. The interface of the developed GUI is shown in Fig. 16.

Conclusions

This study employed six distinct ML models comprising 1089 data-set points extracted from the use of FA in concrete in terms of twelve input parameters to predict the FA concrete CS. The following key findings emerged from this study:

1. The wide range of input variables and the output variable, as evidenced by the statistical analysis and marginal plots, served to validate the reliability of the collected dataset.
2. Correlation analysis revealed that no features were uncorrelated, so all the input features were utilized to increase the accuracy of the developed models.
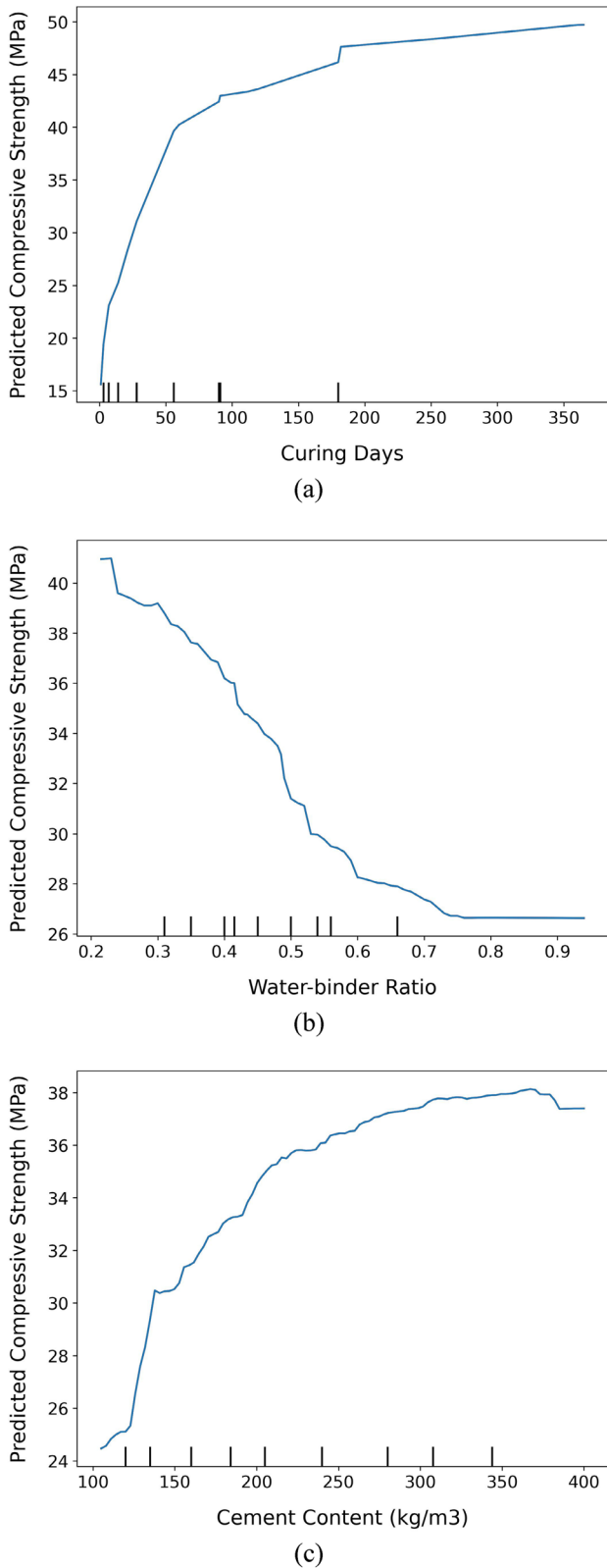
Fig. 16 The interface of the prepared GUI to predict the FA concrete CS

3. The ensemble ML models (RF and XGB) showed better performance, as indicated by higher values of R^2 and lower statistical errors (MSE and MAE), with XGB being the most accurate (r-squared value of 0.95). SVR and ANN performed moderately on both training and testing datasets, meanwhile, DT and LR were the least effective in predicting the results, with R^2 values of 0.80 and 0.70, respectively.

4. K-fold cross-validation, which was utilized to confirm the accuracy of developed models revealed similar results with the XGB regressor showing superior performance across all folds.

5. Based on REC analysis, XGB, RF, SVR, and ANN showed strong performance, with low error rates across various thresholds, while DT and LR performed moderately and poorly, respectively.

6. Based on SHAP analysis, curing days, water-binder ratio, cement content, and replacement percentage were the most critical parameters in FA concrete CS prediction for the given data set.

7. The partial dependence plots for curing days, water-binder ratio, and cement content were consistent with the general trend.

8. A graphical user interface (GUI) was successfully developed, which will enable users to predict the FA concrete CS based on their own set of input values.

Fig. 15 Partial dependence plot for **a** curing days, **b** water-binder ratio, and **c** cement content using the XGB model

Declarations

References

Ahmad, W., Ahmad, A., Ostrowski, K. A., Aslam, F., Joyklad, P., & Zajdel, P. (2021). Application of advanced machine learning approaches to predict the compressive strength of concrete containing supplementary cementitious materials. *Materials, 14*(19), 5762. https://doi.org/10.3390/ma14195762

Alaka, H. A., & Oyedele, L. O. (2016). High volume fly ash concrete: The practical impact of using superabundant dose of high range water reducer. *Journal of Building Engineering, 8*, 81–90. https://doi.org/10.1016/j.jobe.2016.09.008

Albostami, A. S., Al-Hamd, R. K. S., Alzabeebee, S., Minto, A., & Keawsawasvong, S. (2023). Application of soft computing in predicting the compressive strength of self-compacted concrete containing recyclable aggregate. *Asian Journal of Civil Engineering, 25*(1), 183–196. https://doi.org/10.1007/s42107-023-00767-2

Al-Gburi, M., & Yusuf, S. A. (2022). Investigation of the effect of mineral additives on concrete strength using ANN. *Asian Journal of Civil Engineering, 23*, 405–414. https://doi.org/10.1007/s42107-022-00431-1

Allen, D. M. (1971). Mean square error of prediction as a criterion for selecting variables. *Technometrics, 13*(3), 469–475. https://doi.org/10.1080/00401706.1971.10488811s

Andrew, R. M. (2019). Global CO_2 emissions from cement production, 1928–2018. *Earth System Science Data, 11*(4), 1675–1710. https://doi.org/10.5281/zenodo.831454

Atis, C. D. (2003). High-volume fly ash concrete with high strength and low drying shrinkage. *Journal of Materials in Civil Engineering, 15*(2), 153–156. https://doi.org/10.1061/ASCE0899-156120 0315:2153

Balakrishnan, B., & Awal, A. S. M. A. (2014). Durability properties of concrete containing high volume Malaysian fly ash. *International Journal of Research in Engineering and Technology, 3*(4), 529–533. https://doi.org/10.15623/ijret.2014.0304093

Barbhuiya, S. A., Gbagbo, J. K., Russell, M. I., & Basheer, P. A. M. (2009). Properties of fly ash concrete modified with hydrated lime and silica fume. *Construction and Building Materials, 23*(10), 3233–3239. https://doi.org/10.1016/j.conbuildmat.2009.06.001

Bennett, K. P., Bi, J., & Edu, B. (2003). Regression error characteristic curves. *Proceedings, Twentieth International Conference on Machine Learning, 1*, 43–50.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST, 25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bildirici, M. E. (2019). Cement production, environmental pollution, and economic growth: Evidence from China and USA. *Clean Technologies and Environmental Policy, 21*(4), 783–793. https://doi.org/10.1007/s10098-019-01667-3

Chen, H. J., Shih, N. H., Wu, C. H., & Lin, S. K. (2019). Effects of the loss on ignition of fly ash on the properties of high-volume fly ash concrete. *Sustainability, 11*(9), 2704. https://doi.org/10.3390/su11092704

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chicco, D., Warrens, M. J., & 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*, e623. https://doi.org/10.7717/peerj-cs.623

Chindaprasirt, P., Chotithanorm, C., Cao, H. T., & Sirivivatnanon, V. (2007). Influence of fly ash fineness on the chloride penetration of concrete. *Construction and Building Materials, 21*(2), 356–361. https://doi.org/10.1016/j.conbuildmat.2005.08.010

Chopra, P., Sharma, R. K., & Kumar, M. (2016). Prediction of compressive strength of concrete using artificial neural network and genetic programming. *Advances in Materials Science and Engineering*. https://doi.org/10.1155/2016/7648467

Durán-Herrera, A., Juárez, C. A., Valdez, P., & Bentz, D. P. (2011). Evaluation of sustainable high-volume fly ash concretes. *Cement and Concrete Composites, 33*(1), 39–45. https://doi.org/10.1016/j.cemconcomp.2010.09.020

Felekoglu, B. (2006). Utilisation of Turkish fly ashes in cost effective HVFA concrete production. *Fuel, 85*(12–13), 1944–1949. https://doi.org/10.1016/j.fuel.2006.01.019

Golewski, G. L. (2018). Effect of curing time on the fracture toughness of fly ash concrete composites. *Composite Structures, 185*, 105–112. https://doi.org/10.1016/j.compstruct.2017.10.090

Hansen, T. C. (1990). Long-term strength of high fly ash concretes. *Cement and Concrete Research, 20*(2), 193–196. https://doi.org/10.1016/0008-8846(90)90071-5

Hashmi, A. F., Shariq, M., Baqi, A., & Haq, M. (2020). Optimization of fly ash concrete mix—A solution for sustainable development. *Materials Today: Proceedings, 26*, 3250–3256. https://doi.org/10.1016/j.matpr.2020.02.908

Huang, C. H., Lin, S. K., Chang, C. S., & Chen, H. J. (2013). Mix proportions and mechanical properties of concrete containing very high-volume of class F fly ash. *Construction and Building Materials, 46*, 71–78. https://doi.org/10.1016/j.conbuildmat.2013.04.016

Iranmanesh, A., & Kaveh, A. (1999). Structural optimization by gradient base neural networks. *International Journal of Numerical Methods in Engineering, 46*, 297–311.

Jiang, Y., Li, H., & Zhou, Y. (2022). Compressive strength prediction of fly ash concrete using machine learning techniques. *Buildings, 12*(5), 690. https://doi.org/10.3390/buildings12050690

Kaveh, A. (2024). *Applications of artificial neural networks and machine learning in civil engineering, Studies in Computational Intelligence 1168*. Springer.

Kaveh, A., Bahreininejad, A., & Mostafaie, M. (1999). A hybrid graph-neural method for domain decomposition. *Computers and Structures, 70*, 667–674. https://doi.org/10.1016/S0045-7949(98)00209-0

Kaveh, A., Dadras Eslamlou, A., Javadi, S. M., & Geran Malek, N. (2021). Machine learning regression approaches for predicting the ultimate buckling load of variable-stiffness composite cylinders. *Acta Mechanica, 232*(3), 921–931. https://doi.org/10.1007/s00707-020-02878-2

Kaveh, A., Eskandari, A., & Movasat, M. (2023). Buckling resistance prediction of high-strength steel columns using metaheuristic-trained artificial neural networks. *Structures, 56*, 104853. https://doi.org/10.1016/j.istruc.2023.07.043

Khan, G. M. (2018). Artificial neural network (ANNs). *Studies in Computational Intelligence, 725*, 39–55. https://doi.org/10.1007/978-3-319-67466-7_4

Kumar, B., Tike, G. K., & Nanda, P. K. (2007). Evaluation of properties of high-volume fly-ash concrete for pavements. *Journal of Materials in Civil Engineering*. https://doi.org/10.1061/ASCE0899-1561200719:10906

Kumar, M., Sinha, A. K., & Kujur, J. (2021). Mechanical and durability studies on high-volume fly-ash concrete. *Structural Concrete, 22*(1), 1036–1049. https://doi.org/10.1002/suco.202000020

Li, G., Zhou, C., Ahmad, W., Usanova, K. I., Karelina, M., Mohamed, A. M., & Khallaf, R. (2022). Fly ash application as supplementary cementitious material: A review. *Materials, 15*(7), 2664. https://doi.org/10.3390/ma15072664

Mahajan, L., & Bhagat, S. (2022). Machine learning approaches for predicting compressive strength of concrete with fly ash admixture. *Research on Engineering Structures and Materials*. https://doi.org/10.17515/resm2022.534ma0927

Manzoor, B., Othman, I., Durdyev, S., Ismail, S., & Wahab, M. H. (2021). Influence of artificial intelligence in civil engineering toward sustainable development—A systematic literature review. *Applied System Innovation, 4*(3), 52. https://doi.org/10.3390/asi4030052

McCarthy, M. J., & Dhir, R. K. (2005). Development of high volume fly ash cements for use in concrete construction. *Fuel, 84*(11), 1423–1432. https://doi.org/10.1016/j.fuel.2004.08.029

Mehta, P. K., & Gjorv, O. E. (1982). Properties of portland cement concrete containing fly ash and condensed silica-fume. *Cement and Concrete Research, 12*(5), 587–595. https://doi.org/10.1016/0008-8846(82)90019-9

Mukherjee, S., Mandal, S., & Adhikari, U. B. (2013). Comparative study on physical and mechanical properties of high slump and zero slump high volume fly ash concrete (HVFAC). *Global Nest Journal, 15*(4), 578–584. https://doi.org/10.30955/gnj.000801

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics, 18*(6), 275–285. https://doi.org/10.1002/cem.873

Nayak, D. K., Abhilash, P. P., Singh, R., & Kumar, V. (2022). Fly ash for sustainable construction: A review of fly ash concrete and its beneficial use case studies. *Cleaner Materials, 6*(100143), 100143. https://doi.org/10.1016/j.clema.2022.100143

Nochaiya, T., Wongkeo, W., & Chaipanich, A. (2010). Utilization of fly ash with silica fume and properties of portland cement-fly ash-silica fume concrete. *Fuel, 89*(3), 768–774. https://doi.org/10.1016/j.fuel.2009.10.003

Oner, A., Akyuz, S., & Yildiz, R. (2005). An experimental study on strength development of concrete containing fly ash and optimum usage of fly ash in concrete. *Cement and Concrete Research, 35*(6), 1165–1171. https://doi.org/10.1016/j.cemconres.2004.09.031

Pisner, D. A., & Schnyer, D. M. (2019). Support vector machine. *Machine Learning: Methods and Applications to Brain Disorders*. https://doi.org/10.1016/B978-0-12-815739-8.00006-7

Reiner, M., & Rens, K. (2006). *High-Volume Fly Ash Concrete: Analysis and Application*. https://doi.org/10.1061/ASCE1084-0680200611:158

Saha, A. K. (2018). Effect of class F fly ash on the durability properties of concrete. *Sustainable Environment Research, 28*(1), 25–31. https://doi.org/10.1016/j.serj.2017.09.001

Sathiparan, N. (2024). Prediction model for compressive strength of rice husk ash blended sandcrete blocks using a machine learning models. *Asian Journal of Civil Engineering*. https://doi.org/10.1007/s42107-024-01077-x

Sathiparan, N., Jeyananthan, P., & Subramaniam, D. N. (2023). Silica fume as a supplementary cementitious material in pervious concrete: Prediction of compressive strength through a machine learning approach. *Asian Journal of Civil Engineering, 25*(3), 1–15. https://doi.org/10.1007/s42107-023-00956-z

Scrivener, K. L., John, V. M., & Gartner, E. M. (2018). Eco-efficient cements: Potential economically viable solutions for a low-CO_2 cement-based materials industry. *Cement and Concrete Research, 114*, 2–26. https://doi.org/10.1016/j.cemconres.2018.03.015

Shaikh, F. U. A., & Supit, S. W. M. (2015). Compressive strength and durability properties of high volume fly ash (HVFA) concretes containing ultrafine fly ash (UFFA). *Construction and Building Materials, 82*, 192–205. https://doi.org/10.1016/j.conbuildmat.2015.02.068

Siddique, R. (2004). Performance characteristics of high-volume class F fly ash concrete. *Cement and Concrete Research, 34*(3), 487–493. https://doi.org/10.1016/j.cemconres.2003.09.002

Siddique, R., & Khatib, J. M. (2010). Abrasion resistance and mechanical properties of high-volume fly ash concrete. *Materials and Structures, 43*(5), 709–718. https://doi.org/10.1617/s11527-009-9523-x

Singh, S., Bano, S., Singh, V., Singh, A., Kumar, A., & Singh, S. N. (2023). An investigative inquiry into harnessing the capabilities of machine learning for the assessment of compressive strength in red mud-based concrete enriched with fly ash as a viable road construction constituent. *Asian Journal of Civil Engineering, 25*(2), 1571–1585. https://doi.org/10.1007/s42107-023-00862-4

Su, X., Yan, X., & Tsai, C. L. (2012). Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics, 4*(3), 275–294. https://doi.org/10.1002/wics.1198

Suhendro, B. (2014). Toward green concrete for better sustainable environment. *Procedia Engineering, 95*, 305–320. https://doi.org/10.1016/j.proeng.2014.12.190

Suleiman, A. A., Abdullahi, U. A., & Ahmad, U. A. (2015). An analysis of residuals in multiple regressions. *International Journal of Advanced Science and Engineering, 3*(01), 2348–7550.

Sun, J., Shen, X., Tan, G., & Tanner, J. E. (2019). Compressive strength and hydration characteristics of high-volume fly ash concrete prepared from fly ash. *Journal of Thermal Analysis and Calorimetry, 136*(2), 565–580. https://doi.org/10.1007/s10973-018-7578-z

Supit, S. W. M., & Shaikh, F. U. A. (2015). Durability properties of high volume fly ash concrete containing nano-silica. *Materials and Structures, 48*(8), 2431–2445. https://doi.org/10.1617/s11527-014-0329-0

Tkaczewska, E. (2021). Effect of chemical composition and network of fly ash glass on the hydration process and properties of portland-fly ash cement. *Journal of Materials Engineering and Performance, 30*(12), 9262–9282. https://doi.org/10.1007/s11665-021-06129-w

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research, 30*(1), 79–82. http://www.jstor.org/stable/24869236

Woyciechowski, P., Woliński, P., & Adamczewski, G. (2019). Prediction of carbonation progress in concrete containing calcareous fly ash co-binder. *Materials, 12*(17), 2665. https://doi.org/10.3390/ma12172665

Yazici, C., Hasan, H., & Arel, H. (2012). Effects of fly ash fineness on the mechanical properties of concrete. *Sadhana*. https://doi.org/10.1007/s12046-012-0083-3