BILINGUAL PUBLISHING GROUP

Journal of Architectural Environment & Structural Engineering Research

https://journals.bilpubgroup.com/index.php/jaeser

ARTICLE

Comparative Evaluation of Traditional and Deep Learning Based Machine Learning Models for Concrete Compressive Strength Prediction

Atif Khan [1*] **, *Afsar Ali*** [1] **, *Syed Saqib Mehboob*** [2]

[1] *Department of Civil and Environmental Engineering, University of Louisiana at Lafayette, Lafayette, LA 70504, USA*

[2] *Department of Civil Engineering, University of Engineering and Technology, Taxila 47050, Pakistan*

ABSTRACT

Machine learning is widely used to predict concrete compressive strength because it captures nonlinear interactions among binders, aggregates, water content, and chemical admixtures. Tree-based ensemble models such as Random Forest and XGBoost often achieve high numerical accuracy; however, their discrete decision splitting mechanisms inherently produce stepwise response trends that may not reflect the smooth and continuous nature of material behavior. To overcome this limitation, this study proposes a hybrid CNN1D–BiLSTM model that integrates one dimensional convolutional neural networks for localized feature extraction with bidirectional long short term memory units to learn long range compositional dependencies and gradual transitions in mixture design. Linear Regression, Random Forest, XGBoost, and the proposed model were evaluated using 1,030 concrete mixtures from the Yeh dataset. Model performance was assessed using the coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Nash Sutcliffe Efficiency (NSE), Mean Bias Error (MBE), and Percent Deviation Degree (PDD) to provide a comprehensive evaluation of predictive accuracy and systematic bias. The CNN1D–BiLSTM achieved the best testing performance with R^2 and NSE of 0.9324 and RMSE of 4.17 MPa, while exhibiting minimal systematic bias as indicated by low MBE and PDD values. In addition to strong numerical performance, the proposed model generated smooth partial dependence trends consistent with hydration kinetics and continuous strength development. Shapley Additive

*CORRESPONDING AUTHOR:

Atif Khan, Department of Civil and Environmental Engineering, University of Louisiana at Lafayette, Lafayette, LA 70504, USA;
Email: atifkhan076@outlook.com

ARTICLE INFO

Received: 6 November 2025 | Revised: 25 December 2025 | Accepted: 2 January 2026 | Published Online: 8 January 2026
DOI: https://doi.org/10.30564/jaeser.v9i1.13199

CITATION

Khan, A., Ali, A., Mehboob, S.S., 2026. Comparative Evaluation of Traditional and Deep Learning Based Machine Learning Models for Concrete Compressive Strength Prediction. Journal of Architectural Environment & Structural Engineering Research. 9(1): 1–28.
DOI: https://doi.org/10.30564/jaeser.v9i1.13199

COPYRIGHT

Copyright © 2026 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (https://creativecommons.org/licenses/by-nc/4.0/).

Explanations (SHAP) analysis confirmed stable and physically meaningful contributions of key mixture components, demonstrating that the hybrid framework maintains high predictive capability while preserving realistic response behavior across varying mixture compositions.

Keywords: Machine Learning; Concrete Compressive Strength; XGBoost; Sensitivity Analysis; Sustainable Construction; Data-Driven Prediction

1. Introduction

Modern infrastructure is based on concrete, and it is the most commonly manufactured construction material in the world, with the demand projected to surpass eighteen billion tons per year by 2050[1]. Such massive production poses a huge environmental impact[2]. The production of standard Portland cement as the main binding substance in the traditional concrete involves the use of high energy volumes and leads to nearly eight percent of the total anthropogenic carbon dioxide emissions worldwide[3]. Meanwhile, the high rate of natural resources depletion through sand and aggregates adds to the environmental pressure[4,5]. All these difficulties demonstrate the necessity of more sustainable and resource efficient construction methods and the importance of creating more sophisticated tools that can aid in the design of high performance and environmentally conscious concrete mixtures[6].

As a reaction to these developments the construction industry is increasingly looking at the adoption of supplementary cementitious materials (SCM) as a viable approach of minimizing the clinker content of concrete and its environmental consequences. Industrial by-products such as fly ash and ground granulated blast-furnace slag are still popular, although agricultural by-products, including rice husk ash and sugarcane bagasse ash (SCBA), and natural fibers, including date palm fibers, have become increasingly popular[7,8]. Their application, along with the minimization of the reliance on energy-consuming ordinary Portland cement, facilitates the concept of the circular-economy by adding value to the waste streams that will otherwise be used to further the extent of the landfill[9,10]. Nevertheless, the chemical and physical variability of these substitute binders is natural and contributes to the further complexity of the mixture design process, since the accurate prediction of the significant mechanical properties, in particular, compressive strength becomes much harder. This nonlinearity and continuity can cause changes in material behavior that are nonlinear and continuous and cannot be represented in conventional empirical or linear approaches, particularly when the proportions of mixtures are highly variable.

Conventional predictive techniques like empirical equations and linear regression have been extensively applied to predict compressive strength, yet the assumptions tend to simplify the behavior of the material. These methods are based on rigid mathematical correlations that are unable to describe nonlinear interactions between binder components, aggregates, water content, and curing conditions[11]. Consequently, their predictions often fail to match the experimental values when the compositions of mixtures become longer than the calibrated ranges. The initial uses of machine learning such as the groundbreaking work of Yeh with artificial neural networks showed that data-driven models could address these drawbacks by learning complicated patterns directly using experimental information[12].

Machine learning is a useful technique to predict concrete strength with the development of computational procedures because it can consider nonlinear relationships that are not possible to model with conventional models[13]. Support vector machines, random forest, gradient boosting, and XGBoost are the techniques that have become popular due to their ability to offer high predictive accuracy in a wide range of mixture compositions[14]. These models are able to process large data sets, discover complicated interactions and provide better generalization than the traditional regression-based methods[15]. Recent works have further added explainable artificial intelligence tools like SHAP analysis to gain a better insight into the effects of individual variables on model predictions and improve transparency in data driven decision making[16–18].

Although these developments have been made, a number of studies that have been done are mainly aimed at enhancing numerical accuracy as opposed to the evaluation of whether the trends that have been predicted are in agreement

with the actual material behavior. Tree based ensemble models offer good statistical results, but due to the use of discrete decision splits, they result in stepwise and discontinuous predictions that are not reflective of the smooth and gradual development of concrete strength. It is especially clear when considering the case of supplementary cementitious materials and admixtures, the effects of which are formed gradually during hydration and microstructural refinement[19,20]. Due to the ever-growing use of sustainable binders and complex mixtures design, there is an increasing demand of predictive models which are capable of capturing the high dimensional nonlinear interactions and the continuous physical transitions that control the development of strength.

Deep learning is an encouraging avenue to solve these issues since its hierarchical nature allows the derivation of low level and high-level features of complex data sets[21]. Convolutional neural networks can discover local interactions and patterns between mixture components, whereas recurrent networks like long short-term memory networks can learn long range interactions which are reminiscent of the progressive nature of hydration and strength gain[22]. These techniques can be used together to provide the capability of modeling continuous nonlinear behavior more accurately than traditional or tree-based techniques. Nevertheless, there is little use of hybrid deep learning models in concrete strength prediction, and more studies are required to determine their capability to provide high accuracy and physically realistic predictions.

This paper introduces a hybrid version of deep learning model to enhance the predictive power of data driven models by combining one-dimensional convolutional neural network with long-short term memory units. The joint architecture is meant to learn local feature relationships by convolution and at the same time learn sequential dependencies, gradual transitions by using bidirectional memory cells. This model structure allows smooth and continuous predictions to be produced that are closer to actual concrete behavior. Through these complementary advantages, the proposed framework will address the shortcomings of tree-based models and develop a more resilient framework to predict compressive strength over a broad mixture composition space. The novelty of this study lies in three key aspects. First, unlike prior studies that primarily emphasize numerical accuracy, this work systematically evaluates the physical consistency of model response behavior using partial dependence analysis to examine whether predicted trends align with established hydration and mixture interaction mechanisms. Second, a hybrid CNN–BiLSTM architecture is implemented specifically to capture both local compositional interactions and long-range dependency patterns in mixture design, which remains limited in existing concrete strength prediction literature. Third, this study provides a controlled comparative framework in which linear regression, ensemble tree-based models, and deep learning architectures are evaluated under identical preprocessing and training conditions to isolate differences in response continuity and interpretability rather than accuracy alone.

1.1. Research Significance

The challenge of accurately predicting the compressive strength of concrete is very crucial since the mixtures of concrete applied in real-life situations are nonlinear due to the interactions of cementitious materials with the aggregates, water content, and chemical admixtures[23]. Despite the high predictive accuracy of tree-based ensemble models (Random Forest, XGBoost) which have been broadly used, they can frequently fail to reproduce smooth and physically interpretable response patterns when changing the material proportions. Their progressive or discontinuous predictive action is inconsistent with the continuous characteristic of the strength development in concrete, therefore, restricting their applicability to field uses in the context of mixture optimization or inverse design. This reaffirms the requirement of a modeling framework that not only provides high quality of predictive accuracy, but also the physical behavior of materials in the real sense without providing unrealistic discontinuities.

To fill this gap, the current research presents a hybrid deep learning architecture which combines one-dimensional convolutional neural networks with bidirectional long short-term memory units. The 1D-CNN component is important in that it is able to derive multiscale patterns in the input mixture features whereas the BiLSTM layer is able to learn long-range dependencies and smooth transitions that are indicative of material interactions. This design is meant to integrate the nonlinear and interdependent effects on compressive strength of age, cement, water, aggregates, mineral admixtures, and superplasticizer dosage, which are continuous and nonlinear.

Moreover, the paper leads to the interpretability of the model, as it performs SHAP and partial dependence analyses that explain the mechanistic impact of input features and the reason behind the deep learning models providing more realistic trends than ensemble tree-based approaches.

1.2. Research Objectives

The main goal of this study is to develop a robust and physically meaningful predictive framework for estimating the compressive strength of concrete by integrating traditional machine learning techniques with an advanced deep learning architecture (CNN1D–BiLSTM) and by evaluating the consistency, interpretability, and practical applicability of these models for real-world mixture design. The specific objectives of the study are as follows:

1. To develop and train a hybrid deep learning model (CNN1D–BiLSTM) capable of accurately predicting compressive strength using mixture proportions and curing age while maintaining smooth and realistic response behavior.
2. To benchmark the proposed deep learning model against linear and ensemble tree-based models by comparing their predictive accuracy and generalization performance using R^2, RMSE, MAE, and train–test behavior.
3. To evaluate the influence of each input variable using SHAP analysis and determine the dominant contributors to strength development while assessing consistency with established materials science principles.
4. To analyze the behavior of different models using partial dependence plots (PDPs) in order to assess how well each model captures physical and continuous relationships between mixture components and strength.
5. To assess the practical suitability of each modeling approach for real-world mixture design, highlighting why ensemble tree-based models, despite strong accuracy metrics, may not be reliable for field use due to their discontinuous and nonphysical prediction patterns, and why deep learning provides a more robust alternative.

2. Methodology

2.1. Data Collection

The success of the creation of strong and generalized machine learning models is contingent on the presence of high quality and extensive databases. The data used in this research was gathered by Yeh[12] and comprised 1,030 data points collected from seventeen different sources. It had eight input variables, that is, Cement (C), Blast furnace slag (BFS), fly ash (FA), water (W), superplasticizer (SP), coarse aggregate (CA), fine aggregate (FAg) and age (A). The compressive strength of concrete was taken as the output variable.

Table 1 provides a list of 1,030 concrete mix samples with a diverse range of material proportions and curing ages. The cement content was significantly varied between 102.0 and 540.0 kg/m^3 with the mean of 281.17 kg/m^3. The dispersions of blast furnace slag and fly ash were high, which means that the supplementary cementitious material was used differently in the mixes. The water content was 121.8–247.0 kg/m^3 with a standard deviation that was relatively low, indicating more controlled batching conditions than other constituents. The amount of superplasticizer was also found to differ significantly ranging between zero and 32.2 kg/m^3, proving mixes with and without chemical admixtures.

Table 1. Statistical summary of dataset variables used for predicting concrete compressive strength.

Variable	Unit	Count	Mean	SD	Min	25%	50%	75%	Max
Cement	kg/m^3	1,030	281.17	104.51	102.0	192.38	272.0	350.0	540.0
Blast Furnace Slag	kg/m^3	1,030	73.90	86.88	0.0	0.0	22.0	118.3	359.4
Fly Ash	kg/m^3	1,030	54.18	63.99	0.0	0.0	0.0	118.3	200.1
Water	kg/m^3	1,030	181.57	21.35	121.8	164.9	185.0	192.0	247.0
Superplasticizer	kg/m^3	1,030	6.20	5.97	0.0	0.0	6.40	10.2	32.2
Coarse Aggregate	kg/m^3	1,030	972.92	77.75	801.0	932.0	968.0	1,029.4	1,145.0
Fine Aggregate	kg/m^3	1,030	773.58	80.18	594.0	739.96	779.5	829.0	992.6
Age	days	1,030	45.66	63.17	1.0	7.0	28.0	56.0	365.0
Compressive Strength	MPa	1,030	35.82	16.71	2.33	23.71	34.45	46.14	82.60

The coarse and fine aggregates were distributed widely with coarse aggregate varying between 801.0 and 1,145.0 kg/m^3 and fine aggregate between 594.0 and 992.6 kg/m^3, which are the varying gradation and packing characteristics. The age factor was used between 1 and 365 days, and this gave an extensive coverage of early, intermediate and long-term curing. The compressive strength was between 2.33 and 82.60 megapascal which covers low and high strength concrete mixes within the same database.

Figure 1 presents the histograms and density plots of the input variables and compressive strength, showing the distributional characteristics of the dataset. The plots indicate wide variation in cement, blast furnace slag, fly ash, superplasticizer, aggregate contents, curing age, and compressive strength, confirming that the dataset covers a broad range of concrete mixture conditions.

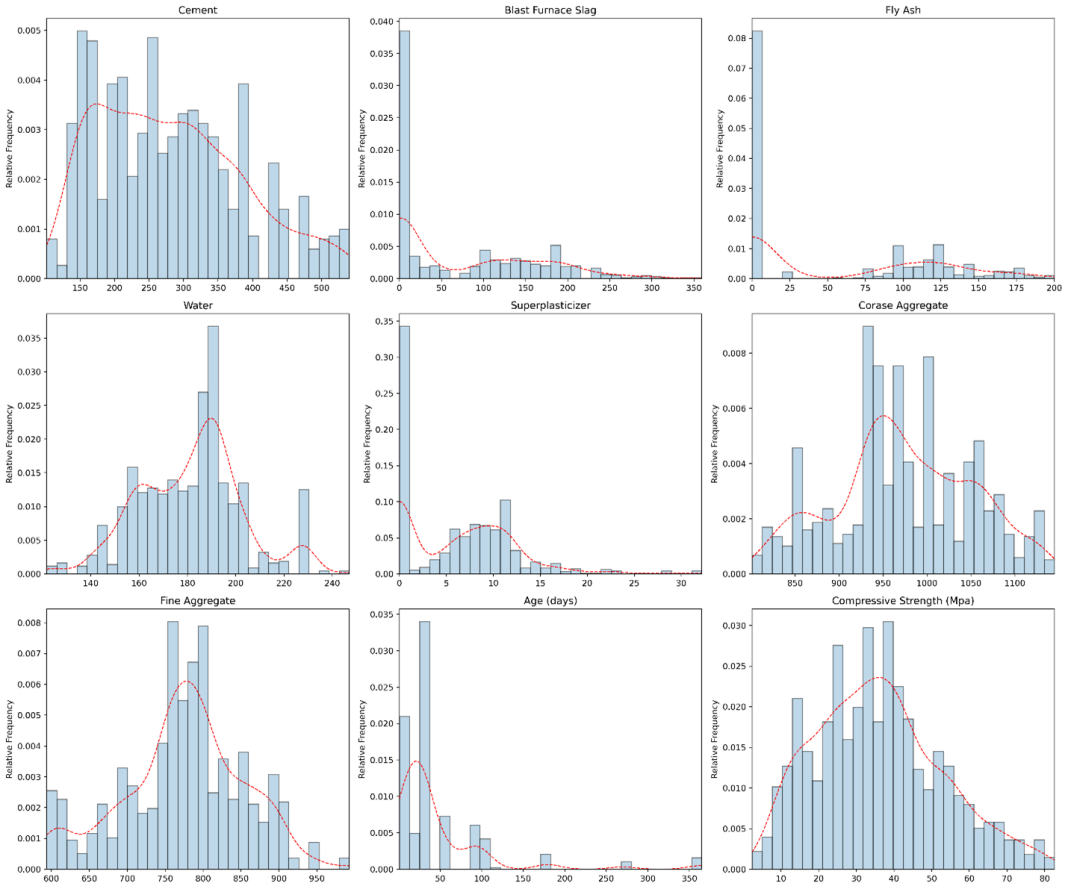


Figure 1. Histograms and Density Plots Illustrating the Distribution of Input Variables and Compressive Strength.

The heatmap of the correlation in **Figure 2** shows the linear relationship between the constituents of the mixture and the compressive strength obtained. The compressive strength is found to have the most positive relationship with cement content, and hence, increased binder content is always related to increased strength development [24]. The age also demonstrates a positive correlation that is noticeable indicating the expected strength increase with the ongoing hydration [25]. Superplasticizer has an average positive relationship, probably because of better dispersion of particles and lower water requirement, which increases the density of the matrix [26]. Conversely, the compressive strength is negatively correlated with water content, which aligns with the established increase of water-to-binder ratio and its effect on pore structure and decreased integrity of the matrix [27]. Fine aggregate has a weak negative correlation with coarse aggregate and fly ash has very small correlations, implying that the effect of these components is nonlinear or may have interactions that are not reflected by Pearson measures. On the whole, the heatmap proves the absence of excessively high levels of correlation between two input variables, which evidences the appropriateness of all of the features to develop a machine learning model.

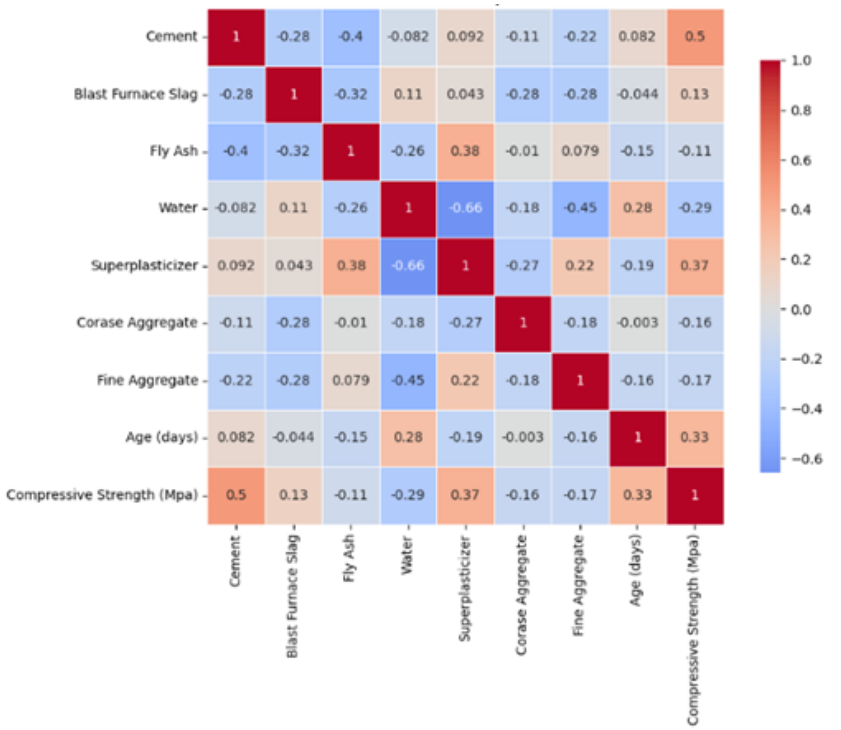


Figure 2. Pearson correlation heatmap showing linear relationships among mix parameters and compressive strength of concrete.

Figure 3 presents the feature importance ranking obtained from the Random Forest model, highlighting the relative contribution of each input variable to the prediction of compressive strength. Age and cement content emerge as the two most influential parameters, together accounting for more than half of the total importance. This reflects the central role of hydration time and binder quantity in strength development. Water content follows as the next most important variable, consistent with its effect on porosity and the water to binder ratio. Blast furnace slag and superplasticizer show moderate importance, indicating their secondary but meaningful influence on hydration reactions and workability. Fine aggregate, coarse aggregate, and fly ash exhibit the lowest importance scores, suggesting that their effects are either marginal or captured indirectly through interactions with more dominant variables. Overall, the ranking provided in **Figure 3** demonstrates that strength prediction is governed primarily by binder related parameters and curing age, while aggregate proportions and supplementary materials play comparatively smaller roles.

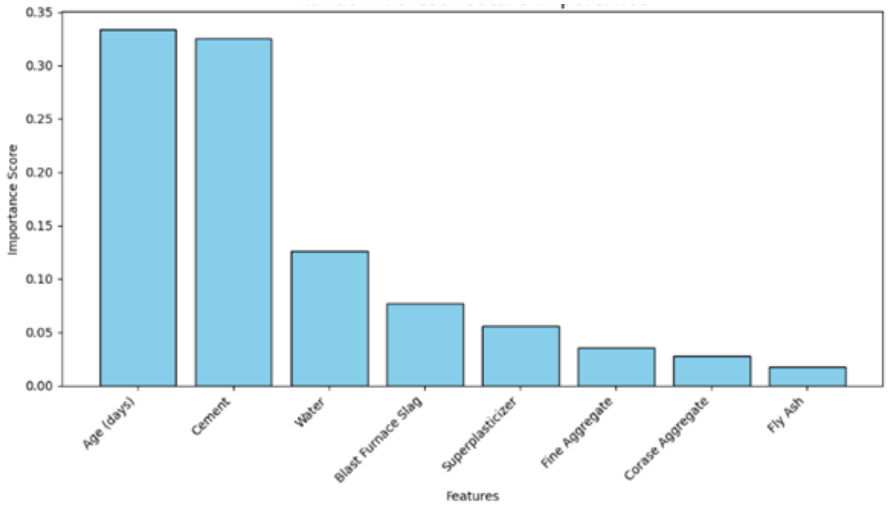


Figure 3. Feature importance ranking derived from the Random Forest algorithm.

3. Ensemble Tree-Based Models

3.1. Random Forest (RF)

Random Forest is an ensemble learning algorithm that integrates a great number of decision trees to enhance the accuracy of prediction and the stability of the model [28,29]. In the forest, each tree is trained on a bootstrap sample which is formed by randomly selecting the observations of the original data with replacement [30]. This sampling process will make sure that every tree is assigned slightly different portion of the data which will encourage diversity in the ensemble. The algorithm also samples a subset of the input variables at each split besides sampling instances, which enables the trees to learn about different combinations of features [31]. These two types of randomness aid in minimizing the variance and ensure that individual trees do not overfit the training data.

The overall prediction is also achieved by averaging the results of the individual trees after all the trees are constructed, which removes noise and enhances generalization [32]. Each bootstrap sample contains around two thirds of the data and the other one third, called out of bag observations are used to internally estimate the model performance without having to use a distinct validation set. This inbuilt evaluation offers an effective evaluation of prediction error and it adds strength to the method. Random Forest will be very appropriate modeling concrete strength since it is capable of nonlinear relationships, complex interactions among variables and is also stable when the dataset has moderate noise or redundant variables [33]. The conceptual structure of the Random Forest model is illustrated in **Figure 4**, where multiple decision trees generate individual predictions that are averaged to produce the final output.

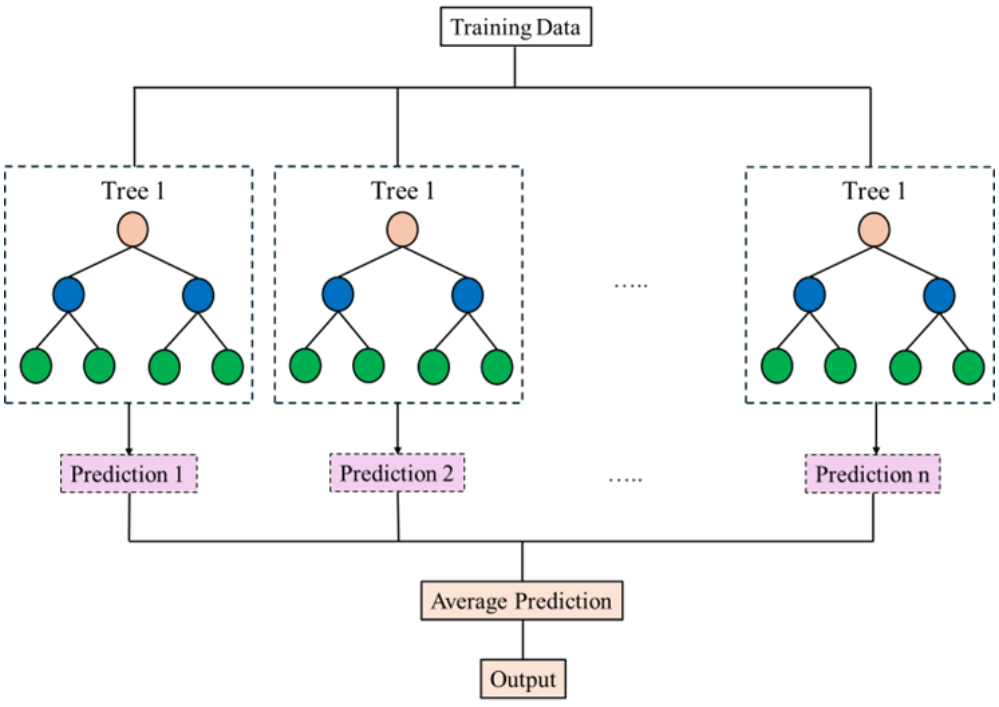


Figure 4. Random Forest Ensemble Structure Showing Individual Tree Predictions and Averaged Output.

3.2. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting, commonly known as XGBoost, is a decision tree-based ensemble method that constructs a sequence of regression trees to progressively minimize prediction errors [34,35]. The algorithm follows the principle of gradient boosting, where each new tree is trained to correct the residuals produced by the previous ensemble. This process creates an additive learning model in which the final prediction represents the cumulative contribution of all fitted trees [36].

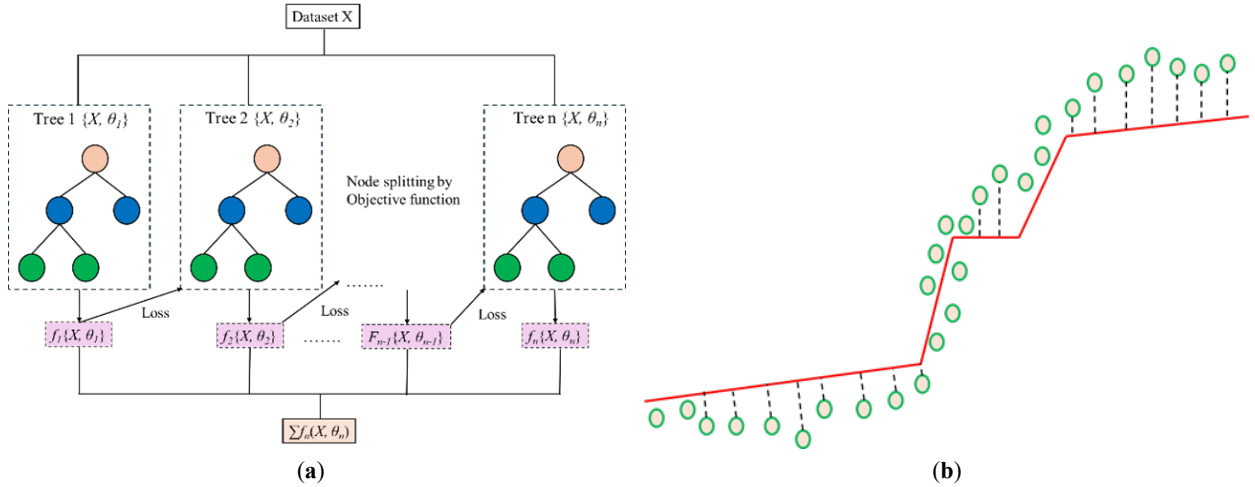
Let y_i^n denote the prediction for sample i after the k-th iteration. XGBoost updates the model by adding a new regression tree $f_n(x_i)$ that minimizes the loss generated by

earlier trees, expressed as:

$$y_i^n = \sum_{j=1}^{n-1} f_j(x_i) + f_n(x_i)$$

The construction of each tree relies on a second order approximation of the objective function, which incorporates both the gradient and the Hessian of the loss [36]. This allows the algorithm to evaluate the quality of each candidate split more accurately and to identify the split that yields the greatest reduction in loss. Regularization terms are also included during optimization to control model complexity, stabilize the training process, and prevent overfitting [37,38].

Figure 5a conceptually illustrates this iterative learning process, where multiple trees are trained sequentially and combined to form a stronger predictive model. **Figure 5b** shows how the ensemble of piecewise constant functions approximates the nonlinear relationship between the input variables and the target strength. These visualizations demonstrate how XGBoost incrementally refines the prediction surface by integrating information from each newly added tree.



(**a**) (**b**)

Figure 5. (**a**) Sequential tree construction through gradient boosting and (**b**) resulting piecewise constant prediction surface.

3.3. Deep Learning Model

3.3.1. Convolutional Neural Network (CNN)

Convolutional Neural Networks are broadly applied deep learning architectures that can extract hierarchical features of sequential or spatial data [39]. The one-dimensional CNN architecture was used in this work, and convolution kernels move in the direction of the input sequence to find significant patterns in the mixture attributes. The convolution layer executes the operations of weighted filtering that convert the raw input into a collection of feature maps, each signifying a different learnt pattern [39]. The feature maps are then subjected to a nonlinear activation function, usually the rectified linear unit, which makes the feature maps nonlinear and allows the network to learn complex relations in the data [40].

After convolution, a pooling layer is used to downsample the dimensionality of feature maps and to keep the most important information [41]. This is often done with max pooling, which picks the best activation in each window, thus enhancing feature invariance and decreasing model complexity. As illustrated in **Figure 6**, the extracted convolutional features are then passed through pooling and dense layers to generate the final compressive strength prediction. Once the convolution and pooling processes have been done, the extracted features are flattened and input into fully connected layers, which learn high level interrelations between the input variables [33]. The last thick layer produces the anticipated compressive strength. The one-dimensional CNN architecture can be used to offer an efficient and effective model to capture nonlinear dependence in concrete mixture data by integrating local pattern extraction with deeper fully connected learning [42].

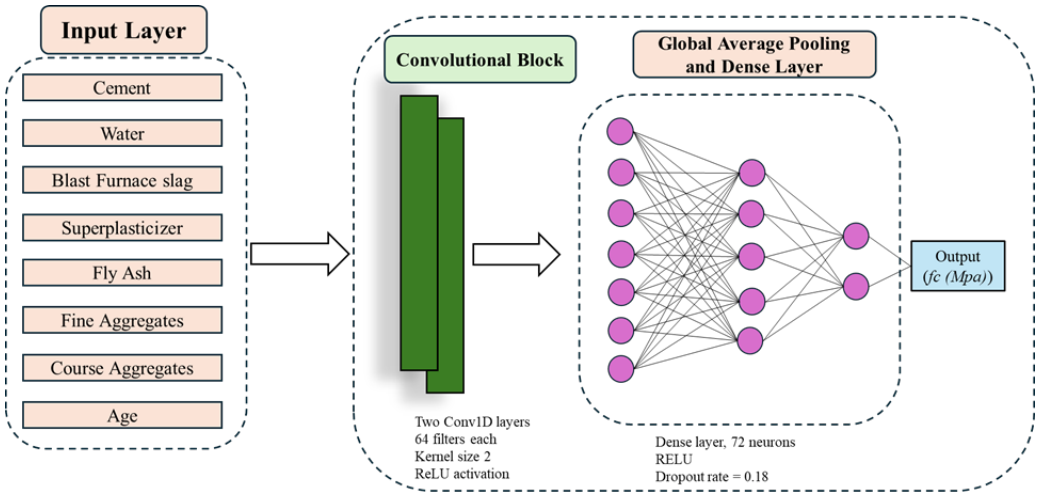


Figure 6. One-dimensional CNN architecture showing feature extraction via convolutional layers, global average pooling, and dense layers for compressive-strength prediction.

3.3.2. Bidirectional LSTM (BiLSTM)

Bidirectional Long Short Term Memory networks are based on the standard LSTM architecture except that the input sequence is processed in both forward and backward directions[43]. This design allows the model-to-model relationships that are dependent on the past values and also the patterns that are dependent on future values in the sequence. Each direction has an independent LSTM layer that learns temporal features by using a set of memory cells and gating mechanisms that regulate information flow[44]. These gates control the amount of the previous state that is retained or updated at any given time which enables the network to learn long range dependencies and overcome the vanishing gradient problem of traditional recurrent networks[45].

Under the bidirectional architecture, the forward and backward LSTM outputs are used to create a more complex representation of the input as shown in **Figure 7**[46]. This combined representation improves the learning capabilities of the model to nonlinear and time dependent relationship in the data. To predict the concrete compressive strength, the BiLSTM framework enables the model to take into account the previous and subsequent steps of the mixture properties, which is effective at capturing the hidden interactions between material proportions. The end result is obtained by a fully connected layer that projects the learned temporal characteristics to the expected values of strength, which allows the network to represent complex dependencies better than traditional LSTM or feed forward systems[47].

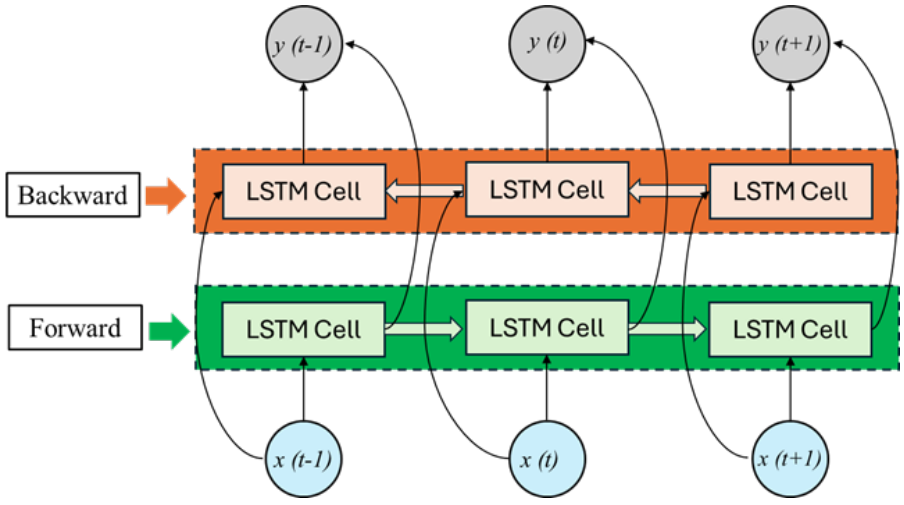


Figure 7. Bidirectional LSTM structure illustrating forward and backward sequence processing.

3.3.3. Ensemble 1D-CNN and BiLSTM

The proposed model starts with a one-dimensional convolutional neural network (1D-CNN), which takes the raw input features (e.g., mix proportions, age, constituents) as input and extracts local patterns and inter-feature interactions. The CNN consists of several convolutional filters with small kernel size (e.g., size = 2) and activation (e.g., ReLU) to produce feature maps, which are learned relationships between input variables. Stacking two successive convolutional layers enhances the network's capacity to learn higher order combinations of input parameters [48]. After convolution and possible pooling or flattening, the resulting feature vector is fed to a recurrent layer for temporal/context learning. The overall workflow of the proposed CNN1D–BiLSTM ensemble model is presented in **Figure 8**.

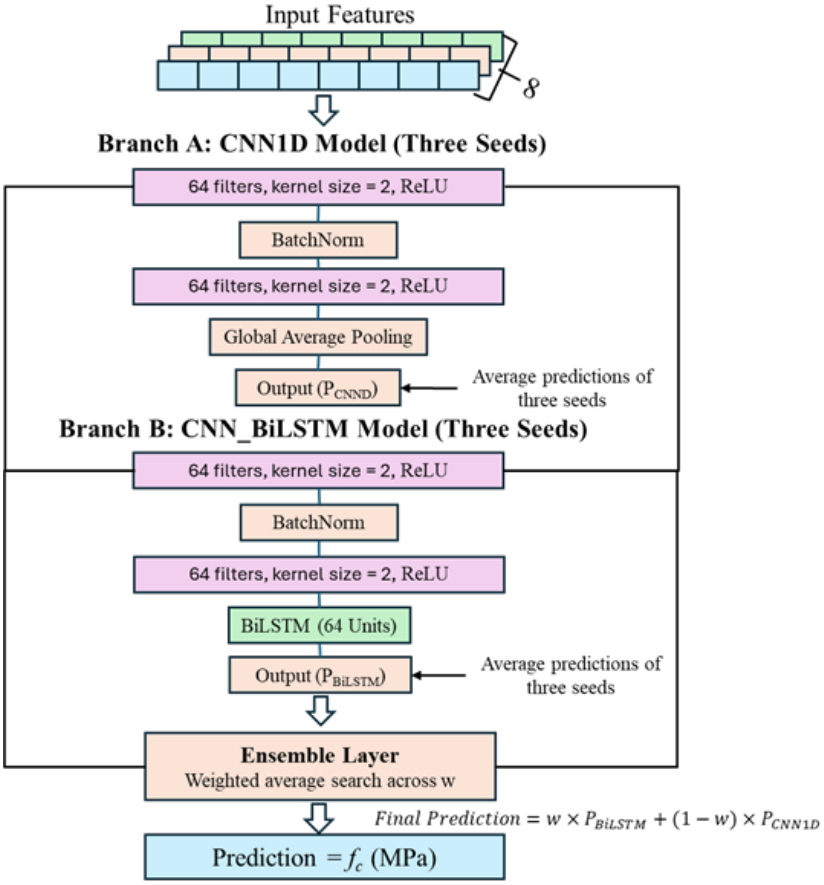


Figure 8. Workflow diagram of the CNN1D–BiLSTM Ensemble with Weighted Fusion Layer.

3.3.4. Model Evaluation

The performance of each model was assessed using multiple statistical indicators to comprehensively evaluate accuracy, efficiency, and systematic bias.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{(\sum (y_i - \hat{y}_i)^2}{n}} \quad (2)$$

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \quad (3)$$

$$MAPE = \frac{100}{n} \sum \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4)$$

In addition, the Nash Sutcliffe Efficiency coefficient was used to further evaluate predictive skill:

$$NSE = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (5)$$

To assess systematic prediction bias, the Mean Bias Error and Percent Deviation Degree were also calculated:

$$MBE = \frac{1}{n} \sum (\hat{y}_i - y_i) \quad (6)$$

$$\text{PDD} = 100 \times \frac{\sum (\widehat{y}_i - y_i)}{\sum \widehat{y}_i} \quad (7)$$

Where, y_i and $\widehat{y}_i$ denote the experimental and predicted compressive strengths, respectively, and n represents the total number of samples. The optimal model was determined based on the highest coefficient of determination (R^2) and the lowest error metrics (RMSE, MAE, and MAPE) for both training and testing datasets.

3.3.5. Hyperparameter Setting for Ensemble Tree Based Models

The hyperparameter tuning process for the ensemble tree-based models showed distinct differences in the response of each algorithm to the variation in model complexity. As shown in **Table 2**, the number of estimators was explored for a range of 300, 500 and 800 trees for both models. Random Forest showed its best performance with 800 estimators with a maximum depth of 15 which gave an R^2 score of 0.8805. XGBoost, on the other hand, achieved optimal performance when the number of estimators was 500 and the maximum depth was 5, showing its efficiency in learning complex nonlinear patterns with fewer trees. The dataset was randomly divided into training and testing subsets using an 80% and 20% split with a fixed random state of 42 to guarantee consistent replication of results. Out of these models, XGBoost achieved the highest predictive accuracy with an R^2 score of 0.9316, followed by RF with an R^2 score of 0.8816. These results contribute to the benefit of boosting based algorithms in capturing complex relationships in the dataset while still achieving good generalization performance.

Table 2. Hyperparameter search space and optimal configurations for ensemble tree based models.

Model	$n_{estimator}$	Optimum $n_{estimator}$	Max_{depth}	Optimum Max Depth
Random Forest	300, 500, 800	800	15, 25, None	15
XGBoost	300, 500, 800	500	5, 7, −1	5

3.3.6. Ensemble Configuration and Full Training Protocol

For ensemble model all input variables were standardized using StandardScaler, where scaling parameters were computed exclusively on the training subset and subsequently applied to the testing subset to prevent data leakage. The target variable, compressive strength, was retained in its original physical units of MPa.

Training was implemented in TensorFlow Keras. The optimizer used was Adam with a learning rate of 0.001 and mean squared error as the loss function. The maximum number of epochs was set to 400 with a batch size of 32. Early stopping was applied by monitoring validation loss with a patience of 30 epochs and restore best weights enabled. A ReduceLROnPlateau scheduler was used to adaptively decrease the learning rate with a factor of 0.5 and patience of 10 to improve convergence stability. The validation subset was obtained from the training data using a validation split of 0.20 while preserving the fixed random seed. To enhance stability and reduce stochastic variability, all random generators including Python, NumPy, and TensorFlow were initialized with fixed seeds, and training was repeated with multiple predefined seeds, with predictions averaged for the final output.

The ensemble prediction was computed as a weighted average of the component model outputs. The ensemble weight was determined exclusively using a validation-based grid search over values between 0.00 and 1.00 with an increment of 0.01. The weight that maximized validation R^2 and minimized validation RMSE on the training data was selected and fixed prior to final testing. The testing dataset was not used at any stage of hyperparameter tuning or weight selection. Final performance metrics including R^2, RMSE, MAE, and MAPE were calculated separately for training and testing sets to assess both goodness of fit and generalization performance.

4. Results and Discussion

4.1. Linear Regression

The Linear Regression model produces a broadly scattered distribution around the 1:1 reference line as shown in **Figure 9**. This indicates the linear formulation cannot fully reflect the nonlinear relationships among mixture constituents. Although the model achieves moderate R^2 values for both training (0.8131) and testing (0.7843), the relatively high RMSE values of 7.287 and 7.455 show that its predic-

tions deviate substantially from measured strengths. The dispersion becomes more evident at higher strengths, where linear assumptions fail to represent saturation behavior and interaction effects. This confirms that while Linear Regression establishes a simple baseline, it does not capture the inherent complexity of concrete mixtures.

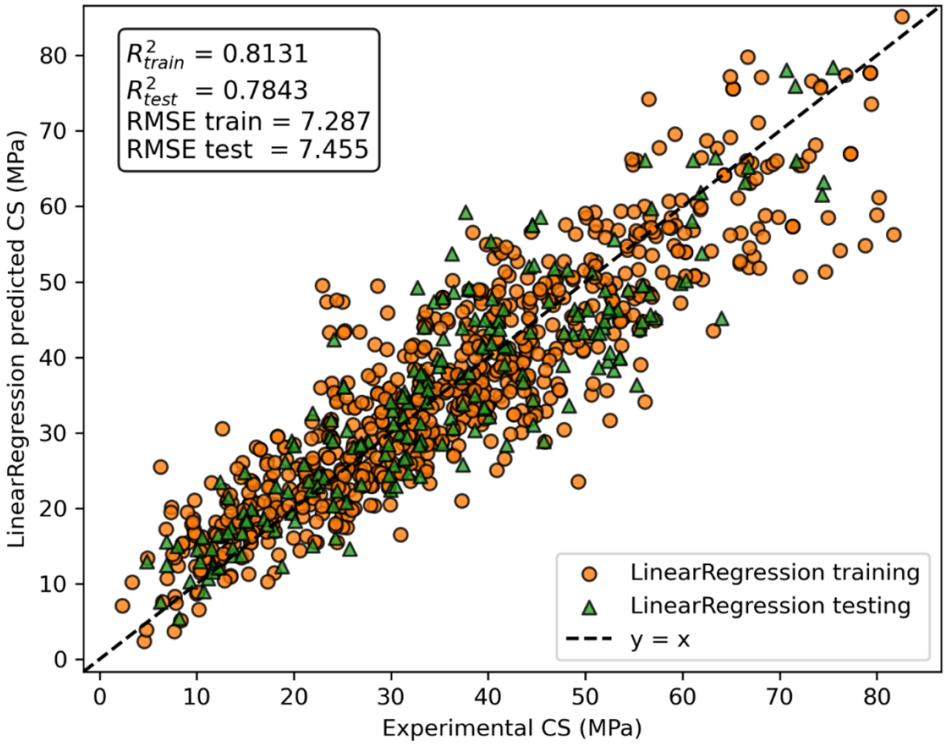


Figure 9. Predicted vs. experimental compressive strength using linear regression.

4.2. Random Forest

In **Figure 10**, the clustering of the points around the 45 degree line is more concentrated in the Random Forest model, which indicates the better way to identify nonlinear patterns and interaction patterns in the dataset. The testing R^2 of 0.8805 and RMSE of 5.548 are higher than the linear model. The scatter is less and more consistent in the range of strength, which proves the responsiveness of the model to the relationships between variables. Nevertheless, the predictions still exhibit small discontinuities and local deviations, which are in line with the stepwise character of tree based partitioning which models behaviour in discrete chunks instead of continuous changes.

4.3. XGBoost

The XGBoost results exhibit one of the closest alignments between predicted and experimental strengths. With a high testing R^2 of 0.9316 and a low RMSE of 4.198, the model accurately captures nonlinear trends across the full spectrum of strength values. The points trace the 1:1 line very closely, especially in mid to high strength regions in **Figure 11**. XGBoost's gradient boosting mechanism enables detailed fitting of interactions, though subtle bands remain visible due to the algorithm's localized decision splits. Despite these small artifacts, the overall pattern shows strong predictive performance and consistent agreement with experimental observations.

4.4. CNN–BiLSTM Ensemble Model

The CNN–BiLSTM ensemble provides highly consistent and smooth predictions across all strength ranges as shown in **Figure 12**. The testing R^2 of 0.9324 and RMSE of 4.173 demonstrate excellent predictive capability comparable to XGBoost. The predictions closely follow the reference line with minimal scatter, indicating that the model effectively captures both short range feature interactions and longer range developmental trends related to cement hydration and mixture

composition. The smoothness of the predicted behavior reflects the model's continuous learning structure, which aligns more closely with real material response than the segmented predictions typical of tree based methods.

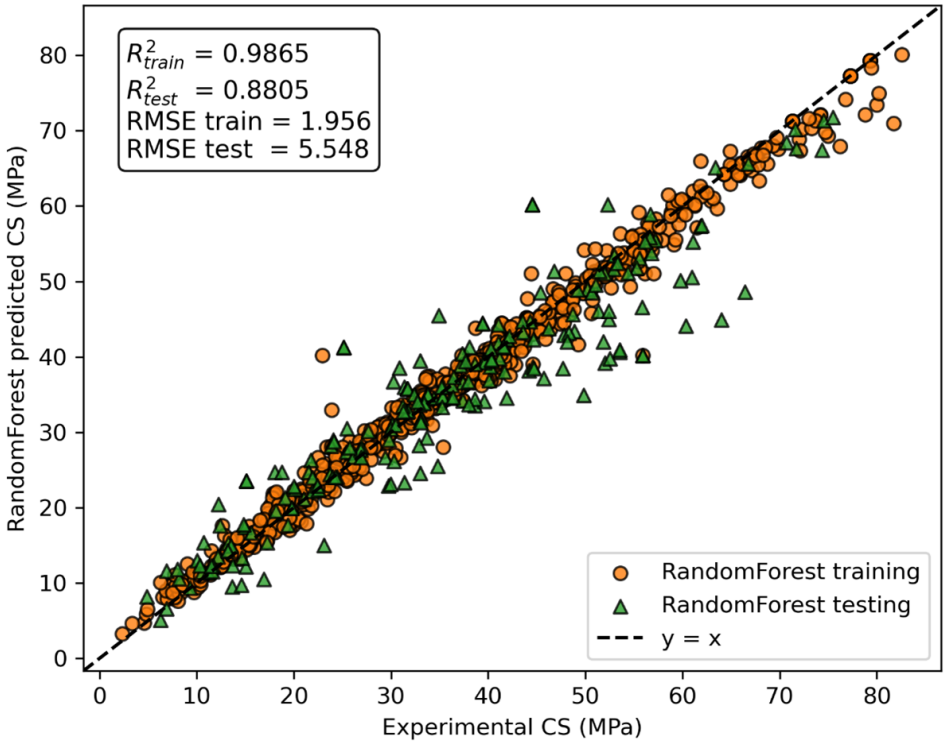


Figure 10. Predicted vs. experimental compressive strength using RF.

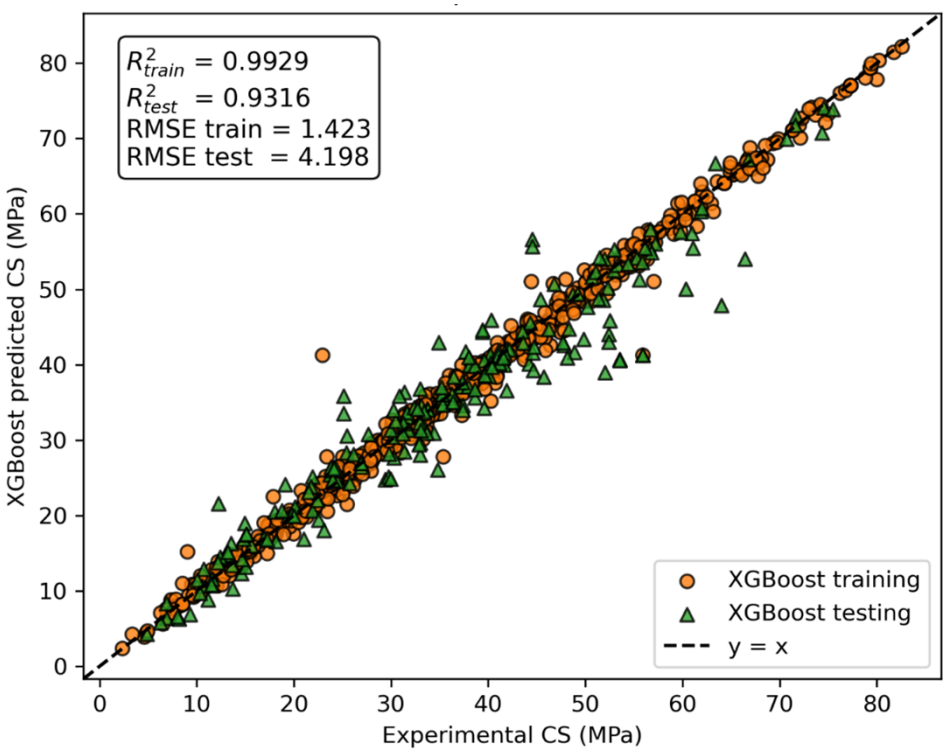


Figure 11. Predicted vs. experimental compressive strength using XGBoost.

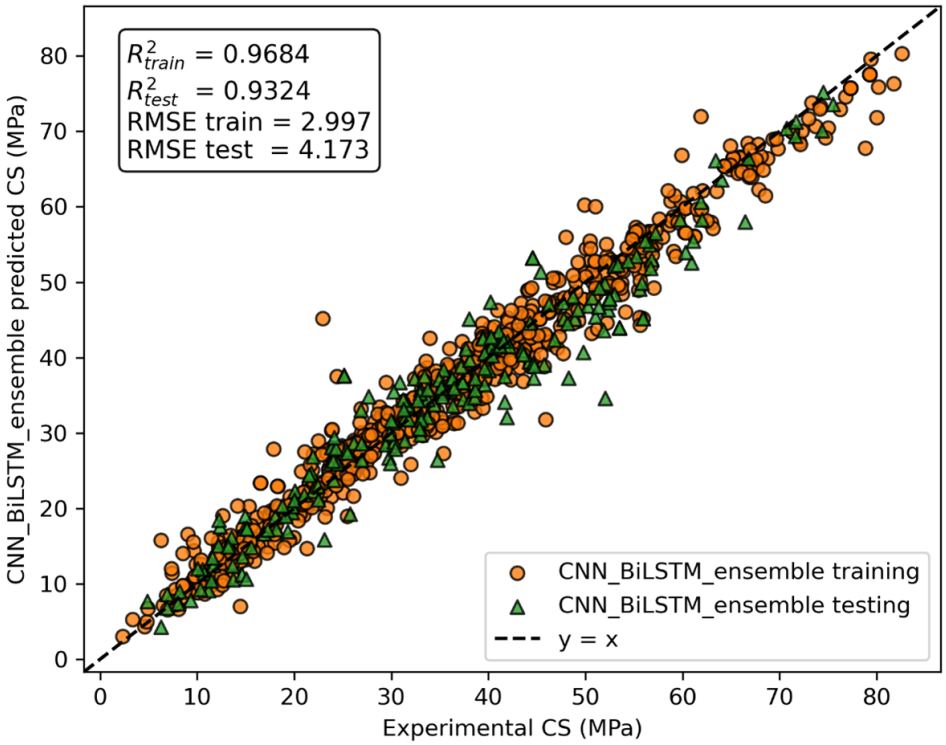


Figure 12. Predicted vs. experimental compressive strength using ensemble CNN–BiLSTM.

The loss curve for the CNN–BiLSTM ensemble in **Figure 13** shows rapid reduction in both training and testing losses within the initial epochs, followed by a gradual and stable decline toward convergence. The close alignment of the two curves indicates consistent learning behavior throughout training. The smooth shape of the loss trajectories reflects the ability of the ensemble framework to progressively refine feature relationships while maintaining stable predictive performance. This behavior confirms that the deep learning architecture efficiently learns from the high dimensional mixture data and achieves reliable convergence without abrupt fluctuations.

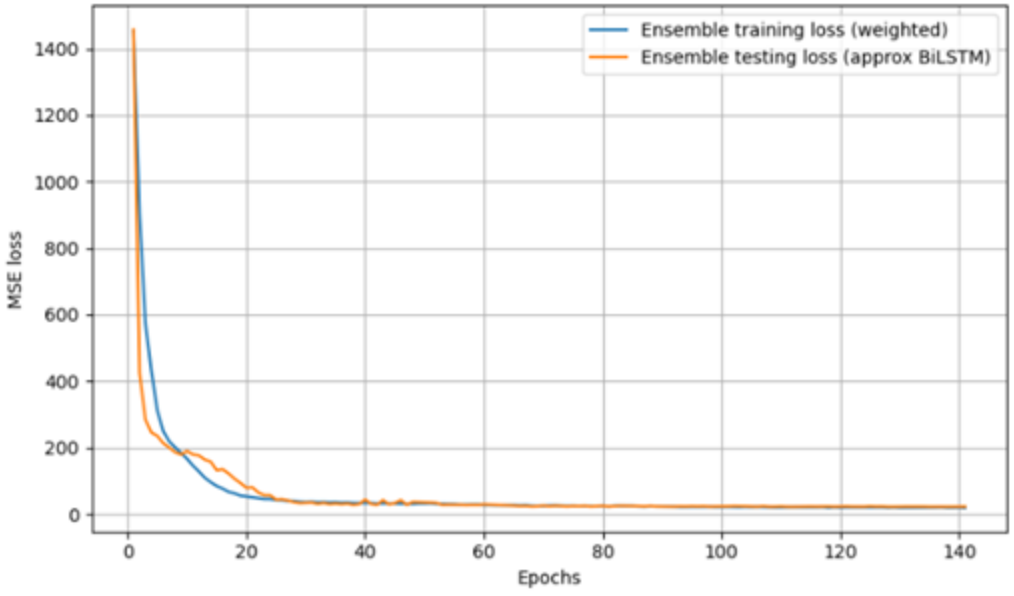


Figure 13. Training and Testing Loss Curves of the Ensemble CNN1D–BiLSTM Model.

4.5. Comparative Analysis of Model Performance

The relative analysis of the four modeling methods in **Table 3** shows that they have definite differences in terms of predictive strength and practical reliability. The Ensemble CNN1D–BiLSTM has the highest generalization capability as it achieves the highest testing accuracy with a test R^2 of 0.9324, an NSE of 0.9324, and a low testing RMSE of 4.17, which indicates strong agreement between predicted and observed values. The near identical values of R^2 and NSE further confirm the robustness of the ensemble model. In addition, the ensemble exhibits a small MBE of −0.61 MPa and a low PDD of −1.72%, indicating minimal systematic bias and only slight underprediction. Even though XGBoost also competes closely with a testing R^2 of 0.9316 and NSE of 0.9316, its RMSE of 4.20 is slightly higher, indicating marginally greater dispersion compared to the ensemble model. Its MBE of −0.64 MPa and PDD of −1.80% similarly indicate low but consistent underprediction. Random Forest, on the other hand, experiences a significant decrease in generalization performance, as testing R^2 and NSE decrease to 0.8816 and 0.8805, respectively, with a higher RMSE of 5.55. Moreover, its larger negative MBE of −0.88 MPa and PDD of −2.46% suggest comparatively greater systematic bias. Linear Regression has the poorest overall performance, with a testing R^2 of 0.7843 and the largest RMSE of 7.46, which confirms that it is not able to capture the nonlinear nature of relationships that determine concrete compressive strength. Although Linear Regression shows the smallest absolute MBE of −0.06 MPa and PDD of −0.18%, this reflects lower systematic bias rather than superior predictive accuracy, as its overall error magnitude remains significantly higher than the other models.

Table 3. Comparative accuracy and bias analysis of concrete compressive strength prediction models.

Model	Train R^2	Test R^2	$Train_{RMSE}$	$Test_{RMSE}$	NSE	PDD (%)	MBE (Mpa)
Ensemble CNN1D–BiLSTM	-	0.9324	3.00	4.17	0.9324	−1.72	−0.61
XGBoost	0.9929	0.9316	1.42	4.20	0.9316	−1.80	−0.64
Random Forest	0.9865	0.8816	1.96	5.55	0.8805	−2.46	−0.88
Linear Regression	0.8131	0.7843	7.29	7.46	0.7843	−0.18	−0.06

4.6. Partial Dependence Plots (PDP)

(a) Cement

The models view the contribution of cement content in a very different way. **Figure 14a** Linear Regression model gives a smooth monotonic increase which is similar to the general anticipation that an increase in cement contents leads to an increase in binder availability. The curve however becomes unrealistically linear at higher dosages where the gain of strength normally decreases as a result of saturation of the paste. **Figure 14b** of Random Forest shows a staircase-like sequence with sudden leaps on the way that are caused by tree splits and not due to any material behavior which restricts the capabilities of this method to reflect the diminishing returns in real mixtures. **Figure 14c** depicts that XGBoost exhibits an upward trend, however, it introduces regular local variations due to sequential boosting, which fits small areas of data and breaks continuity. By comparison, the CNN–BiLSTM ensemble in **Figure 14d** yields a smooth and increasingly flattening curve, which closely resembles the practical behaviour of cement hydration where the strength gain decreases as mixtures become saturated with binder. The ensemble thus gives the most realistic understanding of nonlinear effect of cement.

(b) Age

The PDPs of age depict high contrasts among the four modeling strategies. **Figure 15a** shows a smooth increase in the Linear Regression curve that reflects the elementary concept of strength development with the duration of hydration, but it assumes one mathematical expression and cannot be used to reflect early-age acceleration and maturity-stabilized maturity. Random Forest in **Figure 15b** is characterized by abrupt initial leaps and flat-topped later ages, caused by discrete tree partitions, not by any actual maturity behavior, and leads to discontinuous instead of continuous growth. **Figure 15c** shows similar discontinuities in XGBoost, with the overfitting of the early ages being more pronounced, as small data clusters form artificially imposed transitions. **Figure 15d** CNN–BiLSTM ensemble produces a fast rise in the early age with a slow leveling which is in line with the established cement hydration kinetics. This physically coherent tendency shows that the ensemble model is the best model to represent the hardening process of time.

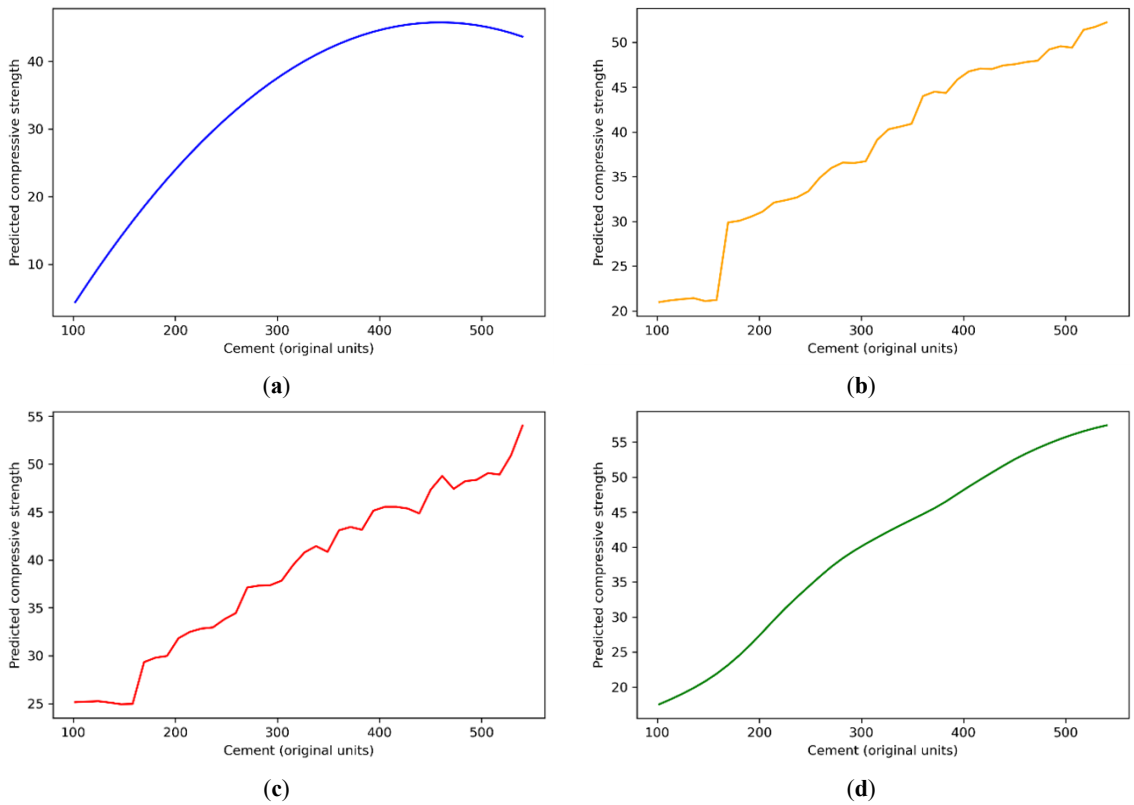


Figure 14. (**a**) PDP for cement using linear regression; (**b**) PDP for cement using RF; (**c**) PDP for cement using XGBoost; (**d**) PDP for cement using CNN–BiLSTM.

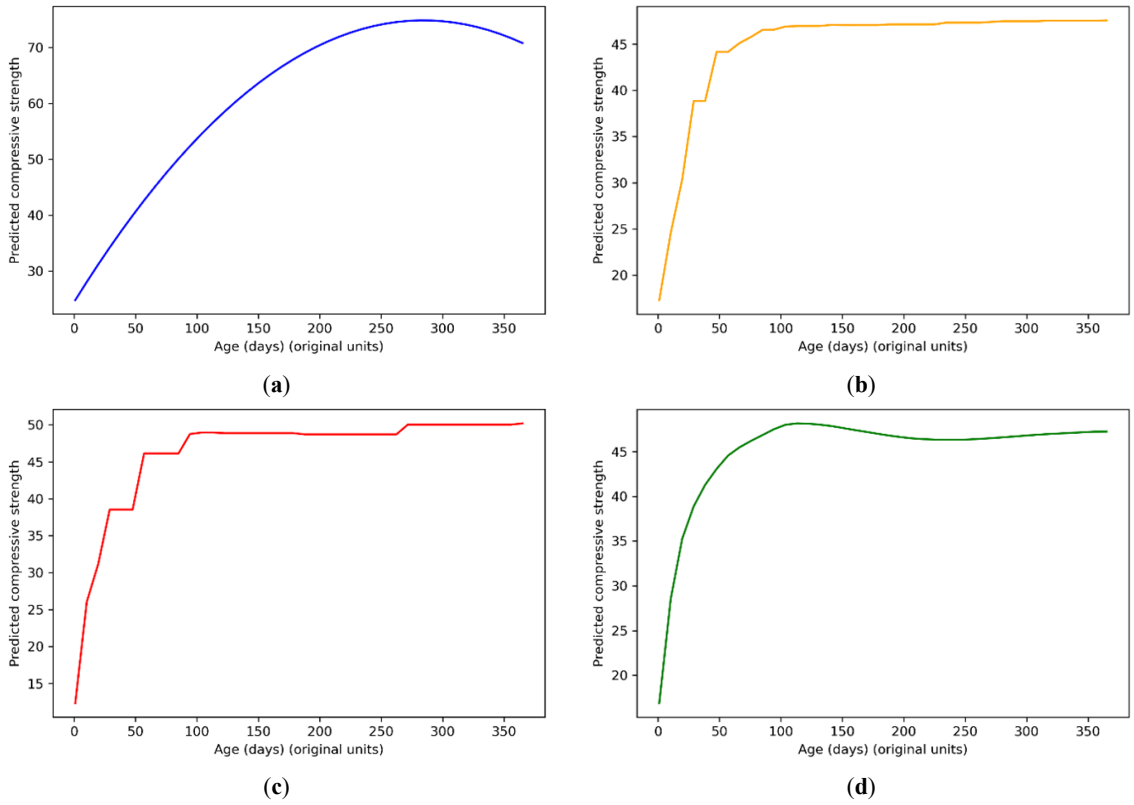


Figure 15. (**a**) PDP for age using linear regression; (**b**) PDP for age using RF; (**c**) PDP for age using XGBoost; (**d**) PDP for age using CNN–BiLSTM.

(c) Water

Water-related PDPs indicate the various capabilities of the models to model the effects of dilution and workability. **Figure 16a** using Linear Regression shows a fluent parabolic decrease, which in effect shows that the concrete water content ultimately weaken as it is over-simplified by the nonlinear sensitivity of water-to-cement ratios at optimum values. **Figure 16b** shows that in Random Forest, a downward trend is formed with a number of flat segments, which do not reflect any real hydration or dilution transition. **Figure 16c** shows that XGBoost predicts more monotonic decrease but introduces sudden drops caused by localized overfitting at negative correlations. The CNN–BiLSTM ensemble in **Figure 16d** reflects a physically consistent behavior, with an initial decrease, a stabilization period with workable mixtures that generate similar strengths, and a subsequent decrease at excessive water contents.

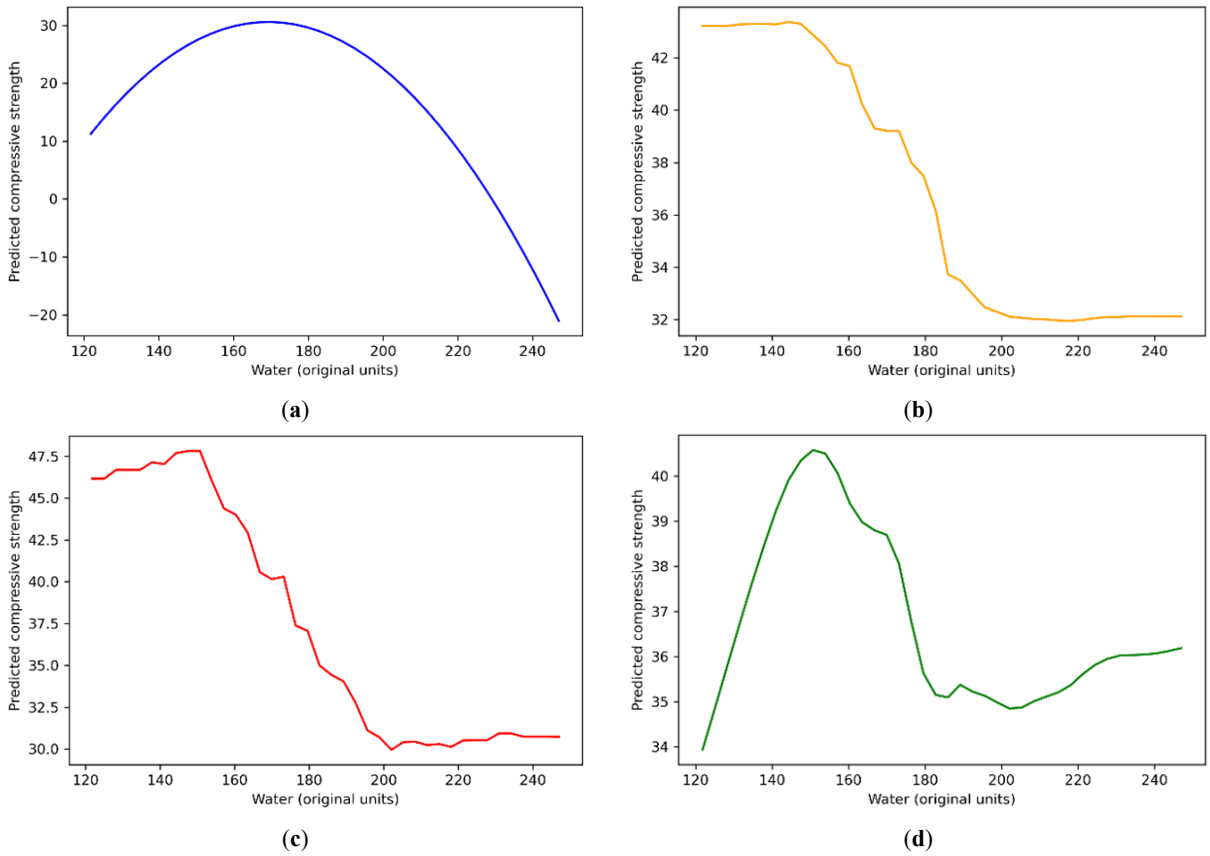


Figure 16. (**a**) PDP for water using linear regression; (**b**) PDP for water using RF; (**c**) PDP for water using XGBoost; (**d**) PDP for water using CNN–BiLSTM.

(d) Blast Furnace Slag

The models give significantly varying definitions of slag contribution. **Figure 17a** indicates that Linear Regression has a smooth upward curve indicating that the strength increases with slag replacement but is unable to portray the latency in early ages that is normally related to slag activation. **Figure 17b** of the random forest presents several increments since the profile of the pozzolanic reaction is discretized into a tree that hides the continuous profile. In **Figure 17c**, XGBoost gives a general upward trend although interrupted by abrupt changes indicating sensitivity to local data blocks. The CNN–BiLSTM ensemble in **Figure 17d** produces a continuous and gradually increasing trend with slight acceleration in the middle of the slag contents, which is in line with the latent hydraulic reaction of slag and the interaction of slag with the available calcium hydroxide. The reaction of the ensemble is closest to the established chemistry of blended systems.

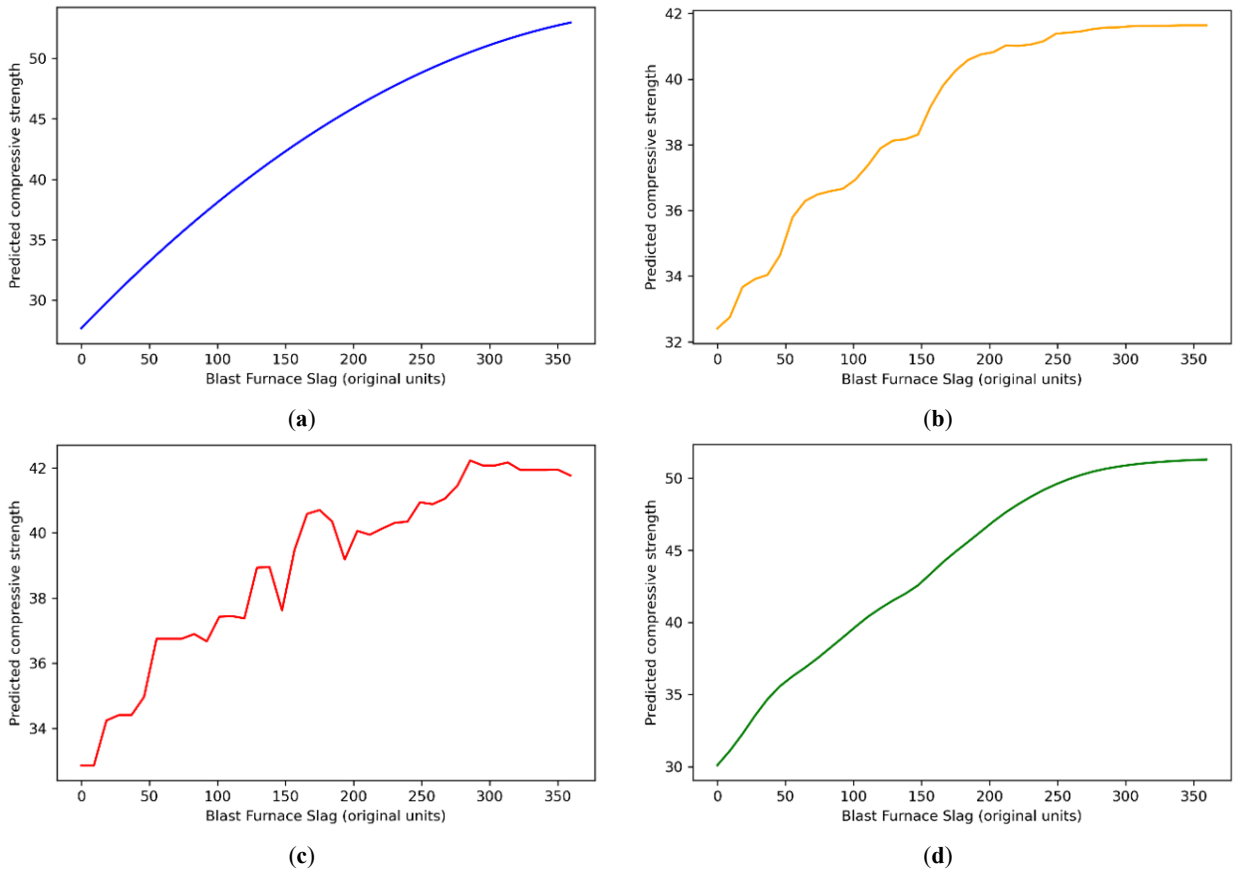


Figure 17. (**a**) PDP for Blast Furnace Slag using linear regression; (**b**) PDP for Blast Furnace Slag using RF; (**c**) PDP for Blast Furnace Slag using XGBoost; (**d**) PDP for Blast Furnace Slag using CNN–BiLSTM.

(e) Superplasticizer

The superplasticizers PDPs indicate significant variations in model interpretation. **Figure 18a** Linear Regression results in a symmetric increase-decrease trend that is not representative of the more complicated rheological behaviour of superplasticizer mixtures. **Figure 18b** of Random Forest displays an increasing trend with a series of plateaus between them caused by the structure of trees and not admixture behavior. **Figure 18c** XGBoost shows small oscillations and sharp fluctuations that are associated with residual fitting, and thus, the trend is not easy to interpret. The CNN–BiLSTM ensemble in **Figure 18d** has the most realistic curve with slight initial downturn due to overdosing in very low doses, then a steady rise and a plateau at higher dosages, which is the expected behavior of optimum dispersion increasing the strength but the extra doses do not give any more significant returns.

(f) Fine Aggregates

Models react to changes in fine aggregate content differently. **Figure 19a** shows that the balance between the availability of sand and the obstruction of paste is predicted by the Linear Regression which assumes an overly symmetric response. Random Forest **Figure 19b** shows a tendency to decrease, with sudden flat plateaus due to the segmentation of trees, which hides the effect of packing density and the distance between particles. **Figure 19c** XGBoost indicates irregular negative fluctuations, which indicate that it is sensitive to sparse areas of data. CNN–BiLSTM ensemble in **Figure 19d** gives a more detailed trend with an initial slight dip, a localized increase towards optimum packing, and a downward trend at high sand contents. This more detailed profile is more consistent with the multifunctional nature of fine aggregates in balancing flowability, packing, and paste demand.

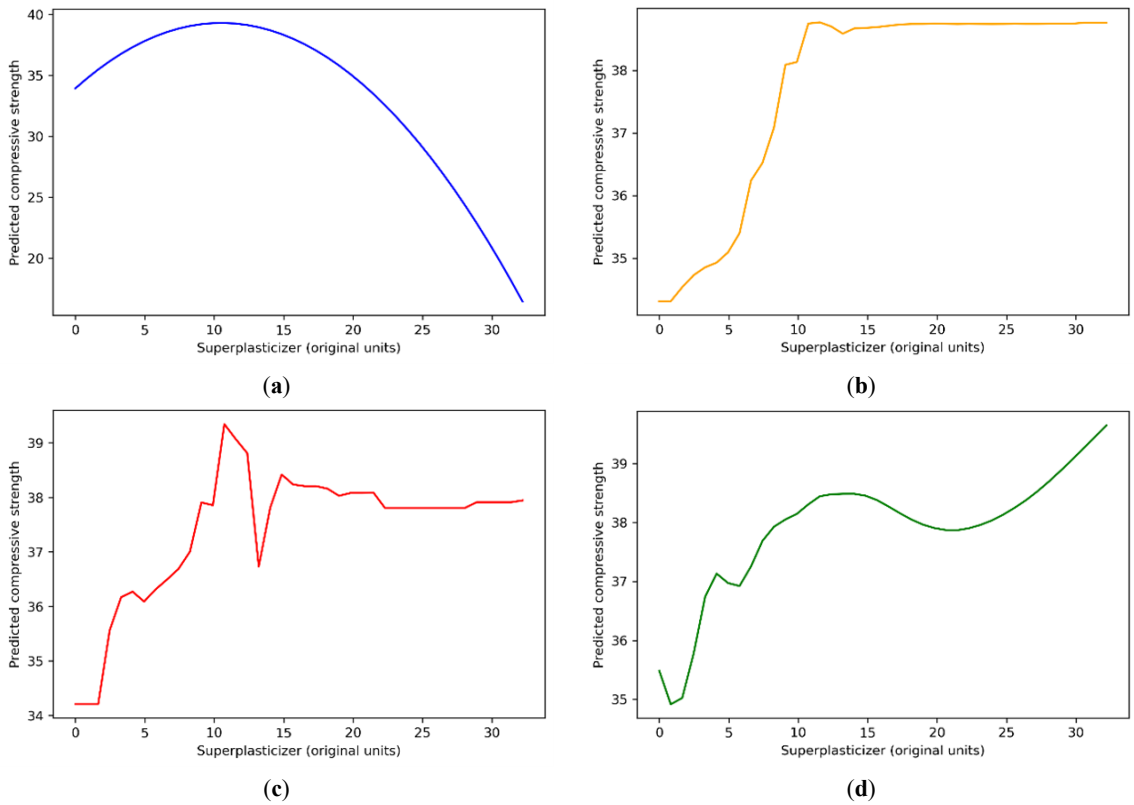


Figure 18. (**a**) PDP for superplasticizer using linear regression; (**b**) PDP for superplasticizer using RF; (**c**) PDP for superplasticizer using XGBoost; (**d**) PDP for superplasticizer using CNN–BiLSTM.

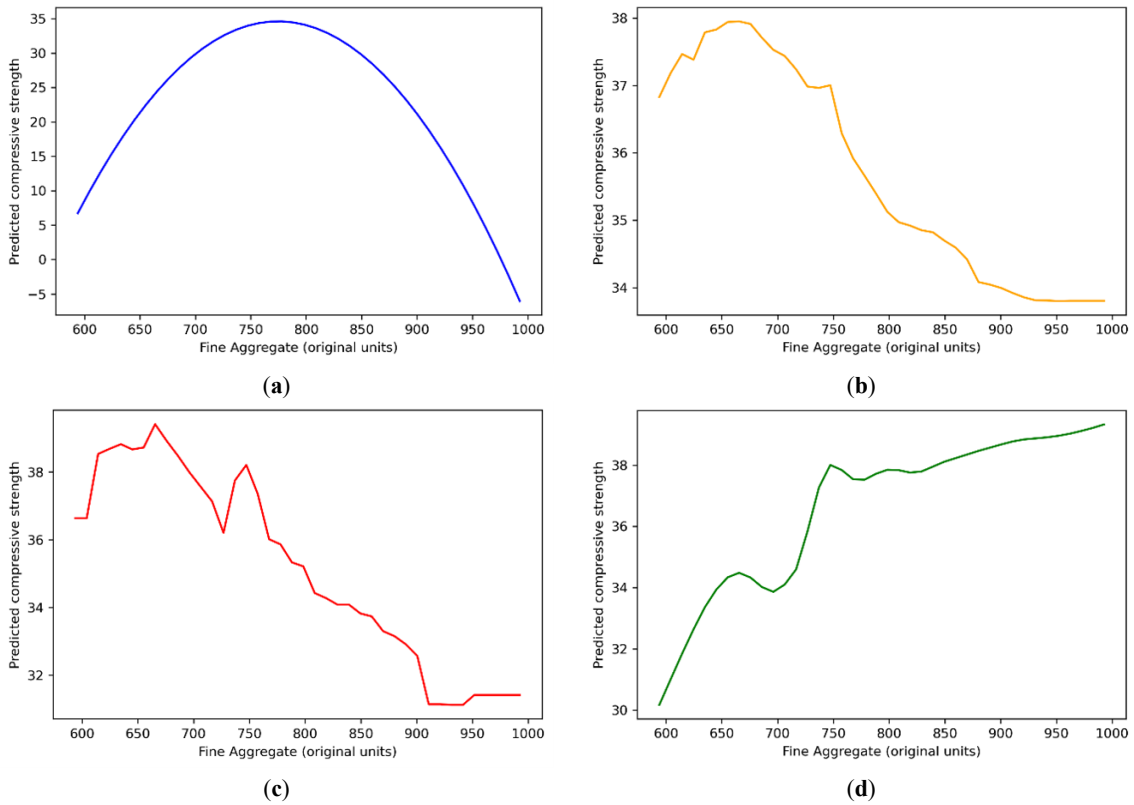


Figure 19. (**a**) PDP for fine aggregate using linear regression; (**b**) PDP for fine aggregate using RF; (**c**) PDP for fine aggregate using XGBoost; (**d**) PDP for fine aggregate using CNN–BiLSTM.

(g) Coarse Aggregates

There are also differences in the interpretation of coarse aggregate effects between models. **Figure 20a** uses Linear Regression, which constrains a parabolic curve, which indicates a single optimum yet does not consider multi-stage changes associated with aggregate stiffness and paste-aggregate interaction. The prediction of the random Forest in **Figure 20b** is of a trend that is predominantly decreasing with artificial plateaus. XGBoost in **Figure 20c** is similar to Random Forest except that it has more pronounced downward variation which is not mechanical. **Figure 20d** CNN–BiLSTM ensemble gives a multi-phase curve with minor initial enhancement, a valley at moderate proportions, and a final increase at high proportions, which is in agreement with the mechanical interaction between the formation of the aggregate skeleton and the quality of paste bonds. This representation is the most realistic material mechanism.

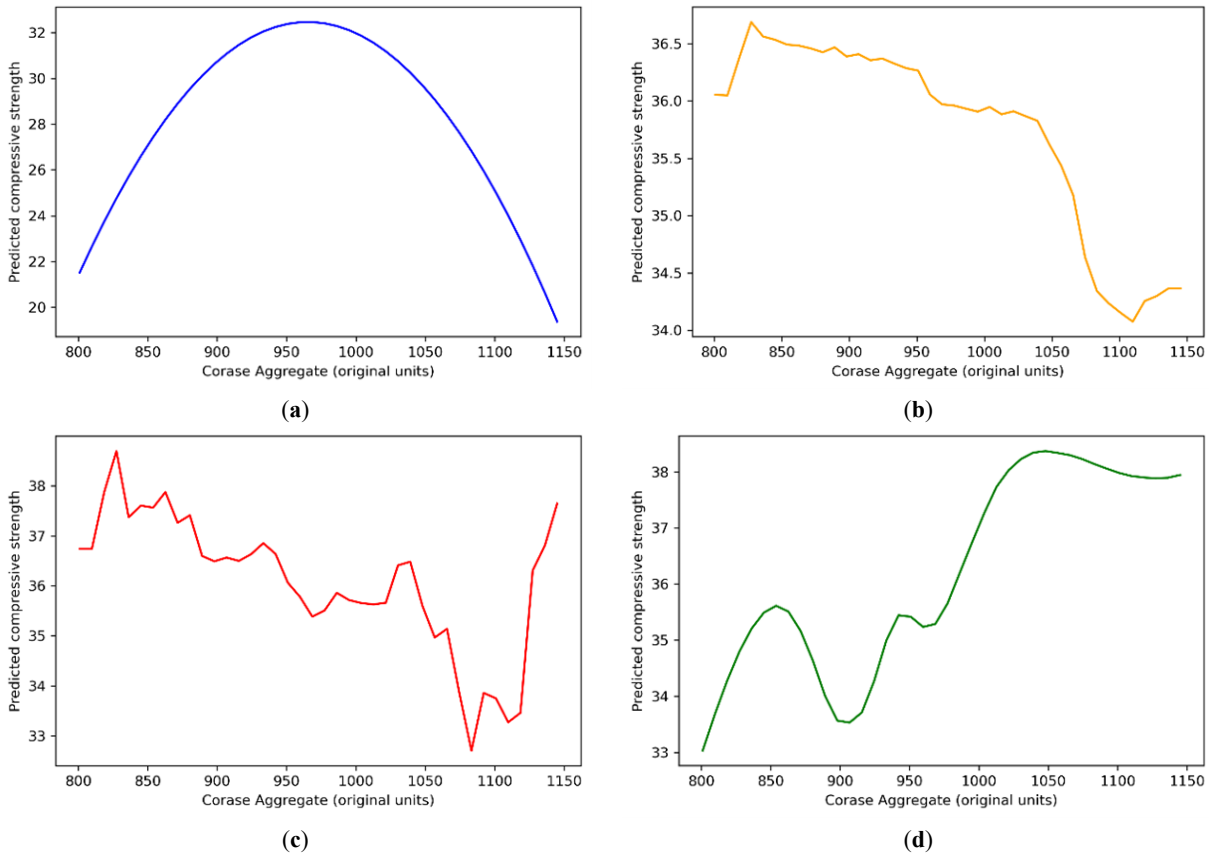


Figure 20. (**a**) PDP for coarse aggregate using linear regression; (**b**) PDP for coarse aggregate using RF; (**c**) PDP for coarse aggregate using XGBoost; (**d**) PDP for coarse aggregate using CNN–BiLSTM.

(h) Fly Ash

There are also significant model-dependent variations in the fly ash PDPs. **Figure 21a** using Linear Regression forecasts a smooth upward increasing trend but does not indicate early-age latency with fly ash contribution being less before secondary hydration. Random Forest in **Figure 21b** gives an irregular curve with sharp falls and partial recovery which are not related to pozzolanic kinetics. In **Figure 21c**, XGBoost shows more prominent decreases and waves generated by overfitting that is boosted. **Figure 21d** shows that the CNN–BiLSTM ensemble produces a physically consistent trend with a minimal initial decrease, and then a steady increase as fly ash strengthens the structure over time. This behavior is consistent with the known behavior of fly ash, and it is the most consistent interpretation of all the models.

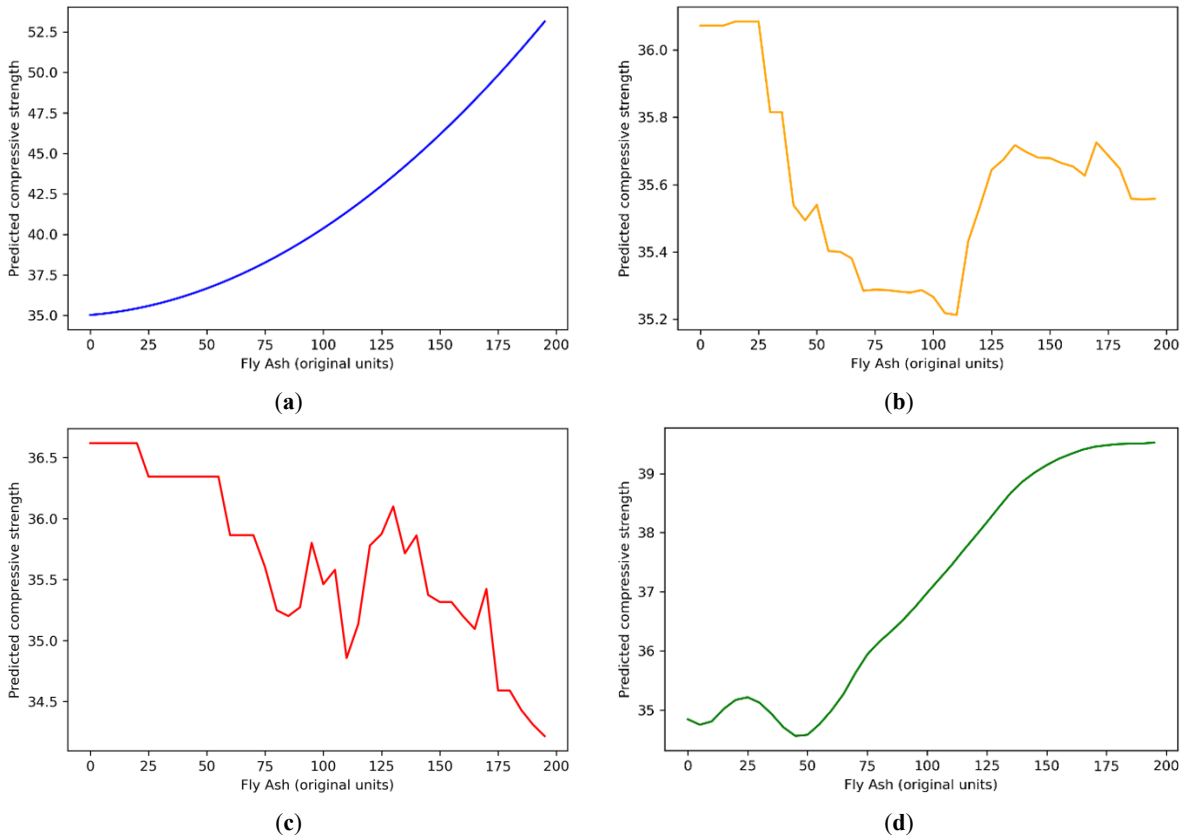


Figure 21. (**a**) PDP for fly ash using linear regression; (**b**) PDP for fly ash using RF; (**c**) PDP for fly ash using XGBoost; (**d**) PDP for fly ash using CNN–BiLSTM.

(i) Influence of Model Type on PDP Behavior

The analysis of the partial dependencies shows clearly that the model structure has a direct influence on the way material behavior can be modeled. Linear Regression always produces smooth, single-curvature responses because its global parametric expression limits its capability to represent local nonlinearities like early-age acceleration, optimum binder contents or admixture saturation. Due to this, it smooths out physical transitions and loses significant mechanistic effects. Random Forest and XGBoost are tree-based models with piecewise and staircase-like trends, which indicate discrete decision boundaries. Random Forest generates sudden jumps and plateaus by averaging tree partitions and XGBoost generates even steeper transitions because of sequential residual fitting. Both techniques are reasonably able to capture the importance of variables, but the PDP curves of both are not physically realistic and discontinuous, which is not suited to processes that are driven by hydration and depend on admixtures, which in real-life scenarios are smooth processes. The CNN–BiLSTM ensemble gives the best continuous, stable and mechanistically meaningful PDP responses. Localized nonlinear features are found by the CNN component and long-range dependencies are found by the BiLSTM, related to cumulative hydration, water sensitivity, and aggregate-paste interactions. The resultant PDPs show continuous transitions, diminishing returns with increasing contents and realistic inflection points that have a strong resemblance to material mechanisms. The PDP analysis reveals that the deep learning models, especially the CNN–BiLSTM ensemble, best reflect the continuous and nonlinearity of concrete strength development, and tree-based and linear models add distortions that do not allow realistic interpretation.

4.7. Shapley Additive exPlanations (SHAP)

The SHAP summary plot can be used to understand the effect of each feature on the XGBoost prediction as it indicates the direction and the strength of the effect. In this study, SHAP analysis was performed specifically for the

XGBoost model to examine feature level contributions and verify that the learned relationships were consistent with known material behavior. In this diagram, red dots placed on the right of the zero line signify strong positive contributions to compressive strength, whereas blue dots on the left signify strong negative contributions. On the other hand, there are blue points on the right and red points on the left which show negative relationship with strength. This directional trend directly points out the variables that have a positive and negative impact on the model output. Age and cement have clear red clusters on the right which proves that they have a positive correlation with compressive strength. Increasing values of these features are always shifting the model towards increasing the predicted strength, which is in line with the known concrete behavior in which curing age and binder content are the factors increasing hydration and strength gain. Water, fine aggregate, and fly ash exhibit red clusters that are mostly located on the left side of the plot implying that increasing values of these variables are associated with less strength than is predicted. This is not surprising because too much water will cause porosity, high fine aggregate content will cause a change in the density of packing and fly ash does not usually add strength in the early age. Blast furnace slag and superplasticizer exhibit mixed but generally positive effects, which is indicative of their effect on increasing hydration and workability when used in the right mix conditions. Coarse aggregate presents red points on both sides of the SHAP axis, meaning that its effect is sample-dependent: in certain combinations, high coarse aggregate content strengthens the material by improving load transfer, whereas in other combinations, it weakens the material by weakening paste-aggregate bonding or by having an inadequate content of binder. The SHAP analysis in **Figure 22** indicates that the relationships obtained by the model are consistent with the behavior of physical concrete and gives a clear picture of the influence of each input variable on the compressive strength predicted by the model. It should be noted that SHAP analysis was not applied to the CNN–BiLSTM model in this study, and therefore interpretability comparisons are limited to the XGBoost framework.

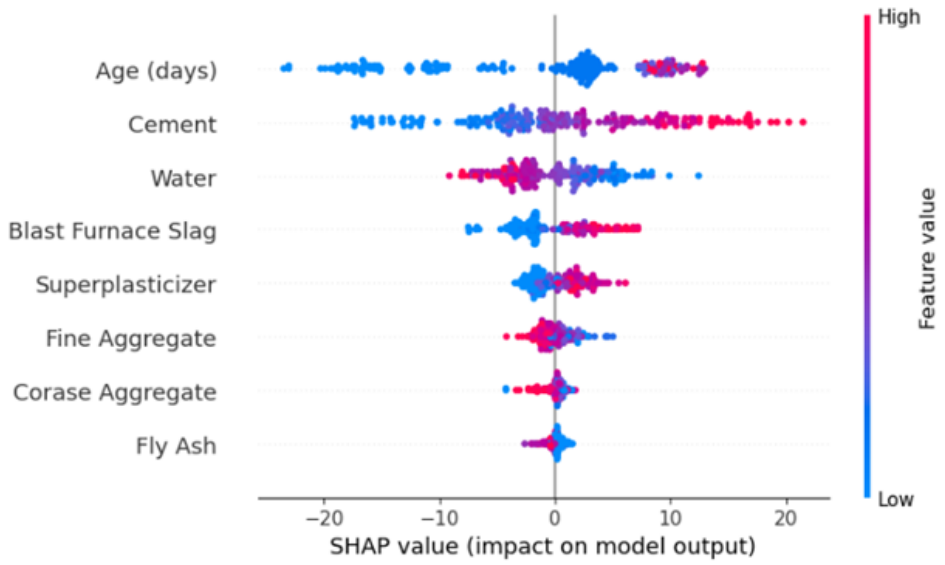


Figure 22. SHAP summary plot illustrating global feature importance and directional influence on the predicted compressive strength.

4.7.1. Analysis of SHAP Interconnection Plots

The SHAP dependence plots indicate that there are strong relationships between the parameters of the mixture and their marginal contributions to compressive strength. The SHAP dependence plots shown in **Figure 23** indicate strong relationships between the mixture parameters and their marginal contributions to compressive strength. These dependence analyses correspond to the XGBoost model and are presented to assess whether the learned feature contributions align with established hydration and mixture interaction principles. The water content is found to have a negative contribution throughout the entire range, with the values of SHAP decreasing with water content. This is due to the water-cement ratio becoming too high and consequently, the

water-cement ratio results in the dilution of binder and the rise in porosity [49]. The color gradient indicates that mixtures that have a higher cement content are more likely to exhibit lower negative SHAP values at similar water content. This shows a moderating effect in which adequate availability of binders partially counteracts the deleterious effect of excess water, as has been previously established in the hydration and microstructural development processes [33].

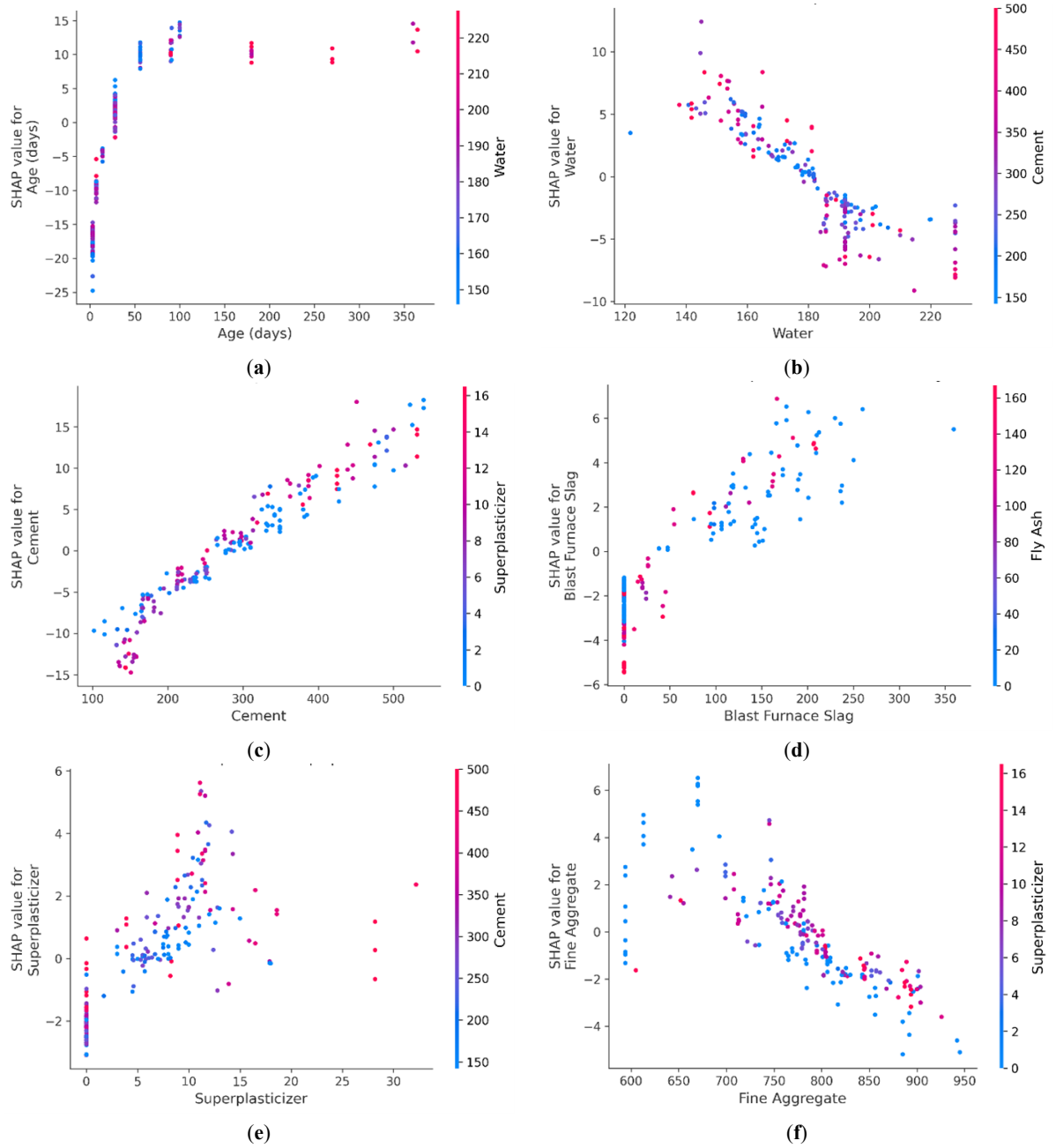


Figure 23. (**a**) SHAP dependence plot for age; (**b**) SHAP dependence plot for water; (**c**) SHAP dependence plot for cement; (**d**) SHAP dependence plot for blast furnace slag; (**e**) SHAP dependence plot for superplasticizer; (**f**) SHAP dependence plot for fine aggregate.

Cement content shows a positive and high contribution to strength, and the values of SHAP increase nearly linearly with cement. This trend is a manifestation of the basic role of cement as the main binder in charge of hydration and strength production. The dosage of superplasticizer as an overlaying color scale has an interesting interaction. Blends with increased superplasticizer levels have marginally better SHAP values at equivalent cement levels, indicating that

increased dispersion and workability increase the efficiency of the binder. This interaction helps to realize that dispersive admixtures may be used to achieve optimal particle packing and hydration efficiency.

The values of SHAP of blast furnace slag have a progressive rise with the increase in slag content. This trend emphasizes the continued role of latent hydraulic reactions, which increases with the replacement of more cement by slag. The fly ash interaction, which is represented by the color gradient, indicates that mixtures containing increased fly ash contents are more likely to exhibit slightly high SHAP values at moderate slag contents. This is indicative of the complementary pozzolanic and hydraulic activity that takes place when the two SCMs are used jointly and improves long-term strength development by creating more C-S-H gel.

The superplasticizer plot shows a slow increase in SHAP values with increase in dosage and contributions level off at higher levels. It means that the better dispersion and less viscosity increases the strength at first by the better packing of the particles and less entrapped air. But high dosages do not lead to any proportionate increase in SHAP values, indicating the benefit levels off at high flow characteristics. The cement color scale supports this explanation since increasing the contents of binders increases the positive effect of admixture efficiency, which is the interaction between the dispersive action and hydration demand.

Fine aggregate content has a low negative trend in SHAP values as sand content rises. The drop indicates that too much fine aggregate upsets the balance between the paste volume and the density of the packing, decreasing the workable paste, and restricting the effective bonding. The dosage response of the superplasticizer shows that mixtures with higher admixture content have a small negative SHAP value, which means that the better the workability, the less negative the high fine aggregate proportions are.

The combination of these SHAP analyses gives a consistent physical explanation of mixture behavior. Parameters that increase the availability of the binder or efficiency of the reaction have progressively positive SHAP contributions whereas variables that dilute binder fraction or disrupt optimal packing have negative trends. The interactions of color also bring out the interplay of combined effects of cementitious materials, water content, and admixture dosage in contributing to the overall strength, which is a complex but systematic process of concrete mixture design.

5. Conclusions

This study provided a comprehensive evaluation of concrete compressive-strength prediction using a broad suite of machine learning and deep learning models, and the findings clearly demonstrate how model architecture influences predictive accuracy, feature interactions, and interpretability. The integrated assessment of statistical models, tree-based ensembles, and hybrid deep-learning architecture revealed distinct strengths and limitations across all approaches.

1. The linear regression model provided a standard of performance and the overall general monotonic trends between mix constituents and compressive strength. Nonlinear interactions were, however, poorly modeled by it, which led to relatively low predictive accuracy and systematic errors in the scatter plots.
2. Random Forest and XGBoost tree-based ensemble models were found to be much more predictive accurate than linear regression. These models were effective in describing heterogeneous, discrete and threshold-based material behavior. However, their partial-dependence profiles were staircase-like showing that they were over-reliant on stepwise approximations as opposed to smooth physical relations. This discontinuous manner of behavior does not enable them to reproduce continuous real-world material responses.
3. XGBoost and RF gave competitive results, which is indicative of their capability to recognize the interaction effects among a variety of compositional variables. However, with good numerical performance, they were limited in their interpretability because of discontinuous PDP behavior and sudden local changes in the responses being predicted.
4. The hybrid CNN–BiLSTM ensemble was always the most predictive and provided a high agreement between experimental and predicted compressive strength in training and testing datasets. The proposed hybrid architecture uniquely integrates one dimensional convolutional feature extraction with bidirectional long short term memory units to simultaneously capture local compositional interactions and long range nonlinear dependencies in mixture design. The model

was able to learn nonlinear relationships, long-range dependencies between features and created smooth partial-dependence curves which corresponded well with known material behavior.

5. The SHAP analysis has proved that the most significant predictors of compressive strength were age, cement, water, and slag. The features with red-to-the-right and blue-to-the-left distributions were associated with strong positive correlations with strength and vice versa. Explanations of deep-learning models included more consistent and physically significant relationships, and tree-based models were more scattered and localized.

6. In all eight variables examined with PDP analysis, the deep-learning ensemble exhibited smooth, continuous and monotonic trends, which were in line with hydration kinetics and mix-design principles. This study uniquely demonstrates that evaluating response continuity and mechanistic alignment provides additional insight beyond conventional error metrics when comparing predictive models for concrete strength.

Despite the strong numerical performance of the tree based models, reflected by high R^2 values and low prediction errors, their piecewise decision structure inherently produces stepwise response surfaces in partial dependence analysis. While this behavior is a normal mathematical property of tree algorithms and does not imply incorrect predictions, it may introduce artificial discontinuities when interpreted in the context of material behavior. In cementitious systems, compressive strength develops progressively through continuous hydration, microstructural densification, and gradual pozzolanic reactions rather than discrete jumps between curing intervals or abrupt transitions caused by small changes in mixture proportions. Therefore, our argument is not that tree-based models are invalid predictors, but that their response representation may be less aligned with the known continuous nature of strength development. The present evaluation was conducted using the established Yeh dataset to enable controlled benchmarking under identical conditions across models. While this provides a rigorous comparative framework, validation on independent experimental or field datasets would further strengthen confidence in broader generalization. In contrast, the CNN–BiLSTM ensemble learns continuous nonlinear mappings that generate smoother and gradually varying trends, which more closely reflect established hydration mechanisms, while maintaining predictive performance comparable to state of the art ensemble tree models rather than claiming substantial numerical superiority.

Author Contributions

Research conception and design, A.K. and A.A.; machine learning and implementation, A.K. and A.A.; data analysis and writing, A.K. and A.A.; review and rewriting, A.K., A.A., and S.S.M. All authors have read and agreed to the published version of the manuscript.

Funding

The authors received no specific funding for this research from any public, commercial, or not for profit funding agency.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The dataset used in this study is publicly available and corresponds to the Yeh concrete compressive strength dataset.

Acknowledgments

The authors acknowledge the UC Irvine Machine Learning Repository for providing the dataset that makes the foundation for this study. The authors also greatly acknowledge the University of Louisiana at Lafayette for providing the research facilities for this study.

Conflicts of Interest

The authors declare no conflict of interest.

AI Use Statement

During the preparation of this work, the authors used ChatGPT by OpenAI only for minor grammar and clarity checks. The authors subsequently reviewed and edited the content as necessary and take full responsibility for the final content of the published article.

References

[1] Moein, M.M., Saradar, A., Rahmati, K., et al., 2023. Predictive Models for Concrete Properties Using Machine Learning and Deep Learning Approaches: A Review. Journal of Building Engineering. 63, 105444.

[2] Ali, A., Khan, Q.U.Z., Mehboob, S.S., et al., 2024. Enhancing Multi-Objective Mix Design for GGBS-Based Geopolymer Concrete with Natural Mineral Blends under Ambient Curing: A Taguchi-Grey Relational Optimization. Ain Shams Engineering Journal. 15(5), 102708.

[3] Ruth Bola Oliveira, D.; Leite, G.; Possan, E.; et al., 2023. Concrete powder waste as a substitution for Portland cement for environment-friendly cement production. Construction and Building Materials. 397, 132382. DOI: https://doi.org/10.1016/j.conbuildmat.2023.132382

[4] Aziz, A., Mehboob, S.S., Tayyab, A., et al., 2024. Enhancing Sustainability in Self-Compacting Concrete by Optimizing Blended Supplementary Cementitious Materials. Scientific Reports. 14, 12326.

[5] Ali, A., Khan, Q.U.Z., Mehboob, S.S., 2024. Experimental Design Optimization Using Taguchi-Analysis for Compositional Parameters of Eco-Friendly GGBS Based Geopolymer Concrete. In Proceedings of the 2nd International Conference on Recent Advances in Civil Engineering and Disaster Management, Peshawar, Pakistan, 15–16 December 2022.

[6] Jain, S., Jain, R., Rao, K.R., et al., 2024. Leveraging Machine Learning to Minimize Experimental Trials and Predict Hot Deformation Behaviour in Dual Phase High Entropy Alloys. Materials Today Communications. 41, 110813.

[7] Khan, A., Khattak, M.J., Pesacreta, T.C., 2025. Development of Geopolymer-Based Green Construction Materials for Highway Infrastructure Using High-Content Sugarcane Bagasse Ash. Transportation Research Record. 2680(1), 121–137.

[8] Khan, A., Khattak, M.J., Pesacreta, T.C., et al., 2025. Microstructural, Mechanical, and Durability Assessment of Sustainable Geopolymers Synthesized Using Rice Husk Ash, a Byproduct of the Rice Industry. Journal of Sustainable Cement-Based Materials. 14, 2519–2536.

[9] Khan, A., 2025. Development of Green Construction Materials Using High-Content Sugarcane Bagasse Ash [Master's Thesis]. University of Louisiana at Lafayette: Lafayette, LA, USA.

[10] Haq, I.U., Elahi, A., Khan, A., et al., 2025. The Addition of Natural Clay and Industrial and Agricultural Waste on the Performance of Green and Sustainable Concrete. Scientific Reports. 15, 17034.

[11] Sun, B., Huang, Y., Liu, G., et al., 2024. Prediction of Compressive Strength of Concrete under Various Curing Conditions: A Comparison of Machine Learning Models and Empirical Mathematical Models. Innovative Infrastructure Solutions. 9, 262.

[12] Yeh, I.-C., 1998. Concrete Compressive Strength. Available from: https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength (cited 3 May 2025).

[13] Ilyas, I., Zafar, A., Afzal, M.T., et al., 2022. Advanced Machine Learning Modeling Approach for Prediction of Compressive Strength of FRP Confined Concrete Using Multiphysics Genetic Expression Programming. Polymers. 14, 1789.

[14] Sun, J., Dang, W., Wang, F., et al., 2023. Prediction of TOC Content in Organic-Rich Shale Using Machine Learning Algorithms: Comparative Study of Random Forest, Support Vector Machine, and XGBoost. Energies. 16, 4159.

[15] Ozcan, G., Gulbandilar, E., Kocak, Y., 2025. Prediction of Compressive Strengths of Portland Cement with Random Forest, Support Vector Machine and Gradient Boosting Models. Neural Computing and Applications. 37, 23495–23511.

[16] Ali, T., Onyelowe, K.C., Mahmood, M.S., et al., 2025. Advanced and Hybrid Machine Learning Techniques for Predicting Compressive Strength in Palm Oil Fuel Ash-Modified Concrete with SHAP Analysis. Scientific Reports. 15, 4997.

[17] Alsharari, F., 2025. Predicting the Compressive Strength of Fiber-Reinforced Recycled Aggregate Concrete: A Machine-Learning Modeling with SHAP Analysis. Asian Journal of Civil Engineering. 26, 179–205.

[18] Ghrici, A.A., Benzaamia, A., Mezzoudj, F., et al., 2025. SHAP-Enhanced Tree-Based Regression for Predicting the Compressive Strength of High Performance Concrete. Construction and Building Materials. 495, 143602.

[19] Berodier, E.M.J., 2015. Impact of the Supplementary Cementitious Materials on the Kinetics and Microstructural Development of Cement Hydration [PhD Thesis]. École Polytechnique Fédérale de Lausanne: Lausanne, Switzerland.

[20] Moradi, N., Tavana, M.H., Habibi, M.R., et al., 2022. Predicting the Compressive Strength of Concrete Containing Binary Supplementary Cementitious Material Using Machine Learning Approach. Materials. 15, 5336.

[21] Alzubaidi, L., Zhang, J., Humaidi, A.J., et al., 2021. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. Journal of Big Data. 8, 53.

[22] Liu, Y., Yu, H., Guan, T., et al., 2025. Intelligent Prediction of Compressive Strength of Concrete Based on CNN-BiLSTM-MA. Case Studies in Construction Materials. 22, e04486.

[23] Ziolkowski, P., 2025. Influence of Optimization Algorithms and Computational Complexity on Concrete Compressive Strength Prediction Machine Learning Models for Concrete Mix Design. Materials. 18, 1386.

[24] Bernal, S.A., de Gutiérrez, R.M., Pedraza, A.L., et al., 2011. Effect of Binder Content on the Performance of Alkali-Activated Slag Concretes. Cement and Concrete Research. 41, 1–8.

[25] Powers, T.C., 1947. A Discussion of Cement Hydration in Relation to the Curing of Concrete. Portland Cement Association: Chicago, IL, USA.

[26] Li, L.G., Kwan, A.K., 2015. Effects of Superplasticizer Type on Packing Density, Water Film Thickness and Flowability of Cementitious Paste. Construction and Building Materials. 86, 113–119.

[27] Yusslee, E., Beskhyroun, S., 2023. The Effect of Water-to-Binder Ratio (W/B) on Pore Structure of One-Part Alkali Activated Mortar. Heliyon. 9(1), e12983.

[28] Salman, H.A., Kalakech, A., Steiti, A., 2024. Random Forest Algorithm Overview. Babylonian Journal of Machine Learning. 2024, 69–79.

[29] Rezzoug, A., Khan, A., Murtaza, N., et al., 2025. Predicting Flood Energy Reduction in Vegetated Open Channel: Comparative Assessment of Hybrid Artificial Intelligence Techniques. Engineering Applications of Artificial Intelligence. 159, 111756.

[30] Özçift, A., 2011. Random Forests Ensemble Classifier Trained with Data Resampling Strategy to Improve Cardiac Arrhythmia Diagnosis. Computers in Biology and Medicine. 41(5), 265–271.

[31] Jaiswal, J.K., Samikannu, R., 2017. Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression. In Proceedings of the 2017 World Congress on Computing and Communication Technologies, Tiruchirappalli, India, 2–4 February 2017; pp. 65–68.

[32] Xia, S., Wang, G., Chen, Z., et al., 2018. Complete Random Forest Based Class Noise Filtering Learning for Improving the Generalizability of Classifiers. IEEE Transactions on Knowledge and Data Engineering. 31, 2063–2078.

[33] Al-Shamasneh, A.R., Kewalramani, M., Mahmoodzadeh, A., et al., 2025. Forecasting Compressive Strength of Concrete Containing Rice Husk Ash Using Various Machine Learning Algorithms. Scientific Reports. 15, 39162.

[34] Sahin, E.K., 2020. Assessing the Predictive Capability of Ensemble Tree Methods for Landslide Susceptibility Mapping Using XGBoost, Gradient Boosting Machine, and Random Forest. SN Applied Sciences. 2, 1308.

[35] Rouhani, M.M., Dolatshahi, A., 2025. Advanced Hybrid Models for Predicting Dynamic Compressive Strength of Rocks: Exploring Various Input Scenarios and Introducing Novel Empirical Approaches. Results in Engineering. 28, 107947.

[36] Nielsen, D., 2016. Tree Boosting with Xgboost—Why Does Xgboost Win Every Machine Learning Competition? Norwegian University of Science and Technology: Trondheim, Norway.

[37] Demir, S., Şahin, E.K., 2022. Liquefaction Prediction with Robust Machine Learning Algorithms (SVM, RF, and XGBoost) Supported by Genetic Algorithm-Based Feature Selection and Parameter Optimization from the Perspective of Data Processing. Environmental Earth Sciences. 81, 459.

[38] Rouhani, M.M., Dolatshahi, A., 2025. Predicting Mode I Fracture Toughness of CCNBD Quasi-Brittle Specimens Using Hybrid Gradient Boosting Algorithms. Theoretical and Applied Fracture Mechanics. 138, 104966.

[39] Khan, A., Sohail, A., Zahoora, U., et al., 2020. A Survey of the Recent Architectures of Deep Convolutional Neural Networks. Artificial Intelligence Review. 53, 5455–5516.

[40] Ding, B., Qian, H., Zhou, J., 2018. Activation Functions and Their Characteristics in Deep Neural Networks. In Proceedings of 2018 Chinese Control And Decision Conference (CCDC), Shenyang, China, 9–11 June 2018; pp. 1836–1841.

[41] Gholamalinezhad, H., Khosravi, H., 2020. Pooling Methods in Deep Neural Networks, a Review. arXiv preprint. arXiv:2009.07485.

[42] Li, C., Zhang, M., Zhang, X., 2023. Enhancing Concrete Creep Prediction with Deep Learning: A Soft-Sorted One-Dimensional CNN Approach. IEEE Access. 11, 139314–139325.

[43] Atef, S., Eltawil, A.B., 2020. Assessment of Stacked Unidirectional and Bidirectional Long Short-Term Memory Networks for Electricity Load Forecasting. Electric Power Systems Research. 187, 106489.

[44] Yu, Y., Si, X., Hu, C., et al., 2019. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. Neural Computation. 31, 1235–1270.

[45] Chandra, N., Ahuja, L., Khatri, S.K., et al., 2021. Utilizing Gated Recurrent Units to Retain Long Term Dependencies with Recurrent Neural Network in Text Classification. Journal of Information Systems and Telecommunication (JIST). 2, 89.

[46] Liu, G., Guo, J., 2019. Bidirectional LSTM with Attention Mechanism and Convolutional Layer for Text Classification. Neurocomputing. 337, 325–338.

[47] Zhang, S., Liu, C., Jiang, H., et al., 2015. Feedfor-

ward Sequential Memory Networks: A New Structure to Learn Long-Term Dependency. arXiv preprint. arXiv:1512.08301.

[48] Lei, F., Liu, X., Li, Z., et al., 2021. Multihop Neighbor Information Fusion Graph Convolutional Network for Text Classification. Mathematical Problems in Engineering. 2021(1), 6665588.

[49] Piasta, W., Zarzycki, B., 2017. The Effect of Cement Paste Volume and w/c Ratio on Shrinkage Strain, Water Absorption and Compressive Strength of High Performance Concrete. Construction and Building Materials. 140, 395–402.