Research on Engineering Structures & Materials
www.jresm.org
Research Group

Research Article

Data-driven prediction of UHPC compressive strength using a hybrid 1D CNN–GRU network

Nguyen Thi Thu Nga [*,a]

Faculty of Engineering, University of Transport Technology, Hanoi, Vietnam

Article Info

Article History:

Received 04 July 2025
Accepted 26 Aug 2025

Keywords:

Ultra-high performance concrete;
Data-Driven;
1D CNN;
Gated recurrent unit

Abstract

Ultra-high-performance concrete (UHPC) is one of the cutting-edge materials in the concrete industry. UHPC possesses superior mechanical properties compared to conventional concrete, bringing many breakthroughs in construction. Compressive strength is one of the most essential properties of UHPC and is determined through destructive laboratory tests. These tests are often costly and time-consuming. In this study, a data-driven approach is proposed to predict the compressive strength of UHPC using a hybrid deep learning model combining a Convolutional Neural Network (CNN) with a Gated Recurrent Unit (GRU). The 1D CNN component effectively extracts local feature patterns among material properties, while the GRU module captures sequential and interdependent relationships. A comprehensive dataset, including various mix designs, was used for model training and validation. The performance of the hybrid CNN–GRU network was compared with the standalone CNN and LSTM models. The results demonstrate that the proposed hybrid model achieves superior accuracy, exhibiting lower mean absolute error (MAE) and mean square error (MSE) on the test dataset. This study highlights the potential of data-driven hybrid neural networks in improving the prediction of UHPC compressive strength, providing practical insights for optimizing mix design in UHPC applications.

1. Introduction

The size and complexity of infrastructure projects have increased recently, leading to a rise in the demands imposed on building materials, especially regarding strength, lifespan, and durability. Ultra-High Performance Concrete (UHPC), a novel material that marks a substantial improvement over traditional concrete technologies, has been developed and adopted as a result of these demands [1,4]. Characterized by its exceptional mechanical properties, UHPC typically exhibits compressive strengths exceeding 150 MPa—approximately three times greater than the concrete used in most critical structures. Furthermore, UHPC demonstrates remarkable durability, with superior resistance to chemical attacks, abrasion, and environmental degradation. These attributes make it an ideal candidate for high-demand applications such as long-span bridges, high-rise towers, offshore platforms, and hydroelectric structures.

The compressive strength of UHPC is frequently assessed using destructive testing techniques in laboratories to guarantee its performance [5,6]. These tests are time-consuming, labor-intensive, and expensive, especially when several trial mixtures are required for optimization, even if they provide accurate and dependable measurements. Furthermore, quality control is challenging due to the diversity in test specimen production and curing. Destructive testing causes substantial material waste and hinders the design process in UHPC, where each batch may contain pricey components and additives.

*Corresponding author: ngantt@utt.edu.vn
[a]orcid.org/0000-0002-9613-6011
DOI: http://dx.doi.org/10.17515/resm2025-1006me0704rs

Predicting the compressive strength of UHPC accurately remains a formidable challenge due to the complex, nonlinear interactions among numerous mix design parameters. Among these are the amounts of fly ash, slag, cement, silica fume, steel or polymer fibers, nano additives, and water-to-binder ratio. These complex linkages are frequently complicated for traditional empirical or statistical regression models to capture, particularly when working with big and diverse datasets or investigating unusual compositions.

Amidst this challenge, Artificial Intelligence (AI) and Machine Learning (ML) techniques have emerged as powerful tools for modeling and prediction tasks in materials science and civil engineering [7,9]. These techniques have shown great potential in addressing the limitations of conventional methods, particularly in problems involving highly nonlinear data. In the context of concrete materials, several studies have demonstrated the efficacy of ML models in predicting mechanical properties with high accuracy. Alaneme et al. [10] explored the use of Artificial Neural Networks (ANN), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and Gene Expression Programming (GEP) for predicting both the compressive and flexural strengths of geopolymer concrete incorporating banana peel ash (BPA) and sugarcane bagasse ash (SCBA), with ANFIS outperforming other models. Similarly, Okasha et al. [11] applied machine learning algorithms including ANN, Support Vector Regression (SVR), and Histogram-based Gradient Boosting (HGB) to estimate the elastic modulus and flexural strength of concrete incorporating carbon nanotubes (CNTs), finding that ANN yielded the best predictions for modulus. At the same time, HGB was superior for flexural strength. Rezvan et al. [12] developed an ANN model using 85 data samples to predict the compressive strength of plastic fiber-reinforced concrete (PFRC) derived from recycled PET bottles, leveraging 12 input variables related to mix design and fiber properties.

The ability of ANN and related models to capture the nonlinear behavior of concrete performance has been shown by numerous other investigations [13,14]. However, in the face of highly complex patterns or interdependencies typical of UHPC datasets, ANN-based methods may find it difficult to generalize because they frequently rely significantly on manually chosen input features. Research into deep learning and hybrid models, which can automatically extract multiscale characteristics and increase resilience in prediction tasks, has been spurred by this.

Recent research has looked to deep learning architectures, which can automatically learn complex, high-dimensional representations from raw data, to overcome the shortcomings of conventional and shallow learning models. Among them, the Convolutional Neural Network–Gated Recurrent Unit (CNN-GRU) hybrid model has drawn interest due to its ability to integrate GRU's temporal modeling capabilities with CNN's feature extraction power. While GRU layers are good at modeling sequential dependencies and nonlinear correlations in the data, CNN layers in this architecture serve as strong local pattern detectors that capture multiscale dependencies and interactions among mixed design elements. CNN-GRU is more resilient to overfitting than traditional Artificial Neural Networks (ANNs), particularly when working with structured but noisy experimental datasets [15,16]. Furthermore, GRU is a lightweight version of Long Short-Term Memory (LSTM) that maintains a high capacity for learning long-term dependencies while lowering computational cost. This improves prediction accuracy and generalization across different mix compositions, making CNN-GRU a viable option for precisely forecasting UHPC compressive strength based on various input factors

2. Methodology

2.1. Data Collection and Pre-processing

The dataset used in this research was compiled from comprehensive experimental results documented in existing literature and publicly accessible repositories focused on UHPC. Specifically, the primary data source utilized was the publicly available UHPC dataset published by Mahjoubi and Bao [17,18] via Mendeley Data, which systematically collates experimental outcomes from numerous international studies. This dataset encompasses 381 distinct UHPC mix compositions, rigorously collected from laboratory tests documented in peer-reviewed research papers and technical reports, thus ensuring reliability and reproducibility. The detailed laboratorytested mix proportions and compressive strength results have been validated through a rigorous peer review process

Each data entry comprises 16 critical mix-design parameters: cement content and type (strength grade), fly ash, slag, silica fume, metakaolin, nano silica, limestone, quartz powder, sand, maximum aggregate size, water content, superplasticizer dosage, steel fiber volume fraction, fiber aspect ratio, and specimen dimensions. The corresponding target output is the compressive strength measured after 28 days of curing, evaluated via standardized laboratory tests according to international guidelines (e.g., ASTM C39/C39M and EN 12390-3). The dataset reflects a diverse range of mix designs, making it suitable for training and evaluating deep learning models. Its public availability ensures transparency, reproducibility, and alignment with scientific best practices in data-driven research. Table 1 lists the input parameters used for the hybrid deep learning model

Table 1. Input parameters

No.	Input Parameter	Description	Range	Median	IQR	Mean±SD
1	Cement	Cement-to-cm ratio	0.174 – 1.000	0.793	0.084	0.752 ± 0.135
2	Cement type	Strength grade of cement (e.g., 42.5 or 52.5)	42.5 – 52.5	52.5	10.0	48.835 ± 4.822
3	Fly ash	Fly ash-to-cm ratio	0 – 0.503	0.0	0.0	0.049 ± 0.11
4	Slag	Slag-to-cm ratio	0 – 0.696	0.0	0.0	0.028 ± 0.094
5	Silica fume	Silica fume-to-cm ratio	0 – 0.257	0.18	0.113	0.158 ± 0.082
6	Metakaolin	Metakaolin-to-cm ratio	0 – 0.286	0.0	0.0	0.009 ± 0.037
7	Nano silica	Nano silica-to-cm ratio	0 – 0.172	0.0	0.0	0.004 ± 0.014
8	Limestone	Limestone-to-cm ratio	0 – 1.088	0.0	0.0	0.019 ± 0.091
9	Quartz powder	Quartz powder-to-cm ratio	0 – 0.789	0.0	0.196	0.078 ± 0.127
10	Sand	Sand-to-cm ratio	0.264 – 2.070	1.0	0.345	1.014 ± 0.422
11	Max aggregate size	Maximum particle size of aggregates (mm)	0.1 – 5.0	2.0	4.4	2.233 ± 1.783
12	Water	Water-to-cm ratio	0.118 – 0.286	0.192	0.03	0.189 ± 0.028
13	Superplasticizer	Superplasticizer-to-cm ratio	0.007 – 0.238	0.023	0.012	0.027 ± 0.02
14	Steel fiber volume	Fiber volume fraction (%)	0 – 6	0.003	2.0	1.053 ± 1.441
15	Fiber aspect ratio	Aspect ratio of steel fibers (length/diameter)	0 – 400	37.5	65.0	37.793 ± 53.519
16	Specimen size	Size of compressive strength specimen (mm)	40 – 110	40.0	10.0	48.936 ± 17.041

Exploratory Data Analysis (EDA) was conducted to ensure the reliability and suitability of the UHPC dataset for predictive modeling. For data quality and missing values: All records were cross-checked and no missing values were found. The initial dataset was culled from peer-reviewed studies, ensuring standardized reporting of mix proportions and compressive strengths. For outlier detection: Box plots and z-score analysis were performed for each numerical characteristic. Less than 2% of the samples exhibited extreme values, primarily related to unusually high fiber frame ratios (>350) or silica fume ratios. These values were retained as they corresponded to valid experimental designs and not data entry errors.

Distribution and Correlation Analysis: Frequency histograms and density plots were generated to examine characteristic distributions. Most features follow a right-skewed distribution typical for mixture proportion data (e.g., predominantly low silica fume and nano-silica content). Figure 1 presents the Pearson correlation matrix among all 16 input features and the target compressive strength. Strong positive correlations are observed for water-to-binder ratio, silica fume, and steel fiber volume, while most other features exhibit weak to moderate correlations (|r| < 0.5). This confirms the necessity of using a nonlinear hybrid model such as CNN–GRU to capture complex feature interactions.

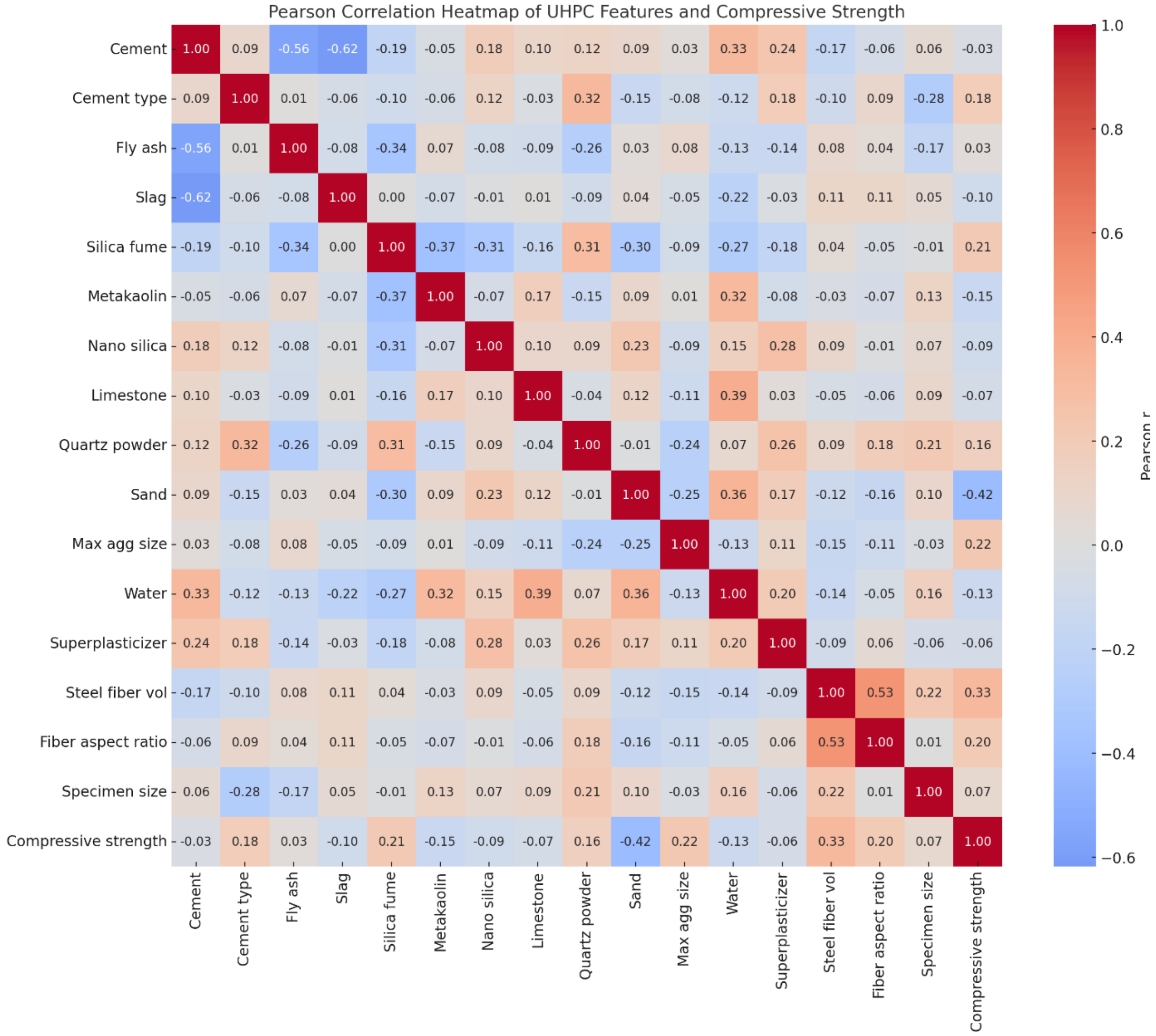


Fig. 1. Pearson correlation heatmap from UHPC dataset

Feature Importance and Dimensionality Reduction: No explicit feature removal or dimensionality reduction (e.g., PCA) was applied because (i) the CNN–GRU model is capable of automatic feature extraction, and (ii) retaining all 16 features preserves potential nonlinear interactions that are important for UHPC performance. Post-permutation importance analysis confirmed that water/binder ratio, silica fume, and steel fiber volume were the three most influential variables in predicting compressive strength. This EDA confirmed that the dataset was complete, contained no erroneous entries, and that all 16 input features contributed meaningful information, either directly or through nonlinear interactions, to predicting UHPC compressive strength.

All input data were normalized in the range of 0-1 using Min-Max scaling to ensure efficient model convergence. The dataset was randomly divided into 80% for training and 20% for testing to evaluate the model's generalization ability.

2.2. Convolutional Neural Networks (CNN)

CNN is a specialized type of deep learning architecture widely applied in pattern recognition and predictive tasks [19,21]. CNNs have demonstrated outstanding capabilities in automatically extracting and learning features from raw data without manual feature engineering [22,23]. A convolution operation between an input feature map X and a filter W produces an output feature map Z, calculated as:

$$Z_{i,j,k}^{(l)} = f\left(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} W_{m,n,k}^{(l)} . X_{i+m-1,j+n-1}^{(l-1)} + b_k^{(l)}\right) \tag{1}$$

where: $Z_{i,j,k}^{(l)}$ is the output value at location *(i, j)* in feature map *k* at layer *l*; $X^{(l-1)}$ is input from the previous layer; $W_{m,n,k}^{(l)}$ is the Weight of the filter (kernel) element at position *(m, n)* for feature map *k*; $b_k^{(l)}$ is bias term for the *k* filter; f*(.)* is activation function (commonly ReLU)

After convolution, CNNs typically apply a pooling layer (e.g., max pooling or average pooling) to reduce spatial dimensions while retaining the most essential features:

$$Z_{i,j}^{(k)} = \max\left\{X_{p,q,k}^{(l-1)}\right\} \tag{2}$$

Where *p and q* represent the indices within the pooling region.

The output of the final convolutional and pooling stages is usually flattened and passed through one or more fully connected (Dense) layers for final prediction:

$$y_i = f\left(\sum_{i=1}^{n} w_{i,j} x_i + b_j\right) \tag{3}$$

With x_i is feature input; $w_{i,j}$ is weight; b_i is bias; *f(.)* is activation function.

2.3. Gated Recurrent Unit (GRU)

Cho et al. [24] presented the Gated Recurrent Unit (GRU), a recurrent neural network (RNN) architecture, as a more straightforward substitute for the Long Short-Term Memory (LSTM) network. GRU was created to solve the vanishing gradient issue that conventional RNNs frequently run into when attempting to model long-term dependencies in sequential data. In contrast to traditional RNNs, GRU uses a gating mechanism that adaptively regulates the input flow, allowing the network to keep pertinent historical data while eliminating unnecessary signals.

The GRU architecture combines the hidden state and cell state of LSTM into a single state vector, and uses two primary gates: the update gate and the reset gate. The update gate z_t determines how much the previous hidden state h_{t-1} should be carried forward to the current state h_t. The reset gate r_t controls how much prior information is ignored in the current computation. The mathematical formulation of GRU is as follows:

$$z_t = \sigma\left(W_z x_t + U_z h_{t-1} + b_z\right) \tag{4}$$

$$r_t = \sigma\left(W_r x_t + U_r h_{t-1} + b_r\right) \tag{5}$$

$$\tilde{h}_t = \tanh\left(W_h x_t + U_h\left(r_t \cdot h_{t-1}\right) + b_h\right) \tag{6}$$

$$h_t = \left(1 - z_t\right) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \tag{7}$$

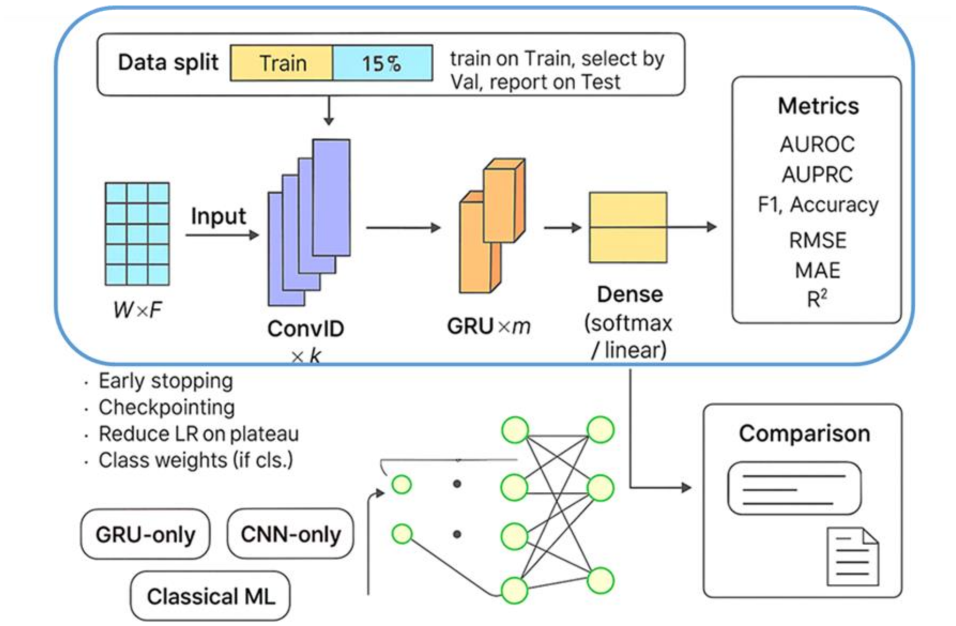
Where; x_t is the input at time step *t*, h_t is the hidden state at time step *t*, *σ* is the sigmoid activation function, . denotes element-wise multiplication, $\tilde{h}_t$ is the candidate hidden state, *W* and *U* are learnable weight matrices, and *b* are biases.

GRU has fewer parameters than LSTM, making it computationally more efficient while maintaining comparable performance in many sequence modeling tasks. In civil engineering applications such as predicting the compressive strength of UHPC, GRU offers an effective solution for modeling nonlinear relationships between input parameters and strength evolution, particularly when temporal or sequential dependencies are embedded in the data.

2.4. Hybrid CNN-GRU

The primary function of CNN layers is to extract significant local characteristics from the input data automatically. By applying one-dimensional convolutions to the input vector, CNNs can efficiently detect spatial patterns, correlations, and interactions between materials, including cement, silica fume, water, fibers, and admixtures. This decreases dimensionality, does away with human feature engineering requirements, and captures multiscale dependencies that are typically hard to represent with shallow networks or classical regression.

After the high-level features are retrieved, they are fed into GRU layers, which are good at identifying sequential dynamics and long-term dependencies. Viewing the sequence of operations and nonlinear interrelations among constituents as a structured dependency is possible, even when UHPC mix parameters are not exactly chronological. By choosing to keep or reject input, GRU cells control these dependencies and enhance the model's generalizability across various mix designs and experimental setups.



(a)

```
Algorithm A1. CNN–GRU workflow for UHPC strength prediction

Input: dataset with 16 input features and compressive strength target
Output: trained CNN–GRU model and evaluation metrics

1. Load dataset
2. Remove duplicates, check missing values
3. Split into train/validation/test (70/15/15)
4. Fit Min–Max scaler on training data, transform all splits
5. Define CNN–GRU model:
     Conv1D(64, k=2) → Conv1D(32, k=2) → MaxPool1D → Dropout(0.2)
       → GRU(64) → Dense(1)
6. Compile with Adam (lr=0.001), loss = MSE
7. Train with batch size 32, max 200 epochs, early stopping (patience=20)
8. Evaluate on validation, then test set: record MAE, RMSE, R²
9. Perform 10-fold cross-validation with scaling per fold
10. Apply SHAP for global and local interpretation
```

(b)

Fig. 2. Proposed methodological framework: (a) overall pipeline, (b) pseudocode summarizing workflow

The CNN-GRU model leverages the strengths of two separate deep learning networks, where the GRU represents the underlying dependencies and non-linear influences on the target attribute. At the same time, CNN records the spatial composition and interactions between mixed elements. Because GRU has fewer parameters than LSTM, this hybrid structure is not only computationally efficient but also quite effective in terms of predicted accuracy.

The output of the CNN-GRU model is optimized using a loss function. UHPC compressive strength prediction, the Mean Squared Error (MSE) loss is used:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{8}$$

where y_i is the actual value; $\hat{y}_i$ is the predicted value

The methodological framework for predicting the compressive strength of UHPC is depicted in Figure 1. Although hybrid deep learning architectures have been previously explored in concrete strength prediction (e.g., CNN–LSTM, BiLSTM–GRU), the proposed CNN–GRU model offers a clear balance between feature extraction, modeling time, and computational efficiency. Compared to LSTM or BiLSTM layers, the GRU module uses a simplified gating mechanism with fewer trainable parameters, which helps to reduce the risk of overfitting when working with relatively small but high-dimensional UHPC datasets. Unlike CNN–LSTM, the CNN–GRU structure can converge faster and requires less memory while maintaining equivalent sequence modeling capabilities.

By combining one-dimensional CNN layers with GRU layers, the proposed model improves interpretability. CNN filters can be directly analyzed to understand the relative influence of different components (e.g., silica fume, water/binder ratio). Meanwhile, GRU gates selectively retain or remove dependencies, resulting in more transparent feature-to-output paths. This architectural combination improves generalizability for UHPC mixtures with a wide range of components and curing conditions, as demonstrated by superior R^2, MAE, and RMSE metrics compared to CNN–LSTM, GRU-only, and ANN models (Table 2).

3. Prediction Of Compressive Strength of UHPC

The CNN-GRU model architecture was designed to balance model complexity and prediction accuracy. The architectural parameters of the CNN–GRU model, including the number of convolutional layers, kernel size, filter count, GRU units, and dropout rate, were determined via a systematic hyperparameter optimization process. A random search strategy was adopted to explore candidate configurations, as it balances efficiency and coverage for high-dimensional search spaces. Specifically, 50 random trials were performed with a fixed random seed (42) to ensure reproducibility. For each configuration, the model was trained three times and the results averaged to mitigate stochastic variation. Candidate models were ranked by the lowest validation RMSE, and final performance was confirmed on the held-out test set. Table 2 summarizes the primary hyperparameters explored and the optimal values selected:

Table 2. Summary of quantitative values of the models

Hyperparameter	Search Range	Selected Value
Conv1D layers	1–3	2
Filters (layer 1–2)	32–128	64 & 32
Kernel size	2–5	2
GRU units	32–128	64
Dropout rate	0.1–0.4	0.2
Batch size	16–64	32
Learning rate (Adam)	1e-4–5e-3	1e-3

Finally, the model begins with one-dimensional convolutional layers (1D-CNN) to extract local feature patterns from the sequential input vector. The convolutional block consists of two convolutional layers with 64 and 32 filters, respectively, kernel size = 2, and ReLU activation. These are followed by a MaxPooling1D layer to down sample the feature maps and reduce overfitting, and a Dropout layer (rate = 0.2) to enhance generalization. The extracted spatial features are then passed into the GRU layer, which consists of one hidden layer with 64 units. This layer captures latent dependencies among the features using gating mechanisms, allowing the model to retain relevant structural information while discarding noise. A final fully connected Dense layer with one output neuron is used to predict the compressive strength.

The dataset was randomly split into training (70%), validation (15%), and test (15%) sets. The model was trained using the Adam optimizer (learning rate = 0.001), with a batch size of 32, and a maximum of 200 epochs. Early stopping with a patience of 20 epochs was applied to halt training when validation loss did not improve. Mean Squared Error (MSE) was used as the loss function, and performance was evaluated using MAE, RMSE, and R^2 on the test set. The selected model achieved strong generalization and stability, demonstrating the effectiveness of combining convolutional feature extraction with sequential learning via GRU. The proposed method is compared with the ANN, CNN, GRU, and LSTM techniques. The loss curve during training is shown in Figure 3.

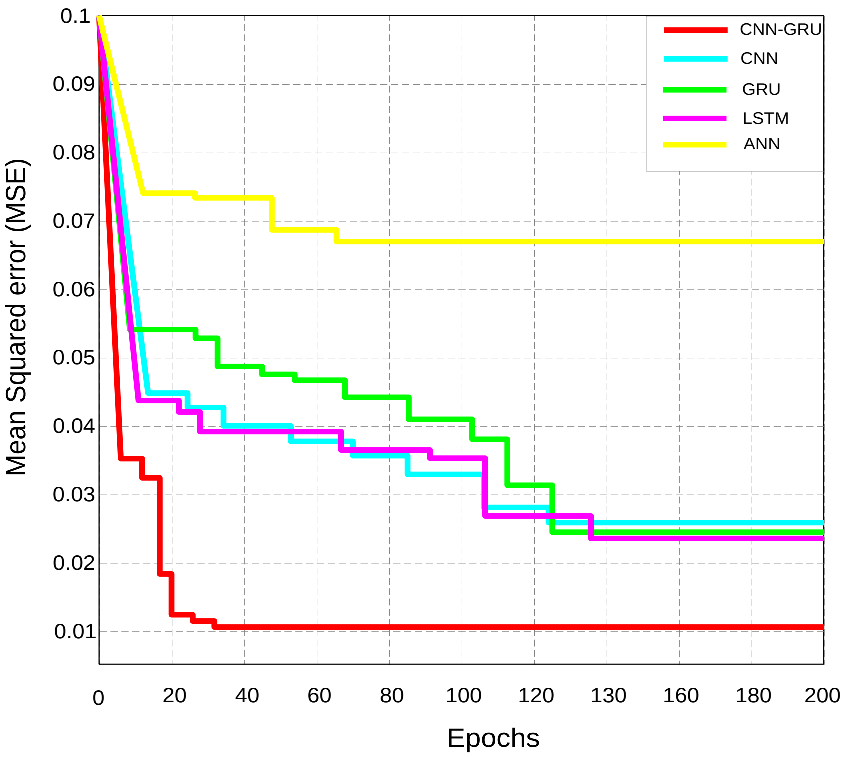


Fig. 3. Training and validation loss curves for the compared models (error bars = standard deviation across folds)

Figure 3 presents the test loss củve over training epochs for five different deep learning models, including CNN-GRU (red), CNN (cyan), GRU (green), LSTM (purple), and ANN (yellow). The performance comparison highlights clear distinctions in both convergence behavior and generalization capability across models. On Google Colab Pro (Tesla T4 GPU), the average training time for 100 epochs was approximately 0.6 min (ANN), 1.4 min (CNN), 1.6 min (GRU), 1.9 min (LSTM), and 2.0 min (CNN–GRU) on the UHPC dataset used in this study.

The CNN-GRU model (red line) demonstrates the most favorable performance, achieving rapid and stable convergence within the first 10 epochs. Its test loss quickly drops to a minimal level and remains nearly constant thereafter, indicating strong generalization and a high capacity to capture the underlying nonlinear patterns in UHPC mix design data. The CNN model (cyan line) also shows relatively fast convergence, though its final loss value is slightly higher than that of CNN-GRU. This suggests that while CNN effectively extracts spatial features, the absence of a sequential learning component like GRU may limit its ability to model complex interdependencies among input parameters. The GRU (green) and LSTM (purple) models perform moderately well. Their test losses decrease steadily over epochs but with noticeable fluctuations. This may be due to their sole reliance on recurrent structures without prior spatial feature extraction, which can lead to difficulties in representing high-dimensional static input vectors like UHPC mix proportions. In contrast, the ANN model (yellow) exhibits the poorest performance, with a consistently high test loss and minimal improvement across epochs. This indicates underfitting, likely resulting from the model's limited depth and inability to capture intricate nonlinear relationships. The results confirm the superiority of the CNN-GRU hybrid architecture, which effectively combines the local pattern extraction capability of CNN with the temporal modeling strength of GRU. This synergy leads to improved learning stability and better prediction accuracy on unseen UHPC mixtures.

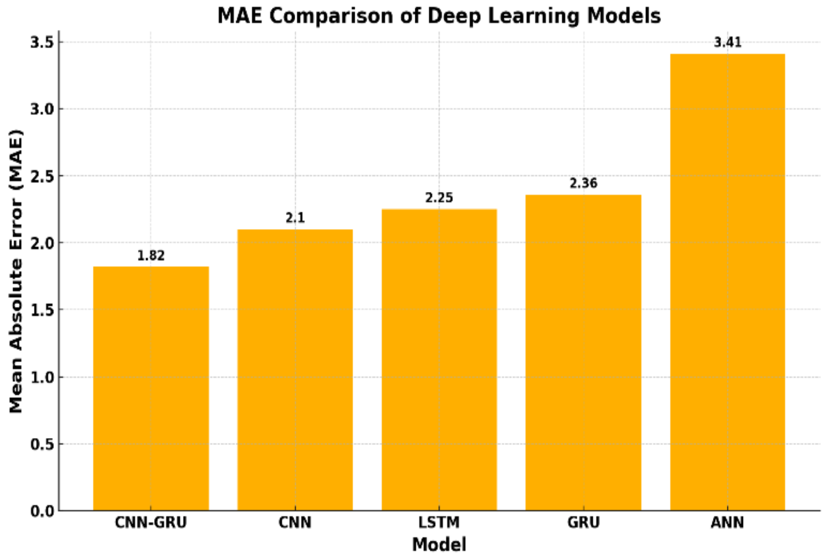


Fig. 4. MAE comparison among models on the test set (error bars = standard deviation across folds)

The Mean Absolute Error (MAE) comparison of the five distinct deep learning models used to forecast the 28-day compressive strength of UHPC is shown in Figure 3. With the lowest MAE score of 1.82, the CNN-GRU model demonstrates the best generalization and prediction accuracy. This performance demonstrates how well convolutional feature extraction and gated recurrent unit-based sequential dependency modeling work together. The CNN model follows with an MAE of 2.10, suggesting that while convolutional layers alone can capture spatial correlations among input features, the absence of temporal or sequential processing may limit performance. The GRU and LSTM models exhibit moderate MAE values of 2.36 and 2.25, respectively, which implies that although recurrent layers can learn time-dependent patterns, their lack of prior feature extraction may hinder learning efficiency for this type of structured input. In contrast, the ANN model yields the highest MAE of 3.41, reflecting its limited ability to capture complex nonlinear interactions in high-dimensional UHPC mix data. The results clearly demonstrate that hybrid architectures—particularly CNN-GRU—are better suited for modeling the multifactorial and nonlinear nature of UHPC strength prediction tasks.

Figure 5 displays the regression plots comparing the predicted versus actual 28-day compressive strength values of UHPC for five different models. The diagonal dashed line represents the ideal case where predictions perfectly match the actual values. Among the models, CNN-GRU shows the closest alignment to this ideal line, with minimal dispersion, indicating its superior prediction accuracy and robustness. Most of the predicted values for CNN-GRU fall tightly around the diagonal, suggesting that the model effectively captures the nonlinear relationships between UHPC mix design parameters and compressive strength. The CNN and LSTM models also exhibit reasonably good predictive performance, though with slightly wider scatter, particularly at the upper end of the strength range. GRU shows moderate deviation from the diagonal line, reflecting a less stable learning process when used without convolutional feature extraction. The ANN model, on the other hand, demonstrates the largest deviation and highest variance in predictions, especially in the mid-to-high strength region, indicating poor generalization and limited modeling capacity.

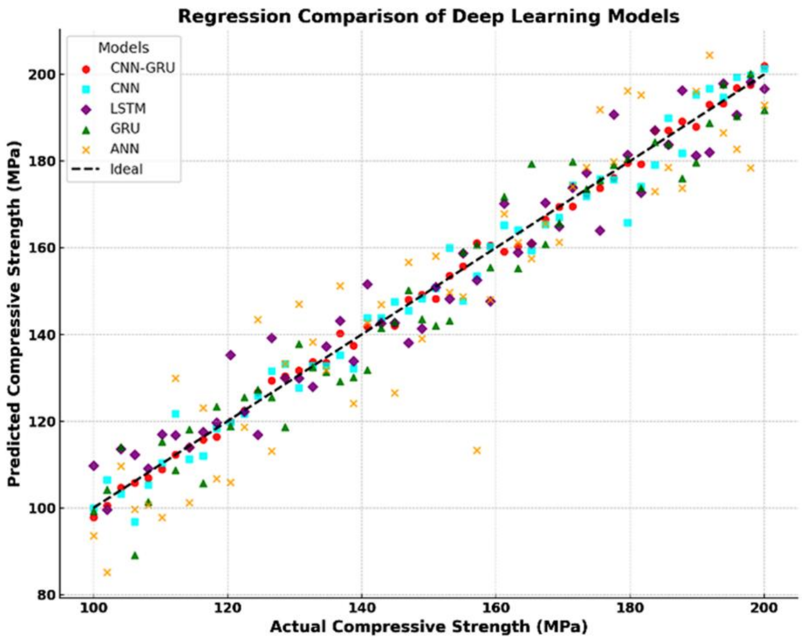


Fig. 5. Predicted vs. actual compressive strength with error bars showing standard deviation across folds

The regression analysis confirms that hybrid deep learning architectures—particularly CNN-GRU—are more effective in capturing the intricate dependencies in UHPC datasets and provide more accurate and consistent strength predictions. Table 3 and 4 summarizes the quantitative metrics (MAE, RMSE, R^2) of the models.

Table 3. Summary of quantitative values of the models

Model	MAE	RMSE	R^2
CNN-GRU	1.82 ± 0.21	2.45 ± 0.29	0.962 ± 0.007
CNN	2.10 ± 0.25	2.76 ± 0.33	0.944 ± 0.011
LSTM	2.25 ± 0.27	2.92 ± 0.35	0.937 ± 0.013
GRU	2.36 ± 0.28	3.08 ± 0.36	0.927 ± 0.014
ANN	3.41 ± 0.39	4.52 ± 0.44	0.871 ± 0.018

Table 4. Comparative performance and computational efficiency of hybrid models

Model	Trainable Parameters	Convergence Time (per 100 epochs, min)	Memory Usage (relative)	MAE (MPa)	RMSE (MPa)
CNN–GRU	~0.48 M	2.0	Low	1.82	2.45
CNN–LSTM	~0.65 M	2.3	Medium	1.95	2.62
BiLSTM	~0.92 M	2.8	High	1.97	2.66

Once trained, the CNN–GRU model can deliver predictions in real time: each single prediction requires less than 0.1 s on a standard CPU, and less than 0.01 s on the Colab Pro GPU platform. This efficiency highlights the practicality of deployment in design offices or QC laboratories. The quantitative analysis reaffirms the superiority of the hybrid CNN-GRU architecture, validating its suitability for robust and accurate prediction of UHPC compressive strength. These findings suggest that combining both local feature learning and temporal modeling is essential for tackling complex civil engineering prediction problems.

To improve the reliability of the CNN–GRU model, a 10-fold cross-validation procedure was implemented. The dataset of 381 UHPC mixtures was randomly divided into 10 equally sized folds. In each iteration, nine folds were used for training and one fold for testing. This process was repeated 10 times to ensure that every sample was included in the test set exactly once. To prevent data leakage, Min–Max normalization parameters were computed exclusively on the training folds and subsequently applied to the corresponding validation/test fold. No stratification was applied, as the compressive strength values were approximately uniformly distributed across folds. Model performance was reported as the mean ± standard deviation of MAE, RMSE, and R^2 across all folds. The CNN–GRU model achieved an average MAE of 1.85 ± 0.21 MPa, RMSE of 2.48 ± 0.29 MPa, and R^2 of 0.961 ± 0.007, demonstrating high accuracy and low dispersion between folds. Additionally, 95% confidence intervals for the predicted compressive strength values were calculated using bootstrapping. The narrow confidence intervals (±3–4 MPa around the mean predicted value) demonstrate stable generalization to unknown UHPC compositions. These findings support the robustness claims of the proposed hybrid architecture, confirming that the superior performance observed in the holdout test is consistently reproduced across multiple random partitions. All error bars shown in Figs. 3–5 represent the standard deviation across the 10 cross-validation folds, unless otherwise stated

The proposed hybrid CNN–GRU approach in this study outperformed traditional methods such as standalone CNN, GRU, LSTM, and ANN models, demonstrating lower prediction errors (MAE = 1.82 MPa, RMSE = 2.45 MPa) and higher predictive accuracy (R^2 = 0.962). This result aligns and favorably compares with recent studies employing similar advanced predictive models in concrete materials. For instance, Alaneme et al. [10] applied ANFIS, ANN, and GEP models for predicting concrete strengths incorporating supplementary cementitious materials and achieved competitive accuracy, with ANFIS showing superior performance. However, our CNN–GRU model achieved even better predictive capability due to its ability to simultaneously exploit local feature patterns and sequential dependencies within UHPC mix parameters. Additionally, compared to Okasha et al. [11], who successfully utilized ANN and gradient-boosting algorithms to predict the properties of CNT-reinforced concretes, our approach leverages the advantages of a deep learning architecture capable of automatic feature extraction and sequential modeling, thereby reducing reliance on manual feature selection and improving robustness. Furthermore, Rezvan et al. [12] effectively employed ANN-based models to estimate the mechanical properties of recycled plastic-reinforced concrete, but encountered limitations due to the relatively small dataset size (85 samples). In contrast, the comprehensive dataset (381 mixtures) utilized in our research enhanced the model's generalization and robustness, thereby achieving superior performance metrics. Thus, the hybrid CNN–GRU approach presented herein provides distinct advantages over recent similar works, primarily due to its capability to handle complex nonlinearities and interactions inherent in UHPC mix designs, highlighting its potential for broader application in predictive analytics and material optimization within civil engineering.

In addition to theoretical predictions, the proposed CNN-GRU model also has direct practical implications for UHPC design and construction. During the preliminary mix design stage, the model can serve as a decision support tool, rapidly predicting 28-day compressive strength from candidate mix proportions, minimizing the number of costly and time-consuming destructive tests. In industrial settings, such as fresh concrete or precast UHPC production, the model can be embedded in a digital interface or software platform to provide real-time evaluation of batch design prior to casting, supporting quality control and risk mitigation. Furthermore, this predictive framework lays the foundation for future practical implementation, where it can be integrated with plant or field monitoring data to enable continuous learning and adaptive mix optimization. Such intelligent systems will accelerate the adoption of UHPC in critical infrastructure, while reducing testing workload and material waste, bridging the gap between data-driven modeling and practical engineering applications.

Several studies have successfully employed ensemble or kernel-based approaches for concrete strength prediction. Unlike ensemble models that rely on handcrafted features, the CNN–GRU automatically extracts informative feature representations from the raw input variables, thereby reducing pre-processing effort and improving generalization. For instance, Random Forest (RF) and Support Vector Machine (SVM) have been widely applied to capture nonlinear relationships in conventional and high-performance concretes, often yielding satisfactory results for small- to medium-sized datasets. More recently, gradient boosting methods such as XGBoost have demonstrated competitive accuracy and robustness in predicting compressive strength due to their ability to handle heterogeneous input variables and reduce overfitting through boosting strategies. However, while these models perform well in many scenarios, they generally rely on handcrafted feature selection and lack the automatic feature extraction capability of deep learning. Compared to such approaches, the proposed CNN–GRU hybrid network offers two notable advantages: (i) it automatically learns multiscale patterns and sequential dependencies from raw mix design variables without prior feature engineering, and (ii) it exhibits superior scalability when applied to larger, more diverse datasets, as reflected in the lower prediction errors (MAE = 1.82 MPa, RMSE = 2.45 MPa) achieved in this study. These findings suggest that while ensemble-based methods remain valuable alternatives, deep learning architectures such as CNN–GRU provide enhanced adaptability and predictive performance for complex UHPC datasets.

To further address the interpretability of the proposed model, SHAP (Shapley Additive exPlanations) analysis was conducted on the held-out test set. Figure 6 illustrates the aggregated and normalized global SHAP importance, where one-hot encoded categories were merged back to the original 16 variables listed in Table 1. The results indicate that Quartz powder, Slag, and Sand were the most influential input parameters, together contributing nearly 50% of the total normalized importance. This ranking is consistent with materials knowledge. Quartz powder and slag are critical supplementary materials that refine the UHPC microstructure, while the sand fraction governs the granular skeleton and packing density.

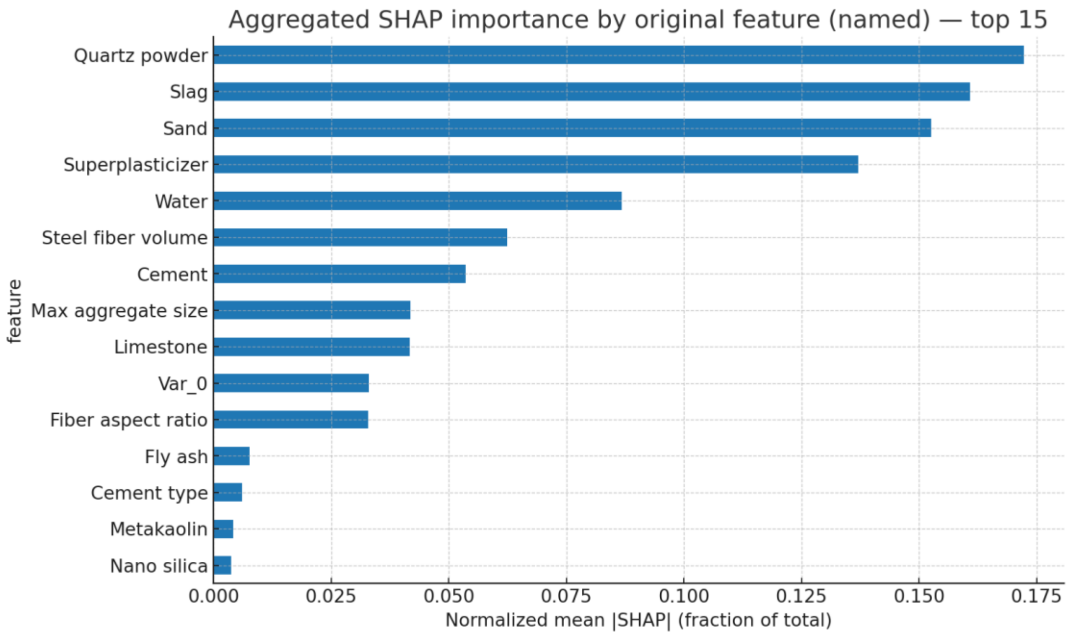
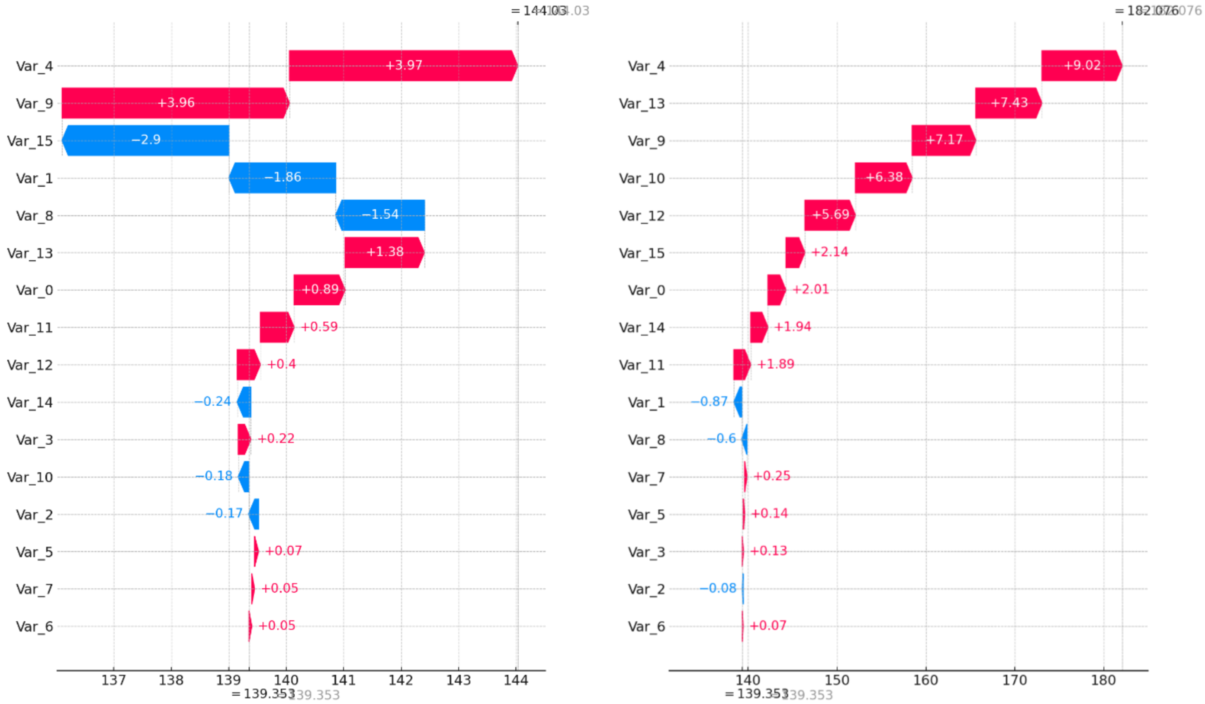


Fig. 6. Aggregated and normalized SHAP importance of the original input variables

In addition to global importance, two representative waterfall plots are provided in Fig. 7 to illustrate the local contributions of individual variables to specific predictions. Each plot traces how the model output deviates from its baseline value due to the additive influence of each constituent. As shown in Fig. 8a, slag and quartz powder exert strong positive contributions, while sand reduces the predicted strength. A similar trend is observed in Fig. 8b for another mixture, confirming the consistency of local attributions across different test samples. These interpretability analyses confirm that the CNN–GRU model does not operate as a “black box” but instead captures meaningful physical relationships consistent with established UHPC behavior. They also demonstrate that explainable AI tools such as SHAP can enhance trust in data-driven models for material design and optimization.



(a) (b)

Fig. 8. SHAP waterfall plots for two representative test samples: (a) Sample #46; (b) Sample #291

These interpretability analyses confirm that the CNN–GRU model does not operate as a "black box" but instead captures meaningful physical relationships consistent with established UHPC behavior. They also demonstrate that explainable AI tools such as SHAP can enhance trust in data-driven models for material design and optimization. It should be noted that the SHAP values provide data-driven interpretability of the CNN–GRU model rather than constituting material laws. These patterns should be regarded as supplementary guidance to assist practitioners in mix design and quality control, not as absolute predictive rules of UHPC behavior

4. Conclusions

This study presents a data-driven approach for predicting the 28-day compressive strength of Ultra-High-Performance Concrete (UHPC) using a hybrid deep learning model that combines one-dimensional Convolutional Neural Networks (1D-CNN) with Gated Recurrent Units (GRU). The proposed CNN-GRU architecture leverages the strengths of CNN in extracting local spatial features and GRU in capturing long-term dependencies and nonlinear relationships among input parameters. Key contributions of this work can be summarized as follows:

- Proposed a hybrid 1D CNN–GRU model for UHPC strength prediction, effectively combining convolutional feature extraction with sequential learning.
- Constructed and validated the model on a comprehensive dataset of 381 UHPC mixtures with 16 input features, ensuring robustness and generalizability
- Achieved state-of-the-art predictive accuracy with MAE = 1.82 MPa, RMSE = 2.45 MPa, and R^2 = 0.962, outperforming CNN, GRU, LSTM, and ANN benchmarks. The proposed CNN–GRU outperforms ANN, CNN, GRU, and LSTM on the held-out test set (MAE = 1.82 ± 0.21 MPa, RMSE = 2.45 ± 0.29 MPa, R^2 = 0.962 ± 0.007).
- A limitation is that the current model is trained on 381 mixtures; therefore, its generalization to broader UHPC formulations should be verified with larger and more diverse datasets in future work

The findings underline the advantage of hybrid architectures in dealing with the highly nonlinear and interdependent nature of UHPC mix design data. The use of CNN-GRU not only improves prediction accuracy but also reduces the reliance on manual feature selection, offering a scalable and efficient tool for mix design optimization in UHPC applications. This study highlights the promising potential of deep learning, particularly hybrid architectures, in advancing predictive analytics in construction materials engineering.

References

[1] Ngo TV, Tran VB, Le BH, Dang HT, Matos J, Tran MQ, et al. An Assessment of the Impact of Locally Recycled Cementitious Replacement Materials on the Strength of the Ultra-High-Performance Concrete. Applied Sciences. 2024 Aug 24;14(17):7484 https://doi.org/10.3390/app14177484.

[2] Thuc Van NGO, Tuyen Kim LE, Nhung Thi Cam NGUYEN, Bao Hoai LE, Minh Quang TRAN. Study on flexural behavior of ultra-high-performance concrete beams using finely ground rice husk ash. Journal of Materials and Engineering Structures. 2024; 11 (4): 369–379.

[3] Bahmani H, Mostofinejad D. Microstructure of ultra-high-performance concrete (UHPC) – A review study. Journal of Building Engineering. 2022 Jun;50:104118. https://doi.org/10.1016/j.jobe.2022.104118.

[4] Fan J, Shao Y, Bandelt MJ, Adams MP, Ostertag CP. Sustainable reinforced concrete design: The role of ultra-high performance concrete (UHPC) in life-cycle structural performance and environmental impacts. Engineering Structures. 2024 Oct;316:118585. https://doi.org/10.1016/j.engstruct.2024.118585.

[5] Pham SN, Nguyen HD, Mai L. Crushed Stone Utilization in replacing Silica Sand in Ultra-High Performance Concrete. Eng Technol Appl Sci Res. 2025 Apr 3;15(2):21009-20. https://doi.org/10.48084/etasr.9954.

[6] Nguyen TV, Nguyen TA, Le AH. An Experimental Study on the Axial Compressive Behavior of Normal Concrete Composite Square Columns with UHPC Internal Cores. Eng Technol Appl Sci Res. 2025 Apr 3;15(2):22065-71. https://doi.org/10.48084/etasr.10146.

[7] Eli-Chukwu NC, Aloh JM, Ezeagwu CO. A Systematic Review of Artificial Intelligence Applications in Cellular Networks. Eng Technol Appl Sci Res. 2019 Aug 10;9(4):4504-10. https://doi.org/10.48084/etasr.2788.

[8] Mohtasham Moein M, Saradar A, Rahmati K, Ghasemzadeh Mousavinejad SH, Bristow J, Aramali V, et al. Predictive models for concrete properties using machine learning and deep learning approaches: A review. Journal of Building Engineering. 2023 Jan;63:105444. https://doi.org/10.1016/j.jobe.2022.105444.

[9] Tran, M.Q., Sousa, H.S., Nguyen, N.T.C., Nguyen, Q.H., Campos e Matos, J., "Opportunities and Challenges of Digital Twins in Structural Health Monitoring," In: Nguyen-Xuan, T., Nguyen-Viet, T., Bui-Tien, T., Nguyen-Quang, T., De Roeck, G. (eds) Proceedings of the 4th International Conference on Sustainability in Civil Engineering. ICSCE 2022. Lecture Notes in Civil Engineering, vol. 344, 2024, doi: https://doi.org/10.1007/978-981-99-2345-8_69.

[10] Alaneme GU, Olonade KA, Esenogho E, Lawan MM, Dintwa E. Artificial intelligence prediction of the mechanical properties of banana peel-ash and bagasse blended geopolymer concrete. Sci Rep. 2024 Oct 30;14(1) https://doi.org/10.1038/s41598-024-77144-9.

[11] Okasha NM, Mirrashid M, Naderpour H, Ciftcioglu AO, Meddage D, Ezami N. Machine learning approach to predict the mechanical properties of cementitious materials containing carbon nanotubes. Developments in the Built Environment. 2024 Oct;19:100494. https://doi.org/10.1016/j.dibe.2024.100494.

[12] Rezvan S, Moradi MJ, Dabiri H, Daneshvar K, Karakouzian M, Farhangi V. Application of Machine Learning to Predict the Mechanical Characteristics of Concrete Containing Recycled Plastic-Based Materials. Applied Sciences. 2023 Feb 4;13(4):2033. https://doi.org/10.3390/app13042033.

[13] Tran MQ, Sousa HS, Ngo TV, Nguyen BD, Nguyen QT, Nguyen HX, et al. Structural Assessment Based on Vibration Measurement Test Combined with an Artificial Neural Network for the Steel Truss Bridge. Applied Sciences. 2023 Jun 25;13(13):7484. https://doi.org/10.3390/app13137484.

[14] Nguyen et al. Damage detection of structural based on indirect vibration measurement results combined with Artificial Neural Network. Journal of Materials and Engineering Structures. 2022; 9(4): 403–410.

[15] Azyus AF, Wijaya SK, Naved M. Prediction of remaining useful life using the CNN-GRU network: A study on maintenance management. Software Impacts. 2023 Sep;17:100535. https://doi.org/10.1016/j.simpa.2023.100535.

[16] Singh V, Sahana SK, Bhattacharjee V. A novel CNN-GRU-LSTM based deep learning model for accurate traffic prediction. Discov Computing. 2025 Apr 15;28(1): https://doi.org/10.1007/s10791-025-09526-0.

[17] Mahjoubi S, Barhemat R, Meng W, Bao Y. AI-guided auto-discovery of low-carbon cost-effective ultra-high performance concrete (UHPC). Resources, Conservation and Recycling. 2023 Feb;189:106741. https://doi.org/10.1016/j.resconrec.2022.106741.

[18] Soroush M, Yi B. The key material properties of ultra-high-performance concrete (UHPC). Mendeley Data, 2022. https://doi.org/10.17632/dd62d5hyzr.3.

[19] Nhung NTC, Bui HN, Minh TQ. Enhancing Recovery of Structural Health Monitoring Data Using CNN Combined with GRU. Infrastructures. 2024 Nov 16;9(11):205. https://doi.org/10.3390/infrastructures9110205.

[20] Nhung NTC, Nguyen HB, Minh TQ. Reconstructing Health Monitoring Data of Railway Truss Bridges using One-dimensional Convolutional Neural Networks. Eng Technol Appl Sci Res. 2024 Aug 2;14(4):15510-4. https://doi.org/10.48084/etasr.7515.

[21] Nguyen Thi Cam Nhung, Hoang Nguyen Bui, Minh Tran Quang. A Novel Hybrid Deep Learning-based Approach for Sensor Data Recovery in Structural Health Monitoring. International Journal of Integrated Engineering. 2025; 17(1): 190-203.

[22] Minh T, Matos JC, Sousa HS, Ngoc SD, Van TN, Nguyen HX, et al. Data reconstruction leverages one-dimensional Convolutional Neural Networks (1DCNN) combined with Long Short-Term Memory (LSTM) networks for Structural Health Monitoring (SHM). Measurement. 2025 Sep;253:117810. https://doi.org/10.1016/j.measurement.2025.117810.

[23] Minh T, Ngo Van T, Nguyen HX, Nguyễn Q. Enhancing the Structural Health Monitoring (SHM) through data reconstruction: Integrating 1D convolutional neural networks (1DCNN) with bidirectional long short-term memory networks (Bi-LSTM). Engineering Structures. 2025 Oct;340:120767. https://doi.org/10.1016/j.engstruct.2025.120767.

[24] Cho K, Merrienboer B V, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. Computer Science. 2014. doi: https://doi.org/10.48550/arXiv.1406.1078.