Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.081794

ARTICLE

Interpretable Deep Learning Framework for Predicting Compressive Strength of Steel Fiber Reinforced Geopolymer Concrete

Quynh-Anh Thi Bui[1,*], Son Hoang Trinh[1], Maryam Sayadi[2] and Reza Khanali[3]

[1]Research group on Industry 4.0 in Transportation (I4T Group), University of Transport Technology, Trieu Khuc, Thanh Liet, Hanoi, Vietnam

[2]Division of Water Resources Engineering, Lund University, Lund, Sweden

[3]Department of Water Resource, University of Tabriz, Tabriz, Iran

*Corresponding Author: Quynh-Anh Thi Bui. Email: quynhanhbt@utt.edu.vn

Received: 09 March 2026; Accepted: 27 April 2026; Published: 27 May 2026

ABSTRACT: Geopolymer concrete has attracted increasing attention as a lower-carbon alternative to ordinary Portland cement concrete because it can utilize aluminosilicate-rich industrial by-products while still achieving satisfactory mechanical performance. However, the 28-day compressive strength of steel fiber-reinforced geopolymer concrete (SFGPC) is governed by multiple interacting mixture variables, which makes reliable prediction difficult, especially for medium-sized experimental datasets. This study developed an interpretable deep-learning framework to predict the 28-day compressive strength (CS28) of SFGPC using an original experimental dataset of 189 mixtures produced under a consistent laboratory protocol in Vietnam. The dataset covered nine mixture variables, including activator chemistry (NaOH, Na_2SiO_3, and Na_2O), binder constituents (Fly Ash (FA), Ground Granulated Blast-Furnace Slag (GGBS), and silica fume (SF), aggregate contents, and steel fiber dosage. Three tabular models, namely Tabular Deep Polynomial Transformer (TabDPT), Tabular Mixer (TabM), and Extreme Gradient Boosting (XGBoost), were trained and evaluated. Among them, TabDPT achieved the best performance on the independent test set, with R^2 = 0.978 and RMSE = 2.214 MPa, while also showing more stable behavior under repeated 5-fold cross-validation than XGBoost. SHapley Additive exPlanations indicated that binder composition and activator chemistry were the dominant variables in the trained model, with GGBS and Na_2O showing the strongest influence, whereas steel fiber dosage had a comparatively smaller contribution to CS28 within the investigated domain. Totally, the proposed framework can support preliminary mixture screening and strength-oriented design of SFGPC within the investigated material and curing domain.

KEYWORDS: Experimental dataset; steel fiber; geopolymer concrete; TabDPT; TabM; SHapley Additive exPlanations (SHAP) interpretation; in-context learning

1 Introduction

In the context of sustainable construction, geopolymer concrete has been widely regarded as a promising alternative to ordinary Portland cement (OPC) concrete because it can utilize industrial by-products, reduce CO_2 emissions associated with cement production, and still provide satisfactory mechanical performance [1,2]. Geopolymers are formed through the alkali activation of aluminosilicate-rich precursors such as fly ash, ground granulated blast-furnace slag (GGBS), metakaolin, and silica fume using alkaline activators, typically sodium hydroxide (NaOH) and sodium silicate (Na_2SiO_3). This process leads to the formation of binding gels and a three-dimensional aluminosilicate network that can provide high strength and good

durability under aggressive environmental conditions [3]. Previous studies have also shown that geopolymer concrete can exhibit favorable resistance to elevated temperature, chemical attack, and weathering compared with conventional OPC-based systems [4,5].

Recent developments in geopolymer technology have extended beyond conventional two-part alkali-activated systems toward one-part binders, hybrid precursor systems, and alternative activator strategies derived from industrial or waste-based sources, aiming to improve practicality and sustainability [6–8]. Nevertheless, conventional NaOH–Na_2SiO_3-based systems remain widely used in laboratory-based strength studies because they provide a relatively well-controlled platform for examining composition–performance relationships [1,2].

Despite these advantages, the compressive strength of geopolymer concrete is influenced by multiple mixture parameters acting simultaneously. The proportions of fly ash and GGBS affect the effective Si/Al and Ca/Si ratios and, therefore, the type and quantity of reaction products formed during geopolymerization [9,10]. The alkaline activator system, including NaOH, Na_2SiO_3, and alkalinity expressed as Na_2O, governs precursor dissolution, polycondensation kinetics, and the relative development of N–A–S–H and C–A–S–H type gels [11,12]. In addition, aggregate proportions and supplementary fine materials such as silica fume influence particle packing, matrix densification, and interfacial bonding [13]. The incorporation of steel fibers may further improve crack resistance and post-cracking behavior, although its contribution to compressive strength is often less direct and depends on dosage, dispersion, and matrix compatibility [14].

The simultaneous presence of these variables, together with their nonlinear interactions, renders mix design and compressive-strength prediction particularly challenging. Traditional linear regression or empirical models, which assume linear and independent relationships, are generally inadequate to capture the nonlinear and interdependent behaviors in geopolymer systems. For example, the beneficial effect of Na_2SiO_3 content often manifests only within specific ranges of NaOH molarity and FA/GGBS ratio; beyond such bounds, the geopolymer gel may develop suboptimally, thereby reducing strength [5,15]. Linear regression typically represents only isolated or linear effects while ignoring interaction effects [2,16]. Several semi-empirical approaches have been proposed [17,18]. However, they remain constrained by pre-assumed functional forms and struggle to reflect the complex relationships between mixture composition and reaction chemistry.

Recent studies have increasingly explored machine learning (ML) techniques to predict the compressive strength of fiber-reinforced cementitious composites, including steel fiber–reinforced concrete, ultra-high-performance fiber-reinforced concrete, and, more recently, fiber-reinforced geopolymer concrete. In conventional steel fiber–reinforced concrete, Li et al. [19] reported that Support Vector Regression (SVR) ensemble with AdaBoost outperformed SVR bagging and standalone SVR for 28-day compressive strength prediction, achieving an R^2 value of 0.96. Pakzad et al. [20] further compared several ML and deep learning models and found that a Convolutional Neural Network (CNN) provided the highest prediction accuracy for SFGPC, with R^2 = 0.928. In the case of ultra-high-performance fiber-reinforced concrete, Abdellatief et al. [21] evaluated five AI models for early-age compressive strength prediction and showed that Gaussian Process regression (GP) and Support Vector Regression (SVR) outperformed Artificial Neural Network (ANN), Random Forest (RF), and Gradient Boosting (GB) models, while also identifying water content, superplasticizer dosage, curing temperature, and fiber content as the most influential variables. Philip and Nidhi [22] investigated fiber-reinforced geopolymer concrete composites using ANN and RF models on a dataset of 110 samples, and reported that random forest performed slightly better than ANN with R^2 = 0.906. Similarly, Hossain et al. [23] applied Gene Expression Programming (GEP) to predict the compressive strength of fiber-reinforced geopolymer concrete using a larger database of 393 experimental samples, incorporating a broad range of mixture, curing, and fiber-related variables.

In parallel, recent advances in machine learning for tabular data have highlighted the growing role of deep neural networks and transformer-based architectures for structured datasets. Recent surveys and benchmark studies have shown that modern deep-learning models can achieve competitive performance with traditional machine-learning methods for tabular data, particularly when nonlinear feature interactions are significant [24–26]. In addition, representation learning approaches have been increasingly explored to improve the modeling of complex tabular relationships [27]. Although classical models such as gradient-boosted decision trees remain strong baselines, recent studies suggest that deep-learning-based approaches can offer improved flexibility and generalization under certain conditions [28]. Given that mixture design data in geopolymer systems are inherently structured and governed by nonlinear interactions among variables, these advances in tabular deep learning provide a relevant framework for investigating strength prediction in SFGPC.

Taken together, the above studies confirm the growing relevance of both experimental–AI integration in fiber-reinforced concrete research and recent advances in tabular foundation models. However, several limitations remain in the current literature. First, compared with the extensive literature on conventional geopolymer concrete, fewer studies appear to have examined CS28 prediction for SFGPC using an internally consistent dataset generated under a single laboratory protocol. A large portion of the existing literature is based on pooled datasets collected from multiple independent studies, which may introduce heterogeneity in raw materials, curing conditions, and testing procedures [19–23]. Second, many existing predictive models rely on pooled datasets collected from multiple independent studies. Although such datasets are useful for enlarging sample size, they may also introduce substantial heterogeneity in raw materials, specimen preparation, curing conditions, and testing procedures, thereby reducing the practical reliability of the resulting models. Third, while model interpretability has improved substantially in recent years, fewer studies have investigated explanatory prediction in controlled SFGPC datasets, especially with respect to the interacting roles of activator chemistry, binder composition, and fiber dosage in strength development. Finally, although mixture-design data are inherently tabular, advanced deep-learning architectures specifically designed for tabular data, such as transformer-based tabular models and parameter-efficient ensemble frameworks, have been less frequently investigated for compressive strength prediction of SFGPC.

To address these gaps, this study develops a predictive framework for the 28-day compressive strength of SFGPC using an original experimental dataset generated under a consistent laboratory protocol in Vietnam. The study has three main objectives. First, an original dataset comprising 189 SFGPC mixtures was established to provide a relatively broad yet internally consistent basis for model development and evaluation. Second, two modern deep-learning architectures, namely the Tabular Deep Polynomial Transformer (TabDPT) and Tabular Mixer (TabM), were developed and compared with XGBoost as a strong tree-based baseline in order to assess their predictive capability and generalization performance on unseen data. Third, SHapley Additive exPlanations (SHAP) were employed to clarify the trained model and to examine the relative importance and nonlinear influence patterns of the input variables with respect to CS28.

The novelty of this study lies in the integrated application of several elements that have rarely been examined together for 28-day compressive-strength prediction of steel fiber-reinforced geopolymer concrete (SFGPC): (i) an original single-source experimental dataset generated under a consistent laboratory protocol, which reduces inter-study heterogeneity; (ii) the application of recent tabular deep-learning models, particularly the in-context-learning-based TabDPT model, to a geopolymer concrete problem; (iii) a multi-level evaluation framework combining hold-out testing, repeated cross-validation, learning-curve analysis, and uncertainty mapping; and (iv) interpretable design-support outputs that visualize both the strength-favorable Na_2O–GGBS region and the associated prediction uncertainty within the investigated design space. Accordingly, the contribution of this work is application-specific and methodological for SFGPC data

analysis, rather than algorithmic in the sense of introducing a new AI architecture. Within this scope, the study aims to provide a more reliable, interpretable, and practically useful framework for strength-oriented mixture screening under an internally consistent experimental setting.

The overall framework of the study consists of four stages. First, experimental data were obtained from laboratory testing of SFGPC mixtures covering activator chemistry, binder composition, aggregate fractions, and steel fiber content. Second, the data were screened, preprocessed, and divided into training and testing subsets, with feature scaling applied when required by the learning algorithm. Third, the models were trained and optimized using cross-validation on the training set. Finally, model performance was evaluated using R^2, RMSE, MAE, and MSE on both cross-validation and independent test results, followed by SHAP-based comprehension of the best-performing model to examine mixture–strength relationships.

2 Experimental Data

2.1 Data Obtained from Laboratory Experiments

2.1.1 Materials and Mix Proportion

All materials were sourced from suppliers commonly available in Vietnam to ensure practical relevance within Vietnamese material supply conditions. Accordingly, the findings are most representative for similar Low-Ca fly ash-based systems and comparable curing/testing conditions. The geopolymer mixture comprised fly ash (FA), ground granulated blast-furnace slag (GGBS), alkaline activators (NaOH and Na_2SiO_3), fine and coarse aggregates, silica fume, and dispersed steel fibers.

Fly Ash (FA)

FA was obtained from the Pha Lai Thermal Power Plant (Vietnam) and classified as Class F according to ASTM C618 (i.e., CaO < 10%). With SiO_2 = 52.8%, Al_2O_3 = 27.1%, Fe_2O_3 = 9.5%, and CaO = 5.7%, and the sum of SiO_2 + Al_2O_3 + Fe_2O_3 was about 89%, meeting ASTM C618-19a requirements for Class-F fly ash used in geopolymers. The median particle size was d50 ≈ 12 μm; Class F fly ash is typically reported to contain predominantly spherical particles, although particle morphology (e.g., SEM) was not explicitly quantified in this study, providing a reactive aluminosilicate backbone under alkali activation.

Ground Granulated Blast-Furnace Slag (GGBS)

GGBS was supplied by the Hoa Phat Dung Quat steelworks and ground to a Blaine fineness of about 4300 cm^2/g. Typical chemistry include of CaO = 41.2%, SiO_2 = 33.6%, Al_2O_3 = 13.2%, MgO = 7.3%.

Alkali Activator Solution

The activator comprised NaOH and Na_2SiO_3. NaOH (sodium hydroxide): prepared by dissolving 98%-purity NaOH pellets in deionized water to reach the target molarity (6–14 M). The solutions were made 24 h in advance to stabilize temperature. Na_2SiO_3 (sodium silicate): commercial sodium silicate solution with silicate modulus Ms = SiO_2/Na_2O = 2.5 and solids content of about 40%. The mass ratio Na_2SiO_3/NaOH was set between 1.5 and 2.5 depending on the mix design. The total alkali content, expressed as Na_2O-equivalent (%Na_2O). In practical geopolymer mix design, Na_2O content is often selected within a moderate range (commonly around 3%–7% by binder mass) to balance reactivity and workability. In the present dataset, Na_2O ranged from 3 to 6.7% depending on mixture.

Aggregates (Fine Aggregate (FiAg) and Coarse Aggregate (CoAg))

Fine aggregate (FiAg) was Red River natural sand (FM ≈ 2.5), clean and free of organics. Coarse aggregate (CoAg) was crushed basalt (5–10 mm). For the conventional geopolymer concrete mixtures containing coarse aggregate, the fine-to-coarse aggregate proportion was selected within a typical practical range (approximately 40%–60% by total aggregate mass), consistent with common mix-design practice. In addition, the experimental program also included fine-grained geopolymer concrete mixtures in which coarse aggregate was intentionally omitted (CoAg = 0).

Silica Fume (SF)

Silica fume (SF) was used as a mineral additive to improve particle packing and reduce capillary porosity; its dosage was selected within a commonly adopted range in geopolymer concrete practice (typically around 3%–8% of binder solids).

Steel Fibers

Straight steel fibers were used in this study because they were the fibers employed in the underlying experimental program. Although hooked-end fibers may provide improved mechanical anchorage and pull-out resistance, straight fibers offer advantages in terms of dispersion, mixing stability, and experimental control. The nominal properties of the fibers used were as follows: length 13 mm, diameter 0.2 mm, aspect ratio 65, and tensile strength approximately 1100 MPa. The fiber was incorporated at a volume fraction of 0%–1.0%.

2.1.2 Sample Preparation and Testing

Sample Preparation

The experimental procedures were conducted in accordance with ASTM C192/C192M. Sodium hydroxide (NaOH) and sodium silicate (Na_2SiO_3) solutions were prepared in advance and allowed to cool and equilibrate for 24 h. Fly ash (FA), ground granulated blast-furnace slag (GGBS), and silica fume were dry-mixed for 2 min to ensure uniform dispersion. The alkaline activator solution was then added and mixed for 3 min until a homogeneous binder paste was obtained. Subsequently, fine and coarse aggregates were introduced, followed by the gradual addition of steel fibers, and mixing continued for an additional 2 min to achieve a consistent geopolymer concrete mixture.

The fresh mixtures were cast into steel cylindrical molds of 150 × 300 mm (diameter × height) in two layers. Each layer was compacted on a vibrating table for 15 ± 5 s to remove entrapped air while avoiding segregation, and the top surface was finished with a trowel. Immediately after casting, specimens were ambient curing at 27 ± 2°C until testing at 28 days.

Compressive Strength Testing

The 28-day compressive strength was measured using a hydraulic compression testing machine in accordance with ASTM C39/C39M for cylindrical specimens. The loading rate was maintained at 0.25 ± 0.05 MPa/s. Some photos of the specimen preparation and testing procedures are presented in Fig. 1.

Figure 1: Specimen preparation and testing.

2.2 Data Preprocessing

For model development, all mixture constituents were expressed on a volumetric basis (kg/m^3). 'NaOH' and 'Na_2SiO_3' denote the masses of the corresponding alkaline solutions per cubic meter of mixture. The Na_2O content (%) was computed from the NaOH and sodium silicate solutions and normalized by the total binder mass (FA + GGBS + SF). 'Fiber' and 'SF' represent their dosages in kg/m^3. Although NaOH solution molarity ranged from 6 to 14 M in the experimental program, molarity was not used as a direct input feature because the model inputs were standardized in kg/m^3.

Raw data were first screened for potential outliers using descriptive statistics and diagnostic plots, including histograms, boxplots, and Q–Q plots. Feature scaling was then applied prior to model training to improve comparability among variables and enhance numerical stability. Min–Max scaling was selected because the input variables spanned markedly different units and numerical ranges, for example Na2O (%) vs. aggregate contents (kg/m^3), and because the deep tabular models considered in this study can be sensitive to feature magnitude. Moreover, several variables were defined at discrete experimental levels rather than following approximately Gaussian distributions. Under these conditions, Min–Max scaling provides a straightforward bounded transformation that preserves the relative ordering of the original design levels while improving numerical stability [8]. Although standardization was also a possible alternative, Min–Max scaling was considered more appropriate for maintaining a consistent preprocessing pipeline across the three models. To prevent data leakage, scaling parameters were estimated using the training folds only and then applied to the corresponding validation or test data.

Feature scaling is generally not required for tree-boosting methods such as XGBoost, but it is beneficial for deep tabular models such as TabDPT and TabM, where optimization can be sensitive to input magnitudes. Therefore, a consistent scaling pipeline was adopted for all models to maintain a fair comparison under identical preprocessing conditions. After preprocessing, the dataset was split into train/test = 80/20 using randomized shuffling with a fixed seed for reproducibility. The 80% training set was used for (i) feature selection (if any), (ii) hyperparameter tuning via GridSearchCV, and (iii) 5-fold cross-validation (k = 5). The 20% test set was held out and used once at the end to estimate generalization performance. All stochastic

procedures (data splitting, model initialization) used a fixed random seed (42). Configurations (80/20 split, k = 5, significance level α, and scaler choice) were kept consistent across the three models (TabDPT, TabM, and XGBoost) to ensure fairness and reproducibility.

The correlation matrix in Fig. 2 shows that many input variables exhibit low-to-moderate pairwise correlations, suggesting that the dataset retains substantial informational diversity across mixture components. The linear correlations between individual predictors and CS28 are generally modest, indicating that compressive strength is unlikely to be explained by any single variable alone and may instead depend on combined and non-linear effects among constituents. Relatively strong correlations are observed mainly within a few chemically or compositionally related variable groups, including Na_2SiO_3–NaOH–Na_2O, FA–GGBS, and FiAg–CoAg. These patterns are expected from the mixture design and underlying material relationships. Thus, the dataset does not appear to exhibit pervasive linear redundancy across all predictors, although several correlated variable clusters remain and should be considered when explaining model behavior and feature importance.

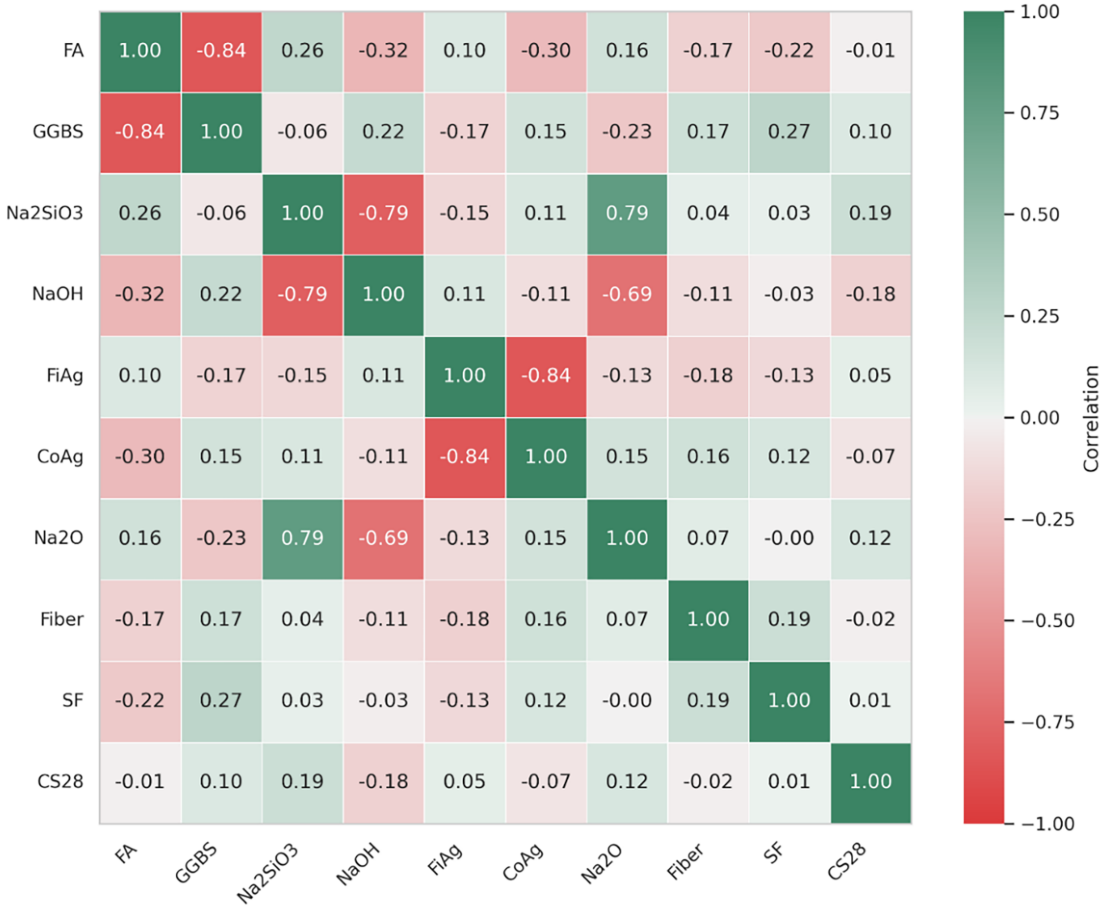

Figure 2: Correlation matrix of parameters.

The distributions of the input variables and the 28-day compressive strength (CS28) are shown in Fig. 3, while their descriptive statistics are summarized in Table 1. The dataset covers broad composition ranges across binder constituents, activator contents, aggregate fractions, and additive dosages. Several variables exhibit discrete or multi-regime distribution patterns, which is consistent with the use of predefined experimental mixture levels rather than unconstrained random sampling.

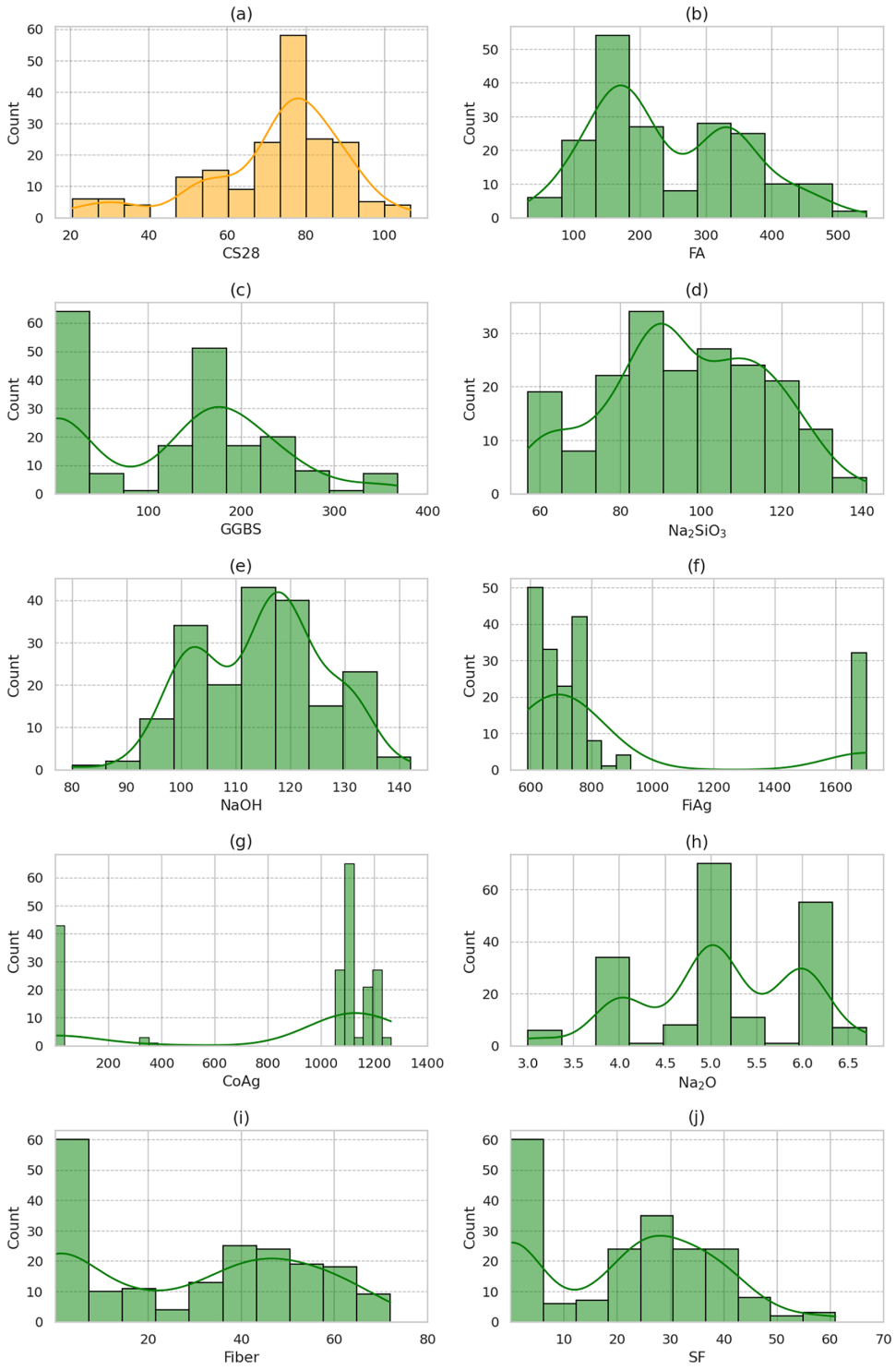

Figure 3: Distribution of input and output variables: (**a**) CS28—28-day compressive strength (MPa); (**b**) FA—fly ash content (kg/m^3); (**c**) GGBS—ground granulated blast furnace slag content (kg/m^3); (**d**) Na_2SiO_3—sodium silicate content (kg/m^3); (**e**) NaOH—sodium hydroxide content (kg/m^3); (**f**) FiAg—fine aggregate content (kg/m^3); (**g**) CoAg—coarse aggregate content (kg/m^3); (**h**) Na_2O—sodium oxide content (%); (**i**) Fiber—steel fiber content (kg/m^3); (**j**) SF—silica fume content (kg/m^3).

Table 1: Statistical summary of dataset.

Type	Unit	Count	Mean	Std	Min	Q1 (25%)	Median	Q3 (75%)	Max
FA	kg/m^3	189	241.82	110.56	30.0	161.00	207.00	334.00	543.0
GGBS	kg/m^3	189	127.23	102.68	0.0	0.00	150.00	201.00	368.0
Na_2SiO_3	kg/m^3	189	96.16	19.55	57.0	82.00	93.00	111.00	141.0
NaOH	kg/m^3	189	114.54	11.45	80.0	104.00	116.00	122.00	142.0
FiAg	kg/m^3	189	862.71	379.68	591.0	637.00	729.00	779.00	1700.0
CoAg	kg/m^3	189	862.09	478.06	0.0	1057.00	1100.00	1160.00	1262.0
Na_2O	%	189	5.12	0.83	3.0	4.70	5.00	6.00	6.7
Fiber	kg/m^3	189	29.46	24.01	0.0	0.00	34.00	50.00	72.0
SF	kg/m^3	189	20.97	16.51	0.0	0.00	24.00	34.00	61.0
CS28	MPa	189	72.10	17.72	20.5	64.92	76.92	84.45	106.6

The target variable, CS28, spans from 20.5 to 106.6 MPa, with a median of 76.92 MPa and an interquartile range from 64.92 to 84.45 MPa. Most observations are concentrated in the moderate-to-high strength range, particularly around 70–85 MPa, while a smaller number of lower-strength mixtures and a limited upper tail above 100 MPa are also present. This distribution provides a reasonably broad response range for regression analysis within the investigated material domain.

Among the binder-related variables, FA shows a broad spread across the dataset, whereas GGBS presents a multi-regime pattern, including mixtures without slag addition and mixtures with moderate-to-high slag contents. This indicates that the experimental program covered both FA-dominant and FA–GGBS blended binder systems.

The activator-related variables also reflect structured design levels. Na_2SiO_3 is concentrated mainly in the intermediate range, while NaOH is distributed over a relatively broad but still discretized range. Na_2O, expressed as a percentage of binder mass, varies from 3.0% to 6.7% and shows visible clustering around several alkalinity bands, especially near 4%, 5%, and 6%.

For the aggregate system, FiAg exhibits a markedly right-skewed distribution, with most mixtures concentrated at lower-to-moderate values and a smaller number of mixtures at substantially higher sand contents. CoAg shows a clear two-regime structure, with one subset corresponding to fine-grained geopolymer concrete mixtures without coarse aggregate (CoAg = 0) and the remaining mixtures concentrated mainly in the range of approximately 1050–1260 kg/m^3. This confirms that the dataset intentionally includes both conventional geopolymer concrete and fine-grained geopolymer concrete.

Finally, Fiber and SF both display mixed distributions with a noticeable proportion of zero values together with broader ranges at moderate dosages, indicating that the experimental design included mixtures both with and without these constituents. Taken together, these distribution patterns suggest that the dataset captures substantial design-driven variability and provides a suitable basis for analyzing multivariable mixture–strength relationships.

3 Machine Learning Framework

3.1 TabM Model

TabM is a deep learning architecture specifically designed for tabular data that enhances a standard multilayer perceptron (MLP) through parameter-efficient ensembling. Instead of training multiple independent neural networks, TabM embeds ensemble behavior within a single model using shared parameters,

which improves generalization performance while maintaining a relatively low computational cost [29]. In practice, the model learns several parallel predictive representations on top of a shared feature extractor, thereby combining the diversity advantage of ensembling with the efficiency of a single network. This mechanism makes TabM particularly suitable for tabular regression problems with nonlinear relationships and limited-to-moderate dataset sizes [30].

In this study, TabM was implemented for regression to predict CS28. The normalized input features were fed into the TabM network, and the model was trained by minimizing the mean squared error loss. Key hyperparameters, including the number of hidden layers, hidden dimensions, dropout rate, learning rate, and ensemble size, were optimized using 5-fold cross-validation on the training set. The optimal configuration was selected based on the average cross-validated error and subsequently retrained on the full training dataset before final evaluation on the test set. Fig. 4 illustrates TabM architecture for ensemble learning with k members. Each member model (MLP) shares weights across the ensemble while having its own member-specific parameters and output head. During inference, the model outputs k individual predictions per sample, and the final prediction is the mean of these outputs. The training loss is computed using the mean squared error across all ensemble members.

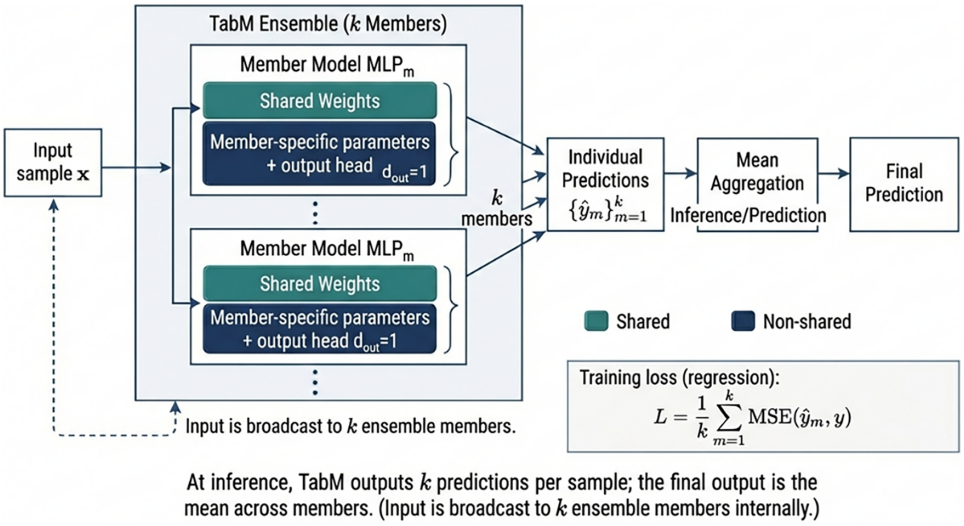

Figure 4: Illustration of the TabM model.

3.2 TabDPT Model

TabDPT is a transformer-based foundation model for tabular data that operates under an in-context learning (ICL) paradigm. Unlike conventional neural networks that require explicit retraining or fine-tuning for each dataset, TabDPT adapts to a given task by conditioning its predictions on a subset of training samples provided as contextual information during inference [31]. In this framework, the model receives both the query sample and a set of context examples drawn from the training data, and it infers the target output by learning the relationship between inputs and outputs within the provided context. This design enables TabDPT to exploit similarities among samples and to capture complex nonlinear dependencies without relying solely on fixed parameter updates for each specific dataset [31].

Architecturally, TabDPT extends the transformer mechanism to structured tabular regression by combining contextual encoding with nonlinear feature interactions. Through its attention-based structure, the model can identify which contextual samples and which input dimensions are most informative for

the prediction of a given query instance. In addition, the polynomial interaction component enhances the model's ability to represent higher-order relationships among variables, which is particularly relevant for material datasets where strength development is governed by coupled effects among binder composition, activator chemistry, aggregate proportions, and fiber content. Owing to these characteristics, TabDPT is well suited for medium-sized experimental tabular datasets in which the relationships between inputs and outputs are nonlinear, heterogeneous, and partially interaction-driven.

In the present work, TabDPT was employed as a regression model for predicting CS28. The input features were first normalized using Min–Max scaling, consistent with the preprocessing applied to other models. Model performance primarily depends on inference-related parameters, notably the context size (number of samples used as context) and the number of ensemble predictions. These parameters were tuned using 5-fold cross-validation on the training data. The optimized configuration was then used to generate predictions for the unseen test set. No explicit fine-tuning of the pretrained model weights was performed, in line with the standard usage of TabDPT as a tabular foundation model. Fig. 5 illustrates general architecture and workflow of the TabDPT model used in this study, including data preprocessing, train–test splitting, five-fold cross-validation, in-context learning–based inference, and performance evaluation using standard regression metrics.

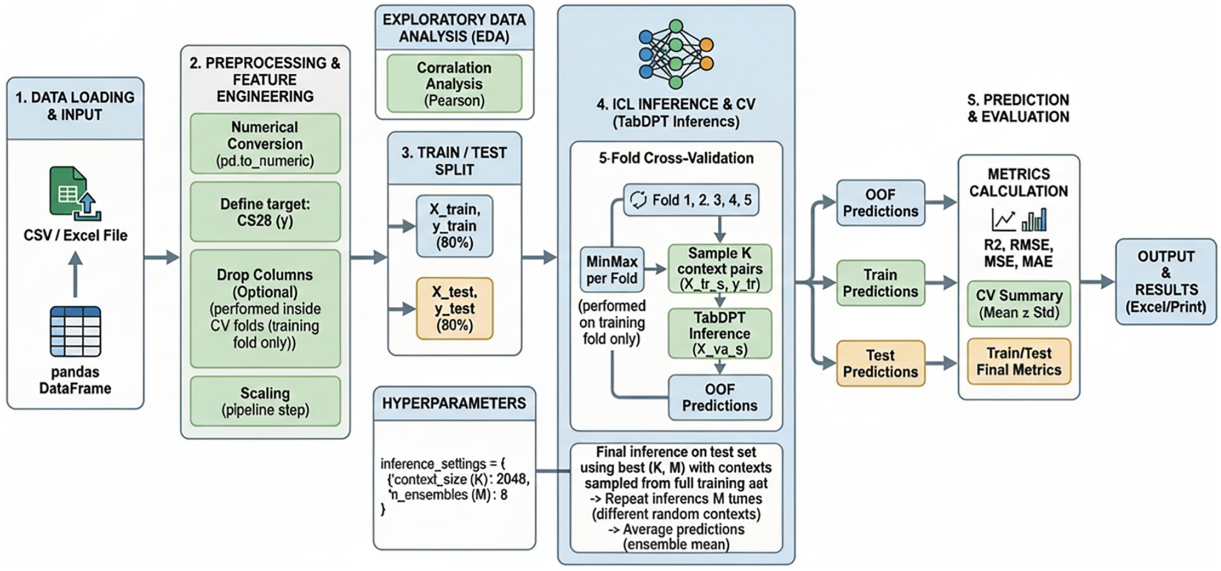

Figure 5: Illustration of the TabDPT model.

3.3 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is a regularized implementation of gradient-boosted decision trees that builds an additive predictive model in a stage-wise manner [32,33]. In this framework, trees are added sequentially, and each new tree is trained to reduce the residual error of the current ensemble. Owing to this boosting strategy, XGBoost can effectively capture nonlinear relationships and high-order interactions among input variables without requiring explicit assumptions about the underlying functional form.

For regression tasks, XGBoost minimizes an objective function that combines a loss term with a regularization term to control model complexity. Its performance is influenced by several key hyperparameters, including the number of trees (n_estimators), maximum tree depth (max_depth), learning rate, subsampling ratios (subsample, colsample_bytree), and regularization terms such as reg_alpha, reg_lambda,

min_child_weight, and gamma. In this study, these hyperparameters were optimized using grid search with 5-fold cross-validation on the training set (as shown in Fig. 6). XGBoost was selected as a strong tree-based baseline because it has shown competitive performance in predicting the mechanical properties of cementitious and geopolymer materials under multivariable nonlinear conditions [34].

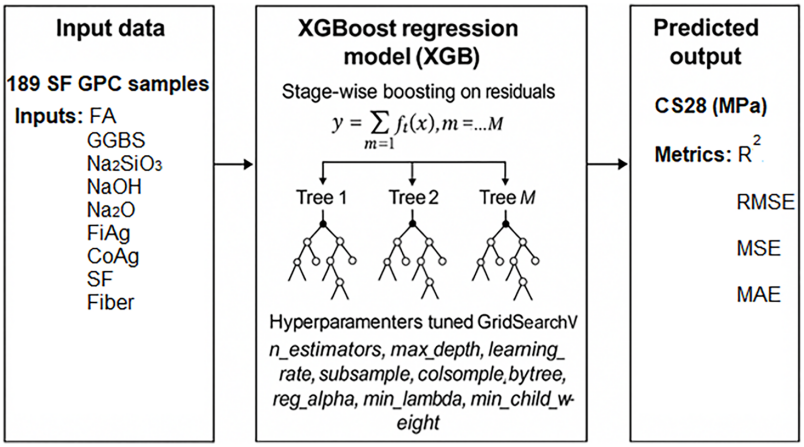

Figure 6: Illustration of the extreme gradient boosting model.

3.4 GridSearchCV

To optimize the hyperparameters of each model in a systematic and reproducible manner, GridSearchCV was applied to the training set [35,36]. For each model, a discrete search space was defined, and all candidate hyperparameter combinations were evaluated using 5-fold cross-validation within the training subset. The root mean square error (RMSE) was used as the primary optimization criterion, while R^2 and MAE were also recorded for reference. The hyperparameter search was conducted in two stages. First, a coarse grid was used to identify promising regions of the search space. Second, a refined grid was constructed around the best-performing configurations obtained from the initial search. This two-stage procedure was adopted to improve search efficiency while retaining adequate coverage of the candidate hyperparameter domain.

To avoid data leakage, all preprocessing operations required by a given model, such as scaling or transformation, were performed separately within each cross-validation fold using only the corresponding training partition. After the optimal hyperparameter configuration had been identified, the model was refitted using the entire training set and was then carried forward for subsequent comparison and evaluation on the independent test set. In this study, TabDPT, TabM, and XGBoost were optimized independently using this procedure.

3.5 Cross-Validation

Cross-validation was used to assess model stability and to reduce sensitivity to a single random data split. In this study, the full dataset of 189 samples was first divided into training and independent testing subsets using a fixed random seed. Hyperparameter optimization was then conducted only on the training subset, while the independent test subset was reserved exclusively for final performance evaluation. Within the training subset, 5-fold cross-validation was employed during model development. In each iteration, the training data were partitioned into five folds, of which four were used for model fitting and one for validation. This process was repeated until each fold had served once as the validation subset (Fig. 7). The resulting

performance metrics were aggregated across folds to provide a more reliable estimate of model behavior during training.

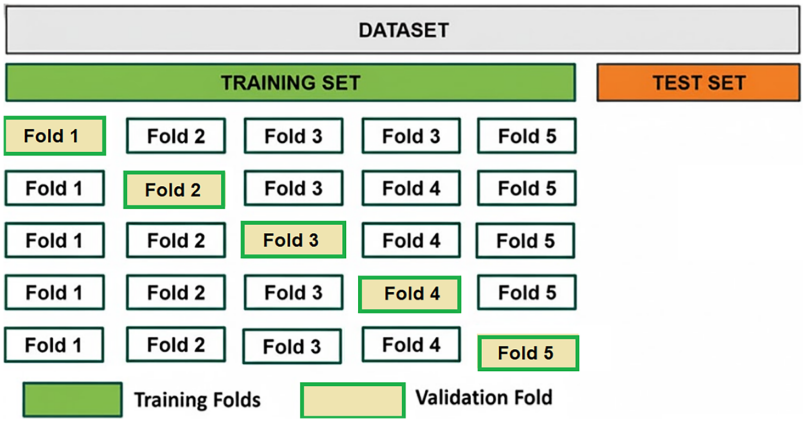

Figure 7: Schematic of the CV algorithm.

In addition to the single 5-fold procedure used for hyperparameter tuning, repeated 5-fold cross-validation with multiple random seeds was further conducted to assess robustness. This additional analysis was intended to examine the consistency of model performance under different fold assignments, which is particularly relevant for medium-sized experimental datasets that may contain inherent material and testing variability [37]. The final selected models were then retrained on the full training subset and evaluated on the independent test subset to assess generalization performance.

3.6 SHAP

To interpret the trained models and quantify the contribution of each input variable to the predicted 28-day compressive strength, SHAP (SHapley Additive exPlanations) was employed as a post hoc explainability method [38]. SHAP is based on the concept of Shapley values from cooperative game theory, where each feature is treated as a contributor to the final prediction. The SHAP value of a feature represents its marginal contribution to the prediction, and the sum of all SHAP values equals the model output relative to a reference baseline.

In this study, SHAP summary plots were used to evaluate global feature importance, whereas SHAP dependence plots were used to examine the direction and nonlinear response pattern of individual variables across the investigated dataset. These plots were used to identify which input variables were most strongly associated with variations in predicted compressive strength and to visualize how their effects changed over different value ranges. For the deep tabular models, namely TabDPT and TabM, DeepSHAP was used to explicate the learned relationships. For XGBoost, standard tree-based SHAP values were calculated. SHAP provides an interpretable description of model behavior within the investigated dataset; however, the inferred feature effects should be understood as model-based associations rather than direct proof of causal material mechanisms.

3.7 Model Evaluation

Statistical indicators

To evaluate the predictive accuracy and generalization performance of the developed models, four standard statistical indicators were employed: the coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). These metrics provide complementary information on the agreement between predicted and measured compressive strength values [21,36,39].

R^2 measures the proportion of variance in the observed data explained by the model, with values closer to 1 indicating better predictive agreement. MSE and RMSE quantify the average squared deviation between predictions and observations, whereas RMSE is more directly clarifiable because it is expressed in the same unit as the target variable (MPa). MAE represents the average absolute deviation and is less influenced by occasional large errors than MSE or RMSE. In this study, model performance was primarily evaluated using cross-validation results on the training set and was subsequently verified on the independent test set for final generalization assessment.

Learning curve analysis

Learning-curve analysis was conducted as a supplementary diagnostic tool to examine model convergence behavior and to identify possible underfitting or overfitting. A learning curve describes how model performance changes as the number of training samples increases by comparing training error with validation error estimated under the same training protocol. When the two curves gradually converge toward stable values, the model can be considered to exhibit more consistent generalization. In contrast, a persistently large gap between training and validation performance may indicate overfitting [40].

In this study, learning curves were constructed for the best-performing model, TabDPT, by progressively increasing the training subset size and estimating validation performance using the same cross-validation setting adopted in the main experiments. Both R^2 and RMSE were used to track convergence trends and error behavior.

Taylor diagram

A Taylor diagram was used as an additional graphical tool to compare the predictive performance of different models in a compact and integrated manner. This diagram simultaneously summarizes three complementary statistics: the correlation coefficient, the centered root means square error, and the standard deviation of model predictions relative to the observed data. As a result, it provides a concise visual assessment of how closely each model reproduces the variability and pattern of the experimental results [41].

In the Taylor diagram, the reference point represents the observed data, and the relative position of each model indicates its similarity to that reference in terms of correlation, dispersion, and error. Models located closer to the reference point generally show better agreement with the measured values. In this study, the Taylor diagram was used as a comparative visualization tool to support the interpretation of model performance beyond conventional scalar metrics.

4 Results and Discussion

4.1 Model Performance Comparison

Table 2 summarizes the hyperparameter search spaces for the three models considered in this study TabM, TabDPT, and XGBoost used to identify optimal configurations under the selected evaluation criteria. For TabM, the search includes architectural and optimization parameters (e.g., number of blocks, block width, dropout, learning rate, batch size, and training epochs). For TabDPT, the search focuses on inference and ensemble-related settings (e.g., context size and the number of ensembles) as well as implementation

options (e.g., Flash attention usage). For XGBoost, key boosting parameters controlling model capacity and regularization are explored, including tree depth, number of estimators, learning rate, subsampling, column subsampling, minimum child weight, gamma, and L1/L2 regularization terms (reg_alpha and reg_lambda). All models are tuned using cross-validation on the training set, and the best configuration is refit before final evaluation on the held-out test set.

Table 2: Hyperparameter search grid.

TabM		TabDPT		XGBoost	
batch_size	32, 64, 128	compile	FALSE	colsample_bytree	0.8, 1.0
d_block	128, 256, 384	context_size	512, 1024, 2048	gamma	0, 0.1, 0.3
dropout	0.0, 0.1, 0.2	n_ensembles	1, 2, 4, 8	learning_rate	0.1, 0.2, 0.03
epochs	150, 250, 350	seed_predict	0, 1, 2, 42	max_depth	3, 5, 7
eval_interval	25	use_flash	TRUE	min_child_weight	1, 3, 5
grad_clip	0.5, 1.0, 2.0	–	–	n_estimators	200, 400, 600
k_members	32, 16, 8	–	–	reg_alpha	0, 0.1, 1
lr	5e−4, 1e−3, 2e−3	–	–	reg_lambda	1, 5, 10
n_blocks	2, 3, 4	–	–	subsample	0.8, 1.0
weight_decay	0, 1e−3, 1e−2, 5e−2	–	–	–	–

Table 3 summarizes the optimized hyperparameter settings obtained for TabM, TabDPT, and XGBoost after the tuning process. Each configuration reflects the learning characteristics and architectural requirements of the corresponding model, aiming to balance predictive accuracy with generalization stability. For the TabM model, the selected hyperparameters indicate a moderately sized architecture with controlled capacity. A block dimension of 128 combined with three blocks provides sufficient representational power to capture nonlinear interactions among mixture variables without introducing excessive complexity. The batch size of 32 and a learning rate of 0.001 support stable optimization, while gradient clipping 1 reduces the risk of unstable updates during training. The selected dropout setting (0.0) suggests that regularization is achieved mainly through architecture and training control rather than stochastic feature dropping, which can be advantageous for relatively small experimental datasets.

Table 3: Optimized hyperparameter configurations.

TabM		TabDPT		XGBoost	
batch_size	32	compile	FALSE	colsample_bytree	1
d_block	128	context_size	512	gamma	0
dropout	0	n_ensembles	4	learning_rate	0.1
epochs	250	seed_predict	42	max_depth	5
eval_interval	25	use_flash	TRUE	min_child_weight	1
grad_clip	1	–	–	n_estimators	200
k_members	16	–	–	reg_alpha	0.1
lr	0.001	–	–	reg_lambda	1
n_blocks	3	–	–	subsample	0.8
weight_decay	0.001	–	–	–	–

The optimized configuration of TabDPT emphasizes contextual representation and variance reduction through ensembling. A context size of 512 enables richer feature conditioning, while using four ensemble members improves robustness by averaging multiple predictions. The fixed prediction seed ensures reproducibility across runs. Enabling Flash computation (use_flash = True) primarily improves computational efficiency and does not change the objective function, while other settings (e.g., compile) follow the implementation choices used in this work. For XGBoost, the optimized hyperparameters correspond to a tree-based model with moderate depth and controlled complexity. A maximum depth of 5 and 200 estimators' captures nonlinear patterns while mitigating overfitting risk associated with overly deep trees. The learning rate of 0.1 provides efficient convergence, and subsampling (0.8) introduces beneficial randomness that can improve generalization. Regularization via reg_alpha (0.1) and reg_lambda (1) further constrains model complexity. Additional parameters (e.g., gamma = 0 and colsample_bytree = 1) indicate that split penalties were not increased and that all features were considered at each tree, which is reasonable for datasets with a limited number of predictors.

Table 4 compares the predictive performance of the tuned TabDPT, TabM, and XGBoost models for estimating the 28-day compressive strength (CS28) on both the training and independent test sets. The use of multiple evaluation metrics, including R^2, RMSE, MSE, and MAE, provides a more comprehensive assessment of model accuracy and generalization behavior. On the training set, XGBoost achieved the strongest fit, with the highest R^2 (0.999) and the lowest error values (RMSE = 0.369 MPa, MAE = 0.264 MPa), indicating a very strong fit to the training data. TabDPT also showed strong training performance (R^2 = 0.997, RMSE = 0.982 MPa, MAE = 0.727 MPa), whereas TabM exhibited a more conservative fit with higher training errors (R^2 = 0.984, RMSE = 2.136 MPa, MAE = 1.646 MPa). More meaningful differences emerged on the independent test set. TabDPT achieved the best generalization performance, with the highest test R^2 (0.978) and the lowest RMSE (2.214 MPa) and MAE (1.806 MPa). TabM ranked second, with a test R^2 of 0.976, RMSE of 2.376 MPa, and MAE of 1.923 MPa. Although XGBoost still maintained relatively strong predictive performance on unseen data (R^2 = 0.968), its error increased more markedly from training to testing, with RMSE rising from 0.369 to 2.755 MPa. This wider training–test gap suggests that XGBoost was more sensitive to the training set and showed weaker generalization stability than TabDPT.

Table 4: Performance comparison of tuned models.

	Train Data				Test Data			
	R^2	**RMSE**	**MSE**	**MAE**	**R^2**	**RMSE**	**MSE**	**MAE**
TabDPT	0.997	0.982	0.964	0.727	0.978	2.214	**4.903**	1.806
TabM	0.984	2.136	4.564	1.646	0.976	2.376	5.646	1.923
XGBoost	0.999	0.369	0.136	0.264	0.968	2.755	7.581	2.139

These trends are further supported by the Taylor diagrams in Fig. 8. On the training set (Fig. 8a), XGBoost is located closest to the reference point, reflecting its extremely high correlation, closely matched standard deviation, and minimal centered error. However, on the independent test set (Fig. 8b), TabDPT appears closer to the reference point than the other models, indicating a more favorable balance among correlation, dispersion, and prediction error. TabM also shows competitive test behavior but remains slightly farther from the reference than TabDPT. In contrast, XGBoost shifts farther away on the test diagram, consistent with the larger increase in prediction error observed in Table 4. A plausible explanation for the better generalization of TabDPT lies in the different ways in which the two models represent tabular structure. XGBoost can fit nonlinear patterns very effectively and therefore achieved an almost perfect fit

on the training data; however, on a medium-sized experimental dataset it may also become more sensitive to partition-specific patterns. By contrast, TabDPT performs prediction through contextual conditioning, allowing each query sample to be interpreted relative to informative training examples, while its polynomial interaction component helps represent coupled effects among binder composition, activator chemistry, and aggregate variables. Within the investigated dataset, this combination appears to yield a more favorable balance between expressiveness and robustness. This explication is made conservatively and is limited to the present dataset rather than being claimed as a universal advantage over tree-based models.

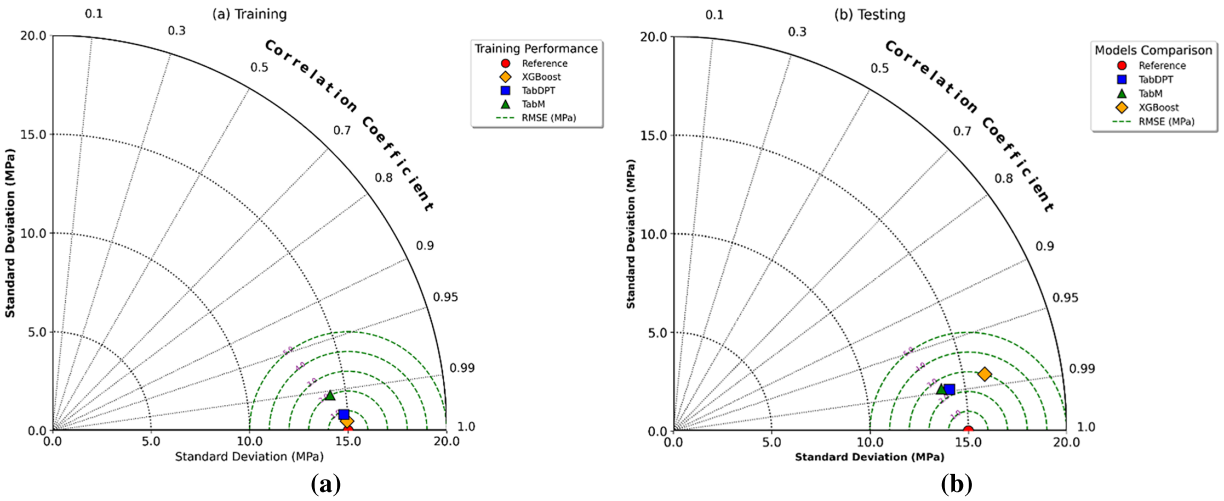

Figure 8: Taylor diagrams comparing the predictive performance of TabDPT, TabM, and XGBoost on the (**a**) training set and (**b**) test set.

Taken together, Table 4 and Fig. 8 indicate that TabDPT achieved the most balanced overall performance, combining high predictive accuracy with more stable generalization to unseen mixtures.

Fig. 9 compares the predictive performance of TabM, TabDPT, and XGBoost for estimating the 28-day compressive strength (CS28) using 5-fold cross-validation and the independent test set. All three models were developed using the same input variables and target output, thereby enabling a consistent comparison of their predictive behavior. Across the four-evaluation metrics, TabDPT shows the most favorable overall performance. It achieves the lowest RMSE, MAE, and MSE values, together with the highest R^2, on both the cross-validation results and the independent test set. These results indicate that TabDPT provides the best balance between predictive accuracy and generalization within the investigated dataset. By contrast, XGBoost exhibits the largest discrepancy between cross-validation and test performance, particularly in terms of RMSE and MSE. This pattern suggests that its predictive behavior is more sensitive to data partitioning and that its generalization to unseen mixtures is less stable than that of TabDPT. TabM shows intermediate performance, with better stability than XGBoost but slightly lower accuracy than TabDPT. The test-set performance is slightly better than the repeated cross-validation average for some metrics. This difference may reflect sampling variability associated with the limited dataset size and the structured nature of the mixture design. For this reason, repeated cross-validation results are considered the more informative indicator of model robustness, whereas the independent test set is used for final external verification.

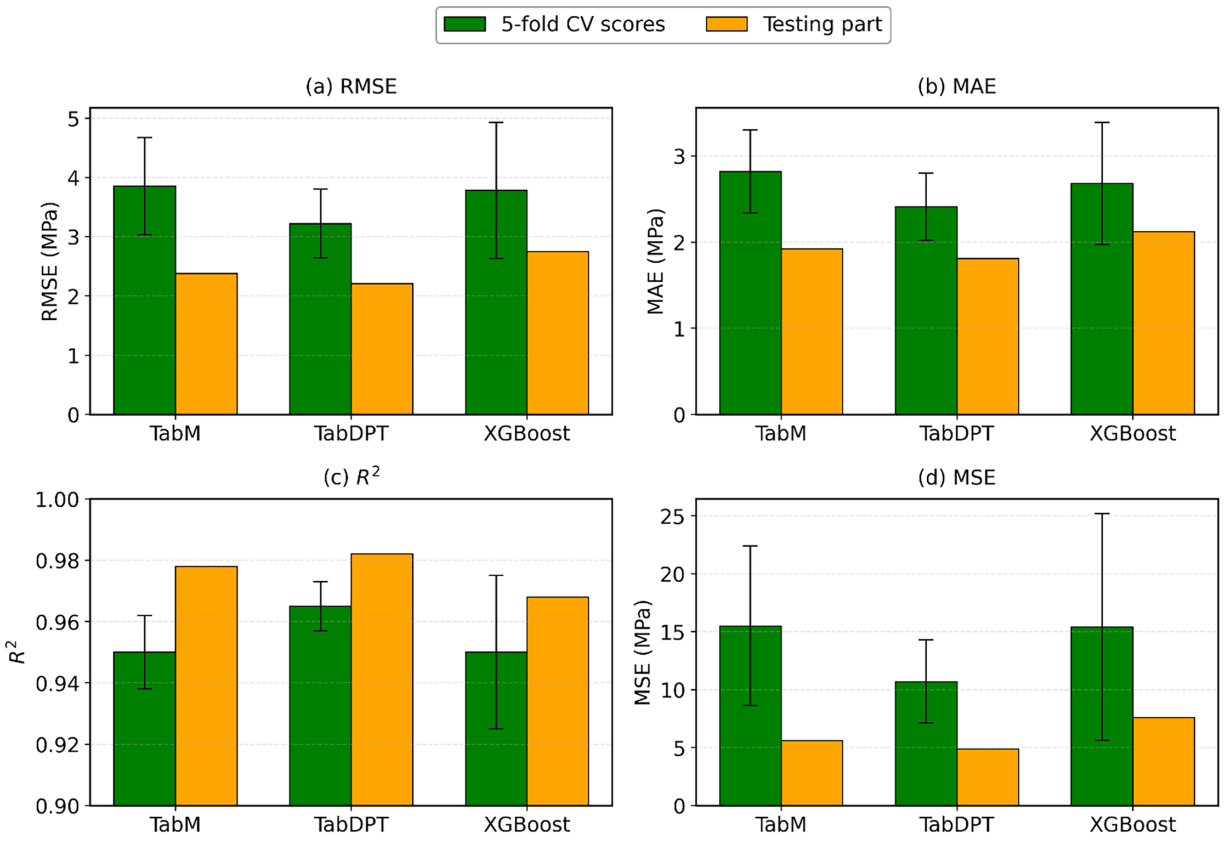

Figure 9: Performance comparison of TabM, TabDPT, and XGBoost on repeated 5-fold cross-validation and the independent test set: (**a**) RMSE, (**b**) MAE, (**c**) R^2, and (**d**) MSE.

Fig. 10 shows the distributions of RMSE values obtained from repeated 5-fold cross-validation for the three models. The box–whisker plots provide additional insight into both average prediction error and variability across different data splits. Among the three models, TabDPT yields the lowest mean RMSE (3.22 MPa) and the smallest standard deviation (0.60 MPa), indicating that its errors are more tightly clustered across repeated runs. This relatively compact distribution suggests a more stable performance under different fold assignments. XGBoost displays the highest mean RMSE (3.79 MPa) and the largest dispersion (standard deviation = 1.17 MPa), with a wider interquartile range and longer whiskers. These features indicate greater sensitivity to the specific partitioning of the data and a higher risk of occasional large prediction errors. TabM again shows intermediate behavior. Its mean RMSE (3.88 MPa) is comparable to that of XGBoost, but its distribution is somewhat more compact, although a noticeable outlier is still observed. Taken together, Fig. 10 confirms that TabDPT is not only the most accurate model on average, but also the most stable across repeated cross-validation runs.

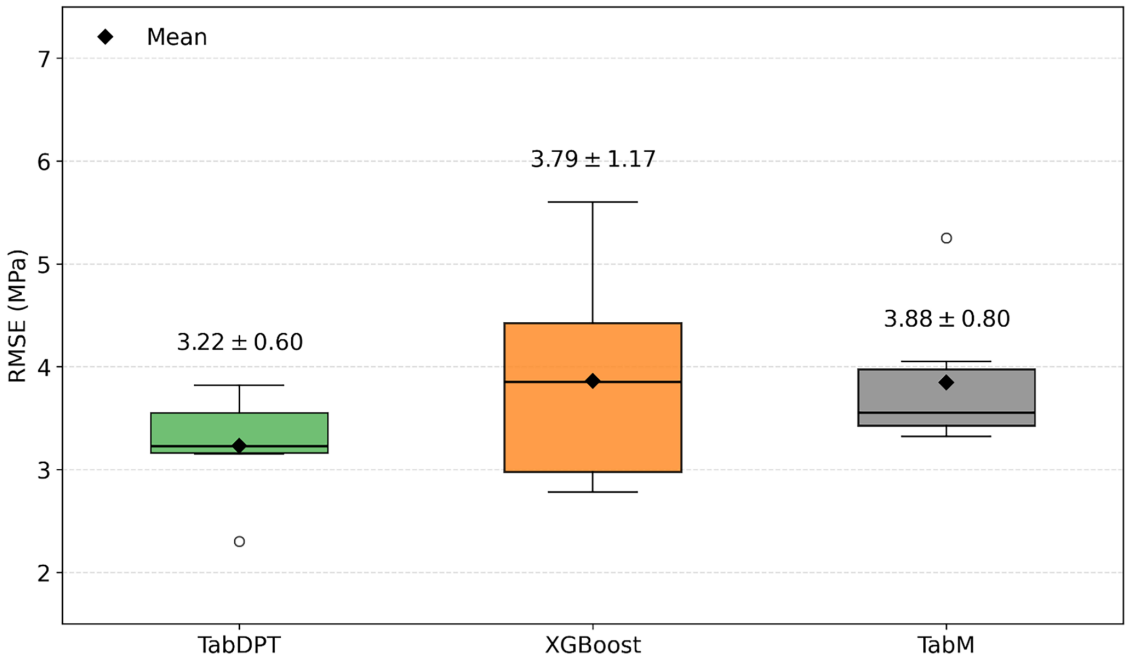

Figure 10: RMSE distributions of TabDPT, XGBoost, and TabM under repeated 5-fold CV.

4.2 Performance of the Best Model

Fig. 11 presents parity plots of the measured vs. predicted 28-day compressive strength (CS28) obtained by the best-performing model TabDPT for both the training and independent test sets. The scatter distributions relative to the 1:1 reference line (y = x) and the fitted regression line provide a direct visual assessment of prediction accuracy, potential bias, and consistency over the full-strength range. For the training set, predictions closely follow the 1:1 line, with data points tightly clustered along the diagonal. The fitted linear trend nearly overlaps the ideal line, indicating negligible systematic bias and confirming that TabDPT captures the dominant relationships in the experimental data with high fidelity. For the test set, the parity plot shows a similarly strong correspondence between measured and predicted values, with a modest increase in dispersion relative to the training set, as expected for unseen samples. Nevertheless, most points remain concentrated near the 1:1 line, and the fitted trend remains close to the ideal relationship, suggesting that the model retains its predictive capability without pronounced overestimation or underestimation when generalizing to new mixture compositions. To sum up, Fig. 11 visually corroborates the quantitative performance metrics reported earlier, supporting the conclusion that TabDPT achieves a favorable balance between accuracy and generalization for CS28 prediction within the investigated experimental domain.

Fig. 12 shows the distribution of relative prediction errors, defined as $|\text{Actual} - \text{Predicted}|/\text{Actual}$, for the optimal TabDPT model on the training and testing datasets. This analysis complements the absolute error metrics in Table 4 by providing a normalized view of prediction accuracy across samples with different compressive-strength levels. For the training set (Fig. 12a), the relative errors remain low for most samples. The majority of values are below approximately 0.02, with only a few isolated peaks approaching about 0.04–0.05. This pattern is consistent with the strong training performance reported in Table 4 and indicates that the model reproduces the training data with only limited normalized deviations. For the testing set (Fig. 12b), the relative errors are more dispersed, as expected for unseen data. Most samples still remain below approximately 0.05, although a small number of cases show larger deviations, with peak values reaching about 0.15–0.18. These higher-error cases appear to be isolated rather than dominant, which is consistent with

the relatively low test RMSE and MAE reported in Table 4. In addition, no obvious trend in relative error is observed with sample index in either subset, suggesting that the prediction deviations are not concentrated in a specific portion of the ordered samples. Therefore, the relative-error analysis indicates that TabDPT provides high predictive precision for most mixtures, while the larger deviations on the test set remain limited to a small number of individual cases.

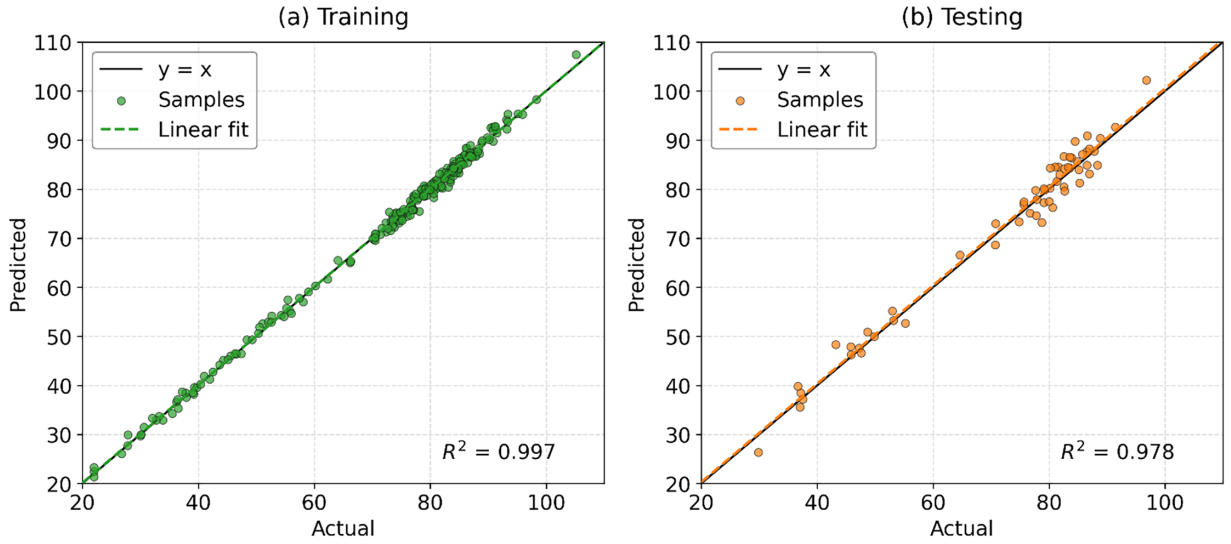

Figure 11: Relationship between measured and predicted values of the optimal TabDPT model: (**a**) training set; (**b**) testing set.

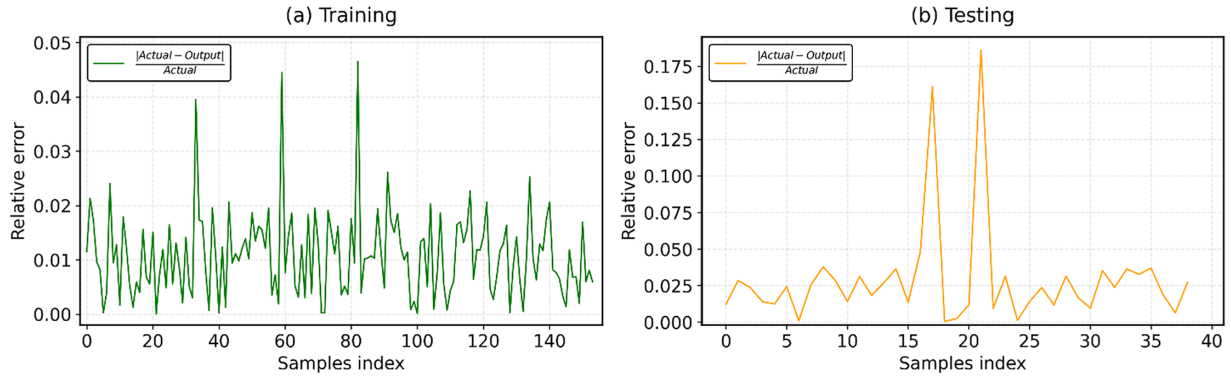

Figure 12: Relative error vs. sample index for the selected model (TabDPT): (**a**) training; (**b**) test.

Fig. 13 presents the residual–density distributions vs. predicted compressive strength for the TabDPT model on the training and test sets. This visualization was used to examine residual concentration, possible bias relative to the zero-residual line, and whether the spread of errors changes systematically over the prediction range. The dashed red line denotes zero residual, while the color intensity represents regions with higher sample density. For the training set (Fig. 13a), the residuals are densely concentrated around the zero-residual line, particularly in the predicted-strength range of approximately 70–85 MPa. Most points fall within a relatively narrow band, and no clear asymmetric clustering or systematic drift is observed. This pattern is consistent with the low training error reported in Table 4 and suggests that the model fits the training data with minimal bias. For the test set (Fig. 13b), the residual distribution becomes broader, which

is expected for unseen data. The highest residual density is observed mainly in the predicted-strength range of about 70–90 MPa, but the dense region is shifted slightly below the zero-residual line. This indicates a mild tendency toward negative residuals in that range, suggesting slight overprediction for part of the test data. A mild negative bias is therefore visible in a subset of the test samples, mainly within this predicted-strength range. Because the independent test subset is relatively small, the present results do not support a strong categorical attribution of this bias to a single mixture family such as high-slag or high-fiber mixtures. More conservatively, the bias appears to arise within a localized response region where several interacting variables overlap. Nevertheless, most test residuals remain concentrated within a relatively limited band, while only a few isolated points reach larger negative deviations. No strong funnel-shaped pattern is evident, so there is no clear visual indication of pronounced heteroscedasticity within the investigated prediction range.

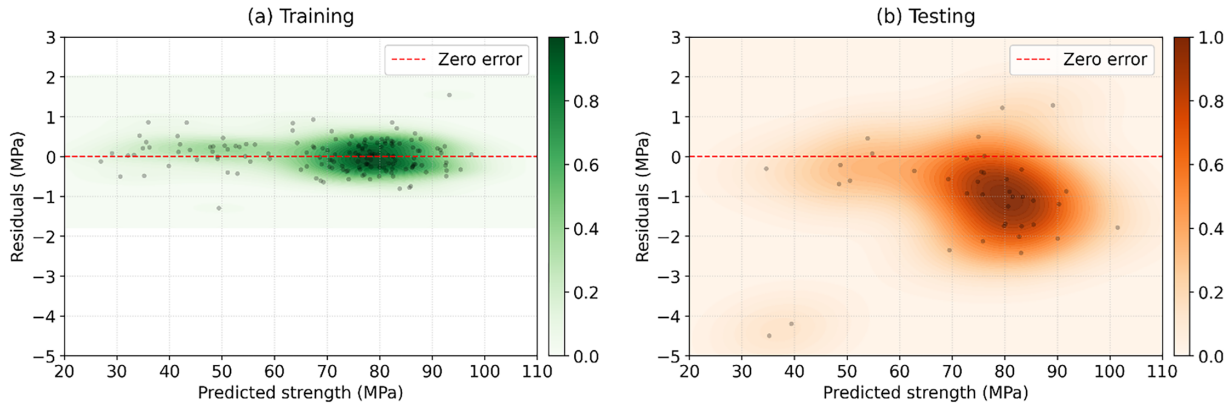

Figure 13: Residual-density distribution vs. predicted values for the TabDPT model: (**a**) training; (**b**) test.

Taken together, the residual–density patterns support the findings from Table 4 and Figs. 12 and 13. TabDPT provides stable predictions with low training bias and acceptable test-set dispersion, although a mild negative bias is visible for a subset of the test samples.

Fig. 14 illustrates the learning behavior of the TabDPT model as a function of training-set size using two complementary metrics, R^2 and RMSE, for both the training data and 5-fold cross-validation. At small training sizes (approximately 15–25 samples), the cross-validated R^2 is relatively low and accompanied by wide uncertainty bands, while the cross-validated RMSE is high, indicating that the model has not yet captured the mixture–strength relationship in a stable manner. As the number of training samples increases to about 40–60, the cross-validation performance improves markedly, with R^2 rising to around 0.90 and RMSE decreasing substantially. This trend suggests that additional data in this range contribute strongly to improved generalization.

Beyond approximately 80 training samples, the learning curves begin to flatten. The cross-validated R^2 continues to increase only gradually, reaching about 0.95–0.97 at the largest training sizes, while the corresponding RMSE decreases more slowly to roughly 2.8–3.8 MPa. At the same time, the uncertainty bands become narrower, indicating more stable performance across folds. Throughout the full range, the training R^2 remains high and the training RMSE remains low, reflecting strong fitting capacity of the model.

Importantly, the gap between training and cross-validation performance becomes smaller as the dataset grows, especially beyond about 80 samples. This pattern suggests that the model benefits substantially from larger training sets and that its generalization behavior becomes more stable as more mixtures are included. Within the investigated experimental domain, the results suggest that a training size on the order

of 100–150 samples may already provide a reasonably stable basis for prediction, although some incremental improvement could still be expected from additional data.

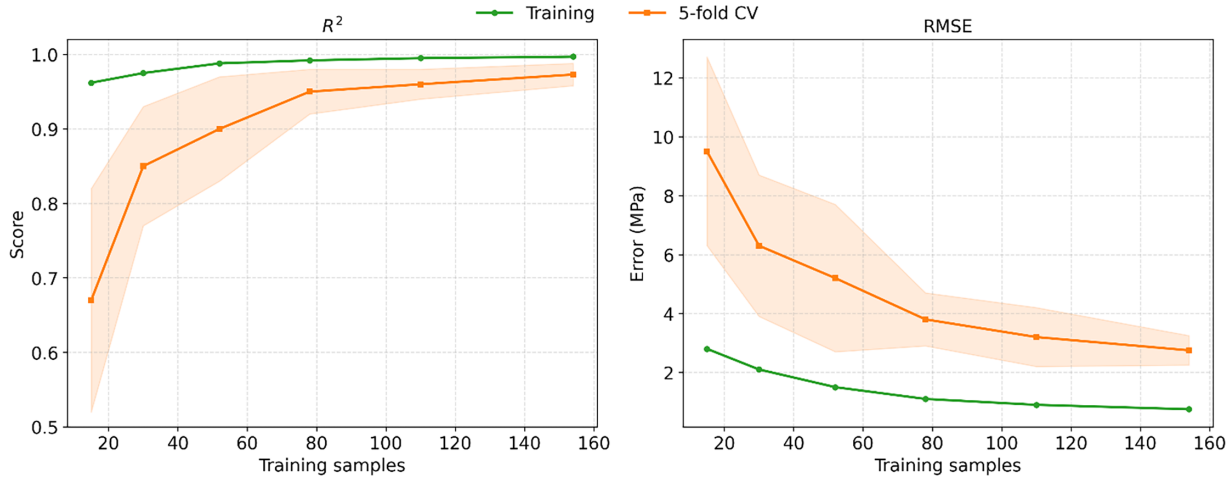

Figure 14: Learning curves of the selected model (TabDPT): R^2 and RMSE vs. training-set size.

To further assess robustness, the stability of the selected TabDPT model was evaluated using repeated 5-fold cross-validation across 20 different random seeds, with performance summarized as the mean ± standard deviation across seeds. Table 5 and Fig. 15 describe the sensitivity of the model to repeated fold partitioning, that is, resampling stability, whereas the fixed hold-out test results reported earlier reflect generalization performance under a single data split.

Table 5: Stability of the TabDPT model under repeated 5-fold CV across 20 random seeds.

Seed	R^2 CV	RMSE CV (MPa)	MAE CV (MPa)	R^2_std_fold	RMSE_std_fold (MPa)	MAE_std_fold (MPa)
0	0.9668	3.0819	2.3548	0.0073	0.4375	0.3124
1	0.9593	3.2563	2.4039	0.0099	0.5696	0.3418
2	0.9680	2.9900	2.2836	0.0047	0.5557	0.2986
3	0.9680	3.0265	2.2912	0.0049	0.3848	0.2647
4	0.9582	3.3645	2.3835	0.0136	0.5086	0.3369
5	0.9582	3.3421	2.4041	0.0146	0.5274	0.3482
6	0.9680	2.9572	2.2566	0.0068	0.4883	0.2815
7	0.9680	2.9500	2.3132	0.0093	0.4300	0.2954
8	0.9554	3.4789	2.5019	0.0138	0.5859	0.3726
9	0.9673	3.0502	2.2435	0.0065	0.4921	0.2768
10	0.9555	3.4366	2.4839	0.0133	0.6258	0.3612
11	0.9555	3.4642	2.4673	0.0165	0.7198	0.4025
12	0.9638	3.2372	2.3628	0.0101	0.4617	0.3019
13	0.9385	4.0705	2.7549	0.0230	0.8600	0.4683
14	0.9407	3.9391	2.7175	0.0119	0.8044	0.4516
15	0.9544	3.5081	2.5370	0.0152	0.5113	0.3574
16	0.9491	3.5689	2.4923	0.0142	0.7035	0.3897
17	0.9646	3.2368	2.3771	0.0095	0.5519	0.3185
18	0.9503	3.5122	2.5151	0.0135	0.6509	0.3748
19	0.9444	3.7207	2.6258	0.0109	0.7216	0.4172

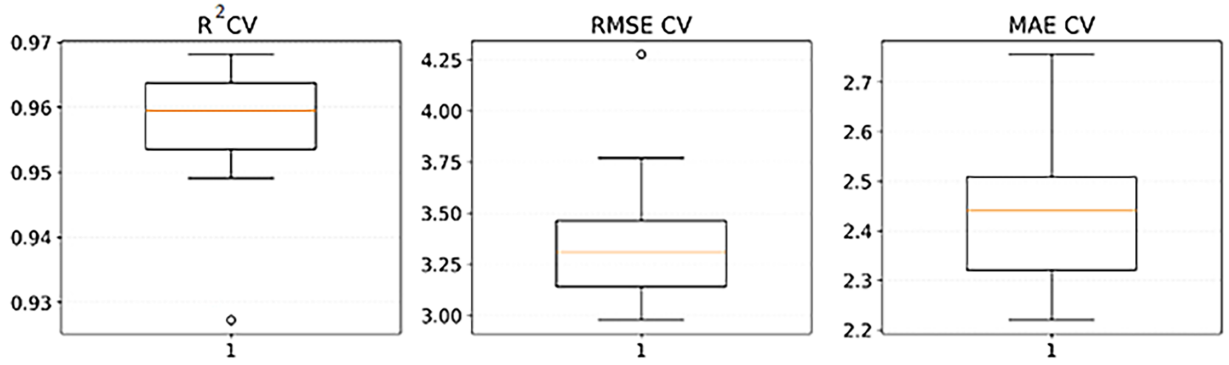

Figure 15: Distribution of repeated 5-fold CV metrics of the TabDPT model across 20 random seeds.

The repeated cross-validation results indicate that TabDPT maintains strong and relatively stable predictive performance across different random partitions, with seed-averaged values of R^2 CV = 0.9577 ± 0.0094, RMSE CV = 3.3596 ± 0.3148 MPa, and MAE CV = 2.4385 ± 0.1438 MPa (Table 5). The compact box–whisker distributions in Fig. 15 further suggest that the model is not overly sensitive to random data partitioning. Although a small degree of variability remains across seeds, the spread is limited, indicating that the observed predictive performance is not driven by a single favorable split but remains reasonably consistent under repeated resampling.

4.3 SHAP-Based Interpretability

4.3.1 Global Influence and Variable Ranking Based on SHAP

Fig. 16 presents the SHAP summary plot of the selected TabDPT model, which provides a global view of feature importance together with the direction and spread of each variable's contribution to the model output. In this plot, the variables are ordered according to their mean absolute SHAP values, while the horizontal dispersion of the points reflects the range of their contributions across the dataset. SHAP values describe model-based associations within the trained predictive framework rather than direct causal effects of the material variables.

Among the investigated inputs, GGBS shows the broadest SHAP spread and the highest impact on model predictions. The distribution of points indicates that variations in GGBS content are associated with substantial changes in predicted compressive strength across the investigated mixtures. Na_2O ranks second and also exhibits a wide SHAP range, suggesting that alkalinity is another major source of predictive variation in the trained model. Na_2SiO_3 and NaOH form a second tier of influential variables, indicating that activator chemistry remains important, although their average contributions are clearly smaller than those of GGBS and Na_2O. By contrast, FA, FiAg, SF, Fiber, and CoAg exhibit narrower SHAP spreads and lower importance, suggesting that their contributions to predicted CS28 are more limited within the present dataset.

The quantitative ranking is summarized in Table 6 using mean absolute SHAP values. GGBS is ranked first with a mean |SHAP| value of 9.382 MPa, followed by Na_2O with 7.778 MPa, Na_2SiO_3 with 3.855 MPa, and NaOH with 3.136 MPa. FA occupies an intermediate position with a mean |SHAP| of 1.980 MPa, whereas FiAg, SF, Fiber, and CoAg contribute relatively small values. This ranking indicates that the predictive behavior of the model is influenced primarily by binder composition and activator-related variables, while aggregate- and fiber-related variables play a comparatively secondary role in CS28 prediction.

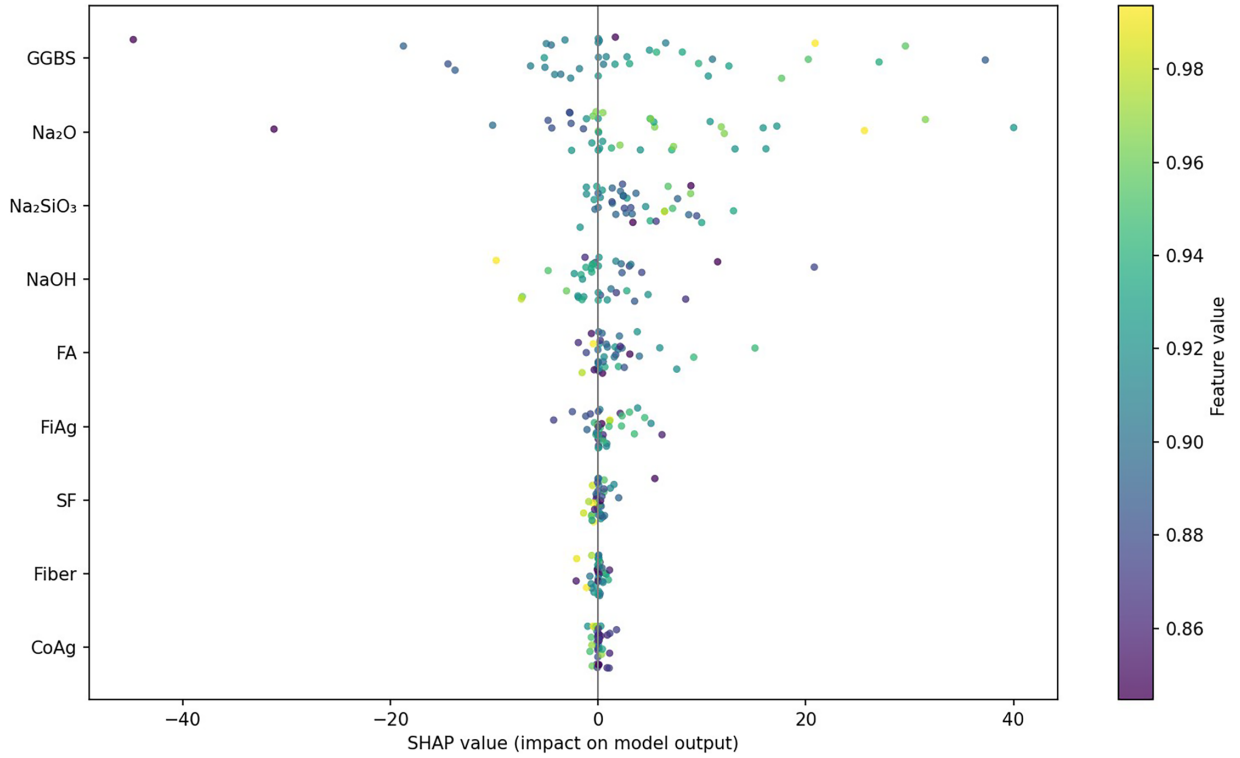

Figure 16: SHAP summary plot showing global feature importance.

Table 6: Average SHAP values and corresponding variable ranks.

Rank	Mean	SHAP
1	GGBS	9.382
2	Na_2O	7.778
3	Na_2SiO_3	3.855
4	NaOH	3.136
5	FA	1.980
6	FiAg	1.296
7	SF	0.558
8	Fiber	0.392

From a materials perspective, these trends are broadly consistent with previous studies on geopolymer systems. Higher GGBS contents are often associated with faster reaction kinetics and denser binding phases, which may contribute to higher compressive strength under suitable curing conditions [5]. Likewise, Na_2O and other activator-related variables are known to influence dissolution and gel formation in alkali-activated binders [12]. Nevertheless, such interpretations should be treated cautiously: the SHAP results indicate how the trained model responds to the available data, rather than establishing universal mechanistic laws. Within this scope, the combined evidence from Fig. 16 and Table 6 suggests that the developed model relies most strongly on chemical composition and activator chemistry when estimating the 28-day compressive strength of the investigated SFGPC mixtures.

4.3.2 Nonlinear Influence Patterns and Design-Relevant Ranges Based on SHAP Dependence Plots

Fig. 17 presents SHAP dependence plots for six selected influential variables, chosen based on the global SHAP ranking and engineering relevance including the main binder, activator, and fiber-related inputs. The order of presentation follows the discussion sequence rather than the ranking order in Fig. 16. It shows how the SHAP value of each variable changes across its investigated range. These plots provide a more detailed view of the nonlinear response patterns learned by the TabDPT model.

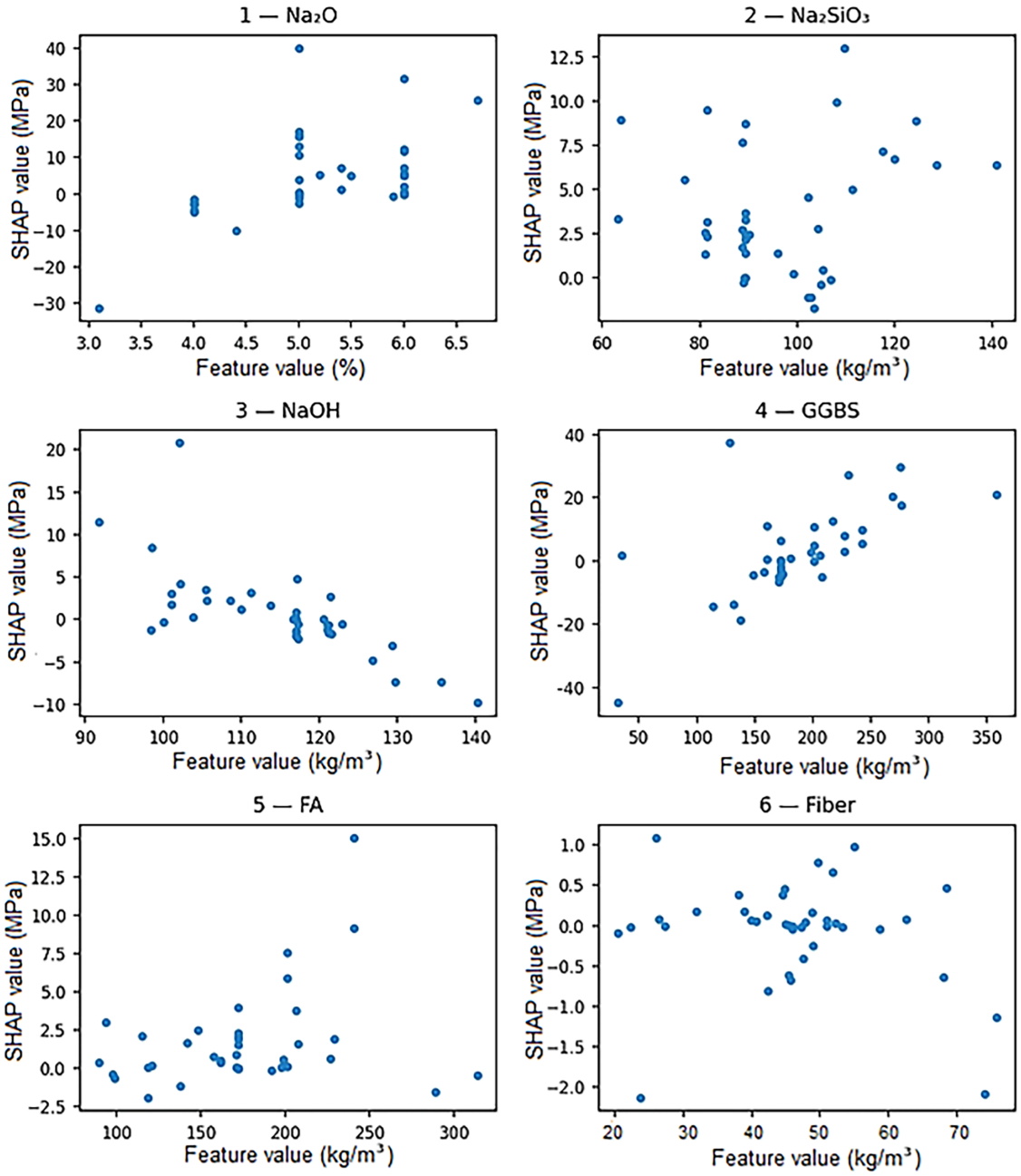

Figure 17: SHAP dependence plots for six selected variables.

Among the activator-related variables, Na_2O exhibits the clearest transition-like behavior. At low alkalinity levels, particularly around 3.0%–4.5%, the SHAP values are mostly negative and in some cases strongly so, indicating that low Na_2O is associated with reduced predicted compressive strength. As Na_2O increases toward approximately 5.0%–6.0%, the SHAP values shift upward and become predominantly positive. This pattern suggests an apparent threshold-like region near 5.0%, above which the model associate's alkalinity with more favorable strength development. However, this should be implied as an indicative range within the investigated mixtures rather than a universal design limit. Na_2SiO_3 shows a generally positive but more scattered pattern. Across much of the investigated range, its SHAP values remain near zero to moderately positive, with some larger positive contributions at higher contents. At the same time, the spread of points indicates that the effect of Na_2SiO_3 is not strictly monotonic and likely depends on interactions with other mixture variables, especially Na_2O, NaOH, and binder composition. In contrast, NaOH displays a different trend. At lower-to-moderate values, the SHAP contributions are near zero or slightly positive, whereas at higher dosages, especially beyond roughly 120 kg/m^3, the SHAP values tend to become negative. Within the trained model, this may indicate that excessive hydroxide dosage is associated with less favorable strength predictions in part of the investigated domain. This pattern is broadly consistent with the possibility that overly high alkalinity does not always translate into improved matrix development, although the present SHAP results alone do not establish the underlying mechanism [5,34]. GGBS shows the strongest and most consistent positive association among all variables. At low slag contents, the SHAP values are mostly negative, while increasing GGBS content is associated with a marked upward shift in SHAP values, especially over the approximate range of 150–280 kg/m^3. This trend indicates that the model relies strongly on slag-related information when predicting compressive strength. From a materials perspective, this is broadly consistent with the known contribution of slag to accelerated reaction kinetics and denser binding phases in alkali-activated systems under suitable curing conditions [13,42,43]. FA presents a more dispersed and context-dependent pattern. Over much of the intermediate range, the SHAP values remain close to zero or mildly positive, whereas higher FA contents are associated with both positive and negative contributions. This suggests that the influence of FA is not determined by its absolute dosage alone, but may depend on its interaction with GGBS and activator chemistry. Finally, steel fiber content exhibits relatively small SHAP magnitudes compared with the chemical and binder-related variables. Most SHAP values remain within a narrow range, indicating a limited contribution of fiber dosage to predicted compressive strength in the present model. This result is compatible with the general understanding that steel fibers contribute more directly to crack control and post-cracking behavior than to compressive strength itself [14]. The relatively low SHAP importance of fiber in the present study should therefore be interpreted cautiously and only within the scope of the current dataset and target variable. For 28-day compressive strength, the model indicates that binder composition and activator chemistry contribute more strongly than fiber dosage to prediction variability within the investigated mixtures. This does not imply that steel fibers are unimportant in general; rather, their contribution may be more pronounced for other responses such as flexural behavior, toughness, crack resistance, or post-cracking performance. In addition, process-related variables such as mixing regime or water-related parameters were not part of the present dataset and may influence the apparent relative importance of the available features.

Generally, the SHAP dependence plots suggest that the most design-relevant nonlinear patterns in the model are associated with GGBS and alkali activator variables, particularly Na_2O and NaOH. Within the investigated experimental domain, the plots indicate that mixtures with Na_2O around 5.0%–6.0% and moderate-to-high GGBS contents tend to be associated with more favorable strength predictions, whereas very low Na_2O and excessively high NaOH appear less favorable. These ranges should be interpreted

as model-informed guidance within the current dataset rather than fixed mechanistic thresholds for all SFGPC systems.

4.3.3 Multivariable Interaction and Design-Relevant Region

Fig. 18 visualizes the combined effect of the two highly influential variables identified by the SHAP analysis, namely alkalinity (Na_2O) and GGBS content, on the predicted 28-day compressive strength. The contour map highlights a favorable prediction region in which CS28 exceeds 60 MPa, as indicated by the hatched area and the corresponding 60 MPa boundary line. The plot suggests a compensatory relationship between these two variables: when Na_2O is relatively low, a higher GGBS content is required to maintain the predicted strength above 60 MPa, whereas at higher Na_2O levels the minimum GGBS requirement becomes lower [5,13].

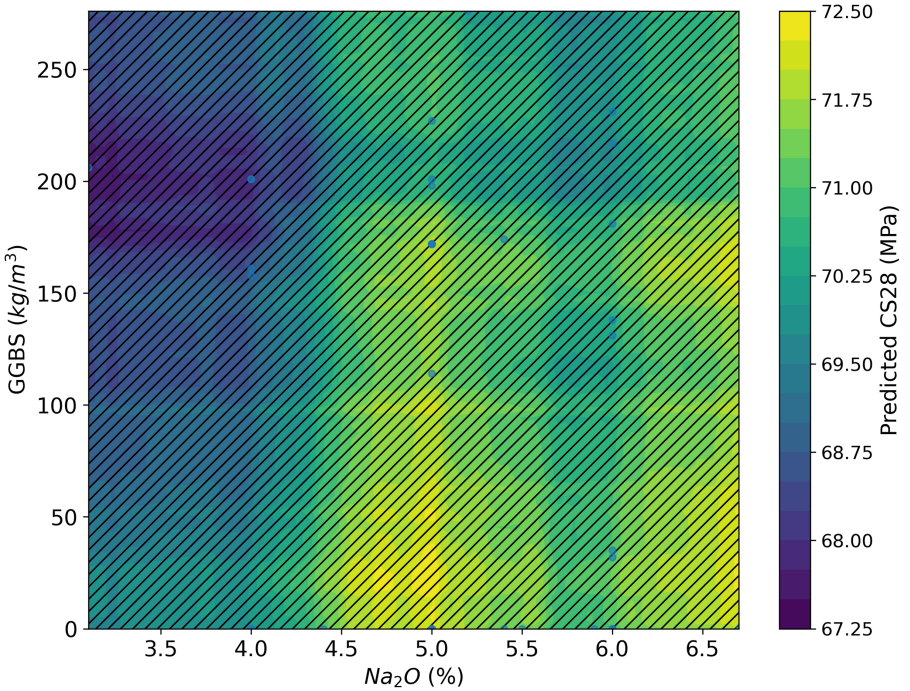

Figure 18: Bivariate interaction plot illustrating the combined effect of alkalinity (Na_2O) and slag content (GGBS) on the predicted CS28 above 60 MPa.

More specifically, within the investigated design domain, when Na_2O is around 4.0%, the model predicts that CS28 values above 60 MPa can be achieved with GGBS contents of approximately 75 kg/m^3 or higher. By contrast, when Na_2O decreases to about 3.6%–3.8%, the model indicates that GGBS may need to increase to above roughly 150 kg/m^3 to reach the same strength threshold. Conversely, the lowest predicted strength values, below approximately 40 MPa, are concentrated in the lower-left portion of the map, where both Na_2O and GGBS remain low. This pattern suggests that simultaneously low alkalinity and low slag content are unfavorable for strength development within the present dataset.

The contour map also shows that predicted strengths above about 80 MPa are mainly located in the region where Na_2O exceeds approximately 5.5% and GGBS is relatively high, typically above about 300 kg/m^3. In addition, the contour geometry suggests a transition-like region around $Na_2O \approx 4.0\%$, beyond which the positive contribution of increasing GGBS becomes more evident in the model predictions. This

trend is broadly consistent with the general understanding that slag content and activator dosage jointly influence strength development in alkali-activated systems [11,12]. However, the numerical boundaries inferred from Fig. 18 should be interpreted as model-informed guidance within the investigated dataset, rather than as universal design thresholds for all SFGPC mixtures.

4.3.4 Uncertainty Map

Based on the TabDPT predictions, Fig. 19 provides an uncertainty map that helps delineate regions of relatively high and low predictive stability within the investigated design space. The dashed boundary indicates the convex hull of the available data. Here, the prediction uncertainty is expressed in terms of the standard deviation (STD) of the model output, thereby offering a practical indication of where the model appears relatively high and low predictive stability.

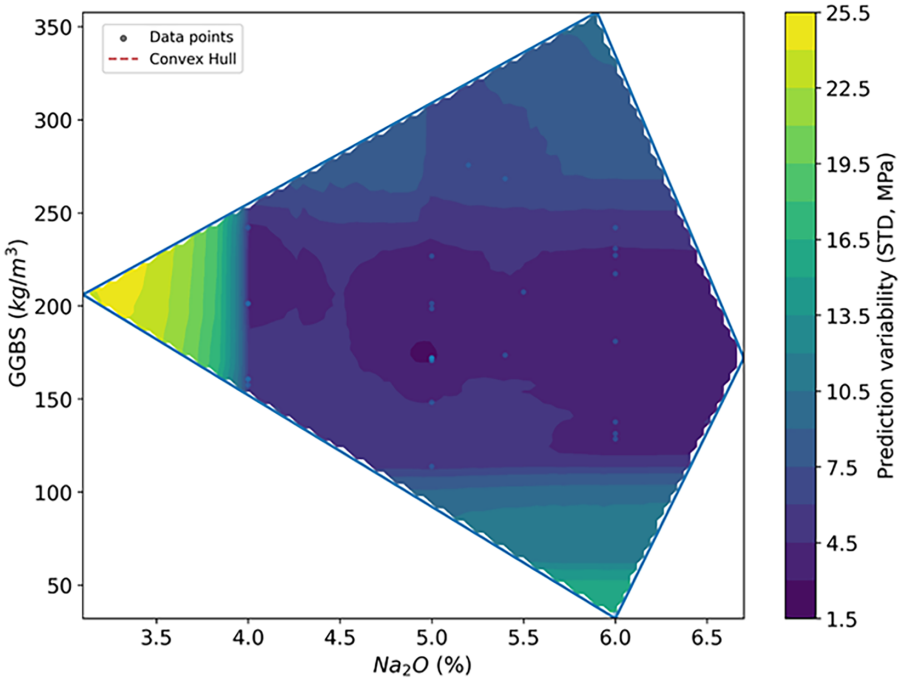

Figure 19: Uncertainty map of the TabDPT model in the Na_2O–GGBS design space, expressed as the standard deviation (STD) of predicted CS28.

The dark purple regions correspond to the lowest prediction variability, with STD values approximately in the range of 1.5–4.5 MPa. These low-uncertainty zones are concentrated mainly in the central part of the design space, particularly around Na_2O values near 5.0% and GGBS contents of roughly 150–200 kg/m^3. This region also appears to coincide with a relatively denser distribution of experimental data, which may explain the improved stability of the model predictions there. This finding aligns with advanced modeling trends where local data density is recognized as a primary driver of epistemic uncertainty reduction in material property prediction [44]. By contrast, the transition toward green and yellow colors indicate substantially higher predictive uncertainty, with STD values rising above about 20–25.5 MPa in some areas. This increase is especially evident along the left boundary of the map, where Na_2O is low, generally below about 4.0%. Such behavior suggests that the model becomes less stable in sparsely sampled or more weakly supported regions of the design domain, and that predictions in these areas should be interpreted with greater caution.

This observation is consistent with the well-known risk of performance degradation when machine learning models are forced to extrapolate beyond the convex hull of their training domain [45].

The convex hull outlines the region spanned by the available data and therefore provides a useful visual boundary for model-supported interpolation. Mixture designs located inside the lower-uncertainty zones are more likely to yield stable predictions, whereas predictions near the outer edges, particularly in high-uncertainty regions, may carry a greater risk of extrapolation or reduced reliability. Accordingly, extreme mixture combinations, such as very low alkali content or very low slag content, should preferably be verified through additional experiments rather than relying solely on model output. This rigorous approach to uncertainty-aware design is crucial for ensuring the structural safety of sustainable concrete applications [24].

Despite the encouraging results, several limitations of the present study should be acknowledged. First, although the dataset was generated under a relatively consistent laboratory protocol, the sample size remains moderate, and broader generalization would still benefit from additional experimental data. Second, the dataset reflects the specific material domain investigated in this study, including the particular fly ash, GGBS, NaOH–Na_2SiO_3 activator system, aggregate sources, silica fume content, and straight steel fibers used in the laboratory program. Therefore, the trained model should not be directly extrapolated to substantially different precursor chemistries, fly ash characteristics, fiber geometries such as hooked-end or hybrid fibers, or markedly different curing and mixing conditions without further validation. Third, some potentially relevant process-related variables, such as water-related effects and mixing parameters, were not available in the current dataset and thus were not included in the modeling framework. Finally, the SHAP results should be interpreted as model-based associations within the investigated experimental domain, rather than universal causal mechanisms. Within these boundaries, the proposed framework is best understood as a reliable tool for prediction and preliminary mixture screening under controlled conditions, while further dataset expansion remains necessary to improve transferability.

5 Conclusion

This study investigated the prediction of 28-day compressive strength (CS28) of SFGPC using an original experimental dataset of 189 mixtures produced under a consistent laboratory protocol in Vietnam. Three models, namely TabDPT, TabM, and XGBoost, were developed and compared. Among them, TabDPT achieved the most balanced performance, combining high predictive accuracy with better generalization on unseen data. On the independent test set, TabDPT yielded the best results, with R^2 = 0.978, RMSE = 2.214 MPa, MSE = 4.903 MPa and MAE = 1.806 MPa, whereas XGBoost, despite its extremely strong fit on the training set, showed a larger degradation in performance from training to testing. Repeated 5-fold cross-validation across 20 random seeds further indicated that TabDPT maintained relatively stable performance under different data partitions, suggesting that its predictive behavior was not driven by a single favorable split.

The interpretability analysis provided additional insight into the response patterns learned by the selected model. Global SHAP ranking indicated that GGBS and Na_2O were the two most influential variables, followed by Na_2SiO_3 and NaOH, whereas aggregate-related variables and steel fiber content played comparatively smaller roles in CS28 prediction within the investigated dataset. SHAP dependence plots suggested nonlinear and interaction-dependent effects, particularly for activator chemistry and slag content. In addition, the bivariate interaction map of Na_2O and GGBS indicated that favorable strength predictions above 60 MPa were associated with moderate-to-high alkalinity and sufficient slag content, while the uncertainty map showed that prediction stability was highest in the more densely sampled central region of the design space and decreased near sparsely supported boundaries. These results suggest that the

proposed framework can provide useful model-informed guidance for mixture screening and preliminary design within the investigated experimental domain.

In conclusion, the present results show that recent tabular deep-learning models, particularly TabDPT, can provide reliable prediction of CS28 for SFGPC when trained on a relatively broad and internally consistent experimental dataset. However, the current dataset remains limited to specific material sources, mixture ranges, and curing conditions, and the SHAP-based results should be interpreted as model-based associations within the investigated domain rather than universal causal mechanisms. Future work should focus on expanding the dataset to additional material systems and exposure conditions, validating the framework on external datasets, and further developing uncertainty-aware prediction for broader and more practical geopolymer mixture design.

Acknowledgement: We are profoundly grateful for the generous support for our research provided by the I4T group, University of Transport Technology, Hanoi, Vietnam.

Funding Statement: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contributions: Quynh-Anh Thi Bui: conceptualization, methodology, investigation (experimental work), formal analysis, data curation, writing—original draft, writing—review & editing, visualization, supervision, project administration. Son Hoang Trinh: investigation (experimental work), formal analysis, data curation, writing—review & editing. Maryam Sayadi: software (model development), methodology (AI framework), validation, writing—review & editing. Reza Khanali: software (model development), methodology (AI framework), validation, writing—review & editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets generated and analyzed during the current study are not publicly available at this stage because they are part of an ongoing research project and are subject to confidentiality and intellectual property restrictions. However, the data can be made available from the corresponding author upon reasonable request, for academic and non-commercial purposes, under a data-sharing agreement.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Davidovits J. Geopolymer chemistry and applications. Saint-Quentin, France: Geopolymer Institute; 2008.
2. Hardjito D, Wallah SE, Sumajouw DJ, Rangan BV. Fly ash-based geopolymer concrete. Aust J Struct Eng. 2005;6(1):77–86. doi:10.1080/13287982.2005.11464946.
3. Provis JL, Van Deventer JS. Geopolymers: structures, processing, properties and industrial applications. Amsterdam, The Netherlands: Elsevier; 2009. doi:10.1533/9781845696382.
4. Lloyd NA, Rangan BV. Geopolymer concrete with fly ash. In: Proceedings of the Second International Conference on Sustainable Construction Materials and Technologies; 2010 Jun 28–30; Ancona, Italy. p. 1493–504.
5. Nath P, Sarker PK. Effect of GGBFS on setting, workability and early strength properties of fly ash geopolymer concrete cured in ambient condition. Constr Build Mater. 2014;66(2):163–71. doi:10.1016/j.conbuildmat.2014.05.080.
6. Provis JL. Alkali-activated materials. Cem Concr Res. 2018;114:40–8. doi:10.1016/j.cemconres.2017.02.009.
7. Singh B, Ishwarya G, Gupta M, Bhattacharyya SK. Geopolymer concrete: a review of some recent developments. Constr Build Mater. 2015;85:78–90. doi:10.1016/j.conbuildmat.2015.03.036.
8. Sathiparan N, Jeyananthan P, Subramaniam DN. A comparative study of machine learning techniques and data processing for predicting the compressive strength of pervious concrete with supplementary cementitious materials and chemical composition influence. Next Mater. 2025;9(6):100947. doi:10.1016/j.nxmate.2025.100947.

9. Zuhua Z, Xiao Y, Huajun Z, Yue C. Role of water in the synthesis of calcined Kaolin-based geopolymer. Appl Clay Sci. 2009;43(2):218–23. doi:10.1016/j.clay.2008.09.003.
10. Kumar S, Kumar R. Mechanical activation of fly ash: effect on reaction, structure and properties of resulting geopolymer. Ceram Int. 2011;37(2):533–41. doi:10.1016/j.ceramint.2010.09.038.
11. Hardjito D, Wallah SE, Sumajouw DMJ, Rangan BV. On the development of fly ash-based geopolymer concrete. Aci Mater J. 2004;101(6):467–72. doi:10.14359/13485.
12. Provis JL, Bernal SA. Geopolymers and related alkali-activated materials. Annu Rev Mater Res. 2014;44(1):299–327. doi:10.1146/annurev-matsci-070813-113515.
13. Deb PS, Nath P, Sarker PK. The effects of ground granulated blast-furnace slag blending with fly ash and activator content on the workability and strength properties of geopolymer concrete cured at ambient temperature. Mater Des. 2014;62(9):32–9. doi:10.1016/j.matdes.2014.05.001.
14. Shaikh FUA, Hosan A. Mechanical properties of steel fibre reinforced geopolymer concretes at elevated temperatures. Constr Build Mater. 2016;114(8):15–28. doi:10.1016/j.conbuildmat.2016.03.158.
15. Provis JL. Geopolymers and other alkali activated materials: why, how, and what? Mater Struct. 2014;47(1):11–25. doi:10.1617/s11527-013-0211-5.
16. Rangan BV. Engineering properties of geopolymer concrete. In: Geopolymers. Cambridge, UK: Woodhead Publishing; 2009. p. 211–26. doi:10.1533/9781845696382.2.211.
17. Yunsheng Z, Wei S, Zongjin L, Yantao J. Study of polycondensation process of metakaolin-based geopolymeric cement using semi-empirical AM1 calculations. Adv Cem Res. 2009;21(2):67–73. doi:10.1680/adcr.2008.00017.
18. Zhang Y, Sun W. Semi-empirical AM1 calculations on 6-memebered alumino-silicate rings model: implications for dissolution process of metakaoline in alkaline solutions. J Mater Sci. 2007;42(9):3015–23. doi:10.1007/s10853-006-0521-x.
19. Li Y, Zhang Q, Kamiński P, Deifalla AF, Sufian M, Dyczko A, et al. Compressive strength of steel fiber-reinforced concrete employing supervised machine learning techniques. Materials. 2022;15(12):4209. doi:10.3390/ma15124209.
20. Pakzad SS, Roshan N, Ghalehnovi M. Comparison of various machine learning algorithms used for compressive strength prediction of steel fiber-reinforced concrete. Sci Rep. 2023;13(1):3646. doi:10.1038/s41598-023-30606-y.
21. Abdellatief M, Hamla W, Hamouda H. AI driven prediction of early age compressive strength in ultra high performance fiber reinforced concrete. Sci Rep. 2025;15(1):20316. doi:10.1038/s41598-025-06725-z.
22. Philip S, Nidhi M. Performance comparison of artificial neural network and random forest models for predicting the compressive strength of fibre-reinforced GGBS-based geopolymer concrete composites. Mater Circ Econ. 2024;6(1):34. doi:10.1007/s42824-024-00128-7.
23. Hossain MA, Uddin MN, Hossain MM. Prediction of compressive strength fiber-reinforced geopolymer concrete (FRGC) using gene expression programming (GEP). Mater Today Proc. 2023;20(4):519. doi:10.1016/j.matpr.2023.02.458.
24. Khan AA, Chaudhari O, Chandra R. A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation. Expert Syst Appl. 2024;244(2):122778. doi:10.1016/j.eswa.2023.122778.
25. Borisov V, Leemann T, Seßler K, Haug J, Pawelczyk M, Kasneci G. Deep neural networks and tabular data: a survey. IEEE Trans Neural Netw Learn Syst. 2024;35(6):7499–519. doi:10.1109/tnnls.2022.3229161.
26. Gorishniy Y, Rubachev I, Khrulkov V, Babenko A. Revisiting deep learning models for tabular data. Adv Neural Inf Process Syst. 2021;34:18932–43.
27. Jiang JP, Liu SY, Cai HR, Zhou QL, Ye HJ. Representation learning for tabular data: a comprehensive survey. IEEE Trans Pattern Anal Mach Intell. 2026;2026(6):1–20. doi:10.1109/TPAMI.2026.3657217.
28. Shwartz-Ziv R, Armon A. Tabular data: deep learning is not all you need. Inf Fusion. 2022;81(1):84–90. doi:10.1016/j.inffus.2021.11.011.
29. Gorishniy Y, Kotelnikov A, Babenko A. TabM: advancing tabular deep learning with parameter-efficient ensembling. arXiv:2410.24210. 2024. doi:10.48550/arXiv.2410.24210.

30. Gorishniy Y, Rubachev I, Babenko A. On embeddings for numerical features in tabular deep learning. Adv Neural Inf Process Syst. 2022;35:24991–5004. doi:10.52202/068431-1812.
31. Ma J, Thomas V, Hosseinzadeh R, Labach A, Kamkari H, Cresswell JC, et al. TabDPT: scaling tabular foundation models on real data. arXiv:2410.18164. 2024. doi:10.48550/arXiv.2410.18164.
32. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189–232. doi:10.1214/aos/1013203451.
33. Alabdullah AA, Iqbal M, Zahid M, Khan K, Amin MN, Jalal FE. Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. Constr Build Mater. 2022;345(3):128296. doi:10.1016/j.conbuildmat.2022.128296.
34. Ahmad A, Ahmad W, Chaiyasarn K, Ostrowski KA, Aslam F, Zajdel P, et al. Prediction of geopolymer concrete compressive strength using novel machine learning algorithms. Polymers. 2021;13(19):3389. doi:10.3390/polym13193389.
35. Kim J. Iterated grid search algorithm on unimodal criteria. Blacksburg, VA, USA: Virginia Polytechnic Institute and State University; 1997.
36. Rathnayaka M, Karunasinghe D, Gunasekara C, Wijesundara K, Lokuge W, Law DW. Machine learning approaches to predict compressive strength of fly ash-based geopolymer concrete: a comprehensive review. Constr Build Mater. 2024;419(7):135519. doi:10.1016/j.conbuildmat.2024.135519.
37. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Statist Surv. 2010;4:40–79. doi:10.1214/09-ss054.
38. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:1–10.
39. Davawala M, Joshi T, Shah M. Compressive strength prediction of high-strength concrete using machine learning. Emergent Mater. 2023;6(1):321–35. doi:10.1007/s42247-022-00409-4.
40. Mohr F, van Rijn JN. Learning curves for decision making in supervised machine learning: a survey. Mach Learn. 2024;113(11):8371–425. doi:10.1007/s10994-024-06619-7.
41. Taylor KE. Summarizing multiple aspects of model performance in a single diagram. J Geophys Res. 2001;106(D7):7183–92. doi:10.1029/2000jd900719.
42. Ma Z, Dan H, Tan J, Li M, Li S. Optimization design of MK-GGBS based geopolymer repairing mortar based on response surface methodology. Materials. 2023;16(5):1889. doi:10.3390/ma16051889.
43. Alhamoud A, Tajmir Riahi H, Ataei A. A practical mix design method of ground granulated blast-furnace slag-based one-part geopolymer concrete. Arab J Sci Eng. 2024;49(4):5447–66. doi:10.1007/s13369-023-08419-y.
44. Al-Shamasneh AR, Mahmoodzadeh A, Karim FK, Saidani T, Alghamdi A, Alnahas J, et al. Application of machine learning techniques to predict the compressive strength of steel fiber reinforced concrete. Sci Rep. 2025;15(1):30674. doi:10.1038/s41598-025-16516-1.
45. Hastie T, Friedman J, Tibshirani R. The elements of statistical learning. New York, NY, USA: Springer; 2001. doi:10.1007/978-0-387-21606-5.