FEDERAL UNIVERSITY OF TECHNOLOGY, AKURE
JOURNAL OF ENGINEERING AND ENGINEERING TECHNOLOGY

Explainable Machine Learning Modelling of Metakaolin Concrete Compressive Strength Using TabPFN, CatBoost, XGBoost, and SVR

[1]Omotayo, O.O.*

[1]*Department of Civil and Environmental Engineering, Federal University of Technology, Akure, Nigeria*
*Corresponding author: ooomotayo@futa.edu.ng

Omotayo O. O. (2026). Explainable Machine Learning Modelling of Metakaolin Concrete Compressive Strength: Benchmarking TabPFN Against Boosted Ensemble and Kernel-Based Algorithms. FUTA Journal of Engineering and Engineering Technology 20 (2), 11-26

Received Date: 15.05.2026 **Accepted Date: 29.05.2026**

Abstract
Predicting concrete compressive strength, particularly in supplementary cementitious material (SCM) systems, remains a critical challenge despite the advances in modelling techniques in the past decade. This is due to concrete's highly nonlinear physicochemical interactions, caused by variability in mix composition, hydration processes, and pozzolanic reactions. This study, therefore, evaluates TabPFN's effectiveness compared to support vector regression (SVR), categorical boosting (CatBoost), and extreme gradient boosting (XGBoost), in predicting metakaolin (MK) concrete compressive strength. A dataset of 328 records was curated from 14 literature studies on MK concrete compressive strength, processed, analyzed, and used in training the selected models. Specific input features in the dataset include metakaolin oxide compositions and content; the water content and water-binder ratio; the cement and aggregate contents; and the curing age, while the target was the compressive strength. Bayesian hyperparameter optimization was conducted on SVR, XGBoost, and CatBoost using Optuna-SearchCV, while TabPFN was implemented without hyperparameter tuning. The models' performance was measured via mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R^2), and benchmarked. The results revealed that TabPFN demonstrated the highest predictive accuracy with MAE, MSE, RMSE, and R^2 values of 2.27, 13.54, 3.68, and 0.98, respectively, outperforming the benchmark models despite its untuned configuration. SHAP analysis revealed that curing age and water content were the most dominant contributors to the compressive strength prediction. While the study employed only literature dataset, future studies could explore TabPFN's robustness in predicting SCM concrete properties with validation from experimental trials.

Keywords: *TabPFN, metakaolin concrete, compressive strength prediction, XGBoost, ensemble learning*.

Introduction

Concrete is generally recognized as the most utilized material in the construction industry, with an annual demand estimated at 14 billion m^3 (Bărbulescu and Hosen, 2025). However, manufacturing this important material comes with a high environmental cost. It has been reported that Portland cement (PC) production, the major constituent binding material in concrete, constitutes about 5-7% of total annual greenhouse emissions (Miller *et al.*, 2021), with projections to about 1.4 - 3.8 Gt by 2050 (Cheng *et al.*, 2023). Furthermore, the unabated increase in the cost of cement and other construction materials due to growing inflation makes cement highly expensive for construction.

Supplementary cementitious materials (SCMs) have been developed as technically viable alternatives to partially replace Portland cement in concrete production. These materials, usually of natural, industrial, or agro-based origins, also known as pozzolans, generally do not possess cementitious properties on their own, but in finely ground form and in the presence of water and calcium hydroxide, they can react to form compounds with cementitious properties (Becerra-Duitama and Rojas-Avellanda, 2022). They show strong capabilities not only in improving cementitious properties but also enhancing concrete mechanical, durability, and rheological properties (Ikumapayi *et al.*, 2021; Alabi and Mahachi, 2022; Arum *et al.*, 2022; Pradhan *et al.*, 2024). Metakaolin (MK), a product of the

calcination of kaolin clay within a range of 650-800°C, particularly has attracted research interest due to its high reactivity, strong mechanical and durability properties. Studies by Ismail *et al.* (2020) and Sharaky *et al.* (2021) showed that concrete compressive strength increased with metakaolin addition within a range of 5 to 15% in partial replacement of cement. Pillay *et al.* (2022) observed that MK enhanced concrete strength and chloride penetration resistance, and prevented significant deterioration in a coastal environment. MK has also shown promise when applied in recycled aggregate concrete systems, and concrete subjected to high temperature (Verma *et al.*, 2023; Zhong *et al.*, 2024; Haider *et al.*, 2025). Haider *et al.* (2025) reported that MK exhibited strong pozzolanic reactivity and microstructural properties, producing higher density and lower porosity while reducing spalling and mitigating the breakdown of the hydration products at elevated temperatures. Inaty *et al.* (2025) showed that MK exhibits strong resistance to carbonation, acid, and sulfate attacks in geopolymer systems.

Despite these many benefits, the variation in mixture conditions and the sensitivity of its performance to replacement levels, water-binder ratio, and curing age make the modelling and prediction of MK concrete compressive strength a non-trivial challenge. Many modelling approaches have been applied in the prediction and optimization of concrete properties. Earlier models, such as linear regression (LR) and multiple linear regression (MLR) (Ahmed *et al.*, 2022), although simple and easy to interpret, were limited in identifying the compound, non-linear relationships inherent in concrete mix design. Over the past decade, mathematical optimization, such as Scheffe's optimization and Taguchi design methodology were adopted to systematize the mix design space (Akeke *et al.*, 2023; Abunassar and Alas, 2025), but these approaches, being largely dependent on experimental programmes, do not generalize well beyond their calibration range. Other soft computing techniques have been explored, including fuzzy logic inference systems, gene expression programming (GEP) (Akin *et al.*, 2020), which showed greater flexibility, but were frequently challenged by the heterogeneous nature and limited size of concrete datasets.

In the past decade, a surge has been witnessed in the application of machine learning algorithms to the prediction of concrete properties. Some extensively studied algorithms include support vector regression (SVR), a kernel-based method that has foundations in statistical learning theory, and tree-based algorithms such as decision trees (DTs), which depend on recursive splitting of the data space, thus creating a tree-like hierarchical structure that operates using decision rules that can be picked from data features (Mienye and Jere, 2024). Many studies have also explored Gradient boosting ensemble techniques, such as Categorical Boosting (CatBoost), eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), in relation to concrete property prediction, with superior predictive performance relative to classical regression models (Pranav, Lahoti and Gopalarathnam, 2023). Furthermore, interpretability has become an emerging requirement for ML models in concrete and materials engineering. The Shapley Additive eXPlanations (SHAP) framework has become one major tool for meaningfully interpreting the direction and magnitude of input features to model predictions (Uddin *et al.*, 2023).

Despite the breadth of ML studies, particularly relating to the prediction of metakaolin concrete, a significant research gap exists in the application of more advanced and computationally efficient transformer-based tabular learning models in MK concrete compressive strength prediction. Such models, like Tabular Prior-data Fitted Network (TabPFN), are highly capable of robust inference on small-to-medium datasets without overfitting.

TabPFN was introduced by Hollmann *et al.* (2022) and represents a fundamental change from conventional ML model paradigms. Unlike tree or kernel-based algorithms that need to be trained and their hyperparameters tuned on every new dataset, TabPFN employs in-context learning to execute Bayesian approximate inference in a single transformer forward pass, without any dataset-specific training (Hollmann *et al.*, 2025). It is conceptually a prior-data fitted network (PFN) explicitly trained to approximate Bayesian model averaging, unlike traditional ensembles, which rely on empirical risk minimization (ERM) that can easily lead to overfitting on small datasets without heavy regularization. Its architecture is well-suited to concrete materials research, where data is scarce, and assembling datasets of a few hundred records requires extensive laboratory work. However, there

is scarcely any published study benchmarking TabPFN in its untuned configuration against tuned boosted ensemble and kernel-based algorithms for MK concrete compressive strength prediction.

Hence, this study addresses this gap by developing and benchmarking four ML models, including XGBoost, CatBoost, SVR, and TabPFN, for forecasting metakaolin concrete compressive strength. The objectives were to (i) curate a comprehensive metakaolin concrete compressive strength database from literature, (ii) train, tune, and evaluate the performance of XGBoost, CatBoost, and SVR using Bayesian optimization and TabPFN in its zero-shot (untuned) learner configuration, (iii) identify the best-performing model based on R^2, RMSE, MSE, and MAE metrics, and (iv) apply SHAP explainability analysis to extract feature importance rankings and the interaction effects for MK concrete mix design.

Materials and Methods

Data collection and pre-processing

A dataset of 328 records was curated from experimental studies in the literature on the compressive strength of concrete with metakaolin (MK) partially replacing ordinary Portland cement (OPC) (Narmatha and Kala, 2016; Weng *et al.*, 2013; Dinakar, Sahoo and Sriram, 2013; Chen *et al.*, 2020; Ding and Li, 2002; Guneyisi, Gesoglu and Mermerdas, 2008; Nadeem, Memon and Lo, 2014; Mermerdaş *et al.*, 2012; Adekitan, Oyerinde and Jaji, 2015; Akin *et al.*, 2020; Ramezanianpour and Jovein, 2012; Sharaky *et al.*, 2021; Johari *et al.*, 2011; Shafiq *et al.*, 2015). The inclusion criteria comprised the availability of data on MK concrete compressive strength, MK content/dosage, curing age, and mix composition. The exclusion criteria included records with missing key variables and records exhibiting implausible values. The key features extracted include SiO_2, Al_2O_3, Fe_2O_3, and CaO contents of the MK; the water-binder (w/b) ratio; the percentage MK addition (%); the cement content (kg/m^3); the MK content (kg/m^3); the fine aggregate content (kg/m^3); the coarse aggregate content (kg/m^3); the water content (kg/m^3); and the curing age, while the target was the compressive strength.

To ensure data consistency across all literature studies, all compressive strength values were converted to equivalent 100 3 100 3 100 mm cube strengths (Graybeal and Davis, 2008) using a factor of 1.05 to convert 150 mm cubes, 0.95 for 50 mm cubes, and 1.25 for 100 3 200 mm cylinder strengths. This was to ensure robust model training without the introduction of any form of artificial variance. Initial correlation analysis and feature elimination were conducted before dataset splitting to prevent data leakage, and the results revealed a strong dependency between MK (kg/m^3) and the percentage MK addition (%), water (kg/m^3) and water-binder ratio, SiO_2, Al_2O_3, Fe_2O_3, and CaO contents. Specifically, the columns of MK (kg/m^3) and water (kg/m^3) were retained while the percentage MK addition (%) and water-binder ratio columns were excluded since retaining both could create an exact or near-exact linear dependence. Meanwhile, Boosted algorithms and TabPFN are more robust in handling multicollinearity than linear models, hence there was no need to drop the SiO_2, Al_2O_3, Fe_2O_3, and CaO columns since their chemical composition may contribute meaningfully to the interpretability and learning capacity of the models. The final dataset thus comprised 11 input features and 1 target variable (compressive strength). A total of 24 records out of the whole dataset had missing values in two features: coarse aggregates and superplasticizer content, arising from incomplete reporting in a source publication. The missing values were addressed using k-Nearest Neighbours (kNN) imputation with k=5, by determining the Euclidean distance between an incomplete record and all the complete records across the non-missing features, identifying the five closest neighbours and using their distance-weighted mean to replace the missing value. To prevent data leakage, this imputation was fitted only on the training dataset and applied to the validation and test sets using the training-derived parameters to prevent data leakage.

Dataset split

A data split of 70/15/15 percent was adopted for training, validation, and testing, respectively, and the random seed was fixed at 42 to ensure reproducibility, giving a total number of 229 samples for training, 49 for validation, and 50 for testing. The training set was exclusively employed for model fitting and Bayesian hyperparameter optimization, while the test set was held out completely from all model development activities, including hyperparameter tuning, and accessed only after all the models had been finalized. This was to prevent data leakage and to give unbiased results of the generalization performance on unseen data.

Feature Scaling

Feature scaling was applied selectively based on the requirements of each model's algorithm. Since the RBF kernel in Support Vector Regression (SVR) uses a Euclidean distance computation and is highly sensitive to differences in feature magnitude and scale, StandardScaler (z-score normalization) was applied to the input features before training the SVR model (Nguyen, 2016). On the other hand, the features were left unscaled for XGBoost and CatBoost because they are invariant to monotonic feature transformations. Meanwhile, TabPFN performs internal normalization as part of its in-context learning mechanism and is therefore inherently scale-invariant. Hence, the features supplied to the TabPFN model were also left unscaled in accordance with its default usage convention.

Model Training

The models explored in this study include support vector regression (SVR), eXtreme gradient boosting (XGBoost), categorical boosting (CatBoost), and TabPFN. These models were selected due to their relative efficiency in handling structured data and high-dimensional features, and provide a good means of benchmarking the performance of TabPFN against kernel-based and ensemble boosting algorithms. The k-fold cross-validation technique, utilizing a 5-fold split, was used during the training process to ensure accuracy during the training. Additionally, a Bayesian optimization approach using Optuna search CV algorithm was employed in tuning the hyperparameters of the SVR, XGBoost, and CatBoost models, while the TabPFN model was left untuned, being a prior data-fitted model. The key hyperparameters tuned for the models include the learning rate, minimum sample leaf, maximum depth, minimum sample split, and number of estimators.

The performance of the models was measured via mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R^2), represented by Equations (1) – (4).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(t-y)^2 \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t-y)^2} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|t-y| \tag{3}$$

$$R^2 = \frac{(\sum_{i=1}^{n}(t-\bar{t})(y-\bar{y}))^2}{\sum_{i=1}^{n}(t-\bar{t})^2\sum_{i=1}^{n}(y-\bar{y})^2} \tag{4}$$

where y represents the predicted value of the compressive strength, t represents the experimental value of the compressive strength, $\bar{y}$ denotes the average of y, $\bar{t}$ denotes the average of t. R^2 values close to 1, and low RMSE, MSE, and MAE values indicate good model performance. SHapley Additive exPlanation (SHAP) analysis was conducted after the model training to evaluate feature significance on the compressive strength prediction.

Machine learning models

Support Vector Regression (SVR)

Support Vector Regression (SVR) was developed by Vapnik (2000), based on the statistical learning theory where a regression function lying within an insensitive tube is sought around the training targets with the simultaneous minimization of the L2 norm of the weight vector. The goal of SVR is to find an optimal plane that minimizes the error between the plane and the training samples. The optimization problem is defined in Equation (5) as:

$$\begin{cases} \min \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ s.t. \begin{cases} y_i - \omega\Phi(x_i) - b \le \epsilon + \xi_i \quad , i = 1,2,\ldots,l \\ -y_i + \omega\Phi(x_i) + b \le \epsilon + \xi_i^* \\ \xi_i \ge 0, \xi_i^* \ge 0 \end{cases} \end{cases} \tag{5}$$

where C is the penalty factor, and the higher its value, the greater is the penalty for the sample error over, and represents the insensitivity margin, and are slack variables (Zhang *et al.*, 2024). The radial basis function (RBF) kernel was evaluated during Optuna hyperparameter optimization, and StandardScaler was used in transforming the input features before SVR training.

eXtreme Gradient Boosting (XGBoost)

XGBoost algorithm is a decision trees-based algorithm with a gradient boosting procedure that has been used by many researchers because of its excellent performance and high speed in calculations (Cui *et al.*, 2021). It creates an additive ensemble of weak learners (shallow trees) in a

sequential, residual-correcting manner, and minimizes losses by computing the second-order gradients of the loss function along with advanced regularization (L1 and L2). This regularization method enhances generalization and reduces overfitting; hence, it shows a strong performance on structured datasets and has been very effective in the prediction of material properties (Zhang *et al.*, 2026).

Categorical Boosting (CatBoost)

CatBoost is a gradient boosting framework, created by Prokhorenkova *et al.* (2018), that makes use of symmetric (oblivious) decision trees and ordered boosting. To mitigate the prediction shift and target leakage issues that plague traditional gradient boosting decision tree algorithms on short datasets, ordered boosting uses a permutation-based target statistics methodology that computes leaf values using only previous training observations at each step (Pranav *et al.*, 2023). On moderately large datasets, the symmetric tree structure produces quicker inference and more stable generalisation by enforcing balanced binary splits at every depth level. To further minimize overfitting on sparse data, CatBoost additionally employs ordered bootstrapping during tree building.

TabPFN

Specifically created for small-to-medium tabular datasets, TabPFN is a transformer-based meta-learning model (Hollmann *et al.*, 2022). In contrast to traditional machine learning algorithms, which need to be retrained on every new dataset, TabPFN uses in-context learning (ICL) to generate predictions in a single forward pass without gradient-based retraining. The entire training set is encoded as a sequence of context tokens and passed into the transformer's multi-head self-attention mechanism alongside each test query. By explicitly conditioning on observed training data, this approach enables TabPFN to approximate Bayesian posterior predictive inference (Hollmann *et al.*, 2025). This study employed TabPFN version 2.0, which was pretrained on over 130 million synthetic datasets collected using Bayesian neural network priors and structural causal models.

Bayesian Hyperparameter Optimization

Optuna (v3), an open-source automatic hyperparameter optimisation (AutoML) framework which uses the Tree-structured Parzen Estimator (TPE) algorithm as its main sequential model-based optimization (SMBO) sampler, was used to perform hyperparameter optimization for XGBoost, CatBoost, and SVR. Two non-parametric density models, $l(x)$, which represent the distribution of hyperparameter configurations producing high objective scores, and $g(x)$, which represent poor-performing configurations, are maintained and updated by TPE. New candidate configurations are then chosen by maximising the acquisition ratio $l(x)/g(x)$ (Akiba *et al.*, 2019). In terms of sample efficiency, this Bayesian method significantly outperforms grid and random search, especially for high-dimensional hyperparameter spaces. Maximizing the R^2 score on the validation set was the optimization goal. MedianPruner was employed to break unpromising trials early based on intermediate R^2 scores, thus reducing unnecessary computation.

SHapley Additive exPlanation (SHAP) Explainability Analysis

In this study, TabPFN and the best-performing model among the other three (SVR, XGBoost, and CatBoost) were subjected to SHapley Additive exPlanation (SHAP) analysis to gain physically interpretable insights into the parameters affecting MK concrete compressive strength. SHAP was developed by Lundberg and Lee (2017) based on cooperative game theory. It evaluates the average marginal contribution of each feature over all potential feature coalitions to quantify each feature's contribution to individual forecasts. SHAP values are a theoretically sound and model-neutral explanation framework because they meet the requirements of local correctness, missingness, and consistency (Lundberg *et al.*, 2020).

A TabPFN-compatible SHAP explainer (KernelExplainer) was used for TabPFN, which approximates Shapley values by evaluating the model's predictive function as a black-box oracle over a background reference dataset. Meanwhile, for XGBoost and CatBoost, SHAP values were computed using the TreeExplainer, an exact, computationally efficient algorithm specifically designed for tree-based ensemble models that exploits the tree structure to compute exact Shapley values in polynomial time, avoiding the exponential complexity of the general Shapley formulation.

In order to provide multi-level interpretability, the following SHAP visualisation outputs were produced: (i) SHAP summary beeswarm plots,

which display SHAP value distributions for each feature across all test samples and their direction of effect; (ii) SHAP mean absolute importance bar charts, which rank features according to their average contribution magnitude; and (iii) SHAP dependency graphs for the top-ranked features, which show possible interaction effects and show how SHAP values change as a function of feature value. Together, these outputs allow ML predictions to be translated into practical mix design recommendations for MK concrete.

Results and Discussion

Descriptive Statistics and Exploratory Data Analysis

Descriptive analysis was conducted on the dataset after general pre-processing steps. The summary statistics of the dataset are presented in Table 1, and a box plot of the compressive strength values is presented in Figure 1. The compressive strength ranged from 11.17 MPa to 126 MPa, with an average value of 53.77 MPa and a standard deviation of 26.83 MPa. No outliers were observed; however, the records show that the dataset comprises both normal and high-strength concrete mixtures (especially where admixtures are used). In general, the compressive strength demonstrated a relatively normal distribution with a moderate skewness of 0.418, and this can favour machine learning generalization.

It is also observed from Table 1 that among the oxide compositions, SiO_2 exhibited a mean value of 57.12% with a positively skewed distribution (1.434), while Al_2O_3 showed a negatively skewed distribution (-1.237). The strong inverse skewness between SiO_2 and Al_2O_3 suggests a compositional tradeoff that is inherent in aluminosilicate systems (Abdeen *et al.*, 2026). The water content and curing age were the most highly skewed variables, with skewness values of 2.383 and 2.942, respectively. The curing age distribution suggests that the majority of the specimens were tested at early ages, while relatively fewer samples were cured for extended durations, such as 180 or 365 days. This observation is common in many concrete databases compiled from multiple experimental studies.

The histogram and kernel density plots are presented in Figure 2. The plot reveals that quite a number of variables exhibit multimodal and non-normal distributions. SiO_2, Al_2O_3, Fe_2O_3, CaO, water content, and curing age in particular showed multimodal tendencies, suggesting that the dataset may have been compiled from multiple experimental studies or mix-design regimes. Such multimodal behaviours are common in concrete databases and reflect variations in the SCM composition, material sourcing, curing, and mix design methodologies. SiO_2 and Al_2O_3 showed bimodal tendencies, and further supports the compositional dependency between these oxides.

Table 1. Summary statistics of the variables in the MK dataset

Parameter		Mean	Standard Deviation	Minimum	Maximum	Skewness
Input Variable	SiO_2	57.12	8.08	51.2	74.3	1.434
	Al_2O_3	37.12	9.42	17.8	45.8	-1.237
	Fe_2O_3	1.66	1.51	0.3	5.3	1.185
	CaO	0.82	1.21	0	3.38	1.449
	Cement (kg/m^3)	410.77	119.16	251.47	834.94	1.895
	MK (kg/m^3)	51.07	47.92	0	256	1.659
	Fine Agg. (kg/m^3)	780.77	143.21	573.5	1017.5	0.192
	Coarse Agg. (kg/m^3)	1044.16	191.34	765	1646.09	0.960
	Water (kg/m^3)	186.65	71.51	126	417.47	2.383
	Curing Age (days)	43.33	56.53	1	365	2.942
	Superplasticizer (kg/m^3)	5.36	5.47	0	18.3	0.743
Output variable	Compressive Strength (MPa)	53.77	26.83	11.17	126	0.418

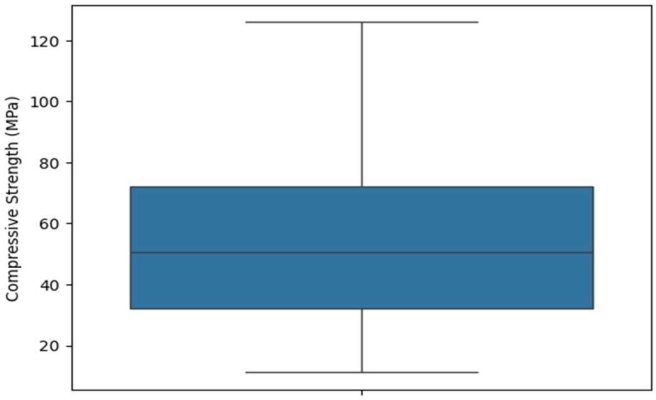

Figure 1. Box plot of the MK dataset compressive strength values

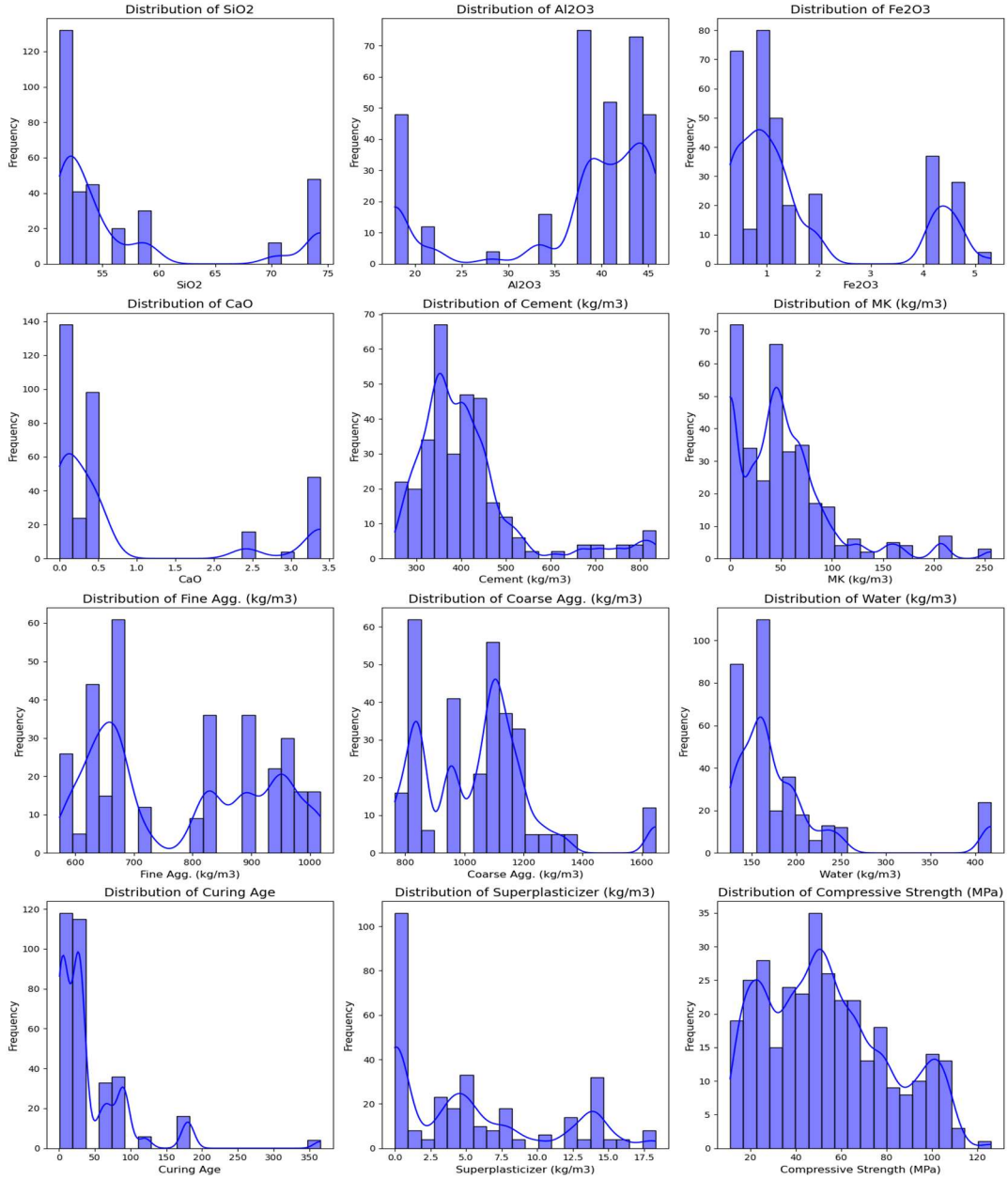

Figure 2. Histograms and KDE plots of the features and target of the MK dataset

Correlation Analysis

A Pearson correlation analysis was carried out to investigate the linear relationships among the variables and identify the primary factors influencing compressive strength, and presented in a correlation heatmap, as shown in Figure 3.

The results showed that curing age had the highest positive correlation with compressive strength (r = 0.45), confirming the strong influence of hydration progress on compressive strength development. Fe_2O_3 also exhibited a moderate positive correlation with compressive strength (r = 0.34) while superplasticizer dosage had a weak positive correlation (r = 0.18). On the other hand, water content demonstrated the strongest negative correlation with compressive strength (r = -0.43), aligning well with Abram's law of water addition, which establishes that an inverse relationship exists between water content and concrete strength (Li *et al.*, 2020). Other variables such as coarse aggregate content (r = -0.097), metakaolin content (r = -0.10), and fine aggregate content (r = -0.057) showed relatively weak linear relationships with compressive strength. The heatmap also shows that substantial interdependence exists among the oxide compositions. A strong negative correlation was noticed between SiO_2 and Al_2O_3 (r = -0.97), and between CaO and Al_2O_3, while there was a noticeable positive correlation between CaO and SiO_2. These relationships suggest a strong interdependence among the oxides.

Variance Inflation Factor (VIF) analysis was further used to assess multicollinearity among the input features, and presented in Table 2. While all remaining variables showed acceptable VIF less than 5, VIF values for Al_2O_3 (87.88), SiO_2 (39.24), and CaO (20.53) showed high multicollinearity among the oxides. Although these VIF values were above the conventional thresholds, the oxide variables were retained in the present study, particularly because they contribute substantially to the hydration behaviour and pozzolanic reactivity in metakaolin concrete systems. Additionally, the boosted ensemble models and TabPFN algorithms adopted in this study are less sensitive to multicollinearity than traditional regression methods.

The findings from the EDA, such as the multimodal behaviour, nonlinear feature interactions, and multicollinearity, show that the choice of advanced nonlinear machine learning techniques for this study is justified.

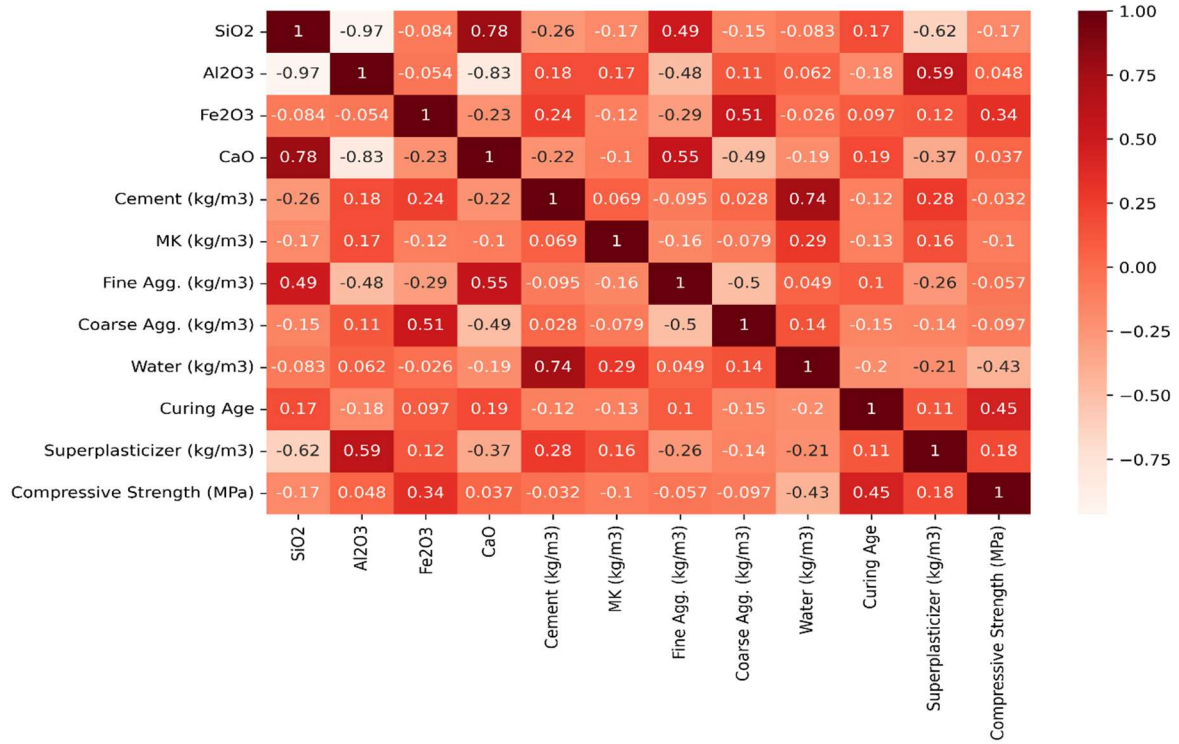

Figure 3: Correlation heatmap of the features of the MK dataset

Table 2. Variance Inflation Factor (VIF) analysis for the input features for the MK dataset

	Feature	VIF
0	const	9470.624
1	SiO_2	39.240
2	Al_2O_3	87.882
3	Fe_2O_3	4.061
4	CaO	20.533
5	Cement (kg/m^3)	2.250
6	MK (kg/m^3)	1.727
7	Fine Agg. (kg/m^3)	3.235
8	Coarse Agg. (kg/m^3)	4.667
9	Water (kg/m^3)	1.897
10	Curing Age (days)	1.175
11	Superplasticizer (kg/m^3)	2.362

Model Performance Comparison

The results of the model training on the baseline parameters are presented in Table 3. The performance evaluation metrics indicate that the boosted ensemble models and the TabPFN performed well, with TabPFN showing the best performance with a mean CV R^2 of 0.978, MAE of 2.278, MSE of 13.543, and RMSE of 3.680.

Hyperparameter Optimization Results

A Bayesian technique was employed in tuning the hyperparameters of the SVR, XGBoost, and CatBoost models via the Optuna search CV algorithm to improve their predictive capabilities, and the results are presented in Table 4. The results revealed an improvement in the predictive capacities of the three models, with all four metrics increasing beyond the performance of the baseline parameters. Support Vector Regression exhibited the greatest improvement in performance with a change in MSE value from 908.08 to 133.18, and mean CV R2 value from -0.0198 to 0.7405. Nevertheless, SVR showed the least performance among all the models. XGBoost and CatBoost also showed moderate improvement upon hyperparameter tuning, with an increase in the mean CV R^2 from 0.9381 to 0.9668 for XGBoost, and 0.9638 to 0.9724 for CatBoost. However, it was observed that none of the tuned kernel-based and ensemble boosting models performed as high as the TabPFN model, which was left untuned. The superior performance of TabPFN compared to the tuned boosting algorithms may be attributed to its transformer-based probabilistic learning framework and its ability to leverage prior-learned representations from large collections of synthetic tabular tasks. Unlike conventional boosting models that require iterative hyperparameter optimization, TabPFN showed strong generalization ability and faster adaptation on the relatively small and heterogeneous dataset. This reduces dependence on the computationally expensive tuning procedures required by boosting models. The findings suggest that TabPFN has significant potential for efficiently predicting concrete properties, especially in concrete material datasets, which are usually characterized by limited sample sizes and complex non-linear relationships. The final tuned hyperparameters are presented in Table 5.

The plots of the predicted vs actual compressive strength values for training and testing of the concrete samples is shown in Figures 4, 6, 8, and 10, while Figures 5, 7, 9, and 11 represent the comparison of the predicted and experimental data values, together with associated error values after the tuning of the hyperparameters. The graphs demonstrate that a significant relationship exists between the predicted and actual compressive strengths. A high statistical correlation can be observed between the prediction and the actual data, with a low range of error in the accuracy of the models, especially for the TabPFN, XGBoost, and

CatBoost models. SVR showed satisfactory performance as well, but not with as high predictive accuracy as the other models. These results suggest that boosted ensemble and prior data-fitted models are more nuanced in handling multimodal and highly variable concrete data compared to standard linear and kernel-based models.

Table 3: Results of model training on baseline parameters

Model	**Train R^2**	**Validation R^2**	**Test R^2**	**MSE (Test)**	**RMSE (Test)**	**MAE (Test)**	**Mean CV R^2**
Support Vector Regression	0.0130	0.0457	-0.0025	908.0787	30.1343	24.7209	-0.0198
XGBoost	0.9999	0.9679	0.9633	33.2214	5.7638	3.9985	0.9381
CatBoost	0.9987	0.9871	0.9756	22.1363	4.7049	3.2555	0.9638
TabPFN	0.9973	0.9912	0.9850	13.5427	3.6800	2.2780	0.9781

Table 4: Results of model training on Bayesian hyperparameter optimization

Model	**Train R^2**	**Validation R^2**	**Test R^2**	**MSE (Test)**	**RMSE (Test)**	**MAE (Test)**	**Mean CV R^2**
Support Vector Regression	0.8906	0.8830	0.8530	133.1800	11.5404	8.2400	0.7405
XGBoost	0.9964	0.9839	0.9736	23.8991	4.8887	3.2483	0.9668
CatBoost	0.9988	0.9862	0.9776	20.2999	4.5055	3.2062	0.9724
TabPFN	0.9973	0.9912	0.9850	13.5427	3.6800	2.2780	0.9781

Table 5. Final tuned hyperparameters used in training the models

Model	Key Hyperparameters	Value
Support Vector Regression	Scaler	StandardScaler
	C	34.9867
	Epsilon	0.00484
	gamma	0.09013
	n_trials	25
XGBoost	n_estimators	310
	learning_rate (eta)	0.12517
	max_depth	3
	colsample_bytree	0.74771
	min_child_weight	2
	n_trials	60
CatBoost	Iterations	541
	learning_rate	0.16494
	depth	4
	l2_leaf_reg	3.96385
	loss_function	RMSE
	n_trials	60
TabPFN	N/A	N/A

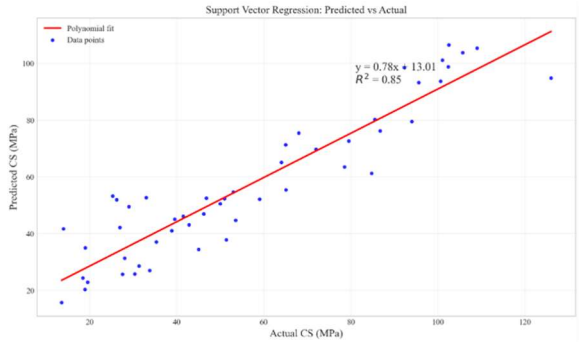

Figure 4. Scatter plot of predicted vs actual values for compressive strength (SVR model)

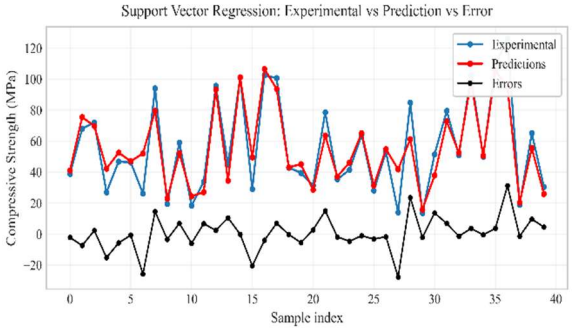

Figure 5. Predicted vs experimental compressive strength with residual error distribution (SVR)

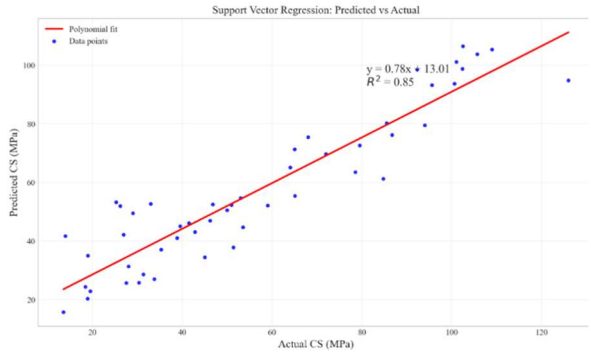

Figure 6. Scatter plot of predicted vs actual values for compressive strength (XGBoost model)

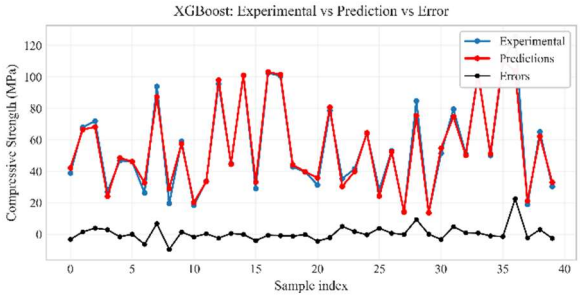

Figure 7. Predicted vs experimental compressive strength with residual error distribution (XGBoost)

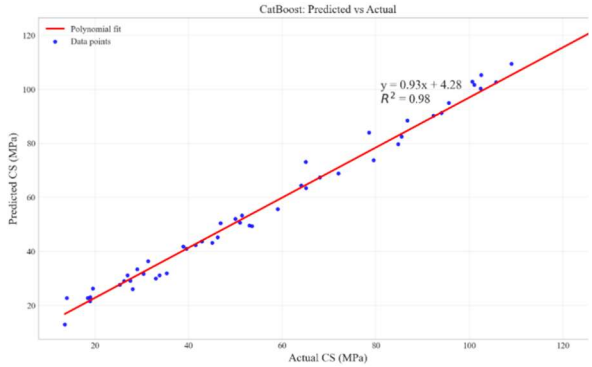

Figure 8. Scatter plot of predicted vs actual values for compressive strength (CatBoost model)

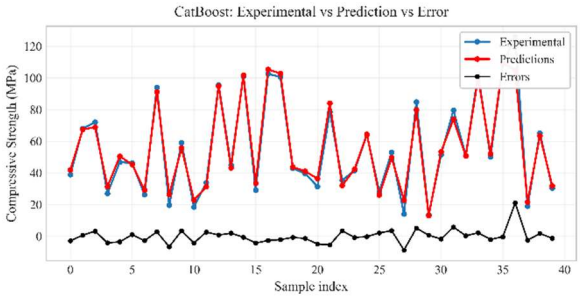

Figure 9. Predicted vs experimental compressive strength with residual error distribution (CatBoost)

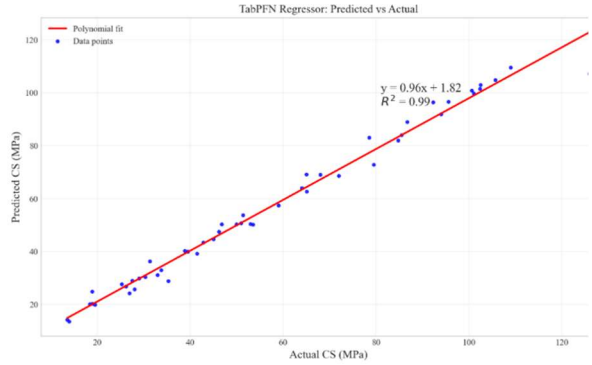

Figure 10. Scatter plot of predicted vs actual values for compressive strength (TabPFN model)

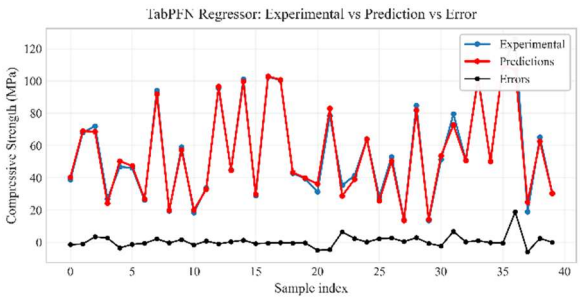

Figure 11. Predicted vs experimental compressive strength with residual error distribution (TabPFN)

SHAP Feature Importance

The feature importance sensitivity analysis for the CatBoost model and the TabPFN model was conducted via SHAP, and the results are plotted in Figures 12 – 17.

The SHAP summary plots for both the CatBoost and TabPFN models (Figures 12 and 13) revealed that the curing age exerted the strongest influence on metakaolin compressive strength prediction, followed by the water content. The beeswarm plots (Figures 14 and 15) further show that higher curing ages tend to bring a significant positive influence on the compressive strength, whereas higher water contents tend to negatively affect the compressive strength. This is a phenomenon that is supported by established concrete and material science, where the time of hydration, water-binder ratio, and long-term pozzolanic reaction contribute more significantly to strength at later ages. In both models, higher values of Fe_2O_3 oxide content showed a positive contribution to the compressive strength. However, TabPFN ranked SiO_2 contribution higher than that of Fe_2O_3, suggesting that TabPFN's transformer architecture may be identifying a more nuanced relationship between silica content and compressive strength. These observations suggest that the oxide interdependencies in metakaolin concrete systems play a substantial role in compressive strength development.

It was also observed that coarse aggregate and metakaolin (MK) content bring a moderate contribution to the model output, with median ranged values of MK content contributing more positively to the strength. This suggests that the optimal range of metakaolin addition for positive effects on the compressive strength should be in the median range of about 10 to 15% partial replacement of cement.

The SHAP dependence plots (Figures 16 and 17) for both models show a rapid increase in strength from 0 to 28 days, which gradually plateaus as it approaches 90 and 360 days. TabPFN's interaction plot shows that at higher curing ages, the silica content is a major strength determinant. This aligns with the pozzolanic mechanistics where silica from metakaolin continues to react with calcium hydroxide over long periods to form C-S-H gels, giving additional strength to the concrete.

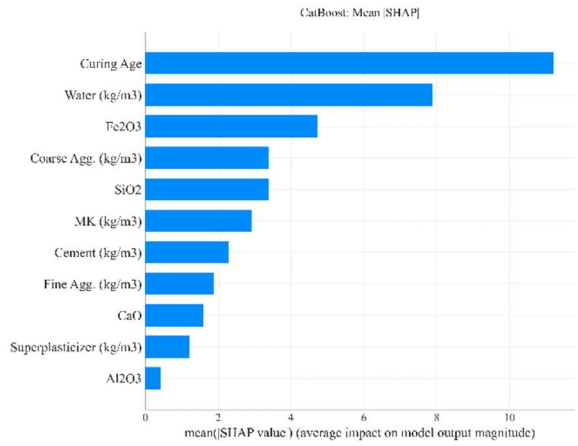

Figure 12. SHAP Summary plot ranking impact of features on compressive strength prediction (CatBoost model)

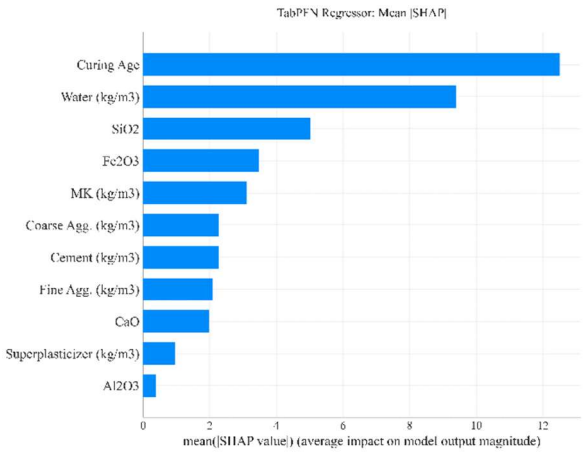

Figure 13. SHAP Summary plot ranking impact of features on compressive strength prediction (TabPFN model)

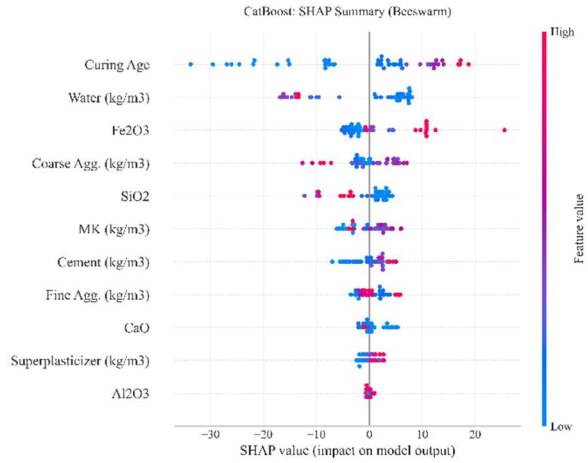

Figure 14. SHAP Beeswarm plot showing feature influence on compressive strength prediction (CatBoost model)

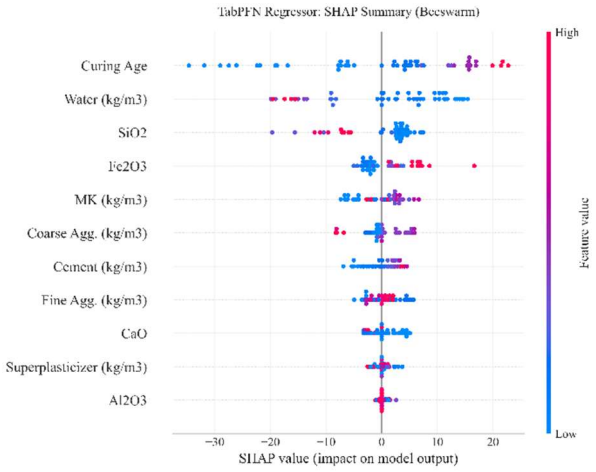

Figure 15. SHAP Beeswarm plot showing feature influence on compressive strength prediction (TabPFN model)

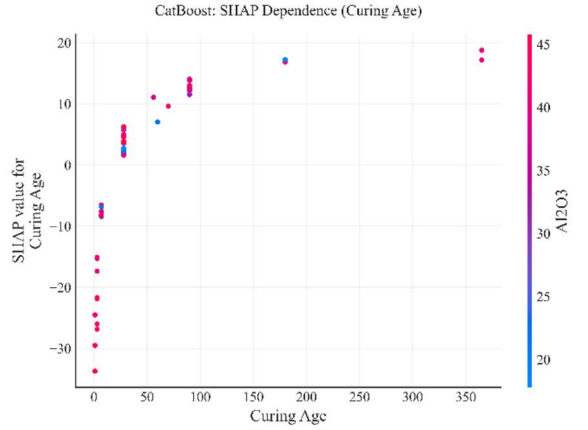

Figure 16. SHAP Dependence plot showing the relationship between curing age and Al_2O_3 (CatBoost model)

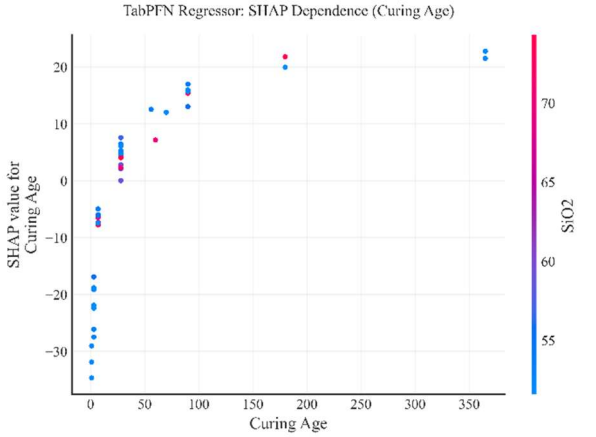

Figure 17. SHAP Dependence plot showing the relationship between curing age and SiO_2 (TabPFN model)

Conclusions

This study demonstrates the capability of a prior data-fitted model (TabPFN) in predicting metakaolin concrete compressive strength in comparison with a kernel-based method (support vector regression) and ensemble boosting machine learning techniques (XGBoost and CatBoost). From the results of the models and analysis, the following conclusions can be drawn:

i. TabPFN, a prior data-fitted model, in its zero-shot learner configuration, outperforms both ensemble boosting models, such as XGBoost and CatBoost, and the kernel-based model considered in this study (Support vector regression).
ii. Correlation analysis indicated that curing age had the highest positive correlation with compressive strength, while the water content showed the strongest negative correlation with strength.
iii. TabPFN, as well as both XGBoost and CatBoost, showed strong predictive capabilities, recording R^2 values greater than 0.96; TabPFN outperformed the other models with a 0.978 R^2 score. SVR showed the least performance after hyperparameter tuning with an R^2 value of 0.74.
iv. SHAP analysis revealed that the curing age and water content contributed most significantly to the compressive strength prediction, with higher values of curing age positively affecting the compressive strength, and higher values of

water content negatively contributing to strength. Furthermore, SHAP summary analysis indicated nonlinear interactions among curing age, water content, and SiO_2 content. Higher SiO_2 values contributed positively to strength development, suggesting that silica-rich metakaolin systems enhance long-term pozzolanic reactivity and strength gain. Conversely, higher water contents produced strongly negative strength contributions, especially at lower curing ages. This highlights the coupled influence of hydration maturity and water demand on strength evolution. These findings demonstrate the capability of explainable AI techniques to capture complex interactions among mix-design parameters.

In general, this research contributes to knowledge by demonstrating the robustness of TabPFN, a prior-data fitted model for metakaolin concrete compressive strength prediction. The model revealed higher predictive accuracy and a more nuanced identification of the interaction between features affecting the compressive strength of concrete.

Limitations and Future Research Direction

One limitation of this research is that the dataset for the study was solely sourced from the literature, without any laboratory data. Future research could explore the potential of prior-data fitted models in the prediction of metakaolin concrete durability properties and other SCM-concrete systems. The impact of stacking and bagging of boosted and prior-data fitted models could also be investigated.

References

Abdeen, H. *et al.* (2026) 'Interdependence of key molar ratios (SiO_2/Al_2O_3 and Al_2O_3/Na_2O) in metakaolin-based geopolymers: phase composition, microstructure and mechanical insights'. Rochester, NY: Social Science Research Network. Available at: https://doi.org/10.2139/ssrn.6282906.

Abunassar, N. and Alas, M. (2025) 'Optimization of strength and durability properties of rubberized concrete mixtures containing silica fume using Taguchi method', *Construction and Building Materials*, 468, p. 140455. Available at: https://doi.org/10.1016/j.conbuildmat.2025.140455.

Adekitan, O., Oyerinde, A.O. and Jaji, M.B. (2015) 'Comparative Compressive Strength Assessment of Cement Concretes Blended With Two Locally Produced Natural Pozzolans', *Procs 4th Applied Research Conference in Africa. (ARCA) Conference, 27-29 August 2015. 4th Applied Research Conference in Africa. (ARCA) Conference*, Ibadan, Nigeria (In: Mojekwu,J.N., Nani G., Atepor, L., Thwala,W.D., Ogunsumi, L., Awere E., Ocran,S.P., and Bamfo-Agyei, E. (Eds)), pp. 130–145.

Ahmed, H.U. *et al.* (2022) 'Statistical Methods for Modeling the Compressive Strength of Geopolymer Mortar', *Materials*, 15(5). Available at: https://doi.org/10.3390/ma15051868.

Akeke, G.A. *et al.* (2023) 'Experimental investigation and modelling of the mechanical properties of palm oil fuel ash concrete using Scheffe's method', *Scientific Reports*, 13(1), p. 18583. Available at: https://doi.org/10.1038/s41598-023-45987-3.

Akiba, T. *et al.* (2019) 'Optuna: A next-generation hyperparameter optimization framework', *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631.

Akin, O.O. *et al.* (2020) 'Prediction of the Compressive Strength of Concrete Admixed with Metakaolin Using Gene Expression Programming', *Advances in Civil Engineering*, 2020(1), p. 8883412. Available at: https://doi.org/10.1155/2020/8883412.

Alabi, S.A. and Mahachi, J. (2022) 'Performance assessment of mechanical and durability properties of cupola slag geopolymer concrete with fly and rice husk ashes', *Nigerian Journal of Technological Development*, 19(1), pp. 27–38.

Arum, R.C., Arum, C. and Alabi, S.A. (2022) 'The highs and lows of incorporating pozzolans into concrete and mortar: A review on strength and durability', *Nigerian Journal of Technology*, 41(2), pp. 197–211.

Bărbulescu, A. and Hosen, K. (2025) 'Cement Industry Pollution and Its Impact on the Environment and Population Health: A Review', *Toxics*, 13(7), p. 587. Available at: https://doi.org/10.3390/toxics13070587.

Becerra-Duitama, J.A. and Rojas-Avellanda, D. (2022) 'Pozzolans: A review', *Engineering and*

Applied Science Research (EASR), 49(4), pp. 495–504.

Chen, J.J. *et al.* (2020) 'Cement Equivalence of Metakaolin for Workability, Cohesiveness, Strength and Sorptivity of Concrete', *Materials*, 13(7). Available at: https://doi.org/10.3390/ma13071646.

Cheng, D. *et al.* (2023) 'Projecting future carbon emissions from cement production in developing countries', *Nature Communications*, 14(1), pp. 1–12. Available at: https://doi.org/10.1038/s41467-023-43660-x.

Cui, L. *et al.* (2021) 'Application of extreme gradient boosting based on grey relation analysis for prediction of compressive strength of concrete', *Advances in Civil Engineering*, 2021.

Dinakar, P., Sahoo, P.K. and Sriram, G. (2013) 'Effect of Metakaolin Content on the Properties of High Strength Concrete', *International Journal of Concrete Structures and Materials*, 7(3), pp. 215–223.

Ding, J.-T. and Li, Z. (2002) 'Effects of metakaolin and silica fume on properties of concrete', *ACI Materials Journal*, 99, pp. 393–398.

Graybeal, B. and Davis, M. (2008) 'Cylinder or cube: strength testing of 80 to 200 MPa (11.6 to 29 ksi) ultra-high-performance fiber-reinforced concrete', *ACI Materials Journal*, 105(6), p. 603.

Guneyisi, E., Gesoglu, M. and Mermerdas, K. (2008) 'Improving strength, drying shrinkage, and pore structure of concrete using metakaolin', *Materials and Structures*, pp. 1–13. Available at: https://doi.org/10.1617/s11527-007-9296-z.

Haider, I. *et al.* (2025) 'Investigating the synergistic effects of Metakaolin and silica fume on the strength and durability of recycled aggregate concrete at elevated temperatures', *Scientific Reports*, 15(1), p. 29510. Available at: https://doi.org/10.1038/s41598-025-11494-w.

Hollmann, N. *et al.* (2022) 'Tabpfn: A transformer that solves small tabular classification problems in a second', *arXiv preprint arXiv:2207.01848* [Preprint].

Hollmann, N. *et al.* (2025) 'Accurate predictions on small data with a tabular foundation model', *Nature*, 637(8045), pp. 319–326.

Ikumapayi, C.M., Arum, C. and Alaneme, K.K. (2021) 'Reactivity and hydration behavior in groundnut shell ash based pozzolanic concrete', *Materials Today: Proceedings*, 38, pp. 508–513.

Inaty, F.E. *et al.* (2025) 'Mechanical and durability performance of metakaolin and fly ash-based geopolymers compared to cement systems', *Results in Engineering*, 27, p. 105788. Available at: https://doi.org/10.1016/j.rineng.2025.105788.

Ismail, M.H., Rusly, N.S.M. and Deraman, R. (2020) 'Strength and Water Absorption of Concrete Containing Metakaolin and Nylon Fiber', *International Journal of Sustainable Construction Engineering and Technology*, 11(1), pp. 230–242.

Johari, M.M. *et al.* (2011) 'Influence of supplementary cementitious materials on engineering properties of high strength concrete', *Construction and Building Materials*, 25(5), pp. 2639–2648.

Li, S., Yang, J. and Zhang, P. (2020) 'Water-Cement-Density Ratio Law for the 28-Day Compressive Strength Prediction of Cement-Based Materials', *Advances in Materials Science and Engineering*. Edited by S.A. Memon, 2020(1), p. 7302173. Available at: https://doi.org/10.1155/2020/7302173.

Lundberg, S.M. *et al.* (2020) 'From local explanations to global understanding with explainable AI for trees', *Nature machine intelligence*, 2(1), pp. 56–67.

Lundberg, S.M. and Lee, S.-I. (2017) 'A unified approach to interpreting model predictions', *Advances in neural information processing systems*, 30.

Mermerdaş, K. *et al.* (2012) 'Strength development of concretes incorporated with metakaolin and different types of calcined kaolins', *Construction and Building Materials*, 37, pp. 766–774. Available at: https://doi.org/10.1016/j.conbuildmat.2012.07.077.

Mienye, I. and Jere, N. (2024) 'A Survey of Decision Trees: Concepts, Algorithms, and Applications', *IEEE Access*, PP, pp. 1–1. Available at: https://doi.org/10.1109/ACCESS.2024.3416838.

Miller, S.A. *et al.* (2021) 'Achieving net zero greenhouse gas emissions in the cement industry via value chain mitigation strategies', *One Earth*, 4(10), pp. 1398–1411. Available at: https://doi.org/10.1016/j.oneear.2021.09.011.

Nadeem, A., Memon, S.A. and Lo, T.Y. (2014) 'The performance of Fly ash and Metakaolin concrete at elevated temperatures', *Construction and Building Materials*, 62, pp. 67–76. Available at: https://doi.org/10.1016/j.conbuildmat.2014.02.073.

Narmatha, M. and Kala, D. (2016) 'Meta kaolin – The Best Material for Replacement of Cement in Concrete', *IOSR Journal of Mechanical and Civil Engineering*, 13, pp. 66–71. Available at: https://doi.org/10.9790/1684-1304016671.

Nguyen, L. (2016) 'Tutorial on support vector machine', *Applied and Computational Mathematics (ACM)* [Preprint].

Pillay, D.L. *et al.* (2022) 'Engineering performance of metakaolin based concrete', *Cleaner Engineering and Technology*, 6, p. 100383. Available at: https://doi.org/10.1016/j.clet.2021.100383.

Pradhan, S.S. *et al.* (2024) 'Effects of rice husk ash on strength and durability performance of slag-based alkali-activated concrete', *Structural Concrete*, 25(4), pp. 2839–2854. Available at: https://doi.org/10.1002/suco.202300173.

Pranav, S., Lahoti, M. and Gopalarathnam, M. (2023) 'Concrete Compressive Strength Prediction Using Boosting Algorithms', in S.B. Singh et al. (eds) *Fiber Reinforced Polymeric Materials and Sustainable Structures*. Singapore: Springer Nature, pp. 307–315. Available at: https://doi.org/10.1007/978-981-19-8979-7_26.

Prokhorenkova, L. *et al.* (2018) 'CatBoost: unbiased boosting with categorical features', *Advances in neural information processing systems*, 31.

Ramezanianpour, A.A. and Jovein, H.B. (2012) 'Influence of metakaolin as supplementary cementing material on strength and durability of concretes', *Construction and Building materials*, 30, pp. 470–479.

Shafiq, N. *et al.* (2015) 'Calcined kaolin as cement replacing material and its use in high strength concrete', *Construction and Building Materials*, 81, pp. 313–323.

Sharaky, I.A. *et al.* (2021) 'Experimental and theoretical study on the compressive strength of the high strength concrete incorporating steel fiber and metakaolin', *Structures*. Elsevier, pp. 57–67.

Uddin, M.N. *et al.* (2023) 'Interpretable XGBoost–SHAP machine learning technique to predict the compressive strength of environment-friendly rice husk ash concrete', *Innovative Infrastructure Solutions*, 8(5), p. 147. Available at: https://doi.org/10.1007/s41062-023-01122-9.

Vapnik, V. (2000) *The Nature of Statistical Learning Theory*. 2nd edn. New York, NY: Springer (Information Science and Statistics). Available at: https://doi.org/https://doi.org/10.1007/978-1-4757-3264-1.

Verma, M. *et al.* (2023) 'Analysis of the properties of recycled aggregates concrete with lime and metakaolin', *Materials Research Express*, 10(9), p. 095508. Available at: https://doi.org/10.1088/2053-1591/acf983.

Weng, T.-L., Lin, W.-T. and Cheng, A. (2013) 'Effect of Metakaolin on Strength and Efflorescence Quantity of Cement-Based Composites', *The Scientific World Journal*, 2013(1), p. 606524. Available at: https://doi.org/10.1155/2013/606524.

Zhang, J. *et al.* (2026) 'Hybrid Explainable Machine Learning Models with Metaheuristic Optimization for Performance Prediction of Self-Compacting Concrete', *Buildings*, 16(1). Available at: https://doi.org/10.3390/buildings16010225.

Zhang, W., Liu, D. and Cao, K. (2024) 'Prediction of concrete compressive strength using support vector machine regression and non-destructive testing', *Case Studies in Construction Materials*, 21, p. e03416. Available at: https://doi.org/10.1016/j.cscm.2024.e03416.

Zhong, C. *et al.* (2024) 'An investigation on mechanical properties and durability of metakaolin reinforced modified recycled concrete', *Case Studies in Construction Materials*, 20, p. e02978. Available at: https://doi.org/10.1016/j.cscm.2024.e02978.