scientific reports

Check for updates

OPEN

Machine learning-based prediction of crack mouth opening displacement in ultra-high-performance concrete

Arsalan Mahmoodzadeh[1], Manish Kewalramani[2], Abdulaziz Alghamdi[3], Anwar Ahmed[4], Shtwai Alsubai[5], Abdullah Alqahtani[5], Abed Alanazi[5] & Sivaprakasam Palani[6]✉

Ultra-high-performance concrete (UHPC) has become an essential construction material due to its exceptional strength, durability, and crack-resistance properties, making it well-suited for long-span bridges, protective structures, and demanding infrastructure applications. However, accurately predicting crack mouth opening displacement (CMOD) in fiber-reinforced UHPC (FR-UHPC) presents significant challenges, as traditional empirical and physics-based models struggle to capture the complex nonlinear relationships between mix design, fiber geometry, and structural parameters. This study assembled a comprehensive experimental database containing eleven mix and material variables that control post-cracking behavior. Nine advanced machine learning algorithms were trained and evaluated using fivefold cross-validation, including kernel-based regressors, ensemble methods, deep neural networks, and the innovative tabular prior-data fitted networks (TabPFN). The modela were ensured interpretability through SHapley Additive exPlanations (SHAP), sensitivity analysis, contour mapping, and interaction diagrams. These analyses consistently showed that fiber volume (FV) and fiber length (FL) were the primary factors controlling CMOD, while silica fume content (SF), fly ash content (FA), initial notch depth (a_0), and water-to-binder ratio (w/b) had secondary effects. Feature selection improved predictive performance by narrowing the input space to the six most influential variables. With this optimized configuration, TabPFN delivered exceptional accuracy (R2 = 0.942, RMSE < 0.072 mm), surpassing ensemble methods (XGBoost, RFR, GBR) and kernel-based approaches (SVR, NuSVR, GPR). Model predictions were validated against experimental CMOD curves, and bootstrap resampling generated 95% confidence intervals, demonstrating that TabPFN provided both high precision and reliable uncertainty estimates. This research contributes three main innovations: (i) creation of a detailed experimental database for FR-UHPC fracture behavior, (ii) first application of TabPFN for CMOD prediction, and (iii) combined focus on interpretability and uncertainty quantification. The transparent, uncertainty-aware predictive framework developed here connects data-driven modeling with fracture mechanics, providing practical tools for designing and optimizing resilient UHPC structures.

Keywords Crack mouth opening displacement, Ultra-high-performance concrete, Machine learning, Statistical analysis

Background and motivation

The reliability of hydraulic engineering works (dams, levees, and spillways) depends fundamentally on their structural integrity, which governs engineering safety, the management of water resources, flood mitigation, and energy production[1]. Yet, the occurrence of concrete cracking, whether limited to fine surface hairlines or

[1]Center of Research and Strategic Studies, Lebanese French University, Erbil, Iraq. [2]Department of Civil Engineering, College of Engineering, Abu Dhabi University, Abu Dhabi, UAE. [3]Department of Civil Engineering, Faculty of Engineering, University of Tabuk, 47512 Tabuk, Saudi Arabia. [4]Department of Civil Engineering, College of Engineering, Northern Border University, 73222 Arar, Saudi Arabia. [5]Department of Computer Science, College of Computer Engineering and Sciences in Al-Kharj, Prince Sattam Bin Abdulaziz University, P.O. Box 151, 11942 Al-Kharj, Saudi Arabia. [6]Department of Mechanical Engineering, College of Engineering, Addis Ababa Science and Technology University, Po Box 16417, Addis Ababa, Ethiopia. ✉email: arissss.de@gmail.com; shiva@aastu.edu.et

extending to serious through-section voids, constitutes a recurrent and troubling issue that shortens the service life of these crucial assets[2,3]. The origins of cracking are rarely isolated; instead, they stem from the interplay of concrete characteristics, changing environmental loads, and variations in workmanship[2].

At the beginning of 1956, researchers formally identified the main drivers of concrete displacement as water pressure, temperature, and aging, all interacting in a nonlinear way[4]. Non-uniform thermal fields during hydration, shrinkage from water evaporation, and uneven loading can set up tensile stresses that pass the concrete's tensile strength, causing cracks to start and then spread[5,6]. As global temperatures rise, the temperature component of this process is likely to grow more pronounced[7].

Recent numerical and experimental investigations have addressed the characteristics of crack opening behavior under eccentric loading[8,9]. The maximal load sustained by the specimen at various loading stages is formulated as the product of the axial compressive force, the slenderness ratio, the eccentricity coefficient, and the reduction factor associated with the slotted cross-section[8]. The experimental work further demonstrates that the incorporation of basalt fiber into the concrete matrix markedly influences the response of columns subjected to heavy eccentric compressive loads[9]. The resultant displacement and moment distribution is captured by a governing differential equation, while the maximum crack opening conditioned by large eccentricity is represented by a derived analytical expression[1].

Crack opening displacement (COD) is widely recognized as a key parameter that reflects both the long-term performance of concrete structures and the fundamental mechanisms of crack growth. Because of this dual importance, researchers have devoted significant effort to quantifying and predicting COD in both experimental and numerical studies[10]. Tada et al.[11] compiled empirical equations for computing COD in homogeneous materials subjected to different loading scenarios. These equations have become standard in studies tracking crack growth in concrete[12–14]. To refine the characterization of concrete fracture, Shah[15] advocated the three-point bending test to obtain COD measurements from concrete beams. Barr et al.[16] performed flexural tests on fiber-reinforced concrete specimens, correlating COD to the observed deflection. Similarly, Ding[17] reevaluated test results from notched steel fiber-reinforced beams, confirming a linear COD–deflection relationship. Aslani et al.[18] assessed the effect of different fiber types on COD through a series of controlled tests. More recently, Zhang and Ansari[19] introduced a method for in-situ COD measurement at the crack tip via embedded optical fiber sensors.

Recent developments in predicting COD have seen the application of the extended finite element method (XFEM) to concrete fracture analyses. Aghajanzadeh and Mirzabozorg[20] employed XFEM to model the fracture process in concrete beams, successfully tracking COD evolution. Following this, Ma et al.[21] studied how varying the initial crack length influences COD, reinforcing the method's capacity for parametric exploration. Yang et al.[22] extended the approach to self-compacting lightweight aggregate concrete, quantitatively mapping COD changes throughout the cracking sequence.

Several researchers have refined purely analytical strategies for COD estimation. Accornero et al.[23,24] and Rubino et al.[25] advanced a bridged crack model, combining fracture mechanics with displacement compatibility to describe the cracking response of steel-fibre-reinforced concrete beams, ultimately linking internal displacements to COD measurements in cracked zones. Drawing on the differing bond properties linking reinforcement and concrete, Fu et al.[10] outlined a technique for calculating COD by tracing the nonlinear strain profile across the two materials and assessing the sectional rotation of the cracked beam. Their findings demonstrated that, with escalating load, the bond stiffness correlating the reinforcement to the concrete progressively declines, triggering nonlinear variations in the tensile strain of the reinforcement.

Progress achieved through experimental, analytical, and numerical methods for quantifying COD has been substantial, yet each method carries technical limitations that restrict its applicability and predictive consistency. While experimental campaigns remain essential for identifying fundamental fracture processes, their inherent demands for equipment, time, and meticulous sample preparation typically confine studies to a restricted range of material grades, geometries, and loading configurations. This narrow experimental domain hampers the extrapolation of results to more complex architectural or large-scale structural applications. Numerical methods such as XFEM yield informative data on crack growth and size, yet their reliability hinges on the fidelity of the chosen material laws, the adoption of appropriate element sizes, and the prohibitive run times, especially for multidimensional or progressive-loading scenarios. Analytical solutions, despite their clarity and concise formulation, are often formulated under idealized scenarios and overlook the influential, yet inseparable, effects of concrete microstructural variability, the toughening imparted by fibers, transient environmental actions, and the kinetic history of the loading. As a result, although each methodology has enriched the discipline of concrete fracture mechanics, their collective capacity to provide rapid and extendable prognostics of crack displacement over a wide range of concrete types and service environments is limited.

In this context, machine learning techniques set a new, highly adaptable standard. Unlike classical methods, machine learning algorithms naturally capture intricate, nonlinear, multivariate dependencies straight from experimental records, bypassing the need for pre-defined constitutive laws or rupture models[26–31]. By harnessing expansive, heterogeneous datasets, machine learning frameworks reveal subtle trends in COD, from microstructure and reinforcement shape to curing schedules and applied loads. This evidence-based strategy permits swift and precise forecasts, creating a broadly applicable route that can match or even exceed the predictive consistency and computational speed of classic fracture mechanics. Additionally, machine learning naturally supports uncertainty assessment, importance-ranking of features, and interpretable artificial intelligence (AI) models, ensuring that the resulting forecasts are both accurate and comprehensible, thereby reinforcing their reliability for engineering decision-making.

Today, machine learning methods have shown significant ability to solve various engineering problems[32–35]. Despite the increasing use of machine learning across structural and materials engineering disciplines[36–38], investigations specifically targeting COD prediction remain limited and unevenly distributed. Earlier

machine learning studies related to concrete mechanics have predominantly concentrated on properties such as compressive strength[39], flexural strength[40], tensile properties[41], stress–strain constitutive model[42], and carbonation depth[43], while COD has attracted relatively minimal focus. As presented in Table 1, the few investigations that do consider COD and similar applications are frequently constrained by small, diverse datasets or narrow analyses restricted to specific types of fiber-reinforced concrete mixes. This uneven coverage reveals a pressing research need: a coordinated effort to compile systematic, expansive, and high-fidelity COD datasets matched with sophisticated machine learning frameworks that can decipher the complex interactions between concrete formulation, reinforcement types, and crack development. Filling this void is essential for the discipline, as accurate COD models underpin reliable service-life predictions, informed fracture toughness assessments, and the optimized design of durable hydraulic infrastructures.

BPNN, Back propagation neural network; SHC, Self-healing concrete; GEP, Gene expression programming; SVR, Support vector regression; XGBoost, Extreme gradient boosting; FRCB, Fibre-reinforced concrete beams; MLP, Multilayer perceptron; neural network, RF, Random forest; ANFIS, Adaptive neuro-fuzzy inference system; LMSR, least mean squares regression; GFRP-RC, Glass fiber-reinforced polymer-reinforced concrete; LightGBM, Light gradient boosting machine; AdaBoost, Adaptive boosting; GB, Gradient boosting; KNN, K-nearest neighbors.

This study introduces several innovations in predicting crack mouth opening displacement (CMOD) in ultra-high-performance concrete (UHPC). UHPC is a composite material built on ordinary Portland cement that blends high-volume fractions of high-strength microsteel fibers with highly cementitious ingredients using a low water-to-binder ratio[48]. UHPC has gained significant importance in construction projects because of its exceptional compressive strength, robust microstructure, and enhanced durability. These properties deliver improved performance and extended service life for structures like bridges, high-rise buildings, defense installations, and specialized facilities[49]. This research develops a new experimental dataset that covers a wide range of material mixtures, fiber geometries, and structural behaviors, which proves essential for training reliable machine learning models. The work employs a novel integration of nine advanced machine learning algorithms, including the Tabular Prior-Data Fitted Networks (TabPFN), which hasn't been applied to CMOD prediction in UHPC before. This approach differs from earlier studies that mostly relied on traditional numerical and experimental methods or simpler machine learning models, often constrained by narrow datasets or specific fiber-reinforced concrete mixtures. The study also emphasizes model interpretability by applying SHapley Additive Explanations (SHAP), which reveals insights into the physical significance of parameters like fiber volume, fiber length, and silica fume content. The research incorporates uncertainty quantification through bootstrap resampling, offering calibrated confidence intervals (CIs) for CMOD predictions. These contributions create a comprehensive, data-driven, and interpretable framework for CMOD prediction that addresses gaps in previous research while providing a robust tool for structural diagnostics and design optimization.

Research significance and novelty

This research represents a significant step forward in the realm of fiber-reinforced UHPC by merging cutting-edge machine learning techniques with fracture mechanics and structural design. Accurate prediction of CMOD is vital for assessing the post-cracking behavior, serviceability, and long-term durability of fiber-reinforced UHPC systems. Conventional empirical and physics-based models often fall short in accounting for the intricate nonlinear interplay among mix proportions, fiber attributes, and geometric configurations, leading to compromise on both the predictive precision and broader applicability of the results.

A significant contribution of this work is the creation of a detailed and novel experimental database that addresses the flexural behavior of fiber-reinforced UHPC. The dataset seamlessly combines microstructural and mechanical variables, illuminating the delicate interactions among binder chemistry, fiber geometry, and CMOD. By acquiring high-resolution results over an extensive range of parameters, the collection delivers a robust platform for the development and verification of next-generation machine learning tools. Its novelty

References	Input(s)	Output	Material	Method	Results
[1]	Water level, Precipitation, Maximum air temperature, Minimum air temperature	CCO	Concrete	BPNN	$R^2 = 0.82$
[44]	Fly-Ash, Silica Fume, Limestone Powder	Crack width	SHC	GEP	$R^2 = 0.938$
[45]	–	COD	Rock	SVR, MLP, XGBoost, RF	R^2 = above 0.97
[46]	Concrete cover, Distance of the outer reinforcement layer, Cross-section width, Reinforcement diameter, Outer area of the tension reinforcement layer, Elastic modulus of steel reinforcement, Tensile strain in the outer tensile reinforcement layer, Lap-spliced length, Diameter of compressive rebars, Length of beam, Strain in compressive rebars, Concrete compressive strength, Yield stress of compressive rebars, Cross-section height, Steel fibres content	Crack spacing	FRCB	MLP, ANFIS, SVR, LMSR	Maximum $R^2 = 0.99$
[47]	Bar strain, Bar spacing, Concrete cover, Reinforcement ratio, Concrete compressive strength, Bar diameter, Crack width	Crack width	GFRP-RC	XGBoost, LightGBM, AdaBoost, GB, RF, KNN, Ridge, Lasso	RMSE: (XGBoost: 0.140, LightGBM: 0.131, AdaBoost: 0.107, GB: 0.125, RF: 0.123, KNN: 0.143, Ridge: 0.130, Lasso: 0.128)

Table 1. Summary of previous studies on machine learning applications for crack opening displacement and related predictions in different materials.

guarantees that conclusions drawn are not merely indicative of predictive fidelity but push forward the collective grasp of fiber-reinforced UHPC response, filling an important void in current literature while establishing a reference point for subsequent data-driven inquiry into high-performance concrete systems.

Leveraging advanced machine learning frameworks (specifically, TabPFN, NuSVR, SVR, GPR, ANN, GBR, DTR, RFR, and XGBoost), this study achieves a level of prediction accuracy and reliability that is previously unattained. Comprehensive benchmarking is conducted through rigorous fivefold cross-validation and evaluation across multiple metrics, including coefficient of determination (R^2), root mean square error (RMSE), and variance accounted for (VAF). Additionally, the research dissects the influence of feature selection on the CMOD prediction. These investigations furnish actionable recommendations for both numerical simulations and experimental investigations, enhancing the practical utility of the models in design and research.

The adoption of SHAP-based interpretability significantly enhances the relevance of this research by furnishing mechanistic foundations for each prediction the model generates. These investigations bridge the gap between data-driven modeling and the physics of fracture mechanics, converting previously opaque "black-box" results into practical guidance for the tailored design and refinement of impact-resistant UHPC systems.

Additionally, the study embeds uncertainty quantification using bootstrapped CIs, thereby enriching the predictive engine with a vital layer of reliability. This approach is indispensable in engineering practice, where safety thresholds and material variability govern success. This dual emphasis on accuracy and uncertainty-aware interpretability positions the study at the forefront of predictive modeling for cementitious materials, enabling engineers to make informed design decisions and paving the way for more resilient and optimized FR-UHPC structures. Overall, this research propels the discipline of concrete fracture prediction by:

- Integrating state-of-the-art machine learning with transparent and physically grounded modeling.
- Employing a novel dataset obtained through experimental test.
- Pinpointing and corroborating the dominant features that govern post-cracking response.
- Delivering a cohesive pipeline that fuses forecasting, mechanistic understanding, and uncertainty assessment.
- Building a solid groundwork for the future data-driven enhancement of UHPC.

The schematic diagram in Fig. 1 provides a clear overview of the research framework, showing the progression through experimental database construction, machine learning implementation, model evaluation, interpretability analysis, and uncertainty quantification.

Materials and methods

Machine learning models

Many complex systems have been mathematically modeled using machine learning. In the viewpoint of civil and structural engineering, machine learning methods have shown much potential[50–52]. They are particularly effective in handling issues involving multiple interconnected factors and nonlinearities. Machine learning techniques can be considered as an efficient technique in various areas of research involving composite structures[52]. Their capacity to handle complex datasets and extract meaningful patterns makes them excellent for simulating the behavior of objects such as reinforced concrete[51].

For this study, a varied array of machine learning techniques was deployed, spanning kernel-based regressors, ensemble strategies, and deep learning frameworks. Each algorithm was chosen for its capacity to reveal nonlinear dependencies, its resilience to small sample sizes, and its unique strengths in forecasting CMOD for the fiber-reinforced UHPC specimens.

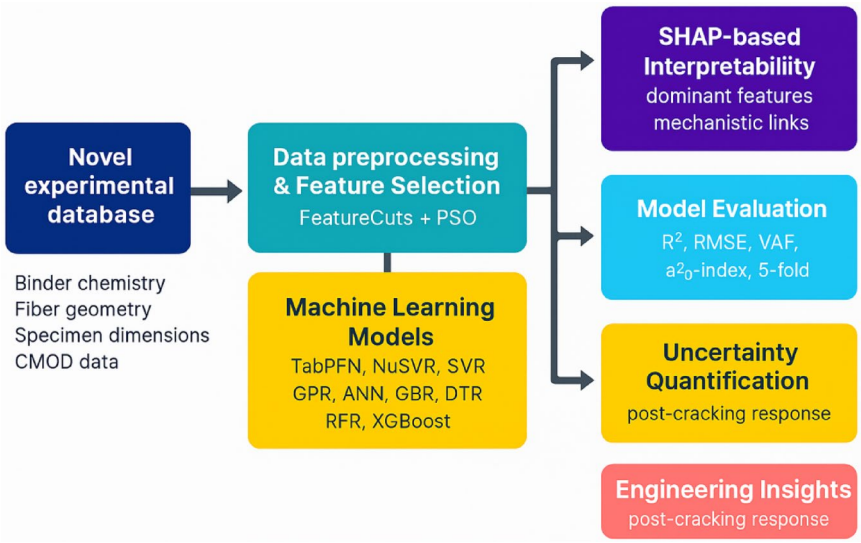

Fig. 1. Schematic framework of the present study.

TabPFN
TabPFN builds on a transformer backbone and is probabilistically trained on a broad spectrum of tabular benchmarks. By utilizing attention layers, it captures intricate feature interactions without explicit feature engineering and, critically, it produces predictive distributions rather than point estimates[53]. This probabilistic output lends itself to reliable uncertainty quantification.

NuSVR and SVR
These kernel regressors map the input space into a higher-dimensional manifold where nonlinear patterns can be treated as linear separable. SVR employs an ε-insensitive loss to tolerate small deviations from target values[54], while NuSVR's parameter ν governs the trade-off between the number of support vectors and the tolerance of errors[55]. Together, they proficiently appreciated the nonlinear interactions between input features and target.

GPR
GPR operates as a flexible, non-parametric framework, placing a Gaussian process prior over the space of functions. The model's output is a predictive mean accompanied by a confidence region, which is critical when estimating uncertainty in intricate material responses[56]. By utilizing a tailored kernel, GPR specifies how input features influence one another, allowing the model to smoothly interpolate experimental observations while simultaneously adapting to the nonlinear dependencies present in the UHPC data.

ANN
Drawing motivation from biological neural networks, an ANN serves as a robust function approximator, adept at modeling intricate, non-linear interactions among input features[57]. Because of its flexible architecture, it can effectively memorize and generalize the intricate dependencies present in concrete mixture formulation and mechanical performance. Nonetheless, the technique mandates deliberate regularization strategies and a disciplined training regimen to mitigate the risks of overfitting, which is particularly salient when working with datasets of intermediate scale.

GBR
GBR builds its predictive capability through a series of weak, shallow trees, each trained to rectify the errors of the ensemble so far[58]. Sequential adjustment of the residuals leads to a strong, coherent prediction. In the realm of UHPC, this method and effectively models the nonlinear coupling between inputs, while its ensemble nature conferrs a desirable balance between bias reduction and resilience to overfitting, especially when the dataset is noisy or of limited size.

DTR
DTR works by repeatedly splitting the data according to the features until the reduction in prediction error becomes minimal[59]. Its structure makes it easy to understand, yet it's prone to overfitting in intricate systems such as fiber-reinforced UHPC. This overfitting becomes evident when the model responds excessively to small variations in data or when crucial variables are intentionally or unintentionally excluded. Each of these factors can lead to unreliable predictions when the model encounters unseen data.

RFR
RFR builds upon the single-tree framework by constructing many such trees, each one trained on a different random subset of both rows and columns[60]. Predictions are then averaged, which smoothes out the noise and variance that a single tree would amplify. This ensemble strategy results in a more robust model that can better handle the complex interactions present in UHPC data.

XGBoost
XGBoost represents a refined and efficient implementation of gradient boosting that targets both predictive accuracy and computational speed. The algorithm builds an ensemble of decision trees sequentially, with each new tree fitted to the residuals of the previous ensemble. This approach progressively reduces bias while improving overall performance. The algorithm incorporates regularization terms into its objective function, helping prevent overfitting by penalizing excessive model complexity[61]. Shrinkage (learning rate) controls how much each tree contributes to the final prediction, while column subsampling boosts robustness by working with feature subsets. Parallelized tree construction and efficient sparse data handling make the algorithm both fast and scalable.

XGBoost has found widespread success in predicting mechanical properties of cementitious materials. Researchers have applied it to estimate splitting tensile strength of basalt fiber reinforced coral aggregate concrete[62], compressive strength of recycled aggregate concretes[63], and flexural strength of steel-fiber-reinforced concretes[64]. These applications demonstrate the algorithm's capacity to capture complex nonlinear relationships in heterogeneous material systems. The XGBoost was selected as one of the benchmarking ensemble learners for CMOD prediction in this study based on these proven capabilities. The algorithm effectively balances accuracy and variance while maintaining robustness when features are perturbed.

Machine-learning model selection and rationale

The modeling stage balanced four competing objectives: (1) capture complex, nonlinear relationships between CMOD and the experimental features; (2) provide reliable uncertainty estimates useful for engineering decisions; (3) maintain interpretable diagnostics that can be linked to fracture mechanics; and (4) benchmark modern, high-performing algorithms against simpler baselines. We deliberately selected representatives from

several algorithmic families to satisfy these goals: kernel methods, probabilistic (Bayesian) models, tree-based ensembles, single decision trees, and neural networks, plus a recent transformer-style tabular prior (TabPFN). Table 2 lists the main motivations for each choice.

Computational environment and hyperparameter tuning

The entire workflow for developing and evaluating the machine learning models was carried out in the Jupyter Notebook environment, running Python 3.7 via the Anaconda Navigator. The environment was based on an Intel Core i7-10750H processor clocked at 2.60 GHz, paired with 32 GB of RAM. This setup comfortably handles the simultaneous training of multiple models, SHAP calculations, and cross-validation without performance degradation, ensuring that the planned analytical procedures can be executed with the reliability they demand.

Each machine learning model received dedicated hyperparameter tuning through exhaustive Grid Search on the training data selecting parameter ranges based on well-established defaults from the literature and some initial trial runs. For instance, we looked at the learning rate within the range of [0.001–1.0], the number of estimators from [50–500], maximum depth between[2–50], and regularization parameters (like α in GPR and C in SVR) across logarithmic scales (10^{-3}–10^{3}). To avoid any data leakage, we used a nested cross-validation strategy for hyperparameter tuning: the inner folds focused on optimizing the hyperparameters, while the outer folds assessed how well the model generalized. This approach made sure that our model selection didn't skew the final metrics we reported. Table 3 presents the final optimized hyperparameters for each model.

Dataset preparation

Experimental program

As in Fig. 2, a dense UHPC blend featuring Portland cement, silica fume, fine quartz or silica sand (< 2 mm), and, when specified, fly ash was prepared according to defined target proportions. The dry constituents (cement, silica fume, and fly ash) were mixed together for 2–3 min to achieve uniform dispersion. About 80% of the total water was introduced at low speed to dampen the fine powders. The superplasticizer was pre-diluted in the remaining water and added gradually over the next 2–3 min, until the blend reached the desired cohesive and highly deformable mortar consistency. Fresh mix temperature was controlled to remain below 30 °C by pre-cooling aggregates or by incorporating chilled mixing water. The workability was checked using a mini-slump or flow spread procedure tailored for UHPC, and any fine-tuning of consistency was carried out by varying the superplasticizer dosage, ensuring that the water-to-binder ratio (w/b) remained unchanged.

Steel fibers meeting the specified length-to-diameter ratio were employed. To determine the required amount per cubic meter, the planned volume fraction was combined with the density of steel, taken as 7850 kg/m3. The fibers were dosed steadily into the mixer over a 2–4 min period (Fig. 3), at low to medium speed, to minimize agglomeration and guarantee a consistent spread. Mixing continued for 2 to 3 more minutes, focusing on the even incorporation of the fibers while preventing any clusters or unintended directional orientation.

Molds were fabricated to the precise beam dimensions outlined in the experimental protocol, with the width and depth adjusted for the measured specimen thickness and overall height. The clear-span length was set to the specified span-to-depth ratio, accurately positioning the load and supports for the test setup. The molds received a light film of form release agent before the pour. The freshly prepared UHPC was placed in two lifts, each compacted on a vibrating table or through brief internal vibration to eliminate air voids and keep the fibers evenly suspended. The upper surface was then struck off and smoothed with a steel trowel.

Right after casting, all specimens were sealed to keep moisture inside. Those meant for standard moist curing had their molds removed after 24 ± 2 h and were then placed in either lime-saturated water or a fog room kept at 20 ± 2 °C with at least 95% relative humidity, maintained until the planned curing ages. For specimens undergoing heat or steam curing, the temperature schedule began once initial setting was complete, progressing smoothly to 90 °C, holding constant for 24 h, and then cooling down to 20 °C before moving to final storage. Once the curing phase was complete, all specimens were acclimatized to the laboratory environment (20 ± 2 °C)

Model	Algorithm family	Why chosen (role in study)	Main strengths	Main limitations
SVR / NuSVR	Kernel methods	Strong small-data nonlinear regressor; good bias/variance control via regularization	Effective with few samples; well-understood regularization	Sensitive to feature scaling; kernel selection needs tuning
GPR	Probabilistic, non-parametric	Provides predictive distributions and interpretable kernels (uncertainty baseline)	Principled uncertainty, good for smooth functions, interpretable hyperparams	Computational cost $O(n^3)$; scalable to moderate n only
RFR	Ensemble—bagging	Robust baseline capturing interactions; reduces variance	Handles mixed features, low tuning sensitivity	May underfit compared to boosting for complex signals
GBR	Ensemble—boosting	High predictive power via sequential residual fitting	High accuracy, captures complex patterns	Sensitive to hyperparameters; risk of overfitting if not tuned
XGBoost	Gradient boosting (efficient)	State-of-the-art boosting with regularization and speed	Highly performant, efficient, robust to missing values	Many tuning knobs; gradient boosting can be data-hungry
DTR	Single decision tree	Interpretable baseline; visualise splitting rules	Simple, transparent	High variance; limited predictive accuracy
ANN	Deep learning	Flexible nonlinear approximator to test deep models on tabular data	Expressive, scalable with data	Requires tuning, risk of overfitting on small datasets
TabPFN	Transformer-based tabular prior	Strong small-data probabilistic learner; provides calibrated predictive distributions	Good small-data performance, probabilistic outputs	Relatively new; architecture hyperparameters and priors matter

Table 2. Summary of the machine learning algorithms employed in this study, their family, rationale for selection, key strengths, and limitations in predicting CMOD of FR-UHPC.

Model	Hyperparameter	Explored range/value	Optimized value/type
TabPFN	Device		"cuda" (if available) else "cpu"
	Batch size	{32, 64, 128}	64
	Max features	{50, 100, 200}	100
	N Ensemble configurations	{16, 32, 64}	32
ANN	Hidden layers	{[16,16], [32,32], [32,32,32], [64,64]}	[32, 32, 32] neurons, ReLU activation
	Output layer		1 neuron, ReLU activation
	Optimizer	{"adam", "sgd"}	"adam"
	Loss		"mse"
	Batch size	{8, 16, 32}	8
	Epochs	{50, 100, 200}	100
NuSVR	Kernel		"rbf"
	C	[1, 10, 100, 200, 500]	200
	Coef0		0
	Cache size	[50, 100, 150, 200, 250]	200
	Gamma	[0.001, 0.01, 0.02, 0.05]	0.02
	Max Iter		− 1 (no limit)
	Nu	[0.1, 0.2, 0.3, 0.4, 0.5]	0.38
	Shrinking		TRUE
	Tolerance (tol)		0.003
SVR	Kernel		"rbf"
	C	[1, 10, 20, 50, 100]	20
	Gamma	[0.001, 0.005, 0.006, 0.01]	0.006
	Epsilon	[0.01, 0.05, 0.1, 0.2]	0.1
GBR	Loss		"squared_error"
	Learning rate	[0.01, 0.05, 0.1, 0.3, 0.65]	0.65
	Max leaf nodes		20
	Max depth	[3, 5, 10, 20, None]	None
	Min samples leaf	[10, 20, 30, 50]	30
	Warm start		FALSE
	Validation fraction		0.1
	Iteration no change		5
	Tol		1.0E-08
	N estimators	[50, 100, 200, 500]	100
	Max features	{"auto", "sqrt", "log2"}	"auto"
	Max depth		50
	Min samples split		5
XGBoost	N estimators	[50, 100, 200, 500]	120
	Max depth	[3, 6, 10, 15]	15
	Objective		"reg:squarederror"
	Learning rate	[0.01, 0.05, 0.1, 0.3]	0.01
	Colsample by tree	[0.5, 0.7, 1.0]	0.7
GPR	Kernel	{RBF, Constant + RBF, Dot Product + RBF}	Constant + RBF
	Alpha	[1e-4, 1e-3, 1e-2, 0.008, 0.01]	0.008
DTR	Criterion	{"mse", "mae", "friedman_mse", "squared_error"}	"squared_error"
	Random state		0

Table 3. Optimized heper parameters of the machine learning models.

for a minimum of 3 h before any testing took place, and their widths and depths were measured to an accuracy of ± 0.1 mm.

A finely cut notch was produced in each specimen with a water-cooled diamond saw. The cut width was consistently held between 2 and 5 mm. The remaining ligament height, $h_{sp} = h - a_0 h$, was evaluated at three positions along the span, and the mean value was then used for further calculations. To monitor off-set opening, knife-edge clips were fixed across the notch at mid-span, positioned perpendicular to the anticipated crack face. Either a clip-on CMOD gauge or a high-precision displacement sensor was then fastened to the knife edges. The entire measurement chain was calibrated immediately before each test, and the gauge was re-centered using a light initial compressive load.

Fig. 2. Laboratory preparation and curing process of UHPC specimens.

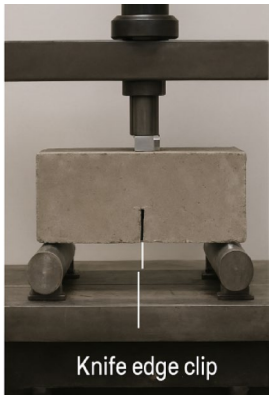

Fig. 3. Three-point bending test.

Testing was carried out using a three-point bending setup where each specimen rested on a pair of steel rollers that defined a clear span matching the specified span-to-depth ratio. A third cylindrical roller was positioned to exert a load directly over the mid-line notch. The procedure used a closed-loop CMOD control strategy to maintain steady crack advancement and to characterize the post-peak reduction in load. Initially, the cross-head speed was set to 0.05 mm/min until the CMOD reached 0.10 mm, after which the rate was raised to 0.20 mm/min until the specified final CMOD value was attained. Load and CMOD readings were logged continuously at a rate of at least 10 Hz, covering the entire duration of each test.

The limit of proportionality load (F_L) was obtained from the measured load-CMOD curves as the peak load recorded for CMOD values up to 0.05 mm. Residual loads (F_i) were recorded at CMOD increments of 0.5 mm, 1.5 mm, 2.5 mm, and 3.5 mm. The associated flexural tensile strengths were then calculated using Eqs. 1 and 2.

$$f_L = \frac{3F_L l}{2bh_{sp}^2} \tag{1}$$

$$f_{R,i} = \frac{3F_i l}{2bh_{sp}^2} \tag{2}$$

where l is the clear span, b is the specimen width, and h_{sp} is the residual ligament height. The complete load–CMOD curves were also used to determine fracture energy values through numerical integration.

Quality control measures involved checking the fresh-state density of every batch, discarding any specimens that showed obvious defects like fiber balling or air voids, and confirming dimensional tolerance before testing. For every mix design and curing age, a minimum of three replicate specimens were tested, and the average values along with the standard deviations were documented. Safety precautions were strictly observed throughout the project. Respirators were mandatory when handling silica fume, cut-resistant gloves and face shields protected personnel when manipulating steel fibers, and protective shields were in place during mechanical tests to prevent injury from spalling fragments.

Data normalization

To make sure every input parameter played its fair role in training the machine learning model, we first standardized the raw experimental data. The dataset included features that were sorted into widely different numerical intervals and physical units: geometric ratios (dimensionless), mechanical strengths (MPa), and blend proportions as mass percentages of the binder. If we had skipped the normalization, the features with the highest absolute values could have distorted the model parameter optimization, drowning out the impact of smaller-scale features that were nevertheless critical for prediction. Normalization harmonizes the numerical range of the inputs while preserving their distribution and relative relationships, thereby allowing all predictors to contribute fairly during learning.

After evaluating several scaling techniques, we opted for min–max normalization because of its straightforwardness, clarity, and compatibility with algorithms that are particularly sensitive to the scale of the data. The transformation for any feature (x) was executed using Eq. 3.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3}$$

where x_{min} and x_{max} represent the lowest and highest observed values of the feature in the complete dataset. This operation compresses every observation into the range [0, 1], guaranteeing that all predictors contribute equally and sit within the same numerical range during learning. In cases anticipated to have outliers, such as the compressive strength of heat-cured specimens, an outlier-detection phase preceded normalization in order to limit distortions of the rescaled data.

To validate the approach, the normalization boundaries (x_{min} and x_{max}) were calculated using only the training split when the model was being fitted; the derived values were then rigidly applied to both the validation and testing splits. This approach unequivocally prevented any leakage of unseen data into the training phase, thereby sustaining the credibility of the model's performance metrics. Consistent and methodical normalization thus shielded the dataset while keeping the original inter-feature relationships intact, free from biases attributable to differing scales. This careful preprocessing was decisive in equipping the learning algorithms to discover the patterns dictating the crack-opening behavior of ultra-high-performance concrete from experimental observations.

Data description

In this investigation, to facilitate rigorous and dependable machine learning modeling, we compiled a complete dataset of 600 samples including eleven features such as water-to-binder ratio (w/b), silica fume (SF), fly ash (FA), superplasticizer (SP), fiber volume content (FV), fiber length (FL), fiber diameter (FD), initial notch depth (a_0), span depth (SD), and specimen thickness (ST). The dataset was created entirely from our own experimental measurements using three-point bending tests, without pulling in any external or previously published data. In contrast, datasets that have been available in this field often face limitations, such as small sample sizes, narrower parameter ranges, and a reliance on simulated rather than experimental values. Our dataset breaks through these barriers by offering a significantly larger number of samples, a wider range of variables, and high-fidelity experimental conditions. To the best of readers knowledge, there isn't a comparable dataset in the literature that combines this level of scope and precision, making it both original and perfectly suited to meet the goals of our current study.

Outlier detection was carried out using a stepwise, statistically grounded method grounded in the Interquartile Range (IQR) formulation. For each variable, we calculated the first quartile (Q1) and the third quartile (Q3), then determined the IQR as IQR = Q3 − Q1. Entries falling outside the bounds [Q1 − 1.5 × IQR, Q3 + 1.5 × IQR] were classified as outliers and marked for extraction. This procedure yielded no anomalous points. As a result, all the 600 data points were carried forward into the machine learning modeling phase. The 600 samples create a substantial experimental database for FR-UHPC research. Most previous machine learning studies in this field have worked with fewer than 500 samples. The descriptive statistics of the processed dataset are presented in Table 4, providing clarity about the dataset's organization and the measures taken to safeguard its integrity.

	w/b	SF	FA	SP	FV	FL	FD	a_0	SD	ST	CMOD
Unit	–	(%)	(%)	(%)	(%)	(mm)	(mm)	(mm)	(mm)	(mm)	(mm)
count	600	600	600	600	600	600	600	600	600	600	600
mean	0.19	20.08	10.18	1.65	1.61	12.34	0.22	32.89	377.43	76.13	0.74
std	0.02	4.21	5.48	0.46	0.61	4.87	0.03	5.90	81.28	16.81	0.23
min	0.15	10.30	0.40	0.64	0.07	6.00	0.15	20.50	204.00	41.00	0.02
25%	0.17	17.00	5.60	1.32	1.16	6.00	0.20	28.58	316.50	63.00	0.58
50%	0.19	19.95	9.10	1.62	1.68	13.00	0.21	31.90	367.00	75.00	0.76
75%	0.21	23.10	14.40	1.97	2.05	16.00	0.24	36.70	434.25	89.00	0.91
max	0.25	29.70	24.20	2.78	2.95	20.00	0.30	49.60	589.00	116.00	1.30

Table 4. A statistical summary of numerical input features used for machine learning analysis. w/b, water-to-binder ratio; SF, silica fume; FA, fly ash; SP, superplasticizer; FV, fiber volume; FL, fiber length; FD, fiber diameter; a_0, initial notch depth; SD, span depth; ST, specimen thickness.

As presented in Table 5, the CA parameter is handled as a categorical feature instead of a continuous value. Such a treatment is appropriate when the numbers reflect separate experimental phases or specific settings rather than a continuum where the gaps represent meaningful variations. Here, CA indicates particular curing lengths (3, 7, 14, 28, 56, 90, 120, and 180 days) that coincide with designated points for testing concrete performance in the lab. To prepare the data, each CA obtains its own binary code using one-hot encoding. This means that for any specimen the digit in the position that matches the age is set to "1" and all other digits are "0." For example, a specimen cured for 28 days translates to 00010000 and one cured to 90 days translates to 00000100. This encoding allows the categorical information to merge seamlessly into machine learning algorithms while preventing the unintentional introduction of misleading ordinal meanings, such as the false notion that the curing age of 28 days is "twice" that of 14 days.

Looking at the retained data, we see that 16% of the specimens were evaluated at the 28-day mark, confirming its enduring role as the traditional benchmark for concrete strength. We see closely spaced groups at 56 days and 14 days at 15% and 12% of the total, respectively, and the 90-day, 120-day, and 180-day points each contribute 12% as well. Measurements taken at 3 days and 7 days represent 11% and 10% of the total, highlighting a sustained interest in early strength for applications requiring quick turnaround or provisional loading. The overall pattern reveals a balanced distribution across the ages we tracked, yet still centered on the well-established 28-day reference. This distribution reflects a deliberate experimental strategy designed to document strength gain across the entire timeline of hydration, from the quick initial set to the intermediate phases and into the long-term durability window. The resulting data set therefore supports more reliable models of strength development, which are increasingly critical for fine-tuning performance-based mix designs and for predicting the durability of concrete under real-service conditions.

The box plots shown in Fig. 4 display 11 primary features following outlier removal, calibrating any outlier previously documented to zero. Each box illustrates the median, the interquartile range, and whiskers terminating at minimum and maximum values confined within $1.5 \times$IQR. Covered features include mix proportions, material characteristics, and curing parameters, collectively charting the data set's spread and distribution following cleansing.

The w/b varies roughly between 0.15 and 0.25, centering at about 0.19. The symmetrical and compact distribution underscores a tightly controlled mix design across the entries, promising uniform hydration kinetics and mechanical performance in the UHPC batches. SF ranging from about 10% to 30% and peaking at a median of 20%, shows a slight right skew, with most results clustering at the 20% substitution mark. This narrow set of values attests to the typical practice of fixing SF dosage, reinforcing the microstructure and durability of the UHPC per established design guidelines. FA exhibits a range from 0.4% to 24%, with a median near 10%. The left-skewed pattern, highlighted by a pronounced lower tail, reveals that a few mixtures incorporate only small amounts of fine aggregate. Such differences may affect workability, packing density, and microstructural behavior. SP fluctuates from 0.64% to 2.78%, with a median of around 1.65%. The nearly symmetric shape and moderate dispersion suggest that adjustments are made cautiously to optimize workability while keeping the mixtures comparably uniform. FV is reported between 0.07% and 2.95%, averaging around 1.61%. The mild right skew and small interquartile range confirm that fiber levels cluster between 1.5% and 2%, aligning well with the typical reinforcement recommendations for UHPC. FL is distributed from about 6 mm to 20 mm, with a median of 12.34 mm. The left-skewed curve shows a preference for longer fibers, while the presence of shorter lengths in some mixtures allows for an exploration of varied mechanical behavior. FD spans 0.15 mm to 0.30 mm, with a median of roughly 0.22 mm. The symmetric and tight distribution reflects careful manufacturing to achieve uniform diameter, which in turn promotes predictable mechanical performance. Parameter a_0 lies between 20.50 mm and 49.60 mm, the median at 32.89 mm. The slight right skew suggests good control over dimensions needed for mechanical assessment. SD runs from 204 to 589 mm, with a median of 377 mm. The nearly centered spread and modest range show depth adjustments from the experimental framework still kept within preset limits. ST extends from 41 to 116 mm and has a median at 76 mm. The balanced spread and moderate variability indicate a standard set point for thickness, with allowances for particular testing protocols.

Correlation between parameters

Exploring how variables relate to one another is an essential part of puzzling out the structure of any dataset, especially when the end goal is to tune a predictive model. By calculating correlations, we can not only grasp how tightly an input is shackled to the target variable, either through a straight line or a more twisted path,

Categorical parameter	State (days)	Percentage of total data	One-hot encoding
CA	3	11	10,000,000
	7	10	01000000
	14	12	00100000
	28	16	00010000
	56	15	00001000
	90	12	00000100
	120	12	00000010
	180	12	00000001

Table 5. A statistical summary of CA parameter considered as a categorical variabe. CA, curing age.

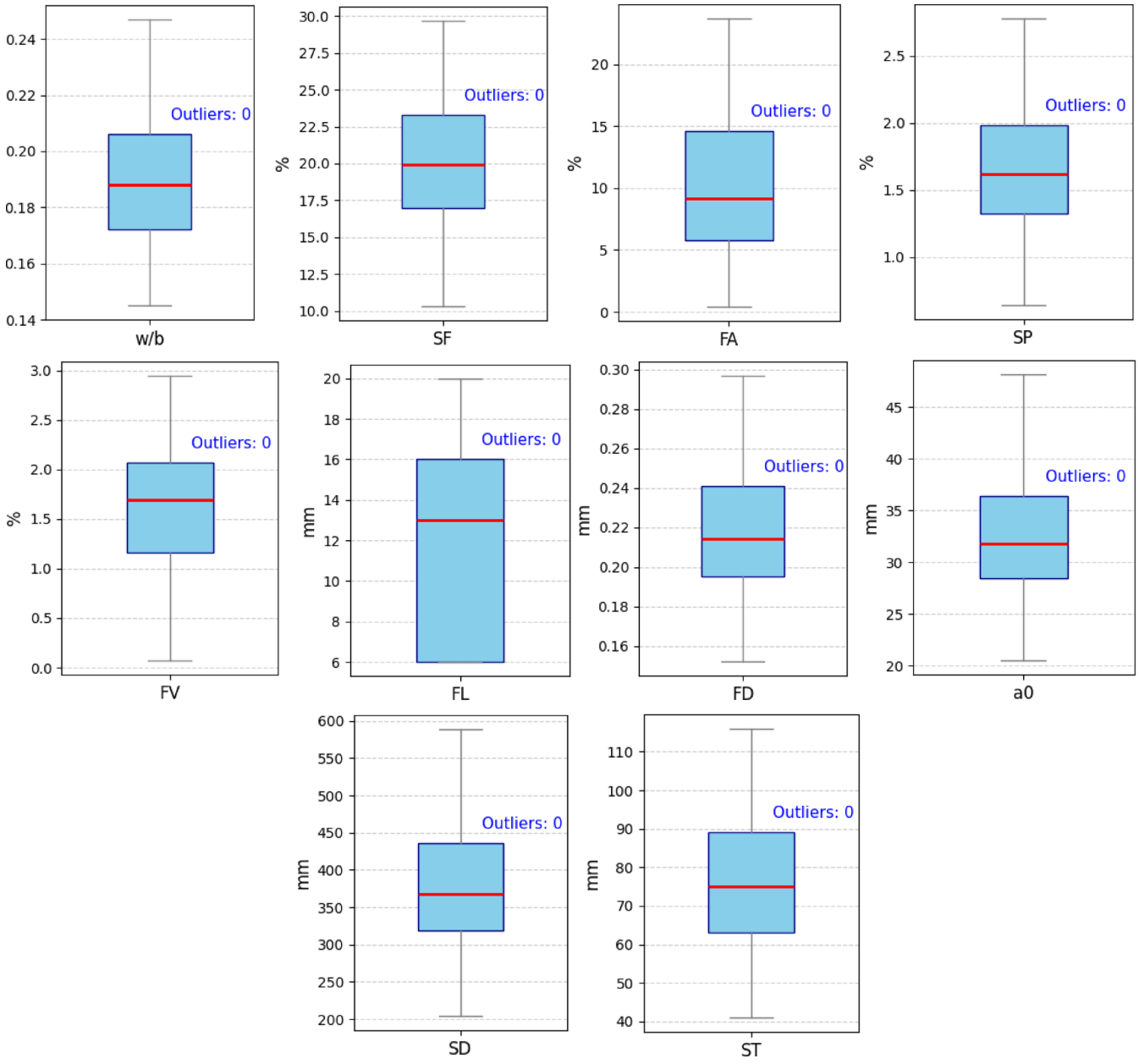

Fig. 4. Box plots of numerical features illustrating the distribution and variability of the cleaned dataset.

but we can also see how pairs of input variables influence one another. The classic Pearson coefficient gives a read on straight-line connections, but when dancers are less rigid, tools like distance correlation and mutual information step in to reveal hidden dependencies, be they linear or not. This phase of the work points us toward the predictors that carry the most weight, flags instances of multicollinearity that might bog down the model, and occasionally offers a window onto the physical processes driving the data. By weaving together several correlation gauges, we can be confident that no vital cue is left lurking in the shadows, especially in landscapes where relations curl and twist out of the linear realm.

Understanding how variables connect with each other forms a basic part of dataset analysis, especially when we're building predictive models. We started by measuring pairwise linear relationships between variables using the Pearson correlation coefficient (r), defined as[65]:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}} \tag{4}$$

where x_i and y_i are paired observations, and $\overline{x}$, $\overline{y}$ are their respective means. The r values fall between -1 and $+1$, where the sign shows direction and the magnitude shows how strong the linear relationship is.

Figure 5 presents the Pearson correlation matrix visualized as a heatmap. Overall, it shows that most of the predictor variables and the CMOD remain loosely linked, with the notable exception of the FV and CMOD pair, whose correlation coefficient of 0.85 indicates a tight, positive dependence. This finding is coherent with the underlying mechanics: increased fiber loading translates directly to improved crack-bridging effectiveness, thereby enlarging the CMOD. A weaker, but positive, link is recorded between SF and CMOD (0.23), as well as between SF and FV (0.05). Variables including w/b, SF, and FL remain nearly unrelated to CMOD, with absolute correlation coefficients below 0.15. The generally low inter-variable correlations throughout the matrix signal a low likelihood of multicollinearity in subsequent linear analyses.

Figure 6 illustrates distance correlation with respect to CMOD. This technique picks up both linear and curvilinear associations and ranks FV highest, with an coefficient of 0.826. The next-most influential variables are SF (0.238), FL (0.170), and a_0 (0.142). Unlike the Pearson analysis, the distance correlation ranks a_0 higher, hinting it may exert a non-linear weighting on CMOD. Variables such as SP, FD, and SD show weak but non-zero associations, suggesting marginal influence.

The mutual information analysis illustrated in Fig. 7 reveals that FV remains the strongest predictor of CMOD with a score of 0.633, decisively outperforming the other variables. The scores for SF (0.044) and FA (0.014) indicate that they add some, though limited, non-linear predictive content. In contrast, variables such as SP, FD, a_0, and ST feature values yield mutual information near zero, suggesting they play almost no role in elucidating the variability of CMOD.

When synthesized, the three analytic techniques continuously rank FV as the dominant influence, followed by SF and FL which hold moderate predictive weight. The varying character of the Pearson, distance correlation, and mutual information rankings further reveals a mix of linear and non-linear dynamics in the response variable, thereby validating the necessity for a diverse suite of metrics to evaluate feature importance comprehensively.

Feature selection

Selecting the right features is vital in machine learning because it boosts model accuracy, cuts training costs, and makes the results easier to explain by removing noise and overlap among variables. When we concentrate on the

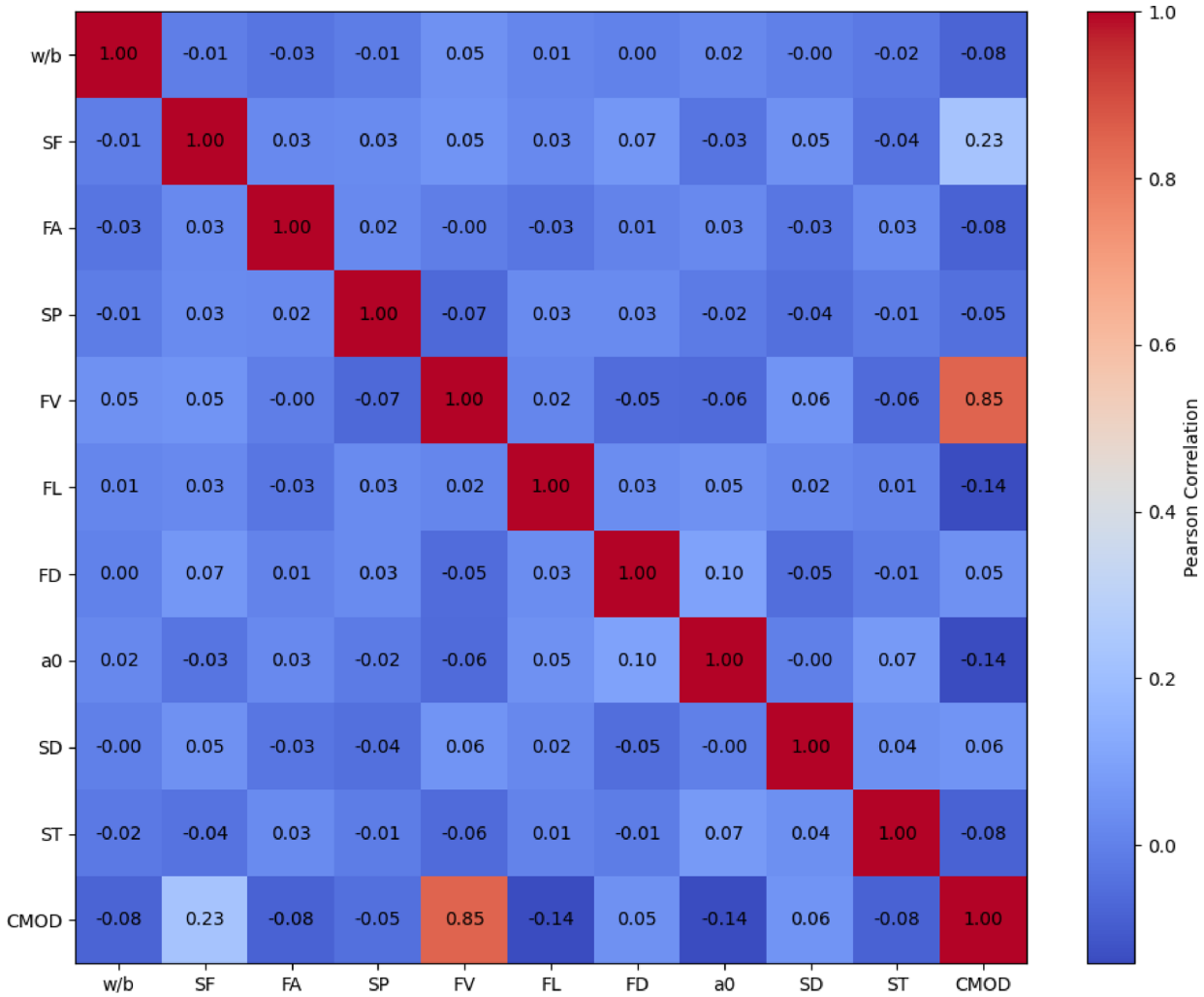

Fig. 5. Pearson correlation matrix of the numerical parameters.

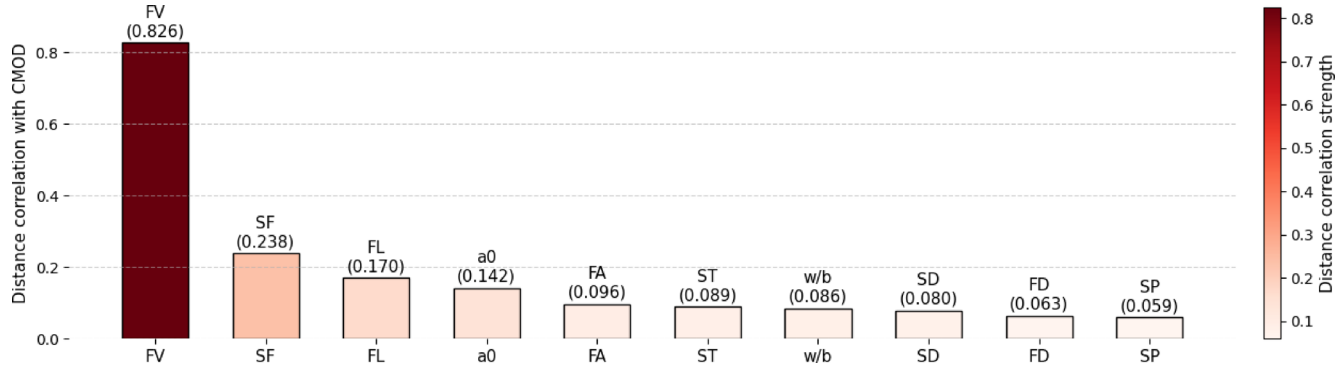

Fig. 6. Distance correlation scores with target variable (CMOD).

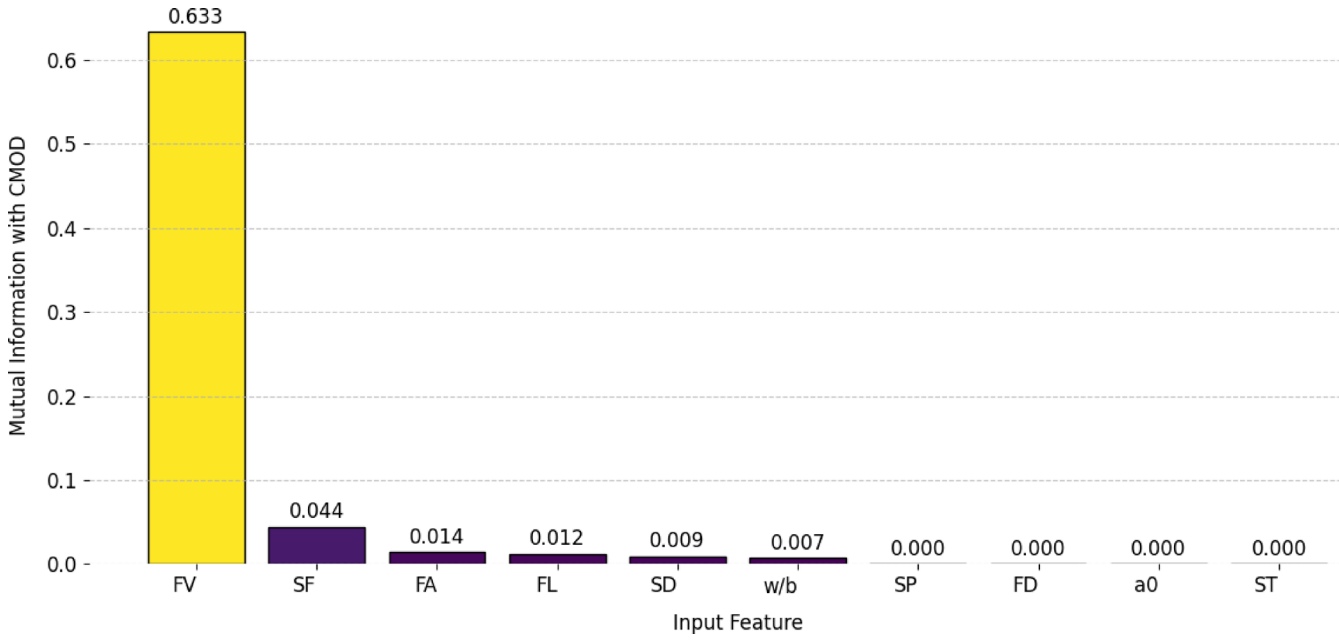

Fig. 7. Feature-wise mutual information scores with target variable (CMOD).

predictors that count, models tend to perform better on new, unseen data, resist overfitting, and can be trained in shorter times. In research settings, careful feature selection also reveals which experimental or physical variables most strongly drive the outcome, enriching the theoretical understanding of the problem.

A two-stage feature selection approach including FeatureCuts and Particle Swarm Optimization (PSO) was applied in this study. 100% of your text is likely AI-generated. First up, FeatureCuts acts as a smart initial filter, efficiently trimming down the dimensionality by getting rid of irrelevant or weakly contributing variables. This helps to reduce the computational load for the next stage. Then, PSO steps in with its powerful optimization technique, adept at navigating complex feature interactions and steering clear of local optima. This hybrid approach takes advantage of the best of both worlds: quick initial reduction thanks to FeatureCuts and solid subset optimization through PSO. On the flip side, methods like Recursive Feature Elimination (RFE) and LASSO, while popular, tend to come with higher computational costs in high-dimensional situations (RFE) or rely on strong linearity assumptions that might not fit the nonlinear relationships in our dataset (LASSO). So, we opted for the FeatureCuts + PSO combo as the best way to strike a balance between efficiency and accuracy in this study.

First, a FeatureCuts-inspired method using an RFR quantified the mean decrease in impurity attributed to each predictor, allowing us to sort them by their importance. Any feature that scored below a preset cutoff of 0.02 was automatically discarded. This initial pruning removed the variables that added little to no value. In the second stage, a PSO algorithm was applied to operate on a binary feature space as a wrapper method. Here, the optimization goal was to minimize the cross-validated mean squared error (MSE) of a random forest model, guaranteeing that the final feature subset achieved the highest possible predictive performance. The resulting hybrid method leverages the speed of the filter stage to quickly pare down the feature set, while the wrapper stage delivers the accuracy that results from a careful, model-guided search.

It Fig. 8 displays the importance scores calculated for all input variables, with green bars representing the features retained in the final model following PSO optimization and red bars indicating those that were removed. FV feature emerges as the preeminent predictor, achieving a score exceeding 0.75 and markedly outperforming all other variables. The significant role of FV in Fig. 8 really matches what we'd expect physically. This is because the amount of fiber used is the key factor that affects crack bridging and the CMOD response in FR-UHPC. The so-called abnormality comes from how we scale the importance scores: FV explains most of the variance, while other factors like FL, FD, and matrix properties take a backseat.

The subsequent ranks are held by SF, FL, and FD, all of which have notably reduced scores. Although FD exceeded the importance ranks of FA, the w/b, and a_0, it was ultimately discarded in the PSO process. This decision stemmed from its diminishing return in overall predictive strength when FV and FL were already accounted for, a situation likely exacerbated by multicollinearity that renders overlapping explanatory capabilities less distinct. Variables such as SP, SD, and CA registered near-zero scores and were also omitted from the final model. The noticeable difference in scores shows that FV has the biggest impact on the model predictions, while other factors play a supporting role that fine-tunes but doesn't overshadow the predictive behavior.

Results analysis and comparison

Holdout cross validation

Figure 9 presents a focused comparison of the estimated outputs from each algorithm with the independently recorded CMOD measurements, exposing the results on a plot that incorporates the a20-index. This index measures the percentage of predictions that fall within ± 20% of the actual observed values, giving us a clear indication of practical accuracy in engineering contexts. Unlike purely statistical measures, the a20-index directly shows whether the model meets the precision levels that are generally accepted for engineering decision-making. In civil engineering, for instance, predictive models are often evaluated based on their ability to stay within certain error tolerances. Typically, deviations of 15–25% are seen as acceptable for strength predictions and material property estimations. When the a20-index is displayed alongside the complementary aα family (especially a10 and a30), it represents a pragmatic middle ground. The a10 is widely deemed excessively severe, punishing deviations that, in the context of concrete performance, are unlikely to jeopardize workability. In contrast, the a30 is often criticized for being overly lenient, letting predictions seem reliable while masking substantial errors. Thus, the a20-index serves as a simple benchmark for assessing whether our predictions are not just statistically valid but also practically reliable.

For this analysis, the data were split prior to any modeling into a training subset containing 80 percent of the records and a held-out testing subset containing the other 20 percent. The partition occurred during the standard holdout phase; the training subset was solely employed for calibrating model parameters, while the testing subset was never exposed to the modeling process until the final evaluation. This careful separation guarantees that the accuracy scores (most notably the a20-index values illustrated in Fig. 9) serve as unbiased indicators of the models' ability to generalize to new cases. By eliminating the risk of information bleed from the training phase into the testing phase, the holdout design delivers a truthful appraisal of how the model is likely to perform on data it has never encountered.

Fig. 9, every sub-plot compares the predicted CMOD values against the measured CMOD values from the testing set. In this analysis, all features (FV, SF, FL, FD, FA, w/b, a_0, ST, CA, SP, SD) were considered. DTR returned the lowest performance, registering an a20-index of 0.71; its spread of predictions around the ideal diagonal reveals excessive variance, indicating it largely memorised the training samples rather than captured a generalisable mapping. In contrast, SVR provided a much tighter clustering of points, raising the a20-index to 0.90 and signalling good generalisation across the test set. The best performance came from the NuSVR, which

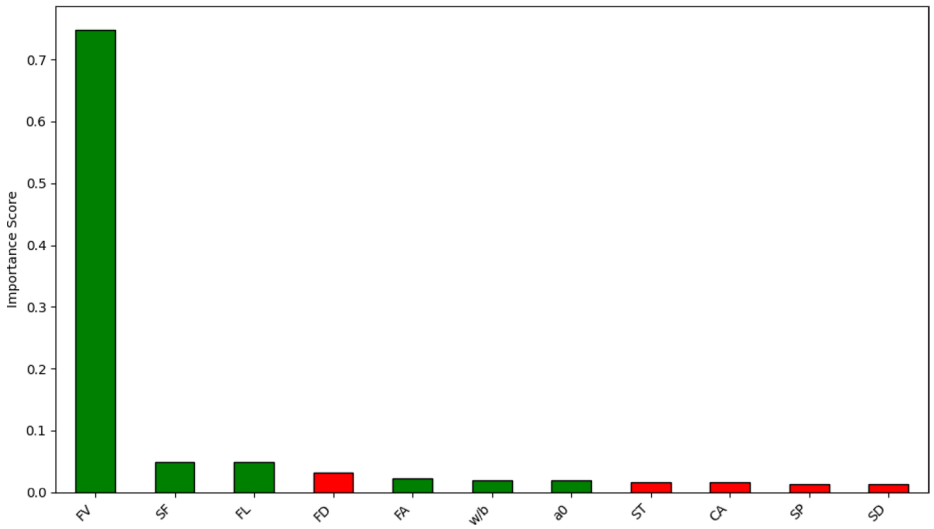

Fig. 8. Feature selection based on FeatureCuts-PSO technique.

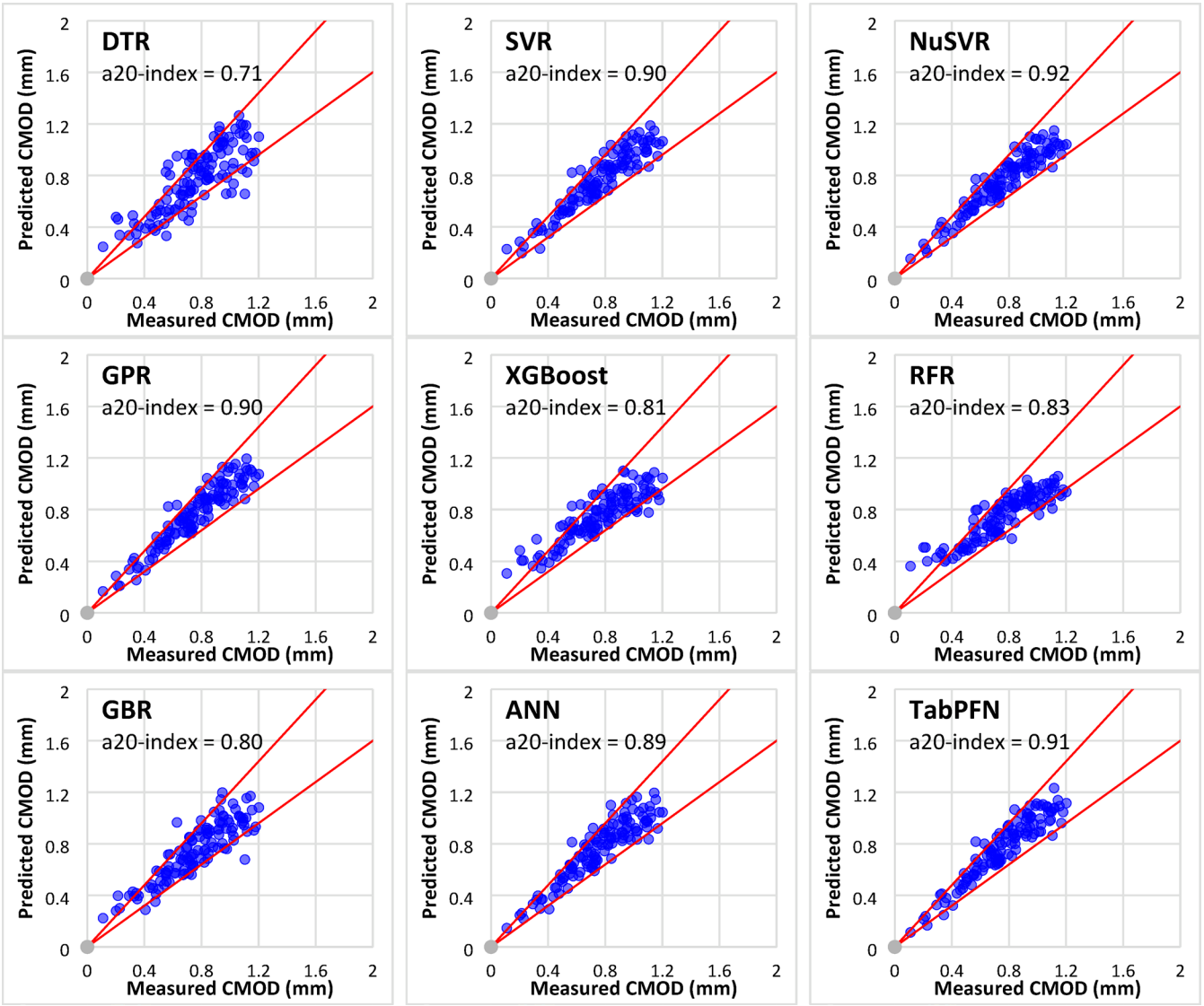

Fig. 9. Evaluating the machine learning algorithms' accuracy in predicting the CMOD using the a20-index (all features are considered).

reached an a20-index of 0.92 and showed a noticeably tighter error spread, suggesting an effective trade-off of bias and variance from carefully tuning the ν term. GPR equalled the SVR on the a20-index of 0.90, presenting predictions with a smooth and well-calibrated envelope. Its probabilistic nature adds interpretive power, though its computational demands may limit scalability with larger datasets.

Among the tree-based ensemble techniques, XGBoost achieved an a20-index of 0.81, while the RFR reached 0.83. Both methods exceeded the performance of the DTR but did not catch the leading kernel-based strategies, a difference likely linked to constraints on tree depth and possibly under-optimized hyperparameter settings. GBR exhibited an a20-index of 0.80, a result aligned with XGBoost and suggesting that either the learning rate was kept conservative or the boosting iterations remained shallow. ANN attained an a20-index of 0.89, positioning it just behind the kernel methods, with NuSVR, SVR, and GPR still ahead. The ANN successfully captured nonlinear patterns, though some residual variance persisted, indicating that deeper layers or better-tuned regularization could extract further gains.

TabPFN produced an a20-index of 0.91, ranking just behind NuSVR yet outperforming nearly all other models. This result underlines the effectiveness of automated model selection and representation learning, demonstrating that one can reach high predictive accuracy without the burdens of extensive manual feature engineering.

The final a20-index ranking, arranged from top to bottom, is: NuSVR (0.92), TabPFN (0.91), SVR (0.90), GPR (0.90), ANN (0.89), RFR (0.83), XGBoost (0.81), GBR (0.80), and DTR (0.71). This pattern shows that kernel-based techniques (especially NuSVR, SVR, and GPR) excelled in CMOD prediction, likely because of their capacity to model smooth, nonlinear structure within the data. Although deep learning models remained competitive, their slightly lower performance may relate to dataset size and feature dimensionality. Meanwhile,

TabPFN demonstrated strong, automated performance, yet classic tree ensembles consistently trailed the kernel methods.

Further gains are conceivable through a stacking ensemble that synergizes NuSVR, TabPFN, and GPR. If implemented and validated using the original 80/20 holdout strategy, this hybrid model could potentially surpass the 0.92 a20-index limit currently set by NuSVR, yielding even greater predictive power.

Now, the a20-index analysis continues to adhere to the established protocol of allocating 80% of the data for training and 20% for testing, yet adopts a pivotal modification: only the six features deemed most influential (FV, SF, FL, FA, w/b, and a_0) guided both the model training and the evaluation routines. By intentionally reducing dimensionality, we aim to observe how each algorithm responds, given that sensitivity to uninformative or only marginally relevant predictors can vary widely among them. Figure 10 presents the predicted CMOD values plotted against the measured CMOD for all algorithms under these restricted feature conditions, facilitating a direct comparison with the corresponding results illustrated in Fig. 9, which utilized the entire feature set.

The overall impact of reducing the feature set yields inconclusive results. The DTR posted a small accuracy dip, with the a20-index sliding from 0.71 in Fig. 9 to 0.68 in Fig. 10. This pattern suggests that some of the variables dropped from the input still contained information that improved the effectiveness of the decision splits. Ensemble tree algorithms recorded similar trends: the RFR dipped slightly from 0.83 to 0.85 (possibly a small improvement from alleviated overfitting) while the GBR fell more sharply from 0.80 to 0.78, pointing to its reliance on a broader feature set.

Kernel-based methods exhibited more differentiated behavior. The NuSVR held its strong performance constant at an a20-index of 0.92, indicating that its kernel transformation still recovered the relevant signal from the compressed feature set. The SVR slipped from 0.90 to 0.87, and the GPR dropped from 0.90 to 0.85, showing

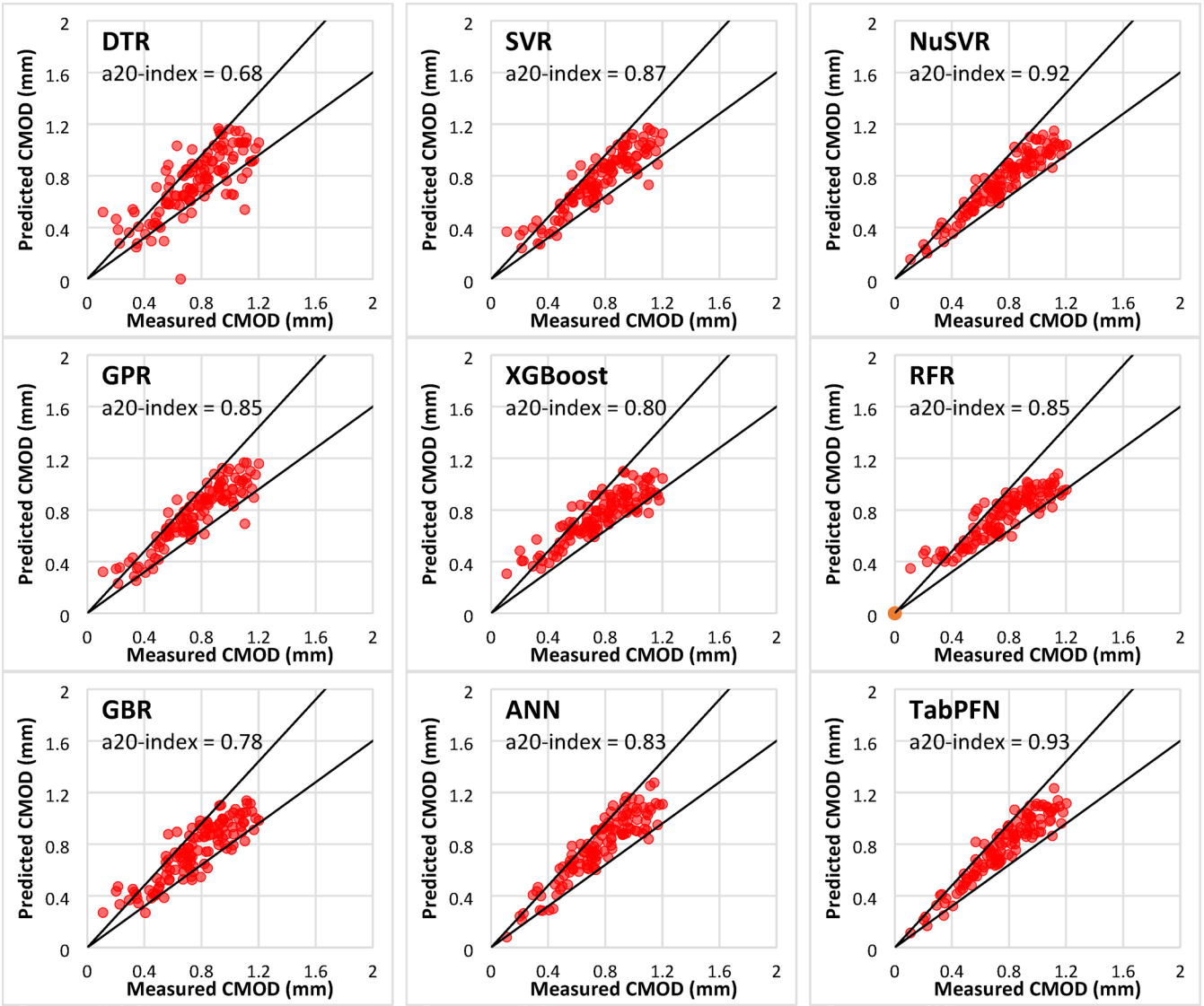

Fig. 10. Evaluating the machine learning algorithms' accuracy in predicting the CMOD using the a20-index (6 features including FV, SF, FL, FA, w/b, and a_0 are considered).

these models derived some utility from the discarded dimensions, even if they were not strictly necessary for prediction accuracy. Boosting methods exhibit mild yet statistically meaningful shifts: XGBoost decreased from 0.81 to 0.80 and the ANN fell from 0.89 to 0.83. The ANN's steeper decline implies its capability to model subtle nonlinearities was more reliant on the expanded feature set offered prior. The striking adjustment comes from the TabPFN. Whereas its prior performance in Fig. 9 reached 0.91, it elevates to 0.93 in Fig. 10, claiming the summit among all contenders in both composites. This regressive gain suggests TabPFN capitalized on the dimensionally pruned input; its architecture and learning-ID are attuned to diminish overfitting when pruned to the salient descriptors.

Across the figures, the most robust models to the contraction are NuSVR and TabPFN, the latter leaping past all preceding markers. This observation implies that in this particular dataset, streamlining to the dominant predictors can enhance performance for select sophisticated algorithms, while the net effect on others can be trivial. Such dynamics advocate for the prospective merit of amalgamating TabPFN and NuSVR in a layered meta-learning schema, capitalizing on their respective advantages when feature selection is fine-tuned.

Table 6 provides a comparison of the predictive performance among the algorithms we evaluated. The TabPFN model stood out with the best point estimates, while GPR, NuSVR, and SVR followed closely behind. The bootstrap 95% CIs are quite narrow, which suggests that the estimates are stable. However, the significant overlap between TabPFN and the next-best models indicates that the practical improvement is only modest.

K-fold cross validation

K-fold cross-validation is a standard practice for checking how well a machine learning model will perform on new data. The complete dataset is divided into K roughly equal-sized groups called folds. During each round, one fold is kept back for testing, while the model is trained on the remaining K-1 folds. This process is repeated K times, ensuring that each fold is used as the test set one time. The evaluation metrics from these K training–testing cycles are then averaged, providing a single score that is less influenced by the random quirks of any individual split. This technique effectively demonstrates the model's ability to generalize across the dataset, since every data point is both trained on and tested against the model throughout the K cycles. K-fold cross-validation helps guard against overfitting by presenting the model with several different training subsets. Rather than memorizing the peculiarities of a single split, the model must identify patterns that are consistent across all the folds. Consequently, the averaged performance reflects a sturdier estimate of the model's predictive power than what a single train-test split could offer.

For our study, we chose a fivefold cross-validation scheme (K = 5) to rigorously assess the machine learning models. The full dataset was partitioned into five equal segments; in each fold, one segment served as the validation set while the other four were merged to create the training set.

Results from evaluating a machine learning model can vary significantly depending on which performance metric is chosen for emphasis. Different metrics reveal distinct dimensions of performance—some focus on predictive error, others on explained variance, yet others on error robustness. To counter the risk of bias from a single viewpoint, a multi-criteria scoring model is preferable. Here, R^2, RMSE, and variance accounted for (VAF) were chosen. Each model received a rank for each metric; the one with the best R^2 earned a position of 9 (first among nine competitors), the second-best 8, and so on, down to the lowest, which received 1. Each model thus competed against the others within the same metric framework. No weighting was applied, so the overall rank results from the additive performance of these metric ranks.

The fivefold cross-validation performance, evaluated across all features, is summarised in Table 7. Table 8 shows the fivefold cross-validation results considering the most influential features. The final columns of these tables show the total rank score, which is the simple sum of the ranks earned on each of the chosen metrics and indicates the overall standing of each model. This straightforward evaluation approach prevents any favoritism toward models by relying on a single performance metric.

By closely examining Tables 7 and 8, we see how switching from all available features in Table 6 to the six most influential features (FV, SF, FL, FA, w/b, and a_0) in Table 8 modifies the performance profile of each algorithm during fivefold cross-validation. Overall, the feature-reduced setup yields significant improvements alongside a few minor losses, varying by model. The cumulative ranking score, calculated from the ordered contributions to R^2, RMSE, and VAF, provides a straightforward metric for evaluating the overall impact of this feature pruning.

Model	R^2 (95% CI)	RMSE (95% CI)	VAF (95% CI)
DTR	0.62 (0.50–0.67)	0.14 (0.13–0.15)	0.62 (0.50–0.68)
SVR	0.87 (0.86–0.89)	0.08 (0.08–0.09)	0.87 (0.87–0.89)
NuSVR	0.88 (0.85–0.89)	0.08 (0.08–0.09)	0.88 (0.86–0.89)
GPR	0.89 (0.87–0.89)	0.08 (0.07–0.09)	0.88 (0.88–0.89)
XGBoost	0.71 (0.67–0.75)	0.12 (0.10–0.15)	0.71 (0.67–0.75)
RFR	0.78 (0.73–0.82)	0.11 (0.09–0.13)	0.78 (0.73–0.82)
GBR	0.80 (0.77–0.82)	0.10 (0.09–0.12)	0.80 (0.77–0.82)
ANN	0.85 (0.82–0.87)	0.09 (0.08–0.10)	0.86 (0.84–0.88)
TabPFN	0.90 (0.89–0.91)	0.07 (0.07–0.08)	0.90 (0.89–0.91)

Table 6. Predictive performance of the evaluated machine learning models for CMOD. Reported values are R^2, RMSE, and VAF, each accompanied by 95% CIs obtained from 1000 bootstrap resamples.

Model	Fold no	R^2	Score	RMSE (mm)	Score	VAF (%)	Score	Ranking score
DTR	1	0.622	1	0.149	1	0.623	1	15
	2	0.643	1	0.151	1	0.644	1	
	3	0.656	1	0.134	1	0.657	1	
	4	0.674	1	0.127	1	0.677	1	
	5	0.499	1	0.144	1	0.502	1	
SVR	1	0.871	6	0.087	6	0.871	6	93
	2	0.867	6	0.092	5	0.867	6	
	3	0.874	6	0.081	6	0.875	6	
	4	0.885	7	0.075	6	0.885	7	
	5	0.861	6	0.076	7	0.865	7	
NuSVR	1	0.887	8	0.081	8	0.887	8	102
	2	0.880	8	0.087	6	0.880	7	
	3	0.875	8	0.080	7	0.876	7	
	4	0.883	6	0.075	6	0.884	6	
	5	0.854	5	0.077	6	0.859	6	
GPR	1	0.884	7	0.083	7	0.884	7	111
	2	0.882	7	0.087	6	0.882	8	
	3	0.882	7	0.078	8	0.883	8	
	4	0.891	8	0.073	7	0.891	8	
	5	0.874	7	0.072	8	0.876	8	
XGBoost	1	0.690	2	0.135	2	0.693	2	30
	2	0.666	2	0.146	2	0.672	2	
	3	0.724	2	0.120	2	0.727	2	
	4	0.721	2	0.117	2	0.721	2	
	5	0.746	2	0.103	2	0.748	2	
RFR	1	0.773	4	0.116	4	0.774	4	51
	2	0.733	3	0.130	3	0.734	3	
	3	0.777	3	0.108	3	0.778	3	
	4	0.822	4	0.094	4	0.822	4	
	5	0.813	3	0.088	3	0.814	3	
GBR	1	0.765	3	0.118	3	0.766	3	54
	2	0.785	4	0.117	4	0.785	4	
	3	0.807	4	0.100	4	0.810	4	
	4	0.797	3	0.100	3	0.797	3	
	5	0.823	4	0.086	4	0.824	4	
ANN	1	0.832	5	0.099	5	0.856	5	74
	2	0.866	5	0.092	5	0.870	5	
	3	0.866	5	0.084	5	0.875	5	
	4	0.823	5	0.093	5	0.836	5	
	5	0.851	4	0.079	5	0.851	5	
TabPFN	1	0.890	9	0.080	9	0.890	9	131
	2	0.887	9	0.084	7	0.888	9	
	3	0.894	9	0.074	9	0.895	9	
	4	0.912	9	0.066	8	0.913	9	
	5	0.893	8	0.067	9	0.893	9	

Table 7. Comparison of models performance through fivefold cross validation considering all features.

The most pronounced gain arises with TabPFN. Its ranking score ascends from 131 to 133, preserving its lead among the models. Beyond this numeric advancement, TabPFN achieves R^2 values in Table 8 that peak at 0.942, eclipsing the former maximum of 0.912 recorded in Table 7, while the RMSE in the top folds consistently sinks below 0.072. These results suggest that TabPFN reaps the rewards of feature selection and then exploits the slimmer input set to boost generalization and stability across the cross-validation folds.

NuSVR stays at the front, with its ranking nudging from 102 to 109. The small slide in average R^2 is offset by strong fold-to-fold steadiness, allowing it to still edge past most rivals. Its performance hints that the kernel-based method remains stable when weaker features drop, as long as the key variables still reveal the problem's main nonlinear patterns. GPR shows a nearly identical story, shifting from 111 to 108 with little real difference. The close score tells us that the GPR can comfortably adjust to the pared-down feature set without noticeable

Model	Fold	R^2	Score	RMSE (mm)	Score	VAF (%)	Score	Ranking score
DTR	1	0.601	1	0.153	1	0.603	1	15
	2	0.623	1	0.155	1	0.627	1	
	3	0.581	1	0.148	1	0.590	1	
	4	0.732	1	0.115	1	0.741	1	
	5	0.558	1	0.136	1	0.558	1	
SVR	1	0.822	7	0.102	7	0.822	7	95
	2	0.791	6	0.115	6	0.791	6	
	3	0.815	6	0.098	6	0.817	6	
	4	0.847	6	0.087	6	0.848	7	
	5	0.814	6	0.088	6	0.816	7	
NuSVR	1	0.828	8	0.100	8	0.828	8	109
	2	0.795	7	0.113	7	0.796	7	
	3	0.819	7	0.097	7	0.821	7	
	4	0.857	8	0.083	8	0.857	8	
	5	0.814	6	0.087	7	0.815	6	
GPR	1	0.821	6	0.103	6	0.821	6	108
	2	0.799	8	0.113	7	0.799	8	
	3	0.824	8	0.096	8	0.825	8	
	4	0.855	7	0.085	7	0.856	6	
	5	0.825	7	0.085	8	0.826	8	
XGBoost	1	0.705	2	0.132	2	0.706	2	30
	2	0.659	2	0.147	2	0.661	2	
	3	0.726	2	0.120	2	0.726	2	
	4	0.731	2	0.115	2	0.731	2	
	5	0.682	2	0.115	2	0.683	2	
RFR	1	0.773	5	0.116	5	0.774	5	66
	2	0.750	4	0.126	4	0.751	4	
	3	0.782	4	0.107	4	0.784	4	
	4	0.830	5	0.092	4	0.830	5	
	5	0.796	5	0.092	5	0.797	5	
GBR	1	0.764	4	0.118	4	0.765	4	51
	2	0.747	3	0.127	3	0.749	3	
	3	0.745	3	0.116	3	0.751	3	
	4	0.791	3	0.101	3	0.792	3	
	5	0.754	4	0.101	4	0.755	3	
ANN	1	0.723	3	0.128	3	0.733	3	63
	2	0.764	5	0.122	5	0.778	5	
	3	0.812	5	0.099	5	0.819	5	
	4	0.822	4	0.094	5	0.823	4	
	5	0.731	3	0.106	3	0.786	4	
TabPFN	1	0.918	9	0.072	9	0.918	9	133
	2	0.910	9	0.076	8	0.910	9	
	3	0.923	9	0.069	9	0.919	9	
	4	0.942	9	0.058	9	0.959	9	
	5	0.925	8	0.067	9	0.923	9	

Table 8. Comparison of models performance through fivefold cross validation considering 6 features (FV, SF, FL, FA, w/b, and a_0).

harm to performance, though its absolute accuracy is still a notch behind the top scorers. ANN loses ground, its ranking score falling from 74 to 63. The cut in feature variety seems to hamper the network's grasp of intricate interactions, and the widening R^2 spread across folds indicates it is still a data-hungry architecture that thrives on abundance.

Among the tree-based techniques, RFR slides from 51 to 66, and GBR from 54 to 51, indicating moderate R^2 declines and negligible RMSE upticks. The pattern likely stems from decision-tree algorithms leveraging a wider array of predictors, including many weak ones, to refine split rules. In contrast, DTR stays anchored at 15 points across both metrics, reiterating its struggle to grasp the problem's inherent complexity, regardless of feature

abundance. XGBoost's score holds steady at 30, with R^2 and RMSE budging only slightly. This stability hints that the method's boosting process, which continually adapts the influence of misclassified instances, offsets the departure of less informative predictors more effectively than the averaging strategy inherent in RFR.

Shifting to the overall order of models, three standout items from the feature selection exercise are:

1. TabPFN now sets the highest R^2 at 0.942, pushing the performance envelope.
2. NuSVR and GPR both maintain their ranking and effectiveness, despite needing fewer predictors.
3. XGBoost's scores are stable, showing it can still thrive on a leaner feature diet.

This side-by-side evaluation clearly indicates that feature selection constitutes a powerful catalyst for improved accuracy in specific advanced architectures, particularly in TabPFN, while exerting little influence or even slight detriments in alternative models. These findings lend considerable weight to the strategy of integrating TabPFN with NuSVR and, potentially, GPR within a meta-stacking arrangement, harnessing the diverse robustness of each method and their aptitude for extracting value from a finely tuned subset of predictor variables.

The comparative analysis illustrated in Fig. 11 quantifies and characterizes how feature selection reshapes the predictive accuracy of several machine learning models over five repeated cross-validation partitions. Each algorithm exhibited a unique sensitivity to the reduction of input dimensions, underscoring the co-evolution of model architecture and the structure of the dataset.

Among the tested methods, the TabPFN architecture attained the strongest validation accuracy before any features were pruned. Once feature selection was performed, it manifested uniform but slight enhancements across R^2, RMSE, and VAF. Importantly, these gains were reproducible across all folds, suggesting that even transformer models with extensive representational power profit from excising irrelevant and weak predictors. The principal mechanism behind the increased robustness appears to be the removal of collinear features, which alleviates redundancy and streamlines the learning of the mapping from a compact feature space to the target outcome.

After the feature selection step, the ANN model exhibited inconsistent results, most notably a decline in most cross-validation folds. The mean R^2 dropped (e.g., Fold 1: 0.832→0.723), RMSE rose (Fold 1: 0.099→0.128), and VAF lowered (Fold 1: 0.856→0.733). These shifts imply that beneficial variables for capturing the network's nonlinear interactions were eliminated, weakening the model's capacity to resolve intricate dependencies. The outcome suggests that the ANN, with its flexible architecture, thrives on a more extensive feature collection to deliver the best generalization on the given dataset.

In contrast, the tree-based models, namely GBR and XGBoost, yielded only modest and variable shifts following the feature selection. The XGBoost implementation registered a tiny R^2 bump in some folds (e.g., Fold 1: 0.690→0.705) while dropping slightly in others. GBR, however, typically recorded a lower R^2 (Fold 3: 0.807→0.745) along with a higher RMSE, confirming that the excluded features offered only marginal enhancements in predictive power. RFR exhibited virtually constant performance across all measures, underscoring its resilience to redundant features due to the combined effects of bagging and the deliberate randomness in feature selection.

Kernel methods (GPR, SVR, and NuSVR) produced variable outcomes. GPR, for instance, recorded an across-the-board decline in R^2 following variable pruning (Fold 1 0.884 falling to 0.821), accompanied by a heightened RMSE. This suggests that the pruned feature set limited GPR's capacity to fit the underlying function. SVR and NuSVR mirrored this pattern (Fold 1 R^2 0.871 dropping to 0.822), with RMSE worsening throughout every split. These results contradict the assumption that kernel methods profit from lowered dimensionality, implying that the eliminated features carried meaningful information rather than mere noise.

DTR delivered uniformly diminishing R^2 (Fold 3 0.656 to 0.581) alongside stagnant RMSE, highlighting its pronounced vulnerability to reduced cue availability. Accuracy metrics confirmed it as the weakest performer, reaffirming the limitations of single-tree strategies in regression problems marked by complexity and noise.

The standout finding is that TabPFN preserved top R^2, minimal RMSE, and peak VAF postpartum, and in all folds improved (Fold 1 R^2 increased from 0.890 to 0.918, RMSE from 0.088 to 0.072). This consistency and modest performance gain in the face of feature compression signal TabPFN's architecture as adept at leveraging compact, high-quality inputs.

The results show that rather than universally enhancing model precision, feature selection led to small declines in nearly all models aside from TabPFN, which either preserved or boosted accuracy, and XGBoost, which gained minor improvements in a few folds. The persistent predictive power of the trimmed features implies that feature selection ought to be conducted judiciously and always validated against the model in use, lest valuable data be discarded.

Statistical significance testing

Rank-based nonparametric tests are recommended instead of multiple pairwise t-tests when comparing several algorithms over a single set of problems or cross-validation folds. This controls for inflation in the Type I error probability and does not rely on assumptions of normality or equal variance, assumptions frequently violated during algorithm benchmarking. The Friedman test is the major omnibus test for this situation, either all algorithms sharing the same performance distribution or at least one differing significantly. Once there is a significant result from the Friedman test, post-hoc pairwise comparisons are made to ascertain which models differ. The Nemenyi test, and its derivatives such as the Bonferroni–Dunn method, are most frequently used at this stage. A critical difference (CD) diagram nicely summarizes these comparisons by plotting average ranks for the algorithms and connecting models that are not significantly different from one another.

In this study, we used Friedman tests on RMSE values across five folds and obtained $\chi_F^2 = 34.818$, ($p = 2.9 \times 10^{-5}$), signifying significant differences among the nine methods. Then, post-hoc Nemenyi testing was

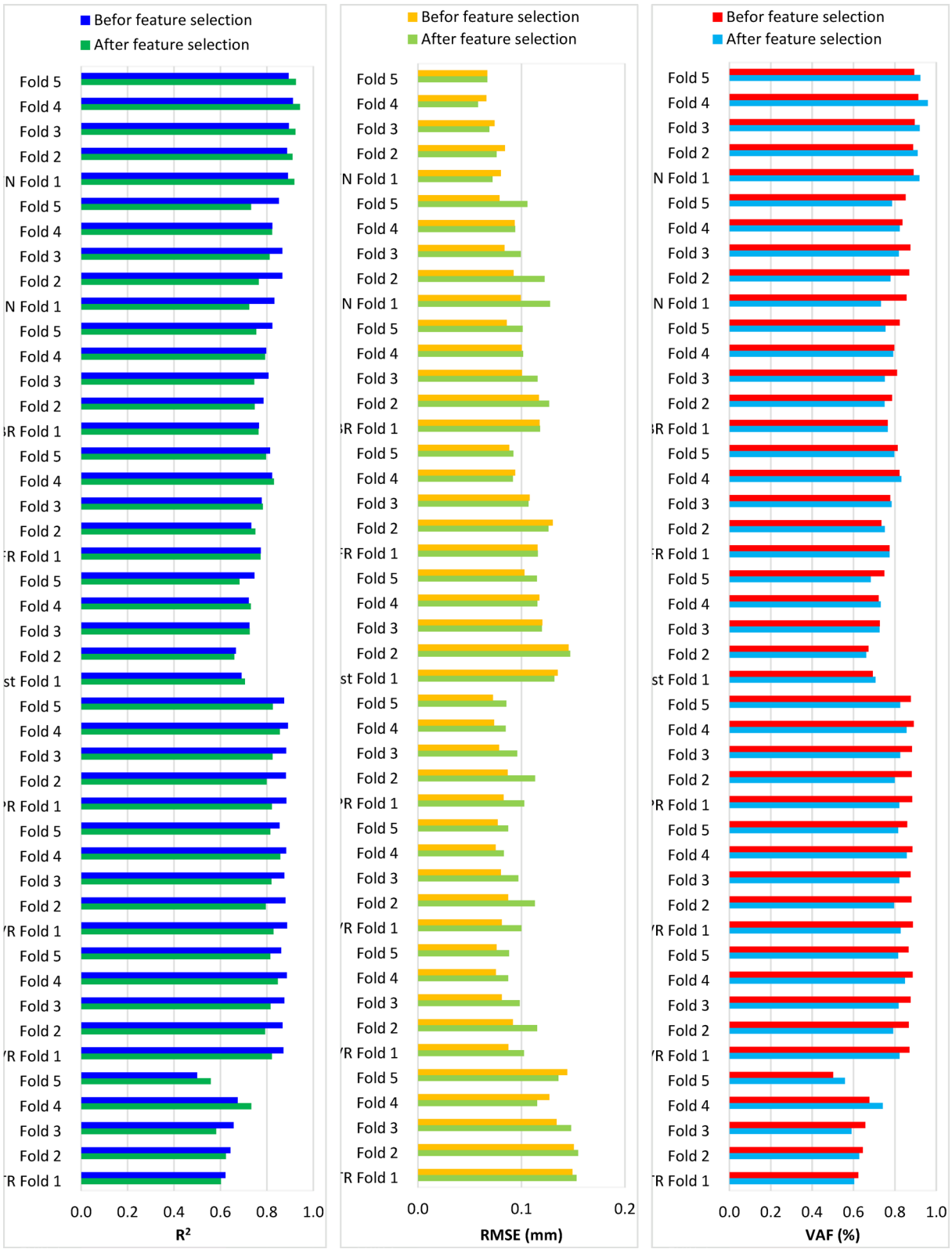

Fig. 11. Impact of feature selection on model performance.Comparative analysis of statistical indices (R^2, RMSE, VAF) before and after dimensionality reduction. After dimensionality reduction, 6 features (FV, SF, FL, FA, w/b, and a_0) were considered.

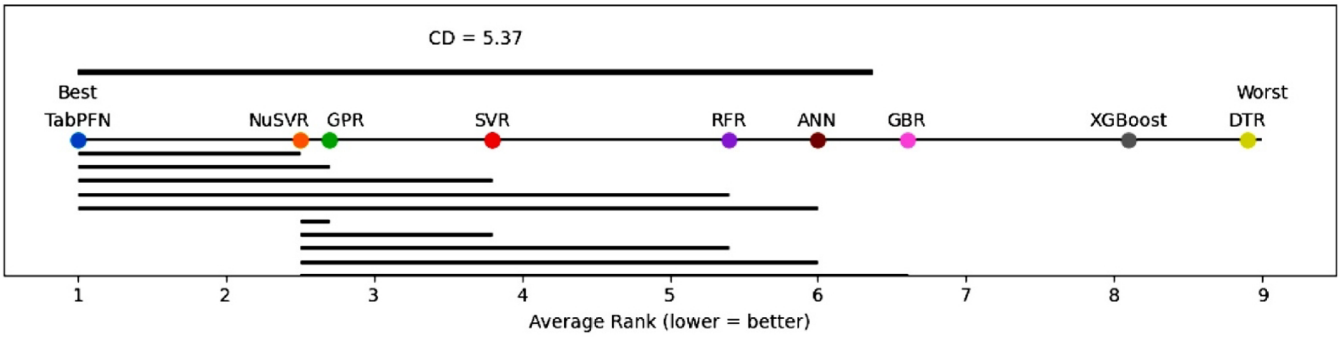

Fig. 12. CD diagram of machine learning models.

	TabPFN	ANN	NuSVR	SVR	GBR	XGBoost	GPR	DTR
TabPFN	—	0.032	0.004	0.007	0.021	0.015	0.041	0.003
ANN	0.032	—	0.118	0.095	0.21	0.134	0.087	0.056
NuSVR	0.004	0.118	—	0.084	0.143	0.109	0.176	0.091
SVR	0.007	0.095	0.084	—	0.126	0.14	0.162	0.073
GBR	0.021	0.21	0.143	0.126	—	0.088	0.074	0.064
XGBoost	0.015	0.134	0.109	0.14	0.088	—	0.097	0.082
GPR	0.041	0.087	0.176	0.162	0.074	0.097	—	0.12
DTR	0.003	0.056	0.091	0.073	0.064	0.082	0.12	—

Fig. 13. Pairwise statistical significance (p values) of performance differences between models, based on RMSE distributions across cross-validation folds.

carried out and the results are summarized in the CD diagram in Fig. 12. The value for CD computed was 5.37, indicating that models with average rank differences less than this threshold cannot be considered at $\alpha = 0.05$. TabPFN, with the lowest average rank, is identified as the best performer in Fig. 12. However, there is no statistically significant difference in the performance of TabPFN versus NuSVR and GPR, as all three make up one statistical group. Models like DTR, XGBoost, and GBR, all had rankings that were substantially lower, and the average ranking differences were more than the CD. These results show that although TabPFN leads in average performance, many high-grade models, especially NuSVR and GPR, output statistically comparable results.

Comparative statistical evaluation of models

To assess the statistical significance of performance differences between models, the ranking scores with hypothesis testing was complemented. Model errors (RMSE values across cross-validation folds) were compared pairwise using both paired t-tests and Wilcoxon signed-rank tests, depending on distributional assumptions. Figure 13 shows the pairwise statistical significance of the differences in model performance based on RMSE distributions. TabPFN demonstrated statistically significant improvements over all other models ($p < 0.05$), highlighting its strength as the top-performing method. On the other hand, the differences among ANN, SVR, NuSVR, GBR, XGBoost, GPR, and DTR were mostly not statistically significant ($p > 0.05$ in most pairwise comparisons), suggesting that their predictive accuracy overlaps considerably. These findings imply that while

TabPFN stands out as the clear winner, we should be cautious when interpreting the rankings of the other models.

Sensitivity analysis of the machine learning models

As in Fig. 5, a Pearson correlation of 0.85 between FV and CMOD indicates a strong positive correlation with a linear relationship. In the context of fracture testing EN 14,651 or ASTM C1609, it can be interpreted that specimens with higher FV tend to demonstrate larger CMOD values during the post-cracking load phase. This points out statistically that FV is one of the most important factors in post-cracking deformation behavior. From a mechanical viewpoint, fiber bridging effect explains this relationship. After the matrix has cracked, fibers within a FRC begin to apply bridging tensile forces. An increase in FV improves the crack bridging due to increased fibers straddling the crack's surface. This results in better post-crack load transfer and is the mechanism for increased crack-bridging effectiveness, which improves the material's ability to carry load and permits the crack to extend further before the load reduction to zero, increasing CMOD. This is similar to the pull-out mechanics where post-crack energy absorption and residual stress improves with increased FV.

Concrete without fibers will crack, but the concrete will fracture suddenly, resulting in small CMOD values. With the provided higher fiber content, separations will be delayed, and fibers will enable the cracks to open wider due to pull-out. Thus, the FV–CMOD correlation is a consequence of statistical coincidence. In summary, CMOD and FV are strongly and positively linked. This link, which is the mechanics of FRC and fracture theory, was tested in this study using experimental research.

To thoroughly quantify how the newly developed machine learning models react to systematic variations in FV, we designed a dedicated experimental campaign that involved 16 distinct UHPC mixtures, summarized in Table 9. Within the campaign, every mixture component was kept consistent apart from FV, allowing the isolatory impact of this parameter on CMOD to be examined cleanly. The FV value was stepped from 0 to 3% in increments of 0.2%, a span and resolution that replicates the conditions imposed during the model-training phase. To maintain a uniform w/b throughout, the water and binder contents were both modulated upwards by 0.1% for every 0.2% rise in FV, thereby keeping the ratio fixed.

On the computational side, we focused on the six features that earlier analyses had confirmed as the main drivers: FV, SF, FL, FA, w/b, and a_0. By limiting the input to this minimized yet fully representative set, we fed each of the pre-trained models the variations of FV obtained in the experiments. This approach allowed us to juxtapose model predictions directly with the experimentally acquired CMOD results across the precisely controlled FV gradient.

Fig. 14 enables a straightforward juxtaposition of empirical CMOD data against the output of each machine learning model for the complete FV variation. The TabPFN configuration tracks the experimental curves without discernible divergence, maintaining a nearly indistinguishable slope and intercept for the entire FV span. The ANN counterpart displays similarly tight fits—though it tends to slightly underestimate CMOD beyond FV ~ 2.4%, hinting at a subtle undercapture of fiber-bridging influences at the higher dosage tail.

The SVR, NuSVR, and GPR variants yield nearly identical linear regressions, attaining R^2 values of 0.99 or above. Nevertheless, they consistently lie below the experimental data for all FV points. The identical vertical shifts across the domain signal a structural phenomenon (possibly the result of kernel-selection penalization or insufficient representation of the extreme-FV samples during training) rather than merely random scatter. Ensemble frameworks (GBR, RFR, and XGBoost) typically track the experimental trend with certain localized departures. GBR provides solid accuracy and limited fluctuation, while RFR shows a gentle saturation beyond FV ≈ 2.0%, resulting in underestimation at the upper tail. XGBoost, despite accurately reflecting the global trajectory, presents more visible oscillations in the mid-range (FV = 0.8%–2.2%), suggesting a heightened sensitivity to the

FV (%)	w/b	SF (%)	FA (%)	FL (mm)	a_0 (mm)	SP (%)	FD (mm)	CA (days)	SD (mm)	ST (mm)	CMOD
0	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	0.29
0.2	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	0.35
0.4	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	0.42
0.6	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	0.48
0.8	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	0.54
1	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	0.6
1.2	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	0.66
1.4	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	0.72
1.6	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	0.78
1.8	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	0.85
2	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	0.92
2.2	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	0.99
2.4	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	1.06
2.6	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	1.11
2.8	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	1.17
3	0.19	20.08	10.18	12.34	32.89	1.65	0.22	28	370	75	1.22

Table 9. New experimental dataset for assessing machine learning model sensitivity to variations in the FV.

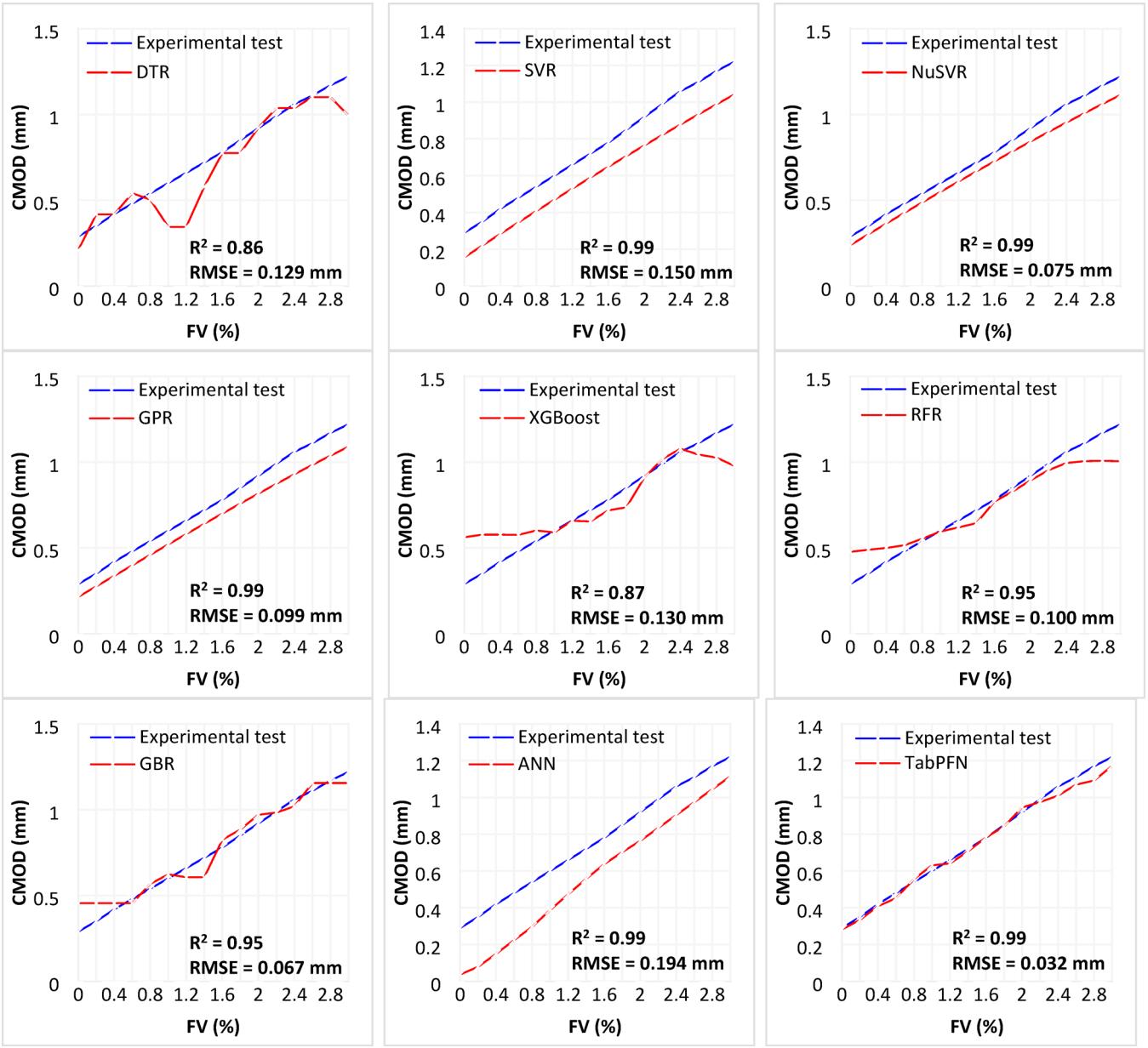

Fig. 14. Comparison of machine learning predictions and experimental results for varying FV with all other parameters held constant.

local variance present in the training set. The DTR model registers sharp local oscillations that contrast with the expected experimental smoothness, especially within the low to mid FV range. This behavior aligns with the known overfitting risk in single-tree formulations, which lack the level of smoothing and generalization that ensemble and kernel methods achieve.

The majority of models underreport CMOD at FV values exceeding 2.4%. This recurring underprediction emphasizes the struggle encountered when algorithms seek to extrapolate the nonlinear effects of fiber-bridging at elevated reinforcement levels, where mechanisms hindering crack propagation increasingly dominate.

The thoughtful design outlined in Table 9, along with meticulous feature selection, established a robust and unbiased foundation for exploring how FV affects CMOD predictions within the model framework. Of all methods tested, TabPFN and ANN delivered the clearest reproduction of the experimental CMOD–FV curve, showing very small bias and consistent performance. Kernel methods and ensemble strategies adequately followed the broad trajectory but revealed persistent offsets or concentrated errors, and the DTR proved markedly volatile. These observations highlight the critical role of sophisticated architectures (especially probabilistic foundation networks and deep neural networks) in accurately tracing CMOD changes as FV varies in ultrahigh-performance concrete.

SHAP analysis

Complex machine learning models need interpretability, especially in engineering applications. SHAP offers a theoretically solid framework that's transparent and rigorous. SHAP builds on Shapley values from cooperative game theory, treating each feature as a "player" in a predictive "game" where the prediction becomes the "payout." Each feature's contribution gets quantified as its average marginal effect on model output across all possible feature coalitions[66]. This approach guarantees two important properties: local accuracy (SHAP values for any instance sum to the model prediction) and consistency (if a model changes so a feature's marginal contribution increases, its SHAP value can't decrease).

Recent civil engineering studies show SHAP works well for understanding input variables in predicting mechanical and durability properties of concretes. Examples include electrical resistivity of fiber-reinforced coral aggregate concrete[67] and compressive strength of coral aggregate concrete[68]. These studies demonstrate that SHAP can connect data-driven predictions with physical understanding, making it well-suited for cementitious composite applications.

Here, SHAP was applied to the developed machine learning models to quantify each input parameter's contribution to predicted CMOD. Unlike conventional feature importance measures, SHAP breaks down predictions into exact feature contributions at both global (overall trends) and local (individual specimens) levels. This provides trustworthy interpretation, identifies the main drivers of fracture behavior, and reveals how their relationships influence predictions across different specimens. SHAP connects data-driven forecasts to mechanistic fracture processes, offering physically meaningful model validation.

Because the TabPFN model consistently delivered higher predictive accuracy and better generalization than other machine learning methods, the SHAP analysis was restricted to that architecture for a thorough examination of the reasoning behind its predictions.

Figure 15 shows SHAP force plots for three selected samples (A, B, and C), detailing how six important features (FA, SF, FV, FL, a_0, and w/b) each affect the forecasted CMOD values. In the plots, red bars represent features that raise the prediction above the baseline, while blue bars represent features that lower it. The length of a bar shows the size of the effect, indicating how strongly that feature influences the model for the particular sample.

Sample A has a predicted CMOD of 0.72 mm. The strongest influences come from SF at 0.983% and FL at 1.0 mm. SF decreases the CMOD, reflecting its success in limiting post-crack deformation and thus its contribution to stronger matrix confinement. At the same time, FL, FV, and a_0 raise the prediction: the full FL of 1 mm aids in bridging the crack, and the FV of 0.535% adds to that effect. FA and a_0 also slightly increase CMOD. The final CMOD prediction of roughly 0.7 mm shows how fiber arrangement and the geometry of the specimen work together.

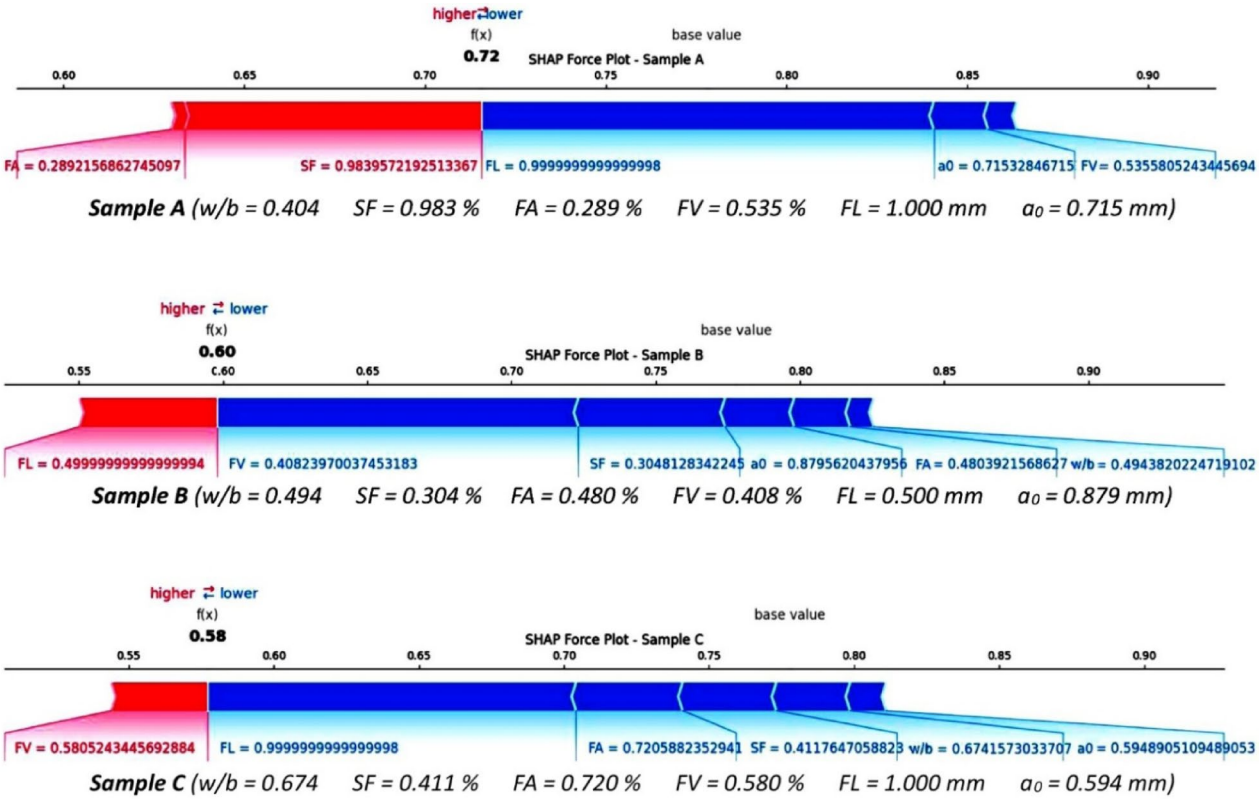

Fig. 15. Force plot for local interpretation of three sample points based on the LabPFN model.

In Sample B, the predicted CMOD is 0.60 mm. Here, FL (0.5 mm) and a_0 (0.879 mm) rank as the crucial parameters, while FV and FA exert moderate positive influences. SF consistently shows a negative role. The relatively short FL = 0.5 mm limits the bridging action, and thus lowers the CMOD, whereas the larger a_0 pushes the predicted displacement upward. A modest SF (0.304%) weakens confinement, which only partly offsets the benefits of the other features. This example shows how the interplay of geometric and fiber characteristics shapes the post-cracking behaviour.

For Sample C, the predicted CMOD is 0.58 mm. FV (0.58%) and FA (0.720%) make the most substantial positive contributions, while SF (0.411%) actively detracts. FL and a_0 still help, but their effects come in at smaller magnitudes. The considerable FV boosts load transfer across the crack, accounting for its leading positive role. A moderate FA further encourages CMOD growth, and SF, while stiffening the matrix, keeps displacement in check. This scenario underscores that FV is vital for dictating post-cracking deformation, especially when FL is at its upper limit (1 mm).

Across all three experiments, FV and FL of the fibers consistently emerge as strong positive drivers of the crack-bridging performance, backing earlier lab results that underscored their decisive roles. In contrast, SF tends to exert a negative pull, which hints that too much spacing or too little fiber interaction can hinder the capacity to resist further crack widening. Examination of the SHAP force visualizations uncovers fine, sample-responsive dependencies, indicating that the relevance of each feature can shift even for the same model, dictated by the input feature subsets. Crucially, the TabPFN architecture captures these intertwined, non-linear relationships: the varying roles of the factors FV, FL, a_0, and SF across the three experiments track closely with the core tenets of fracture mechanics.

Taken together, the local SHAP diagnostics yield a mechanistic lens on how each variable steers the CMOD for UHPC. The fiber indices (especially FV and FL), emerge as the primary drivers of crack-bridging capacity, while aggregate gradation and spacing terms adjust the intensity of that response. Beyond reinforcing the TabPFN model's explanatory power, these interpretations provide a systematic basis for refining fiber-reinforced concrete designs, leveraging transparent, evidence-based guidance grounded in the experimental data.

Expanding on the localized SHAP force plots, the global sensitivity ranking presented in Fig. 16 consolidates the influence of the six input variables onto a single axis. The analysis shows that FV significantly influences the model output, boasting a mean SHAP value of about 0.17, which is considerably higher than the impact of other features. FL takes the second position and yields a comparatively lower, yet still significant, impact. The contributions of SF and w/b are of moderate magnitude, effectively coupling mechanical and microstructural responses. Minimal yet positive SHAP values are attributed to FA and a_0, verifying their marginal effect on the computed CMOD within the LabPFN architecture. The obtained ordering aligns precisely with fracture mechanics doctrines, whereby the spatial density of the fibers controls the area available for bridging, the length influences the kinetics of pull-out processes, and spacing together with the w/b ratio mediate the confinement of the cement matrix and, consequently, its toughness.

The SHAP heatmap in Fig. 17 deepens our understanding of feature contributions at the instance level. The FV feature displays a clear monotonic behavior: when FV is low (blue), CMOD predictions drop, whereas at higher FV values (red), CMOD predictions rise sharply. This observation is consistent with tests showing that increased FV allows cracks to remain open longer without immediate coalescence, thereby permitting greater openings to develop. The FL feature, conversely, presents alternating zones of red and blue, suggesting effects that vary with the case and confirming the well-documented balance between fiber bridging and pull-out resistance provided by FL. Variables SF, w/b, and FA exert moderate influences, with patterns scattered across instances, indicating their conditional relevance that depends on overall mixture design. Finally, a_0 contributes little overall, though isolated cases reveal localized impacts on CMOD.

The joint evaluation of Figs. 16 and 17 indicates that FV and FL dominate CMOD predictions in fiber-reinforced UHPC, whereas SF and w/b remain subordinate yet structurally relevant ancillary variables. The LabPFN architecture accurately encodes these mechanistic links and, crucially, resolves the subtle interactions

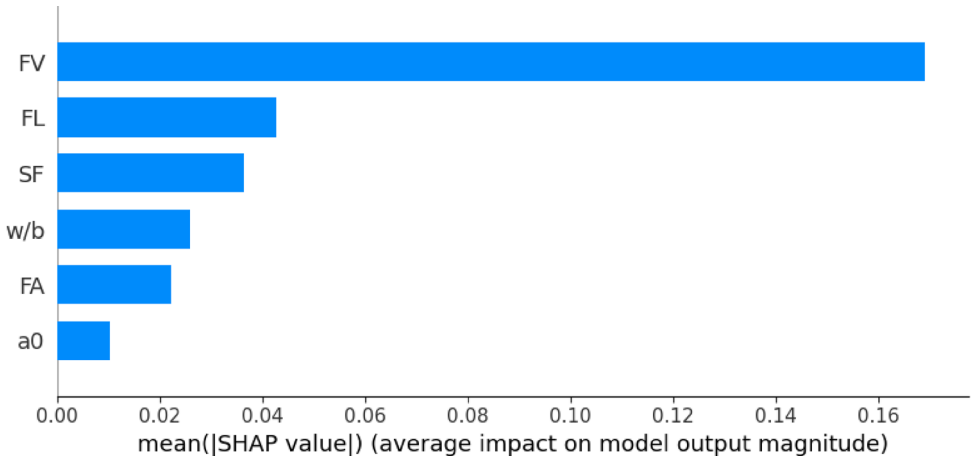

Fig. 16. Total sensitivity index of input features on the LabPFN model.

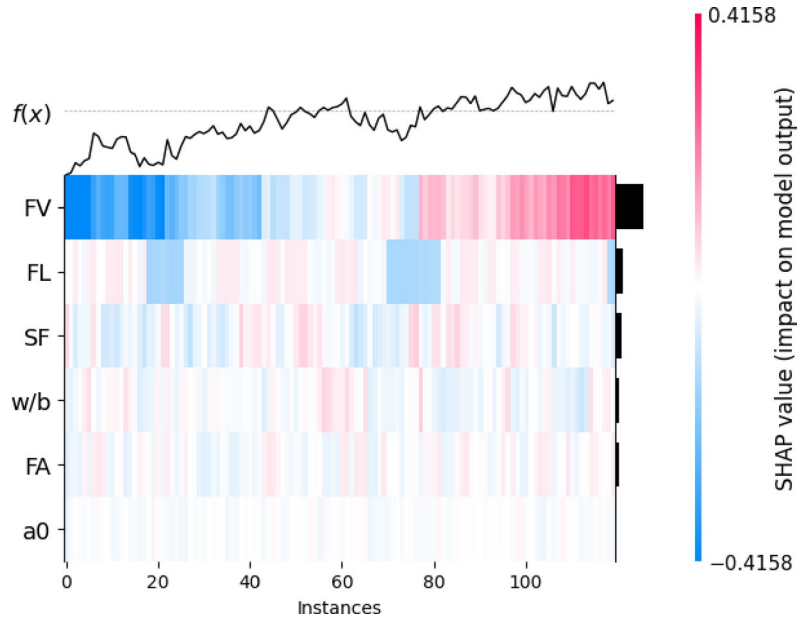

Fig. 17. SHAP heatmap based on the LAbPFN model.

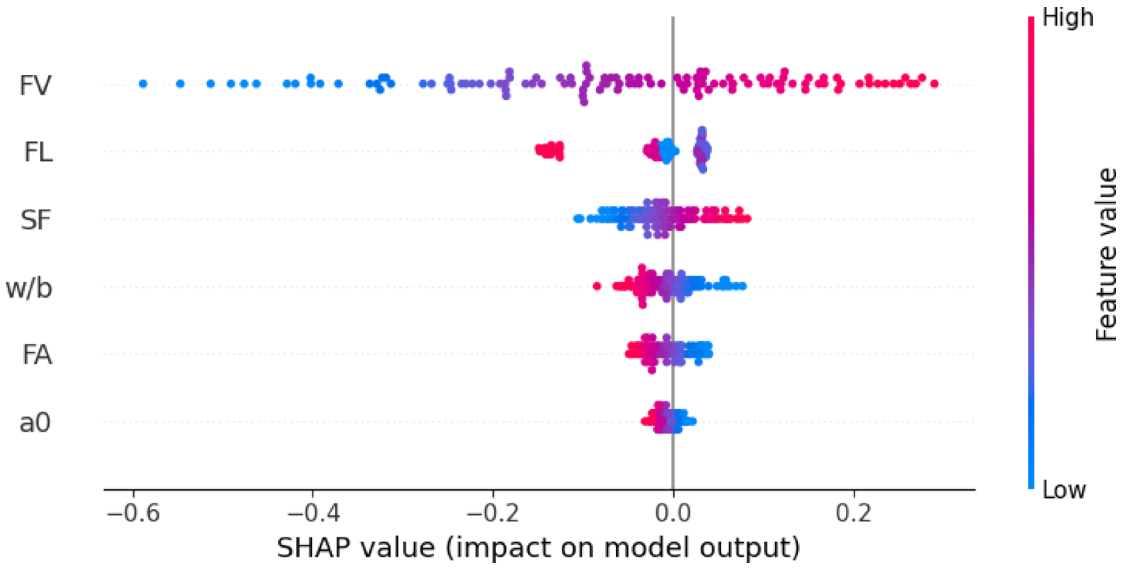

Fig. 18. Global SHAP values based on the LabPFN model.

among the ensemble of mixture parameters. The consistent trends seen at both the population level and individual instances back up the model's statistical reliability, laying the groundwork for data-driven optimization of FR-UHPC.

Figure 18 presents a global SHAP summary plot in which each dot represents an individual prediction. The horizontal position stands for SHAP value (both size and sign of influence on CMOD), and color conveys actual feature value (blue for low and red for high). The layout thus permits a simultaneous view of feature significance, influence direction, and variation inside the same feature.

As before, FV remains the leading source of influence. High FV instances (red dots) congregate on the positive x-axis, indicating a substantial positive effect on CMOD predictions. Conversely, low FV cases (blue dots) cluster on the negative x-axis, suppressing CMOD outputs. The consistent separation between red and blue exemplifies the expected mechanism: increased FV improves crack bridging, leading to larger crack openings and consequently elevating CMOD.

FL exhibits a contrasting distribution. Both high and low FL values cluster close to zero, occasionally displaying slight negative SHAP values, which signals that length effect on CMOD varies with context rather than follows a strict trend. This variability can be interpreted in terms of the dual role that longer fibers can play:

while their bridging capacity tends to reduce crack propagation, a weak matrix bond may provoke pull-out, thus countering the potential benefit and yielding smaller or even negative contributions in some cases.

SF demonstrates a scattered centroid around zero, favoring negative effects from small interactions while wider (red) values, though sparse, incline toward small positive effects. This indicates that tighter spacing curtails crack advance by enhancing bridging, whereas expanded distances diminish reinforcement action and elevate CMOD values.

The w/b parameter is distributed more or less evenly about zero, though the excess water side (red) shifts the prediction lower. This behavior matches the mechanistic insight that surplus water dilutes the binder matrix, weakens interfacial attachment at the fiber–matrix junction, and subsequently allows wider crack openings.

FA shows comparatively modest spread around the zero mean, and samples with both small and large aspect ratios yield minor away from zero shifts. Hence, FA is comparatively subordinate relative to FV and FL, but the spread indicates that pull-out efficiency can be within reach of influencing behavior under targeted design regimes.

Lastly, a_0 induces a small but uniformly negative contribution in the higher error end (red): deeper notches reduce CMOD predictions, probably by augmenting crack-tip angular stress orientation and lessening reinforcement effects at the bridging segment.

In Fig. 19 the SHAP dependency and interaction plots produced by the LabPFN model elucidate the role of FV and its coupling with other key variables (SF, FA, FL, a_0, and w/b) on the modelled CMOD outcome. Each subplot quantifies the change in SHAP value for FV against its absolute value, while a second feature introduces a graded interaction, represented by the background hue.

Fig. 19(a) plots FV's SHAP value against the measured FV magnitude, with color serving as a marker for SF. The scatter reveals a near-uniform linear ascent, indicating that increments in FV consistently drive the SHAP value higher; the robust rising slope conclusively affirms that FV exerts a predominant beneficial adjustment to CMOD. Superimposed color indicates that larger SF values compress the slope, thereby suggesting that more closely spaced or more abundant fibers temper the magnitude of the core FV effect without negating its direction.

Fig. 19(b) retains the FV axis and now overlays FA as the conditional feature. The same upward trade-line accompanies a colorized gradient that deepens as FA increases, thereby reinforcing the FV dividend. The change in color directly infers that higher aggregate fractions complement fiber content, magnifying the negative incremental linear through a combined influence that ultimately appears to fortify fracture energy dispersion and post-failure dilation.

In Fig. 19(c) we compare FV to FL. Once again the relationship is linear; SHAP values rise steadily with increasing FV. The accompanying color scale reveals that longer fibers, corresponding to higher FL, intensify the beneficial response to FV. This observation agrees with fracture mechanics where longer fibers provide a more effective bridging mechanism, increasing the degree of crack deflection and absorbing more dissipated energy. The combined effect of increasing FV and FL illustrates a synergistic enhancement of toughness in the post-cracking stage.

Fig. 19(d) correlates FV with a_0. The linearity is retained, and FV still exerts the principal positive influence. The color gradient, however, identifies larger SHAP values as a_0 increases, signifying that the beneficial effect of fiber addition is accentuated in specimens already pre-cracked. This reinforced response arises because the same fiber distribution now encounters a wider opening, and fibers thus play a decisive role in restraining further extension of the crack and inhibiting unstable crack growth.

The analysis in Fig. 19(e) corroborates the role of FV when examined alongside the w/b. Although the sportype remains linear, the observed curvature indicates a distinct nonlinear coupling. Low FV contexts exhibit negative or nearly neutral SHAP values, shifting sharply to dominant positive influence under higher FV. Furthermore, the color gradient establishes that elevated w/b ratios scale up the fissural strength of FV, clarifying that in less compact, more permeable regions, supplementary fiber is requisite for efficient micro-crack control.

Collectively the SHAP dependency and interaction illustrations consistently relegate FV as the foremost positive driver of CMOD across all tested couplings. Nevertheless, the strength of that advantage is systematically modulated by additional constituents (SF, FA, FL, a_0, and w/b) whose effects, though secondary, remain significant. The integrated findings validate the primacy of FV and FL in generating crack-bridging force, while the mixture of aggregate type (FA, w/b), pre-existing micro-geometry (a_0), and SF govern overall reinforcement performance. This thorough interpretability examination supplies compelling evidence that the LabPFN framework not only assimilates linear relationships but also accurately captures the nonlinear and interaction rules prescribed by failure mechanics, thereby furnishing a physically consistent and data-validated reference for fine-tuning fiber-reinforced UHPC formulations.

The SHAP analysis not only sheds light on how model predictions work but also pinpoints the most crucial parameters for managing cracks in UHPC. To make these findings more useful, we translated the key features identified by SHAP into practical advice for mix design, which you can find summarized in Table 10. The analysis shows that fiber-related factors are the main drivers for controlling CMOD. It turns out that increasing the FV has the most significant impact, enhancing crack bridging and toughness after cracking. However, using too much fiber can make the mix less workable and increase the chances of fiber clumping, which is why incorporating admixtures like superplasticizers and adding fibers in stages is essential for practical use.

Additionally, FL and aspect ratio play a vital role, especially when it comes to bridging larger cracks. While longer fibers boost toughness, they can also make the mix harder to work with, suggesting that a combination of short and long fibers in hybrid systems might strike a better balance. The SHAP results also emphasize the significance of the bond properties between the fibers and the matrix. A strong bond improves load transfer, but if the bond is too strong, it can lead to brittle pull-out or fiber breakage. Therefore, selecting the right surface treatments and coatings is crucial to encourage stable, energy-absorbing pull-out behavior.

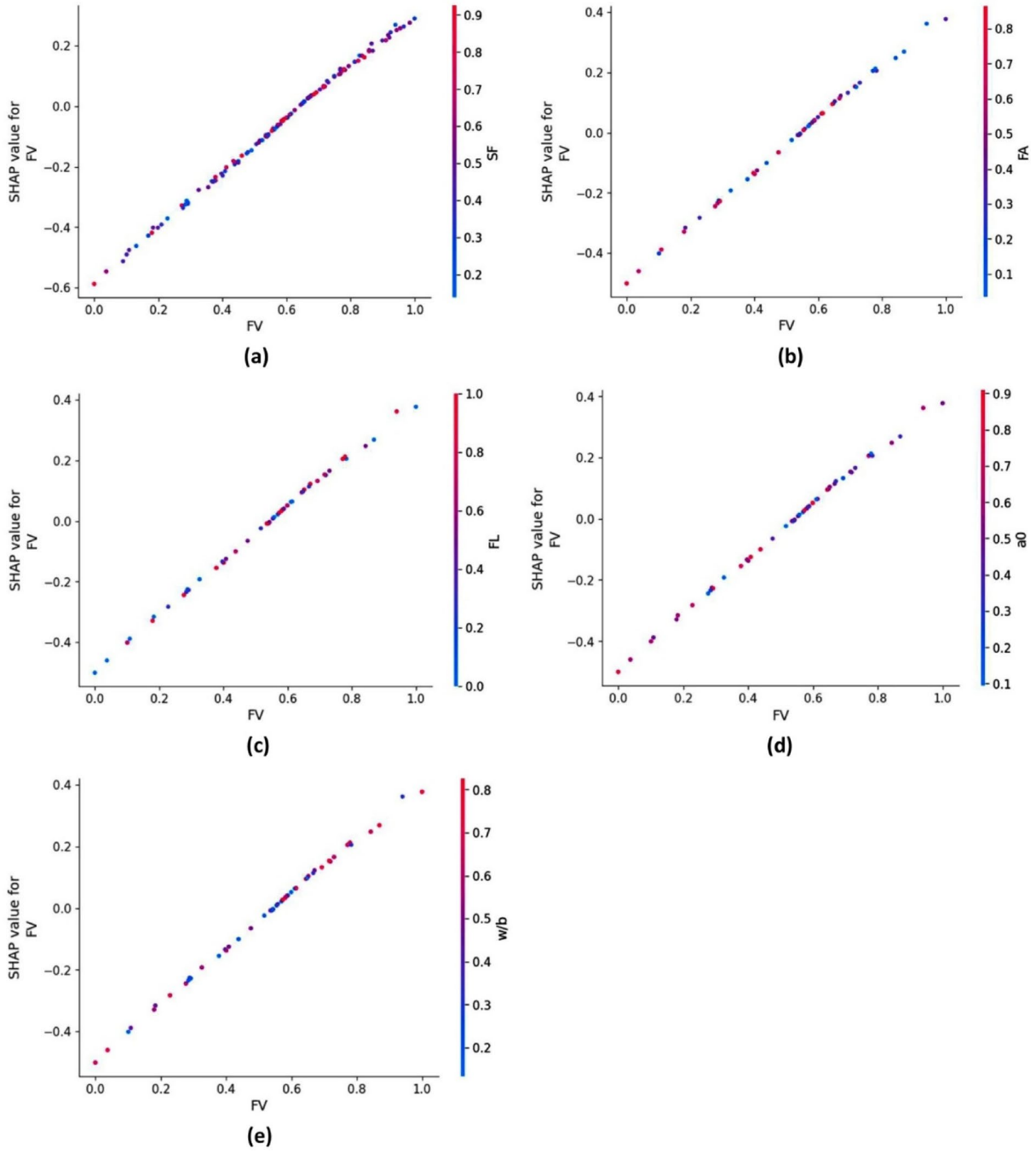

Fig. 19. SHAP dependency and interaction plots based on the LabPFN model.

Beyond the fibers, the properties of the matrix like the w/b ratio, fines content, and the addition of SF also play a role in cracking, mainly through shrinkage and stress development. Lowering the w/b ratio and optimizing fines can help reduce shrinkage, but these changes need to be paired with suitable admixtures to keep the mix workable. Likewise, proper curing and early-age practices are vital: extended moist curing or using curing membranes can significantly cut down on shrinkage-induced cracking, which in turn boosts the effectiveness of the fibers.

These findings show that while SHAP effectively points out the most significant predictors from a statistical standpoint, their true engineering value comes from helping us navigate practical trade-offs. In Table 10, a clear summary of these implications, connecting each key parameter to its mechanistic impact, possible limitations, and suggested engineering actions can be find.

Dominant parameter	Practical implication	Trade-offs / caveats	Recommended actions
FV	Primary lever for reducing CMOD; improves crack bridging and post-crack toughness	High volumes reduce workability; risk of fiber balling	Incrementally increase fiber volume; use superplasticizers and staged addition
FL / Aspect ratio	Longer fibers improve bridging of wider cracks; aspect ratio increases energy absorption	Overly long fibers reduce workability; dispersion issues	Select optimal length; consider hybridizing long + short fibers
FT	Surface/bond affects pull-out behavior; stable pull-out preferred over brittle rupture	Excessive bond strength causes brittle pull-out or fiber rupture	Use coatings/treatments to balance bond strength; perform pull-out tests
Matrix Properties (w/b, fines)	Lower w/b and optimized fines reduce shrinkage; admixtures maintain workability	Too low w/b impairs workability; high fines increase admixture demand	Use admixtures to ensure workability; validate shrinkage performance
CT	Better curing reduces shrinkage-induced cracking, enhancing fiber effectiveness	Requires stricter site control and longer curing times	Apply moist curing, curing membranes, or temperature control
Hybrid fiber strategies	Combining micro- and macro-fibers balances micro-crack suppression and large-crack control	More complex design and mixing; may increase cost	Adopt hybrid mixes for multi-objective optimization; validate experimentally

Table 10. Engineering guidance for UHPC crack control derived from SHAP analysis.

Feature	Mean SHAP	Std. dev
FV	0.142	0.088
FL	0.036	0.033
SF	0.030	0.022
CA	0.026	0.019
w/b	0.025	0.018
FD	0.019	0.017
FA	0.019	0.014
a_0	0.013	0.011
ST	0.008	0.008
SD	0.007	0.006
SP	0.007	0.007

Table 11. Quantitative SHAP feature importance analysis for CMOD prediction.

Quantitative SHAP analysis

Table 11 presents the average absolute SHAP values along with their standard deviations for all samples, offering a numerical complement to the visual SHAP plots. The findings clearly indicate that the FV is the most significant predictor of CMOD, boasting a mean SHAP value of 0.142 ± 0.088, which is notably higher than any other feature. This prominence underscores the crucial role that fiber dosage plays in influencing crack bridging and post-cracking behavior in UHPC. Following FV, the next key contributors are FL at 0.036 ± 0.033 and SF at 0.030 ± 0.022, both of which impact the refinement of microstructure and the interactions between the matrix and fibers. Other features like CA, w/b, FD, and FA seem to have a secondary effect, each showing mean SHAP values in the range of 0.019 to 0.026. On the lower end of the influence spectrum, we find predictors such as ST, SP, and SD, all of which contribute SHAP values below 0.01. While these factors do help fine-tune the model's output, their impact pales in comparison to the dominant influence of FV.

In summary, the quantitative SHAP analysis reveals a clear ranking of feature importance that aligns well with the principles of fracture mechanics: fiber dosage stands out as the primary factor, followed by fiber geometry and binder composition, with other mix parameters having a minimal effect. This hierarchy remains consistent across different folds and is in agreement with previous experimental results.

Uncertainty quantification

Uncertainty quantification plays a pivotal role in enhancing both the reliability and the practical deployment of machine learning approaches in engineering and materials science. Conventional models produce single-valued responses, yet they often omit any indication of the associated confidence or dispersion. In critical fields like fracture mechanics and structural longevity, relying solely on these deterministic figures can expose designers to undue risk or unrecognized failure pathways. Uncertainty quantification remedies this by introducing statistically valid confidence bounds around the outputs, thereby upgrading predictions from mere accuracy to actionable trustworthiness. By systematically quantifying the uncertainty in predictions, machine learning frameworks gain a level of interpretability that is indispensable in risk-governed environments, ensuring that probabilistic safety margins and variations in material behavior are appropriately integrated.

Figure 20 illustrates the diagnostic evaluation of the TabPFN architecture specifically tuned to estimate the CMOD in concrete samples. The TabPFN's predictions (depicted as an orange line) are juxtaposed against experimentally acquired values (shown in blue), with bootstrapped 95% CIs cast in light gray bands. This incorporation of uncertainty quantification transitions the evaluation from accuracy alone to a dual exposition of both predictive capability and reliability, clearly illuminating statistical dispersion alongside the mean estimate.

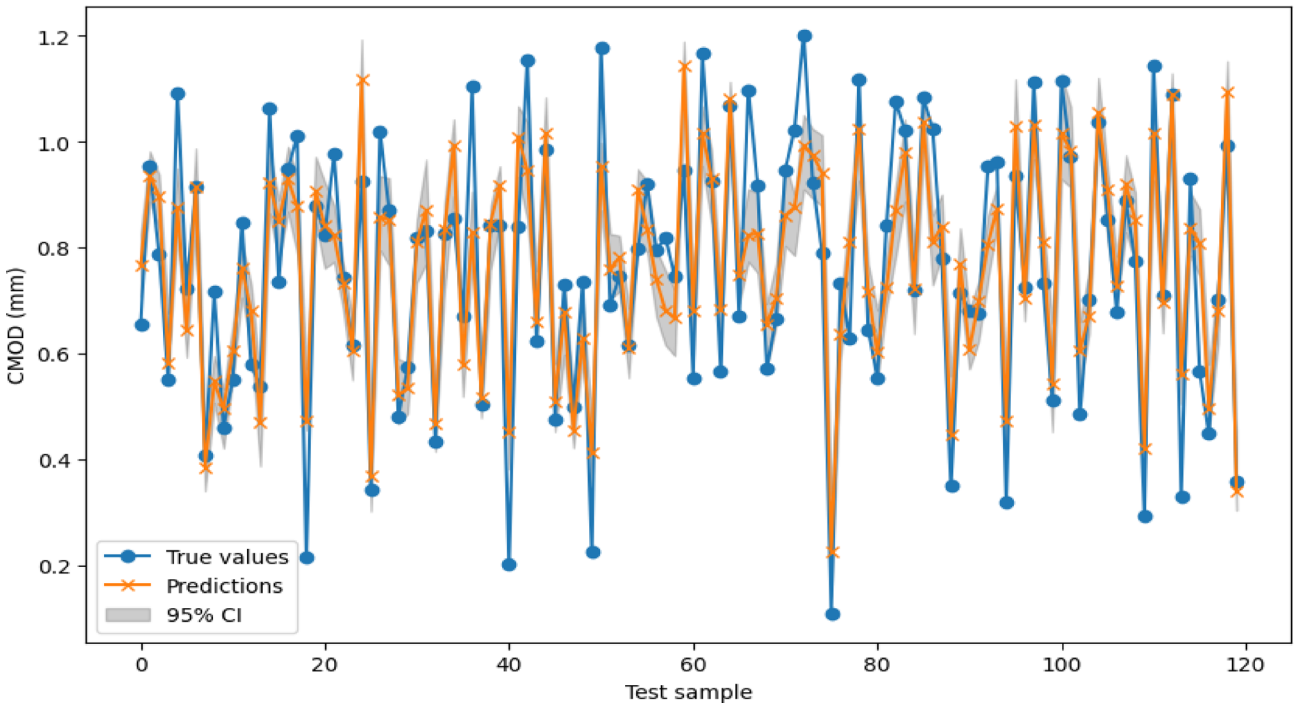

Fig. 20. TabPFN predictions with bootstrapped 95% CIs.

Evaluating prediction accuracy, the TabPFN model performs impressively, with predicted values tracking measured CMOD data very closely. It follows the data's fluctuations and nonlinear trends, showing that the model effectively learns the underlying physics. The overlap between the predicted and observed values across nearly all test samples underscores the model's broad generalization ability. While a few large deviations appear at the extremes, the strong overall alignment suggests that the model consistently manages the noisy and nonlinear properties of the CMOD measurements.

Bootstrapped 95% CIs shed light on the model's reliability. The narrow confidence bands observed throughout most test samples signal high stability and strong confidence in the point predictions. Wider bands at select points, however, indicate regions of elevated uncertainty, where data variability or model constraints exert more influence. Crucially, the true measurements lie within the confidence bands in nearly every instance, confirming that the uncertainty estimates are both realistic and well calibrated. Bootstrapping strengthens model robustness by repeatedly resampling the training dataset, thus better capturing the variability intrinsic to the prediction task. This method guarantees that the resulting CIs reflect both the uncertainty introduced by the model and the variability of the data itself (an essential capability in engineering fields where minute errors can compromise safety or the longevity of structures). Consequently, the TabPFN framework means that, in addition to delivering precise predictions, it offers a clear, uncertainty-aware output that enhances its trustworthiness for real-world applications.

From a scientific viewpoint, the addition of uncertainty quantification makes TabPFN both more interpretable and more relevant to engineering practice. In structural engineering applications involving fiber-reinforced UHPC, CMOD serves as a critical indicator of fracture toughness and overall serviceability. Being able to furnish not only precise CMOD forecasts but also reliable uncertainty bands is vital for verifying safety margins and for the ongoing refinement of design standards. In contrast to techniques like SVR or RFR, TabPFN uses prior domain knowledge along with a Bayesian-inspired framework, enabling it to perform competently with limited training data while still quantifying uncertainty with confidence.

Figure 21 offers a direct comparison between the widths of bootstrap CIs and the experimental scatter observed in CMOD measurements. For models like TabPFN, ANN, NuSVR, SVR, and XGBoost, the bootstrap CI widths (ranging from 9 to 12%) closely match the experimental scatter (between 9 and 10%). This suggests that the statistical resampling method effectively captures the variability seen in experiments. The alignment is particularly notable for TabPFN, where both metrics are identical at 9%, showcasing a strong agreement between the uncertainty derived from the model and the variability found in physical measurements. On the other hand, tree-based models such as GBR, RFR, GPR, and DTR show bootstrap CI widths (14% to 18%) that are considerably larger than the experimental scatter (10% to 12%). This indicates that these models are more influenced by how the training data is divided, resulting in predictive uncertainties that surpass what is typically observed in repeated physical tests. Overall, these results underscore that while bootstrap-based uncertainty estimation can align well with experimental scatter for certain types of models, it tends to provide more conservative (wider) bounds for others. This highlights the importance of interpreting both statistical and experimental perspectives when evaluating model reliability.

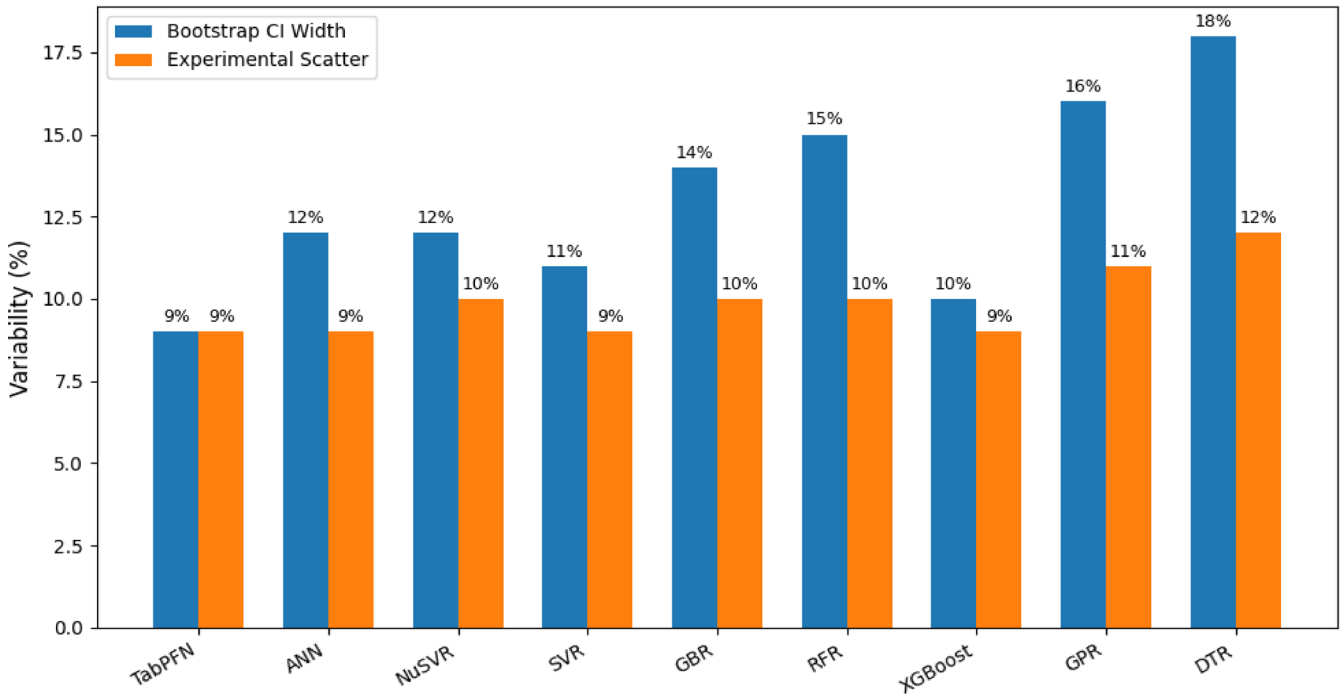

Fig. 21. Comparison of bootstrap CI widths and experimental scatter in CMOD measurements.

Conclusions, limitations, and suggestions

This research explored how machine learning techniques can be used to predict CMOD in FR-UHPC. A variety of algorithms were tested, and the more sophisticated methods, especially TabPFN and NuSVR, consistently outshone traditional models. During fivefold cross-validation, these techniques achieved impressive results. By carefully selecting features, the predictor set was narrowed down to six key variables: FV, SF, FL, FA, w/b, and a_0. Notably, FV and FL stood out as the most influential factors affecting post-cracking performance, with transformer-based models performing particularly well when trained on this refined feature set.

To enhance model interpretability, SHAP analysis was employed, confirming the critical roles of FV and FL in crack-bridging behavior, while the other four variables served as significant secondary contributors. This consistency with established fracture mechanics principles and experimental findings boosts confidence that the models are capturing genuine physical behavior rather than mere statistical flukes. Sensitivity analyses further indicated that TabPFN was more adept at handling nonlinear post-cracking behavior, especially at higher fiber dosages, while kernel-based and ensemble methods showed some residual prediction errors in these scenarios.

Uncertainty quantification was carried out using bootstrap confidence intervals, which provide probabilistic bounds on model predictions. These evaluations highlight how crucial it is to have predictions that account for uncertainty in engineering applications, as they help establish realistic safety margins. The findings confirmed that the models developed show strong predictive performance within the experimental domain studied. However, the applicability of these models to other types of UHPC or structural systems is limited by the current dataset, which lacks diversity in mix proportions, fiber contents, curing conditions, and geometrical variations. To enhance generalizability, it would be beneficial to expand datasets from various sources and employ techniques like transfer learning and domain adaptation. Until these initiatives are pursued, the existing models should be viewed as most reliable under the conditions examined.

From a practical standpoint, there are several key recommendations to consider. First off, FV should be seen as the main design factor: increasing FV can boost crack resistance, but it's important to strike a balance with workability issues. This can be managed by using superplasticizers and adding fibers in stages to prevent clustering. Next, customizing FL and aspect ratio is vital. Longer fibers and higher aspect ratios enhance the ability to bridge larger cracks, while a mix of micro- and macro-fibers can help control early-age cracking and boost toughness at the same time. Additionally, tweaking the matrix, like lowering the w/b ratio and optimizing fine materials with SF, can help reduce shrinkage, although we'll need extra admixtures to keep the mix workable. Furthermore, improving the bond between the fiber and matrix, especially through surface treatments that promote stable pull-out rather than brittle rupture, can enhance overall performance. Lastly, proper curing methods, such as moist curing, using membranes, or creating controlled curing environments, are crucial to minimize shrinkage-related cracking and maximize the effectiveness of the fibers.

As we look to the future, there are several promising research paths to explore. Merging predictive machine learning models with digital twins and structural health monitoring could pave the way for real-time, adaptive evaluations of UHPC performance. Additionally, we need to conduct systematic tests under various environmental conditions (like humidity changes, temperature fluctuations, and freeze–thaw cycles) to truly assess durability and how well models can transfer. By creating larger and more diverse datasets that cover a

wider range of UHPC formulations and shapes, we can leverage deep learning and foundational models more effectively. Moreover, employing advanced uncertainty quantification techniques, such as conformal prediction and Bayesian methods, could enhance the calibration of predictive intervals across different models. Lastly, integrating these predictive frameworks into multi-objective optimization processes would help strike a balance between crack control, cost efficiency, sustainability, and workability.

To conclude, this study illustrates that integrating advanced machine learning models, interpretability techniques, and uncertainty quantification can lead to accurate predictions of CMOD in FR-UHPC. What's even more significant is that the insights we've gained extend beyond mere predictive accuracy, providing actionable recommendations for mix design, fiber optimization, and structural applications. These contributions lay the groundwork for a more reliable and practical framework for using machine learning in UHPC crack control, while also highlighting the potential for broader integration with digital tools and experimental setups down the line.

Data availability

Data not available due to restrictions imposed by research sponsors, ongoing analysis for future studies, and the necessity to maintain data confidentiality until further validation and publication. However, are available from the corresponding author on reasonable request.

Received: 15 September 2025; Accepted: 8 October 2025
Published online: 14 November 2025

References

1. Sun, F. et al. Concrete crack opening forecasting by back propagation neural network and differential equation. *Sci. Rep.* **15**(1), 25452. https://doi.org/10.1038/s41598-025-11216-2 (2025).
2. Singh, P., Yogesh, R., Bhowmik, S. & Kishen, J. M. C. Insights into the fracturing process of plain concrete under crack opening. *Int. J. Fract.* **241**(2), 153–170. https://doi.org/10.1007/s10704-023-00692-0 (2023).
3. Sheng, D., Lou, Y., Sun, F., Xie, J. & Yu, Y. Reengineering and its reliability: An analysis of water projects and watershed management under a digital twin scheme in China. *Water* **15**(18), 3203. https://doi.org/10.3390/w15183203 (2023).
4. Tonini, D. Observed Behavior of Several Italian Arch Dams. *J. Power Div.* https://doi.org/10.1061/JPWEAM.0000062 (1956).
5. Zhang, J., Song, F., Zhang, L., Wang, J. & Liu, C. Analysis on hydraulic fracturing of concrete in super-high arch dam based on the thermodynamic principle of minimum energy consumption rate. *Int. J. Heat Technol.* **40**(2), 383–389. https://doi.org/10.18280/ijht.400204 (2022).
6. Richard, H. A., Fulland, M. & Sander, M. Theoretical crack path prediction. *Fatigue Fract. Eng. Mater. Struct.* **28**(1–2), 3–12. https://doi.org/10.1111/j.1460-2695.2004.00855.x (2005).
7. Xu, Y., Huang, Y., Xu, X. & Xiao, F. Improved hybrid model for predicting concrete crack openings based on chaos theory. *Math. Probl. Eng.* **2022**, 1–14. https://doi.org/10.1155/2022/5147744 (2022).
8. Rong, C., Peng, Y., Shi, Q. & Wang, P. Eccentric compression performance of concrete filled steel tube slotted columns: Experiment and simulation analysis. *Structures* **74**, 108580. https://doi.org/10.1016/j.istruc.2025.108580 (2025).
9. Wang, X. et al. Experimental study on the mechanical properties of short-cut basalt fiber reinforced concrete under large eccentric compression. *Sci. Rep.* **15**(1), 10845. https://doi.org/10.1038/s41598-025-94964-5 (2025).
10. Fu, C., Liu, Y., Lao, Y. & Wang, J. An analytical model based on cross-sectional geometric characteristics for calculating crack opening displacement in reinforced concrete beams. *Theoret. Appl. Fract. Mech.* **135**, 104770. https://doi.org/10.1016/j.tafmec.2024.104770 (2025).
11. Tada, H., Paris, P. C. & Irwin, G. R. The stress analysis of cracks handbook. *Third Edition. ASME Press* https://doi.org/10.1115/1.801535 (2000).
12. Wu, Z., Yang, S., Hu, X. & Zheng, J. An analytical model to predict the effective fracture toughness of concrete for three-point bending notched beams. *Eng. Fract. Mech.* **73**(15), 2166–2191. https://doi.org/10.1016/j.engfracmech.2006.04.001 (2006).
13. Wang, Y.-J., Wu, Z.-M., Zheng, J.-J., Yu, R. C. & Liu, Y. Analytical method for crack propagation process of lightly reinforced concrete beams considering bond-slip behaviour. *Eng. Fract. Mech.* **220**, 106654. https://doi.org/10.1016/j.engfracmech.2019.106654 (2019).
14. Mi, Z., Li, Q., Hu, Y., Xu, Q. & Shi, J. An analytical solution for evaluating the effect of steel bars in cracked concrete. *Eng. Fract. Mech.* **163**, 381–395. https://doi.org/10.1016/j.engfracmech.2016.06.002 (2016).
15. Shah, S. P. Determination of fracture parameters (K Ic s and CTODc) of plain concrete using three-point bend tests. *Mater. Struct.* **23**(6), 457–460. https://doi.org/10.1007/BF02472029 (1990).
16. Barr, B. I. G. et al. Round-robin analysis of the RILEM TC 162-TDF beam-bending test: Part 2 - Application of delta from the CMOD response. *Mater. Struct.* **36**(263), 621–630. https://doi.org/10.1617/13954 (2003).
17. Ding, Y. Investigations into the relationship between deflection and crack mouth opening displacement of SFRC beam. *Constr. Build. Mater.* **25**(5), 2432–2440. https://doi.org/10.1016/j.conbuildmat.2010.11.055 (2011).
18. Aslani, F. & Bastami, M. Relationship between deflection and crack mouth opening displacement of self-compacting concrete beams with and without fibers. *Mech. Adv. Mater. Struct.* **22**(11), 956–967. https://doi.org/10.1080/15376494.2014.906689 (2015).
19. Zhang, Z. & Ansari, F. Crack tip opening displacement in micro-cracked concrete by an embedded optical fiber sensor. *Eng. Fract. Mech.* **72**(16), 2505–2518. https://doi.org/10.1016/j.engfracmech.2005.03.007 (2005).
20. Aghajanzadeh, S. M. & Mirzabozorg, H. Concrete fracture process modeling by combination of extended finite element method and smeared crack approach. *Theoret. Appl. Fract. Mech.* **101**, 306–319. https://doi.org/10.1016/j.tafmec.2019.03.012 (2019).
21. Ma, Y., Qin, Y., Chai, J. & Zhang, X. analysis of the effect of initial crack length on concrete members using extended finite element method. *Int. J. Civil Eng.* **17**(10), 1503–1512. https://doi.org/10.1007/s40999-019-00413-6 (2019).
22. Yang, Y., Lei, Z., Huang, C. & Guo, X. A simulation study on the fracture process for self-compacting lightweight concrete based on extended finite element method. *Sci. Adv. Mater.* **11**(4), 547–554. https://doi.org/10.1166/sam.2019.3489 (2019).
23. Accornero, F., Rubino, A. & Carpinteri, A. Ultra-low cycle fatigue (ULCF) in fibre-reinforced concrete beams. *Theoret. Appl. Fract. Mech.* **120**, 103392. https://doi.org/10.1016/j.tafmec.2022.103392 (2022).
24. Accornero, F., Rubino, A. & Carpinteri, A. Post-cracking regimes in the flexural behaviour of fibre-reinforced concrete beams. *Int. J. Solids Struct.* **248**, 111637. https://doi.org/10.1016/j.ijsolstr.2022.111637 (2022).
25. Rubino, A., Accornero, F. & Carpinteri, A. Fracture mechanics approach to minimum reinforcement design of fibre-reinforced and hybrid-reinforced concrete beams. *Int. J. Damage Mech* **34**(6), 900–919. https://doi.org/10.1177/10567895241245865 (2025).
26. Albaijan, I. et al. Several machine learning models to estimate the effect of an acid environment on the effective fracture toughness of normal and reinforced concrete. *Theoret. Appl. Fract. Mech.* **126**, 103999. https://doi.org/10.1016/j.tafmec.2023.103999 (2023).

27. Mohamed, H. S. et al. Compressive behavior of elliptical concrete-filled steel tubular short columns using numerical investigation and machine learning techniques. *Sci. Rep.* **14**(1), 27007. https://doi.org/10.1038/s41598-024-77396-5 (2024).
28. Tarawneh, A., Saleh, E., Almasabha, G. & Alghossoon, A. Hybrid data-driven machine learning framework for determining prestressed concrete losses. *Arab. J. Sci. Eng.* **48**(10), 13179–13193. https://doi.org/10.1007/s13369-023-07714-y (2023).
29. Long, X., Mao, M., Su, T., Su, Y. & Tian, M. Machine learning method to predict dynamic compressive response of concrete-like material at high strain rates. *Def. Technol* **23**, 100–111. https://doi.org/10.1016/j.dt.2022.02.003 (2023).
30. Chen, T. Machine learning based grey wolf optimal controller of modified algorithm for reinforced concrete structures. *Comput. Concr.* **34**(6), 649–657. https://doi.org/10.12989/cac.2024.34.6.649 (2024).
31. Nassif, N., Al-Sadoon, Z. A., Hamad, K. & Altoubat, S. Cost-based optimization of shear capacity in fiber reinforced concrete beams using machine learning. *Struct. Eng. Mech.* **83**(5), 671–680. https://doi.org/10.12989/sem.2022.83.5.671 (2022).
32. Mahmoodzadeh, A. et al. Forecasting sidewall displacement of underground caverns using machine learning techniques. *Autom. Constr.* **123**, 103530. https://doi.org/10.1016/j.autcon.2020.103530 (2021).
33. Hashempour, S., Boostani, R., Mohammadi, M. & Sanei, S. Continuous scoring of depression from eeg signals via a hybrid of convolutional neural networks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **30**, 176–183. https://doi.org/10.1109/TNSRE.2022.3143162 (2022).
34. Khan, N. A., Mohammadi, M. & Djurović, I. A modified viterbi algorithm-based if estimation algorithm for adaptive directional time-frequency distributions. *Circuits Syst. Signal Process.* **38**(5), 2227–2244. https://doi.org/10.1007/s00034-018-0960-z (2019).
35. Mahmoodzadeh, A. et al. Decision-making in tunneling using artificial intelligence tools. *Tunn. Undergr. Space Technol.* **103**, 103514. https://doi.org/10.1016/j.tust.2020.103514 (2020).
36. Sun, Z. et al. Electrical resistivity prediction model for basalt fibre reinforced concrete: hybrid machine learning model and experimental validation. *Mater. Struct.* **58**(3), 89. https://doi.org/10.1617/s11527-025-02607-y (2025).
37. Sun, Z. et al. Pipeline deformation monitoring based on long-gauge FBG sensing system: Missing data recovery and deformation calculation. *J. Civ. Struct. Heal. Monit.* **15**(7), 2433–2453. https://doi.org/10.1007/s13349-025-00943-9 (2025).
38. Sun, Z. et al. Pipeline deformation prediction based on multi-source monitoring information and novel data-driven model. *Eng. Struct.* **337**, 120461. https://doi.org/10.1016/j.engstruct.2025.120461 (2025).
39. Bolbolvand, M., Tavakkoli, S. M. & Alaee, F. J. Prediction of compressive and flexural strengths of ultra-high-performance concrete (UHPC) using machine learning for various fiber types. *Constr. Build. Mater.* **493**, 143135. https://doi.org/10.1016/j.conbuildmat.2025.143135 (2025).
40. Qian, Y., Sufian, M., Hakamy, A., Farouk Deifalla, A. & Elsaid, A. Application of machine learning algorithms to evaluate the influence of various parameters on the flexural strength of ultra-high-performance concrete. *Front. Mater.* https://doi.org/10.3389/fmats.2022.1114510 (2023).
41. Diab, A. & Ferche, A. C. Prediction of tensile properties of ultra-high-performance concrete using artificial neural network. *ACI Struct. J.* https://doi.org/10.14359/51740245 (2024).
42. Xu, L. et al. Estimation of stress–strain constitutive model for ultra-high performance concrete after high temperature with an deep neural network based method. *Constr. Build. Mater.* **408**, 133690. https://doi.org/10.1016/j.conbuildmat.2023.133690 (2023).
43. Nunez, I. & Nehdi, M. L. Machine learning prediction of carbonation depth in recycled aggregate concrete incorporating SCMs. *Constr. Build. Mater.* **287**, 123027. https://doi.org/10.1016/j.conbuildmat.2021.123027 (2021).
44. Althoey, F. et al. Machine learning based computational approach for crack width detection of self-healing concrete. *Case Stud. Constr. Mater.* **17**, 01610. https://doi.org/10.1016/j.cscm.2022.e01610 (2022).
45. Zhao, S., Tan, D.-Y., Wang, J. & Yin, J.-H. Deep learning-based adaptive denoising method for prediction of crack opening displacement of rock from noisy strain data. *Int. J. Rock Mech. Min. Sci.* **190**, 106112. https://doi.org/10.1016/j.ijrmms.2025.106112 (2025).
46. Rezaiee-Pajand, M., Karimipour, A. & Abad, J. M. N. Crack spacing prediction of fibre-reinforced concrete beams with lap-spliced bars by machine learning models. *Iran. J. Sci. Technol. Trans. Civil Eng.* **45**(2), 833–850. https://doi.org/10.1007/s40996-020-00441-6 (2021).
47. Habibi, O., Gouda, O. & Galal, K. Machine learning-based prediction of crack width and bond-dependent coefficient (k) in GFRP-reinforced concrete beams. *Case Stud. Constr. Mater.* **23**, 05005. https://doi.org/10.1016/j.cscm.2025.e05005 (2025).
48. Kazemi, F., Shafighfard, T., Jankowski, R. & Yoo, D.-Y. Active learning on stacked machine learning techniques for predicting compressive strength of alkali-activated ultra-high-performance concrete. *Arch. Civil Mech. Eng.* **25**(1), 24. https://doi.org/10.1007/s43452-024-01067-5 (2024).
49. Özyüksel Çiftçioğlu, A., Kazemi, F. & Shafighfard, T. Grey wolf optimizer integrated within boosting algorithm: Application in mechanical properties prediction of ultra high-performance concrete including carbon nanotubes. *Appl. Mater. Today* **42**, 102601. https://doi.org/10.1016/j.apmt.2025.102601 (2025).
50. Shafighfard, T., Asgarkhani, N., Kazemi, F. & Yoo, D.-Y. Transfer learning on stacked machine-learning model for predicting pull-out behavior of steel fibers from concrete. *Eng. Appl. Artif. Intell.* **158**, 111533. https://doi.org/10.1016/j.engappai.2025.111533 (2025).
51. Çiftçioğlu, A. Ö., Delikanlı, A., Shafighfard, T. & Bagherzadeh, F. Machine learning based shear strength prediction in reinforced concrete beams using Levy flight enhanced decision trees. *Sci. Rep.* **15**(1), 27488. https://doi.org/10.1038/s41598-025-12359-y (2025).
52. Shafighfard, T., Bagherzadeh, F., Rizi, R. A. & Yoo, D.-Y. Data-driven compressive strength prediction of steel fiber reinforced concrete (SFRC) subjected to elevated temperatures using stacked machine learning algorithms. *J. Market. Res.* **21**, 3777–3794. https://doi.org/10.1016/j.jmrt.2022.10.153 (2022).
53. Hollmann, N. et al. Accurate predictions on small data with a tabular foundation model. *Nature* **637**(8045), 319–326. https://doi.org/10.1038/s41586-024-08328-6 (2025).
54. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**(3), 273–297. https://doi.org/10.1007/BF00994018 (1995).
55. Prasad, D. V. V. & Jaganathan, S. Null-space based facial classifier using linear regression and discriminant analysis method. *Clust. Comput.* **22**(S4), 9397–9406. https://doi.org/10.1007/s10586-018-2178-z (2019).
56. Rasmussen,C. E. "Gaussian Processes in Machine Learning,"pp. 63–71.(2004) https://doi.org/10.1007/978-3-540-28650-9_4.
57. Gad, A. F. Artificial neural networks. In *Practical Computer Vision Applications Using Deep Learning with CNNs* 45–106 (Apress, 2018). https://doi.org/10.1007/978-1-4842-4167-7_2.
58. Gayathri, R., Rani, S. U., Čepová, L., Rajesh, M. & Kalita, K. A comparative analysis of machine learning models in prediction of mortar compressive strength. *Processes* **10**(7), 1387. https://doi.org/10.3390/pr10071387 (2022).
59. Quinlan, J. R. Induction of decision trees. *Mach. Learn.* **1**(1), 81–106. https://doi.org/10.1007/BF00116251 (1986).
60. Ho, T. K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(8), 832–844. https://doi.org/10.1109/34.709601 (1998).
61. Chen, T. and Guestrin C. XGBoost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794. (2016) https://doi.org/10.1145/2939672.2939785.
62. Sun, Z., Li, Y., Yang, Y., Su, L. & Xie, S. Splitting tensile strength of basalt fiber reinforced coral aggregate concrete: Optimized XGBoost models and experimental validation. *Constr. Build. Mater.* **416**, 135133. https://doi.org/10.1016/j.conbuildmat.2024.135133 (2024).
63. Ahmed, A. H. A., Jin, W. & Ali, M. A. H. Prediction of compressive strength of recycled concrete using gradient boosting models. *Ain Shams Eng. J.* **15**(9), 102975. https://doi.org/10.1016/j.asej.2024.102975 (2024).

64. Zheng, D. et al. Flexural strength prediction of steel fiber-reinforced concrete using artificial intelligence. *Materials* **15**(15), 5194. https://doi.org/10.3390/ma15155194 (2022).
65. J. J. Berman, "Understanding Your Data," in *Data Simplification*, Elsevier, 2016, pp. 135–187. https://doi.org/10.1016/B978-0-12-803781-2.00004-7.
66. Koushik, A., Manoj, M. & Nezamuddin, N. SHapley additive explanations for explaining artificial neural network based mode choice models. *Trans. Dev. Econ.* **10**(1), 12. https://doi.org/10.1007/s40890-024-00200-6 (2024).
67. Sun, Z. et al. Investigation of electrical resistivity for fiber-reinforced coral aggregate concrete. *Constr. Build. Mater.* **414**, 135011. https://doi.org/10.1016/j.conbuildmat.2024.135011 (2024).
68. Sun, Z., Li, Y., Li, Y., Su, L. & He, W. Investigation on compressive strength of coral aggregate concrete: Hybrid machine learning models and experimental validation. *J. Build. Eng* **82**, 108220. https://doi.org/10.1016/j.jobe.2023.108220 (2024).

Acknowledgements

The author would like to thank Prince Sultan University for their support. The authors would also like to acknowledge The Office of Research and Sponsored Programs, Abu Dhabi University, Abu Dhabi (U.A.E) for offering Research, Innovation, and Impact Grant (Cost Center # 19300933). The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2025-1161-04". This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2025/R/1447).

Author contributions

A.M. and A.Ah. developed the study concept and designed the methodology. A.M. carried out the machine learning modeling and data analysis. Ab.A. and S.P. contributed to dataset preparation and experimental validation. S.A. and A.Al. assisted with statistical analysis and interpretation of the results. Ab.Al. and A.Ah. prepared the figures and tables. M.K. contributed substantially to the revision stage by enhancing the methodological rigor, refining data interpretation, and improving the clarity of the discussion and conclusions. A.M. drafted the main manuscript text. All authors reviewed the manuscript, provided critical feedback, and approved the final version.

Funding

The author would like to thank Prince Sultan University for their support. The authors would also like to acknowledge The Office of Research and Sponsored Programs, Abu Dhabi University, Abu Dhabi (U.A.E) for offering Research, Innovation, and Impact Grant (Cost Center # 19300933). The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2025–1161-04". This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2025/R/1447).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.