



Prediction of shear behavior of glass FRP bars-reinforced ultra-highperformance concrete I-shaped beams using machine learning

Asif Ahmed · Md Nasir Uddin · Muhammad Akbar · Rania Salih ·
Mohammad Arsalan Khan · Hossein Bisheh · Timon Rabczuk

Received: 1 April 2023 / Accepted: 12 August 2023 / Published online: 30 August 2023
© The Author(s) 2023

Abstract This study focuses on using various machine learning (ML) models to evaluate the shear behaviors of ultra-high-performance concrete (UHPC) beams reinforced with glass fiber-reinforced polymer (GFRP) bars. The main objective of the study is to predict the shear strength of UHPC beams reinforced with GFRP bars using ML models. We use four different ML models: support vector machine (SVM), artificial neural network (ANN), random forest (R.F.), and extreme gradient boosting (XGBoost). The experimental database used in the study is acquired from various literature sources and comprises 54 test observations with 11 input features.

These input features are likely parameters related to the composition, geometry, and properties of the UHPC beams and GFRP bars. To ensure the ML models' generalizability and scalability, random search methods are utilized to tune the hyperparameters of the algorithms. This tuning process helps improve the performance of the models when predicting the shear strength. The study uses the ACI318M-14 and Eurocode 2 standard building codes to predict the shear capacity behavior of GFRP bars-reinforced UHPC I-shaped beams. The ML models' predictions are compared to the results obtained from these building code standards. According to the findings, the

A. Ahmed
Department of Geotechnical Engineering, College of Civil Engineering, Tongji University, Shanghai 200092, China
e-mail: asif.ahmed@alumni.sjtu.edu

M. N. Uddin
Department of Disaster Mitigation for Structures, College of Civil Engineering, Tongji University, Shanghai 200092, China
e-mail: nasiruddin@tongji.edu.cn

M. Akbar (✉)
Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 61000, China
e-mail: akbarmohammad0092@gmail.com

R. Salih
Department of Civil Engineering, Red Sea University, Port-Sudan, Sudan
e-mail: rania.salih2@rsu.edu.sd

M. A. Khan
Department of Civil Engineering, Z.H. College of Engineering and Technology, Aligarh Muslim University, Aligarh 202002, India
e-mail: arsalan.khan@ifg.uni-kiel.de

H. Bisheh (✉)
Department of Mechanical Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada
e-mail: hossein.bisheh@ubc.ca

T. Rabczuk (✉)
Institute of Structural Mechanics, Bauhaus-Universität Weimar, Marienstr. 15, 99423 Weimar, Germany
e-mail: timon.rabczuk@uni-weimar.de

XGBoost model demonstrates the highest predictive test performance among the investigated ML models. The study employs the SHAP (SHapley Additive exPlanations) analysis to assess the significance of each input parameter in the ML models' predictive capabilities. A Taylor diagram is used to statistically compare the accuracy of the ML models. This study concludes that ML models, particularly XGBoost, can effectively predict the shear capacity behavior of GFRP bars-reinforced UHPC I-shaped beams.

Keywords Glass fiber reinforced polymer rebar · Shear behavior · Machine learning · Ultra-high-performance concrete · SHAP analysis

Abbreviations

UHPC	Ultra-high-performance concrete
SVM	Support vector machine
ANN	Artificial neural network
XGBoost	Extreme gradient boosting
GFRP	Glass fiber-reinforced polymer
IEPANN	Improved eliminate particles swamp optimization hybridized artificial neural network
E.T.R.	Extra tree regression
GBRT	Gradient boosts tree regression.
NN-FFA	Neural network Firefly algorithm
S.V.R.	Support vector regressor
N.S.C.	Normal strength concrete
OOB	Out of bag

1 Introduction

Ultra-high-performance concrete (UHPC) has gained significant attention and practical applications in the construction industry due to its exceptional compressive strength and durability. Its use extends to various structures, including super high-rise buildings, long-span bridges, maritime structures, bridge girders, bridge decks, and pre-stressed girders maintenance (Xue et al. 2020; Nematollahi et al. 2012). The mechanical performance, toughness, ductility, and long-term durability of UHPC make it a preferred material for critical infrastructure. The key factors in constructing advanced UHPC include developing a highly dense particle-carrying architecture and incorporating steel fibers for concrete reinforcement. Achieving an optimal steel fiber content in the UHPC

composition is crucial to establish the desired characteristics of the material. These steel fibers enhance the UHPC's tensile durability and ductility, resulting in enhanced crack resistance and overall performance. (Yu et al. 2014). Indeed, the material and structural characteristics of UHPC differ significantly from those of conventional concrete structures. UHPC is famous for its exceptional physical characteristics, including higher compressive strength, higher tensile strength, and excellent durability. These unique properties make UHPC well-suited for various applications, especially in situations where enhanced shear and flexural behavior are essential for structural stability and design (Yavas and Goker 2020a). In such stochastic settings, understanding and predicting shear behavior become even more critical for ensuring the structural integrity of UHPC members. Proper consideration of shear behavior in the design process helps prevent shear failures, shear cracks, and other potential structural issues that may arise due to the complex randomness in geometrical materials. The shear behavior of UHPC has been the subject of numerous studies to understand its mechanical response and ensure the safety and reliability of structures made from this advanced material. Some of the key studies and findings related to the shear behavior of UHPC are summarized. Simwanda et al. (2022) focused on the structural reliability of UHPC fiber-reinforced concrete beams with stirrups. They conducted a numerical analysis using non-linear finite element methods and compared the results with standard design codes to assess the accuracy of existing design approaches. The inclusion of fibers and stirrups in UHPC beams significantly influenced their shear behavior, and the study aimed to validate and enhance the design guidelines for UHPC structures. In a research by Qi et al. (2016), UHPC fiber-reinforced concrete beams underwent a four-spot bending trial to investigate their flexural response, especially changes in bending resistance. Graybeal et al. (2006) used pre-stressed I-UHPC beam and conducted three shear tests without any reinforcement involvement. The findings revealed that conventional code requirements often underestimate shear resistance of UHPC beam. This highlights importance of studying and understanding specific shear behavior of UHPC to develop accurate design guidelines. Some researchers (Yang et al. 2022; Liu et al. 2020) explored interaction characteristics with steel bars and UHPC,

particularly in context of geopolymer-based UHPC (G-UHPC). Understanding the bond characteristics is crucial for assessing the comprehensive shear behavior of UHPC structures, as the interaction of concrete with reinforcement significantly influences the load transfer capacity. Understanding the complexities of shear behavior, especially in challenging scenarios such as intense pressures and early concrete cracking (Elsayed et al. 2022; Yoo and Banthia 2016), is essential for optimizing UHPC's performance and promoting its use in practical engineering applications (Raheem et al. 2019; Schmidt and Fehling 2005). Properly designed UHPC structures with enhanced shear behavior can lead to more durable and resilient infrastructure, offering significant benefits in terms of safety and performance. The outstanding mechanical characteristics of glass FRP, such as its stiffness and strength, make it a valuable material for reinforcement (Silva and Rodrigues 2006). Understanding the bond behavior and stress-slip relationship between UHPC and GFRP is essential for using this combination in practical engineering applications. Such research could lead to the commercialization of UHPC and its successful implementation in real-world structures (Tian et al. 2019; Liao et al. 2022). Overall, these studies highlight the importance of investigating the shear behavior of UHPC and its interaction with various reinforcement materials.

The complexity of the shear mechanism in UHPC presents a significant challenge in developing a proper theoretical calculation method for predicting shear behavior accurately (Chen et al. 2019). Existing UHPC guidelines from different countries, such as the French UHPC guideline AFGC-1013 (Resplendino 2011), Japan Society of Civil Engineers guideline (Uchida, et al. 2005), and Switzerland UHPC standard (S.I.A.) (Graybeal et al. 2020), often rely on classical Strut and Tie theory and employed semi-theoretical shear strength and hybrid empirical calculation methods to estimate shear capacity. Despite efforts by researchers worldwide to propose various shear calculation methods, such as plasticity theory, Strut and Tie method, limit equilibrium theory, and pressure field theory (Nielsen and Hoang 2016), these formulas often lack consistency and may even yield conflicting results. Consequently, the reliability of these shear strength calculation methods remains questionable. Given the limitations of traditional theoretical methods, there is a need for a more reliable

and efficient approach to predict shear behavior in UHPC structures. ML models offer a sophisticated solution to this challenge. Numerous fields of Civil Engineering domain, for example, Bridge engineering (Melhem and Cheng 2003), Tunnel engineering (Xu et al. 2019), and Transportation engineering (Arciszewski et al. 1994) problems, have been successfully solved using the ML model. ML models, such as artificial neural networks (ANN) (Ghafari et al. 2012), support vector machines (SVM) (Umar et al. 2022), and tree-based ensembles (Umar et al. 2022) like XGBoost and random forest, deep learning (DL) (Wu et al. 2022), etc. have shown great promise in solving complex engineering problems. ML models can capture non-linear relationships between input parameters and shear behavior in UHPC structures. By training on experimental data from various UHPC beams, ML models can learn patterns and correlations that conventional methods may overlook. This data-driven approach enables ML models to provide accurate predictions of shear capacity while minimizing the computational effort compared to traditional numerical methods. In summary, due to the complexity and inconsistencies in existing theoretical calculation methods for predicting the shear behavior of UHPC, a more sophisticated approach like ML models is needed. By leveraging the power of data and pattern recognition, ML models can offer reliable predictions of shear behavior with reduced computational requirements, making them a valuable tool for designing and analyzing UHPC structures.

Various research studies and methodologies leverage artificial intelligence and advanced numerical techniques to address complex engineering problems. For example, Tran et al. (2023) focused on using damage indicators in combination with the ALOANN (Augmented Lagrangian-based Optimal Artificial Neural Network) method to detect damage in structures. By incorporating damage indicators, it becomes easier to identify the location and extent of damage in the structure. ALOANN is likely an optimization-based approach that utilizes ANN to efficiently and accurately detect structural damage. Nguyen and Wahab (2023) proposed a method to enhance the calibration process of 2D VARANSVOF (Variational Asymptotic Navier–Stokes–Volume of Fluid) models. These models are used to simulate how waves interact in fluid–structure interaction problems. The proposed methodology

likely improves the accuracy and reliability of the simulation results, enabling better predictions of wave interactions in complex scenarios. In research by Dang et al. (2023), ANN is integrated with balancing composite motion optimization (BCMO) to address optimization tasks related to the vibration and buckling behavior of mathematically varied microplates in unknown physical attributes. The combination of ANN and BCMO likely offers an efficient and effective approach to solve optimization problems in the context of composite materials with varying properties. Convolutional neural networks (CNN) and region-based CNN (R-CNN) were hired to forecast various types of damage and accurately predict bounding boxes encompassing the affected regions (An efficient stochastic-based

coupled model for damage identification in plate structures). By using CNN and R-CNN, a robust and accurate damage forecasting in images or visual data can be achieved.

The existing ML methods for predicting the shear performance are concisely presented in Table 1. The ML model evaluation involved eight ML models and a few hybrid ML models. Among these models, XGBoost showed the highest test accuracy with a high R^2 value ($R^2 > 0.99$), and there was relatively low variance between the training and test datasets. Additionally, ANN, R.F., and SVM also demonstrated good accuracy in both training and test datasets, with minimal variance in R^2 values. Other hybrid ML models achieved high accuracy in both training and test results.

Table 1 Previous literature for predicting C.S. and F.S. of concrete materials

Author	Problem	ML method	Training R^2	Testing R^2
Farouk et al. (2022)	Prediction of Interface Bond Strength between UHPC and NSC	SVM	0.96	0.884
Li et al. (2022)	Estimation of bond strength between UHPC and reinforcing bars	ANN	0.97	0.93
Faruk et al. (2022)	Predicting uncertainty quantification of ultimate bond strength between UHPC and steel bar	IEPANN	0.979	0.973
Abuodeh et al. (2020)	Assessment of compressive strength of UHPC	ANN		0.801
Jiang et al. (2021)	Modelling Shear strength of UHPC	SVR	0.98	0.96
Marani et al. (2020)	Predicting UHPC Compressive Strength	ETR	0.989	0.942
Rahman et al. (2021)	Shear strength prediction of steel fiber reinforced concrete beams	XGBoost	0.998	0.739
Fu et al. (2021)	Predicting the shear strength of C.R.C. beams	GBRT	0.98	0.95
Ly et al. (Ly et al. 2020)	Prediction of Shear Capacity for Steel Fiber R.C. beams	NN-FFA	0.96	0.95
Wakjira et al. (2022a)	Flexural capacity prediction of R.C. beams	XGBoost	99.3	99.3
Sun et al. (2021)	Prediction of Flexural Strength for Waste Glass Cementitious Composite	FS	0.98	0.95
Zhang et al. (2020)	Shear strength prediction of reinforced Concrete beams (Check all)	RF	0.97	0.94
Sarothi et al. (2022)	Predicting the bearing capacity of double shear bolted connections	ANN	0.817	0.563
Dai et al. (2022)	Prediction of the moment capacity of cold-formed steel channel beams	XGBoost	0.99	0.99
Liu et al. (2021)	Shear transfer strength of concrete joints	SVR	0.97	0.95
Wakjira et al. (2022b)	Shear capacity prediction of FRP RC beam	XGBoost	0.994	0.953
Hu et al. (2022)	Predicting the Plate-End Debonding of FRP-Strengthened RC Beams in Flexure	BP	0.97	0.90
Su et al. (2021)	Identification of the interfacial cohesive law parameters of FRP strips	ANN	0.99	0.99
Salem et al. (2022)	Evaluation of the Strength of Slab-Column Connections with FRPs	Boosted SVR	0.98	0.97
Wang et al. (2022)	predicting the effective stiffness of precast concrete column	SVR	0.98	0.86
Guo et al. (2021)	Predicting Mechanical Properties of High-Performance FRCC	XGBoost	0.993	0.995

As far as the author knowledge, no previous research has been carried out specifically on UHPC with GFRP bars using standard country code and compared with ML models. To address this research gap, the study used four different ML models to estimate shear response of I-UHPC GFRP and compared the ML results with conventional country code results.

Employing the experimental data collected from the literature review, proposed ML models outperformed the traditional country code results in predicting the shear response of IUHPC-GFRP. This suggests that the ML models provide greater accuracy in predicting shear response of UHPC beams, especially in combination with GFRP reinforcement, compared to the conventional methods based on standard country codes.

In summary, the study demonstrates the potential of ML models to enhance the accuracy of predicting shear behavior in UHPC structures, particularly when considering the use of GFRP reinforcement. The results provide valuable insights into the application of ML techniques for optimizing the design and performance of UHPC structures in practical engineering applications.

2 Shear strength models of UHPC beams for different codes and proposed equations

Figure 1 depicts the interconnection between the maximum force and the central deflection for UHPC beams without strips (a) and with stirrups (b). The figure displays an ideal model of GFRP UHPC beam specimens with and without stripes.

The available design code, such as ACI318M-14 and Eurocode2, established functional equations to estimate the response of the UHPC I-shaped beam in terms of material characteristics and geometric range.

2.1 ACI318M-14

$$V_u = \left[\sqrt{f'_c} + 120\rho_s \left(\frac{d}{a} \right) \right] \frac{b_w d}{7} + \frac{A_{sw} f_{yv} d}{s} \tag{1}$$

The f'_c is denoted as compressive strength, and the longitudinal reinforcement ratio is termed as ρ_s . A , S_w and w is regarded as distance of shear span, operative width, and operative beam depth. Additional

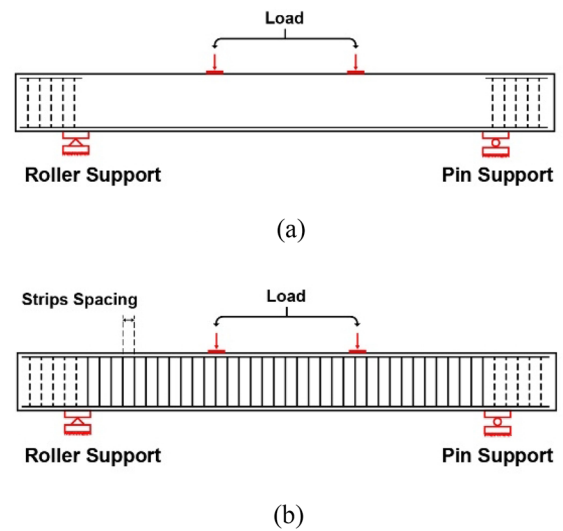


Fig. 1 A model of GFRP-reinforced UHPC beam specimens

parameters include A_{sv} the stirrup’s cross-sectional area and its yield strength, and s denoted as stirrup spacing.

2.2 Eurocode 2

$$V_{RD} = \left[0.18k(100\rho_s f'_c)^{\frac{1}{3}} \right] b_w d + \frac{A_{sw} f_{yv} Z}{s} \tag{2}$$

Here,

$$k = 1 + \sqrt{\frac{200}{d}} \tag{3}$$

Other values are the same as the ACI318M-14 specifications. Deviation), StandardScaler was adopted to handle the missing data processing in 3DP-FRC.

3 Database development

The formation of ML prediction models highly relies on constructing an authentic dataset. For this reason, a comprehensive literature study was conducted to compile information from previously published studies. Therefore, 54 UHPC beams were collected from 8 different published literature (Akbar et al. 2023; Yavas and Goker 2020b; Chenggong et al. 2022; Mészöly and Randl 2018; Ashour et al. 1992; Lee

et al. 2012; Bahij et al. 2018; Jin et al. 2020) to understand the shear performance of glass FRP bars reinforced UHPC I-shaped beam. The datasets stand for varying situations involving shear performance conditions of FRP bars reinforced UHPC I-shaped beam, and all the beams ranged from having 0% additive to having 2% admixture by weight. The input variables were used to forecast the shear performance, and eleven input variables are comprised in our database, which is also used in predicting shear capacity through the county code conventional formula. The input parameters are cubic compressive strength (f_c), cylinder compressive strength (f_{cc}), the operative width of beam (b_w), the operative depth of the beam (d), ratio of shear span to depth(λ), longitudinal reinforcement proportion (ρ_s), the angle between compressive stress and beam axis (θ), the cross-sectional area of stirp (A_{sv}), stirrup spacing (s), yield strength of stirrup (f_{yv}) and proportion of steel fiber volume (p_f). The final output included in this study's database was the experimental value (V_e) of shear behavior of glass FRP bars. Table 2 presents the analytical properties of the dataset's input variables, output variable, and model performance metrics, in average, standard variation, lowest, and hishest values.. The (f_c) ranges between 108.2 and 193 MPa vary between 96.8 MPa and 173.7 MPa while remaining between 225 and 640 MPa, respectively. Another parameter, such as (λ) is 3.5 and (P_f) can be added to 4.5%.

Figure 2 demonstrates the correlation matrix (Li et al. 2021) between inputs and outputs variables, and a small square represents the correlation between the parameters on each axis. The same two parameters

are linked together in those squares; as a result, the plot shape becomes symmetrically diagonal. The correlation is calculated in the range of negative one to positive one, and no evident linear connection involving the two variables if the values are closer to zero. The correlation drawing nearer to the value of 1 demonstrated that inputs are more closely linked with one another. If a specific inputs value rises, another does accordingly. Corresponding results can be obtained with a correlation nearer to a negative one, but instead of both variables increasing simultaneously, one will decrease when the other rises.

The highest correlation coefficient's absolute value is about 0.65, thereby revealing that there is a notable linear relationship between f_c and f_{cc} . The f_{yv} and the A_{sv} have a strong positive correlation. Moreover, with a correlation value range of -0.87, b_w has a negative correlation with λ .

3.1 Data preparation and handling the missing value

The entire database undergo z-score normalization to standardize the values as experimental data collected from multiple origin exhibits diverse unit and range.

$$Z_{in} = \frac{x_{in} - \bar{x}_n}{\delta_j} \tag{4}$$

The Z_{in} represents the standardized value of x_{in} , the nth variable of lth data's inputs, \bar{x}_n is the average value of nth input parameters, and n is the nth dispersion measure of nth input parameters.

Table 2 A detailed examination of analytical features

Input	Unit	Mean	Standard	Minimum	Max	Operation
f_c	MPa	145.744	24.155	108.2	193.0	Input
f_{cc}	MPa	125.565	25.580	96.8	173.7	Input
b_w	mm	85.185	38.349	50.0	150.0	Input
d	mm	365.00	118.29	225.0	640.0	Input
λ		2.433	0.876	1.2	3.5	Input
ρ_s	%	3.730	1.710	0.8	6.0	Input
θ	Degree	38.981	8.852	23.0	67.0	Input
A_{sv}	mm ²	57.666	65.876	0.0	157.0	Input
S	mm	118.750	118.107	0.0	500.0	Input
f_{yv}	MPa	272.475	269.124	0.0	594.0	Input
ρ_f	%	1.300	1.093	0.0	4.5	Input
V_e	kN	353.512	217.141	0.0	845.0	Output



Fig. 2 Correlation between input and output parameters

$$\delta_j = \sqrt{\frac{1}{D-1} \sum_{i=1}^D x_{in} - \bar{x}_n} \tag{5}$$

where D denotes the number of data entries. Z normalization method control the mean of every inputs about zero and standard variation about 1.

Numerous difficulties were encountered during the data-gathering process, including insufficient details and missing data. Several publications address the issue of missing values as a critical problem in dealing with machine learning models. Few studies have been conducted on dealing with missing data through statistical approaches(Audet et al. 2022; Fei

et al. 2023; Enders 2022; Memon et al. 2022); other scholars have taken different approaches to fill up the missing data by replacing missing values with zero, mean values, and medians (Enders 2022; Somer et al. 2022; Little and Rubin 2019). Several numerical models were conducted by a few authors, such as Chakraborty and Gu (Chakraborty and Gu 2019) proposed a mixed numerical model to deal with a high percentage of missing values, Wei et al. (2018) impended eight imputation models. They compared the models for handling different types of missing values, and Nazbzul et al. (2020) suggested a general framework for adapting insufficient heterogeneous missing data.

For this study, we used XGBoost to fill up the missing data, and XGBoost is a popular machine-learning model for dealing with missing values before normalization. Some other imputation approaches, such as K.N.N. and decision trees, are also employed to deal with missing data (Emmanuel et al. 2021), but XGBoost is more accurate when handling missing data (Aydin and Ozturk 2021). Built on the internal sparsity-aware algorithm, the XGBoost method can handle missing values in a more sophisticated way (Chen and Guestrin 2016).

The XGBoost machine learning model will estimate the split object for the non-missing data throughout the model training process, choose the largest split object, and appoint it as a split node by selecting a specified threshold for a specific variable. A sparsity-aware split-finding algorithm can efficiently manage missing values. Therefore, only non-missing samples are evaluated to determine the best direction, reducing the algorithm's complexity. The flow chart in Fig. 3 will graphically illustrate the process.

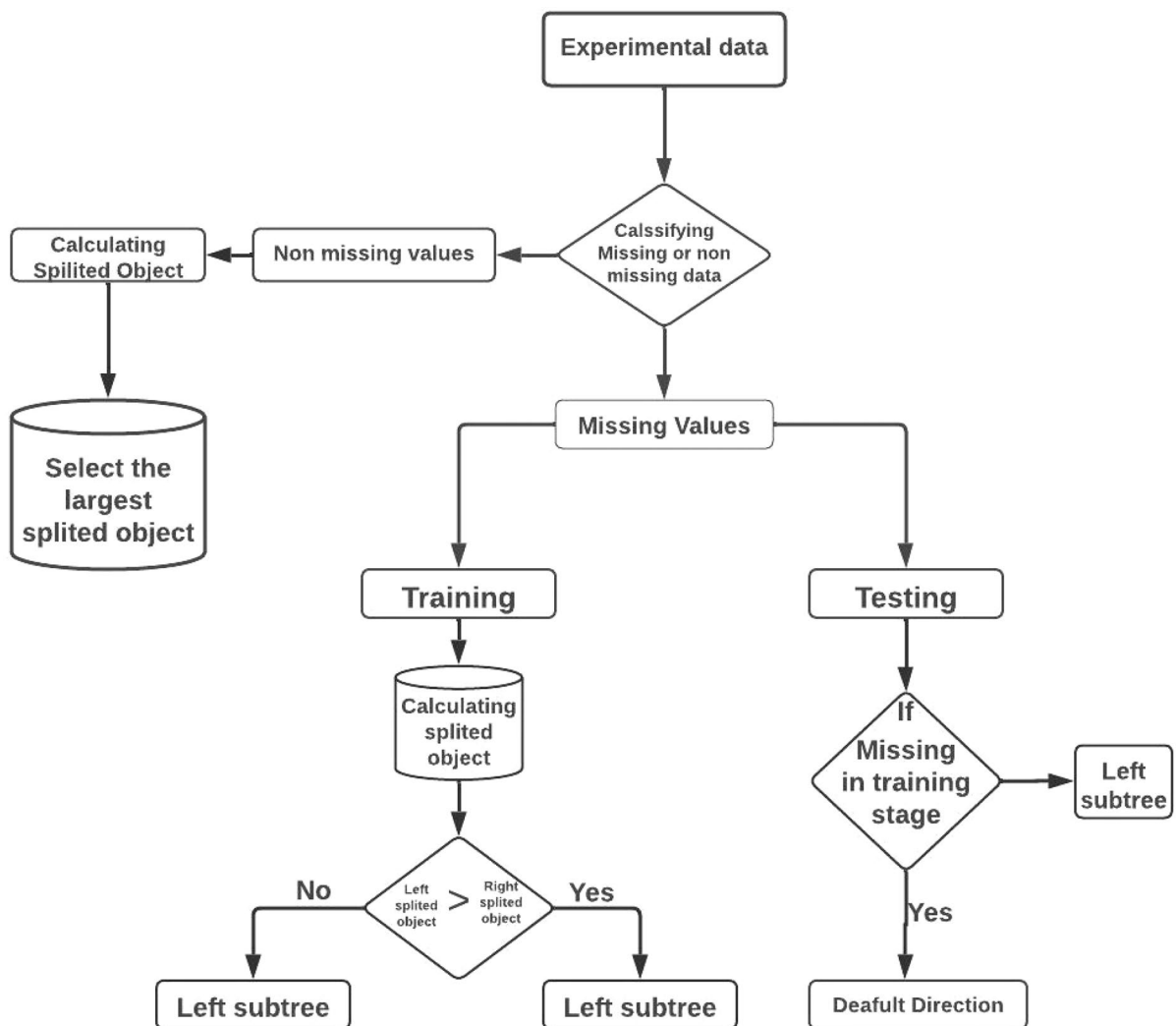


Fig. 3 Flowchart of Handling Missing Data Through XGBoost

4 Applying machine learning (ML) model

4.1 Support vector machine (SVM)

Structured risk minimization approach provides a basis for the support vector regression method, where the empirical risk needs to be reduced for every section of the structure(Awad and Khanna 2015). S.V.R. seeks to minimize the generalization error’s rather than revealing empirical faults. Let’s presume a data set

$$\{G = (x_1, y_1).....(x_n, y_n)\}, \text{ where } |n = 1.....Z \tag{6}$$

Input data base $x \in R^{G \times Z}$ is our devised matrix and $y \in R^n$ is our designed vector output which assigned for training sets. Here, x symbolizes the higher dimensional input pattern space. SVM regression aims to construct a function $p(x)$ with the highest error derivation from the actual targets for all training data.

The standard equation of the non-linear support vector regression function is

$$p(x) = w^T \sigma(x) + k \tag{7}$$

Here in Eq. (7), σ is a non-linear transformation operator, while w indicates weight, k suggests bias, and T indicates transpose vector. W and k are computed using data by reducing the normalized risk function of the variables.

$$M(c) = \frac{1}{2} ||W||^2 + C \left[\frac{1}{D} \sum_{l=1}^D L\sigma(y_i, (Px_i)) \right] \tag{8}$$

The ‘regularized term’ is the previous term on the right-hand side of Eq. (8) to evaluates the function’s flattening. It minimizes flattens level to a maximum extended level. The equation has an empirical error, and there has a loss function to calculate this error.

$$L_\sigma(y_i, P(x_i)) = \begin{cases} y_i - p(x_i) - \sigma; \text{ where, } \sigma \leq |y_i - p(x_i)| \\ y_i - p(x_i); \text{ where, } \sigma > |y_i - p(x_i)| \end{cases} \tag{9}$$

The y_i represents the actual value and $p(x_i)$ shows the estimated value. From Eq. (8), the variable C, also known as the regularized constant, is merely a variation between empirical error and the normalized function. Together C and σ hyperparameter can be specified by us where the value ϵ determine the training

accuracy. Data sets inside the function are used as a support vector. Therefore, the training set doesn’t play any role in making decisions.

The final form of the non-linear Support vector regression will be shaped as below:

$$P(x) = \sum_{l=1}^D (a_l - a_l^*) S(x_i, x_j) + b \tag{10}$$

Here a_i and a_i^* is Lagrange multipliers, which require to satisfy the positive constraints. In certain ways, the function depends on the input size and input dimension. The $k(x_i, x_j)$ denoted as kernel operator whose value is determined by the inner product $\phi(x_i, x_j)$. There are numerous varieties of kernel operator. Gaussian radial basis function (RBF) stands out as one of the most frequently utilized kernel functions and appropriate for dealing with non-linear regression problems as it can translate the training data into multi-dimensional space vector non-linearly.

4.2 Random forest (R.F.)

The Random Forest (R.F.) model combines many decision tree nodes with bootstrapping and aggregation concepts, has been successfully used for decision tree optimization methods. In this model, often known as a random forest, a randomized collection of predictor factors determines how each tree develops. Each tree is constructed using random forests employing sample replacement. Let’s assume generating z tree with tree $k_z(x)$ the formula is shown below yields the R.F. regression model.

$$U(x) = \frac{\sum_{l=1}^l K_d(x)}{d} \tag{11}$$

The nonparametric regression method known as RF regression utilizes an ensemble to generate K outputs for every tree within a set of Z trees, $\{J_1(i), J_2(i), \dots, J_z(i)\}$ where $I = \{i_1, i_2, \dots, i_b\}$ is a vector of Y dimensions that generates an output vector of size i for each tree in the forest, $\beta_i = (i = 1, 2, \dots, J)$. Each decision tree is built using a bootstrap sample created by randomly choosing examples from the previous training database in a repetition process.

The remaining one-third of the training set forms the out-of-bag (OOB) sample. A combined amount of two-thirds of the training set are used to obtain the regression function. A regression tree is built every time through a arbitrary chosen training set taken from original database.

$$JI(k_{d(x)}) = 1 - \sum_{l=1}^n U(k_{d(x)}, n)^2 \tag{12}$$

When using unknown test data, the built-in validation features enhance the random forests' capacity to predict more accurately. The random forest prediction model's precision is evaluated based on its accurate predictions of unknown test sets.

4.3 Artificial neural networks (ANN)

To detect patterns, learn from the estimated error, and forecast outcomes in higher-dimensional space, artificial neural networks (ANNs) are complicated data processing methods that are successfully performed with great accuracy in the research area of classification, non-linear regression, and time series forecasting analysis. Employing a network of interlinked neurons generated from complicated non-linear relationships

between input variables x_l , weights function w_l , and accepted bias b_l , in the l_{th} hidden layer with a target value of y in n_{th} neuron through n_{th} hidden layer. Each neuron's estimated weights and biases are added to generate a network architecture below.

$$N = \sum_{l=1}^D W_{ln}X_l + b_l \tag{13}$$

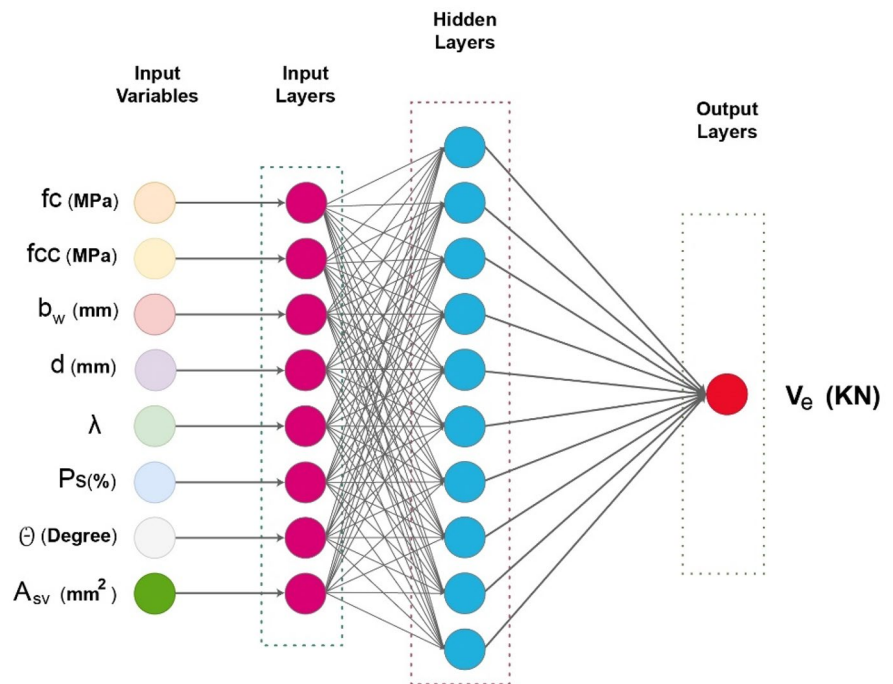
A hyperbolic tangent function was employed to improve the output. The formula's values fall within the range of 0 and 1, and its definition is as follows:

$$y = V(N) = \frac{2}{1 + e^{-2N}} - 1 \tag{14}$$

The V is an active function of the output layer, which enables the network to extract relevant data and control non-linearity to the output layer. Figure 4 shows the ANN with eight input parameters hidden layer and an output layer for predicting V_e (K.N.).

The multilayer perceptron's weights and biases are updated to quantify and minimize the error value of ANN. The process is done on each multilayer perceptron's layers until the final layer produces the output signal. The network is optimized continuously in the training process, which is carried out using various

Fig. 4 ANN architecture



optimization algorithms through repetitive training samples. The quantity of hidden layers and a total number of neuron counted in every layer are the principal hyperparameters of ANN. In our model, we implemented three hidden layers and the RELU activation function to control the input weighted sum and convert it into an output result.

4.4 Extreme gradient boosting (XGBoost)

XGBoost provide exceptional performance for regression model. The technique use "Parallel boosting" concept, that constructs potent learner by amalgamating additive training techniques. XGBoost proves invaluable for prediction, as it effectively mitigates over-fitting and enhances computational efficiency through the integration of regularization terms. By minimizing objective functions, it ensures optimal processing performance. In practical applications, it can be regarded as a soft comprehensive software library that synergistically merges distinctive algorithms with the decision tree approach, leading to enhanced estimation accuracy. Presume the same data base as shown in Eq. 6, the input data set and designed vector output which assigned for training sets.

The suggested ensemble of trees employs n adaptive functions to approximate the system's behaviour.

$$\hat{y}_l = \phi(x_l) = \sum_{n=1}^n f_n(x_l), f_n \in T \tag{15}$$

The set of regression trees is denoted by T and characterized as follows

$$C = \{f(x) = A_{c(x)}\} (t : R^n \rightarrow J, A \in R^m) \tag{16}$$

here, U denotes the tree configurations, J and A indicate leaf node quantities and corresponding values. Furthermore, the term f_n illustrates the association between A and t for the individual tree.

The OA of XGBoost can be reduced as follows,

$$y^z = \sum_{l=1}^n E \left(y_n, \hat{y}_n^{(t+1)} \right) + f_n(x_n) + Q(f_t) \tag{17}$$

Here, y_n is considered as a measured value, \hat{y}_n stands as a estimated value and E is utilized as a convex function (loss function) to assess and compare the precise and predicted values. The iteration count t is

employed for error reduction, while Q represents the retribution for the regression tree model's intricacy.

$$F = y^j + \frac{1}{2}d||r||^2 \tag{18}$$

where, r represents for the number of Leaf vector ratings, d manages regularization process, while y computes the minimum loss required to split the leaf node calculated in aforementioned equation.

5 Results and discussions

5.1 Hyper-parameters tuning

Generally, To build an effective machine learning model, it is essential to choose the optimal algorithm and carefully adjust the hyperparameters to attain the most favourable model architecture. (Chen and Guestrin 2016). Hyperparameters control the behavior and structure of the training period in ML models. A few common hyperparameters, for example, the support vector machine's penalty parameter (C), neural network's learning rate, and kernel type of Support Vector Machine are applied to optimize the algorithm designed and minimize the loss function (Awad and Khanna 2015). In this study, Random Search (R.S.) methods (Schratz et al. 2019) are employed to fine-tune the chosen hyperparameters of models. As possible hyper-parameter settings, R.S. randomly chooses a predetermined samples number from a range between the minimum and maximum values. These possibilities are subsequently trained up until the budget allocation is reached. As per R.S.'s conceptual framework, concerning the configuration space is sufficiently large, it is viable to identify global optimums or, at least, close approximations (Schratz et al. 2019). The primary benefit of R.S. is that it can quickly assign resources and processes parallelly compared with the grid search algorithm (Awad and Khanna 2015). When adjusting hyper-tuning, cross-validation is a potential rescaling approach used to regulate prediction accuracy. Out of many training and test splits set, this database was divided into training and testing group using an 70/30 split. In this study, five subsets of the data are divided into the five-fold cross-validation method, with one subset being kept evaluating the trained model after every iteration. The training data were divided into five segments, with the first segment of every iteration

is being employed for testing purpose while the remaining segments employed to train the model. The ANN, SVM, R.F., and XGBoost algorithms given in Table 3 had their best parameter settings for our database determined through fine tuning. The average of the predictions is employed to assess the model’s performance. This research employs four distinct statistical performance metrics, including mean absolute percentage error (MAPE), coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (M.A.E.).

$$R^2 = 1 - \frac{\sum_{l=1}^n (y_l - \hat{y}_l)^2}{\sum_{l=1}^n (y_l - \bar{y})^2} \tag{19}$$

$$MSE = \frac{1}{n} \sum_{l=1}^n (y_l - \hat{y}_l)^2 \tag{20}$$

$$MAE = \frac{1}{n} \sum_{l=1}^n |y_l - \hat{y}_l| \tag{21}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{l=1}^n (y_l - \hat{y}_l)^2} \tag{22}$$

where y_l is the true value (experimental value) and \hat{y} is the ML predicted value, and \bar{y} is the arithmetic mean of y values.

5.2 Machine learning (ML) model results

The utilized computational environment used in this study, involved popular programming languages like Python and reliable ML libraries commonly applied in the machine learning domain. The ML models’ findings show that among various sets of training testing split, one model that employed 70% of the data for training while 30% of data employed for testing phase produced the best regression accuracy. An in-depth examination of the model is conducted as the performance results improve through hyperparameter tuning, enhancing the overall analysis.

Table 4 displays the accuracy of ML models based on the R^2 , M.A.E., RMSE, and MAPE. When the error value is reasonably low, it is considered to have

Table 3 Fine-tuned hyperparameters of GFRP reinforced UHPC beam ML model

Models	Parameters
ANN	Batch size = 10, input neurons = 11, output neuron = 1, Number of hidden layers = 3, Number of neurons = 128, activation function = RELU, Output layer activation function = Linear, Optimizer = adam, Number of epochs = 128, learning rate = 0.001
SVM	C = 500, kernel = RBF, gamma = 0.077, Estimator = SVR, n_jobs = -1, Coef0 = 0.0, tol = 0.001, epsilon = 0.1, max_iter = -1
XGBOOST	Number of estimators = 3730, CV = 5, eta = 0.1, gamma = 900, Learning rate = 0.001, Maximum depth = 8, min child weight = 1, sub sample = 1, colsample by tree = 1, colsample by node = 1, colsample by level = 1
RF	CV = 5, Number of estimators = 50, Maximum depth = 50, Maximum features = 10, n jobs = -1. Minimum number of samples = 2, Minimum number of sample Leaf = 1, Bootstrap samples = True, Minimal cost-complexity pruning = 0

Table 4 ML model’s performance

Model	Output	Training-70% R^2	Testing-30%						
			MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
ANN	Vexp	0.9	47.4	6225.2	78.89	0.82	64.2	7068.5	84.07
SVM	Vexp	0.8	38.5	9796.6	98.97	0.76	38.5	9555.3	97.75
XGBoost	Vexp	1	13.9	312.14	17.66	0.85	57.8	5864.2	76.57
RF	Vexp	1	32.7	1993.5	44.64	0.83	68.9	6715	81.94

Bold represent the ML models’ predictions are compared to the results obtained from these building code standards. According to the findings, the XGBoost model demonstrates the highest predictive test performance among the investigated ML models

a high-performance model. The R^2 value demonstrates the ML model's ability to evaluate the experimental data. The cost function, usually known as the RMSE, is critical to how an ML system learns. Parameters of both RMSE and M.A.E. measure the precision and quality of fit.

Figure 5 displays the performance of the training dataset's ML models with estimated and experimental value. The distance of data points from the adjusted curve proved high consistency between experimental findings and calculated values across all

datasets. Among all the predicted models, XGBoost predicted most accurately ($R^2=0.99$). Other models also have high prediction accuracy, such as Random Forest has $R^2=0.95$, ANN has $R^2=0.87$, and SVM has $R^2=0.85$. Similar results are shown for the testing dataset. Figure 6 displays significant XGBoost prediction accuracy for the test dataset with $R^2=0.85$ while R.F., ANN, and SVM have an R^2 value 0.832, 0.823, and 0.762.

Additionally, in Fig. 7, the solid lines evaluate the relationship between predicted value with

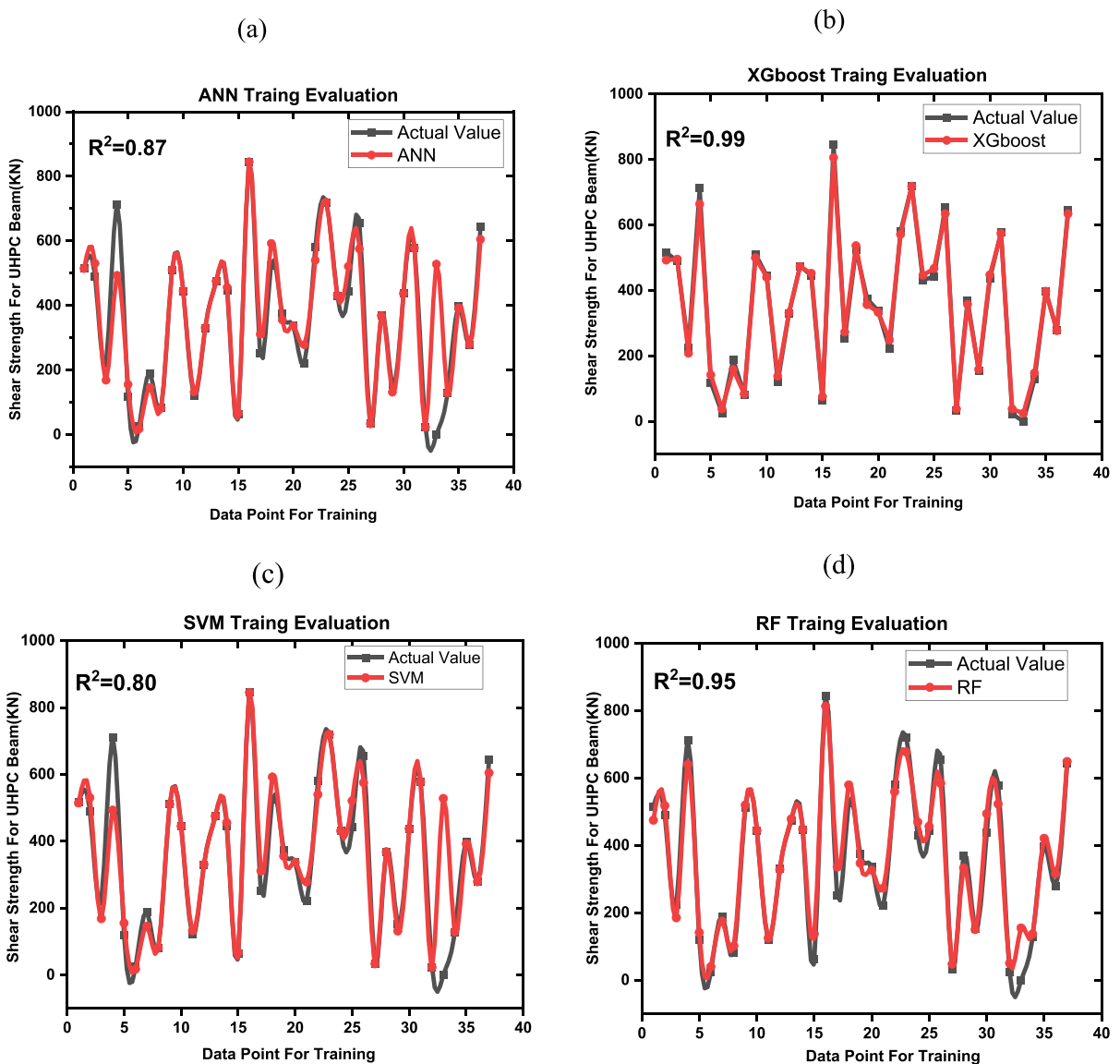


Fig. 5 Performance of the ML model from training evaluation

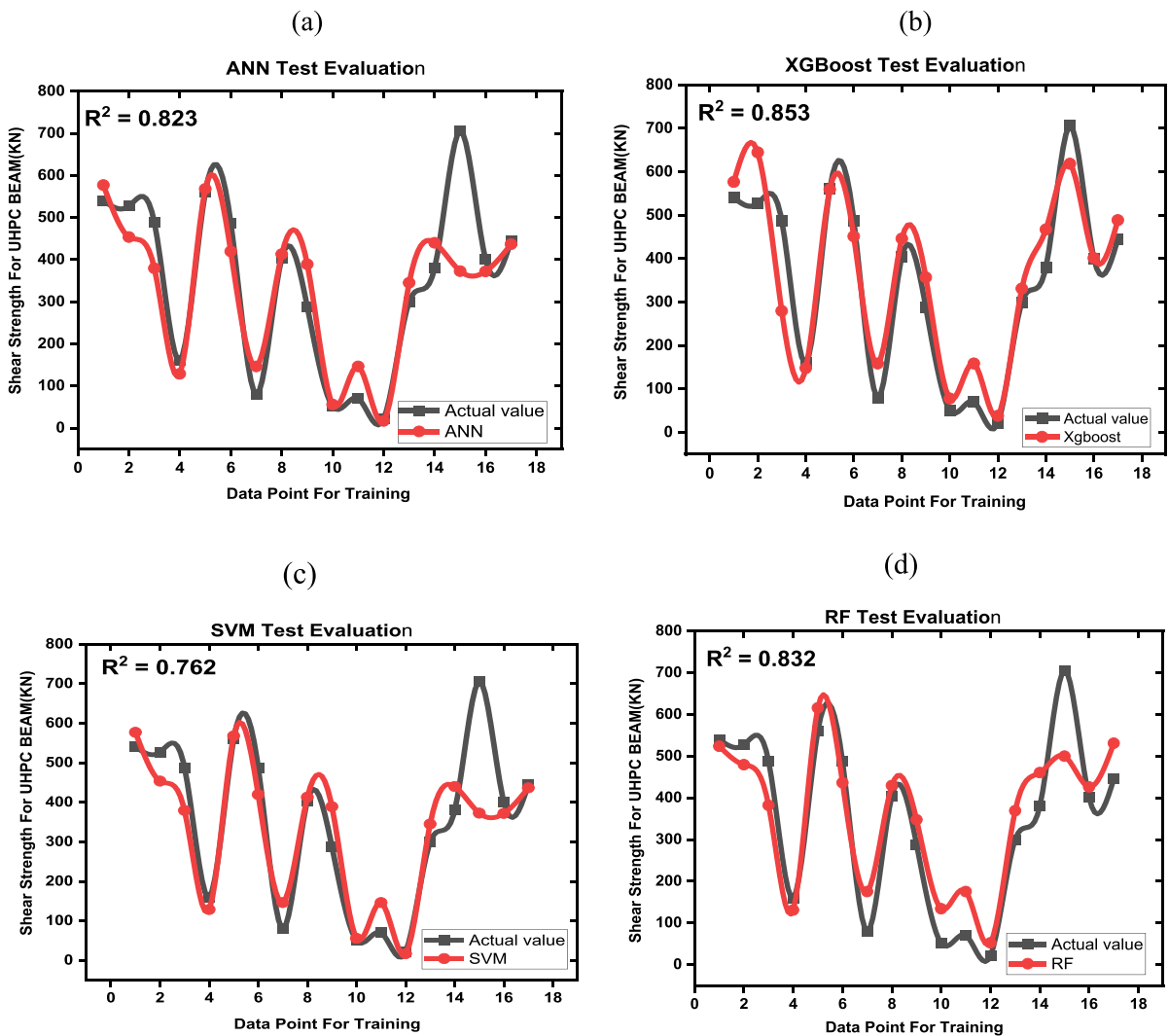


Fig. 6 Performance of the ML model from testing evaluation

the experimental value for train and test data based on suggested ML methods. The grey curve depicted the shear capacities of 95% overestimation or underestimation. These figures show a strong correlation between the predictions made by all models and the relevant experimental values (average R^2 0.81 for all models). The machine learning outcome indicated that the XGBoost model outperformed all other models in estimating shear behavior. The XGBoost model's prediction accuracy and shear capabilities of experimental data are strongly correlated, as shown by the training and test dataset's respective lower error values of MAE 13.89 KN and 57.80 KN.

Furthermore, accuracy assessment of the ML methods revealed a minimally low variation between actual strength capacity and the estimated RMSE values. The XGBoost method prediction accuracy in training and test stage are between 17.66 KN and 76.57 KN, respectively. For the SVM, ANN, and R.F. models, these values were 98.97 KN, 78.89 KN, and 44.64 KN for the training dataset and 84.07 KN, 97.75 KN, and 81.94 KN for the test dataset (Table 4). According to Fig. 7 and Table 4, the R.F. model acquired second place in predictive ability, having the second highest R^2 and the comparative

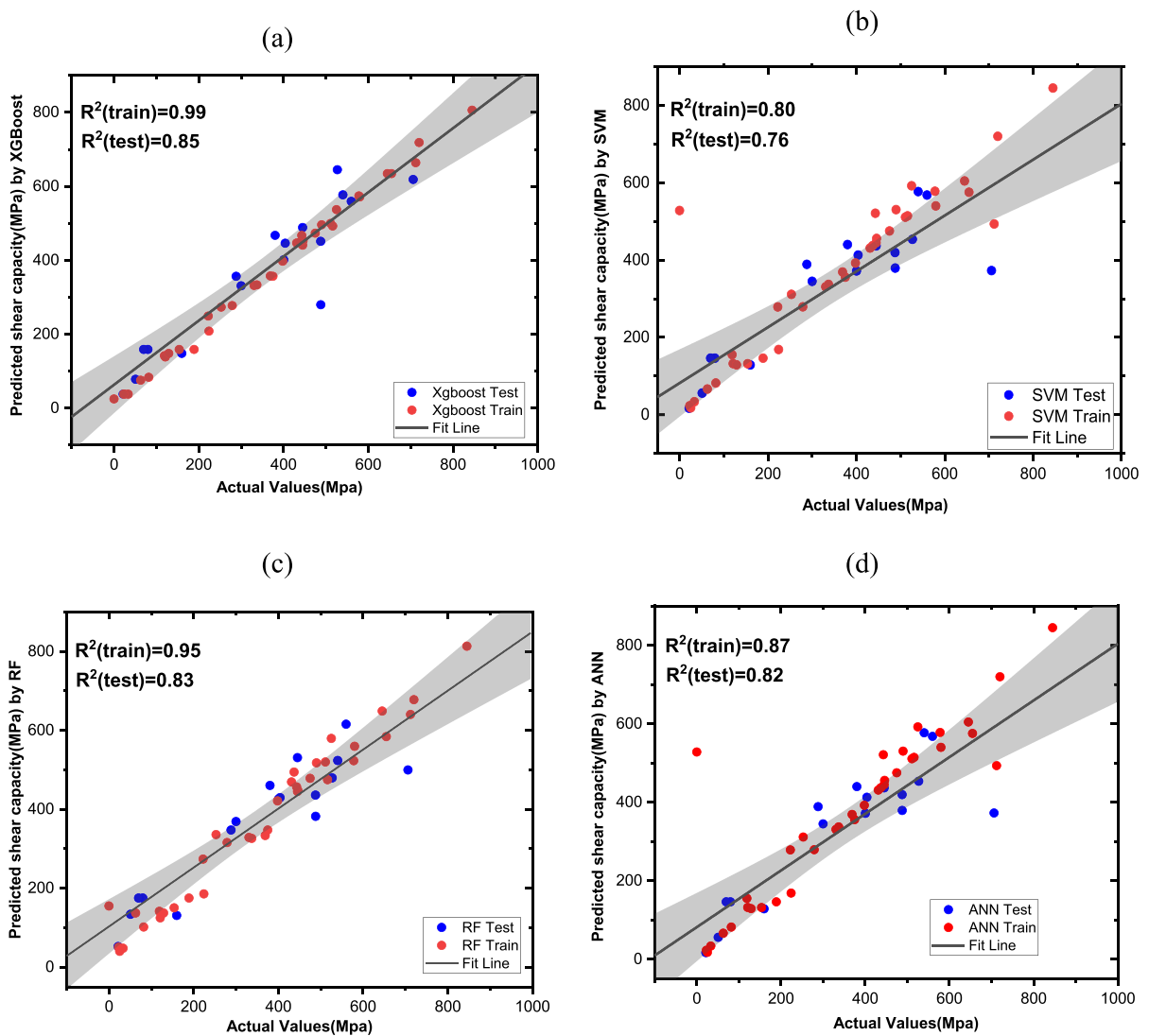


Fig. 7 Evaluation of training and testing performance based on linear regression analysis

lowest M.A.E., MAPE, and RMSE value, followed by ANN.

Figure 8 illustrates the residual assumption of the shear behavior of a glass F.R.P. I-shaped UHPC beam. The ML model’s residual dispersion is roughly the same across different models along the x-axis of λ . All machine learning models display almost the same trend and pattern for XGBoost, ANN, SVM and R.F.. The model indicates that the mean is normally allocated, without significant curvature or non-normality, showing that the regression model is not inadequate.

5.3 Standard country code resultsV

Figure 9a and b depicted the correlation between observed shear capacity and predicted shear force using the conventional formula of Eurocode2 and AFGC-2013 country codes. These outcomes are calculated in R squared value, while best-fitting data values are found in AFGC-2013 between them. The regression model is used to assess intensively the predicted shear behavior from various models and design codes listed in Table 5. Table 5 also shows that between the two design codes, ACI318M-14 has the highest prediction accuracy ($R^2=0.39$), with the

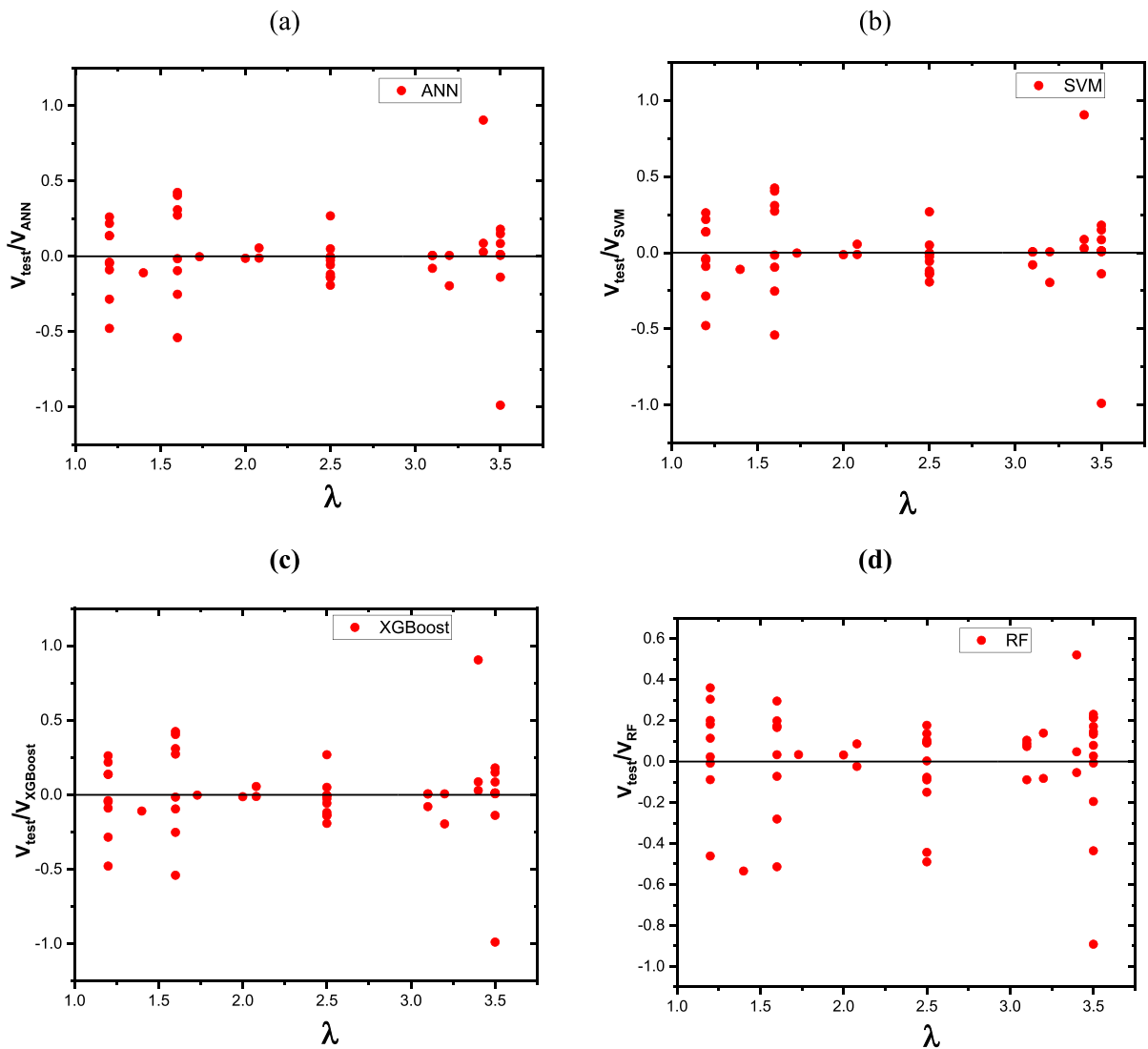


Fig. 8 Residual analysis of ML model in term of λ

lowest average of 0.486 where the lowest standard variation of 0.223. ML model results show that the average and standard variation for SVM and ANN are same, with a value of 0.996 and 0.224, respectively. Thus, it demonstrates a strong relationship between the predicted value and the suggested ML model in this study.

In this dataset, XGBoost demonstrate a mean of 0.941 and a variance of 0.915, while R.F. models have a mean and standard variation of 0.191 and 0.229. XGBoost for determining shear behavior performs better than all other models. The M.S.E.

and RMSE are also calculated in Table 5 to fulfill the robust machine learning model. Therefore, the strongest interconnection between the experimental and predicted value for training and testing phase can be seen in R^2 value. Furthermore, the future utilization of larger datasets is recommended to enhance prediction accuracy, allowing the integration of hybrid ML algorithms for improved forecasting.

Figure 9c and d illustrate the standard residual error. The data along the x-axis essentially evenly distribute the residual assumption with a consistent

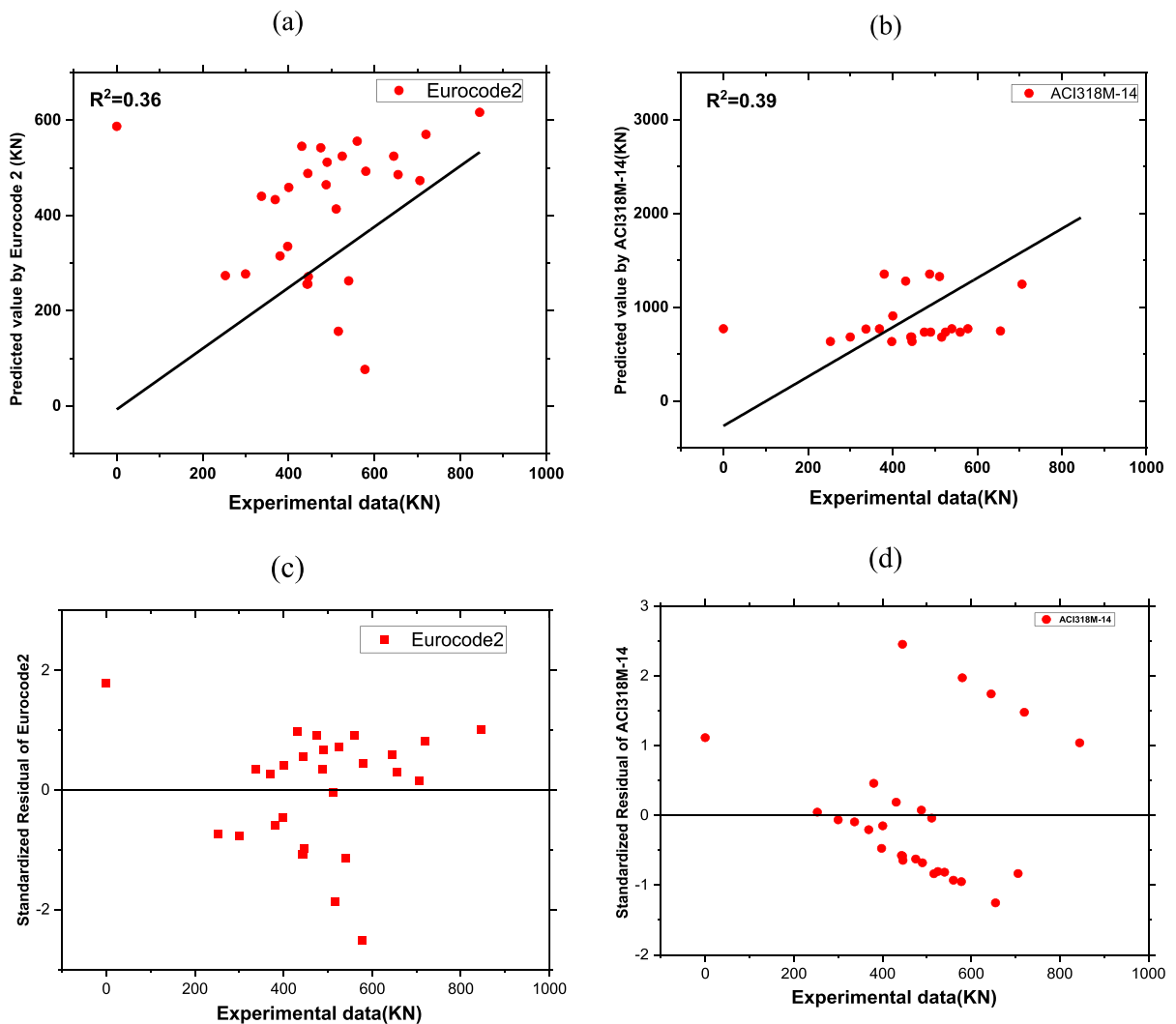


Fig. 9 Standard country code prediction results

Table 5 Comparison between standard country code and ML prediction results

Compared results	Statistical approach	
	Mean	Standard deviation
ANN	0.996	0.224
XGBoost	0.941	0.191
SVM	0.996	0.224
RF	0.915	0.229
ACI318M-14	0.486	0.223
Euro code2	1.449	1.270

variation and variability. This figure verifies the validity of the assumption model by showing that the mean is uniformly distributed and free of any systematic curvature or non-normality.

5.4 Taylor diagram

Figure 10 represents a Taylor diagram that statistically compares the four-machine learning models. The diagram presented a concise statistical evaluation of the similarity between trend parameters based on correlation, root-mean-square variance, and variation proportion. This diagram displays three popular statistical goodness of fit measurement parameter, such

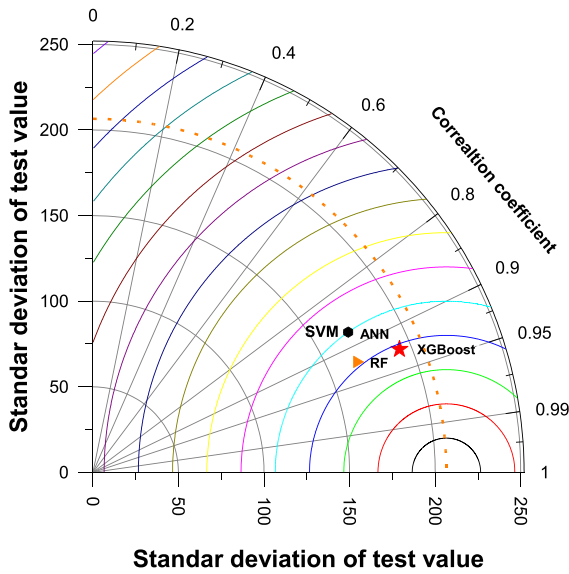


Fig. 10 Taylor Diagram presents a statistical assesment among various ML models

as R^2 , RMSE, and S.T.D to evaluate the accuracy among ML model prediction result with experimental values of shear behavior. As observed in the figure, all machine learning models have a higher rate of prediction accuracy (higher R^2 value) along with lower RMSE, and XGBoost shows the best prediction accuracy among all models with the highest correlation value (0.928) and least RMSE (74.627) value. The ANN demonstrated the lowest prediction accuracy in a correlation coefficient around 0.877, identical to SVM with the same high RMSE (96.461) value. The R.F. model showed satisfactory prediction accuracy while maintaining a position between the best and least accurate model. However, with considerable accuracy, all four ML models can generally predict the shear behavior of glass F.R.P. bars-reinforced ultra-high-performance concrete I-shaped beams.

5.5 SHAP analysis

The Shapley Additive explanations (SHAP) algorithm is used to determine the relative importance of the factors that affect the shear strength behavior of UHPC computed by XGBoost. As a result, it is possible to statistically analyze the input–output correlations that are inherently hidden in the conventional machine learning model. According to the analysis

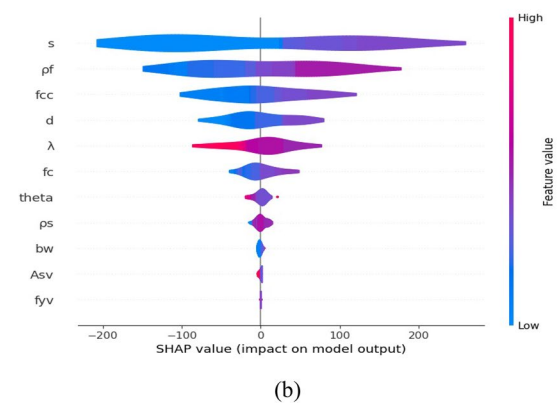
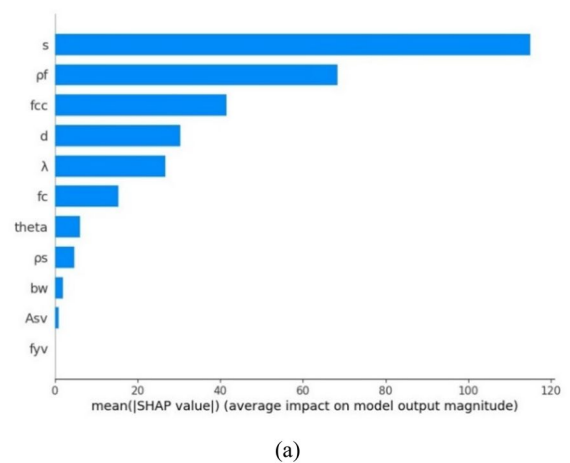


Fig. 11 SHAP summary plot and parameter significance distribution results

of this study, the XGBoost model has shown higher accuracy than other ML models. Therefore, the SHAP value may be used to evaluate the XGBoost predictions in various ways. It shows the significance of each parameter and quantifies how each parameter affects each-others for determining model prediction accuracy. Figure 11 provides the input parameters' feature importance, which illustrates each feature's overall influence on the model predictions.

It can be comprehended from Fig. 11a that Stirrup Spacing (S) significantly influences the shear behavior of glass FRP bars reinforced UHPC, while they f_{yv} are of the minor importance. f_{cc} also affects the shear behavior but the significance is much less than ρ_f and slightly higher than d . Among all other parameters, λ , f_c and θ has an inadequate amount of influence on output results, and other remaining variables have very poor effects as well. A SHAP summary

plot of the features is shown in Fig. 11b through a beea swarm plot. The distribution plot demonstrated the SHAP values for every feature with underlying influence patterns to demonstrate density and dots piled up along every feature sequence. The figure shows the individual SHAP value on the x-axis and the input parameters ranked in order of importance on the y-axis. The database's outcome results are represented in a violin plot, and the color warm indicates the rising or decreasing of a particular value in input features, and values range from small to large in colors blue to red. The plot demonstrates that, on average, stirrup spacing (s) is the most significant feature. Consequently, the summary plot provides information about which feature is highly essential and how one feature affects another, determining shear behavior. The general observation from the figure shows that the shear strength of glass fiber I-shaped UHPC will rise with an increase in the features of ρ_f , f_{cc} and d . On the other hand, shear strength leans to decline as parameters for features like λ and θ increase in values.

6 Conclusions

This research's main driving force and motivation is utilizing the soft computing method to predict the shear behavior of UHPC I-shaped glass bars. Based on experimental data from eight pieces of literature, this study made a machine-learning model to predict the shear strength and performance behavior of UHPC I-shaped beams and compared the ML results to traditional country code results. This study looked at how well four machine learning models, such as ANN, XGBoost, and SVM, could predict the shear strength of a glass FRP UHPC I-shaped beam. The results show that the integrated machine learning approach is very good at predicting UHPC shear behavior, and we conclude that ANN, R.F., SVM, and XGBoost can all be used to figure out how an R.C. beam is sheared. The results of the study led to the following conclusions.

1. During the training phase, the ideal hyperparameters of ML-based models were chosen using a hybrid search with fivefold cross-validation. Comparing with experimental results revealed that the four ML-based approaches with input features and appropriate hyper-parameters can predict the shear behavior of a glass FRP-reinforced UHPC I-shaped beam.
2. XGBoost was the most reliable and efficient model with an $R^2=0.99$, $RMSE=17.66$ KN, and $M.A.E.=13.89$ KN among all other machine learning models used in this study.
3. In addition, this study incorporates current design criteria and empirical methods, which were developed to evaluate the shear behavior of glass FRP UHPC I-shaped beams. Consequently, the machine learning model can predict the actual strength with greater accuracy and reliability than conventional design code models such as Eurocode2 and ACI318M-14.
4. The SHAP algorithm was implemented to show the feature dependency of ML results and indicate feature importance in shapely values. It shows that stirrup spacing(s) is the most significant feature among all input parameters.
5. The findings of this study indicated that all of the essential design factors that affect the shear behavior of glass F.R.P. bars-reinforced UHPC I-shaped beams should be considered to improve the empirical model and standard equations' ability to estimate the shear behavior.
6. UHPC is a relatively new material with limited tests done so far; therefore, future limitations and challenges will need further study.

Acknowledgements This paper was supported by the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) (Grant No. 2019QZKK0902), and National Natural Science Foundation of China (Grant No. 42077275). It was also supported by Youth Innovation Promotion Association of the Chinese Academy of Sciences (2018405).

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The datasets analysed during the current study are not publicly available as of now as they are still being used for subsequent research but are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits

use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abuodeh, O.R., Abdalla, J.A., Hawileh, R.A.: Assessment of compressive strength of ultra-high performance concrete using deep machine learning techniques. *Appl. Soft Comput.* **95**, 106552 (2020)
- Akbar, M., Hussain, Z., Huali, Z., Imran, M., Thomas, B.S.: Impact of waste crumb rubber on concrete performance incorporating silica fume and fly ash to make a sustainable low carbon concrete. *Struct. Eng. Mech.* **85**, 275–287 (2023). <https://doi.org/10.12989/sem.2023.85.2.275>
- An efficient stochastic-based coupled model for damage identification in plate structures. <https://doi.org/10.1016/j.engfailanal.2021.105866>
- Arciszewski, T., et al.: Machine learning in transportation engineering: a feasibility study. *Appl. Artif. Intell. Int. J.* **8**(1), 109–124 (1994)
- Ashour, S.A., Hasanain, G.S., Wafa, F.F.: Shear behavior of high-strength fiber reinforced concrete beams. *Struct. J.* **89**(2), 176–184 (1992)
- Audet, L.-A., Desmarais, M., Gosselin, É.: Handling missing data through prevention strategies in self-administered questionnaires: a discussion paper. *Nurse Res.* **30**(2), 1 (2022)
- Awad, M., Khanna, R.: Support vector regression. In: *Efficient Learning Machines*, pp. 67–80. Springer, Cham (2015)
- Aydin, Z.E., Ozturk, Z.K.: Performance analysis of XGBoost classifier with missing data. *Manch. J. Artif. Intell. Appl. Sci. (MJAIAS)* **2**(02), 2021 (2021)
- Bahij, S., et al.: Numerical investigation of the shear behavior of reinforced ultra-high-performance concrete beams. *Struct. Concr.* **19**(1), 305–317 (2018)
- Chakraborty, H., Gu, H.: *A Mixed Model Approach for Intent-to-Treat Analysis in Longitudinal Clinical Trials with Missing Values*. RTI Press, Triangle Park (2019)
- Chen, X., et al.: Experimental studies and microstructure analysis for ultra high-performance reactive powder concrete. *Constr. Build. Mater.* **229**, 116924 (2019)
- Chen, T., Guestrin, C.: Xgboost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining* (2016)
- Chengcong, C., Qing, W., Pinggen, S., Hao, Z., Muhammad, A., Shiliang, M.: Study on diffusion of oxygen in coral concrete under different preloads. *Constr. Build. Mater.* **319**, 126147 (2022)
- Dai, Y., et al.: A novel machine learning model to predict the moment capacity of cold-formed steel channel beams with edge-stiffened and un-stiffened web holes. *J. Build. Eng.* **53**, 104592 (2022)
- Dang, B.-L., Nguyen-Xuan, H., Wahab, M.A.: An effective approach for VARANS-VOF modelling interactions of wave and perforated breakwater using gradient boosting decision tree algorithm. *Ocean Eng.* (2023). <https://doi.org/10.1016/j.oceaneng.2022.113398>
- Elsayed, M., et al.: Shear behaviour of ultra-high performance concrete beams with openings. In: *Structures*. Elsevier, Amsterdam (2022)
- Emmanuel, T., et al.: A survey on missing data in machine learning. *J. Big Data* **8**(1), 1–37 (2021)
- Enders, C.K.: *Applied Missing Data Analysis*. Guilford Publications, New York (2022)
- Farouk, A.I.B., Jinsong, Z.: Prediction of interface bond strength between ultra-high-performance concrete (UHPC) and normal strength concrete (N.S.C) using a machine learning approach. *Arab. J. Sci. Eng.* **47**(4), 5337–5363 (2022)
- Farouk, A.I.B., et al.: Prediction and uncertainty quantification of ultimate bond strength between UHPC and reinforcing steel bar using a hybrid machine learning approach. *Constr. Build. Mater.* **345**, 128360 (2022)
- Fei, Y., Wu, W., Ma, H., Yin, A.M.: Study on anti-corrosion performance of composite modified magnesium ammonium phosphate cement-based coatings. *J. Build. Eng.* **71**, 106423 (2023). <https://doi.org/10.1016/j.job.2023.106423>
- Fu, B., Feng, D.-C.: A machine learning-based time-dependent shear strength model for corroded reinforced concrete beams. *J. Build. Eng.* **36**, 102118 (2021)
- Ghafari, E., et al.: Design of UHPC using artificial neural networks. In: *Brittle Matrix Composites 10*, pp. 61–69. Elsevier, Amsterdam (2012)
- Graybeal, B.A.: *Structural behavior of ultra-high performance concrete prestressed I-girders*. United States. Federal Highway Administration. Office of Infrastructure, Washington, DC (2006)
- Graybeal, B., et al.: International perspective on UHPC in bridge engineering. *J. Bridge Eng.* **25**(11), 04020094 (2020)
- Guo, P., et al.: Predicting mechanical properties of high-performance fiber-reinforced cementitious composites by integrating micromechanics and machine learning. *Materials* **14**(12), 3143 (2021)
- Hu, T., Li, G.: Machine learning-based model in predicting the plate-end debonding of FRP-strengthened RC beams in flexure. *Adv. Civil Eng.* (2022). <https://doi.org/10.1155/2022/6069871>
- Jiang, C.-S., Liang, G.-Q.: Modeling shear strength of medium-to ultra-high-strength concrete beams with stirrups using S.V.R. and genetic algorithm. *Soft Comput.* **25**(16), 10661–10675 (2021)
- Jin, L.-Z., et al.: Shear strength of fibre-reinforced reactive powder concrete I-shaped beam without stirrups. *Mag. Concr. Res.* **72**(21), 1112–1124 (2020)

- Lee, C., et al.: Shear strength of ultra high performance fiber reinforced concrete (UHPFRC) precast bridge joint. In: High Performance Fiber Reinforced Cement Composites 6, pp. 413–420. Springer, Cham (2012)
- Li, Z., Gao, X., Lu, D.: Correlation analysis and statistical assessment of early hydration characteristics and compressive strength for multi-composite cement paste. *Constr. Build. Mater.* **310**, 125260 (2021)
- Li, Z., et al.: Estimation of bond strength between UHPC and reinforcing bars using machine learning approaches. *Eng. Struct.* **262**, 114311 (2022)
- Liao, J., et al.: Design-oriented stress-strain model for FRP-confined ultra-high performance concrete (UHPC). *Constr. Build. Mater.* **318**, 126200 (2022)
- Little, R.J., Rubin, D.B.: *Statistical Analysis with Missing Data*, vol. 793. Wiley, Hoboken (2019)
- Liu, Y., et al.: Development of ultra-high performance geopolymer concrete (UHPGC): influence of steel fiber on mechanical properties. *Cement Concr. Compos.* **112**, 103670 (2020)
- Liu, T., et al.: Machine-learning-based models to predict shear transfer strength of concrete joints. *Eng. Struct.* **249**, 113253 (2021)
- Ly, H.-B., et al.: Computational hybrid machine learning based prediction of shear capacity for steel fiber reinforced concrete beams. *Sustainability* **12**(7), 2709 (2020)
- Marani, A., Jamali, A., Nehdi, M.L.: Predicting ultra-high-performance concrete compressive strength using tabular generative adversarial networks. *Materials* **13**(21), 4757 (2020)
- Melhem, H.G., Cheng, Y.: Prediction of remaining service life of bridge decks using machine learning. *J. Comput. Civil Eng.* **17**(1), 1–9 (2003)
- Memon, S.M., Wamala, R., Kabano, I.H.: Missing data analysis using statistical and machine learning methods in facility-based maternal health records. *SN Comput. Sci.* **3**(5), 1–15 (2022)
- Mészöly, T., Randl, N.: Shear behavior of fiber-reinforced ultra-high performance concrete beams. *Eng. Struct.* **168**, 119–127 (2018)
- Nazabal, A., et al.: Handling incomplete heterogeneous data using vaes. *Pattern Recogn.* **107**, 107501 (2020)
- Nematollahi, B., et al.: A review on ultra high performance 'ductile' concrete (UHPdC) technology. *Int. J. Civil Struct. Eng.* **2**(3), 994 (2012)
- Nguyen, D.H., Wahab, M.A.: Damage detection in slab structures based on two-dimensional curvature mode shape method and Faster R-CNN. *Adv. Eng. Softw.* (2023). <https://doi.org/10.1016/j.advengsoft.2022.103371>
- Nielsen, M.P., Hoang, L.C.: *Limit Analysis and Concrete Plasticity*. CRC Press, New York (2016)
- Qi, J.N., et al.: Post-cracking shear strength and deformability of HSS-UHPFRC beams. *Struct. Concr.* **17**(6), 1033–1046 (2016)
- Raheem, A.H.A., Mahdy, M., Mashaly, A.A.: Mechanical and fracture mechanics properties of ultra-high-performance concrete. *Constr. Build. Mater.* **213**, 561–566 (2019)
- Rahman, J., et al.: Data-driven shear strength prediction of steel fiber reinforced concrete beams using machine learning approach. *Eng. Struct.* **233**, 111743 (2021)
- Resplendino, J.: Ultra high performance concrete: new AFGC recommendations. In: *Designing and Building with UHP-FRC*, pp. 713–722. Wiley, Hoboken (2011)
- Salem, N.M., Deifalla, A.: Evaluation of the strength of slab-column connections with F.R.P.s using machine learning algorithms. *Polymers* **14**(8), 1517 (2022)
- Sarothi, S.Z., et al.: Predicting bearing capacity of double shear bolted connections using machine learning. *Eng. Struct.* **251**, 113497 (2022)
- Schmidt, M., Fehling, E.: Ultra-high-performance concrete: research, development and application in Europe. *ACI Spec. Publ.* **228**(1), 51–78 (2005)
- Schratz, P., et al.: Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecol. Model.* **406**, 109–120 (2019)
- Silva, M.A., Rodrigues, C.C.: Size and relative stiffness effects on compressive failure of concrete columns wrapped with glass F.R.P. *J. Mater. Civil Eng.* **18**(3), 334–342 (2006)
- Simwanda, L., et al.: Structural reliability of ultra high-performance fibre reinforced concrete beams in shear. *Struct. Concr.* **24**(2), 2862–2878 (2022)
- Somer, E., Gische, C., Miočević, M.: Methods for modeling autocorrelation and handling missing data in mediation analysis in single case experimental designs (SCEDs). *Eval. Health Prof.* **45**(1), 36–53 (2022)
- Su, M., et al.: Identification of the interfacial cohesive law parameters of FRP strips externally bonded to concrete using machine learning techniques. *Eng. Fract. Mech.* **247**, 107643 (2021)
- Sun, J., et al.: Machine-learning-aided prediction of flexural strength and A.S.R expansion for waste glass cementitious composite. *Appl. Sci.* **11**(15), 6686 (2021)
- Tian, H., et al.: Experimental investigation on axial compressive behavior of ultra-high performance concrete (UHPC) filled glass F.R.P. tubes. *Constr. Build. Mater.* **225**, 678–691 (2019)
- Tran, V.-T., Nguyen, T.-K., Nguyen-Xuan, H., Wahab, M.A.: Vibration and buckling optimization of functionally graded porous microplates using BCMO-ANN algorithm. *Thin-Walled Struct.* (2023). <https://doi.org/10.1016/j.tws.2022.110267>
- Uchida, Y. et al.: Outlines of 'recommendations for design and construction of ultra high strength fiber reinforced concrete structures' by JSCE. In: *Proceedings Intnetnational Workshop on High Performance Fiber Reinforced Cementitious Composites in Structural Applications*, Citeseer (2005)
- Umar, T., Yousaf, M., Akbar, M., Abbas, N.: An experimental study on non-destructive evaluation of the mechanical characteristics of a sustainable concrete incorporating industrial waste. *Materials* **15**, 7346 (2022). <https://doi.org/10.3390/ma15207346>
- Wakjira, T.G., et al.: Explainable machine learning model and reliability analysis for flexural capacity prediction of R.C. beams strengthened in flexure with FRCM. *Eng. Struct.* **255**, 113903 (2022a)
- Wakjira, T.G., et al.: Shear capacity prediction of FRP-RC beams using single and ensemble explainable machine learning models. *Compos. Struct.* **287**, 115381 (2022b)

- Wang, Z., et al.: A machine-learning-based model for predicting the effective stiffness of precast concrete columns. *Eng. Struct.* **260**, 114224 (2022)
- Wei, R., et al.: Missing value imputation approach for mass spectrometry-based metabolomics data. *Sci. Rep.* **8**(1), 1–10 (2018)
- Wu, D., Zhang, H., Yang, Y.: Deep learning-based crack monitoring for ultra-high performance concrete (UHPC). *J. Adv. Transp.* (2022). <https://doi.org/10.1155/2022/4117957>
- Xu, H., et al.: Supervised machine learning techniques to the prediction of tunnel boring machine penetration rate. *Appl. Sc.* **9**(18), 3715 (2019)
- Xue, J., et al.: Review of ultra-high performance concrete and its application in bridge engineering. *Constr. Build. Mater.* **260**, 119844 (2020)
- Yang, T., et al.: Experimental and numerical investigation of bond behavior between geopolymer based ultra-high-performance concrete and steel bars. *Constr. Build. Mater.* **345**, 128220 (2022)
- Yavas, A., Goker, C.O.: Impact of reinforcement ratio on shear behavior of I-shaped UHPC beams with and without fiber shear reinforcement. *Materials (basel)* **13**(7), 1525 (2020a)
- Yavas, A., Goker, C.O.: Impact of reinforcement ratio on shear behavior of I-shaped UHPC beams with and without fiber shear reinforcement. *Materials* **13**(7), 1525 (2020b)
- Yoo, D.-Y., Banthia, N.: Mechanical properties of ultra-high-performance fiber-reinforced concrete: a review. *Cement Concr. Compos.* **73**, 267–280 (2016)
- Yu, R., Spiesz, P., Brouwers, H.J.H.: Mix design and properties assessment of ultra-high performance fibre reinforced concrete (UHPFRC). *Cement Concr. Res.* **56**, 29–39 (2014)
- Zhang, J., et al.: Machine-learning-assisted shear strength prediction of reinforced concrete beams with and without stirrups. *Eng. Comput.* (2020). <https://doi.org/10.1007/s00366-020-01076-x>

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