

Article

Not peer-reviewed version

Prediction of Flexural Ultimate Capacity for Reinforced UHPC Beams Using Ensemble Learning and SHAP Method

[Zhe Zhang](#) , [Xuemei Zhou](#) , [Ping Zhu](#) * , [Zhaochao Li](#) , [Yichuan Wang](#)

Posted Date: 22 February 2025

doi: [10.20944/preprints202502.1747.v1](https://doi.org/10.20944/preprints202502.1747.v1)

Keywords: Performance Prediction; Flexural Ultimate Capacity; Reinforced UHPC Beam; Machine Learning (ML); Ensemble Learning (EL); SHAP



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Prediction of Flexural Ultimate Capacity for reinforced UHPC Beams Using Ensemble Learning and SHAP Method

Zhe Zhang ¹, Xuemei Zhou ¹, Zhu Ping ^{2,3,*}, Zhaochao Li ¹ and Yichuan Wang ¹

¹ School of Civil Engineering, Hunan University of Technology, Zhuzhou 412007, China; zhangzhe@hut.edu.cn (Z.Z.)

² Research Institute of Hunan University in Chongqing, Chongqing 401135, China

³ National Key Laboratory of Bridge Safety and Resilience, College of Civil Engineering, Hunan University, Changsha 410082, China

* Correspondence: pingzhu@hnu.edu.cn

Abstract: Ensemble learning (EL) models are designed to enhance the accuracy and efficiency in predicting the flexural ultimate capacity of reinforced ultra-high-performance concrete (UHPC) beams with aim to provide more reliable and efficient design experience for structural applications in this study. For model training and testing, a comprehensive database is initially established for flexural ultimate capacity of reinforced UHPC beams, composed of 339 UHPC-based specimens with varying design parameters compiled from 56 published experimental investigations. Furthermore, multiple machine learning (ML) algorithms, including both traditional and EL models, are employed to develop optimized predictive models for the flexural ultimate capacity of reinforced UHPC specimens derived from the established database. Four statistical indicators of model performance are utilized to assess the accuracies of the prediction results with ML models used. Subsequently, a highly efficient evaluation of ML models is taken by analyzing the sensitivity of ML models to varying data subsets. Finally, a Shapley Additive Explanations (SHAP) method is employed to interpret several EL models, thereby substantiating their reliability and determining the extent of influence exerted by each feature on the prediction results. The present ML models predict accurately the flexural ultimate capacity M_u of reinforced UHPC beams after optimization, with EL models providing a higher level of accuracy than the traditional ML models. The present study also underscores the significant impact of the database division ratios of training-to-testing set on the effectiveness of performance prediction for the ML models. The optimal model functionality may be accomplished by properly considering the effects of database subset distribution on the performance prediction and model stability. The CatBoost model demonstrates superior performance in terms of predictive accuracy, as evidenced by its highest R^2 value, the lowest RMSE value, lowest MAE value, and lowest MAPE value. This substantial improvement in performance prediction of flexural capacity for reinforced UHPC beams is notable when compared to existing empirical methods. The CatBoost model displays a more uniform distribution of SHAP values for all parameters, suggesting a balanced decision-making process and contributing to its superior and stable model performance. The current study identifies a significant positive relationship between the increases in height, and reinforcement ratio of steel rebars and the growth in normalized SHAP values. These findings contribute to a deeper understanding of the role played by each feature in the prediction of flexural ultimate capacity of reinforced UHPC beams, thereby providing a foundation for more accurate model optimization and for a more refined feature section strategy.

Keywords: performance prediction; flexural ultimate capacity; reinforced UHPC beam; machine learning (ML); ensemble learning (EL); SHAP

1. Introduction

Ultra-High-Performance Concrete (UHPC) has emerged as a significant and innovative construction material in mid-1990s [1]. Richard and Cheyrezy made a substantial contribution to the development of UHPC through their creation of Reactive Powder Concrete (RPC), an exceptionally enhanced material that represents a major advance in the evolution of construction material [2]. With an ever-increasing research, UHPC has led to its widespread application globally, particularly in the construction of bridges, infrastructure, and other critical structures [3–5]. Compared with conventional concrete, UHPC exhibits ultra-high compressive strength (commonly > 120 MPa), post-cracking strength (typically > 5 MPa), and remarkable durability. These outstanding properties are primarily attributed to a low water-to-binder ratio (usually < 0.2), a high fineness of supplementary materials, a discontinuous pore structure, and a high-volume fraction of high-strength steel fibers [6–8]. Numerous studies have shown that it is crucial to understand the mechanical responses of UHPC under different loading conditions for further investigation into its structural performance [9], especially for the UHPC with superior bending and tensile properties [10]. Graybeal et al. [11] developed a direct tension testing method, which facilitates a more comprehensive understanding of the tensile response of UHPC and its implications for structural design. A comprehensive literature review reveals that extensive research has been conducted on the flexural behaviors of reinforced UHPC beams. This research has investigated a multitude of variables, including specimen size [12,13], compressive strength of UHPC [14–16], reinforcement ratios of steel rebars [6,12,17–19], and the types/shapes and volume fractions of steel fibers [6,15,16,20]. The findings of these studies have considerably advanced both the design optimization and structural application of UHPC beams [6,12,15–21].

However, existing research regarding the flexural performance of reinforced UHPC elements is frequently based on a limited number of specimens considering a narrow range of parameter variables. Consequently, it is time-consuming and labor-intensive, and the conclusions obtained may be overestimated or insufficient to comprehensively describe the exact influence of these parameters. Furthermore, design codes and structural standards for UHPC beams remain relatively limited [22,23], despite some analytical methods and finite element models having been presented on the basis of some simplification assumptions with certain limitations. Therefore, additional experimental and analytical investigations are required to develop an efficient and energy-saving method to predict the flexural properties of UHPC-based structural elements [13,24–26].

In recent years, machine learning (ML) has emerged as a powerful and versatile tool with a wide range of applications within the field of civil engineering, particularly in the context of predicting the performance of advanced building materials such as UHPC. The application of ML provides an effective and robust platform for predicting structural response of UHPC-based elements, thereby significantly reducing the time and effort required for experimentation and modeling [27]. Numerous studies have demonstrated that ML methods have been employed to predict basic properties of UHPC such as compressive strength, flexural strength, workability, and shrinkage performance, as well as to forecast the interface bonding strength and thus develop interpretable models that optimize UHPC mix designs [28–31]. Particularly, ML techniques offer significant advantages in predicting diverse properties of UHPC under various loading conditions, which include compressive strength, flexural strength, ultimate capacity, and fracture characteristics. Moreover, ML techniques have been utilized to determine the structural performance of reinforced concrete or UHPC beams [30–33]. A gradient boosting regression tree (GBRT) was used by Fu and Feng [29] to forecast the residual shear strength of corroded reinforced concrete beams at different service periods. Feng et al. [32] applied an ensemble learning method to predict the shear strength of reinforced concrete deep beams, demonstrating that the ensemble ML models outperformed traditional mechanics-based models in terms of improved prediction accuracy and reduced bias. Similarly, a variety of ML algorithms, including support vector machines (SVM), artificial neural networks (ANN) and ensemble learning (EL) methods, have also been used to identify failure modes and predict shear capacity of UHPC beams under combined bending and shear forces, achieving a high prediction accuracy [31–33].

Despite the presence of application of ML methods to predict the shear performance of UHPC beams, there is a growing interest in employing ML technologies to accurately and efficiently predict the flexural behavior of reinforced UHPC beams. However, this remains an emerging area of research, with few published studies exploring its application. Solhmirzaei et al. [33] used support vector regression (SVR) and genetic programming to predict the flexural capacity of UHPC beams with varying cross-sectional dimensions and material properties. Ergen and Katlav [34] explored the potential of deep learning (DL) models for predicting the flexural capacity of UHPC beams with and without steel fibers. Nevertheless, the effectiveness of ML models is largely contingent upon the quality of database acquisition, frequently challenged by the selection of input variables. It is therefore important to expand and optimize the database of reinforced UHPC beam specimens. Specifically, the selection of input parameters for the optimized database should be both comprehensive and concise to achieve more reliable and convenient predictions of the bending performance of reinforced UHPC beams. The employment of excessive input parameters is impractical for real-world design applications, while the inclusion of interrelated input parameters unnecessarily inflates the input features without adding a unique or distinct value to the ML model. Further, the versatility of ML algorithms has been shown to result in notable discrepancies in both the accuracy and efficiency of the performance predictions of reinforced UHPC specimens. A comprehensive assessment and comparison of the accuracy and efficiency of various ML models for predicting the flexural ultimate capacity of reinforced UHPC beams is a crucial gap in the current research. Moreover, in the context of machine learning, the division of the original database into training and testing sets represents a fundamental stage in the data processing. The extent to which the training set is divided affects the performance of the ML model in terms of both the accuracy of training and its capacity to generalize to new data. The optimal division ratio of training set to testing set depends on the subset size and characteristics of the database. It is necessary to analyze the model performance on varying subsets of data to ensure an efficient evaluation of both the model and the data quality. Besides statistical evaluations of ML techniques, an adequate discussion is required regarding the physical and structural principles governing reinforced UHPC beams. For practical engineering applications, the comparison of EL models with physical principles alongside statistical model evaluations is essential. Previous research has shown that the Categorical Gradient Boosting (CatBoost) model has excellent predictive stability and generalization ability [35]. To evaluate the accuracy and reliability of EL methods, the CatBoost method is exemplified and compared with existing empirical methods and design standards [36–40]. Additionally, it could be reasonably argued that the differences between various ML algorithms can significantly affect the reliability of parameter analysis in model interpretation. Therefore, it would appear prudent to undertake further research into a comparison of different ML models. SHAP (Shapley Additive Explanations) offers a promising method to clarify the contributions of features to predictions and has been widely used for model interpretation [22,30,35,41]. An in-depth analysis using SHAP method should be carried out by taking account into the impact of key design parameters on performance prediction to provide valuable insights for structural design purposes. Therefore, the prediction of flexural ultimate capacity for reinforced UHPC beams using ensemble learning and SHAP methods is promising.

The objective of this study is to address aforementioned limitations by expanding the database and optimizing ML algorithms, thereby achieving greater accuracy and efficiency in predicting the flexural performance of reinforced UHPC beams and providing more reliable and efficient design recommendations for future applications. To be more specific, a comprehensive database containing 339 testing data of reinforced UHPC beams with various design parameters is initially established. To balance model accuracy and practical implementation, a reliable and efficient approach involving 9 input parameters is considered in this study. Furthermore, several ML algorithms are presented to develop optimized models for precisely predicting the flexural ultimate capacity (M_u) of reinforced UHPC specimens derived from the established database. Traditional models, including ANN, SVR, and K-Nearest Neighbors (K-NN), are first applied to make predictions. Additionally, ensemble learning models, such as Classification and Regression Trees (CART), Random Forest (RF), Adaptive

Boosting (AdaBoost), and Gradient Boosting Regression Trees (GBRT), are utilized for further optimization. To enhance prediction accuracy, advanced models like Light Gradient Boosting Machine (LightGBM), CatBoost, and Extreme Gradient Boosting (XGBoost) are also employed. The performance of ML models used is then evaluated using four statistical indicators to comprehensively assess and compare their prediction accuracies and capabilities for the flexural ultimate capacity of reinforced UHPC specimens. Subsequently, the sensitivity of ML models to varying data subsets is analyzed to ensure a highly efficient evaluation of ML models used and the established database. Moreover, the CatBoost model is exemplified to compare the predictions with several existing empirical formulas alongside statistical evaluations for practical engineering applications. Finally, the SHAP method is employed to interpret multiple EL models, thereby substantiating their reliability and determining the extent of influence exerted by each feature on the prediction results of the flexural capacity of reinforced UHPC beams.

2. Acquisition of the Database

The establishment of a database represents a fundamental stage in the initial process of machine learning, which involves the collection, organization and cleansing of data for model training. By conducting a comprehensive review of the published literature, an ultimate capacity database of reinforced UHPC specimens under bending loads is developed and summarized by integrating test results from diverse experimental studies in the present study (see **Table 1**). The database comprises measured results of 339 UHPC-related specimens with varying design parameters sourced from 56 different experimental investigations [12,13,18–21,42–47]. As previously mentioned, flexural behaviors of reinforced UHPC beams are highly dependent on the specimen geometry, the material properties of UHPC, the shape and volume fraction of steel fibers, and the amount and strength of steel rebars. In the database of **Table 1**, the height (H), width (B) of a given cross-section and the length of shear span (L_a) are considered to represent the geometrical size of the specimen. Additionally, the cylinder compressive strength of UHPC material (f_c) and mechanic characteristics of blended fibers including the shape, length (L_f), diameter (d_f) and volume fraction (V_f) are involved. Furthermore, the yielding strength (f_y) and reinforcement ratio (ρ_r) of steel rebars are also presented. A total of nine performance-sensitive parameters are incorporated as input variables into the established database, while the ultimate capacity of bending moment (M_u) is selected as the output variable. **Table 2** provides detailed information on the statistical characteristic values of the parameters involved.

The presence of longitudinal tensile reinforcement in plain concrete beams has been proven to enhance the load-carrying capacity and stiffness of the structure. Accordingly, more than 95 percent of the flexural specimens in the database are equipped with longitudinal tensile reinforcement. Furthermore, the incorporation of steel fibers into UHPC matrix also improves its tensile strength and toughness. Consequently, 94.6% of the UHPC specimens selected in the database are blended with steel fibers, and the effects of various fiber characteristic parameters on their structural performance are explored. The inclusion of versatile steel fibers is particularly advantageous for enhancing the flexural capacity of UHPC structures. Specifically, the distribution percent of steel fiber shape of UHPC specimens included in the database is presented as follows: 79.5% of straight fibers, 2.5% of hooked-end fibers, 2.8% of corrugated fibers and 9.7% of hybrid fibers. Note that T is Steel fibers with different shapes. T is encoded as numbers to make it easier for models to process the data. Using numbers instead of words helps with calculations and analysis. Each number represents a different shape of fiber: 1 denotes straight fibers, 2 denotes corrugated fibers, 3 denotes hooked-end fibers, 4 denotes hybrid fibers, and 0 denotes specimens without steel fibers. In addition, 5.4% of the specimens without steel fibers are included, thus providing a basis for comparison in regard to the sensitivity of steel fibers. Overall, the database comprises a substantial number of experimental parameters, which may enhance the adaptability of machine learning models for training and evaluation.

Figure 1 illustrates the frequency histograms of each parameter, as well as the dependence between the input variable and the target output variable of ultimate bending moment M_u . It is evident that the estimation of flexural capacity M_u for UHPC specimens is a highly intricate and challenging process. As shown in Figure 1, an increase in the value of flexural capacity M_u is observed with a growing parameter of H , B , L_a , f_c , f_y , and ρ_t . This trend is consistent with the fundamental principles of structural design and material properties. The regression curves for the parameters H and B in Figure 1 display greater values of slopes, indicating that these parameters exert a more pronounced influence on flexural load-carrying capacity M_u . In contrast, the linear slopes of regression curves for the volume fraction V_f and aspect ratio L_f/d_f of steel fibers are approximately zero, making it challenging to assess their impact on M_u . The relatively similar shapes of steel fibers employed in the bending tests may be responsible for the phenomenon, and additional research is required to confirm this hypothesis.

The application of simple linear regression is inadequate for clarifying the inherently complex relationship between ultimate bending moment M_u and an individual input variable. As a result, finite element analysis methods and nonlinear numerical modeling have emerged as a significant development of prediction tools for structural evaluation, providing optimization solutions to an ever-increasing number of complicated structures. The purpose of this study is to estimate the flexural load-carrying capacity M_u of reinforced UHPC specimens in the afore-mentioned database using several ML-based algorithms, including both traditional ML models and EL models. These methods are capable of accommodating a range of complexities, which are user-friendly to employ, and thus facilitate highly nonlinear modeling. This methodology of ML will enable the design of UHPC-based structures with reduced environmental impact and enhanced sustainability, as well as an improved accuracy and efficiency of performance prediction.

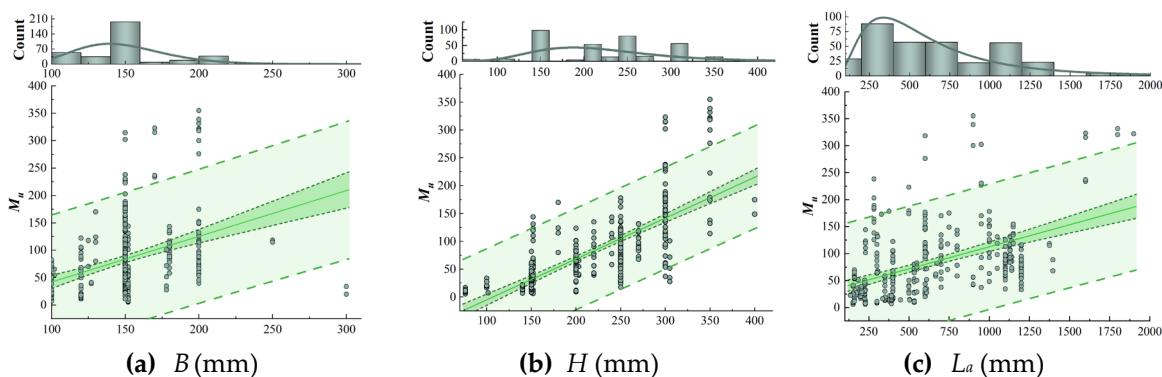
Table 1. Summary of literature on flexural tests of UHPC beams.

Year	Ref.	Specimen number	Design parameters	Moment capacity	Year	Ref.	Specimen number	Design parameters	Moment capacity M_u
2010	[18]	10	ρ_t/f_c	83.3~131.7	2019	[48]	5	V_f/f_c	118~154.5
2011	[49]	7	$H/V_f/f_y/\rho_t$	26.6~222.9	2019	[50]	1	ρ_t	38.5
2011	[51]	5	ρ_t/f_y	11.1~101	2019	[44]	6	$\rho_t/V_f/f_y/f_c$	16.7~33.9
2012	[52]	10	ρ_t/f_y	32.5~144	2019	[53]	9	$\rho_t/V_f/T/f_c/(L_f/d_f)$	40~88.3
2012	[54]	5	ρ_t/f_c	27.6~100.8	2019	[55]	4	ρ_t/f_c	233.6~323.2
2013	[56]	4	V_f/f_c	23.7~29.1	2020	[57]	9	$\rho_t/V_f/f_c$	53.8~116
2013	[58]	4	$\rho_t/(L_f/d_f)$	122~178	2020	[59]	6	$\rho_t/V_f/g$	11.2~21.5
2013	[60]	1	L_a	320.4	2020	[61]	3	V_f/f_c	126~152.5
2014	[17]	2	ρ_t	8.1~9.1	2020	[62]	4	V_f/f_c	102~120
2015	[63]	4	ρ_t/f_y	48.1~101.6	2020	[64]	15	$V_f/f_y/\rho_t/f_c/L_a$	7~22.1
2015	[47]	5	$\rho_t/(L_f/d_f)/T$	39.3~56.1	2020	[65]	4	V_f/f_c	35.7~38.7
2015	[66]	5	ρ_t	90.6~171.6	2021	[67]	8	$\rho_t/V_f/f_y/f_c$	37.1~314.5
2016	[21]	4	ρ_t	72.5~131	2021	[68]	2	ρ_t	58.1~61.9
2016	[69]	1	ρ_t	322	2021	[43]	18	ρ_t	9.1~80.2
2017	[70]	6	V_f/f_c	15.6~19.1	2021	[71]	12	ρ_t	16.5~50.7
2017	[19]	4	ρ_t/L_a	33~118.3	2021	[72]	6	$V_f/f_y/T/L_a$	114~331.7
2017	[73]	2	ρ_t/f_y	70.4~117.8	2021	[74]	10	$\rho_t/f_c/L_a$	22.2~30.8
2017	[75]	6	ρ_t/f_c	13~30.1	2021	[76]	5	ρ_t	28.6~82.5
2018	[77]	8	ρ_t/f_y	43.3~135	2021	[78]	13	$\rho_t/V_f/T/f_c/f_y$	50.8~98.7

2018	[20]	8	$\rho_t / V_f / T / (L_f / d_f)$	37.5~134.4	2022	[6]	8	ρ_t / T	34.2~125.1
2018	[13]	4	ρ_t / H	6.1~12.5	2022	[79]	4	ρ_t / f_y	40.1~58.5
2018	[46]	2	ρ_t	67.8~88.4	2022	[80]	4	ρ_t / f_y	44.1~62
2018	[81]	2	V_f	148.9~174.9	2022	[42]	4	ρ_t / f_y	79.9~170
2018	[45]	14	$\rho_t / f_c / f_y$	11.4~69.8	2023	[82]	5	f_y / ρ_t	69~123.5
2018	[83]	11	$\rho_t / V_f / f_y / f_c / L_a$	125.5~238.3	2023	[15]	5	$\rho_t / V_f / f_y / f_c$	104~171.5
2018	[84]	13	ρ_t	29.9~122.2	2023	[16]	6	$\rho_t / V_f / f_c$	52.8~143.4
2019	[12]	5	ρ_t	5.6~40.9	2023	[85]	2	ρ_t	95.1~111.6
2019	[86]	4	ρ_t	23.8~51.2	2023	[87]	5	$V_f / f_c / \rho_t$	110.6~176.5

Table 2. Statistical information of the parameters chosen.

Parameters	Description	Unit	Mean	Minimum	Maximum	Standard deviation	Median	Skewness	Kurtosis
H	Height of cross section	mm	219.65	76	400	66.11	220	0.10	-0.70
B	Width of cross section	mm	148.81	100	300	31.09	150	0.75	2.73
ρ_t	Ratio of longitudinal reinforcement	%	2.68	0	16.4	2.23	1.9	1.91	5.66
f_y	Yield strength of longitudinal reinforcement	MPa	477.57	0	1395	186.09	456	1.74	9.96
f_c	Compressive strength	MPa	138.07	74.7	216	28.18	134.425	0.78	0.27
V_f	Volume fraction of steel fiber	%	1.81	0	4	0.72	2	-0.69	1.02
L_f/d_f	Aspect ratio of steel fiber	-	64.30	0	150	17.97	65	-1.46	6.05
L_a	Shear span length	mm	627.72	135	1900	377.48	533.3	0.71	-0.28
M_u	Ultimate bending moment	kN·m	82.90	5.6	3552	67.03	68.18	1.66	3.42



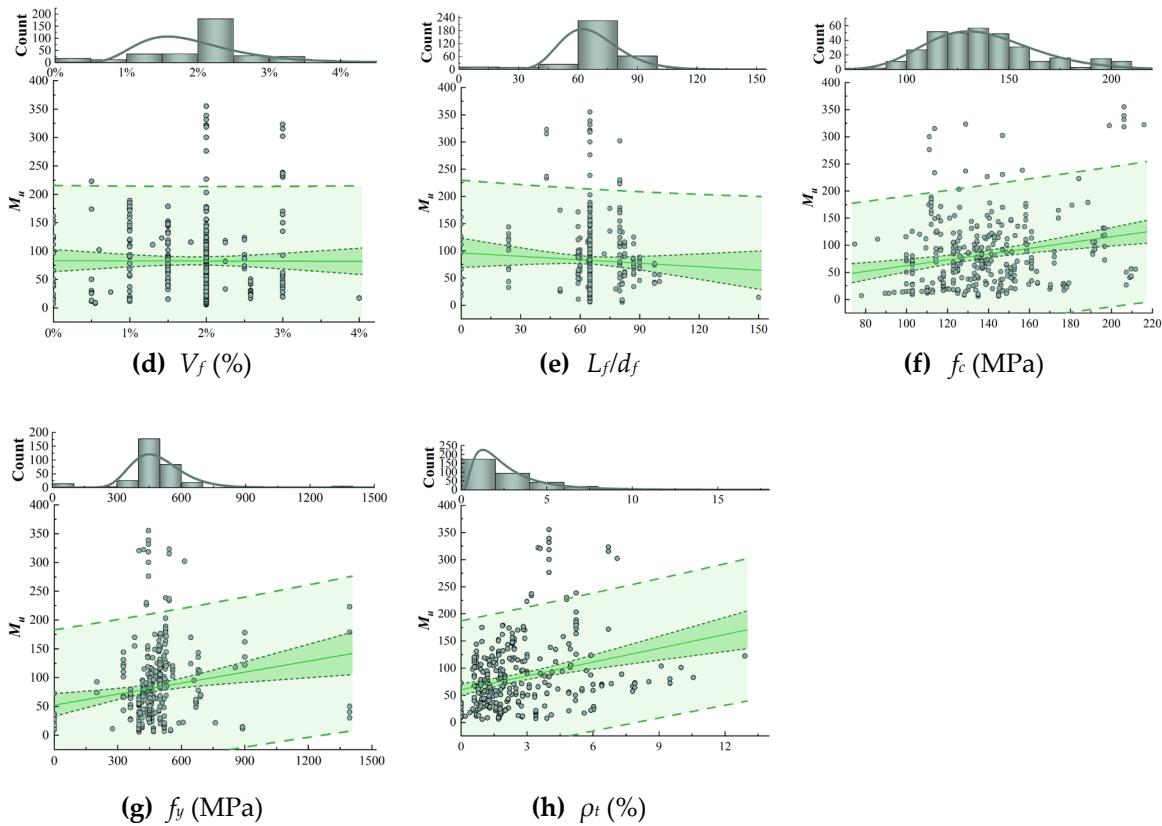


Figure 1. Dependence between the ultimate flexural capacity M_u and input variables.

3. Machine Learning Model

3.1. General Framework

The general framework of this study is illustrated in **Figure 2**. Given the time-consuming and labor-intensive experimental effort of UHPC beams with a great multitude of specimens designed with versatile influencing factors, the machine learning models can be employed to predict the ultimate moment capacity with efficient and accurate. The main technical procedures undertaken in the course of this study are presented as follows:

- First, the experimental results from multiple reinforced UHPC beams are compiled into the database, which are then utilized as input values for the subsequent stage. The database is then divided into two distinct sets for training and testing respectively.
- Second, 10 ML models, composed of traditional ML methods and ensemble learning models, are constructed for the analysis of the established database.
- Third, the hyper-parameters of the 10 ML models are computed and self-adjusted to enhance their prediction accuracy.
- Fourth, the prediction accuracy and efficiency of the 10 ML models are evaluated individually and comparatively.
- Fifth, the stability of the various ML models is investigated by dividing the database into subsets of different sizes for training and testing.
- Further, a comparison analysis is conducted between the calculated values from several existing empirical formulas and the predicted values of the CatBoost model.
- Last, a Shapley Additive exPlanations (SHAP) analysis is employed to interpret the ML models. This allows for the identification of the dependency of each parameter on the ML model and the interactions between parameters.

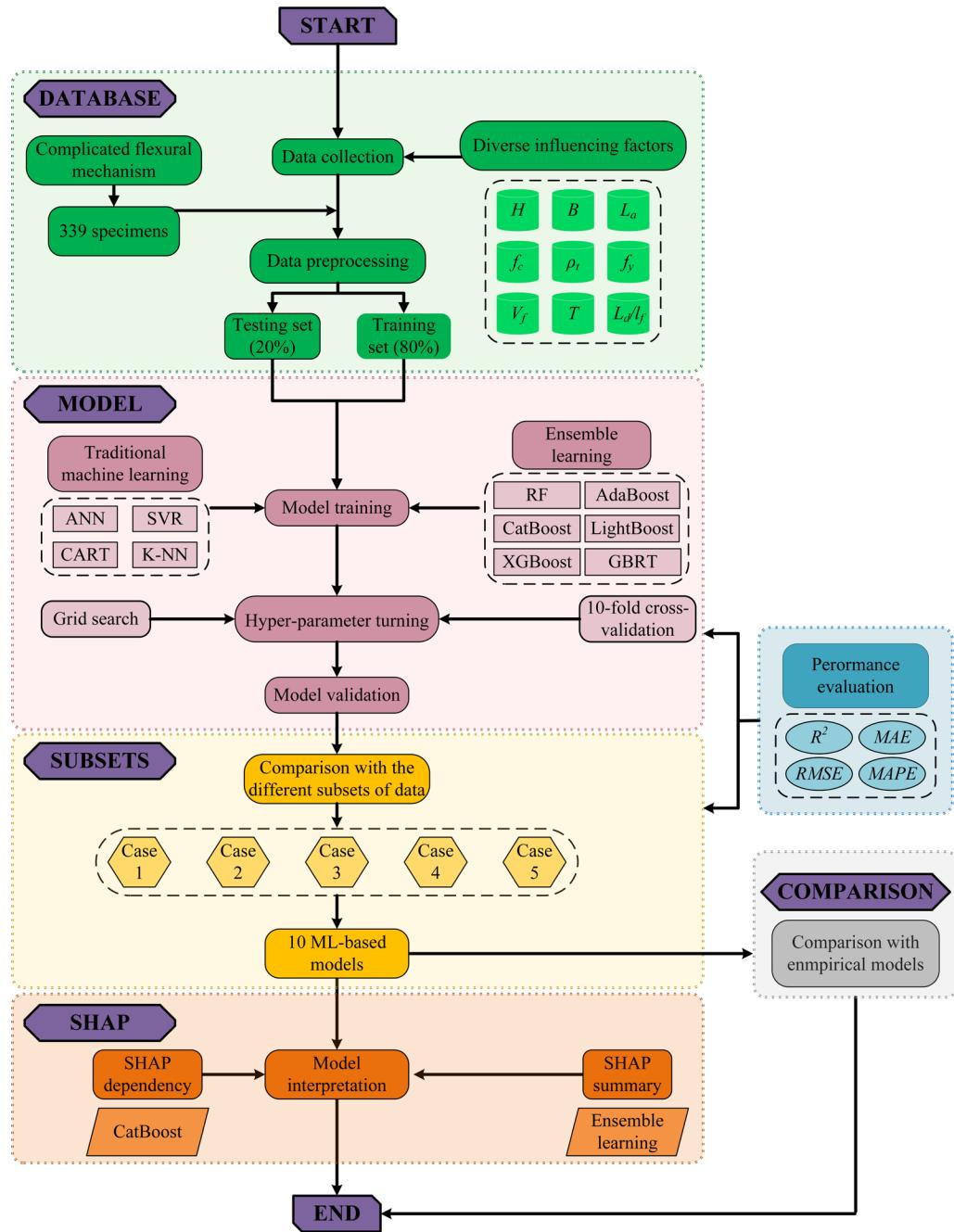


Figure 2. Workflow of this study.

3.2. Traditional Machine Learning Models

The prevailing approach to machine learning is to seek an optimal classifier to achieve maximal data separation. This methodology has the advantage of low computational complexity and a broad applicability. However, there are also notable limitations. For example, it is reliant on domain-specific knowledge for feature extraction and is unable to autonomously learn higher-order features, which presents challenges with complex database structures. These limitations highlight the need for innovative approaches. Several traditional ML models covered in the present investigation are outlined below.

3.2.1. Artificial Neural Network (ANN)

An artificial neural network is a computational model that emulates the structure of the human brain through the interconnection of artificial neurons. It is composed of three principal layers: an

input layer, which receives data and passes it to other layers; a set of hidden layers, which processes the data; and an output layer, which produces the output of network. This interconnected network allows for complex processing, facilitating learning through the adjustment of connection weights, thereby optimizing performance [34,35,88].

The input layer is responsible for processing external data, while the hidden layers are tasked with the extraction of features through the application of non-linear transformations. The output layer generates the final predictions. The behavior of a neuron is defined by the weights (w), which are multiplied by the input data (x) to yield the weighted sum with bias (b). This weighted sum is determined by an activation function, and is adjusted incrementally to yield the desired outputs for the hidden layer. For reference, the subscripts i and j denote the i th layer of the network, and j th unit in a layer, respectively. Therefore, the output for a neuron in the hidden layer is represented as follows:

$$h_i = f \left(\sum_{j=1}^n w_j^i x_j + b^i \right) \quad (1)$$

where h , f , w , x , b , b is the output, activation function, weight, input feature, and bias, respectively. Note that the input features ($x_1, x_2, x_4, \dots, x_{11}$) in this paper are the aforementioned 11 parameters potentially affecting the bending capacity of reinforced UHPC beams, and the target output is the ultimate bending moment of reinforced UHPC beams M_u . The architecture of the neural network is shown graphically in **Figure 3**.

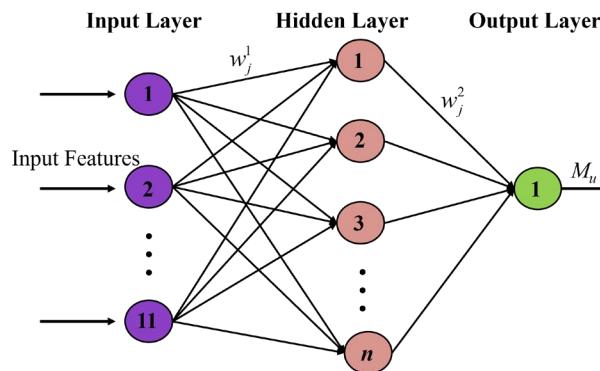


Figure 3. This is a figure. Schemes follow the same formatting.

3.2.2. Support Vector Regression (SVR)

The support vector machine (SVM), known for its accuracy and simplicity, is a widely used algorithm that was first introduced for classification tasks by Boser [89]. SVM performs data classification by the identification of an optimal decision boundary. Further, the support vector regression (SVR) is an extension of the primary principles to regression problems. As shown in **Figure 4**, SVR uses a linear regression model to seek a hyperplane that best fits the data within a decision boundary. The optimal hyperplane maximizes the numbers of data points within a certain margin of tolerance (ε). A prediction $h(x)$ within ε of actual value y incurs no penalty, which promotes robustness and generalizes to unseen data[33,35,88]. The objective of the final optimization is

$$\min_{w, b, \xi_i, \xi_i^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \quad (2)$$

$$s.t. \begin{cases} \varepsilon + \xi_i \geq y_i - (w \cdot x_i + b) \\ \varepsilon + \xi_i^* \geq -y_i + (w \cdot x_i + b) \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (3)$$

where w is the weight defining the separation boundary, b is the bias, and ξ is the relaxation variable; the parameter C controls the balance between maximizing the margin and minimizing classification errors. A larger C enforces stricter avoidance of misclassification, which may lead to overfitting, while a smaller C allows for more errors, potentially increasing generalization. By default, C is set to 1.0. The slack variables ξ_i and ξ_i^* represent the distances from the predicted values to the upper and lower margin boundaries, respectively.

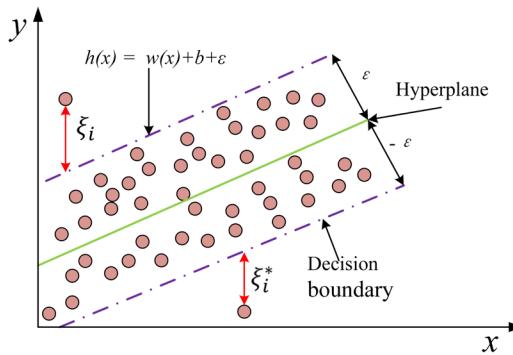


Figure 4. Graphical representations of the SVR used in this study.

3.2.2. K-Nearest Neighbor (K-NN)

K-nearest neighbor, introduced by Fix and Hodges in 1951[90], is a fundamental supervised learning algorithm that is used in both classification and regression tasks. K-NN is non-parametric, which means that it does not assume a specific distribution of the data and postpones learning until testing, and is therefore often referred to as a 'lazy' algorithm. K-NN predicts the class or value of a data point based on the proximity of its neighboring points, typically using the Euclidean distance as the metric[35,88]. Given two points $A (x_1, x_2, x_3, \dots, x_n)$ and $B (y_1, y_2, y_3, \dots, y_n)$, the Euclidean distance between them is determined as follows:

$$d(A, B) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

where $d(A, B)$ is the distance between two points; n is the number of dimensions; x_i and y_i are the coordinates of points A and B in the i th dimension, respectively. It is a description of the method for the computation of Euclidean distance in a multi-dimensional space. Alternative distance metrics, such as the Manhattan distance and Minkowski distance, are also applicable.

In K-NN regression, it averages the values of the nearest neighbors to produce predictions. This involves the identification of the nearest neighbors for the given sample and then the calculation of the average of their labels to determine the prediction value, as shown in **Figure 5**. Consider a sample X_q that is expected to predict, where N_q represents the K -nearest neighbors of X_q :

$$\hat{Y}_q = \frac{1}{K} \sum_{i \in N_q} y_i \quad (5)$$

where \hat{Y}_q is the prediction value for the sample X_q , and y_i denotes the label of the i th neighbor.

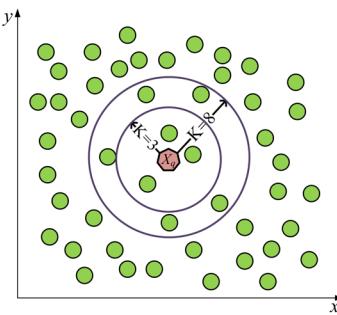


Figure 5. Graphical representations of the K-NN used in this study.

3.2.4. Classification and Regression Trees (CART)

A decision tree (DT), introduced in the 1960s, is a widely used decision-making model structured like a tree with nodes representing decision points and leaf nodes representing outcomes. As an important variant, the classification and regression Tree (CART) extends this approach of DT [35,88,91].

DTs work by recursively dividing the data into increasingly homogeneous subsets, and this process continues until a stopping criterion is met. In the CART algorithm, features and thresholds are selected at each node to maximize the purity or minimize the impurity, as shown in **Figure 6**. For regression tasks, the mean squared error (MSE) is commonly employed to measure the difference between predicted and actual values, which is defined by

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

where MSE signifies the mean squared error across all observations, y_i is the true value, \hat{y}_i is the predicted value, and n is the sample size. The metric of CART algorithm is critical in evaluating the accuracy of the regression model to ensure that the decision tree not only captures the essence of the data but makes accurate predictions.

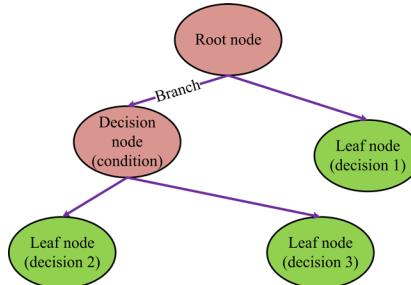


Figure 6. Graphical representations of the CART used in this study.

3.3. Ensemble Learning (EL)

Ensemble learning is the combination of multiple weak learners to build a powerful predictor, and is commonly used in classification, regression, and anomaly detection tasks. The two prevalent technologies of ensemble learning are the bagging and boosting [92], as shown in **Figure 7**. Bagging (bootstrap aggregating) works in parallel, where each learner is trained independently on the bootstrap samples. The final predictions are made by aggregating all of the learners, which reduces the variance and prevents overfitting through voting or averaging [32,35,88].

Boosting, on the other hand, is sequential, where each learner is a corrector of previous errors. Learners are interdependent, and the final predictions are weighted on the basis of accuracy. Boosting reduces both bias and variance, thereby improving the performance of the EL model [93].

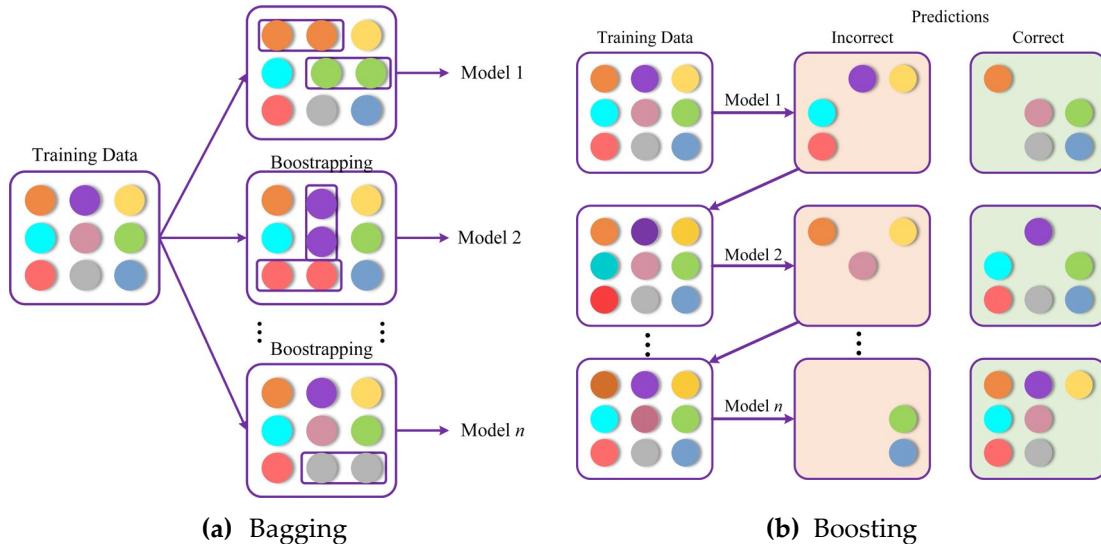


Figure 7. Two technologies for ensemble learning models.

3.3.1. Random Forest (RF)

Random forest (RF) employs the bagging technique to construct multiple decision trees in parallel, each of which is trained on a randomly selected subset of the data and features. This randomization ensures the diversity among the trees, which enhances the robustness of the model [94].

When making predictions, RF aggregates the outputs from all trees, as shown in **Figure 8**. For regression tasks, this is realized by averaging the predictions, which results in more accurate and stable estimates. The random sampling of data and features helps prevent overfitting and increases generalizability. This provides a balance between variance and bias for more reliable predictions. A RF model can be written as:

$$\hat{R}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (7)$$

where $T_b(x)$ is a basic learner, and B is the number of basic learners.

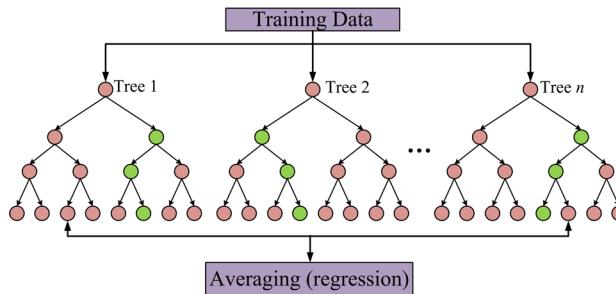


Figure 8. Flowchart of RF for parallel training.

3.3.2. Adaptive Boosting (AdaBoost)

Adaptive boosting (AdaBoost) is an important boosting algorithm known for its application of the exponential loss function. Its core idea is the sequential training of weak learners, where the data weights are adjusted after each iteration. Misclassified samples are given higher weights, which encourages the next classifier to focus on cases that are harder to predict [93]. Each iteration recalibrates the dataset and refines predictions by emphasizing the difficult sample, as shown in **Figure 9**. In particular, AdaBoost adjusts the weights of the dataset, thereby increasing the importance

of the observations that have been misclassified in the previous iteration, while decreasing the influence of those that have been predicted correctly. AdaBoost combines weak learners $h_t(x)$ into a strong ensemble $H(x)$, and improves performance in a variety of applications by adaptively focusing on challenging instances.

$$H(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (8)$$

$$\alpha_t = \frac{1}{2} \log \frac{1 - e_t}{e_t} \quad (9)$$

$$e_t = \sum_{i=1}^N \omega_{ti} I(h_t(x_i) \neq y_i) \quad (10)$$

where α_t is the weight of a weak learner $h_t(x)$; e_t is the error rate, where a lower value leads to a higher weight and vice versa; I is the indicator function, which returns 1 if the prediction $h_t(x_i)$ and the actual value y_i do not match (i.e., a misclassification has occurred), and 0 if they are equal (i.e., the prediction is correct); ω_{ti} is the weight of sample t in the i th iteration. T is the total number of weak learners, and N is the total number of training samples.

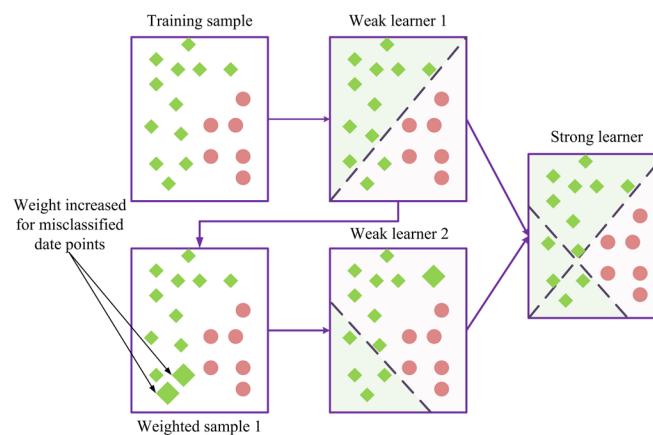


Figure 9. Graphical representation of implementation of AdaBoost with two weak learners.

3.3.3. Gradient Boosting Regression Decision Tree (GBRT)

Gradient Boosting Regression Trees (GBRT) and AdaBoost differ mainly in their updating strategies. AdaBoost adapts sample weights, whereas GBRT updates the regression targets on the basis of the residuals from the previous rounds[95]. As illustrated in **Figure 10**, GBRT employs the gradient of these residuals (r_m) to rebuild new weak learners and iteratively improve prediction accuracy. GBRT typically uses CARTs as weak learners, and the GBRT model can be represented as Equation (11). For regression tasks, it uses binary trees and loss functions such as mean squared error (MSE), absolute loss, and Huber's loss. Huber's loss compensates for MSE and absolute loss, applying absolute loss to outliers and MSE to points near the center.

$$r_m = - \left[\frac{\partial \text{Loss}(y, F_{m-1}(x))}{\partial F_{m-1}(x)} \right] \quad (11)$$

$$F_M(x) = \sum_{m=1}^M T(x; \Theta_m) \quad (12)$$

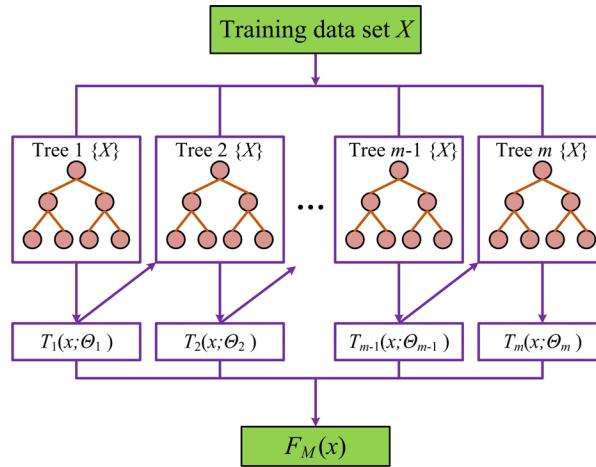


Figure 10. Illustration of the GBRT model.

3.3.4. Extreme Gradient Boosting (XGBoost)

Extreme gradient boosting (XGBoost), introduced by Chen and Guestrin[96], is an advanced gradient-boosting algorithm. It improves upon traditional GBRT by adding a regularization term to the objective function, decreasing overfitting, and using a second-order Taylor expansion to optimize computational efficiency[96]. While XGBoost is similar to GBRT, the essential difference lies in its improved objective function, which is designed to provide faster and more accurate predictions. If K trees and n samples are given, the objective function can be expressed as

$$\mathcal{L}(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (13)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (14)$$

where $\sum_{i=1}^n l(y_i, \hat{y}_i)$ is a loss function that measures the difference between the predicted value \hat{y}_i and the true value y_i of the model; $\Omega(f_k)$ is the regularization term, which is used to prevent the model from overfitting; T is the number of leaf nodes in the tree, and w is the weight of the leaf node; γ and λ are the regularization parameters, and θ is the set of model parameters.

3.3.5. Light Gradient Boosting Machine (LightGBM)

Compared to the traditional gradient boosting algorithm, light gradient boosting machine (LightGBM) also introduces several improvements [97]. First, it employs the Histogram-based Algorithm (HBA) to discretize continuous features into buckets, thereby reducing computational effort. Second, it utilizes a Leaf-wise Growth (LWG) strategy, which selects only the leaf node with the highest loss for splitting, unlike a traditional LWG strategy. This approach results in a more rapid reduction in error and an increase in enhanced model performance, as illustrated in **Figure 11**. The gain formula for the determination of the optimal splitting point is

$$Gain = \frac{1}{2} \left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right) - \gamma \quad (15)$$

where G_L and G_R are the gain sums of the left and right subtrees respectively; H_L and H_R are the second order gradient sums of the left and right subtrees respectively; λ is the regularization parameter and γ is the penalty term for the number of leaf nodes.

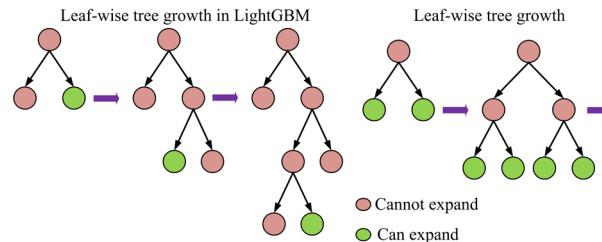


Figure 11. Tree growth methods used in LightGBM and other BAs.

3.3.6. Categorical Gradient Boosting (CatBoost)

Categorical gradient boosting (CatBoost) is first proposed in 2017 by a search company named Yandex to better deal with categorical features. CatBoost improves traditional gradient boosting algorithms in several ways, most notably by integrating an innovative algorithm that automatically converts categorical features into numerical ones. This approach is based on the estimation of target statistics by means of stochastic permutations, also known as ordered target statistics (OTS).

$$\text{OTS}(x_i) = \frac{\sum_{j=1}^{i-1} y_j}{i-1} \quad (16)$$

where x_i is the i th eigenvalue of the sample, and y_j is the j th target value of the sample. CatBoost optimizes the efficiency of training by using an oblivious tree structure, in which each node at the same level is split symmetrically using the same features and points, as demonstrated in **Figure 12a**. This approach ensures uniformity and significantly reduces the amount of computation required. In addition, the residuals of CatBoost shown in **Figure 12b** are computed excluding the current sample to reduce the prediction bias, thereby increasing the accuracy of the model.

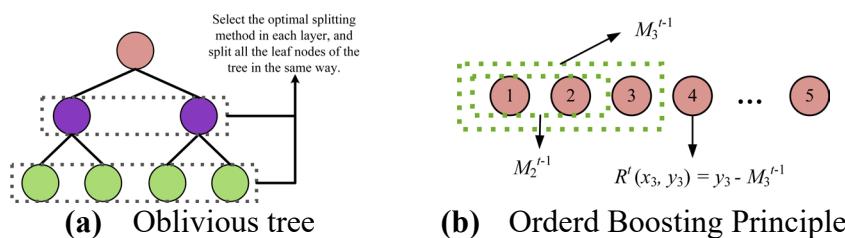


Figure 12. Symmetric oblivious trees and enhanced residual calculation of CatBoost Model.

3.4. Hyper-Parameter Tuning and Modelling Evaluation

The addition of more data is a pervasive approach to refining a machine learning model, but the generation of high-quality data is often a time- and energy-consuming process. A more efficient way to improve performance and save time and resources consumption is to optimize hyperparameters [98]. Unlike parameters learned during training, hyperparameters are manually set and require careful tuning[28,34,35,99]. Grid searching is a popular method for hyperparameter optimization, which involves systematically testing combinations to find the best settings. In this study, grid search is employed to fine-tune hyperparameters for ten ML algorithms, which are informed by previous models and tailored to our database, as shown in **Table 3** and **Table 4**.

Table 3. Hyperparameter tuning of traditional machine learning models.

ML Mode 1	Hyper-parameter	Optimized value	Range	ML Mode 1	Hyper-parameter	Optimized value	Range
ANN	Hidden_layer_sizes	(80,)	(10,) ~ (150,)	SVR	C	1500	1000 ~ 1500

	max_iter	10000	0~15000	gamma	0.15	0.001 ~ 1
	activation	relu	{'relu', 'tanh'}	kernel	rbf	{'linear', 'rbf', 'poly'}
	learning_rate	adaptive	{'constant', 'adaptive'}	leaf_size	20	10 ~ 50
	alpha	0.1	0.0001~0.01	K-NN	n_neighbors	2
CAT	min_samples_leaf	1	1 ~ 8		weights	distance
	min_samples_split	2	1 ~ 8			{'uniform', 'distance'}

Note: 'relu' means using the Rectified Linear Unit activation function; 'tanh' means using the hyperbolic tangent activation function; 'constant' means keeping a fixed learning rate; 'adaptive' means automatically adjusting the learning rate based on the training process; 'linear' means using a linear kernel; 'rbf' means using the Radial Basis Function kernel; 'poly' means using a polynomial kernel; 'uniform' means assigning equal weights to neighbors; 'distance' means dynamically adjusting the weights based on the distance of neighbors.

Table 4. Hyperparameter tuning of ensemble learning models.

ML Model	Hyper-parameter	Optimized value	Range	ML Model	Hyper-parameter	Optimized value	Range
LightGBM	learning_rate	0.1	0.001 ~ 0.1	GBRT	learning_rate	0.1	0.05~0.2
	min_child_samples	5	5~ 20		min_samples_leaf	2	1 ~ 4
	n_estimators	400	100 ~ 400		min_samples_split	10	2 ~ 10
	num_leaves	15	5~ 20		n_estimators	300	100 ~ 300
RF	learning_rate	0.1	0.001 ~ 0.1	AdaBoost	learning_rate	0.01	0.01~0.1
	min_samples_leaf	1	1 ~ 3		loss	square	{'linear', 'square', 'exponential'}
	min_samples_split	2	2 ~ 10		n_estimators	600	100 ~ 600
	n_estimators	300	100 ~ 500		depth	4	4~8
XGBoost	learning_rate	0.1	0.1 ~ 0.2	CatBoost	iterations	1000	500 ~ 1500
	min_child_weight	1	1 ~ 3		l2_leaf_reg	0.1	0.1 ~ 1.0
	n_estimators	300	200 ~ 400		learning_rate	0.1	0.01 ~ 0.1

Note: 'linear' refers to a linear loss function; 'square' refers to a squared loss function; 'exponential' refers to an exponential loss function.

To evaluate model performance after hyperparameter tuning, K -fold cross-validation is employed. It divides the data into K subsets, running K rounds of training and testing. The average performance across the folds gives a reliable metric, with $K=10$ often providing a balance between computational efficiency and prediction accuracy. Cross-validation is critical for the evaluation of performance across different folds and for the selection of the best model [100].

To thoroughly evaluate the prediction results of the ML-based models selected in the present study, four statistical indicators of model performance are utilized hereafter. These evaluation

indicators include the coefficient of determination (R^2), root mean square error (RMSE), mean relative error (MRE), and mean absolute percentage error (MAPE). The definition and calculation formula for each statistical indicator is detailed in **Table 5** below.

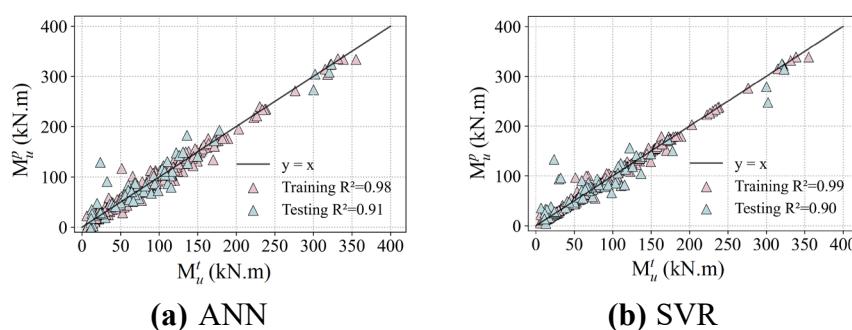
Table 5. Hyperparameter tuning of ensemble machine learning models.

Evaluation indicator	Equation	Note
R^2	$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{Y}_{a,i} - Y_{p,i})^2}{\sum_{i=1}^N (\hat{Y}_{a,i} - \bar{Y}_{m,i})^2}$	The R^2 value ranges from 0 to 1, with values closer to 1 indicating better model performance.
RMSE	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^N (\hat{Y}_{a,i} - Y_{p,i})^2}$	RMSE measures the difference between observed and predicted values. Lower RMSE values, closer to zero, indicate higher model accuracy in predictions.
MAE	$MAE = \frac{1}{n} \sum_{i=1}^N \hat{Y}_{a,i} - Y_{p,i} $	MAE, on the other hand, measures the average magnitude of errors without considering their direction, with lower MAE signifying higher model accuracy.
MAPE	$MAPE = \frac{1}{n} \sum_{i=1}^N \left \frac{\hat{Y}_{a,i} - Y_{p,i}}{\hat{Y}_{a,i}} \right \times 100\%$	MAPE expresses the error as a percentage, offering a relative measure of prediction accuracy. Lower MAPE values indicate more precise predictions.

4. Results and Discussions

4.1. Model Performance: A Comparison Across Diverse ML Algorithms

Ten different algorithms are employed to develop machine learning models based on the aforementioned database to predict the ultimate bending moment M_u of reinforced UHPC beams. The dataset has been divided into 80% for the training and 20% for the testing. **Figures 13 and 14** compare the predicted bending ultimate moments $M_{u,p}$ from the ML models, both traditional ML models and ensemble learning models respectively, with the corresponding tested results $M_{u,t}$ from the established database. The relationships between the predicted ultimate moments and the measured values follow a linear fitted law with a slope of 1.0. Detailed results of this comprehensive evaluation for the bending moment capacity of reinforced UHPC specimens using various ML-based models are presented in **Figure 15**.



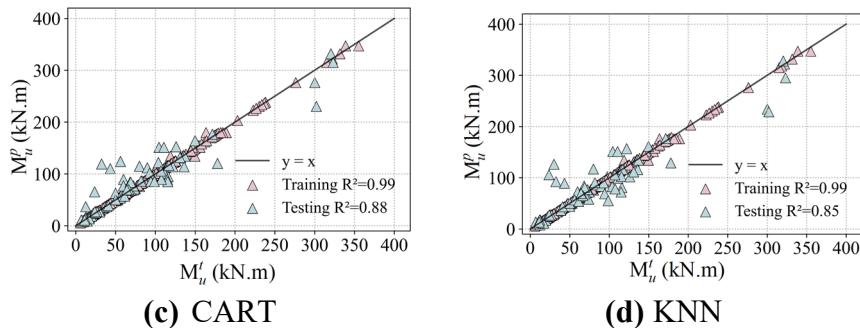


Figure 13. Comparison of the predicted ultimate moments M_u^p from the traditional machine learning models with the corresponding tested results M_u^t from the established database.

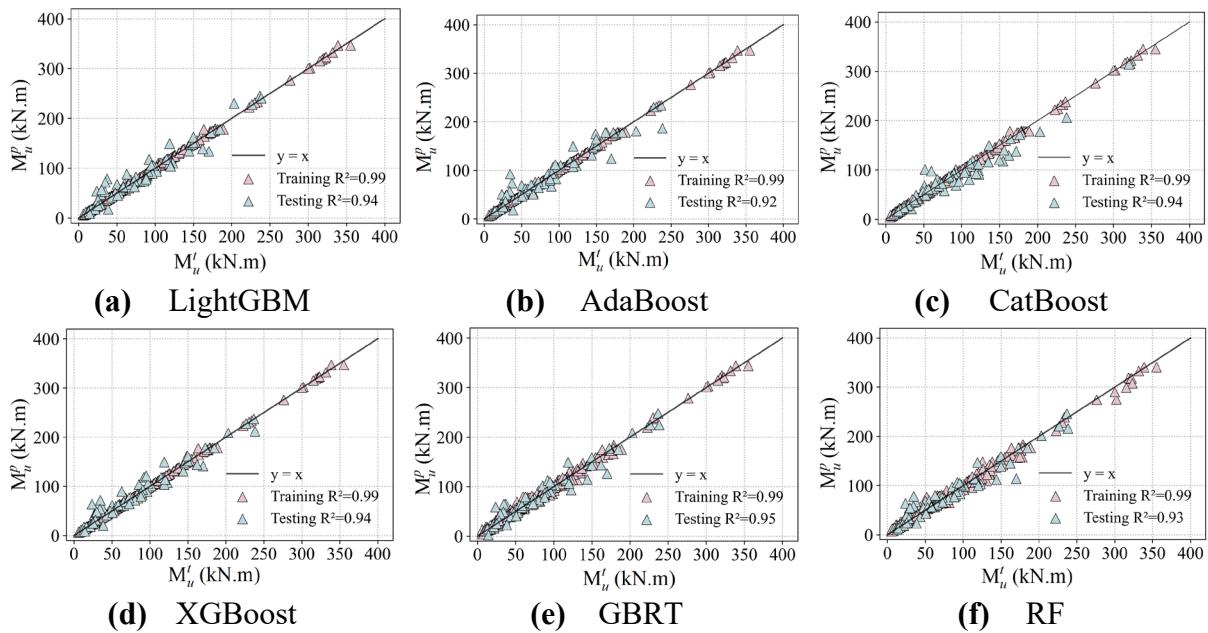


Figure 14. Comparison of the predicted ultimate moments M_u^p from the ensemble learning models with the corresponding tested results M_u^t from the established database.

On the training set, the coefficient of determination R^2 for all ML models except the ANN is greater than 0.99, highlighting their excellent fitting abilities. The ANN model still shows a commendable performance, although it has a slightly lower R^2 value of 0.98. In terms of RMSE, the KNN, AdaBoost, CatBoost, and XGBoost models have relatively lower values compared to other ML models, indicating the minimized discrepancy between their predicted values and tested results, with an average error margin of approximately 2.0. This underlines their exceptional model accuracy in predicting flexural performance of reinforced UHPC beams. In contrast, higher RMSE values of 9.7 and 6.7 are recorded for the ANN and GBRT models have been recorded, respectively. Despite these higher values, the model accuracy is still within an acceptable range.

Further analysis of the MAE reveals that the KNN, AdaBoost, CatBoost, LightGBM, and XGBoost models all maintain values below 2, indicating a negligible average deviation between the predicted and measured values, and thus a high degree of prediction accuracy. Moreover, the evaluation of MAPE clearly shows that the KNN, CART, AdaBoost, and XGBoost models keep the values below 1%, confirming their accurate prediction capabilities. Although the ANN model has a MAPE of 16%, indicating a reduced predictive accuracy—possibly affected by its network structure and hyperparametric settings—it nevertheless meets the fundamental predictive benchmarks.

Considering the testing set, the coefficients of determination R^2 for the LightGBM, CatBoost, XGBoost, and GBRT models are all larger than 0.94, demonstrating their exceptional prediction

potentials. This outstanding performance is primarily due to the inherent advantages of ensemble learning, which includes the reduction of bias and variance in predictions by combining multiple models, thereby enhancing their ability to generalize to new datasets. Conversely, the KNN model gives the lowest R^2 value of 0.85 on the testing set. Its performance limitations may be related to its decision mechanism, which relies on nearest-neighbor voting or averaging. This may fail in the presence of high-dimensional data or uneven data distributions, where the concept of "nearest neighbor" may be somewhat indeterminate.

The KNN, ANN, SVR, CART and AdaBoost models present relatively high values when analyzing the three evaluation indicators of RMSE, MAE and MAPE, indicating a decrease in prediction accuracy on the testing set. The increased sensitivity of these models to the distribution of data features and the presence of noise may be responsible for this trend. In stark contrast, the GBRT and CatBoost models outperform on all three of these indicators, further underscoring the superior effectiveness of ensemble learning models in improving the accuracy of predictions. Specifically, GBRT and CatBoost develop their superiorities from the construction of multiple decision trees and the synthesis of the prediction insights of each tree to reduce potential errors inherent in singular models.

To summarize, the excellent performance of ensemble learning models such as LightGBM, CatBoost, XGBoost, and GBRT on the testing set is fundamentally related to the strategy of ensemble learning with model aggregation. Those approaches effectively reduce the bias and variance, while improving the generalization ability of the models. On the other hand, while the KNN model presents admirable results on the training set, its modest performance on the testing set highlights the importance of considering data characteristics and the compatibility of the logic of model decision with the given problem during model selection. Overall, the ten ML-based models evaluated are capable of accurately predicting the ultimate bending moment values M_u of reinforced UHPC beams, confirming the profound potentials of machine learning models to address challenging structural demands.

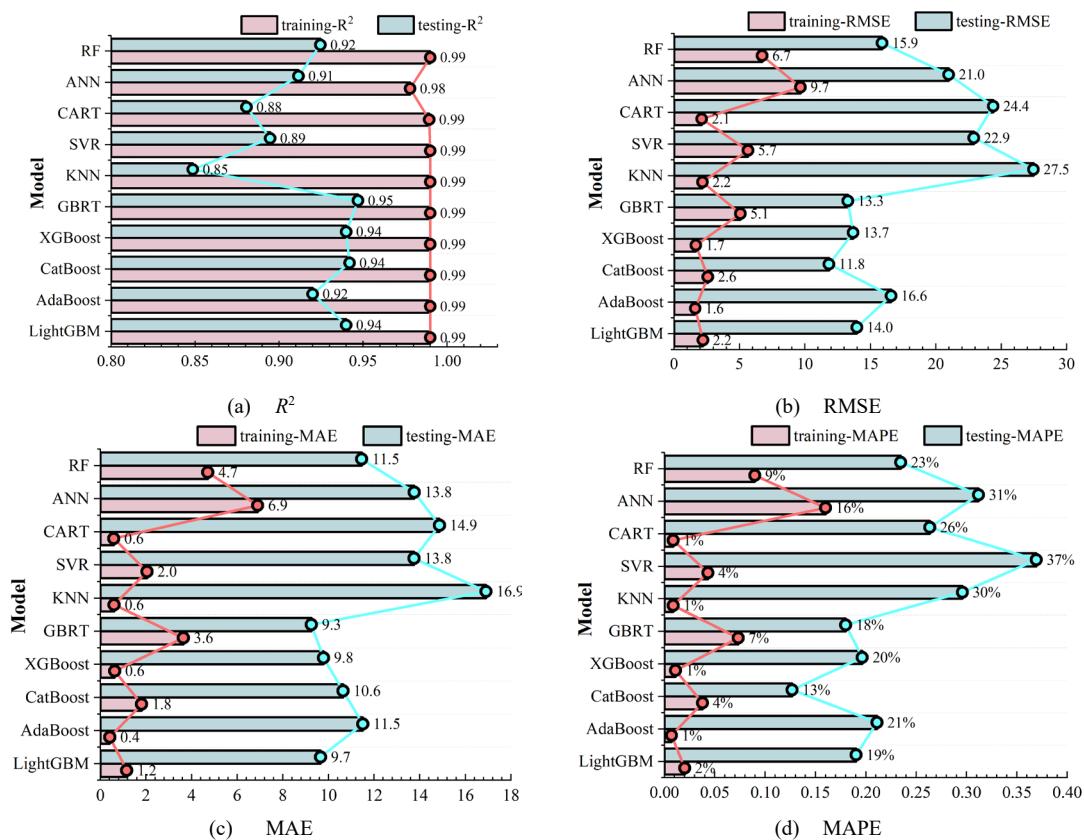


Figure 15. Performance comparison of the ML-based models used.

4.2. Data Subset Analysis for Model Performance and Stability

To systematically evaluate the qualities of ML-based models and database given, as well as to explore model stability, a methodical approach is taken by dividing the database previously established into subsets of varying sizes. This strategy makes it possible to examine model performance across a spectrum of dataset sizes, thereby providing insightful perspectives on how model performance varies with different dataset sizes. Based on findings from previous research and empirical evidence, five different cases of data subsets, as shown in **Figure 16**, have been identified for in-depth analysis.

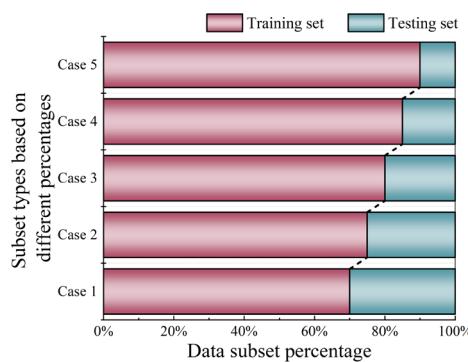


Figure 16. Identification of data subsets for in-depth analysis.

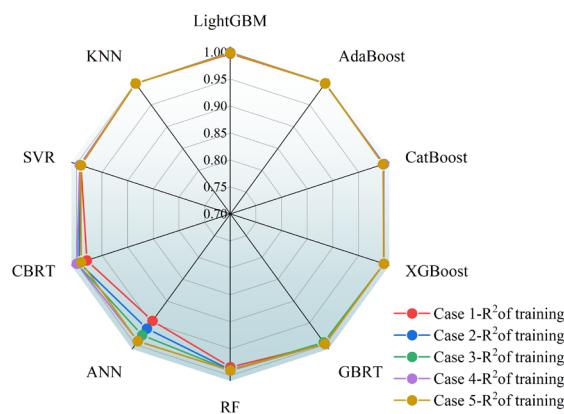
Figure 17 presents a comparative analysis of model performances with various cases of data subsets. An examination of the coefficient of determination R^2 for all models reveals that, across different cases of data subsets, the R^2 values associated with the training set are predominantly greater than 0.98, while those R^2 -values of the testing set are generally larger than 0.9. These results underscore the overall robust performance of ML models. Nevertheless, the ensemble learning models exhibit relatively superior performance compared to the counterparts of traditional ML models. Specifically, the CatBoost model achieves the highest R^2 -value of 0.97 on the testing set at the case 1, and reaches a maximum R^2 -value of 0.96 for testing set at the case 2. For the third to the fifth case of data subsets, the R^2 -value of testing sets peak at 0.94, 0.96 and 0.96 with the ensemble models of GBRT, CatBoost and GBRT, respectively.

This analysis highlights the superior performances of ensemble learning models over traditional ML models in most cases, and explains the variance in model effectiveness when dealing with data subsets of different divisions. The sustained high R^2 -values of ensemble learning models across a variety of data subset configurations can be attributed to their elaborate structures and algorithms, which are able to capture data correlations and patterns in a more effective way. As a result, the accuracy of predictions is improved. Furthermore, integrated models enhance the prediction accuracy by combining several weak learners or regressors. This strategy is especially beneficial when dealing with large and diverse data sets. On the contrary, due to the relatively simple algorithmic structure, traditional ML models may be unable to fully represent the intricacies of data relationships, which impacts to some extent their overall performance.

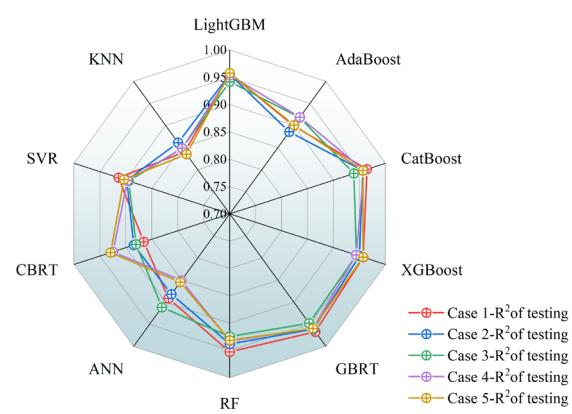
To conduct a thorough evaluation of model performance, three statistical performance indicators of ML models including RMSE, MAE and MAPE are discussed here. When evaluating the training set, a majority of ML models show exceptional and consistent proficiency across these performance indicators. Nonetheless, the KNN models demonstrate suboptimal performance under various cases of data subsets, especially for Case 1. It could be attributed to the fact that the KNN models encounter a deficit in training sample size within the data subset division for the case 1. This leads to an overfitting of training data with details and noise thereby decreasing their generalization abilities. Moreover, the MAE for most of the models is around 10, suggesting a mean absolute deviation of approximately 10 units between model predictions and measured results.

Having analyzed the model performance with statistical indicators, it becomes evident that CatBoost and GBRT models significantly outperform the traditional ML models. The KNN, AdaBoost, SVR, and ANN models display inferior performance in various data subset arrangements. For instance, in the fourth case of data subsets, the ANN model registers a dramatically high MAPE of 41%, suggesting an insufficient prediction accuracy. This may be due to the model not being trained on a sufficiently diverse or large dataset, which may have resulted in inadequate generalization to unseen data. However, in the fifth cases of data subsets, the MAPE values decrease to approximately 18%, revealing a reduction in the average percentage deviation between the prediction values and measured results to about 18%. The significant variation in MAPE values highlights the pronounced differences in the adaptability of diverse models to specific data subsets. It therefore emphasizes the need to consider the sensitivity and adaptability of a model to varying data subsets during the model selection and optimization process.

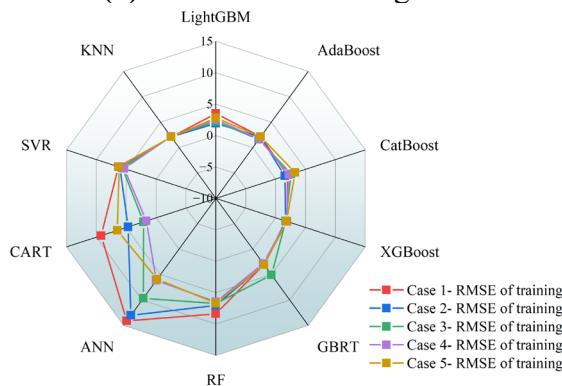
An in-depth evaluation of the ensemble learning models reveals that the second case is found to be the most effective and efficient database division strategy across all of data subset cases. To be more specific, 75% of the database is allocated to the training set, while the remaining 25% of the data served as the testing set. In contrast, the optimal data subset configuration for traditional ML models is identified in the case 3, where the data distribution percentages of the training set and testing set are 80% and 20%, respectively. The findings underscore the considerable influence of data division ratios on model effectiveness. Further investigation reveals that among ensemble learning algorithms, the CatBoost and GBRT models present a remarkable consistency with a varying data subset configuration. On the front of traditional ML models, the CART model stands out for its stability and robustness. Notably, the CatBoost model is distinguished by its superior division strategy of data subsets, considering both model efficiency and stability of data acquisition.



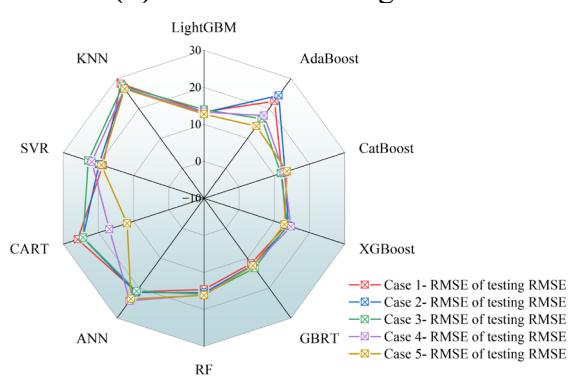
(a) R^2 value of training sets



(b) R^2 value of testing sets



(c) RMSE value of training sets



(d) RMSE value of testing sets

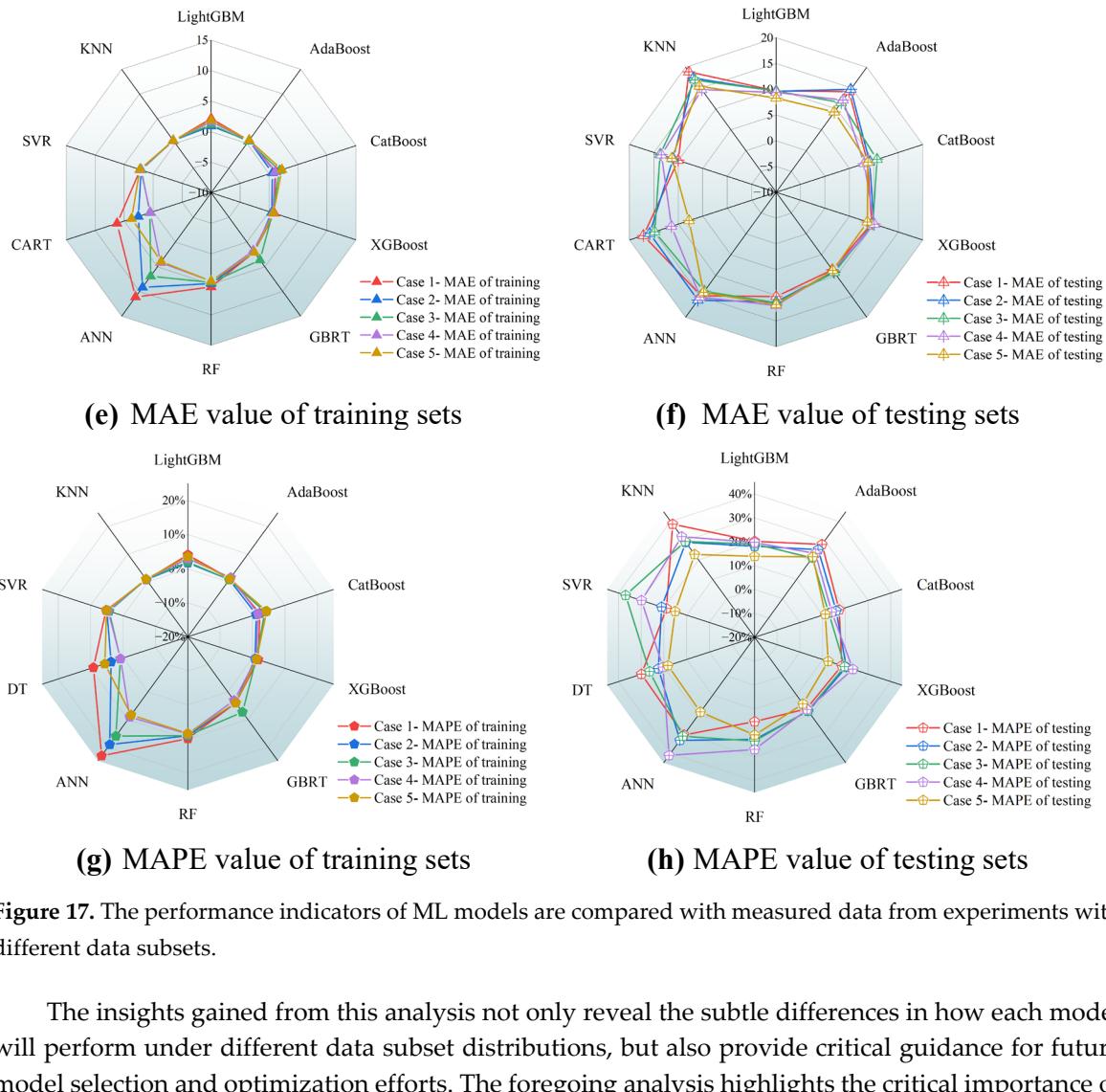


Figure 17. The performance indicators of ML models are compared with measured data from experiments with different data subsets.

The insights gained from this analysis not only reveal the subtle differences in how each model will perform under different data subset distributions, but also provide critical guidance for future model selection and optimization efforts. The foregoing analysis highlights the critical importance of proper data acquisition in improving model performance. Specifically, in the context of ensemble learning models, the selection of an appropriate data subset configuration is of paramount importance for the realization of peak performance. Furthermore, the model stability plays a crucial role in determining how well it performs. Therefore, the effects of data configuration and model stability should be properly considered during the model selection and optimization phases to ensure optimal model functionality in real-world applications.

4.3. Comparison with Existing Empirical Equations

Given the increasing utilization of UHPC-based materials in civil engineering, a multitude of standards and guidelines have emerged worldwide to facilitate the design of UHPC structures [36–39]. The prevailing standards in the field, the French standard NF P 18-710 [36] and Swiss recommendation SIA 2052 [37], provide guidelines for the design of UHPC-based structures. However, these standards face limitations in terms of their practical application and accuracy precision. The French standard emphasizes strain-based failure criteria, requiring iterative calculations without explicitly defined formulas, whereas the Swiss recommendation simplifies compressive stress distribution and applies a reduction factor to tensile contributions. The existing empirical formulas are presented in **Table 6**, with symbol definitions available in the referenced literature, respectively. Similarly, the US design guides of ACI 544.4R-18 [38] and FHWA HIF-13-032 [39], which are based on the equilibrium and strain compatibility, fail to fully capture the nonlinear behavior of UHPC element. The calculation model proposed by Li et al. [40] is derived from

experiments and incorporates UHPC's tensile contribution with an assumption of uniform stress distribution, thereby reducing its applicability under varying reinforcement ratios. For reference, the key formulas for these empirical methods are presented in **Table 6**. Despite the prevalence of existing empirical or code-based methods, numerous studies reveal that the empirical formulas provided for estimating the flexural capacity of reinforced UHPC beams frequently exhibit excessive conservatism, resulting in significant discrepancies between predicted values and experimental observations [12,13]. This study aims to demonstrate the superior predictability of the CatBoost model by comparing its model performance with several widely recognized models based on empirical formulas.

Table 6. Calculation formulas of flexural capacity of reinforced UHPC-based beams.

Empirical equations	Formula Expression
Swiss Recommendation SIA 2052 [37]	$\frac{1}{2} f_c b x_c = 0.9 f_t b (h - x_c) + A_s f_s$ $M_u = \frac{1}{2} f_c b x_c \left(h_0 - \frac{1}{3} x_c \right) - 0.9 f_t b (h - x_c) \cdot \left[\frac{1}{2} \times 0.9 (h - x_c) - a_s \right]$
ACI 544.4R-18 [38]	$M_u = A_s f_y \left(d - \frac{a}{2} \right) + f_t b (h - e) \left(\frac{h + e - a}{2} \right)$ $c = \frac{A_s f_y + f_t h}{f_t \frac{\varepsilon_f + 0.003}{0.003} + 0.85 \beta_1 f_c}$ $e = \frac{\varepsilon_f + 0.003}{0.003} c$ $a = \beta_1 c$ $\sigma_{fs} = 2 \tau_f \left(\frac{l_f}{d_f} \right) \leq (\sigma_{fy})$ $\varepsilon_f = \frac{\sigma_{fs}}{E_{fs}}$
FHWA HIF-13-032 [39]	$M_u = f_{tu} b (h - c) \left(\frac{3h - c}{6} \right) + \rho_s f_y b h \left(d - \frac{c}{3} \right)$ $c = \left(\frac{\rho_s f_y + f_{tu}}{f_{tu} + 0.0035 E_{UHPC} \left(\frac{c}{h - c} \right)} \right)$ $E_{UHPC} = 4200 \sqrt{f_c} \text{ (MPa)}$
Reference [40]	$0.9 f_c b x = 0.25 f_t b \left(h - \frac{x_c}{0.77} \right) + A_s f_y$ $M_u = 0.9 f_c b x \left(h_0 - \frac{x}{2} \right) - 0.25 f_t b \left(h - \frac{x_c}{0.77} \right) \cdot \left[0.5 \left(h - \frac{x_c}{0.77} \right) - a_s \right]$

As shown in **Table 7**, the comparison results reveal that the CatBoost model significantly outperforms the five representative empirical formulas in predicting the flexural capacity of UHPC beams. Empirical design models such as the NF P 18-710 and SIA 2052 provide standardized approaches to the design of UHPC beam; however, they frequently rely on simplified assumptions about material behaviors, such as strain distributions or reduction factors, leading to conservative or

inconsistent predictions. For instance, an examination of the calculation method proposed by Li et al. reveals an average predicted-to-measured flexural capacity ratio of 0.916, thus indicating a tendency to underestimate flexural capacity in practical applications. Conversely, the CatBoost model achieves a mean predicted-to-measured flexural capacity ratio of 1.022, the closest to 1, thereby signifying a higher degree of agreement with actual values.

Table 7. Performance of the empirical method and the CatBoost model.

Models	M_u^p / M_u^t			Quantitative performance			
	Min	Max	Mean	R ²	RMSE	MAE	MAPE
NF P 18-710	0.796	1.911	1.146	0.914	15.724	12.871	18.606%
SIA 2052	0.680	1.665	1.121	0.879	18.674	14.787	19.190%
ACI 544.4R-18	0.565	1.781	1.156	0.863	19.929	16.574	22.557%
FHWA HIF-13-032	0.775	1.809	1.257	0.711	28.923	24.251	29.534%
Reference [40]	0.358	1.244	0.916	0.851	20.781	15.471	19.309%
CatBoost	0.823	1.382	1.022	0.993	4.396	2.055	3.704%

In terms of quantitative performance, the CatBoost model demonstrates superior performance, attaining an R² value of 0.993. This indicates its superior predictive accuracy and fitting capability compared to the existing empirical methods. For example, the recommendation SIA 2052 and the method presented by Li et al. exhibit R² values of 0.925 and 0.851, respectively; while the FHWA method a significantly lower R² value of 0.711. Furthermore, the CatBoost model demonstrates the lowest RMSE value of 4.396, MAE value of 2.055 and MAPE value of 3.704%, exhibiting a substantial improvement in performance compared to empirical methods such as the ACI 544 and FHWA models. These models exhibit significantly higher RMSE values of 19.929 and 28.923, respectively. These findings underscore the efficacy of the CatBoost model in minimizing prediction errors and ensuring consistent accuracy across diverse datasets.

As illustrated in **Figure 18** the predicted data points with the CatBoost model are closely distributed around the baseline $y = x$, suggesting that the model demonstrates reliable and robust performance. The polynomial fitting curve (green) of the CatBoost model exhibits a strong alignment with the observed values and provides a reliable representation of the underlying data. In contrast, traditional empirical models demonstrate notable deficiencies. Specifically, the ACI 544 and the FHWA method exhibit substantial deviations between their curves and the observed values, with a greater degree of scattered data points. It is important to note that the FHWA method, with a high RMSE of 28.923, experiences significant challenges in accurately capturing the complex mechanical behavior of UHPC. Similarly, while the NFP 18-710 model exhibits marginal enhancements, its computational process is complex and its applicability limited. The ACI 544 model, conversely, excessively simplifies the tensile zone contribution of steel fiber-reinforced concrete, resulting in significant prediction errors. These limitations further underscore the advantages of data-driven approaches.

Overall, the CatBoost model demonstrates superior performance in terms of predictive accuracy in comparison to empirical methods. Its enhanced applicability and adaptability are particularly notable, as it is capable of incorporating complex feature interactions and producing highly reliable results, which makes it an invaluable tool for practical engineering applications. The employment of data-driven methodologies by the CatBoost model presents a promising alternative to existing empirical methods, thereby paving the way for enhanced accuracy and efficiency in the field of UHPC structural design.

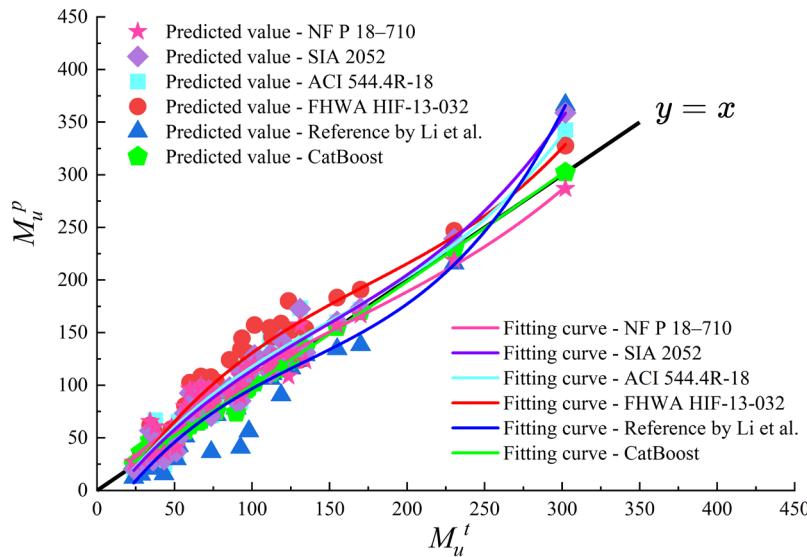


Figure 18. Comparison between the empirical methods and the CatBoost model.

5. Model Interpretation

Advanced machine learning models, such as deep learning, are often considered "black boxes" because of the complexity and nonlinear nature of the models involved that makes it difficult to interpret their decision-making processes. The lack of transparency can have a negative impact on confidence in model predictions.

While techniques such as Local Interpretable Model-Agnostic Explanations (LIME) and Interpretive Decision Trees provide a degree of interpretability, they are limited. A Shapley Additive Explanations (SHAP) has a potential to address these challenges by clarifying the contribution of features to model predictions. SHAP has been widely adopted for model interpretation, since it enhances transparency and credibility through consistency, local interpretability and model independence [35,41,101,102]. The explanatory model of SHAP, $g(a')$, is defined as

$$g(a') = \phi_0 + \sum_{z=1}^Z \phi_z a'_z \quad (17)$$

where ϕ_0 is the baseline value of the model, usually the average of all sample predictions; ϕ_z is the SHAP value of the feature z , which indicates the contribution of the feature z to the prediction; a'_z is a binary indicator for the feature z , indicating whether feature z is in the explanatory model.

5.1. Analysis of Feature Importance Using SHAP

In exploratory analysis of ML models, reliance on the SHAP interpretation for a single model alone may not adequately capture the delicate effects of features on predictions. This limitation arises from the varying dependencies and interactions that different models have with the same set of characteristic parameters. To gain a deeper and more accurate understanding of feature significance and their influence on predictions, it is essential to perform a SHAP analysis with multiple models.

Taking advantages of the global interpretability and powerful visualization capabilities offered by SHAP, a global feature importance analysis across six ensemble learning models is conducted. To illustrate the impact and importance of each feature on model output, **Figure 19** presents the SHAP values for each ensemble learning model. In this figure, the horizontal axis displays the SHAP values, which indicates the extent to which each feature affects the prediction accuracy of the model. Meanwhile, the vertical axis enumerates the features in order of importance, with a color gradient from blue to red representing the progression from lower to higher feature values. It is evident to

note that there are marked discrepancies in how features rank in importance and the direction in which they affect different models.

From **Figure 19**, the SHAP analysis for five models highlights the reinforcement ratio of longitudinal rebars ρ_t , the yielding strength of reinforcement f_y , and the beam height H as the most important features, each of which contributes positively to the prediction results. In contrast, the GBRT model emphasizes the beam height H , reinforcement ratio ρ_t , and the length of shear span L_a as critical, demonstrating the inherent variability in feature prioritization between different models. These observations highlight the critical role of longitudinal reinforcement ratio ρ_t in predicting the ultimate bending moment M_u of reinforced UHPC beams, consistent with its recognized importance in the enhancement of steel rebars for reinforced concrete beams. Despite of a limited tensile strength of concrete-based materials, the longitudinal reinforcements in UHPC beams overcome the limitation by providing essential tensile strength to negative bending moments. The steel reinforcement in UHPC beams effectively carries the tensile load during bending, while the lower section of the beam is longitudinally under tension.

Moreover, several features such as the beam height H , and the beam width B —attribute representatives of the cross-sectional properties of UHPC beams—are highlighted. This emphasizes the critical role of cross-sectional characteristics in determining the flexural performance UHPC beams. The feature B is highly significant in the CatBoost, XGBoost, and GBRT models, but it is of much less importance in the LightGBM model, where its influence is ranked remarkably lower. This discrepancy suggests that feature importance ratings vary due to the unique mechanisms that each model utilizes to process features. For the six ensemble learning models, the yielding strength of reinforcement f_y and the length of shear span L_a show more consistency in both importance and direction of impact. The SHAP values for these features are more concentrated, which is an indication of a more uniform influence on the prediction results. It is noteworthy that the yielding strength of reinforcement f_y is recognized in the classical formulations, while the influence of shear span length L_a is absent in the design specifications. This discrepancy indicates that traditional empirical formulas may not fully capture the complexity associated with flexural capacity prediction M_u for reinforced UHPC beams.

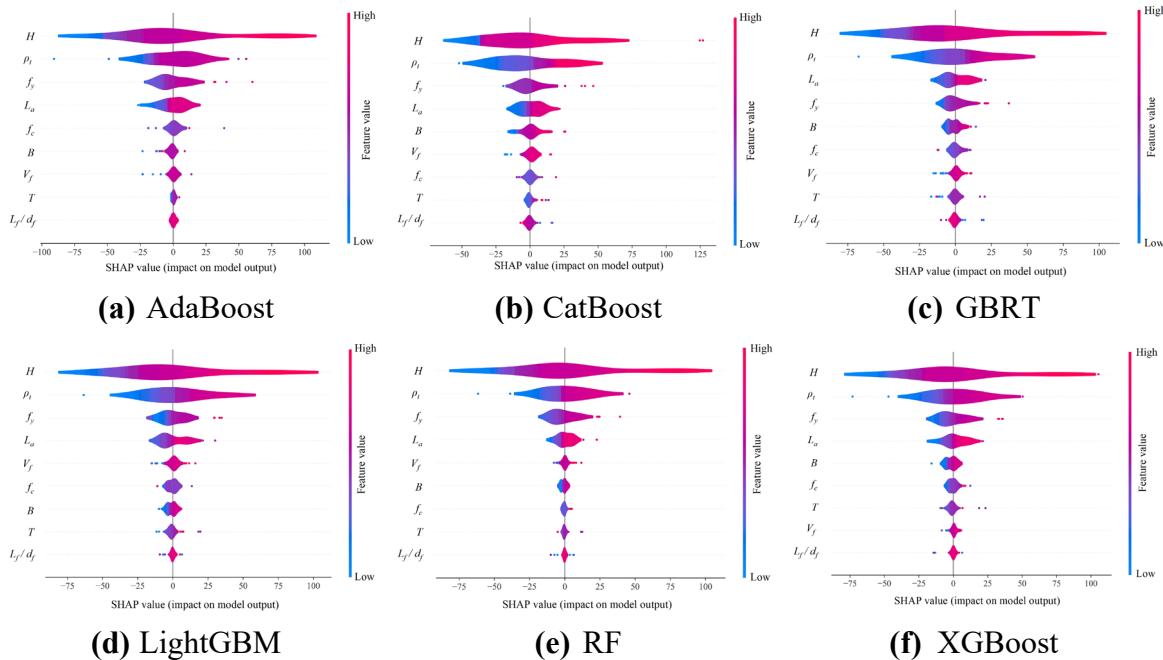


Figure 19. SHAP summary plots of the six ensemble learning models.

Figure 20 presents the SHAP bar plot for the six ensemble learning models, illustrating the average impact magnitude of each feature on the predictions of those models. The CatBoost model features a relatively more uniform distribution of SHAP values across all features, suggesting a

balanced consideration in the decision-making process without excessive dependence on specific features. This equilibrium potentially contributes to the superior performance and stability of the model, explaining its consistent performance across various data subsets among ten ML models analyzed. In contrast, the SHAP bar plots for other ensemble learning models reveal that some features have significantly higher SHAP values, indicating a stronger reliance on particular features, such as in the AdaBoost, RF, and XGBoost models. This dependency might result in fluctuating model performance across different data subsets.

While an interpretation of SHAP values for a single model may not fully delve into the importance and impact of features on predictions, the analysis of SHAP interpretations across multiple models provides a more comprehensive and accurate understanding of feature importance and interactions. This approach facilitates model selection, optimization, and interpretation analysis with a solid theoretical foundation and practical insights.

The purpose of key feature interpretation is to clarify the explanation of how the importance of features varies across different models and how SHAP analysis can provide deeper insights into model behavior and feature impact, thereby supporting more informed decision making in predictive modelling.

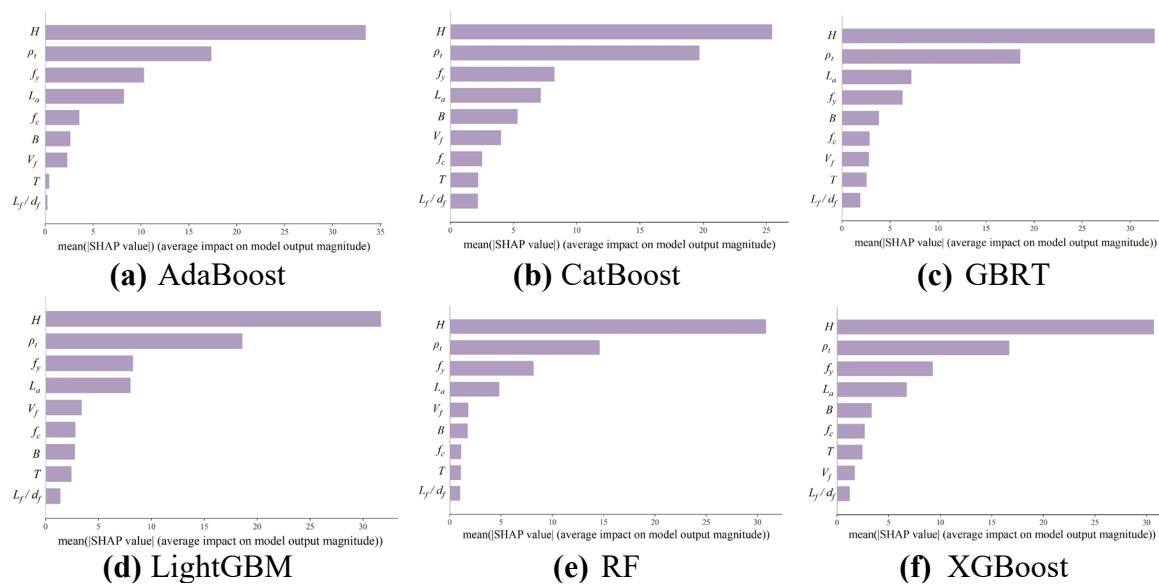


Figure 20. Feature importance of the six ensemble learning models based on SHAP.

5.2. Key Feature Interpretation

The improvement of the transparency and interpretability of ML models is essential for the understanding of their decision processes. To assess the influence of features in the CatBoost model, visualization techniques are applied. A uniform distribution of SHAP values across features suggests a balanced influence, which leads to a focused analysis of the top five most influential features. This approach highlights the key features that drive the predictions of the model [101]. Feature normalization, achieved by subtracting the mean and dividing by the standard deviation, is used to ensure uniform scaling, stabilize the model and accelerate convergence.

$$\mathbb{Z} = \frac{X - \mu}{\sigma} \quad (18)$$

where X is the raw data, μ is the mean, σ is the standard deviation, and \mathbb{Z} is the normalized data. The objective of this normalization process is to neutralize the scaling differences among various features, thereby enhancing the stability of the model during its training phase and allowing for faster convergence of the algorithm.

The normalized SHAP values of the CatBoost model are represented in **Figure 21**. The analysis of normalized SHAP value reveals a prevailing trend that increases in the eigenvalues of H , ρ_t , B , and L_a are associated with increases in the normalized SHAP values. This pattern suggests that increased eigenvalues of these features substantially increase flexural ultimate capacity M_u , as reflected in the greater SHAP values. This notable positive correlation, particularly evident for parameters of H , ρ_t , B , and L_a , highlights their considerable influence on M_u . Nevertheless, the feature f_y shows a clearly nonlinear relationship with flexural ultimate capacity M_u . It is attributed to the ability of the CatBoost model to capture the intricate interactions and nonlinear dynamics between the features. In the context of reinforced concrete beams, the steel reinforcement and the surrounding concrete work together to resist bending moments. The nonlinear influence of f_y is partly due to the complex stress-strain behavior of concrete, in particular its tendency to crack in tension and its ultimate compressive strength limit. An observed increase in the variability of normalized SHAP values for the yielding strength of reinforcement f_y deviating from zero—especially towards positive values—suggests a pronounced influence of reinforcement strength f_y on flexural ultimate capacity M_u in these regions. Deviating from a straightforward linear relationship, the contribution of f_y in increasing flexural ultimate capacity M_u is further evaluated by the distribution and depth of steel reinforcements within the concrete cross-section. As such, the feature of yielding strength of reinforcement f_y , especially in certain regions, deserves a more in-depth analysis.

An interesting observation from **Figure 21(d)** is the prevalence of more red dots at higher values of ρ_t , which suggests a simultaneous increase in H values in these areas, potentially amplifying their influence on flexural ultimate capacity M_u . Thus, when evaluating the effect of ρ_t on flexural ultimate capacity M_u , it is crucial to consider the interactions with other features. The increases in the values of L_a , and H are significantly beneficial to flexural ultimate capacity M_u , while the effect of f_y is nonlinear and more pronounced in certain regions. Furthermore, the interaction of H with high values of ρ_t deserves special attention. These findings allow for a deeper understanding of how each feature contributes to M_u , thus laying a foundation for more accurate model optimization and feature engineering strategies.

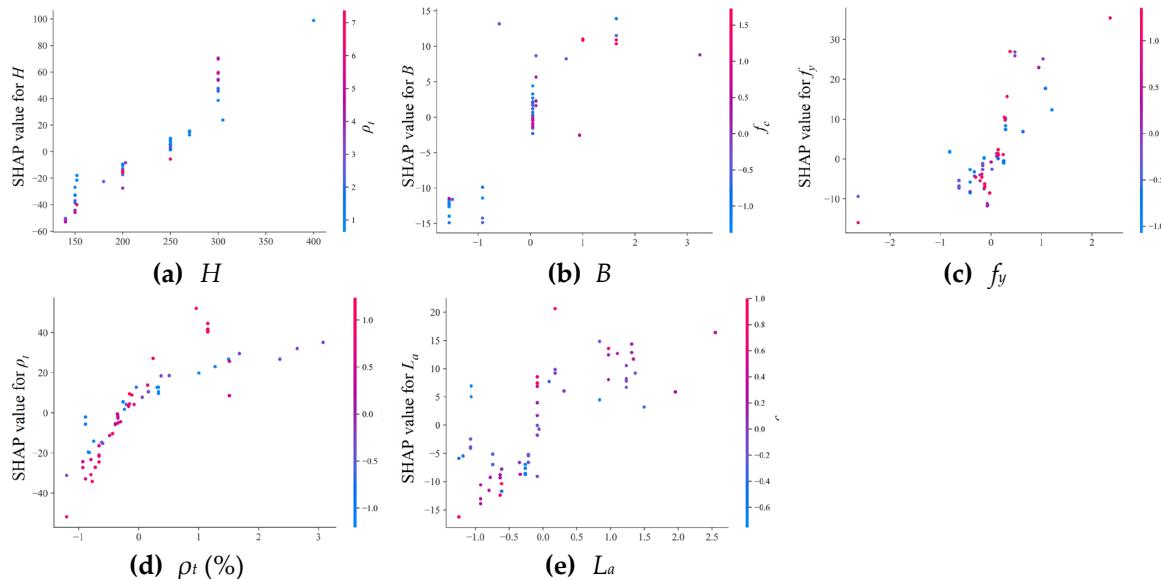


Figure 21. SHAP dependency plots for 5 critical features in the CatBoost model.

6. Conclusions

A comprehensive review of existing literature is conducted to initially establish a database of flexural ultimate capacity for reinforced UHPC composed of measured results of 339 UHPC-based specimens with varying design parameters from 56 different experimental investigations. Ten ML algorithms including ANN, SVR, K-NN, CART, RF, AdaBoost, GBRT, LightGBM, CatBoost and

XGBoost are then employed to develop optimized models for predicting the flexural ultimate capacity M_u of reinforced UHPC specimens. Four statistical indicators of model performance are utilized to evaluate the prediction results of ML-based models presented. Moreover, the performance of ML-based models with different data subset sizes is analyzed to thoroughly assess qualities of model and database given. A comparison analysis is conducted between the calculated values of from the several existing empirical formulas and the predicted values of the CatBoost model. The SHAP method is finally used to interpret the ML models, validating their reliability and examining the effect of each feature on the prediction results. The following conclusions can be drawn:

1. For model training, the models of CART, RF, XGBoost, CatBoost, AdaBoost, GBRT, and LightGBM show superior model performance, characterized by higher R^2 values and lower values of RMSE, MAE and MAPE compared to other ML-based models. The top three models for predicting the flexural ultimate capacity of reinforced UHPC beams are ranked in the order of GBRT > CatBoost > LightGBM. On the testing set, the GBRT model perform the best prediction results with an R^2 of 0.95, RMSE of 13.3, MAE of 9.3 and MAPE of 18%, respectively. On the contrary, the ANN model is found to perform the least effectively. Overall, the developed ML models accurately predict the flexural ultimate capacity M_u of reinforced UHPC beams after optimization, with ensemble learning models typically providing a higher level of accuracy than traditional individual ML models.
2. An in-depth database subset analysis reveals that the second case is the most efficient data subset configuration for ensemble learning models with 75% and the remaining 25% of the data divided as the training set and testing set. For traditional ML-based models, the optimal data subset configuration is identified in the case 3, where the data distribution percentages of the training set and testing set are 80% and 20%, respectively. This underscores the significant impact of the database division ratios of training-to-testing set on the effectiveness of performance prediction for the ML models. Among the ensemble learning models, CatBoost and GBRT show a remarkable consistency with a varying data subset configuration, while CatBoost model presents a distinguished performance prediction using a superior division strategy of data subsets. These insights highlight the importance of proper data acquisition in improving model performance prediction, providing crucial guidance for future model selection and optimization. The effects of database subset distribution on the performance prediction and model stability should be properly considered during the model selection and optimization process to ensure optimal model functionality in real-world applications.
3. The CatBoost model demonstrates superior performance in terms of predictive accuracy, as evidenced by its highest R^2 value of 0.993, the lowest RMSE value of 4.396, lowest MAE value of 2.055, and lowest MAPE value of 3.704%. This substantial improvement in performance prediction of flexural capacity for reinforced UHPC beams is particularly notable when compared to existing empirical methods. Notably, the model's enhanced applicability and adaptability enable it to handle complex feature interactions, leading to highly reliable results. This renders the EL model a potentially invaluable tool for practical engineering applications. The employment of data-driven methodologies by the CatBoost model presents a promising alternative to existing empirical methods, thereby paving the way for enhanced accuracy and efficiency in the field of UHPC structural design.
4. The SHAP-based feature importance analysis indicates that ρ_t , L_a , and H are the most critical features for the determination of the flexural ultimate capacity M_u of reinforced UHPC beams. This finding is consistent across ML models including RF, XGBoost, CatBoost, GBRT, AdaBoost, and LightGBM. In contrast, the ratio of fiber length-to-diameter (L_f/d_f) are the least important characteristics. The CatBoost model displays a more uniform distribution of SHAP values for all parameters, suggesting a balanced decision-making process and contributing to its superior and stable model performance. Several ensemble learning models such as AdaBoost, RF, and XGBoost show higher SHAP values for certain features, indicating a greater dependence on those features and potentially leading to more variable performance for different subsets of database.
5. The analysis of normalized SHAP value reveals a prevailing trend that increases in the eigenvalues of B , ρ_t , L_a and H are associated with the growing of the normalized SHAP values. This notable positive correlation highlights their substantial influence on predicting the flexural

ultimate capacity M_u of reinforced UHPC beams. Nevertheless, the yield strength of longitudinal reinforcement f_y shows a clearly nonlinear relationship with ultimate capacity M_u . Furthermore, the interaction of moment of inertia I with high values of reinforcement ratio ρ_t deserves special attention. These findings allow for a deeper understanding of how each feature contributes to the prediction of flexural ultimate capacity M_u of reinforced UHPC beams, thus laying a foundation for more accurate model optimization and feature engineering strategies.

Author Contributions: Conceptualization, methodology, funding acquisition, Z.Z.; investigation, visualization, writing – original draft, X.Z.; supervision, writing – review & editing, Z.P.; software, data curation, Z.L.; investigation, formal analysis, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Hunan Province, China (Grant No. 2023JJ30216), the Research Foundation of Education Department of Hunan Province, China (Grant No. 23B0576), the National Natural Science Foundation of China (Grant No. 51808212), and Natural Science Foundation of Chongqing, China (CSTB2024NSCQ-MSX1206), respectively. The authors would like to express their gratitude for the financial supports.

Data Availability Statement: Data are contained within the article.

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

References

1. Naaman, A.E.; Wille, K. The Path to Ultra-High Performance Fiber Reinforced Concrete (UHP-FRC): Five Decades of Progress. *Proceedings of Hipermat* **2012**, 3–13.
2. Richard, P.; Cheyrez, M. Composition of Reactive Powder Concretes. *Cem. Concr. Res.* **1995**, *25*, 1501–1511, doi:10.1016/0008-8846(95)00144-2.
3. Russell, H.G.; Graybeal, B.A.; Russell, H.G. *Ultra-High Performance Concrete: A State-of-the-Art Report for the Bridge Community*; United States. Federal Highway Administration. Office of Infrastructure Research and Development, 2013;
4. Voo, Y.; Foster, S.; Voo, C. Ultrahigh-Performance Concrete Segmental Bridge Technology: Toward Sustainable Bridge Construction. *Journal of Bridge Engineering* **2014**, *20*, doi:10.1061/(ASCE)BE.1943-5592.0000704.
5. Yoo, D.-Y.; Yoon, Y.-S. A Review on Structural Behavior, Design, and Application of Ultra-High-Performance Fiber-Reinforced Concrete. *International Journal of Concrete Structures and Materials* **2016**, *10*, 125–142, doi:10.1007/s40069-016-0143-x.
6. Qiu, M.; Hu, Y.; Shao, X.; Zhu, Y.; Li, P.; Li, X. Experimental Investigation on Flexural and Ductile Behaviors of Rebar-reinforced Ultra-high-performance Concrete Beams. *Struct. Concrete* **2022**, *23*, 1533–1554, doi:10.1002/suco.202100794.
7. Akeed, M.H.; Qadi, S.; Ahmed, H.U.; Faraj, R.H.; Mohammed, A.S.; Emad, W.; Tayeh, B.A.; Azevedo, A.R.G. Ultra-High-Performance Fiber-Reinforced Concrete. Part I: Developments, Principles, Raw Materials. *Case Stud. Constr. Mater.* **2022**, *17*, e01290, doi:10.1016/j.cscm.2022.e01290.
8. Du, J.; Meng, W.; Khayat, K.H.; Bao, Y.; Guo, P.; Lyu, Z.; Abu-obeidah, A.; Nassif, H.; Wang, H. New Development of Ultra-High-Performance Concrete (UHPC). *Composites Part B: Engineering* **2021**, *224*, 109220, doi:10.1016/j.compositesb.2021.109220.
9. Compressive Behavior of Ultra-High-Performance Fiber-Reinforced Concrete. *Aci Mater. J.* **2007**, *104*, doi:10.14359/18577.
10. Habel, K.; Denarie, E.; Brühwiler, E. Experimental Investigation of Composite Ultra-High-Performance Fiber-Reinforced Concrete and Conventional Concrete Members. *ACI Structural Journal* **2007**, *104*.
11. Graybeal, B.A.; Baby, F. Development of Direct Tension Test Method for Ultra-High- Performance Fiber-Reinforced Concrete. *Aci Mater. J.* **2013**.
12. Pourbaba, M.; Sadaghian, H.; Mirmiran, A. A Comparative Study of Flexural and Shear Behavior of Ultra-High-Performance Fiber-Reinforced Concrete Beams. *Adv. Struct. Eng.* **2019**, *22*, 1727–1738, doi:10.1177/1369433218823848.

13. Shafieifar, M.; Farzad, M.; Azizinamini, A. A Comparison of Existing Analytical Methods to Predict the Flexural Capacity of Ultra High Performance Concrete (UHPC) Beams. *Constr. Build. Mater.* **2018**, *172*, 10–18, doi:10.1016/j.conbuildmat.2018.03.229.
14. Yang, I.H.; Joh, C.; Kim, B.-S. Structural Behavior of Ultra High Performance Concrete Beams Subjected to Bending. *Eng. Struct.* **2010**, *32*, 3478–3487, doi:10.1016/j.engstruct.2010.07.017.
15. Zhang, Y.; Zhu, Y.; Qiu, J.; Hou, C.; Huang, J. Impact of Reinforcing Ratio and Fiber Volume on Flexural Hardening Behavior of Steel Reinforced UHPC Beams. *Eng. Struct.* **2023**, *285*, 116067, doi:10.1016/j.engstruct.2023.116067.
16. Guo, Y.-Q.; Wang, J.-Y. Flexural Behavior of High-Strength Steel Bar Reinforced UHPC Beams with Considering Restrained Shrinkage. *Constr. Build. Mater.* **2023**, *409*, 133802, doi:10.1016/j.conbuildmat.2023.133802.
17. Kamal, M.M.; Safan, M.A.; Etman, Z.A.; Salama, R.A. Behavior and Strength of Beams Cast with Ultra High Strength Concrete Containing Different Types of Fibers. *HBRC Journal* **2014**, *10*, 55–63, doi:10.1016/j.hbrcj.2013.09.008.
18. Yang, I.H.; Joh, C.; Kim, B.-S. Structural Behavior of Ultra High Performance Concrete Beams Subjected to Bending. *Engineering Structures* **2010**, *32*, 3478–3487, doi:10.1016/j.engstruct.2010.07.017.
19. Singh, M.; Sheikh, A.H.; Mohamed Ali, M.S.; Visintin, P.; Griffith, M.C. Experimental and Numerical Study of the Flexural Behaviour of Ultra-High Performance Fibre Reinforced Concrete Beams. *Constr. Build. Mater.* **2017**, *138*, 12–25, doi:10.1016/j.conbuildmat.2017.02.002.
20. Hasgul, U.; Turker, K.; Birol, T.; Yavas, A. Flexural Behavior of Ultra-high-performance Fiber Reinforced Concrete Beams with Low and High Reinforcement Ratios. *Struct. Concrete* **2018**, *19*, 1577–1590, doi:10.1002/suco.201700089.
21. Yoo, D.-Y.; Banthia, N.; Yoon, Y.-S. Experimental and Numerical Study on Flexural Behavior of Ultra-High-Performance Fiber-Reinforced Concrete Beams with Low Reinforcement Ratios. *Canadian Journal of Civil Engineering* **2017**, *44*, 18–28, doi:10.1139/cjce-2015-0384.
22. Katlav, M.; Turk, K.; Turgut, P. Research into Effect of Hybrid Steel Fibers on the V-Shaped RC Folded Plate Thickness. *Structures* **2022**, *44*, 665–679, doi:10.1016/j.istruc.2022.08.046.
23. *Building Code Requirements for Structural Concrete (ACI 318-14) [and] Commentary on Building Code Requirements for Structural Concrete (ACI 318R-14)*; 2014; ISBN 978-0-87031-930-3.
24. Farouk, A.I.B.; Jinsong, Z. Prediction of Interface Bond Strength between Ultra-High-Performance Concrete (UHPC) and Normal Strength Concrete (NSC) Using a Machine Learning Approach. *Arab. J. Sci. Eng.* **2022**, *47*, 5337–5363, doi:10.1007/s13369-021-06433-6.
25. Aaleti, S.; Petersen, B.; Sritharan, S. *Design Guide for Precast UHPC Waffle Deck Panel System, Including Connections*; 2014; ISBN ISBN-10: 1494894386 ISBN-13: 978-1494894382.
26. Design Considerations for Steel Fiber Reinforced Concrete. *SJ* **1988**, *85*, doi:10.14359/3144.
27. Salehi, H.; Burgueño, R. Emerging Artificial Intelligence Methods in Structural Engineering. *Eng. Struct.* **2018**, *171*, 170–189, doi:10.1016/j.engstruct.2018.05.084.
28. Sun, C.; Wang, K.; Liu, Q.; Wang, P.; Pan, F. Machine-Learning-Based Comprehensive Properties Prediction and Mixture Design Optimization of Ultra-High-Performance Concrete. *Sustainability* **2023**, *15*, 15338, doi:10.3390/su152115338.
29. Fu, B.; Feng, D.-C. A Machine Learning-Based Time-Dependent Shear Strength Model for Corroded Reinforced Concrete Beams. *Journal of Building Engineering* **2021**, *36*, 102118, doi:10.1016/j.jobe.2020.102118.
30. Cakiroglu, C.; Aydin, Y.; Bekdaş, G.; Geem, Z.W. Interpretable Predictive Modelling of Basalt Fiber Reinforced Concrete Splitting Tensile Strength Using Ensemble Machine Learning Methods and SHAP Approach. *Materials* **2023**, *16*, 4578, doi:10.3390/ma16134578.
31. Sun, G.; Du, M.; Shan, B.; Shi, J.; Qu, Y. Ultra-High Performance Concrete Design Method Based on Machine Learning Model and Steel Slag Powder. *Case Stud. Constr. Mater.* **2022**, *17*, e01682, doi:10.1016/j.cscm.2022.e01682.
32. Feng, D.-C.; Wang, W.-J.; Mangalathu, S.; Hu, G.; Wu, T. Implementing Ensemble Learning Methods to Predict the Shear Strength of RC Deep Beams with/without Web Reinforcements. *Engineering Structures* **2021**, *235*, 111979, doi:10.1016/j.engstruct.2021.111979.

33. Solhmirzaei, R.; Salehi, H.; Kodur, V. Predicting Flexural Capacity of Ultrahigh-Performance Concrete Beams: Machine Learning-Based Approach. *J. Struct. Eng.* **2022**, *148*, 04022031, doi:10.1061/(ASCE)ST.1943-541X.0003320.

34. Ergen, F.; Katlav, M. Machine and Deep Learning-Based Prediction of Flexural Moment Capacity of Ultra-High Performance Concrete Beams with/out Steel Fiber. *Asian Journal of Civil Engineering* **2024**, *25*, 4541–4562, doi:10.1007/s42107-024-01064-2.

35. Ye, M.; Li, L.; Yoo, D.-Y.; Li, H.; Zhou, C.; Shao, X. Prediction of Shear Strength in UHPC Beams Using Machine Learning-Based Models and SHAP Interpretation. *Construction and Building Materials* **2023**, *408*, 133752, doi:10.1016/j.conbuildmat.2023.133752.

36. *National Addition to Eurocode 2-Design of Concrete Structures: Specific Rules for Ultra-High Performance Fibre-Reinforced Concrete (UHPFRC)*. Association Francaise de Normalisation; 2016;

37. Epfl, M. Recommendation: Ultra-High Performance Fibre Reinforced Cement-Based Composites (UHPFRC) 2016.

38. Conittee, A.C.I. Design Considerations for Steel Fiber Reinforced Concrete. *ACI Structural Journal* **1988**, *563–580*.

39. Aaleti, S.; Petersen, B.; Sritharan, S. *Design Guide for Precast UHPC Waffle Deck Panel System*; 2013;

40. Li, L. Mechanical Behavior and Design Method for Reactive Powder Concrete Beams. PhD dissertation, Harbin Institute of Technology, 2010.

41. Mangalathu, S.; Hwang, S.-H.; Jeon, J.-S. Failure Mode and Effects Analysis of RC Members Based on Machine-Learning-Based SHapley Additive exPlanations (SHAP) Approach. *Eng. Struct.* **2020**, *219*, 110927, doi:10.1016/j.engstruct.2020.110927.

42. Gu, J.-B.; Wang, J.-Y.; Lu, W. An Experimental Assessment of Ultra High Performance Concrete Beam Reinforced with Negative Poisson's Ratio (NPR) Steel Rebar. *Constr. Build. Mater.* **2022**, *327*, 127042, doi:10.1016/j.conbuildmat.2022.127042.

43. Huang, J.; He, Z.; Khan, M.B.E.; Zheng, X.; Luo, Z. Flexural Behaviour and Evaluation of Ultra-High-Performance Fibre Reinforced Concrete Beams Cured at Room Temperature. *Sci. Rep.* **2021**, *11*, 19069, doi:10.1038/s41598-021-98502-x.

44. Yin, H.; Shirai, K.; Teo, W. Finite Element Modelling to Predict the Flexural Behaviour of Ultra-High Performance Concrete Members. *Eng. Struct.* **2019**, *183*, 741–755, doi:10.1016/j.engstruct.2019.01.046.

45. Pourbaba, M.; Joghataie, A.; Mirmiran, A. Shear Behavior of Ultra-High Performance Concrete. *Constr. Build. Mater.* **2018**, *183*, 554–564, doi:10.1016/j.conbuildmat.2018.06.117.

46. Kodur, V.; Solhmirzaei, R.; Agrawal, A.; Aziz, E.M.; Soroushian, P. Analysis of Flexural and Shear Resistance of Ultra High Performance Fiber Reinforced Concrete Beams without Stirrups. *Eng. Struct.* **2018**, *174*, 873–884, doi:10.1016/j.engstruct.2018.08.010.

47. Yoo, D.-Y.; Yoon, Y.-S. Structural Performance of Ultra-High-Performance Concrete Beams with Different Steel Fibers. *Engineering Structures* **2015**, *102*, 409–423, doi:10.1016/j.engstruct.2015.08.029.

48. Bao-chun, C.; Huang, Q.-W. Study of Steel Fiber Content Influence on Flexural Behavior of R-UHPC Beam. *Journal of Ningxia University (Natural Science Edition)* **2019**, *40*, 130–136.

49. Stürwald, S. Versuche Zum Biegetragverhalten von UHPC Mit Kombinierter Bewehrung. **2011**.

50. Yavaş, A.; Hasgul, U.; Turker, K.; Birol, T. Effective Fiber Type Investigation on the Shear Behavior of Ultrahigh-Performance Fiber-Reinforced Concrete Beams. *Adv. Struct. Eng.* **2019**, *22*, 1591–1605, doi:10.1177/1369433218820788.

51. Wenzhong, Z.; Li, L.; Shanshan, L. Experimental Research on Mechanical Performance of Normal Section of Reinforced Reactive Powder Concrete Beam. *Journal of Building Structures* **2011**, *32*, 125.

52. Visage, E.T.; Perera, K.; Weldon, B.D.; Jauregui, D.V.; Newtson, C.M.; Guaderrama, L. Experimental and Analytical Analysis of the Flexural Behavior of UHPC Beams. *Proceedings of Hipermat* **2012**, 403–410.

53. Yang, I.-H.; Joh, C.; Bui, T.Q. Estimating the Tensile Strength of Ultrahigh-Performance Fiber-Reinforced Concrete Beams. *Advances in Materials Science and Engineering* **2019**, *2019*, 1–16, doi:10.1155/2019/5128029.

54. Wahba, K.; Marzouk, H.; Dawood, N. Structural Behavior of UHPFRC Beams without Stirrups. *Proceedings, Annual Conference - Canadian Society for Civil Engineering* **2012**, *3*, 2487–2496.

55. Liang, X.-W.; Wang, P.; Xu, M.-X.; Wang, Z.-Y.; Yu, J.; Li, L. Investigation on Flexurel Capacity of Reinforced Ultra High Performance Concrete Beams. *Engineering Mechanics* **2019**, *36*, 110–119.

56. Khalil, W.; Tayfur, Y. Flexural Strength of Fibrous Ultra High Performance Reinforced Concrete Beams. *ARPN Journal of Engineering and Applied Sciences* **2013**, *8*, 200–214.

57. Yang, I.-H.; Park, J.; Bui, T.Q.; Kim, K.-C.; Joh, C.; Lee, H. An Experimental Study on the Ductility and Flexural Toughness of Ultrahigh-Performance Concrete Beams Subjected to Bending. *Materials* **2020**, *13*, 2225, doi:10.3390/ma13102225.

58. Rndl, N.; Simon, C.; Mészöly, T. *Experimental Investigations on UHP (FR) C Beams with High Strength Reinforcement*; RILEM-fib-AFGC International Symposium on Ultra-High Performance Fibre-Reinforced Concrete; RILEM Marseilles, France, 2013;

59. Wang, J.; Qi, J.; Liu, J. Flexural Analysis of UHPC Beams Based on a Mesoscale Constitutive Model. *Journal of Building Structures* **2020**, *41*, 137–144, doi:10.14006/j.jzjgxb.2018.0665.

60. Bae, B.I.; Choi, H.K.; Choi, C.S. Flexural and Shear Capacity Evaluation of Reinforced Ultra-High Strength Concrete Members with Steel Rebars. *Key Eng. Mater.* **2013**, *577*–578, 17–20, doi:10.4028/www.scientific.net/KEM.577-578.17.

61. Long, P.; Huang, L.; Qiao, H. RPC Constitutive Relation and Ultimate Flexural Capacity of Rectangular RPC Beams. *China Concrete and Cement Products* **2020**.

62. Kareem, R.R.; Deyab, H.M. Flexural Action of Continuous Reinforced Reactive Powder Concrete Beams. *Iop Conf. Ser.: Mater. Sci. Eng.* **2020**, *888*, 012041, doi:10.1088/1757-899X/888/1/012041.

63. Deng, Z.-C.; Wang, Y.-C.; Xiao, R.; Lan, M.; Chen, X. Flexural Test and Theoretical Analysis of UHPC Beams with High Strength Rebars. *J. Basic Sci. Eng* **2015**, *23*, 68–78.

64. Ridha, M.M.S.; Al-Shaabaf, I.A.S.; Sarsam, K.F. Experimental Study on Shear Resistance of Reactive Powder Concrete Beams without Stirrups. *Mech. Adv. Mater. Struct.* **2020**, *27*, 1006–1018, doi:10.1080/15376494.2018.1504258.

65. Xinyue, W. Research on Mechanical Properties of Reinforced Ultra-High Performance Concrete Beams. PhD Thesis, Xi'an University of Architecture and Technology Xi'an, China, 2021.

66. Lingzhi, J.; Lai, H.; Xinke, W. Experimental Study on Flexural Property of Reactive Powder Concrete Beams with HRB500 Steel. *Building Structure* **2015**, *45*, 87–92.

67. Feng, Z.; Li, C.; Yoo, D.-Y.; Pan, R.; He, J.; Ke, L. Flexural and Cracking Behaviors of Reinforced UHPC Beams with Various Reinforcement Ratios and Fiber Contents. *Eng. Struct.* **2021**, *248*, 113266, doi:10.1016/j.engstruct.2021.113266.

68. Gao, Y.; Zhu, W.; Luo, Y. Research on bending behavior of reinforced ultra-high performance concrete (UHPC) beams. *China Concrete and Cement Products* **2021**, *000*, 67–70.

69. Bae, B.-I.; Choi, H.-K.; Choi, C.-S. Flexural Strength Evaluation of Reinforced Concrete Members with Ultra High Performance Concrete. *Advances in Materials Science and Engineering* **2016**, *2016*, 1–10, doi:10.1155/2016/2815247.

70. Kahanji, C.; Ali, F.; Nadjai, A. Structural Performance of Ultra-high-performance Fiber-reinforced Concrete Beams. *Struct. Concrete* **2017**, *18*, 249–258, doi:10.1002/suco.201600006.

71. Khan, M.I.; Fares, G.; Abbas, Y.M. Behavior of Non-Shear-Strengthened UHPC Beams under Flexural Loading: Influence of Reinforcement Depth. *Applied Sciences* **2021**, *11*, 11168, doi:10.3390/app112311168.

72. Bae, B.-I.; Lee, M.-S.; Choi, C.-S.; Jung, H.-S.; Choi, H.-K. Evaluation of the Ultimate Strength of the Ultra-High-Performance Fiber-Reinforced Concrete Beams. *Applied Sciences* **2021**, *11*, 2951, doi:10.3390/app11072951.

73. Li, Y.; Guertin-Normoyle, C.; Algassem, O.; Aoude, H. *Effect of Ultra-High Performance Fibre-Reinforced Concrete and High-Strength Steel on the Flexural Behavior of Reinforced Concrete Beams*; Proceedings of the RILEM-fib-International Symposium on Ultra-High Performance Fibre-Reinforced Concrete (UHPFRC 2017), Montpellier, France; 2017;

74. Khan, M.I.; Fares, G.; Abbas, Y.M.; Alqahtani, F.K. Behavior of Non-Shear-Strengthened UHPC Beams under Flexural Loading: Influence of Reinforcement Percentage. *Applied Sciences* **2021**, *11*, 11346, doi:10.3390/app112311346.

75. Su, J.; Fu, Y.; Huang, Q. Experimental Study and Finite Element Analysis of Flexural Behavior of Reinforced Ultra-High Performance Concrete Beams. *Journal of China & Foreign Highway* **2017**, *37*, 99–105, doi:10.14048/j.issn.1671-2579.2017.06.022.

76. Zhong, H.; Zheng, X.; Song, Z. Experimental Study on the Effect of Reinforcement Ratio on the Flexural Capacity of UHPC Beams. *Henan Science* **2021**, *39*, 595–603, doi:10.3969/j.issn.1004-3918.2021.04.011.

77. Chen, S.; Zhang, R.; Jia, L.-J.; Wang, J.-Y. Flexural Behaviour of Rebar-Reinforced Ultra-High-Performance Concrete Beams. *Mag. Concr. Res.* **2018**, *70*, 997–1015, doi:10.1680/jmacr.17.00283.

78. Qiu, M. Study on the Basic Performance and Calculation Theory of Reinforced UHPC Members. PhD dissertation, Hunan University, 2021.

79. Ma, H. The Calculating Method of the Crack Width of Flexural Members of UHPC Beams with Ultra-High-Strength Steel Bars. MS thesis, Beijing Jiaotong University, 2022.

80. Liu, Y. Study on Flexural Behavior of Ultra High Performance Concrete Beams with HRB600 Reinforcement. MS thesis, Beijing Jiaotong University, 2022.

81. Smarzewski, P. Hybrid Fibres as Shear Reinforcement in High-Performance Concrete Beams with and without Openings. *Applied Sciences* **2018**, *8*, 2070, doi:10.3390/app8112070.

82. Wang, K.; Gao, L.; Wang, L. Comparison of Methods to Calculate Flexural Capacity of Reinforced UHPC Girder in Chinese, French and Swiss Specifications. *World Bridges* **2023**, *51*, 7–13.

83. Chen, B.; Wu, Q.; Huang, Q.; Ma, X.; Su, J. Experimental Study on Shear Behavior of Reinforced Ultra-High Performance Concrete Beams. *J. Fuzhou Univ* **2018**, *46*, 512–517.

84. Sun, M. Study on Flexural Behavior and Stability Performance of High-Strength Reinforced Reactive Powder Concrete Members. PhD dissertation, Beijing Jiaotong University, 2018.

85. Zhang, J.; Zhao, X.; Rong, X. Flexural Experimental Study and Capacity of Ultra High Strength Bar Reinforced UHPC Beams. *Earthquake Engineering and Engineering Dynamics* **2023**, *43*, 57–66.

86. Pourbaba, M.; Sadaghian, H.; Mirmiran, A. Flexural Response of UHPFRC Beams Reinforced with Steel Rebars. *Advances in Civil Engineering Materials* **2019**, *8*, 411–430, doi:10.1520/ACEM20190129.

87. Hou, C. Experimental and Theoretical Study on Flexural Behavior of Ultra High Performance Concrete (UHPC) Rectangular Beams. MS thesis, Hunan University, 2023.

88. Thai, H.-T. Machine Learning for Structural Engineering: A State-of-the-Art Review. *Structures* **2022**, *38*, 448–491, doi:10.1016/j.istruc.2022.02.003.

89. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A Training Algorithm for Optimal Margin Classifiers. In Proceedings of the Proceedings of the fifth annual workshop on Computational learning theory; ACM: Pittsburgh Pennsylvania USA, July 1992; pp. 144–152.

90. Fix, E.; Hodges, J.L. Discriminatory Analysis. *Nonparametric discrimination: Small sample performance. Report A* **1951**, *193008*.

91. Breiman, L.; Friedman, J.; Olshen, R.; Storne, C. Classification and Regression Trees. Belmont, CA: Wadsworth International Group; 1984. *Google Scholar* **1984**.

92. Zhou, Z.-H. *Ensemble Methods: Foundations and Algorithms*; CRC press, 2012;

93. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. System Sci.* **1997**, *55*, 119–139, doi:10.1006/jcss.1997.1504.

94. Breiman, L. Random Forests. *Machine learning* **2001**, *45*, 5–32.

95. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of statistics* **2001**, 1189–1232.

96. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; ACM: San Francisco California USA, August 13 2016; pp. 785–794.

97. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A Highly Efficient Gradient Boosting Decision Tree. *Advances in neural information processing systems* **2017**, *30*.

98. Bergstra, J.; Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.

99. Solhmirzaei, R.; Salehi, H.; Kodur, V.; Naser, M.Z. Machine Learning Framework for Predicting Failure Mode and Shear Capacity of Ultra High Performance Concrete Beams. *Eng. Struct.* **2020**, *224*.

100. Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In Proceedings of the International Joint Conference on Artificial Intelligence; January 1995; Vol. 14.
101. Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions 2017.
102. Xu, J.-G.; Chen, S.-Z.; Xu, W.-J.; Shen, Z.-S. Concrete-to-Concrete Interface Shear Strength Prediction Based on Explainable Extreme Gradient Boosting Approach. *Constr. Build. Mater.* **2021**, *308*, 125088, doi:10.1016/j.conbuildmat.2021.125088.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.