

## Research Article

# Shear Strength of Internal Reinforced Concrete Beam-Column Joints: Intelligent Modeling Approach and Sensitivity Analysis

De-Cheng Feng <sup>1</sup> and Bo Fu <sup>2,3</sup>

<sup>1</sup>Key Laboratory of Concrete and Prestressed Concrete Structures of the Ministry of Education, Southeast University, Nanjing 211189, China

<sup>2</sup>School of Civil Engineering, Chang'an University, Xi'an 710061, China

<sup>3</sup>State Key Laboratory of Green Building in Western China, Xi'an University of Architecture and Technology, Xi'an 710055, China

Correspondence should be addressed to Bo Fu; 90\_bofu@chd.edu.cn

Received 1 June 2020; Revised 6 July 2020; Accepted 8 August 2020; Published 25 August 2020

Academic Editor: Moacir Kripka

Copyright © 2020 De-Cheng Feng and Bo Fu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, an intelligent modeling approach is presented to predict the shear strength of the internal reinforced concrete (RC) beam-column joints and used to analyze the sensitivity of the influence factors on the shear strength. The proposed approach is established based on the famous boosting-family ensemble machine learning (ML) algorithms, i.e., gradient boosting regression tree (GBRT), which generates a strong predictive model by integrating several weak predictors, which are obtained by the well-known individual ML algorithms, e.g., DT, ANN, and SVM. The strong model is boosted as each weak predictor has its own weight in the final combination according to the performance. Compared with the conventional mechanical-driven shear strength models, e.g., the well-known modified compression field theory (MCFT), the proposed model can avoid the complicated derivation process of shear mechanism and calibration of the involved empirical parameters; thus, it provides a more convenient, fast, and robust alternative way for predicting the shear strength of the internal RC joints. To train and test the GBRT model, a total of 86 internal RC joint specimens are collected from the literatures, and four traditional ML models and the MCFT model are also employed as comparisons. The results indicate that the GBRT model is superior to both the traditional ML models and MCFT model, as its degree-of-fitting is the highest and the predicting dispersion is the lowest. Finally, the model is used to investigate the influences of different parameters on the shear strength of the internal RC joint, and the sensitivity and importance of the corresponding parameters are obtained.

## 1. Introduction

Reinforced concrete (RC) beam-column joint or connection is one of the most critical and vulnerable components in RC structures. The failure of the RC beam-column joints could seriously affect the overall safety of the structures. Especially, it will suffer from the shear failure if there are insufficient transverse reinforcements and/or the material properties are deteriorated due to the aging effects. As it is known to all, shear failure is a brittle failure type without any warnings. Therefore, it is vital to accurately predict the shear strength of the RC beam-column joints to avoid shear failure in design procedures in order to ensure the safety of the structures.

In general, there are three commonly used approaches to assess shear strength of the RC joints, i.e., experimental study, numerical simulation, and theoretical analysis. The experimental study is the most direct and classical way, which can be traced back to 1970s [1]. However, it is costly in both time and money and difficult to operate. The numerical simulation, e.g., finite element method (FEM), is also widely adopted for its low cost [2, 3]. Nevertheless, it usually has several simplifications and some of the mechanisms are hard to be reflected in the FEM framework, e.g., multistress state behavior, shear behavior, and interfacial bond-slip behavior. Apart from the experimental and numerical studies, numerous theoretical models were also proposed to evaluate the performance of the RC beam-column joints, for instance,

the well-known modified compression field theory (MCFT) [4], the strut-and-tie method (STM) [5], etc. These models are actually derived based on the shear mechanisms of fundamental RC elements and can be widely used to evaluate the behavior of any type of shear-dominated RC members, including the beam-column joints [6]. A detailed review of the theoretical and empirical models for the RC joints can be found in [7].

In recent five years, there are some latest development on RC joint models. Eom et al. [8] developed an energy-based hysteresis model for RC beam-column joints by using the energy function and the existing backbone curve of ASCE/SEI 41-06 [9]. Hwang et al. [10] proposed a shear strength degradation model for performance-based design of interior beam-column joints. In the model, all possible failure mechanisms of beams and joints, including flexural yielding of the beam end, diagonal cracking and concrete crushing in the joint panel, bar bond-slip, and bar elongation, are considered. Later, Hwang and Park [11] developed design equations of the joint shear strength and hoop requirement for the performance-based design of interior RC beam-column joints by considering the diagonal strut mechanism and truss mechanism. The target drift ratio and bar bond parameters are defined as the requirements of the joint shear strength and hoop strength. More recently, Hwang and Park [12] modified the shear strength degradation model for interior RC joints and applied it to exterior RC joints with standard hooked bars. Hwang et al. [13] simplified the softened strut-and-tie model to facilitate design practice for the strength prediction of discontinuity regions such as the RC beam-column joints. The shear-resisting mechanisms as suggested by the softened strut-and-tie model are considered in the simplified model. Similarly, Huang and Kuang [14] proposed a shear strength model for exterior RC wide beam-column joints by introducing the softened strut-and-tie concept. Hassan and Moehle [15] collected a database of exterior and corner beam-column joints without transverse reinforcement. Based on the database, they evaluated several existing shear strength models and developed a strut-and-tie model based on the ACI 318 [16] strut-and-tie modeling provisions and an empirical model by considering the effects of joint aspect ratio, column axial load, and concrete compressive strength.

Although the above empirical or theoretical approaches offer simple and clear explanation of the shear mechanism, they also introduce empirical assumptions which will reduce their accuracy. Moreover, the derivations seem to be complicated since the iteration process is likely involved and some of the parameters are empirical that needed to be determined according to the users' experience.

In recent years, with the flourishing of artificial intelligence (AI), a brand new way is come to people's horizons, i.e., using machine learning (ML) techniques to predict the shear strength of the RC beam-column joints. ML is a type of AI, which has various functions, e.g., classification, regression, and clustering. ML can learn the characteristics of a certain type of data according to the existing database and then classify, summarize, and predict the things of interest. Prediction of the shear strength of the RC joints is

essentially a regression problem. There are already some successful applications of prediction using ML in structural engineering, for instances, evaluating the cement strength via fuzzy logic, artificial neural network (ANN), and gene expression programming (GEP) [17, 18], modeling the concrete properties via ANN and support vector machine (SVM) [19–23], simulating the failure of brittle anisotropic materials such as masonry via ANN [24, 25], predicting the structural member capacities via hybrid ML algorithms [26, 27], detecting the structural damage via GEP [28, 29], etc. A detailed state-of-art of the application of ML in structural engineering was summarized in [30].

However, the majority of the ML algorithms used in the previous studies were individual-type learning algorithms such as ANN family [31], SVM family [32], and decision tree (DT) family [33]. The disadvantages of the individual-type learning algorithms are instable and with low accuracy. To improve their performance, a new type of learning algorithms known as ensemble learning algorithms has been recently proposed and successfully applied in various fields. The basic idea of the ensemble learning is to combine several weak learners generated by individual learning algorithms into a strong one. In brief, the ensemble learning algorithms are more stable and accurate compared to the individual learning algorithms [34]. There are mainly two categories of ensemble learning algorithms: bagging and boosting. For the bagging family, the weak learners are produced in parallel while they are produced in sequence for the boosting family. Theoretically, bagging is more efficient and can effectively reduce the variance of the prediction, and boosting is relatively less efficient in reducing the bias. In practice, boosting is superior to bagging in terms of accuracy for general cases. Therefore, one of the most typical boosting ensemble learning algorithms referred to as gradient boosting regression tree (GBRT) [35] algorithm is used in this study.

In this paper, we aim to develop a GBRT-based intelligent method for predicting the shear strength of the RC beam-column joints and make comparisons between the proposed data-driven model and some traditional ML-based models as well as the conventional mechanical-driven MCFT model. Firstly, some individual-type ML techniques, including linear regression (LR), SVM, ANN, and DT, are briefly reviewed. Then, the mathematical background and implementation of GBRT are introduced. Afterwards, the shear strength data of 86 internal RC beam-column joints are collected from the literature. Based on the database, the prediction results from the GBRT-based model are verified by a 10-fold validation test and compared with those from the individual-type ML models. In addition, one of the representative conventional mechanical-driven approaches, i.e., MCFT, is briefly summarized and also used as comparison with the GBRT model. Finally, sensitivity analysis of input variables is conducted for the GBRT model to quantify the influences of different parameters.

## 2. Review of the Traditional ML Techniques

*2.1. Linear Regression (LR).* Linear regression (LR) is one of the most widely used statistical analysis techniques in

determining the qualitative relationship between two or more variables. In general, the least square method is adopted to solve the LR problem. If only one independent variable and one dependent variable are considered and the relationship between them is approximately linear, then this type of regression analysis is called simple linear regression (SLR). On the contrary, if two or more independent variables are included and the relationship between the independent and dependent variables are approximately linear, then this regression analysis is called multiple linear regression (MLR). For the prediction problem considered in this study, more than two input parameters should be assigned as the independent variables, so it belongs to MLR.

**2.2. Support Vector Machine (SVM).** Based on the statistical learning theory proposed by Vapnik [36], the support vector machine (SVM) is an effective optimizing tool to improve the generalization performance and obtain the globally optimal and unique solution. In implementing the SVM regression, the primary goal is to minimize an upper bound of the generalization error based on the structural risk minimization. The essence of the SVM regression is to map the input variables into a high-dimensional feature space by a nonlinear mapping and then conduct linear regression in the space.

**2.3. Artificial Neural Networks (ANN).** The artificial neural network (ANN) is a complex information processing system composed of a huge number of interconnected processing elements (neurons) arranged in layers. It is the abstraction, simplification, and simulation of the structure and mechanism of biological nervous systems such as human brains. Just as the learning process in biological systems, the ANN involves adjustments to the synaptic connections between the neurons. When it is applied to solve engineering problems, a neural network can be a vector mapper which maps input vector(s) to an output one(s).

**2.4. Decision Trees (DT).** Decision tree (DT) is one of the basic classification and regression methods. The DT regression approach mainly refers to one of the binary tree structures, i.e., classification and regression tree (CART) algorithm, in which the characteristic values of internal nodes are only yes or no. The main task for CART is to divide the characteristic space into several units. Every unit has a certain output. As each node is judged by yes or no, the divided boundary is parallel to the coordinate axis. Any testing data can be fallen into a unit according to its characteristic and thus obtain its corresponding output.

### 3. Boosted ML Approach: Gradient Boosting Regression Tree (GBRT)

Though the abovementioned traditional ML methods have already been applied in several aspects of structural engineering, including predicting the behavior of structural

members, there still exist some drawbacks. For some cases, a “best” model may not be easily obtained using those algorithms. Meanwhile, models by different algorithms will have their own hypotheses, which may lead to great model uncertainty. Therefore, this paper employs the ensemble learning algorithms to generate the predictive model for the joint shear strength. Specifically, the boosting family gradient boosting regression tree (GBRT) is adopted. The ensemble learning method is superior to the individual learning method since it offers a powerful framework to obtain a strong estimator (or learner) by integrating several weak estimators (or learners) produced by the individual learning method, so the accuracy and robustness are both enhanced. The boosting idea is reflected in the weights of the weak learners: the one with higher score will get higher weight in the final strong learner. The fundamental and theoretical backgrounds, as well as the implementation procedure, are all presented herein.

**3.1. Gradient Boosting Framework of Ensemble ML.** As mentioned before, ensemble learning is not an individual-type ML method. It is accomplished by integrating multiple weak learners into a strong one. Boosting is a major group of ensemble learning algorithms, which generates the weak learners subsequently and can be interpreted as an optimization algorithm on a suitable cost function. The basic idea of boosting is to update the weight of each weak learner by its learning error. If a weak learner has a large learning error, it will be assigned a large weight so that it could be paid more attention in the subsequent training process. Like other boosting methods, the gradient boosting integrates several weak learners into a single strong learner in an iterative way.

Supposing it requires  $M$  steps to find out the final strong learner and at the  $m$ -th ( $m \in [1, M]$ ) step we have an imperfect model  $f_m(\mathbf{x})$  which is the sum of weak learners in the previous steps,

$$f_m(\mathbf{x}) = \sum_{i=1}^{m-1} \alpha_i h_i(\mathbf{x}), \quad (1)$$

where  $\mathbf{x}$  is the vector containing the input variables;  $h_i(\mathbf{x})$  and  $\alpha_i$  are the weak learner and the corresponding weight at step  $i \in [1, m]$ .

The imperfect model can be improved by adding a new weak learner  $h_m(\mathbf{x})$  as  $f_{m+1}(\mathbf{x}) = f_m(\mathbf{x}) + h_m(\mathbf{x})$ . Then, the optimization problem becomes how to find  $h_m(\mathbf{x})$ . The solution of gradient boosting starts with the observation that a perfect  $h_m(\mathbf{x})$  would imply

$$f_{m+1}(\mathbf{x}) = f_m(\mathbf{x}) + h_m(\mathbf{x}) = y, \quad (2)$$

where  $y$  is the target output or the tested value of the output. Equation (2) can be equivalently expressed as

$$h_m(\mathbf{x}) = y - f_m(\mathbf{x}). \quad (3)$$

Therefore, in the following gradient boosting algorithm fits  $h_m(\mathbf{x})$  with the residual  $y - f_m(\mathbf{x})$ . Like other members of the boosting algorithms,  $f_{m+1}(\mathbf{x})$  is attempted to correct

the errors of its predecessor  $f_m(\mathbf{x})$ . It is observed that the residual  $y - f_m(\mathbf{x})$  is the negative gradient of the squared loss function  $1/2[y - f_m(\mathbf{x})]^2$ , so the negative gradient can be extended to other kinds of loss functions. In other words, the gradient boosting algorithm is a gradient descent algorithm, which can be generalized by varying the loss function and the gradient.

**3.2. Gradient Boosting Regression Tree (GBRT).** As can be seen in the previous section, gradient boosting is actually a framework for ensembling numerous weak learners, rather than a specific learning algorithm. Theoretically, any individual algorithms from the ANN, SVM, and DT families can be used to train the weak learners. However, unlike other boosting algorithms, the individual algorithm for training the weak learners in gradient boosting is restricted to the DT algorithms, thus it is called as GBRT. In each step (or iteration), a new DT is established by fitting the negative gradient of the loss function. The number of DT is determined by the iteration number.

The GBRT model superimposes multiple DTs and is expressed as

$$f_M(\mathbf{x}) = \sum_{m=1}^M T(\mathbf{x}; \Theta_m), \quad (4)$$

where  $T(\mathbf{x}; \Theta_m)$  represents the weak learner by DT;  $\Theta_m$  denotes the parameters of  $m$ -th DT model;  $M$  is the number of DTs, respectively.

For a dataset  $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$  where  $N$  denotes the number of the samples, the essence of training the boosting DT model is selecting the optimal parameters of DTs  $\Theta = \{\Theta_1, \Theta_2, \dots, \Theta_M\}$  to minimize the loss function  $\sum_{i=1}^N L[y_i, f_M(\mathbf{x}_i)]$ , i.e.,

$$\arg \min_{\Theta} \sum_{i=1}^N L[y_i, f_M(\mathbf{x}_i)] = \arg \min_{\Theta} \sum_{i=1}^N L\left[y_i, \sum_{m=1}^M T(\mathbf{x}_i; \Theta_m)\right]. \quad (5)$$

Here, the loss function is used to reflect the difference between the sample real value  $y_i$  and the output of the GBRT  $f_M(\mathbf{x}_i)$ .

Note that the GBRT model in equation (4) can also be written in a forward step way and expressed as

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + T(\mathbf{x}; \Theta_m), \quad m = 1, 2, \dots, M. \quad (6)$$

Therefore, training of the GBRT model can be achieved by  $M$  iteration steps. Specifically, at the initial step, we define  $f_1(\mathbf{x}) = 0$ , and for the  $m$ -th iteration step, a new  $T(\mathbf{x}; \Theta_m)$  is generated. The parameters  $\Theta_m$  of  $T(\mathbf{x}; \Theta_m)$  should be obtained to minimize the loss function

$$\hat{\Theta}_m = \arg \min_{\Theta_m} \sum_{i=1}^N L[y_i, f_{m-1}(\mathbf{x}_i) + T(\mathbf{x}_i; \Theta_m)], \quad (7)$$

where  $\hat{\Theta}_m$  are the optimal DT parameters.

If the squared loss function is used, then one obtains

$$\begin{aligned} L[y_i, f_{m-1}(\mathbf{x}_i) + T(\mathbf{x}_i; \Theta_m)] &= [y_i - f_{m-1}(\mathbf{x}_i) - T(\mathbf{x}_i; \Theta_m)]^2 \\ &= [r_{m,i} - T(\mathbf{x}_i; \Theta_m)]^2, \end{aligned} \quad (8)$$

where  $r_{m,i} = y_i - f_{m-1}(\mathbf{x}_i)$  represents the residual of the model  $f_{m-1}(\mathbf{x})$ .

Therefore, the solution of equation (8) converts to the selection of appropriate  $\Theta_m$  to minimize the difference of the residual  $r_{m,i}$  of the DT and the output  $T(\mathbf{x}; \Theta_m)$  or, equivalently,  $\{(\mathbf{x}_i, r_{m,i})\}_{i=1,2,\dots,N}$  can be used as the sample set of the decision tree  $T(\mathbf{x}; \Theta_m)$ , and the optimal parameters  $\hat{\Theta}_m$  are obtained according to the conventional DT generation process.

Moreover, in a more generalized sense, the negative gradient of the loss function can be used to represent the residual of the model, i.e.,

$$r_{m,i} = -\left[\frac{\partial L(y_i, f(\mathbf{x}_i))}{\partial f(\mathbf{x}_i)}\right]_{f(\mathbf{x})=f_{m-1}(\mathbf{x})}. \quad (9)$$

With  $\{(\mathbf{x}_i, r_{m,i})\}_{i=1,2,\dots,N}$ , we can fit the  $m$ -th DT  $h_m$ , whose leaf nodes can be represented by  $R_{m,j}$ ,  $j = 1, 2, \dots, J$ , where  $J$  indicates the number of leaf nodes of the DT. For each leaf node of the regression tree  $h_m$ , calculate the optimal fitting value  $c_{m,j}$ :

$$c_{m,j} = \arg \min_c \sum_{\mathbf{x}_i \in R_{m,j}} L[y_i, f_{m-1}(\mathbf{x}_i) + c], \quad j = 1, 2, \dots, J. \quad (10)$$

Then, the weak learner for this step can be written as

$$T_m(\mathbf{x}) = \sum_{j=1}^J c_{m,j} I(\mathbf{x} \in R_{m,j}), \quad (11)$$

and the updated strong learner till this step is

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + T_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \sum_{j=1}^J c_{m,j} I(\mathbf{x} \in R_{m,j}). \quad (12)$$

After  $M$  steps, the strong learner is finally obtained by

$$f_M(\mathbf{x}) = \sum_{m=1}^M \sum_{j=1}^J c_{m,j} I(\mathbf{x} \in R_{m,j}). \quad (13)$$

The procedure of the GBRT algorithm can be summarized as follows:

- (1) Initialization of the function for the weak learner  $f_0(\mathbf{x}) = \arg \min_c \sum_{i=1}^N L(y_i, c)$
- (2) For the  $m$ -th iteration ( $m = 1, 2, \dots, M$ ):
  - (a) For each sample  $(\mathbf{x}_i, y_i)$ ,  $i = 1, 2, \dots, N$ , the negative gradient is calculated using equation (9)
  - (b) Train the  $m$ -th DT  $h_m$  by using  $\{(\mathbf{x}_i, r_{m,i})\}_{i=1,2,\dots,N}$  and the corresponding areas

of the  $i$ -th leaf nodes are denoted as  $R_{m,j}$ ,  $j = 1, 2, \dots, J$

(c) For each leaf node of the regression tree  $h_m$ , calculate the optimal fitting value  $c_{m,j}$  using equation (10)

(d) Update the learner  $f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \sum_{j=1}^J c_{m,j} I(\mathbf{x} \in R_{m,j})$

(3) After  $M$  iterations, the strong learner is obtained using equation (13)

**3.3. Implementation of GBRT.** In this study, one of the most widely used DT, i.e., CART, is employed as the individual learning algorithm. The implementation of the GBRT can be summarized as the following four steps:

- (1) Collect and process the data, such as the setting of input/output variables and the grouping of the training/testing datasets
- (2) Train the regression model using the GBRT with the training dataset
- (3) Validate the trained model with the testing dataset
- (4) Apply the model to the realistic problems

The flowchart of the abovementioned procedure is depicted in Figure 1.

Another important issue associated with the implementation of GBRT is the determination of model parameters, which have two levels, i.e., the framework level and the level for the individual learning algorithm. At the framework level, there are two main parameters, i.e., the number of iteration (number of weak learners) and the learning rate, which is used to avoid the overfitting problem. At the single learning algorithm level, there are four primary parameters, i.e., the maximum depth of the tree, the minimum samples for split, the minimum samples of leaf node, and the minimum change in impurity. The selected values of these parameters are determined based on previous studies in literature and practical modeling experience, as shown in Table 1.

#### 4. Collection of Experimental Data for Shear Strength of Internal RC Beam-Column Joints

In implementing the ML techniques for prediction of the shear strength for RC joints, an experimental database is required to train the predictive model and validate the model. Therefore, a database including the experimental results of 86 internal RC beam-column joints was collected for this purpose in this study. In the database, there are 10 input parameters covering material properties and geometric dimensions and reinforcing details of the test specimens, i.e., the concrete strength  $f_c$ , the section width of column  $b_c$ , the section height of column  $h_c$ , the section width of beam  $b_b$ , the section height of beam  $h_b$ , the yielding strength of beam longitudinal bar  $f_{y,b}$ , the yielding strength of column longitudinal bar  $f_{y,c}$ , the yielding strength of joint transverse bar  $f_{y,v}$ , the transverse bar ratio  $\rho_v$ , and the axial load ratio  $n$ . The only output is

the joint shear strength  $\tau$ . The statistical information of these parameters, e.g., mean and standard deviation (St.D.), and the distributions of the aforementioned parameters are illustrated in Table 2 and Figure 2. The details of the tested specimens in the database are given in Table 3.

## 5. Results and Discussion

**5.1. 10-Fold Cross-Validation Results.** To validate the proposed method, the 10-fold cross-validation method is firstly used to evaluate the model's performance. The 10-fold cross-validation method is developed to minimize the bias associated with random sampling of the training and testing datasets. It divides the experimental data samples into 10 subsets, and for each run, 9 are set as training subsets and 1 is set as validating subset. It is believed that repeating this for 10 times is able to represent the generalization and reliability of the predictive model. Moreover, three commonly used indicators are introduced to assess the prediction performance, which are respectively defined as

Coefficient of determination  $R$ -squared ( $R^2$ ):

$$R^2 = 1 - \frac{\sum_{i=1}^N (P_i - T_i)^2}{\sum_{i=1}^N (T_i - \bar{T})^2} \quad (14)$$

Root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (P_i - T_i)^2}{N}} \quad (15)$$

Mean absolute error (MAE):

$$\text{MAE} = \frac{\sum_{i=1}^N |P_i - T_i|}{N} \quad (16)$$

where  $P_i$  and  $T_i$  are the predicted and tested values, respectively;  $\bar{T}$  is the mean value of all the tested values;  $N$  is the total number of the samples in the dataset.

Among the three indicators,  $R^2$  indicates the degree of the linear correlation between the predicted and tested values. RMSE shows the deviation between the predicted and tested values. MAE reflects the ratio of the prediction error to the tested values. The closer the  $R^2$  to 1, the smaller the RMSE or MAE, the better performance the prediction model possesses. Table 4 shows the 10-fold cross-validation statistic results of the GBRT model.

It can be drawn from Table 4 that the average determination coefficient  $R^2$  for the 10-fold results is 0.875, which is close to 1; the average RMSE and MAE are 0.948 MPa and 0.722 MPa, respectively, which are small. The standard derivations (St.D.) for  $R^2$ , RMSE, and MAE

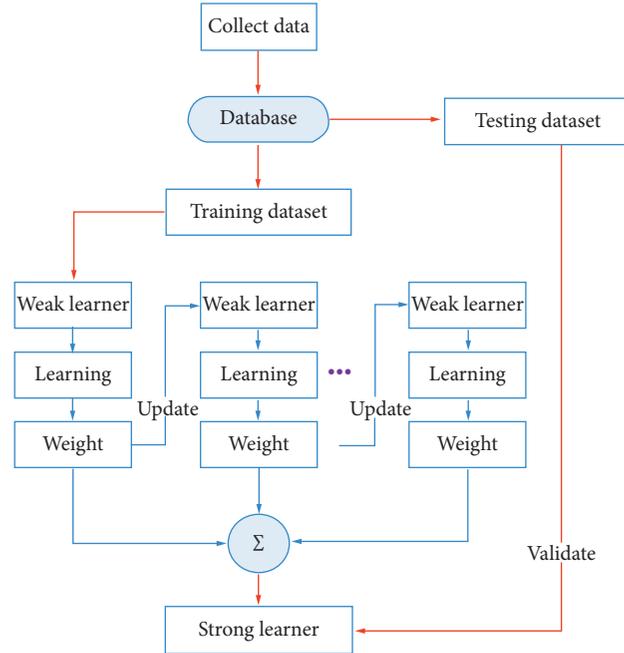


FIGURE 1: Flowchart of implementation of GBRT.

TABLE 1: Model parameters of GBRT.

Level	Parameter	Setting
Framework	Maximum iteration number	100
	Learning rate	0.03
Single learning algorithm	Maximum depth of the tree	10
	Maximum leaf nodes	200
	Minimum samples for split	6
	Minimum samples of leaf node	5

TABLE 2: Statistics of parameters for the internal RC beam-column joints.

Parameter	Unit	Maximum	Minimum	Mean	St.D.	Type
$f_c$	MPa	116.90	16.57	43.31	19.07	Input
$b_c$	mm	458.00	200.00	268.12	66.03	Input
$h_c$	mm	458.00	200.00	277.23	81.72	Input
$b_b$	mm	406.00	150.00	194.74	64.60	Input
$h_b$	mm	500.00	250.00	322.59	56.96	Input
$f_{y,b}$	MPa	1456.00	314.00	457.48	179.87	Input
$f_{y,c}$	MPa	1456.00	314.00	458.09	150.46	Input
$f_{yv}$	MPa	1374.00	276.00	451.73	262.08	Input
$\rho_v$	%	2.00	0.18	0.66	0.38	Input
$n$	—	1.18	0.00	0.25	0.23	Input
$\tau$	MPa	14.87	2.85	7.30	2.81	Output

are 0.082, 0.347, and 0.245, respectively, which means the prediction performance has low variance. All of these indices demonstrate that the proposed method has excellent performance in predicting the shear strength of internal RC joints.

**5.2. Prediction Results of Different ML Models.** To demonstrate the prediction performance of the GBRT model, the four general ML models, i.e., LR, SVM, ANN, and DT, are also used to predict the shear strength of the 86 specimens. The optimized parameters of the four models are determined

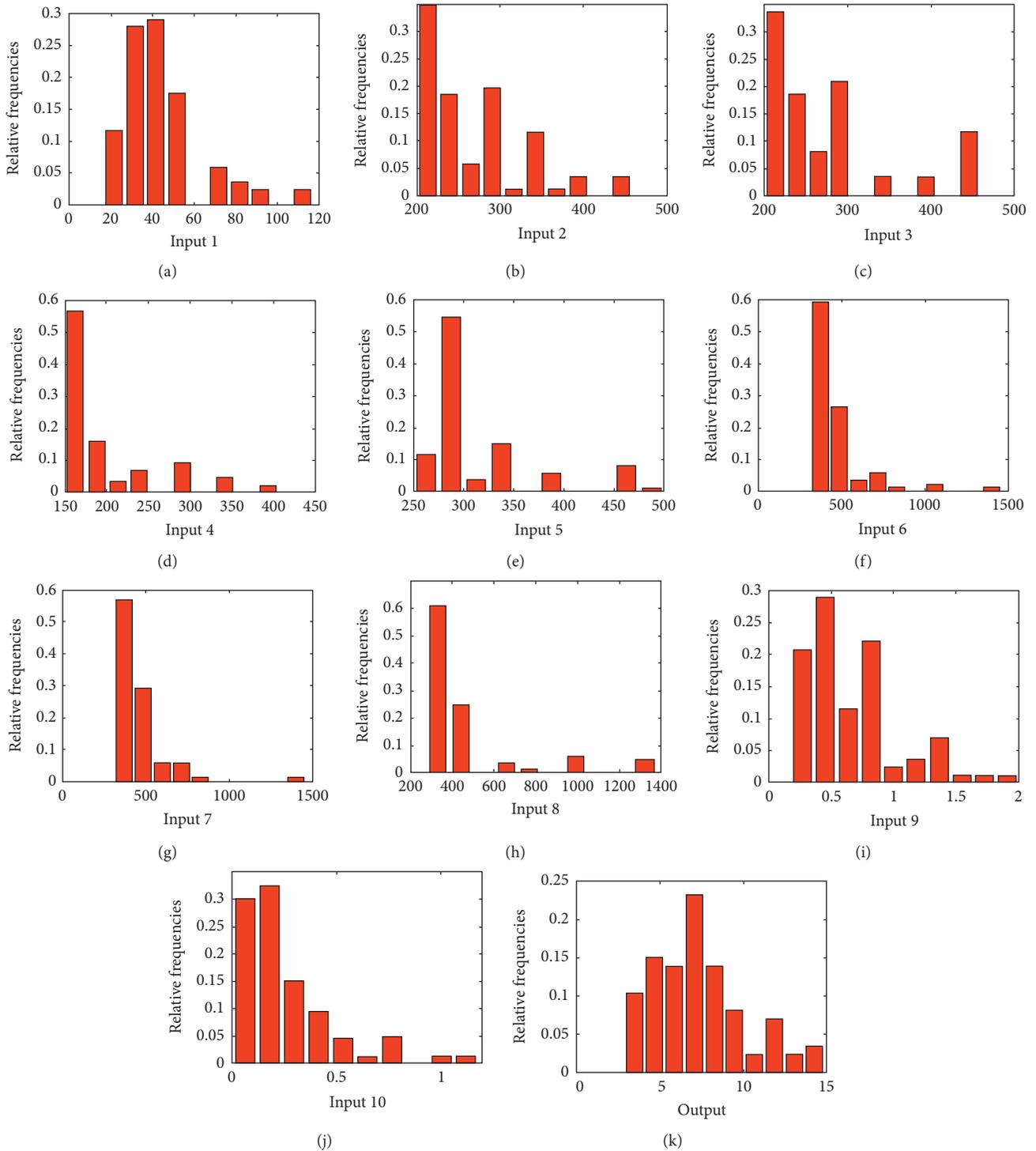


FIGURE 2: Statistic distribution of the input and output parameters. (a) Concrete strength (MPa). (b) Section width of column (mm). (c) Section height of column (mm). (d) Section width of beam (mm). (e) Section height of beam (mm). (f) Yielding strength of beam longitudinal bar (MPa). (g) Yielding strength of column longitudinal bar (MPa). (h) Yielding strength of joint transverse bar (MPa). (i) Transverse bar ratio (%). (j) Axial load ratio. (k) Joint shear strength (MPa).

TABLE 3: Main data of the internal RC beam-column joint tests.

References	Specimen	$f_c$ (MPa)	$b_c \times h_c$ (mm $\times$ mm)	$b_b \times h_b$ (mm $\times$ mm)	$f_{y,b}$ (MPa)	$f_{y,c}$ (MPa)	$f_{y,v}$ (MPa)	$\rho_v$	$n$	$\tau$ (MPa)
[37]	I	24.70	250 $\times$ 250	150 $\times$ 350	391.1	353.8	277.5	0.36	0.226	4.15
	I-1	27.20	250 $\times$ 250	150 $\times$ 350	379.4	366.0	301.1	0.36	0.204	4.16
	I-2	31.95	250 $\times$ 250	150 $\times$ 350	379.4	366.0	301.1	0.36	0.173	5.10
	SJ1-1	31.78	250 $\times$ 250	150 $\times$ 350	366.3	349.0	300.9	0.42	0.173	5.53
	SJ1-2	34.48	250 $\times$ 250	150 $\times$ 350	366.3	349.0	300.9	0.42	1.180	5.82
	SJ1-4	29.80	250 $\times$ 250	150 $\times$ 350	380.3	380.3	300.9	0.42	0.440	7.32
	SJ2-2	28.40	250 $\times$ 250	150 $\times$ 350	366.3	349.0	300.9	1.50	0.174	6.17
	SJ5-1	37.88	250 $\times$ 250	150 $\times$ 350	366.3	349.0	300.9	0.42	0.178	6.47
	SJ-3	34.48	250 $\times$ 250	150 $\times$ 350	462.0	404.6	340.0	1.26	0.125	6.41
	SJ-B	27.00	450 $\times$ 450	250 $\times$ 500	326.0	388.0	384.0	0.76	0.167	5.49
[38]	J3-50	46.12	200 $\times$ 200	150 $\times$ 300	390.0	385.0	307.0	0.83	0.285	7.14
	J4-50	47.40	200 $\times$ 200	150 $\times$ 300	390.0	385.0	307.0	0.83	0.361	7.83
	J4-30	46.74	200 $\times$ 200	150 $\times$ 300	390.0	385.0	307.0	1.33	0.374	7.92
	J5-80	46.71	200 $\times$ 200	150 $\times$ 300	328.0	405.0	307.0	0.66	0.467	8.80
	J5-50	48.16	200 $\times$ 200	150 $\times$ 300	328.0	405.0	307.0	0.83	0.449	8.98
	J6-80	46.94	200 $\times$ 200	150 $\times$ 300	328.0	405.0	307.0	0.66	0.558	9.26
	J6-50	43.60	200 $\times$ 200	150 $\times$ 300	328.0	405.0	307.0	0.83	0.514	8.29
	J6-30	48.12	200 $\times$ 200	150 $\times$ 300	328.0	405.0	307.0	1.33	0.540	9.07
	J7-80	43.60	200 $\times$ 200	150 $\times$ 300	328.0	405.0	307.0	0.66	0.716	7.91
	J7-50	48.42	200 $\times$ 200	150 $\times$ 300	328.0	405.0	307.0	0.83	0.636	8.98
	J7-30	43.81	200 $\times$ 200	150 $\times$ 300	328.0	405.0	307.0	1.33	0.712	8.84
	J8-50	49.45	200 $\times$ 200	150 $\times$ 300	328.0	405.0	307.0	0.83	0.744	9.22
	J8-30	43.81	200 $\times$ 200	150 $\times$ 300	328.0	405.0	307.0	1.33	0.814	8.38
[39]	ZHJ4	27.36	200 $\times$ 200	150 $\times$ 250	360.2	360.2	350.0	1.00	0.330	6.91
	ZHJ5	27.36	200 $\times$ 200	150 $\times$ 250	360.2	360.2	350.0	1.21	0.330	6.93
	ZHJ6	27.36	200 $\times$ 200	150 $\times$ 250	360.2	360.2	350.0	1.81	0.330	7.07
[40]	J-1	27.36	200 $\times$ 200	150 $\times$ 250	380.0	380.0	350.0	1.04	0.330	6.78
	J-2	27.36	200 $\times$ 200	150 $\times$ 250	380.0	380.0	350.0	1.12	0.330	6.84
	J-3	16.57	200 $\times$ 200	150 $\times$ 250	473.0	473.0	375.0	1.30	0.270	6.49
	J-4	19.53	200 $\times$ 200	150 $\times$ 250	473.0	473.0	375.0	2.00	0.23	6.55
	JL1	21.43	250 $\times$ 250	150 $\times$ 300	314.0	314.0	276.0	0.58	0.21	5.32
[41]	SJ-3	29.34	350 $\times$ 350	175 $\times$ 350	354.7	348.0	381.5	0.50	1.06	6.88
[42]	J4	25.7	300 $\times$ 300	200 $\times$ 300	401	401	368	0.28	0.30	4.22
	J5	28.7	300 $\times$ 300	200 $\times$ 300	401	401	368	0.28	0.07	5.04
	C2	25.6	300 $\times$ 300	200 $\times$ 300	324	422	368	0.90	0.08	5.36
[43]	U1	26.2	331 $\times$ 458	280 $\times$ 458	449	457	409	0.50	0.40	5.38
	U2	41.8	331 $\times$ 458	280 $\times$ 458	449	449	409	0.50	0.25	7.87
	U3	26.6	331 $\times$ 458	280 $\times$ 458	449	402	409	0.50	0.39	6.03
	U4	36.1	458 $\times$ 331	406 $\times$ 458	449	449	409	0.50	0.29	7.17
	U5	35.9	331 $\times$ 458	280 $\times$ 458	449	449	409	0.50	0.04	7.57
	U6	36.8	331 $\times$ 458	280 $\times$ 458	449	449	409	0.50	0.48	8.15
	U7	37.2	458 $\times$ 331	406 $\times$ 458	449	449	409	0.50	0.46	7.24
[44]	J1	70.0	300 $\times$ 300	200 $\times$ 300	718	718	955	0.75	0.12	10.84
	J3	107.0	300 $\times$ 300	200 $\times$ 300	718	718	955	0.75	0.12	13.75
	J4	70.0	300 $\times$ 300	200 $\times$ 300	718	718	955	0.75	0.12	11.44
	J5	70.0	300 $\times$ 300	200 $\times$ 300	718	718	955	0.75	0.12	11.26
	J6	53.5	300 $\times$ 300	200 $\times$ 300	718	718	955	0.75	0.12	9.98
[45]	J1	81.2	300 $\times$ 300	240 $\times$ 300	638	638	1374	0.38	0.11	11.58
	J2	81.2	300 $\times$ 300	240 $\times$ 300	1456	1456	1374	0.38	0.11	12.43
	J4	72.8	300 $\times$ 300	240 $\times$ 300	515	515	1374	0.38	0.13	11.94
	J5	72.8	300 $\times$ 300	240 $\times$ 300	839	839	1374	0.38	0.13	13.38
	J8	79.2	300 $\times$ 300	240 $\times$ 300	370	370	775	0.38	0.12	14.01

TABLE 3: Continued.

References	Specimen	$f_c$ (MPa)	$b_c \times h_c$ (mm $\times$ mm)	$b_b \times h_b$ (mm $\times$ mm)	$f_{y,b}$ (MPa)	$f_{y,c}$ (MPa)	$f_{y,v}$ (MPa)	$\rho_v$	$n$	$\tau$ (MPa)
[46]	Aa-4	30.4	200 $\times$ 200	150 $\times$ 300	419	419	350	0.22	0.06	3.19
	Aa-7	38.1	200 $\times$ 200	150 $\times$ 300	400	400	350	0.22	0.05	3.08
	Aa-8	38.1	200 $\times$ 200	150 $\times$ 300	400	400	350	0.22	0.10	3.19
	Aa-1	41.1	200 $\times$ 200	150 $\times$ 300	400	400	350	0.22	0.05	3.10
	Aa-2	41.1	200 $\times$ 200	150 $\times$ 300	400	400	350	0.22	0.10	2.97
	Ab-1	41.1	200 $\times$ 200	150 $\times$ 300	400	400	350	0.22	0.05	3.06
	Ab-2	41.1	200 $\times$ 200	150 $\times$ 300	400	400	350	0.22	0.10	2.85
[47]	HNO3	88.7	400 $\times$ 400	300 $\times$ 400	442	442	681	0.86	0.17	13.01
	HNO4	88.7	400 $\times$ 400	300 $\times$ 400	605	610	681	0.86	0.17	14.87
	HNO5	116.9	400 $\times$ 400	300 $\times$ 400	623	610	681	0.86	0.13	12.39
[48]	S3	24	300 $\times$ 300	200 $\times$ 300	465	450	390	0.36	0.05	3.64
[49]	A1	40.2	220 $\times$ 220	160 $\times$ 250	1069	644	291	0.43	0.08	5.25
	A2	40.2	220 $\times$ 220	160 $\times$ 250	409	388	291	0.43	0.08	4.87
	A3	40.2	220 $\times$ 220	160 $\times$ 250	1069	644	291	0.43	0.23	5.25
[50]	LJ1	25.5	343 $\times$ 457	343 $\times$ 343	440	470	400	0.61	0	3.79
	LJ2	32.8	343 $\times$ 457	343 $\times$ 343	440	470	400	0.61	0	4.86
	LJ3	31.1	343 $\times$ 457	343 $\times$ 343	440	470	400	0.61	0	4.61
	LJ4	34.3	343 $\times$ 457	343 $\times$ 343	440	470	400	0.61	0	5.03
[51]	BCJ2	30.3	254 $\times$ 254	203 $\times$ 305	414	448	414	0.52	0	5.55
	BCJ3	27.4	305 $\times$ 254	203 $\times$ 305	414	448	414	0.43	0	5.08
	BCJ4	27.2	356 $\times$ 254	203 $\times$ 305	414	448	414	0.37	0	5.11
[52]	HSLCJ-1	48.6	260 $\times$ 260	150 $\times$ 300	435	421	449	0.18	0.15	7.25
	HSLCJ-2	48.9	260 $\times$ 260	150 $\times$ 300	435	421	449	0.22	0.15	7.5
	HSLCJ-3	49.5	260 $\times$ 260	150 $\times$ 300	435	421	449	0.18	0.25	7.76
	HSLCJ-4	48.2	260 $\times$ 260	150 $\times$ 300	435	421	449	0.22	0.25	7.82
[53]	JD1	41.9	250 $\times$ 250	150 $\times$ 300	427	443	318	0.33	0.147	7.28
	JD2	43.95	250 $\times$ 250	150 $\times$ 300	427	443	318	0.84	0.147	7.38
	JD3	43.34	250 $\times$ 250	150 $\times$ 300	427	443	318	1.44	0.147	7.46
	JD5	41.06	250 $\times$ 250	150 $\times$ 300	427	443	318	0.84	0.059	6.35
	JD6	44.35	250 $\times$ 250	150 $\times$ 300	427	443	318	0.84	0.235	8.05
[54]	MZJD-1	37.2	300 $\times$ 300	200 $\times$ 400	320	331	318	0.67	0.1	5.85
	MZJD-2	37.2	300 $\times$ 300	200 $\times$ 400	320	331	318	0.67	0.3	7.29
[55]	8-3	54.17	250 $\times$ 250	200 $\times$ 300	479.6	479.6	479.6	0.49	0.1	9.86
	8-5	55.34	250 $\times$ 250	200 $\times$ 300	479.6	479.6	479.6	0.49	0.1	11.29

TABLE 4: 10-fold cross-validation statistic results.

Fold	$R^2$	RMSE (MPa)	MAE (MPa)
1	0.827	1.217	0.937
2	0.926	0.816	0.704
3	0.937	0.726	0.619
4	0.682	1.621	1.008
5	0.819	0.967	0.800
6	0.910	0.873	0.579
7	0.952	0.531	0.394
8	0.962	0.637	0.460
9	0.905	0.649	0.525
10	0.833	1.443	1.192
Average	0.875	0.948	0.722
St.D.	0.082	0.347	0.245

by using the grid search after setting the initial values. The total dataset is divided into for training and testing as 8-2, i.e., 80% of the data is used for training and 20% of the data is used for testing. Figure 3 shows the prediction results of the GBRT model and the four general ML models for the testing

dataset. It is clear that the GBRT model has stronger linear correlation compared with other four ML models. The reason is that the GBRT is an ensemble learning algorithm with strong learner, while other four models use individual-type learning algorithms with weak learners.

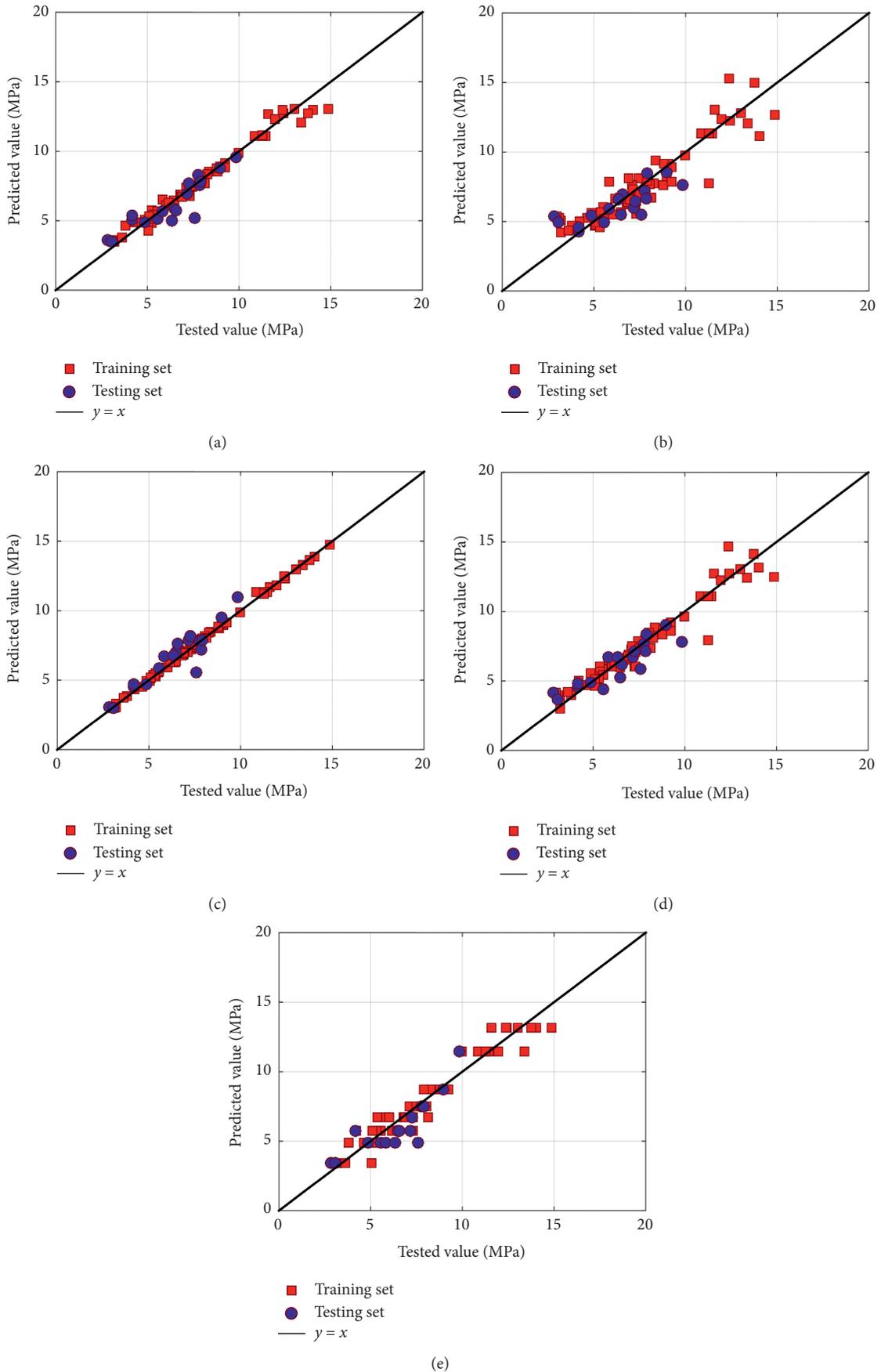


FIGURE 3: Prediction results of different ML models: (a) GBRT. (b) LR. (c) SVM. (d) ANN. (e) DT.

Table 5 exhibits the prediction performance of the five ML models by providing the average statistical indices of the 10-fold cross-validation results. Obviously, the GBRT model has the closest  $R^2$  to 1 and smallest values of RMSE and MAE among the five ML models. It further verifies the superiority of the GBRT model over the general individual-type ML models.

## 6. Comparison with Conventional Mechanical-Driven Approach

**6.1. Typical Mechanical-Driven Approach: MCFT.** In this section, the derivation of MCFT is briefly summarized as it is a representative conventional mechanical-driven shear strength prediction method. A basic assumption for MCFT is that the crack direction of a RC plane element is in accordance with the principal compressive stress and varies accordingly. The definitions of stress, strain, rotational angle, and principal direction are illustrated in Figure 4, where the  $x$ - $y$  coordinate system is the local system and the 1-2 coordinate system indicates the principal tensile strain-principal compressive strain system. The strain vector and stress vector of the RC element in the local system are denoted as  $[\varepsilon_x \ \varepsilon_y \ \gamma_{xy}]^T$  and  $[f_x \ f_y \ v_{xy}]^T$ , respectively.

The derivation of the MCFT includes three parts, i.e., compatibility equations, equilibrium equations, and constitutive equations. The detailed formulations are given as follows.

**6.1.1. Compatibility Equations.** According to Mohr's circle of strain, the principal tensile strain  $\varepsilon_1$  and the principal compressive strain  $\varepsilon_2$  of the element are calculated as

$$\begin{cases} \varepsilon_1 = \frac{(\varepsilon_x + \varepsilon_y)}{2} + \sqrt{\frac{(\varepsilon_x - \varepsilon_y)^2}{4} + \frac{(\gamma_{xy})^2}{4}}, \\ \varepsilon_2 = \frac{(\varepsilon_x + \varepsilon_y)}{2} - \sqrt{\frac{(\varepsilon_x - \varepsilon_y)^2}{4} + \frac{(\gamma_{xy})^2}{4}}. \end{cases} \quad (17)$$

Accordingly, the rotational angle  $\theta$  from the principal strain direction to the  $x$ -axis can be obtained by

$$\tan^2 \theta = \frac{\varepsilon_x - \varepsilon_2}{\varepsilon_y - \varepsilon_2} = \frac{\varepsilon_1 - \varepsilon_y}{\varepsilon_1 - \varepsilon_x} = \frac{\varepsilon_1 - \varepsilon_y}{\varepsilon_y - \varepsilon_2} = \frac{\varepsilon_x - \varepsilon_2}{\varepsilon_1 - \varepsilon_x}. \quad (18)$$

**6.1.2. Equilibrium Equations.** The basic element consists of a steel bar and concrete such that its equilibrium condition can be derived from the stress state as shown in Figure 4, which can be expressed as

$$\begin{cases} f_x = \sigma_{cx} + \rho_{sx} \sigma_{sx}, \\ f_y = \sigma_{cy} + \rho_{sy} \sigma_{sy}, \\ v_{xy} = \tau_{cxy}, \end{cases} \quad (19)$$

where  $\sigma_{cx}$  and  $\sigma_{cy}$  are the normal stresses of concrete in the  $x$  and  $y$  directions, respectively;  $\tau_{cxy}$  is the shear stress of concrete;  $\rho_{sx}$  and  $\rho_{sy}$  denote the reinforcement ratios in the  $x$

and  $y$  directions, respectively;  $\sigma_{sx}$  and  $\sigma_{sy}$  are the normal stresses of the steel bar in the  $x$  and  $y$  directions, respectively.

Considering the condition of Mohr' circle of stress, the normal stresses and shear stress of concrete are obtained by

$$\begin{cases} \sigma_{cx} = \frac{\sigma_{c1} - \tau_{cxy}}{\tan \theta}, \\ \sigma_{cy} = \sigma_{c1} - \tau_{cxy} \tan \theta, \\ \tau_{cxy} = (\sigma_{c1} - \sigma_{c2}) \sin \theta \cos \theta, \end{cases} \quad (20)$$

where  $\sigma_{c1}$  and  $\sigma_{c2}$  are the principal stresses in the 1 and 2 directions.

**6.1.3. Constitutive Equations.** With equations (19) and (20), it is found that the stress vector of the RC element can be obtained by the stress states of concrete and steel. Therefore, the constitutive stress-strain relations of these two materials are necessary for the state determination of the element. Especially, the steel bars are assumed in uniaxial stress state and the concrete is subjected to biaxial stress state, which can be described in the two principal directions.

For reinforcement steel, the uniaxial elastic perfectly-plastic model is adopted, which is given by

$$\begin{cases} \sigma_{sx} = E_{sx} \varepsilon_{sx} \leq f_{yx}, \\ \sigma_{sy} = E_{sy} \varepsilon_{sy} \leq f_{yy}, \end{cases} \quad (21)$$

where  $E_{sx}$ ,  $\varepsilon_{sx}$ , and  $f_{yx}$  are the elastic modulus, strain, and yielding strength of the steel bar in the  $x$  direction, respectively;  $E_{sy}$ ,  $\varepsilon_{sy}$ , and  $f_{yy}$  are the elastic modulus, strain, and yielding strength of the steel bar in the  $y$  direction, respectively.

For concrete, the shear stress state is distinctly different from the uniaxial stress state. In consideration of the tensile stress perpendicular to the principal compressive direction having influences on the compressive behavior of concrete, it is recommended using the modified uniaxial stress-strain relationships to represent the stress-strain relationship of the RC plane element subjected to combined stress state, which are

Stress-strain relationship in the tensile principal direction

$$\sigma_{c1} = \begin{cases} E_c \varepsilon_1, & \varepsilon_1 \leq \varepsilon_{cr}, \\ \frac{f_t}{1 + \sqrt{200\varepsilon_1}}, & \varepsilon_1 > \varepsilon_{cr}. \end{cases} \quad (22)$$

Stress-strain relationship in the compressive principal direction

TABLE 5: Comparisons of predictive performance between GBRT and general ML models.

Model	$R^2$	RMSE (MPa)	MAE (MPa)
GBRT	0.875	0.948	0.722
LR	0.747	1.443	1.038
SVM	0.798	1.257	0.813
ANN	0.626	1.729	1.019
DT	0.802	1.285	0.840

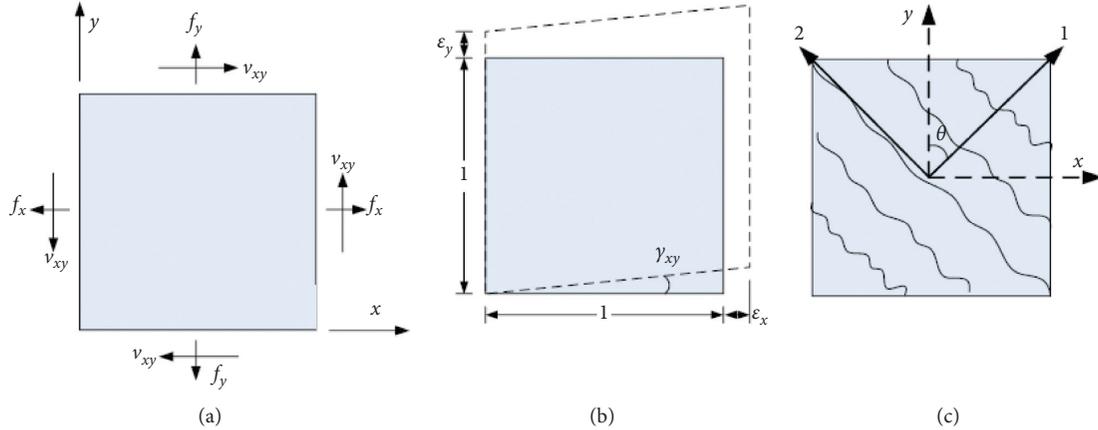


FIGURE 4: The basic RC element state. (a) Stress state. (b) Strain state. (c) Principal direction.

$$\sigma_{c2} = f_{c2\max} \left[ 2 \left( \frac{\varepsilon_2}{\varepsilon_0} \right) - \left( \frac{\varepsilon_2}{\varepsilon_0} \right)^2 \right], \quad (23)$$

with

$$\frac{f_{c2\max}}{f'_c} = \frac{1}{0.8 - 0.34(\varepsilon_1/\varepsilon_0)} \leq 1, \quad (24)$$

where  $E_c$  is the elastic modulus of concrete;  $f_t$  and  $f'_c$  are the tensile and compressive strengths of concrete, respectively;  $\varepsilon_{cr}$  and  $\varepsilon_0$  are the strains corresponding to the tensile strength and the compressive strength, respectively;  $f_{c2\max}$  is the maximum compressive stress in the principal compressive direction. It is clear that the modification equation (24) considers the reduction of concrete compressive strength due to the existence of tensile stress.

**6.1.4. Crack Check.** Note that the abovementioned equations handle with the global behavior of the element in an average sense, while it cannot provide the local behavior description. The local equilibrium across a crack should also be satisfied, say,

$$\begin{cases} \rho_{sx}(\sigma_{sxcr} - \sigma_{sx}) = \frac{\sigma_{c1} + f_{ci} + v_{ci}}{\tan \theta}, \\ \rho_{sy}(\sigma_{syrcr} - \sigma_{sy}) = \sigma_{c1} + f_{ci} + v_{ci} \tan \theta, \end{cases} \quad (25)$$

where  $\sigma_{sxcr}$  and  $\sigma_{syrcr}$  are the steel stress at the crack;  $f_{ci}$  and  $v_{ci}$  are the local compressive and shear stresses at the crack,

respectively. The abovementioned equation can be satisfied if there are no local compressive and shear stresses, say,

$$\rho_{sx}(\sigma_{sxcr} - \sigma_{sx}) = \rho_{sy}(\sigma_{syrcr} - \sigma_{sy}) = \sigma_{c1}. \quad (26)$$

However, a constrain should be ensured that the steel stresses at the crack should not exceed the yield strength of the steel, i.e.,  $(\sigma_{sxcr}/\sigma_{syrcr}) < (f_{yx}/f_{yy})$ . Therefore, if this condition is not satisfied, the local stresses should be calculated iteratively. The expressions for the local stresses are

$$v_{ci} = 0.18v_{ci,\max} + 1.64f_{ci} - 0.82 \frac{f_{ci}^2}{v_{ci,\max}}, \quad (27)$$

$$v_{ci,\max} = \frac{\sqrt{-f_c}}{(0.31 + 24w)/(a + 16)},$$

where  $w$  is the crack width;  $a$  is the maximum aggregate size;  $f_{ci}$  is calculated according to ref [4].

The whole process of using MCFT applied to the shear strength of internal RC joints can be depicted in Figure 5. More details can also be found in [4].

**6.2. Comparison between GBRT and MCFT.** To further evaluate the performance of the GBRT model, the conventional MCFT is also used to predict the shear strength of the 86 RC internal beam-column joints. The statistic results from the MCFT model are compared with the GBRT model and shown in Table 6. Note that to fairly compare the performance of the two models, the prediction results in the 10 testing sets of the 10-fold cross-validation process are used for the GBRT model.

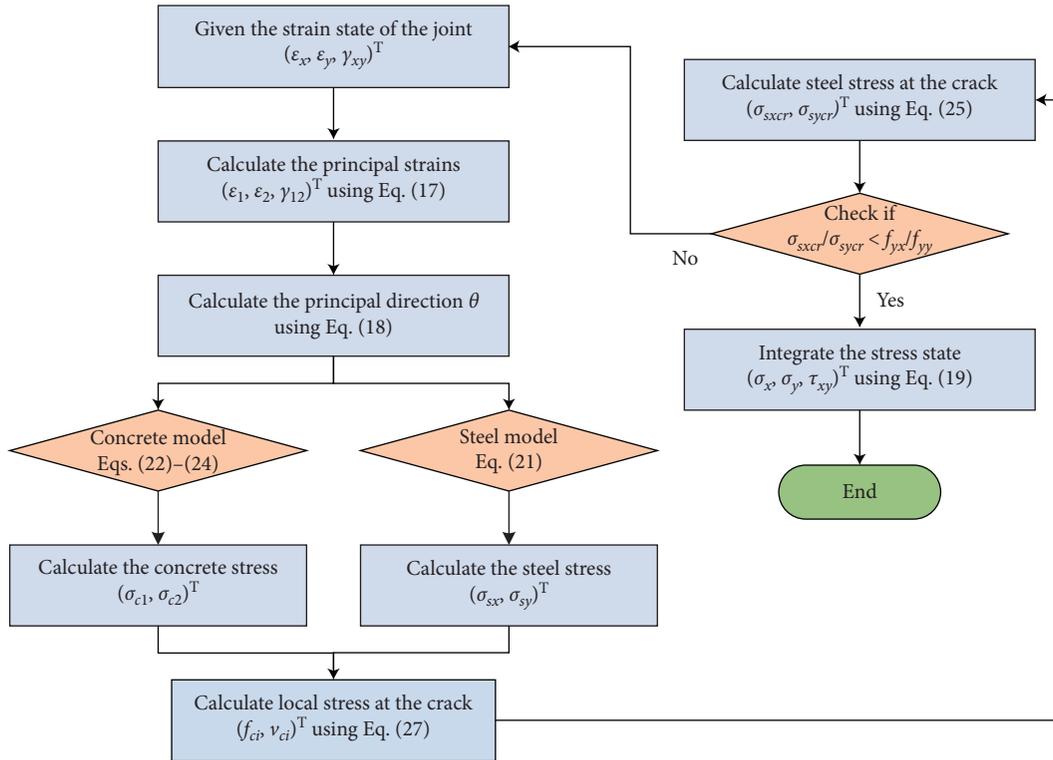


FIGURE 5: Flowchart of implementation of MCFT in calculating shear strength of internal RC joints.

TABLE 6: Comparisons of predictive performance between GBRT and MCFT models.

Model	$R^2$	RMSE (MPa)	MAE (MPa)
MCFT	0.765	1.355	1.066
GBRT	0.960	0.562	0.380
(GBRT-MCFT)/MCFT (%)	25.4	-58.5	-64.3

As can be seen from Table 6, the determination coefficient of the GBRT model has been improved by 25.4% and closer to 1 compared to the mechanical-driven MCFT model, and all the other two indicators have been dropped more than 50%. In other words, the ML-based method has obviously better performance than the MCFT-based method. Furthermore, the predicted and tested values are also plotted in Figure 6. Evidently, the GBRT results match the experimental results much better than those of the MCFT model.

Table 7 gives the statistic results of predicted value/tested value ratios for the MCFT and GBRT models. It can be concluded from Table 7 that the GBRT model statistically underestimates the shear strength because the mean value is less than 1, while the MCFT model slightly overestimates the shear strength. Apparently, the mean predicted value/tested value ratios for the GBRT approach is closer to 1 with less dispersion (St.D.) compared to the MCFT method.

Figure 7 further illustrates the predicted value/tested value ratios for the GBRT and MCFT models. The solid line, the top dotted line, and the bottom line represent the mean value, mean value plus St.D., and mean value minus St.D., respectively. Evidently, better prediction performance is achieved by the GBRT model.

## 7. Model Sensitivity Analysis

**7.1. Sensitivity of Input Parameters.** With the developed GBRT model, it is convenient for us to investigate the influences of different parameters on the shear strength of the internal RC joint and even quantify the influences. In this study, 10 input variables with different value ranges are adopted to conduct a comprehensive parametric analysis. In the parametric analysis, the control variable method is used, i.e., one control parameter varies, while other parameters are fixed. The specimen J6 of [44] is used as the reference model. The numerical ranges of the 10 inputs are shown in Table 8.

Figure 8 shows the predicted shear strength of the internal RC joints with different input variables by using the GBRT model. It can be drawn from Figure 8 that among all the input variables, concrete strength  $f_c$  is the most significant parameter affecting the shear strength. With the increasing of concrete strength  $f_c$ , beam width  $b_b$ , column width  $b_c$ , column height  $h_c$ , yielding strength of column longitudinal bar  $f_{yc}$ , yielding strength of joint transverse bar  $f_{yv}$ , transverse bar ratio  $\rho_v$ , or axial load ratio  $n$ , the shear strength has a general ascending trend. On the contrary, yielding strength of beam longitudinal

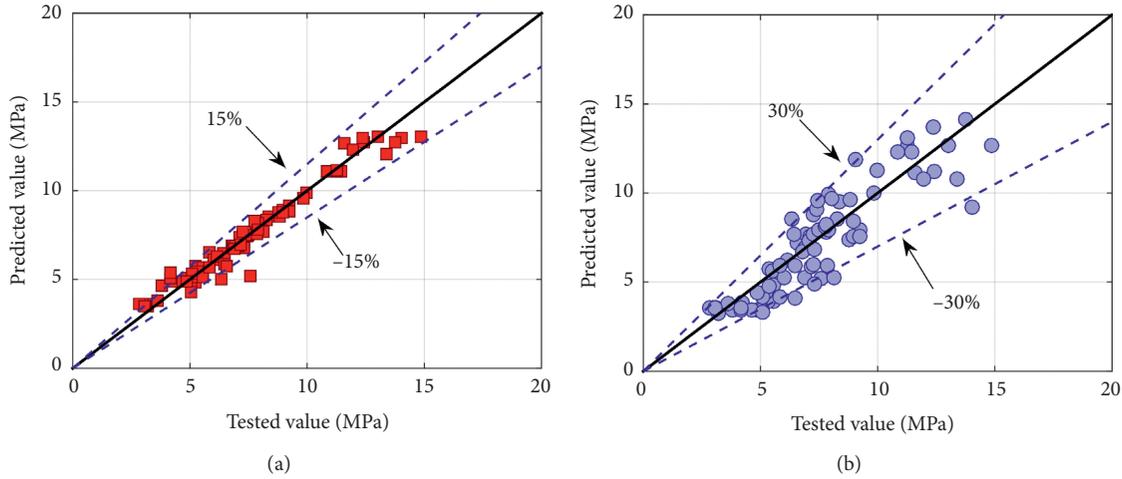


FIGURE 6: Comparisons of shear strength of RC joints by the GBRT predictive model and MCFT model. (a) GBRT results. (b) MCFT results.

TABLE 7: Statistic results of predicted value/tested value ratios for the MCFT and GBRT models.

Model	Mean value	St.D.
MCFT	0.961	0.180
GBRT	1.012	0.092

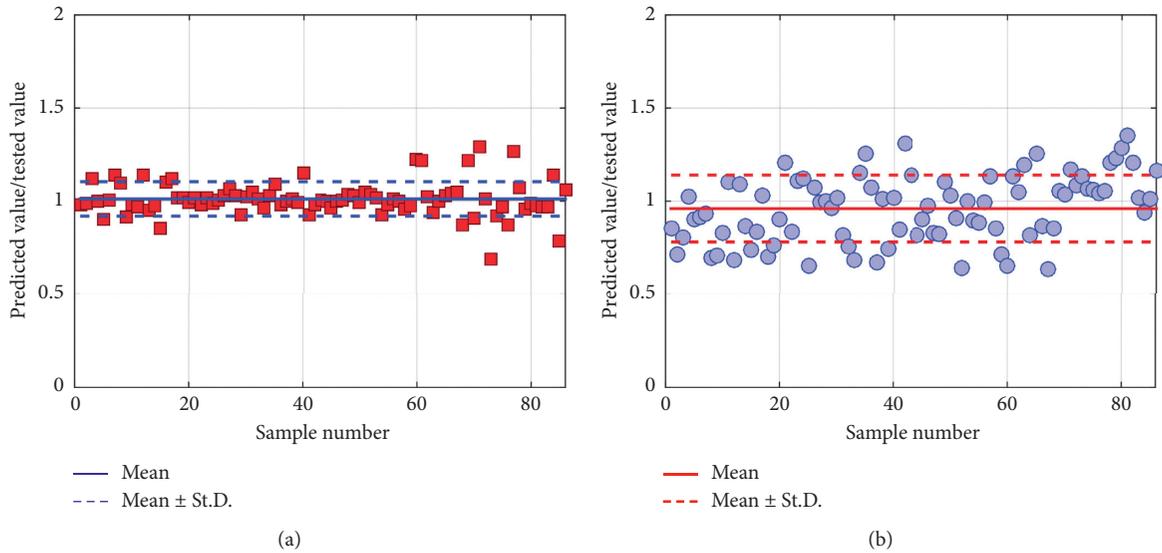


FIGURE 7: Comparisons of the predicted value/tested value for the GBRT and MCFT models. (a) GBRT results. (b) MCFT results.

TABLE 8: Numerical ranges of input variables.

$f_c$ (MPa)	$b_c$ (mm)	$h_c$ (mm)	$b_b$ (mm)	$h_b$ (mm)
[20 : 20 : 100]	[200 : 100 : 500]	[200 : 100 : 500]	[150 : 100 : 450]	[200 : 100 : 500]
$f_{y,b}$ (MPa)	$f_{y,c}$ (MPa)	$f_{yv}$ (MPa)	$\rho_v$ (%)	$n$
[300 : 140 : 1000]	[300 : 140 : 1000]	[300 : 140 : 1000]	[0.2 : 0.2 : 1.0]	[0 : 0.1 : 0.6]

bar  $f_{y,b}$  has negative effects on the shear strength. The influence of beam height  $h_b$  on the shear strength is negligible.

**7.2. Feature Importance.** Feature importance, which is used to quantify the importance of the input variables (or features), is conducted to further investigate the sensitivity of

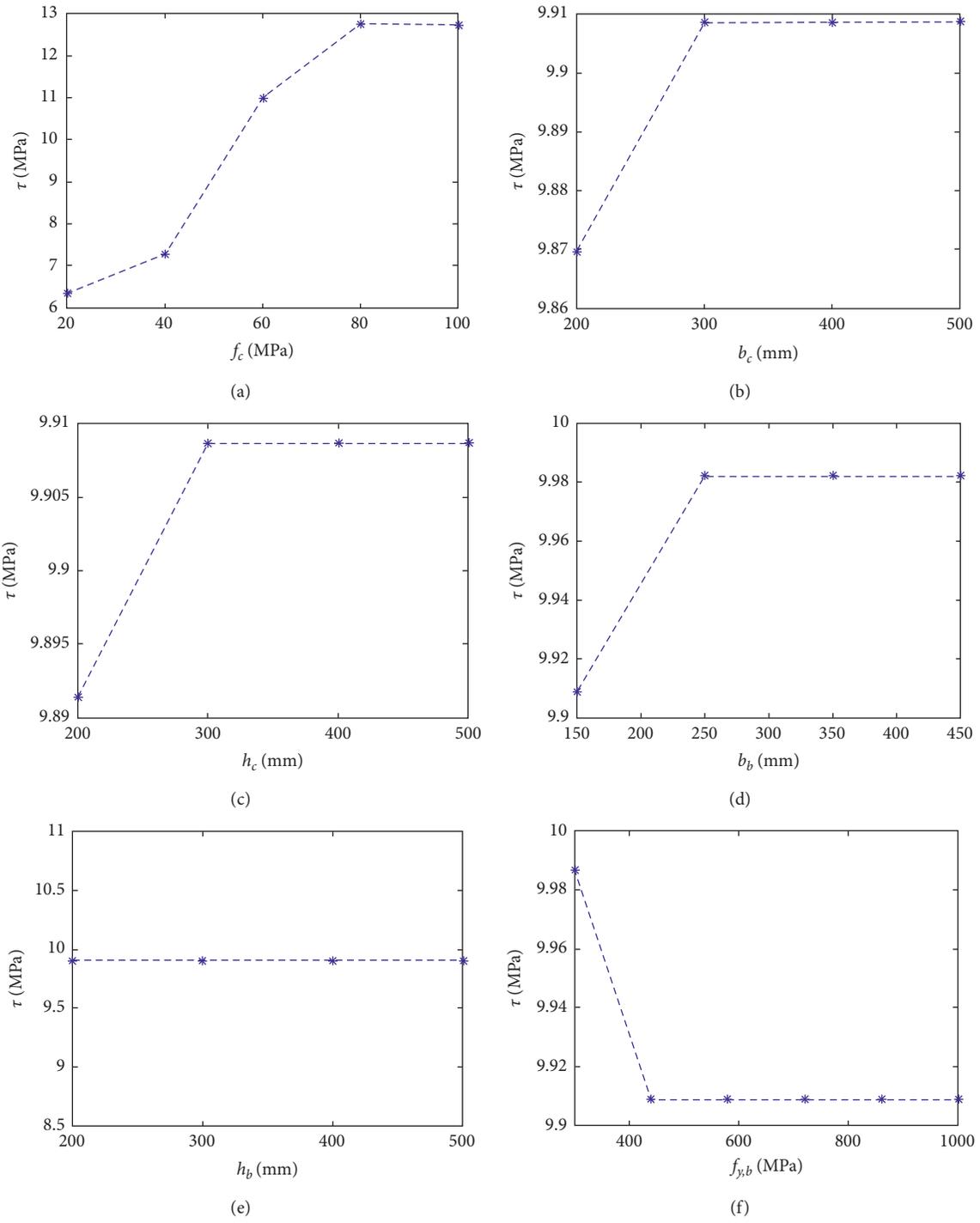


FIGURE 8: Continued.

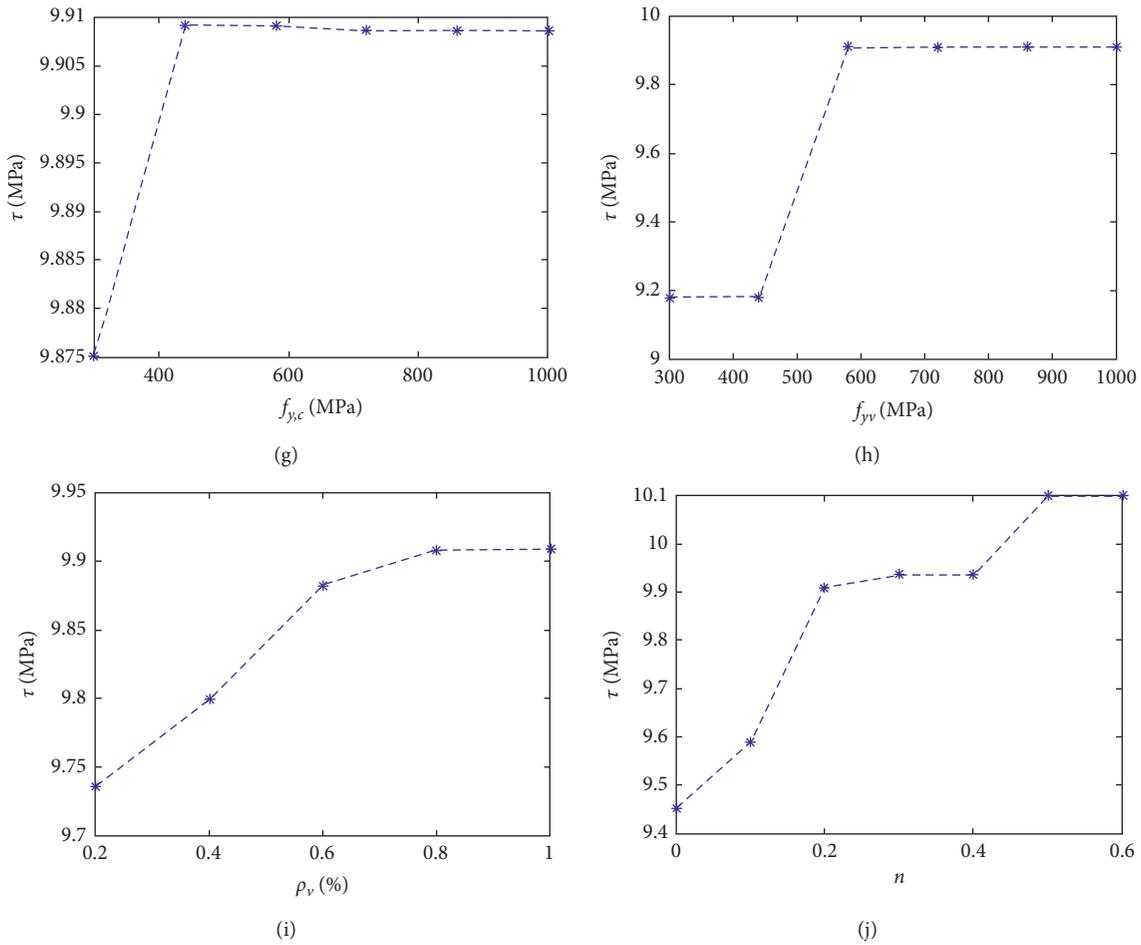


FIGURE 8: Prediction results of parametric analysis. (a) Concrete strength (MPa). (b) Section width of column (mm). (c) Section height of column (mm). (d) Section width of beam (mm). (e) Section height of beam (mm). (f) Yielding strength of beam longitudinal bar (MPa). (g) Yielding strength of column longitudinal bar (MPa). (h) Yielding strength of joint transverse bar (MPa). (i) Transverse bar ratio (%). (j) Axial load ratio.

each input variable on the shear strength of the internal RC joints. The calculation of feature importance can be summarized as follows. Firstly, some out-of-bag (OOB) samples are selected. Secondly, the values of the target input variable are randomly shuffled while other inputs remain unchanged. Then, the feature importance can be calculated as the accuracy difference of the two predictions using the GBRT model. Figure 9 shows the relative feature importance of all input variables. It is clear that concrete strength  $f_c$  is the key feature determining the shear strength of the internal RC joints, which is in accordance with the conclusion obtained from the previous subsection. The influences of the yielding strength of joint transverse bar  $f_{yv}$ , transverse bar ratio  $\rho_v$ , and axial load ratio  $n$  on shear strength are subdominant. The remaining input variables are insignificant features. The feature importance results are also in accordance with the sensitivity results performed before.

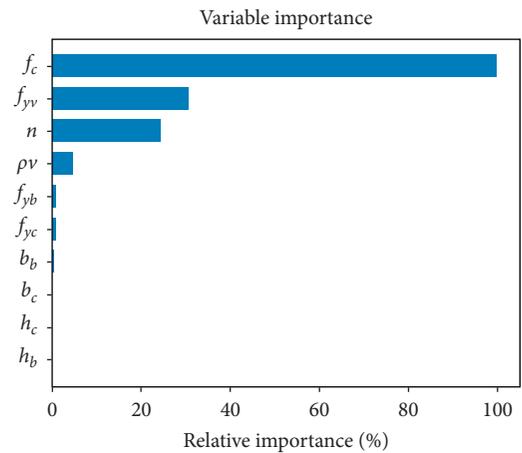


FIGURE 9: Relative feature importance of all input variables.

## 8. Conclusions

This paper presents a ML-based approach to predict the shear strength of internal RC beam-column joints. One of the famous ensemble learning methods, GBRT, is employed to solve the prediction problem. A database of 86 sets of internal RC joint tests is collected from the literature. Some individual-type ML methods and the conventional MCFT method are adopted for comparisons of the developed GBRT prediction model. The model sensitivity analysis of input parameters is conducted for the proposed GBRT-based model. Based on the results, the following conclusions can be drawn:

- (1) The GBRT model can accurately and efficiently predict the shear strength of internal RC beam-column joints with given input variables.
- (2) If 80% of the whole dataset is used to train the GBRT model, the average determination coefficient  $R^2$  of the 10-fold cross-validation is 0.875, which means that the prediction error is low. Meanwhile, the average RMSE and MAE are 0.948 MPa and 0.722 MPa, indicating that the prediction model has a low prediction deviation.
- (3) Among all the ML-based prediction models used in this study, the GBRT model performs best with the closest  $R^2$  to 1 and smallest values of RMSE and MAE. It indicates that the GBRT model is superior to the individual-type ML algorithms.
- (4) The GBRT model has better prediction performance compared with the conventional MCFT model in both average sense and variance sense and exhibits a significant superiority in terms of the three performance indicators.
- (5) Among all the input variables, concrete strength  $f_c$  is the most critical feature affecting the shear strength of the internal RC joints. With the increasing of the concrete strength, the shear strength significantly increases. Other input variables are relatively subordinate or even unimportant.

## Data Availability

The data will be available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The authors greatly appreciate the financial supports from the Natural Science Foundation of Jiangsu Province (Grant no. BK20170680), the National Natural Science Foundation of China (Grant nos. 51708106 and 51908048), the Natural Science Foundation of Shaanxi Province (Grant nos. 2019JQ-021), the Fundamental Research Funds for the

Central Universities, CHD (Grant no. 300102289301), and the Open Project of State Key Laboratory of Green Building in Western China (Grant no. LSKF202007).

## References

- [1] N. W. Hanson and H. W. Connor, "Seismic resistance of reinforced concrete beam-column joints," *Journal of Structural Division ASCE*, vol. 11, no. 56, pp. 533–559, 1967.
- [2] H. Dabiri, A. Kheyroddin, and A. Kaviani, "A numerical study on the seismic response of RC wide column-beam joints," *International Journal of Civil Engineering*, vol. 17, no. 3, pp. 377–395, 2019.
- [3] R. Costa, P. Providência, and A. Dias, "Component-based reinforced concrete beam-column joint model," *Structural Concrete*, vol. 18, no. 1, pp. 164–176, 2017.
- [4] F. J. Vecchio and M. P. Collins, "The modified compression-field theory for reinforced concrete elements subjected to shear," *ACI Structural Journal*, vol. 83, no. 2, pp. 219–231, 1986.
- [5] T. Paulay and M. J. N. Priestley, *Seismic Design of Reinforced Concrete and Masonry Building*, Wiley, New York, NY, USA, 1992.
- [6] X. Long and C. K. Lee, "Improved strut-and-tie method for 2D RC beam-column joints under monotonic loading," *Computers and Concrete*, vol. 15, no. 5, pp. 807–831, 2015.
- [7] C. Lima, E. Martinelli, and C. Faella, "Capacity models for shear strength of exterior joints in RC frames: state-of-the-art and synoptic examination," *Bulletin of Earthquake Engineering*, vol. 10, no. 3, pp. 967–983, 2012.
- [8] T. S. Eom, H. J. Hwang, and H. G. Park, "Energy-based hysteresis model for reinforced concrete beam-column connections," *ACI Structural Journal*, vol. 112, no. 2, pp. 157–166, 2015.
- [9] ASCE/SEI 41, *Seismic Rehabilitation of Existing Buildings*, American Society of Civil Engineers, Reston, VA, USA, 2007.
- [10] H.-J. Hwang, T.-S. Eom, and H.-G. Park, "Shear strength degradation model for performance-based design of interior beam-column joints," *ACI Structural Journal*, vol. 114, no. 5, pp. 1143–1154, 2017.
- [11] H. J. Hwang and H. G. Park, "Requirements of shear strength and hoops for performance-based design of interior beam-column joints," *ACI Structural Journal*, vol. 116, no. 2, pp. 245–256, 2019.
- [12] H.-J. Hwang and H.-G. Park, "Performance-based shear design of exterior beam-column joints with standard hooked bars," *ACI Structural Journal*, vol. 117, no. 2, pp. 67–80, 2020.
- [13] S.-J. Hwang, R.-J. Tsai, W.-K. Lam, and J. P. Moehle, "Simplification of softened strut-and-tie model for strength prediction of discontinuity regions," *ACI Structural Journal*, vol. 114, no. 5, pp. 1239–1248, 2017.
- [14] R. Y. C. Huang and J. S. Kuang, "Predicting strength of exterior wide beam-column joints for seismic resistance," *Journal of Structural Engineering*, vol. 146, no. 2, Article ID 04019209, 2020.
- [15] W. M. Hassan and J. P. Moehle, "Shear strength of exterior and corner beam-column joints without transverse reinforcement," *ACI Structural Journal*, vol. 115, no. 6, pp. 1719–1727, 2018.
- [16] ACI Committee 318, *Building Code Requirements for Structural Concrete (ACI 318-14) and Commentary (ACI 318R-14)*, American Concrete Institute, Farmington Hills, MI, USA, 2014.

- [17] S. Akkurt, G. Tayfur, and S. Can, "Fuzzy logic model for the prediction of cement compressive strength," *Cement and Concrete Research*, vol. 34, no. 8, pp. 1429–1433, 2004.
- [18] A. Baykasoğlu, T. Dereli, and S. Tanış, "Prediction of cement strength using soft computing techniques," *Cement and Concrete Research*, vol. 34, no. 11, pp. 2083–2090, 2004.
- [19] G. Trtnik, F. Kavčič, and G. Turk, "Prediction of concrete strength using ultrasonic pulse velocity and artificial neural networks," *Ultrasonics*, vol. 49, no. 1, pp. 53–60, 2009.
- [20] R. Siddique, P. Aggarwal, and Y. Aggarwal, "Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks," *Advances in Engineering Software*, vol. 42, no. 10, pp. 780–786, 2011.
- [21] A.-D. Pham, N.-D. Hoang, and Q.-T. Nguyen, "Predicting compressive strength of high performance concrete using metaheuristic-optimized least squares support vector regression," *Journal of Computing in Civil Engineering*, vol. 30, no. 3, Article ID 06015002, 2015.
- [22] B. A. Omran, Q. Chen, and R. Jin, "Comparison of data mining techniques for predicting compressive strength of environmentally friendly concrete," *Journal of Computing in Civil Engineering*, vol. 30, no. 6, Article ID 04016029, 2016.
- [23] P. G. Asteris and V. G. Mokos, "Concrete compressive strength using artificial neural networks," *Neural Computing and Applications*, vol. 32, no. 15, pp. 11807–11826, 2019.
- [24] V. Plevris and P. G. Asteris, "Modeling of masonry failure surface under biaxial compressive stress using Neural Networks," *Construction and Building Materials*, vol. 55, pp. 447–461, 2014.
- [25] P. G. Asteris and V. Plevris, "Anisotropic masonry failure criterion using artificial neural networks," *Neural Computing and Applications*, vol. 28, no. 8, pp. 2207–2229, 2017.
- [26] J.-S. Jeon, A. Shafieezadeh, and R. DesRoches, "Statistical models for shear strength of RC beam-column joints using machine-learning techniques," *Earthquake Engineering & Structural Dynamics*, vol. 43, no. 14, pp. 2075–2095, 2014.
- [27] D.-T. Vu and N.-D. Hoang, "Punching shear capacity estimation of FRP-reinforced concrete slabs using a hybrid machine learning approach," *Structure and Infrastructure Engineering*, vol. 12, no. 9, pp. 1153–1161, 2016.
- [28] W. Zheng and F. Qian, "Promptly assessing probability of barge-bridge collision damage of piers through probabilistic-based classification of machine learning," *Journal of Civil Structural Health Monitoring*, vol. 7, no. 1, pp. 57–78, 2017.
- [29] A. Santos, E. Figueiredo, M. Silva, R. Santos, C. Sales, and J. C. Costa, "Genetic-based EM algorithm to improve the robustness of Gaussian mixture models for damage detection in bridges," *Structural Control and Health Monitoring*, vol. 24, no. 3, p. e1886, 2017.
- [30] H. Salehi and R. Burgueño, "Emerging artificial intelligence methods in structural engineering," *Engineering Structures*, vol. 171, pp. 170–189, 2018.
- [31] R. J. Schalkoff, *Artificial Neural Networks*, McGraw-Hill, New York, NY, USA, 1997.
- [32] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and Their Applications*, vol. 13, no. 4, pp. 18–28, 1998.
- [33] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
- [34] Z.-H. Zhou, "Ensemble learning," *Encyclopedia of Biometrics*, S. Z. Li, Ed., pp. 411–416, Springer Press, New York, NY, USA, 2015.
- [35] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, Cambridge, MA, USA, 2012.
- [36] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.
- [37] Research Group on Frame Joints, "Shear strength of reinforced concrete beam-column joints under low reversed cyclic loading," *Journal of Building Structures*, vol. 4, no. 6, pp. 1–17, 1983, in Chinese.
- [38] C. Zhao, D. Zhang, T. Wang, and D. Chen, "Experimental study on the seismic property of the beam-column joints in high strength concrete frame under alternating load," *Journal of Shenyang Architectural and Civil Engineering Institute*, vol. 9, no. 3, pp. 260–268, 1993, in Chinese.
- [39] X. Lu, Z. Guo, and Y. Wang, "Experimental study on seismic behavior of beam-column sub-assemblages in RC frame," *Journal of Building Structures*, vol. 22, no. 1, pp. 2–7, 2001, in Chinese.
- [40] Q. Yu and S. Li, "Research on frame's joint that concrete strength of core is inferior to that of column," *Journal of Tongji University*, vol. 32, no. 12, pp. 1583–1588, 2004, in Chinese.
- [41] B. Xu, M. Cheng, M. Zhang, and J. Qian, "Experimental study on behavior of reinforced concrete beam-column joint with lower core concrete strength," *Building Structures*, vol. 36, no. 6, pp. 18–23, 2006, in Chinese.
- [42] S. Otani, K. Kitayama, and H. Aoyama, "Reinforced concrete interior beam-column joints under simulated earthquake loading," in *Proceedings of the US-New Zealand-Japan Seminar on Design of Reinforced Concrete Beam-Column Joints*, Monterey, Canada, 1984.
- [43] D. E. Meinheit and J. O. Jirsa, "The shear strength of reinforced concrete beam-column joints," CESRL Report No. 77-1, University of Texas, Austin, TX, USA, 1977.
- [44] H. Noguchi and T. Kashiwazaki, "Test on high-strength concrete interior beam-column joints," in *Proceedings of the Tenth World Conference on Earthquake Engineering*, Madrid, Spain, 1992.
- [45] K. Oka and H. Shiohara, "Test on high strength concrete interior beam-column sub-assembly," in *Proceedings of the Tenth World Conference on Earthquake Engineering*, Madrid, Spain, 1992.
- [46] Y. Higashi and Y. Ohwada, "Failing behaviors of reinforced concrete beam-column connections subjected to lateral load," *Memories of Faculty of Technology Tokyo Metropolitan University*, vol. 19, pp. 91–101, University of Virginia, Charlottesville, VA, USA, 1969.
- [47] M. Teraoka, Y. Kanoh, K. Tanaka, and I. C. Hayashi, "Strength and deformation behavior of RC interior beam-and-column joints using high strength concrete," in *Proceedings of the Second US-Japan-New Zealand-Canada Multilateral Meeting on Structural Performance of High Strength Concrete in Seismic Regions*, Honolulu, HI, USA, 1994.
- [48] S. S. Zaid, "Behavior of reinforced concrete beam-column connections under earthquake loading," Doctoral dissertation, University of Tokyo, Tokyo, Japan, 2001.
- [49] S. Fujii and S. Morita, "Comparison between interior and exterior RC beam-column joint behavior," *Design of Beam-Column Joints for Seismic Resistance*, SP-123, pp. 145–165, ACI, Detroit, MI, USA, 1991.
- [50] D. E. Abrams, "Scale relations for reinforced concrete beam-column joints," *ACI Structural Journal*, vol. 84, no. 6, pp. 502–512, 1987.

- [51] R. T. Leon, "Shear strength and hysteretic behavior of interior beam-column joints," *ACI Structural Journal*, vol. 87, no. 1, pp. 3–11, 1990.
- [52] T. Wu, X. Liu, and H. Wei, "Experimental study on seismic behavior of frame interior joints with high-strength lightweight aggregate reinforced concrete," *China Civil Engineering Journal*, vol. 51, no. 6, pp. 32–42, 2018, in Chinese.
- [53] H. H. Zhang, Z. L. Lu, and S. Z. Su, "Experimental study on seismic behavior of high-strength ceramic concrete frame joints," *Building Science*, vol. 32, no. 1, pp. 81–87, 2016, in Chinese.
- [54] B. Qu, "Experimental study on seismic performance of full lightweight aggregate concrete beam-column joints," *Master's thesis*, Jilin Jianzhu University, Changchun, China, 2015, in Chinese.
- [55] C. L. Decker, M. A. Issa, and K. F. Meyer, "Seismic investigation of interior reinforced concrete sand-lightweight concrete beam-column joints," *ACI Structural Journal*, vol. 112, no. 3, pp. 287–297, 2015.