## **ORIGINAL ARTICLE**

Journal of Wood Science

**Open Access** 

# Regression algorithms-driven mechanical properties prediction of angle bracket connection on cross-laminated timber structures

Zhe Wu<sup>1</sup>, Lin Chen<sup>1\*</sup> and Haibei Xiong<sup>1\*</sup>

## Abstract

The construction of structures using cross-laminated timber (CLT) has grown in popularity as a result of its environmentally friendly and high-strength characteristics. The primary function of angle bracket connections is to resist the force of CLT structures under horizontal forces, which is essential to ensure the seismic resilience and ductility of CLT structures. A regression algorithms-driven method for predicting the mechanical performance of angle bracket connections is introduced in this study. As input parameters, the geometric dimensions of the angle bracket connector, the connection method of the connector with the wall and floor slabs, and the properties of the CLT panel are utilized to predict the yield load, the maximal load, the initial stiffness, and the ductility ratio of the angle bracket connection. Prediction models were developed using the collected data from 110 angle bracket experiments, and each prediction model's performance was discussed in depth. Lastly, the permutation importance and SHapley Additive exPlanations (SHAP) value analysis were used to interpret the prediction models. The results showed that the extreme gradient boosting (XGB) algorithm could accurately predict the maximum and yielding load of the angle bracket, the XGB algorithm performed the best with an average ratio of predicted to actual values of 0.985. The results indicated that this study proposed an accurate and efficient method for angle bracket connection to predicting its mechanical properties and confirmed the trustworthiness and feasibility of the prediction models.

**Keywords** Machine learning, Regression algorithm, Cross-laminated timber structure, Angle bracket connection, Mechanical property

## Introduction

The timber structure is frequently cited as the sustainable alternative to concrete and steel construction for its green resources, low gravity, and carbon storage. Mass

\*Correspondence: Lin Chen 22310298@tongji.edu.cn Haibei Xiong xiong204@outlook.com <sup>1</sup> Department of Disaster Mitigation for Structures, Tongji University, Shanghai, China timber projects with high construction efficiency, such as cross-laminated timber (CLT) structures, are increasing globally [1]. CLT structures have been prefabricated wood solutions that have excellent seismic response due to the lightness of engineered CLT panels and the dissipative capacity of the connections. Over the last decade, various application examples of CLT could be found worldwide, like the 18-story building Mjøstårnet, completed in Brumunddal, Norway [2]. As for the codification, European Committee for Standardization (CEN) drafted the second generation of Eurocode in 2012 [3], in which much work has gone into implementing the design



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.gov/licenses/by/4.0/.

rules of the CLT structure. Therefore, it can be observed that CLT structures have positive development prospects, and it is meant to improve the learning of the work properties of CLT structures and their components.

Connections are crucial in providing timber structures with strength, rigidity, stability, and ductility. Extensive research has shown that deformation in CLT structures arises mainly from the bending and slippage of metal connections [4]. Angle bracket connection is a type of CLT structural connection typically evenly arranged along the wall to provide stiffness and strength in the shear direction. Therefore, it is of utmost significance to identify a method for predicting the mechanical performance of angle bracket connection to optimize the design of CLT structures and ensure their seismic safety. However, there are multiple damage mechanisms for angle bracket connections during operation, including the tearing of wood, deformation of steel members, and loss of nail bearing capacity, making evaluating the mechanical properties of angle bracket connections more complicated.

Previously, a considerable amount of research has been performed on the mechanical functionality of angle bracket connections. Gavric et al. [5] accomplished monotonic and low-cycle reversed loading tests on angle bracket connections of different sizes and with varying numbers of fasteners to evaluate and discuss their mechanical characteristics, such as energy dissipation, loss of strength, rigidity, stiffness, and ductility. The study suggested that it is necessary to put the resistance of the fasteners and the characteristics of the CLT walls into consideration when predicting the shear strength of the angle bracket connection. Mahdavifar et al. [6] investigated the influence of various wood densities on the properties of angle bracket joints by conducting shear and uplift experiments on angle bracket connections with two sets of conventional CLT panels and eight sets of hybrid CLT panels. The test findings revealed that if the damage of the screws or bolts penetrated the low-density core layer of the CLT panel, there was a substantial difference in connection efficacy between hybrid and conventional CLT panels. Rezvani et al. [7] built a three-dimensional (3D) numerical model of the angle bracket connection using commercial finite element software ABAQUS to simulate its mechanical properties under different loading combinations. They also introduced a 3D model of the fasteners to conduct a preliminary numerical simulation study of the angle bracket connection. The numerical modeling analysis indicated that replacing nails with screws and adding larger-sized screws did not noticeably improve the shear resistance of the connection. Pošta et al. [8] performed shear experiments on three types of angle bracket connections and compared the experimental results with Eurocode 5 (EC 5) [9]. The results showed that the maximum loads obtained by EC 5 calculations were much higher than those obtained experimentally. The difference was even more remarkable for the angle bracket without a rib, which could be dangerous in practical applications. The above study shows that there is still much room for optimization of mechanical property prediction of angle bracket connections. Mechanical property tests and authentic numerical simulations are time-consuming and costly, so finding more efficient and accurate prediction methods is significant.

Machine learning (ML), a data-driven analytical approach, has become widely used in building construction design and performance evaluation in recent decades [10]. Zhang et al. [11] used nine ML algorithms to build a reinforced concrete (RC) wall seismic performance prediction model based on 429 sets of RC wall test data, including the classification prediction of wall damage modes and the regression prediction of wall lateral stiffness and lateral displacement. Suzuki et al. [12] successfully classified wood damage locations using vibration waveforms combined with ML methods. The specimen waveforms were obtained by piezoelectric sensors, and a classification model was built using a neural network (NN). The results showed that NN could effectively improve the applicability of the wood health monitoring system, with an accuracy of 83.3% for the classification of damaged or undamaged locations. Luo et al. [13] proposed a local ML model named locally weighted least squares support vector regression machine (LWLS-SVMR) to enhance and generalize the estimation of drift capability of RC columns, and the effectiveness of LWLS-SVMR method was verified by comparing it with traditional empirical formulas. The above study shows that using ML in civil engineering has symbolic advantages. However, mechanical property prediction of CLT metal connections using ML algorithms has yet to be reported.

Based on the above research problems of angle bracket connections and the advantages of ML methods, this paper selects input variables and uses 110 sets of angle bracket connection tests and numerical simulation data collected to estimate the mechanical properties of angle bracket connections using four ML regression algorithms: random forest, support vector regression, gradient boosting and extreme gradient boosting. Furthermore, the prediction performance of ML for yielding load, maximum load, initial stiffness, and ductility ratio of angle bracket connections is evaluated. Lastly, this paper provides a parameter importance analysis of the input parameters, and the interpretability analysis of the prediction models is performed to validate the reliability of the prediction models. The method presented in this paper can automatically and efficiently predict the performance of the angle bracket connection, taking into consideration various factors that may affect the performances of the angle bracket connection; the parameter importance analysis and the interpretability analysis of the model serve as an optimization and guide for designing this connection in practical engineering.

The selection of input and output parameters is presented in the next section, and the statistical distribution of each parameter in the database used in this study is described. Then, the process of the proposed ML regression algorithm-driven method for predicting is detailed in the following section. The four regression algorithms and evaluation coefficients used in this study are also described in this section. Subsequently, the outcome of each algorithm and evaluation of the outcomes are discussed by assessing the coefficients. Finally, an interpretability analysis of the prediction model proposed in this paper is provided.

## Preparation of experimental database

#### Selection of input and output parameters

Numerous studies have revealed that when the external load increases, the three phases of mechanical behavior that the angle bracket connection typically exhibits are the elastic, elastoplastic, and failure stages [14]. Accordingly, the four mechanical properties of yielding load  $(F_y)$ , maximum load  $(F_m)$ , initial stiffness  $(K_e)$ , and ductility ratio (D) are selected as the output variables for the angle bracket connection. The yielding displacement  $(v_y)$  and maximum displacement  $(v_m)$  can be obtained based on these output variables, and the simplified bilinear constitutive relationship of this angle bracket connection can be derived (Fig. 1), which will be fundamental to in both the design and research of the angle bracket connection.

To comprehensively quantify the angle bracket connection features, ten variables are selected as inputs in this paper. Four categories can be made up of the input feature variables. The first set of features corresponds to the geometric features of the angle bracket, including the width (B), length (P), height (H), and thickness



(*t*) of the connector (Fig. 2); the second group of features is the thickness of CLT wall panel (*T*); the third group is related to the connection fasteners to the wall, including the self-tapping screw diameter ( $S_r$ ), the self-tapping screw length ( $S_l$ ), the number of self-tapping screws ( $S_n$ ); the last group of features is related to the connection fasteners to the floor, with the ground connection bolts (or screws) diameter ( $B_r$ ), the number of bolts ( $B_n$ ). The units of variables *B*, *P*, *H*, *t*, *T*, *S\_r*, *S\_l*, and  $B_r$  are mm.

#### Description of experimental database

This study collected 110 sets of shear tests [5, 6, 15-36] for angle bracket connections, including 107 sets of experimental data and 3 sets of numerical simulation data (Additional file 1). The distribution of the dataset is



Fig. 2 Geometric parameters of angle bracket



Fig. 3 The distribution of the dataset

shown in Fig. 3. The experimental data within the dataset stem from shear loading tests conducted along the direction of angle brackets. Monotonic or cyclic loading procedures were executed to derive a comprehensive set of mechanical performance parameters for angle brackets. This approach guarantees that the data within the database are amenable to integration for subsequent regression analyses. The minimum error between the numerical simulation results in the dataset and the experimental results is 0.1%, while the maximum error is 35.5%, with an average error of 19.0%. These data were utilized to create ML models to forecast the mechanical characteristics of angle bracket connections.

Figure 4 illustrates the statistical distribution of the input and output variables. The number of data points within the relevant interval is shown on the left y-axis., while the x-axis displays the range of values for the chosen variables. Accordingly, the equivalent cumulative probability is shown on the right y-axis.

Data preprocessing is required before training the prediction model using Scikit-learn to increase the prediction model's accuracy and stability. The data collected were normalized to a range of [0,1] to ensure comparability between features and eliminate the influence of magnitudes. The mean value was utilized to fill in the missing values based on the central tendency of the sample to handle any missing values in the collected information. In this study, while considering the detailed description of angle bracket connections concerning input parameters, a series of measures were taken to account for the influence of non-independent variables. Apart from feature engineering, such as parameter selection and data normalization, in the choice of regression algorithms, an emphasis was placed on selecting ensemble algorithms that exhibit robustness in handling correlation. Model performance with respect to correlation was further improved and validated through hyperparameter optimization and cross-validation techniques.

## Regression algorithms-driven methodology for mechanical estimation of angle bracket

At present, experimental or numerical modeling analysis methods are mainly used to study the mechanical characteristics of angle brackets. In the shear test of angle bracket connection, the variety of parameters is easily limited due to the cost and time required. The mechanical characteristics of angle bracket connections can be effectively simulated using finite element analysis. However, detailed numerical simulations take much time for modeling calculations, and the parameters of metal fasteners and connected wood units are usually missing when performing simulations to improve efficiency. In contrast, the prediction model of mechanical characteristics of angle bracket connection under shear established by ML can better analyze the interrelationship between parameters and accurately and quickly predict the mechanical properties of angle bracket connection under various conditions.

### Framework

Figure 5 gives the proposed framework for predicting the angle bracket connection's mechanical properties using ML. The data collected are split into two parts at random: the training set, which makes up 70% of the total, is used to develop the prediction model, and the test set, which makes up 30% of the total, is used to measure the performance of the prediction model. The same training and test datasets are used for all algorithms to guarantee compatibility among ML algorithms. During model training, hyperparameter optimization is performance out of the algorithms. Finally, the test set not involved in model training is used for prediction model performance evaluation, and the 10 feature values are analyzed for permutation importance and SHAP value.

#### Regression algorithm Random forest

Random forest (RF) is a bagging algorithm based on decision trees [37], in which each iteration selects a subset of data with replacement and a subset of characteristics as inputs randomly (Fig. 6a). In regression, the "ensemble predictor" is created by averaging the output of individual decision trees ( $h_1(x) \\ h_2(x) \\ h_k(x)$ ) (Eq. 1). In each decision tree, the root node determines, in accordance with predefined criteria and conditions, which branch to follow, leading to the internal nodes. Based on the available features, these internal nodes perform assessments to create homogeneous subsets, which are denoted by leaf nodes (or terminal nodes). Since each decision tree, a random forest reduces the possibility of overfitting and improves generalization ability:

$$f(x) = \frac{1}{K} \sum_{i=1}^{K} h_i(x).$$
 (1)

#### Support vector regression

Support vector regression (SVR) is a variant of support vector machine (SVM) and has been extensively used in regression issues [38]. The SVR model is used to find a suitable high-dimensional hyperplane that minimizes the total deviation of all samples from the hyperplane:





Fig. 4 Statistical distribution of input and output variables, depicting minimum (Min), maximum (Max), mean (Mean), and standard deviation (St. Dev) values for a comprehensive overview of the dataset's characteristics



Fig. 5 Prediction workflow of mechanical properties of angle bracket connection



Fig. 6 Diagram of ML algorithm. In **a**, **c**, and **d**, the blue dots represent the root nodes of the decision tree, initiating branching based on specific conditions. These branches, termed 'directed edges' in decision tree terminology, are visualized as one-way arrows. Branches from root nodes lead to internal nodes, as represented with yellow dots, and subsequently to the next level of internal nodes (green dots in **d**) until a stopping condition is met. The final level of nodes, denoted by red dots, represents terminal nodes

$$f(x) = \langle w, x \rangle + b. \tag{2}$$

The notation  $\langle \bullet, \bullet \rangle$  denotes dot product, where *w* is the normal vector of the hyperplane and *b* is the bias term.

In the SVR model, a certain degree of tolerance deviation  $\varepsilon$  is given. When the absolute difference between f(x) and y is within  $\varepsilon$ , the loss value is not calculated, which is equivalent to creating a "margin strip" on both sides of the hyperplane (as shown in Fig. 6(b)), and only the samples falling outside the margin strip are used to calculate the loss.

$$\begin{cases} \min i minimize \frac{1}{2} ||w||^2\\ s.t. |y_i - (wx_i + b)| \le \epsilon, \forall i \end{cases}$$
(3)

#### Gradient boosting

Gradient boosting (GB) is a supervised ML algorithm that trains new weak learners by using the negative gradient information of the loss function of the present model [39]. The existing model is then additively integrated with the trained weak learners (as shown in Fig. 6(c)). For a given training set  $(x, y)_{i=1}^N$ , the GB algorithm uses K weak learners to fit the model  $f_k(x)$ :

$$f_k(x) \leftarrow f_{k-1}(x) + \rho_t h(x, \theta_t), \tag{4}$$

$$(\rho_t, \theta_t) = \operatorname{argmin} \sum_{i=1}^N L(y_i, f_{k-1}(x_i) + \rho h(x_i, \theta)), \quad (5)$$

where  $h(x, \theta)$  is a straightforward parameterized function of input variables *x*, defined by parameters  $\theta_t$ ; the optimal step-size  $\rho$  should be given at each iteration.

#### Extreme gradient boosting

Extreme gradient boosting (XGB) is a scalable ML system for tree boosting. The target function is where it diverges most from the GB algorithm [40].

$$OBj = \sum_{i=1}^{n} l(y_i, \overline{y}_i) + \sum_{k=1}^{K} \Omega(f_k),$$
(6)

where *l* is the loss function used to measure the difference between the true value  $y_i$  of the i-th sample and its predicted value  $\overline{y}_i$ . The model complexity function is represented by  $\Omega$ . The XGB method adds a regularization parameter compared to the GB algorithm to address generalization problems and lessen model complexity (as shown in Fig. 6d).

#### Page 7 of 16

#### Hyperparameter optimization

In the ML model-building process, besides the model parameters estimated by the model from the given data, some parameters that cannot be estimated from the given data, and these parameters are called hyperparameters. The choice of hyperparameters, which are used to control the ML procedure, can impact on the algorithm's robustness, stability, and generalization. Hyperparameter optimization is the task of finding the optimal combination of hyperparameter values to achieve the optimal performance of the model in a reasonable time. The grid search approach is applied for hyperparameter optimization in this paper. For the grid search method, a grid of possible values is created for the hyperparameters, and each iteration is tried in a specific order for the hyperparameter combinations. The performance of the trained model produced by each combination is recorded, and the best model with the best hyperparameters is returned at the end. Table 1 lists the results of hyperparameter optimization for each prediction model in this paper.

#### Evaluation metrics and model interpretability

In this work, the predictive capability of the models was assessed using an impartial testing set. As shown in Table 2, four evaluation metrics—mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), and coefficient of determination (R squared,  $R^2$ )—were used to evaluate the accuracy of each regression model.

However, ML models generally boost their accuracy by increasing model complexity [41], which makes their operation uncertain. Therefore, in addition to verifying the certainty of prediction results, it is equally important to understand why the model makes such predictions and prevent model bias. Permutation importance and SHAP value analysis are techniques used to determine which characteristics affect model fitting most. Permutation importance randomly orders each feature and calculates model changes to overcome the drawback of default feature importance computed with mean impurity decrease. The SHAP explanation method, inspired by cooperative game theory [42], builds an additive explanation model that can reflect each feature's influence and positive or negative effects in each sample.

## **Predicted results and assessment** Prediction results for yielding load

Figure 7 compares the tested normalized yielding loads  $(F_{yt})$  and predicted normalized yielding loads  $(F_{yp})$  for both training and testing sets. The model performs more predictably the closer the data points are near the

**Table 1** Hyperparameter optimization for prediction models

	RF	SVR	GB	XGB
Fy	-	-	-	max_depth = 9 colsample_bytree = 0.8 min_child_weight = 1 subsample = 0.6 eta = 0.5
F <sub>m</sub>	n_estimators=20 max_depth=10	kernel = "linear"	n_estimators=750 min_samples_split=8 max_depth=1 learning_rate=0.05	max_depth = 8 colsample_bytree = 0.8 min_child_weight = 1 subsample = 0.8 eta = 0.1
Ke	n_estimators=20 max_depth=10	kernel = "linear"	-	max_depth = 8 colsample_bytree = 0.8 min_child_weight = 1 subsample = 0.8 eta = 0.4
D	_	_	n_estimators = 100 min_samples_split = 2 max_depth = 3 learning_rate = 0.01	max_depth = 5 colsample_bytree = 0.5 min_child_weight = 3 num_boost_round = 10 subsample = 0.5 eta = 0.1

 Table 2
 Evaluation metrics for the regression model

Evaluation metrics	Definition
MAE	$MAE(y, \widehat{y}) = \frac{1}{n} \sum_{i=1}^{n}  y_i - \widehat{y}_i $
MSE	$MSE(y, \widehat{y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$
RMSE	$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$
R <sup>2</sup>	$R^{2}(y,\widehat{y}) = 1 - \frac{\sum_{i=1}^{n} (y_{i} - \widehat{y}_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \overline{y}_{i})^{2}}$

y,  $\hat{y}$  are the target value and the predicted value for the normalized test dataset, respectively; n is the number of samples in the test dataset; and  $y_{i}$ ,  $\hat{y}_{i}$  refer to the i-th target value and the i-th predicted value for the normalized test dataset, respectively

black line. The figure also includes lines for relative error of  $\pm$  15% and  $\pm$  30% in the testing set. Among the 110 test data points, the RF, SVR, GB, and XGB algorithms have 96, 56, 101, and 105 data samples within the  $\pm$  30% limit. Figure 8 presents the evaluation coefficients of MAE, MSE, RMSE, and R<sup>2</sup> for these models on the test dataset.

Based on the prediction results of angle bracket yielding load,  $R^2$  values for the three ensemble models (RF, GB, and XGB) are all greater than 0.746, indicating that the ensemble models outperform the single model (SVR) (as shown in Fig. 7). However, relying solely on  $R^2$  to determine prediction accuracy may not be proper, and additional evaluation metrics need to be considered. According to overall veracity and performance, XGB is an effective machine learning model for predicting angle bracket yielding load.

It is noteworthy that the RF model has MAE, MSE, RMSE, and  $R^2$  of 0.0207, 0.0008, 0.0297, and 0.956 on the

training set, respectively. However, the predicted results for the testing set have MAE, MSE, RMSE, and  $R^2$  of 0.0514, 0.0096, 0.0983, and 0.746. This indicates that the RF model has poor generalization in predicting the yielding load of angle bracket connections. Therefore, more accuracy in the training set does not always equate to greater accuracy in the test set. Thus, evaluating the prediction model using the training set can reduce the tendency of model overfitting.

#### Prediction results for maximum load

The tested normalized maximum load  $(F_{mt})$  and predicted normalized maximum load  $(F_{mp})$  for the training and testing sets under various ML methods are displayed in Fig. 9. The figure also presents lines representing the relative error of  $\pm 15\%$  and  $\pm 30\%$  in the testing set. Among the 110 test data points, the RF, SVR, GB, and XGB algorithms have 97, 58, 89, and 103 data points within the  $\pm 30\%$  limit, respectively. The distribution of samples indicates that the predictions of RF, SVR, and GB have remarkable errors when the actual values are low, while the overall prediction accuracy of the XGB algorithm is high.

In addition, Fig. 10 shows the evaluation metrics, including MAE, MSE, RMSE, and  $R^2$ , of these models on the testing set. The  $R^2$  of the trained regression models (i.e., RF, SVR, GB, and XGB) are 0.872, 0.839, 0.846, and 0.939, indicating good prediction results [43]. It ought to be noted that the MAE value of the XGB model (0.0362) is significantly better than that of the SVR model (0.0677) because the maximum load values have been normalized



Fig. 7 Comparison of tested yielding load and predicted yielding load of different algorithms



to a small range. Based on the predicted outputs in Fig. 9 and the generalization metrics in Fig. 10, it can be declared that the XGB model performs more accurately than the other three methods in predicting the maximum load of angle bracket connections, with MSE and RMSE values of 0.0024 and 0.0491.

#### Prediction results for initial stiffness

The actual normalized initial stiffness  $(K_{et})$  and the predicted normalized initial stiffness  $(K_{ep})$  of the

training and testing sets under various ML methods are displayed in Fig. 11. The figure also presents straight lines indicating the relative error of  $\pm 15\%$ and  $\pm 30\%$  in the testing set. Among the 110 testing data markers, the RF, SVR, GB, and XGB algorithms have 85, 57, 94, and 96 data points within the  $\pm 30\%$ limit. The distribution of data points reveals that the XGB method has a more robust overall forecast accuracy, while the SVR prediction results have larger errors when the actual values are low.

Moreover, Fig. 12 describes the performance metrics of MAE, MSE, RMSE, and  $R^2$  for these models on the test database. When predicting the initial stiffness of the connection, the MSE and RMSE of the RF algorithm are 0.0164 and 0.1283, and those of the GB algorithm are 0.0148 and 0.1217, which are significantly inferior to those of the SVR and XGB algorithms. The  $R^2$  values of the four regression models (i.e., RF, SVR, GB, and XGB) are 0.642, 0.602, 0.678, and 0.809, respectively. The XGB algorithm has far better generalization performance than the other three algorithms. Therefore, among these four regression models, the XGB algorithm performs best for predicting the initial stiffness of the angle bracket connection.



Fig. 9 Comparison of tested maximum load and predicted maximum load of different algorithms

#### Prediction results for ductility ratio

Figure 13 depicts the predicted normalized ductility ratio  $(D_p)$  and the tested normalized ductility ratio  $(D_t)$  of training and testing sets. The testing set's 15% and 30% relative error lines are also shown in the image, with RF, SVR, GB, and XGB algorithms having 81, 78, 81, and 87 data sets that fall inside the 30% limitations, respectively. It can be observed from the distribution of data sets that SVR and GB models have larger errors in predicting when the actual values are low, while the overall prediction accuracy of the XGB algorithm is higher.



Fig. 10 Evaluation metrics of various ML models for predicting maximum load

Figure 14 shows the assessment indices of MAE, MSE, RMSE, and  $R^2$  of these models on the testing set. When predicting the ductility ratio of the angle bracket connection, the  $R^2$  value of the single model (SVR) is 0.623, which is better than that of the ensemble models (RF, GB, XGB) in terms of generalization performance. Based on the general performance and precision, SVR is a valuable ML model for estimating the ductility ratio of angle bracket connections.

#### Interpretability of prediction model

The XGB model was used to conduct permutation importance and SHAP analysis on the predicted results of the mechanical shear properties of angle bracket connections based on the prediction models in the previous section. The analysis results are shown in Figs. 15 and 16. From the permutation importance analysis results, it can be seen that the width of the angle bracket connector and the number of screws connecting it to the wall panel have the greatest impact on the maximum load and initial stiffness of the connection, with importance coefficients of 44.5% and 24.5%, respectively. Furthermore, the number of bottom anchoring devices and the thickness of the angle bracket have the most significant impact on the yielding load and ductility ratio.



Fig. 11 Comparison of tested initial stiffness and predicted initial stiffness of different algorithms

Based on the SHAP analysis results (Fig. 16), it is evident that the number of fastens used to connect the angle bracket to the wall panel has an enormous influence on the yield strength of the connection. The width of the angle bracket connection is found to have the highest sensitivity with regard to the maximum load and initial stiffness. Additionally, the thickness of the angle bracket connector is observed to have the greatest effect on the ductility coefficient. By comparing the SHAP values of various parameters, it can be concluded that, for maximum load and initial stiffness, the number of self-tapping



Fig. 12 Evaluation metrics of various ML models for predicting initial stiffness

screws, the width of the angle bracket, and the number of bolts have a more significant influence compared to the other parameters. Regarding the maximum load of the connector, the width of the angle bracket has a greater influence than the length and thickness of the angle bracket. However, for the ductility coefficient, the length and thickness of the angle bracket have a more substantial effect than the angle bracket width.

It is crucial to note that permutation importance evaluates the impact of a feature on model performance by randomly shuffling the feature values, while the core idea of SHAP is to calculate the marginal contribution of a characteristic to the model output. When predicting the yielding load of the angle bracket connection, a comparison of the two methods reveals a considerable difference in the impact of the thickness of the angle bracket and the number of self-tapping screws connecting it to the wall panel. Because permutation feature importance mainly measures the model prediction error through single perturbation to determine the importance of features, it cannot consider the correlation between factors. In addition, the thickness of the angle bracket connection in the database is relatively fixed, so the model error will be smaller when perturbing this feature. As a post hoc explanation method for



Fig. 13 Comparison of tested ductility ratio and predicted ductility ratio of different algorithms



Fig. 14 Evaluation metrics of various ML models for predict ductility ratio

models, SHAP analysis provides local and global explanations for the "black box". According to the SHAP analysis approach, the angle bracket connection's thickness affects yield load less than the number of screws used to attach it to the wall panel. Self-tapping screws play a role in bearing shear forces when the angle bracket connection is subjected to shear, so the number of self-tapping screws more strongly influences the yielding load of the angle bracket connection than by its thickness.

#### Discussion

In previous studies on the mechanical properties of angle bracket connections, experimental or numerical modeling analysis methods were typically used. However, the impact of each feature value on mechanical properties was challenging to quantify while incurring high computational time and cost. This work demonstrated that it is feasible to develop a prediction model for the mechanical shear properties of angle bracket connectors using regression algorithms. When performing regression predictions on the yielding load, maximum load, and initial stiffness, ML showed good generalization performance. But when predicting the ductility ratio, the best model achieved an  $\mathbb{R}^2$  of 0.623, indicating significant room for improvement, likely due to limited sample diversity in the current dataset. In addition, due to the limitation of the dataset, the type of steel used for angle bracket, anchor and bolt were not considered with regard to the type of wood used for CLT panels, but this study confirms the feasibility of the prediction method by analyzing the available data. Therefore, expanding the dataset in future studies can effectively improve accuracy.

In the context of this study, the principal objective was predicting the mechanical performance of angle brackets, and the algorithms employed were unsuitable



Fig. 15 Relative importance of input features

for addressing the classification of failure modes. Consequently, this study did not include predictions related to the classification of connection failure modes. Nevertheless, the practical significance of this research is emphasized, as these predictive models can offer valuable insights for engineering design and performance optimization. With the limitations, future research is encouraged to delve deeper into the classification prediction of connection component failure modes, thereby facilitating the provision of more comprehensive insights for engineering applications.

#### Conclusions

This study was based on a database containing 110 sets of angle bracket shear test data and used ML methods to establish predictive models for the yielding load, maximum load, initial stiffness, and ductility ratio of angle bracket connections under shear. The generalization performance and prediction accuracy of different ML methods were analyzed and compared, and thrpretability of ML methods was studied. The results of this study show that:

- 1. XGB algorithm has the highest accuracy in predicting the yielding load and initial stiffness of angle bracket connections, with  $R^2$  values of 0.969 and 0.809. In addition, higher certainty in the training dataset does not automatically imply higher certainty in the test dataset. Evaluating the predictive models with an independent training set can reduce the tendency of model overfitting.
- 2. RF, SVR, GB, and XGB algorithms perform well in predicting the maximum load of angle bracket connections, with evaluation coefficients MSE smaller than 0.068 and  $R^2$  greater than 0.830.
- 3. A single model (SVR) has better generalization performance than ensemble models (RF, GB, XGB) in predicting the ductility ratio of angle bracket connections and is an effective machine learning model for predicting the ductility ratio.



Fig. 16 SHAP value of input features

- 4. Based on the XGB model, permutation feature importance and SHAP value analysis are used to determine the effect of different features on the mechanical properties of angle bracket connections under shear, providing the machine learning model a logical interpretation and enhancing the reliability of the model.
- 5. A database of angle bracket shear test data is established, and the experimental data are systematically organized and shared in Additional file 1, facilitating future research.

#### Abbreviations

CLT	Cross-laminated timber
SHAP	SHapley Additive exPlanations
XGB	Extreme gradient boosting
CEN	European Committee for Standardization
3D	Three-dimensional
EC 5	Eurocode 5
ML	Machine learning
RC	Reinforced concrete
NN	Neural network

LWLS-SVMR	Locally weighted least squares support vector regression
RF	Random forest
SVR	Support vector regression

GB Gradient boostin	g
---------------------	---

- MAE Mean absolute error
- MSE Mean squared error
- RMSE Root mean square error
- R<sup>2</sup> Coefficient of determination

### **Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s10086-023-02110-4.

Additional file 1. The dataset of angle bracket shear tests for prediction model training.

## Acknowledgements Not applicable.

#### Author contributions

ZW conceived the research methodology, executed data collection and organization, performed data analysis and interpretation, and was a major contributor in writing the manuscript. LC contributed to research conceptualization and its scholarly contextualization, and undertaking manuscript review and revision. HX contributed through manuscript review and revision as well. All authors read and approved the final manuscript.

This study is funded by the National Natural Science Foundation of China (51978502) and the National Natural Science Foundation of China (52308327). The financial support is greatly appreciated.

#### Availability of data and materials

The dataset supporting the conclusions of this article is included within the article and its additional file. The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

#### Declarations

Ethics approval and consent to participate Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

Received: 30 August 2023 Accepted: 15 November 2023 Published online: 05 January 2024

#### References

- Ahn N, Dodoo A, Riggio M, Muszynski L, Schimleck L, Puettmann M (2022) Circular economy in mass timber construction: state-of-the-art, gaps and pressing research needs. J Build Eng 53:104562. https://doi.org/ 10.1016/j.jobe.2022.104562
- Tupenaite L, Zilenaite V, Kanapeckiene L, Gecys T, Geipele T (2021) Sustainability assessment of modern high-rise timber buildings. Sustainability 13:8719. https://doi.org/10.3390/su13168719
- Rinaldi V, Casagrande D, Fragiacomo M (2022) Verification of the behaviour factors proposed in the second generation of Eurocode 8 for crosslaminated timber buildings. Earthquake Eng Struct Dynam 52:910–931. https://doi.org/10.1002/eqe.3792
- Ceccotti A, Sandhaas C, Okabe M, Yasumura M, Minowa C, Kawai N (2013) SOFIE project—3D shaking table test on a seven-storey full-scale crosslaminated timber building. Earthquake Eng Struct Dynam 42:2003–2021. https://doi.org/10.1002/eqe.2309
- Gavric I, Fragiacomo M, Ceccotti A (2014) Cyclic behaviour of typical metal connectors for cross-laminated (CLT) structures. Mater Struct 48:1841–1857. https://doi.org/10.1617/s11527-014-0278-7
- Mahdavifar V, Barbosa AR, Sinha A, Muszynski L, Gupta R, Pryor SE (2019) Hysteretic response of metal connections on hybrid cross-laminated timber panels. J Struct Eng. https://doi.org/10.1061/(asce)st.1943-541x. 0002222
- Rezvani S, Zhou L (2019) Numerical modelling analysis of angle bracket connections used in cross laminated timber constructions. Modular and Offsite Construction (MOC) Summit Proceedings 421–428. https://doi. org/10.29173/mocs122
- Pošta J, Hataj M, Jára R, Ptáček P, Kuklík P (2019) Comparison of the use of angle brackets in timber joints with eurocode 5. Constr Build Mater 205:611–621. https://doi.org/10.1016/j.conbuildmat.2019.02.053
- CEN (2004) Eurocode 5: Design of timber structures Part 1–1: General

   Common rules and rules for buildings. European Committee for
   Standardization. https://eurocodes.jrc.ec.europa.eu/EN-Eurocodes/euroc
   ode-5-design-timber-structures. Accessed 26 Aug 2023.
- Sun H, Burton HV, Huang H (2021) Machine learning applications for building structural design and performance assessment: State-of-the-art review. J Build Eng 33:101816. https://doi.org/10.1016/j.jobe.2020.101816
- Zhang H, Cheng X, Li Y, Du X (2022) Prediction of failure modes, strength, and deformation capacity of RC shear walls through machine learning. J Build Eng 50:104145. https://doi.org/10.1016/j.jobe.2022.104145
- 12. Suzuki K, Ito T, Koike K, Kawahara T, Ke M, Mori K (2020) Improvement of generalization performance for timber health monitoring using machine learning. In: APCCAS 2020: Proceedings Of The 2020 IEEE Asia Pacific

Conference On Circuits And Systems (APCCAS 2020). pp 197–200. https://doi.org/10.1109/APCCAS50809.2020.9301662

- Luo H, Paal SG (2019) A locally weighted machine learning model for generalized prediction of drift capacity in seismic vulnerability assessments. Comput Aided Civil Infrastruct Eng 34:935–950. https://doi.org/10. 1111/mice.12456
- Cao J, Xiong H, Chen LX (2020) Procedure for parameter identification and mechanical properties assessment of CLT connections. Eng Struct 203:109867–109867. https://doi.org/10.1016/j.engstruct.2019.109867
- Shen Y, Schneider J, Tesfamariam S, Stiemer SF, Chen ZW (2021) Cyclic behavior of bracket connections for cross-laminated timber (CLT): assessment and comparison of experimental and numerical models studies. J Build Eng 39:102197–102197. https://doi.org/10.1016/j.jobe.2021.102197
- Kržan M, Azinović B (2021) Cyclic response of insulated steel angle brackets used for cross-laminated timber connections. Eur J Wood Wood Prod 79:691–705. https://doi.org/10.1007/s00107-020-01643-5
- Bora S, Sinha A, Barbosa AR (2022) Effect of short-term simulated rain exposure on the performance of cross-laminated timber angle bracket connections. J Arch Eng. https://doi.org/10.1061/(asce)ae.1943-5568. 0000560
- Bora S, Sinha A, Barbosa AR (2021) Effect of wetting and redrying on performance of cross-laminated timber angle bracket connection. J Struct Eng. https://doi.org/10.1061/(asce)st.1943-541x.0003074
- Mahr K, Sinha A, Barbosa AR (2020) Elevated temperature effects on performance of a cross-laminated timber floor-to-wall bracket connections. J Struct Eng. https://doi.org/10.1061/(asce)st.1943-541x.0002737
- Tomasi R, Smith I (2015) Experimental characterization of monotonic and cyclic loading responses of clt panel-to-foundation angle bracket connections. J Mater Civil Eng. https://doi.org/10.1061/(asce)mt.1943-5533. 0001144
- 21. Rezvani S, Zhou L, Ni C (2021) Experimental evaluation of angle bracket connections in CLT structures under in- and out-of-plane lateral loading. Eng Struct 244:112787. https://doi.org/10.1016/j.engstruct.2021.112787
- Xing Z, Zhang J, Zheng C, Lu C (2022) Experimental study and finite element analysis on residual carrying capacity of CLT wall-floor angle bracket connections after fire. Constr Build Mater 328:127113. https://doi. org/10.1016/j.conbuildmat.2022.127113
- Liu J, Lam F (2018) Experimental test of coupling effect on CLT angle bracket connections. Eng Struct 171:862–873. https://doi.org/10.1016/j. engstruct.2018.05.013
- 24. Flatscher G, Bratulic K, Schickhofer G (2015) Experimental tests on cross-laminated timber joints and walls. Proc Inst Civil Eng Struct Build 168:868–877. https://doi.org/10.1680/stbu.13.00085
- Shen Y-L, Schneider J, Tesfamariam S, Stiemer SF, Mu Z-G (2013) Hysteresis behavior of bracket connection in cross-laminated-timber shear walls. Constr Build Mater 48:980–991. https://doi.org/10.1016/j.conbuildmat. 2013.07.050
- Mahdavifar V, Barbosa AR, Sinha A, Muszynski L, Gupta R (2017) Hysteretic behaviour of metal connectors for hybrid (high- and low-grade mixed species) cross laminated timber. ArXiv:171007825 [Physics]. https://doi. org/10.48550/arXiv.1710.07825
- Xing Z, Zhang J, Zheng C (2022) Material model development and fire resistance research on CLT wall-floor angle bracket connections in Open-Sees. Constr Build Mater 347:128605–128605. https://doi.org/10.1016/j. conbuildmat.2022.128605
- Tomasi R, Sartori T (2013) Mechanical behaviour of connections between wood framed shear walls and foundations under monotonic and cyclic load. Constr Build Mater 44:682–690. https://doi.org/10.1016/j.conbu ildmat.2013.02.055
- Gavrić I, Ceccotti A, Fragiacomo M (2011) Experimental cyclic tests on cross-laminated timber panels and typical connections. In: XIV Convegno ANIDIS. https://www.researchgate.net/publication/332529065. Accessed 26 Aug 2023.
- 30. Ma Y (2022) Properties of cross-laminated timber panel with lowvalue lumber and performance of cross-laminated timber wall system subjected to seismic and sequential seismic-wind loadings (Order No. 29161967). Available from ProQuest Dissertations & Theses Global; ProQuest Dissertations & Theses Global A&I: The Sciences and Engineering Collection. (2666888565). https://www.proquest.com/dissertati ons-theses/properties-cross-laminated-timber-panel-with-low/docview/ 2666888565/se-2. Accessed 26 Aug 2023.

- Omar Amini M, Rammer John DR, Pei S (2021) Rocking behavior of high-aspect-ratio cross-laminated timber shear walls: experimental and numerical investigation. J Build Eng. https://doi.org/10.1061/(ASCE)AE. 1943-5568.0000473
- Brown JR, Li M, Sarti F (2021) Structural performance of CLT shear connections with castellations and angle brackets. Eng Struct 240:112346. https://doi.org/10.1016/j.engstruct.2021.112346
- D'Arenzo G, Rinaldin G, Fossetti M, Fragiacomo M, Flavio F, Manuela M (2018) Tensile and shear behaviour of an innovative angle bracket for CLT structures. In: WCTE 2018 - World Conference on Timber Engineering, 2018. https://www.researchgate.net/publication/327112626. Accessed 26 Aug 2023.
- Cao JX, Xiong H, Zhang F-L, Lin C, Carlos RC (2020) Bayesian model selection for the nonlinear hysteretic model of CLT connections. Eng Struct 223:11118–11118. https://doi.org/10.1016/j.engstruct.2020.111118
- Pozza L, Saetta A, Savoia M, Talledo D (2018) Angle bracket connections for CLT structures: experimental characterization and numerical modelling. Constr Build Mater 191:95–113. https://doi.org/10.1016/j.conbu ildmat.2018.09.112
- D'Arenzo G, Rinaldin G, Fossetti M, Fragiacomo M (2019) An innovative shear-tension angle bracket for Cross-Laminated Timber structures: experimental tests and numerical modelling. Eng Struct 197:109434. https://doi.org/10.1016/j.engstruct.2019.109434
- Cutler A, Cutler DR, Stevens JR (2012) Random forests. In: Ensemble machine learning. Springer, New York, pp 157–175. https://doi.org/10. 1007/978-1-4419-9326-7\_5
- Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14:199–222. https://doi.org/10.1023/b:stco.0000035301.49549. 88
- Natekin A, Knoll A (2013) Gradient boosting machines, a tutorial. Front Neurorobot. https://doi.org/10.3389/fnbot.2013.00021
- Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16. pp 785–794. https:// doi.org/10.1145/2939672.2939785
- Linardatos P, Papastefanopoulos V, Kotsiantis S (2020) Explainable AI: a review of machine learning interpretability methods. Entropy 23:18. https://doi.org/10.3390/e23010018
- Lundberg S, Lee S-I (2017) A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 30:4765–4774. https://proceedings.neurips.cc/paper/2017/hash/8a20a 8621978632d76c43dfd28b67767-Abstract.html. Accessed 26 Aug 2023.
- Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science 7:e623. https:// doi.org/10.7717/peerj-cs.623

#### **Publisher's Note**

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

## Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- ▶ Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at > springeropen.com