

Seismic response prediction of a damped structure based on data-driven machine learning methods

Tianyang Zhang ^{a,b,c}, Weizhi Xu ^{a,b,c,*}, Shuguang Wang ^c, Dongshen Du ^c, Jun Tang ^b

^a China-Pakistan Belt and Road Joint Laboratory on Smart Disaster Prevention of Major Infrastructures, Southeast University, Nanjing, China

^b Quakesafe Technologies Co., Ltd, Kunming, China

^c College of Civil Engineering, Nanjing Tech University, Nanjing, China

ARTICLE INFO

Keywords:

Damped structure
Structural response prediction
Interpretable machine learning
Deep learning
Data-driven

ABSTRACT

Damping technology has been widely used because of its good vibration control effect. However, due to the strong nonlinearity of the added dampers, accurately predicting the seismic response of damped structures remains a challenge. This study investigates the application of interpretable machine learning (ML)-based and deep learning-based approaches to the prediction of the maximum inter-storey displacement of a damped structure. A comprehensive database consisting of 13,855 structural responses to ground motions was collected. Seven traditional interpretable ML algorithms including random forest and extreme gradient boosting (XGBoost), a convolutional neural network based on large receptive field, and seismic wave transformer (SWT) model based on a transformer network were developed. The predictions show that the error of the SWT based on unsupervised feature extraction is reduced by 50.90% compared with that of the optimal XGBoost in ensemble learning. Although the SWT has the highest global accuracy, XGBoost is found to have a smaller error when the structure is in a linear state with peak ground acceleration as partition index, so an aggregation model (AM)-based structural response prediction method was also proposed. The accuracy of the AM improved by 27.95% compared with that of the SWT. In contrast with other ML models, the proposed AM is more advantageous in terms of computational efficiency and accuracy.

1. Introduction

Earthquakes are among the most catastrophic natural events that can cause substantial losses in human productivity and livelihood. Their impact is significant and poses a severe threat to human life and property, resulting in adverse consequences, such as social instability [1–3]. In the past two decades, 313 earthquakes with a magnitude of 7.0 or higher, have been recorded globally; most have occurred in developing countries. For these countries, strengthening and retrofitting old buildings to improve seismic resistance are an important approach to achieve regional resilience [4–8]. Utilising metal or viscous dampers [9–12] to enhance energy dissipation in structures is crucial for reducing the probability that structural damage surpasses specified limits. However, the structural response of damped structures to seismic forces is typically intricate owing to the non-linear feature of dampers. The use of finite element (FE) modelling and the conduct of numerous time history analyses by incremental dynamic analysis [13–16] can accurately

determine the structural response of damped structures. Nevertheless, for large-scale community buildings, the computational expense of employing this approach is considerable owing to the inherent computational burden.

Recent efforts in structural response prediction have focused on the development of data-driven methodologies employing machine learning (ML) algorithms to extract features from tests or numerical simulations [17–19]. These algorithms are relatively straightforward to implement and can efficiently generate maps between input and output parameters [20–24]. Xu et al. [25] proposed an ML-based method for the regional seismic performance assessment of reinforced concrete bridges throughout their lifecycle. Furthermore, they developed ML models by considering 480 bridge models and evaluated their performance. Results indicated that interpretable ML methods exhibited high sample classification accuracy [26]. Wang et al. [27] used five ML methods to predict the structural responses of tall pile-supported bridges. Their results indicated that artificial neural networks and gradient boosting trees

* Corresponding author at: China-Pakistan Belt and Road Joint Laboratory on Smart Disaster Prevention of Major Infrastructures, Southeast University, Nanjing, China.

E-mail address: xuwz@njtech.edu.cn (W. Xu).

could effectively predict the seismic response of structures and rubber bearings. Hou et al. [28,29] used ML algorithms to predict the bearing capacity of 2045 circular concrete-filled steel tube columns, and the results showed that the Gaussian process regression model could better predict the strength of concrete-filled steel tubes, with higher accuracy and wider applicability than the existing methods in current design standards. Kazemi et al. [30–34] conducted a series of studies on the seismic response of RC structures and SMRF structures. The results showed that the data-driven methods for structural response prediction solved the problems of computationally expensive, time-consuming, and complex analysis that exist in the current traditional methods.

Among deep learning (DL) algorithms, convolutional neural networks (CNNs) exhibit inherent adaptive design, rendering them highly suitable as unified frameworks for simultaneous feature extraction and classification tasks. Since Krizhevsky et al. [35] proposed AlexNet, CNN has become more popular in the fields of image recognition and semantic segmentation, tending to replace other ML algorithms. Currently, the use of long short-term memory (LSTM) networks for structural response prediction is increasingly common [36–38]. Anirban et al. [39] utilised a CNN and recurrent neural network (RNN) to determine the response uncertainty of non-linear stochastic dynamic systems. They proposed a DL algorithm based on LSTM for seismic response uncertainty quantification. The method effectively resolved the randomness of dynamic loads and uncertainty of structural system parameters. To reduce the computational cost of the stochastic seismic response analysis, Huang et al. [40] employed a one-dimensional (1D) CNN and an LSTM network based on the seismic response characteristics of an underground structure. They implemented seismic response modelling for a two-layer three-span subway station using a data-driven approach. Under the same training sample conditions, the 1D CNN outperformed the LSTM and baseline multi-layer perceptron models in terms of predictive performance and extrapolation ability; the LSTM model had the worst performance.

Although the 1D CNN demonstrates an exceptional capability in predicting structural responses, the receptive field of the convolutional operation is extremely small and better at extracting local features. In contrast, transformer networks offer significant advantages in extracting features from long time series data. The transformer network [41] utilises attention mechanisms to improve the model training speed. It has been widely applied to natural language processing. Generative pre-trained transformers [42] are currently attracting considerable attention in various fields. Chen et al. [43] proposed a transformer network based on a customised classification network for a structural damage assessment framework. Compared with the RNN and CNN, transformer networks have been proven to be more effective in predicting the elastic-plastic responses of structures. Xu et al. [44] selected eight datasets of different scales as well as analysed and compared typical models. They also compared the long-term and short-term memory capabilities of LSTM networks and transformers. They proved that LSTM networks had excellent short-term memory capabilities. Attention mechanism networks, such as transformers, have excellent global (long-term) memory.

Although numerous studies have used different ML methods to predict the seismic response behaviour of structures, seismic response predictions for damped structures are currently lacking. Therefore, this study selects several commonly used interpretable ML models and develops a CNN model based on a large receptive field and a response prediction model based on transformers. Finally, a fast detection model combining different ML methods is devised to improve prediction accuracy.

2. Overview of prototype structure and dataset

To evaluate the potential of the proposed data-driven ML methods, shaking table tests are conducted on a two-storey damped steel frame structure [45]. A numerical model is established based on test results,

and time history analysis is conducted to create a dataset.

2.1. Damped prototype structure and numerical simulation

The two-storey prototype structure used in this study is a typical damped steel frame with metal and viscous dampers installed on the first and second floors, respectively. Each damper type is connected to steel beams using asymmetric steel braces and bolts. The plan dimensions of the frame are 3250 mm × 2640 mm, and the floor height is 2940 mm. The floor of each level in the prototype structure is 80 mm thick and made of 30 MPa grade concrete. The mass of each level is 4000 kg, resulting in a total structure mass of 15140 kg. The steel frame columns and beams have a nominal yield strength of 235 MPa. The beams and columns are HW200 × 150 × 6 × 9 mm and HW150 × 150 × 7 × 10 mm, respectively. The yield strength and displacement of the metal damper are 13.97 kN and 3.09 mm, respectively. The damping coefficient (C_d) of the viscous damper is 587.76 N/(mm/s) $^\alpha$, and the velocity exponent (α) is 0.44. During the experiment, the accuracy of the numerical model was validated by monitoring the dynamic response of the structure using linear variable displacement transformers and acceleration sensors installed on each layer.

A three-dimensional FE model representing the damped structure is built using the OpenSeesPy framework. To model the non-linear properties of the beams and columns, force-based fibre elements employing the trapezoidal quadrature rule with six integration points along their lengths are utilised. The non-linear behaviour of the slabs of the frame is modelled using ShellMITC4 elements, and the dampers of the structure are modelled using TwoNodeLink elements. The metal damper in the numerical model is modelled using the Steel02 constitutive model, whereas the viscous damper is made of a Maxwell material. Before modelling, the parameters of the dampers are verified through individual damper tests. The simulation results of the viscous and metal dampers are found to be in good agreement with the test results. Then, the TwoNodeLink elements with the assigned properties are installed into the numerical model of the steel frame. The measured first and second mode periods of the structure are 0.367 and 0.125 s, respectively. The numerical simulation periods are 0.397 and 0.124 s. Under the excitation of a Terminal Island wave with a magnitude of 0.6 g, the peak displacement and acceleration errors between the numerical model and prototype structure are both less than 10%. The damping structure and detailed procedure of the numerical simulation are illustrated in Fig. 1.

2.2. Dataset and intensity measures

To establish an ML regression dataset with a large amount of data, 13,855 natural earthquake records were obtained from the Pacific Earthquake Engineering Research Centre database. The acceleration, velocity, and displacement spectra of ground motion with a damping ratio of 2% based on 13,855 records are shown in Fig. 2. The median, 16/84 percentile, and 2.5/97.5 percentile are marked in the figure to illustrate the uncertainty of the ground motion.

In conventional interpretable ML models, 21 seismic intensity measures (IMs) are extracted as input variables into the regression models. The methods used to define each IM are summarised in Table 1. The Pearson product moment correlation coefficient (r) is a statistical indicator reflecting the degree of correlation among variables; it is defined as follows:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \sqrt{n\sum y^2 - (\sum y)^2}}. \quad (1)$$

The correlations among the 21 seismic IMs are shown in Fig. 3. When $r > 0$, a positive correlation between two variables is observed, and when $r < 0$, the variables are negatively correlated. Note that $|r| < 0.4$, $0.4 \leq |r| < 0.7$, and $0.7 \leq |r| < 1$ represent low, moderate, and high

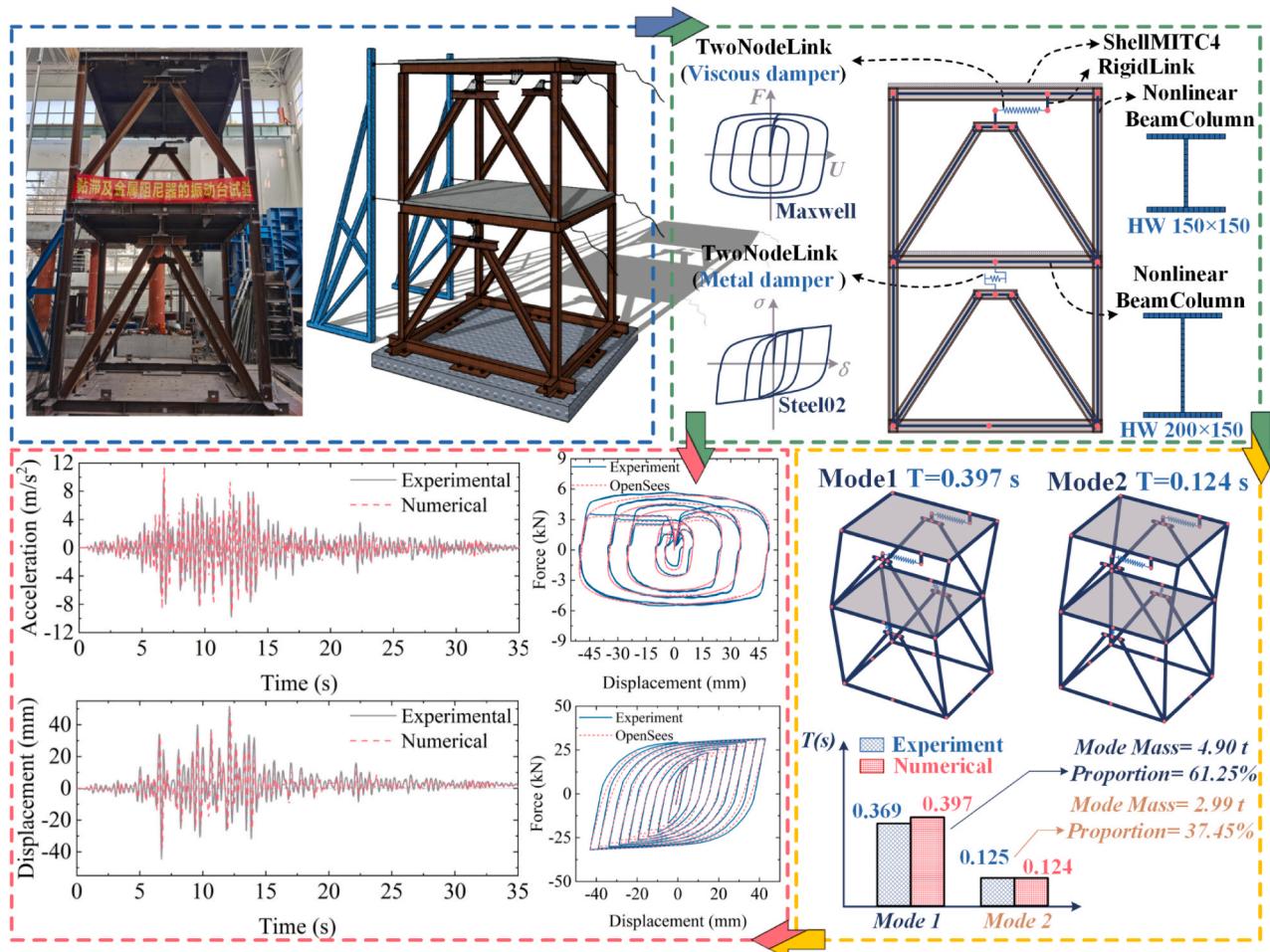


Fig. 1. Prototype structure and FE model.

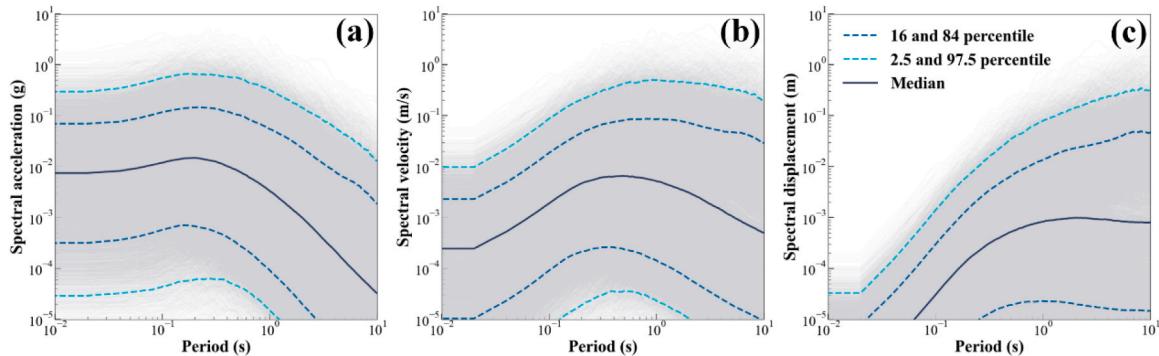


Fig. 2. Response spectrum of 13 855 ground motion records (2% damping ratio): (a) acceleration, (b) velocity, and (c) displacement.

degrees of linear correlation, respectively.

To ensure the richness of the input parameters for interpretable ML models, the 21 IMs in this study are used for predicting the maximum inter-storey displacement of structures and compared with the method of directly using seismic wave inputs for DL models. The normalisation method was used for the ML input of ground motion intensity measures with the following formula:

$$x_{new} = \frac{x - \mu}{\sigma} \quad (2)$$

where x is the feature input to the model, μ is the mean of the features, and σ is the standard deviation of the features. With this standardisation

method, the features are treated as distributions with a mean of 0 and a standard deviation of 1. The integrated learning model and the DL model based on direct input of ground motion in this study do not require additional processing of input features.

3. Regression methods: interpretable ML and DL models

Seven traditional supervised interpretable ML models and two DL models were used to predict the maximum inter-storey displacement responses of the prototype structures. Traditional ML algorithms include ridge regression (Ridge), lasso regression (Lasso), support vector machine (SVM), k-nearest neighbour (K-NN), decision tree (DT), random

Table 1
Seismic IMs for conventional interpretable ML models.

IM	Name	Definition
EPV	Effective peak velocity	$\frac{1}{2.5}S_v$ ($T = 1.0$); S_v is the 2% damped spectral velocity
AI[27]	Arias intensity	$\frac{\pi}{2g} \int_0^{T_{\text{total}}} a^2(t) dt$; $a(t)$ is the time history of acceleration
HI	Housner intensity	$\int_{0.1}^{2.5} PS_v(\xi = 2\%, T) dt$; PS_v is the pseudo-spectral velocity
CAV [46]	Modified cumulative absolute velocity	$\int_0^{T_{\text{total}}} a(t) dt$
PGA	Peak ground acceleration	$\text{Max} a(t) $
PGV	Peak ground velocity	$\text{Max} v(t) $; $v(t)$ is the velocity time history
PGD	Peak ground displacement	$\text{Max} d(t) $; $d(t)$ is the displacement time history
S_{ai}	Spectral acceleration at i s	$S_a(\xi = 2\%, T = i)$; S_a is the 2% damped spectral acceleration
S_{vi} S_{amax}	Spectral velocity at i s Maximum spectral acceleration	$S_v(\xi = 2\%, T = i)$ $\text{Max}(S_a)$
S_{vmax} Period of S_{amax}	Maximum spectral velocity Period corresponding to maximum value of acceleration response spectrum	$\text{Max}(S_v)$ $T[\text{argmax}(S_a)]$; $\text{argmax}(S_a)$ is the index of maximum spectral acceleration
Period of S_{vmax}	Period corresponding to maximum value of velocity response spectrum	$T[\text{argmax}(S_v)]$; $\text{argmax}(S_v)$ is the index of maximum spectral velocity
$S_{\text{amax}}/S_{\text{aT1}}$	Ratio of maximum spectral acceleration to spectral acceleration in first mode period	$\text{Max}(S_a)/S_a(\xi = 2\%, T = T1)$
$S_{\text{amax}}/S_{\text{aT2}}$	Ratio of maximum spectral acceleration to spectral acceleration in second mode period	$\text{Max}(S_a)/S_a(\xi = 2\%, T = T2)$
S_{vmax}/S_{VT1}	Ratio of maximum spectral velocity to spectral velocity in first mode period	$\text{Max}(S_v)/S_v(\xi = 2\%, T = T1)$
S_{vmax}/S_{VT2}	Ratio of maximum spectral velocity to spectral velocity in second mode period	$\text{Max}(S_v)/S_v(\xi = 2\%, T = T2)$

forest (RF), and extreme gradient boosting (XGBoost). In addition, two DL-based structural response prediction methods are proposed in this paper: CNN and seismic wave transformer (SWT).

3.1. Ridge and Lasso

In the multivariate linear regression (MLR) model [47], the relationship between the predicted output variable, \hat{y} , and set of input variables, X , is defined as follows:

$$\hat{y} = w_0 + w_1x_1 + \dots + w_nx_n = w_0 + \sum_{i=1}^n w_i x_i. \quad (3)$$

When multi-collinearity exists in the design matrix, X , parameter w has an extremely high numerical value. A small change in the input variable (x) causes a significant change in the output. Ridge and Lasso [47] reduce the sensitivity of the model to the noise present in the input variable. To limit parameter w , a penalty term ($\lambda \sum_{j=1}^p w_j^2$) is added to the MLR, and the w term in the objective function is defined using the L2 norm, as follows:

$$\hat{w}^{\text{Ridge}} = \text{argmin}_w \left\{ \sum_{i=1}^n \left(y_i - w_0 - \sum_{j=1}^p w_j x_i \right)^2 + \lambda \sum_{j=1}^p w_j^2 \right\}. \quad (4)$$

The L1 norm is used as the penalty term for Lasso; it is expressed as

$$\hat{w}^{\text{Lasso}} = \text{argmin}_w \left\{ \sum_{i=1}^n (y_i - w^T x_i)^2 + \lambda \sum_{j=1}^p |w_j| \right\}. \quad (5)$$

3.2. SVM and K-NN

SVM algorithm [48] uses kernel function to transform nonlinear problem into linear problem by increasing dimension. Common kernel functions include linear basis, sigmoid basis, polynomial basis and radial basis. The radial basis kernel function used in this research, because of its good generalisation performance, it is expressed as:

$$K(x_i, x_j) = \exp \left(-\frac{1}{2} \left(\frac{\|x_i - x_j\|}{\sigma} \right)^2 \right) = \exp(-\gamma \|x_i - x_j\|^2) \quad (6)$$

where x_i and x_j are two variables, and σ is the width of the radial basis function.

K-NN is based on some distance measure to find out the k nearest training samples in the training set to the test sample, and then use the “voting method” to determine the category of the test sample based on the k training samples found. The distance measure we usually choose is L_p distance, which is calculated as follows:

$$L_p(x_i, x_j) = \left(\sum_{l=1}^n \left| (x_i)^{(l)} - (x_j)^{(l)} \right|^p \right)^{\frac{1}{p}} \quad (7)$$

when $p = 1$ the L_p is called the Manhattan distance, and when $p = 2$ the L_p is called the Euclidean distance.

3.3. DT and RF

The DT algorithm [49] is a non-parametric ML model widely used in regression tasks. As the name suggests, DT provides classification results through a tree-like decision-making process. The tree structure comprises a root node and several internal and terminal nodes. In regression problems, the classification and regression tree algorithm uses least square deviation to measure the splitting result of the model and selects the branch that minimises the result of all possible options for the split.

The RF model [49] is an extension of the bagging algorithm used for ensemble learning. It combines multiple DTs to determine the final output instead of relying on a single DT. From the perspective of deviance-variance decomposition, bagging mainly focuses on reducing variance; therefore, it is more effective for unpruned DTs, neural networks, and other learners susceptible to sample perturbation. The RF algorithm provides predictions by averaging individual predictions from N DTs in regression tasks or by voting in classification tasks. The algorithm is given by

$$F(x) = \frac{1}{N} \sum_{n=1}^N f_n(x), \quad (8)$$

where $F(x)$ represents the RF model; N represents the number of DTs; $f_n(x)$ represents a single DT; and x is the input feature.

3.4. XGBoost

Different from RF and gradient boosting DT (GBDT), the loss function of XGBoost [50] adds a regularisation term to the GBDT to control the complexity of the model and prevent overfitting. During the optimisation process, the loss function is expanded using a second-order Taylor series and incorporates both first-order and second-order derivative information. Consequently, significant improvement is observed relative to traditional ensemble learning methods. The objective function solution for XGBoost is as follows:

$$\mathcal{L} = \sum_{i=1}^T \left[G_i w_i + \frac{1}{2} (H_i + \lambda) w_i^2 \right] + \gamma T, \quad (9)$$

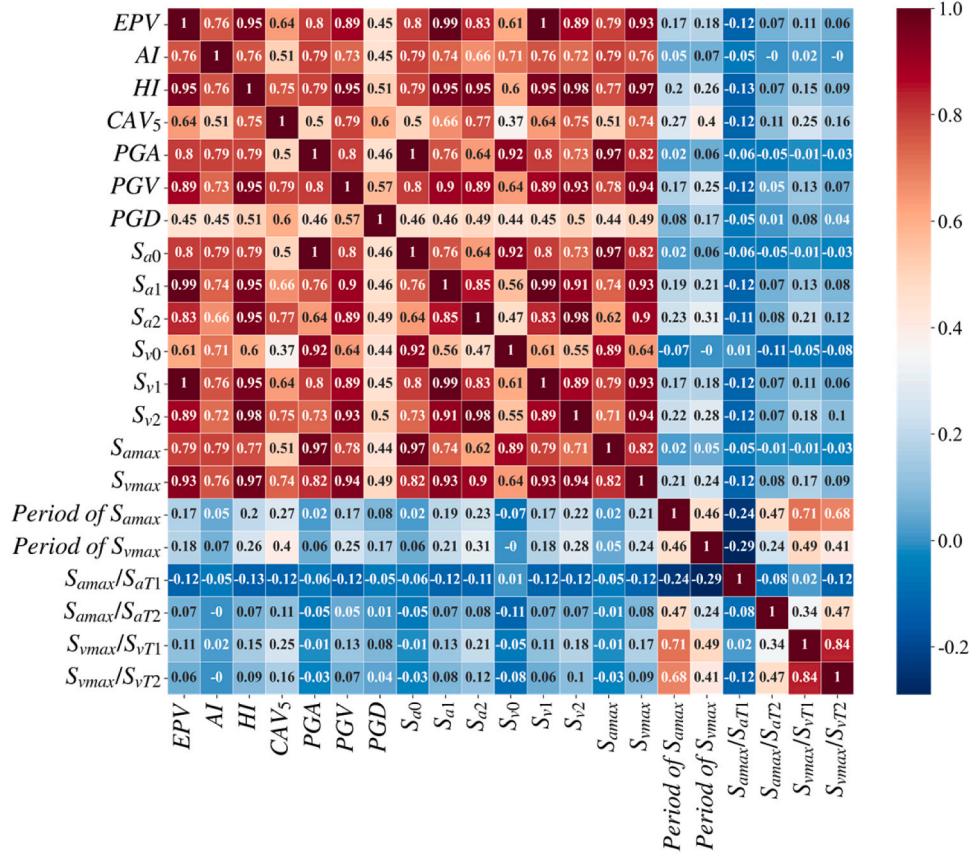


Fig. 3. Pearson product moment correlation coefficient matrix.

where G_i is the sum of the first partial derivatives of the sample in leaf node I ; H_i is the sum of the second partial derivatives; and w_i is the weight of the leaf node.

3.5. CNN and SWT

In contrast to traditional interpretable ML algorithms, CNN models

based on DL technology [51] can directly extract features from input seismic waves without requiring manual feature selection. The structure of the proposed CNN model is shown in Fig. 4. By padding the seismic waves, the lengths of 13,855 waves are aligned and inputted into the convolutional layer. Because the duration of the seismic waves after filling exceeds 240 s and the amount of data is considerable, the proposed CNN model adopts large convolutional kernels to obtain a large

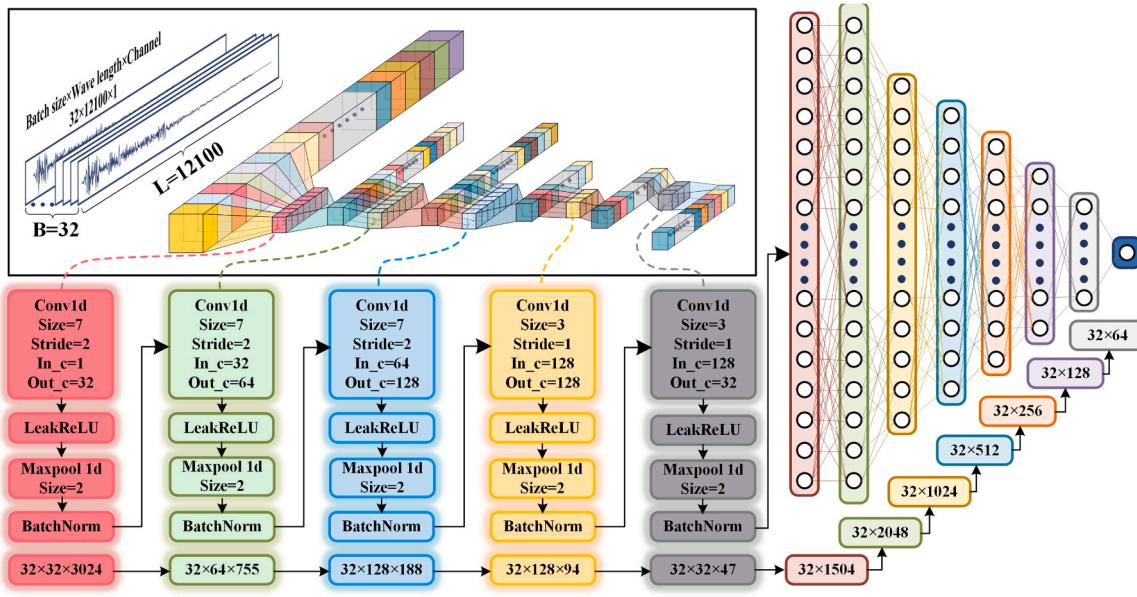


Fig. 4. CNN structure.

receptive field for the convolutional layer. Further, after each convolution operation, activation, pooling, and normalisation layers are added to enhance the non-linearity of the network to prevent overfitting.

When training data are sufficient, the performance of the transformer [52] is typically superior to that of the CNN. The proposed SWT divides the input seismic waves into multiple patches and projects each patch as a vector of fixed length into the transformer encoder. In addition to patch embeddings, the SWT also requires another special type of embedding called position embedding. Unlike CNNs, transformers require position embeddings to encode the positional information of tokens. This is mainly because self-attention is permutation-invariant, which means that rearranging the tokens in a sequence does not change the output. Without providing the model with the position information of patches, the model would need to learn the layout based solely on the semantics of the patches, which would increase the learning cost. In the encoder layer, a multi-head self-attention mechanism is used to capture the internal features of the seismic waves. In the SWT, the seismic wave data of each patch are mapped through fully connected layers, resulting in three matrices: Q , K , and V . The outputs of the self-attention layer [43] are calculated using Eq. (10):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (10)$$

where the dimensions of Q and K are both d_k , and the dimension of V is d_v . In the multi-head sublayer, the attention function is performed in parallel for each mapped version. The output values are then concatenated and re-mapped, resulting in the final values. Multi-head attention enables the model to attend to information from various representation subspaces simultaneously and at different positions. The multi-head layer can be calculated using Eqs. (11) and (12):

$$\text{Head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), \quad (11)$$

$$\text{Multihead} = \text{Concat}(\text{Head}_1, \text{Head}_2, \dots, \text{Head}_h)W^0, \quad (12)$$

where i is a natural number from 1 to h . The mapping results are the parameter matrices: $W_i^Q \in \mathbb{R}^{k \times d_k}$, $W_i^K \in \mathbb{R}^{k \times d_k}$, $W_i^V \in \mathbb{R}^{k \times d_v}$, and $W^0 \in \mathbb{R}^{k \times hd_v}$. As the network depth increases, several problems, such as gradient dissipation and explosion, arise. When stacking the encoder

layers 12 times, skip connection and multiple normalisation are adopted to resolve the problem of gradient dissipation in the backpropagation process, facilitating the training of deep networks. The structure of the proposed SWT model is shown in Fig. 5.

The CNN model had 6,011,567 trainable parameters, and the SWT model proposed in this research had 17,214,721 trainable parameters in total. In order to prevent overfitting of the model, dropout layers are added to the CNN and SWT networks to randomly turn off some neurons to prevent the model from over-relying on certain specific features. In addition, the batch normalization layer was added to make the data distribution of each layer unchanged and do normalisation processing, which speeded up the convergence speed of the model, avoided the disappearance of gradient and improved the accuracy.

The loss function in the regression problem evaluates the dissimilarity between the predicted values derived from the network and observed displacement. Mean square error (MSE), given by the following, is adopted in this research:

$$\mathcal{L}_{\text{MSE}}(\theta_j) = \frac{\|\hat{x}(\theta_j) - x\|_2^2}{N}, \quad (13)$$

where θ_j represents all network parameters at epoch j , and N is the number of batches. Additionally, the adaptive moment optimiser [53], which is an extension of stochastic gradient descent, is adopted in this study. It calculates the adaptive learning rates for different parameters using the estimates of the first and second moments of the gradients.

4. Regression results of the ML models

The reliability of the proposed ML models is verified by training 80% of the data in the seismic wave dataset and using the remaining 20% for testing. Since the RMSE calculation results in the same dimension as the prediction target, the RMSE result is more representative. Therefore, this study focuses on comparing the differences in RMSE of each model. The coefficient of determination (R^2) is used to quantify the strength of the relationship between the observed and predicted displacements. The foregoing is defined as follows:

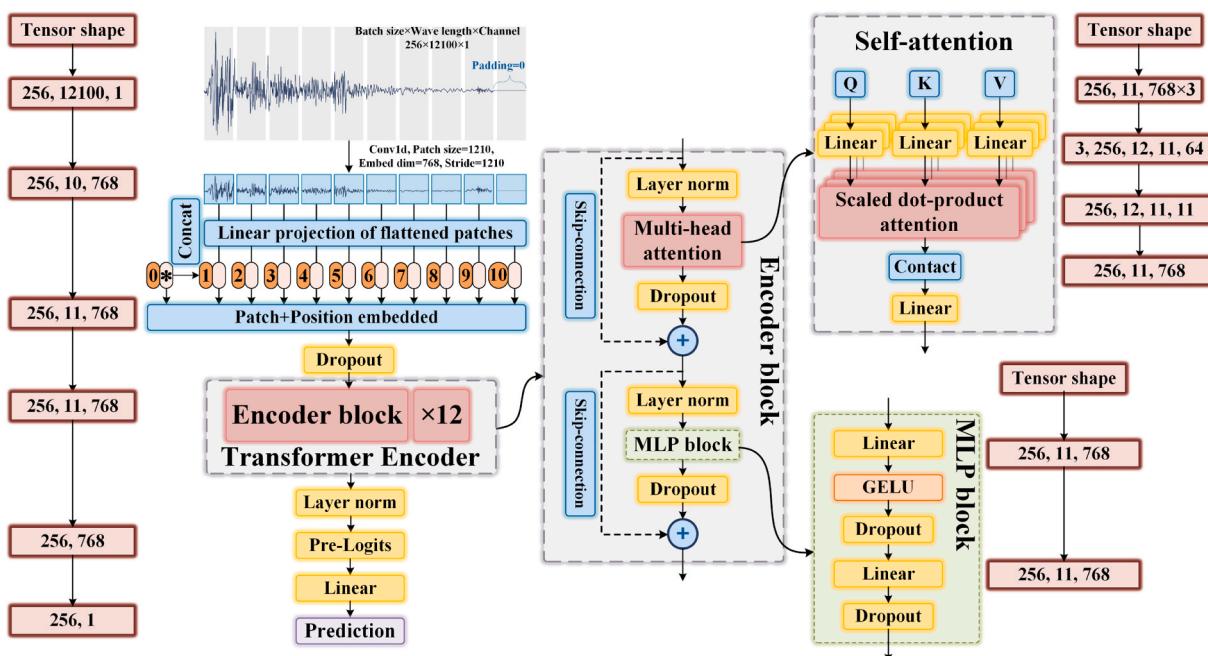


Fig. 5. SWT structure.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{N}}, \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (P_i - \bar{O})^2}. \quad (15)$$

where P_i is the predicted results, O_i is the observed results, N is the total dataset volume.

4.1. Prediction of maximum inter-storey displacement response

Comparisons between the test and predicted results from the regression models for the training and testing data are shown in Fig. 6. The prediction results indicate that K-NN has the worst accuracy because the number of neighbours (K) significantly affect the prediction

results. If K is extremely small, the presence of noisy components has a considerable impact on prediction; otherwise, the model becomes simple, resulting in underfitting. In general, the Ridge, Lasso, and SVM demonstrate relatively similar performance. The RMSE values for the three methods are 0.75, 0.76, and 0.64 mm, respectively. The superior regression effect of SVM on non-linear problems can be attributed to its utilisation of a Gaussian kernel, as shown in Fig. 6(c). The DT model is prone to overfitting, resulting in low generalisation and susceptibility to unbalanced samples. This can lead to computation results that are locally optimal rather than globally optimal. Therefore, the use of a single DT typically leads to inadequate predictions with an RMSE of 1.2 mm, which is only slightly less than the RMSE of K-NN, as shown in Fig. 6(e).

The predicted results of interpretable ensemble learning methods, such as RF and XGBoost, are shown in Fig. 6(f) and (g). The RF algorithm uses statistical probability models, random sampling, and random

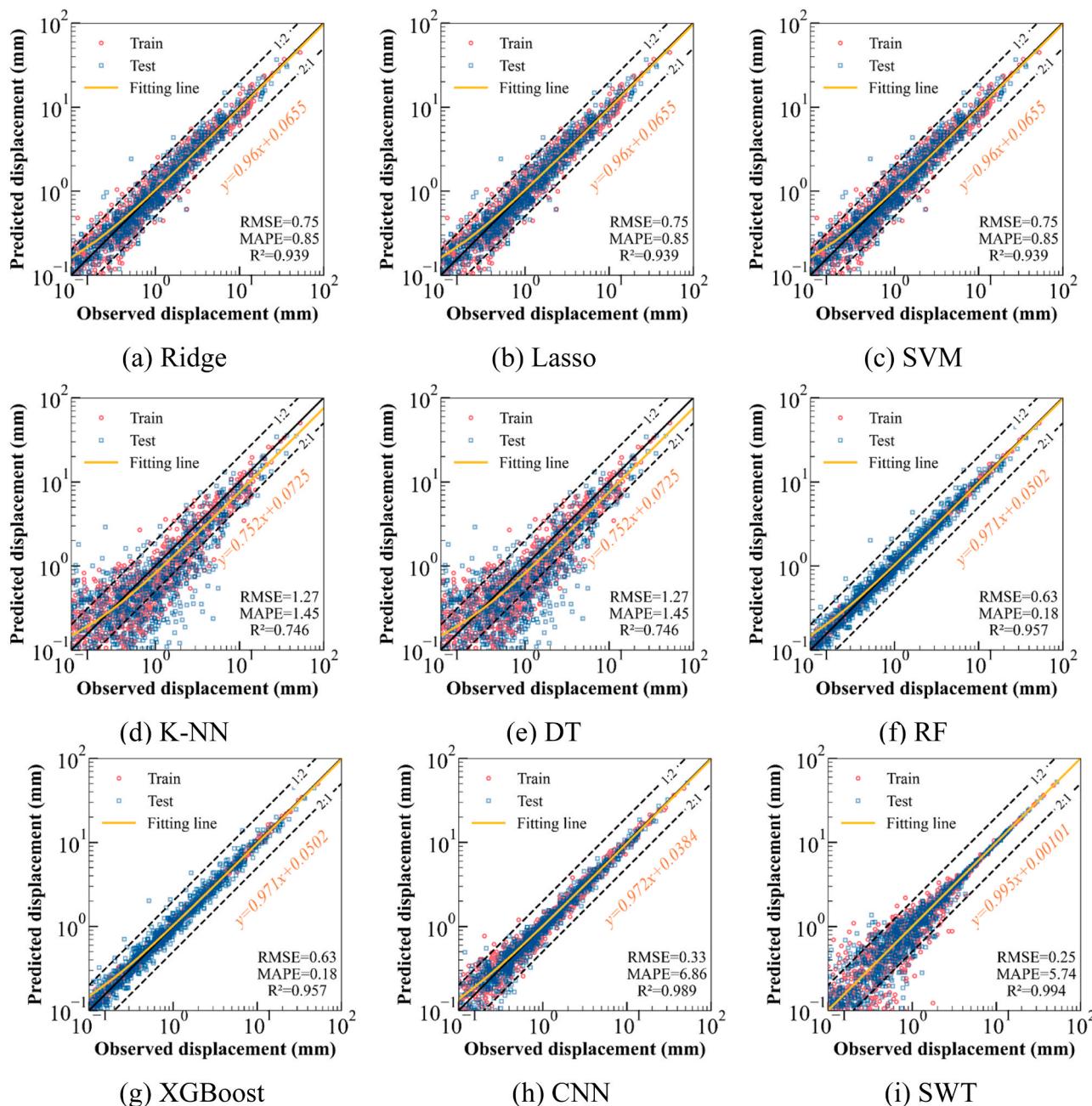


Fig. 6. Predicted results of regression models.

construction of tree nodes to enhance generalisation. The XGBoost algorithm improves accuracy by continuously enhancing learning from incorrectly learned knowledge by boosting the residuals; however, this may lead to overfitting. The prediction results indicate that the two algorithms have similar accuracy; however, XGBoost has more hyperparameters and a higher iteration speed, thus providing more possibilities for tuning the parameters.

The predicted results of the relative inter-storey displacement response based on the DL models are shown in Fig. 6(h) and (i). Compared with other traditional ML models, the XGBoost and RF models show significant reductions in RMSE. In contrast to XGBoost, which performs well in interpretable ML models, the RMSEs of the CNN and SWT decrease by 46.77% and 59.68%, respectively. The CNN has an excellent prediction ability when the relative displacement is small. When the displacement is greater than 5 mm, the prediction accuracy of the SWT is better than those of all the other models.

4.2. Model interpretation

The decision-making process for the five-depth DT model is illustrated in Fig. 7. The first depth of the model uses S_{a0} for decision-making because S_{a0} is virtually equal to the peak ground acceleration (PGA) in terms of redundant information. In the five-depth structure, PGA can be used for most decisions; however, for predicting large inter-storey deformation (S_{amax} and S_{vmax}), PGD and other indicators require support for regression prediction. The RMSE of the five-depth structure is 1.18 mm, R^2 is 0.87, and the fitting results show a ladder-like distribution. As the depth of the tree structure increases, the number of leaf nodes increases, and the regression results are more accurate.

To identify the significant factors for the seismic responses of inter-storey displacement, the relative importance of the input variables

derived from the different ML approaches is shown in Fig. 8. For linear regression algorithms (Ridge regression or Lasso regression), this research directly uses regression coefficient as the feature importance, which reflects the contribution of each input feature to the prediction result. For the RF algorithm, each feature calculates a Gini coefficient. The importance of the feature is the normalised value of the decrease in this coefficient. For the Ridge model, S_{a2} is the most important feature of the model, followed by PGV; both are considered equally important for the PGA and S_{a0} . For the Lasso model, PGA, PGV, and S_{amax} are the three most important features, accounting for 39.97%, 22.88%, and 22.25%, respectively. The regression results of the RF model show that the main variable involved in the decision is S_{a0} . The RF model cannot effectively filter redundant information, and its PGA accounts for a relatively high proportion, whereas S_{amax} only accounts for 6.03%. Although the RF model obtains relatively accurate regression results, its single feature accounts for a high proportion, and the frequency characteristics of the ground motions are ignored.

Shapley additive explanation (SHAP) is an effective solution based on a game theory for explaining supervised ML models by computing the contribution of each feature when added to the model. The distribution of SHAP values for each feature parameter is shown in Fig. 9(a); the figure also indicates the corresponding influence trend. The x axis represents specific SHAP values, and the y axis represents the input features sorted by their importance. Each point represents an individual sample in the entire dataset with the colour transitioning from blue to red to indicate increasing values. Positive and negative SHAP values indicate positive and negative correlations, respectively, between features and predictions. Parameters PGA, S_{amax} , PGV, and S_{vmax} have the greatest influence on the output of the XGBoost model. The hierarchical clustering of features is performed based on the results of input features. As calculated by SHAP, PGA and S_{a0} are more than 50% redundant. By

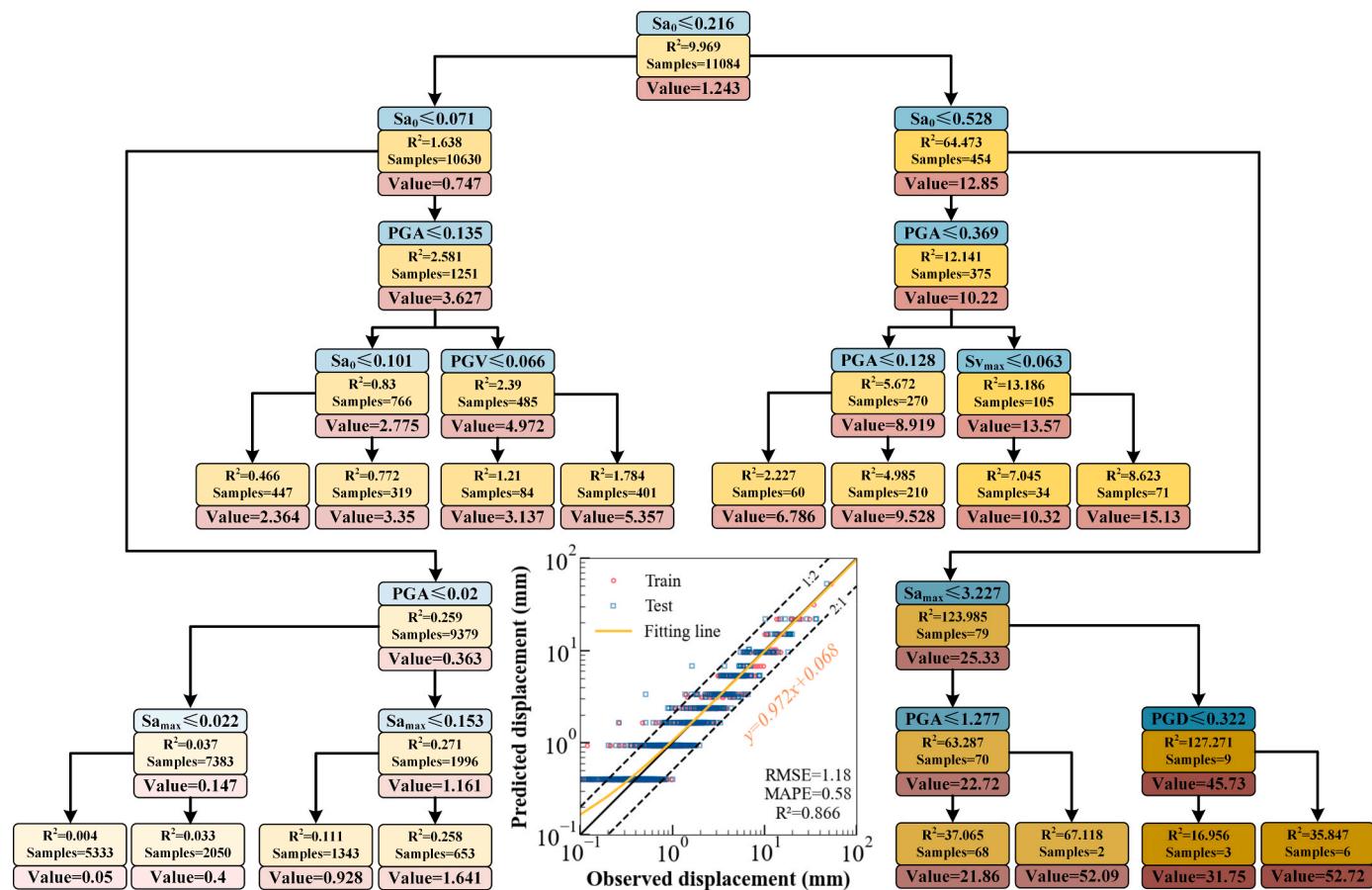


Fig. 7. Structure of five-depth DT model.

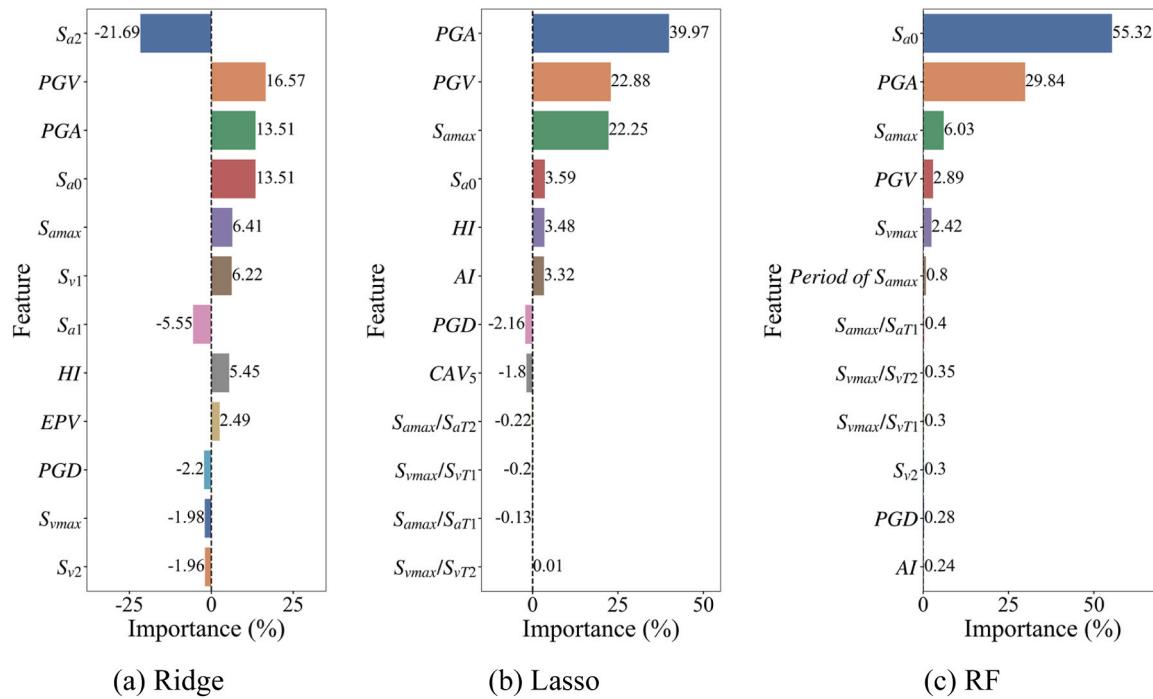


Fig. 8. Relative importance of input variables to regression models.

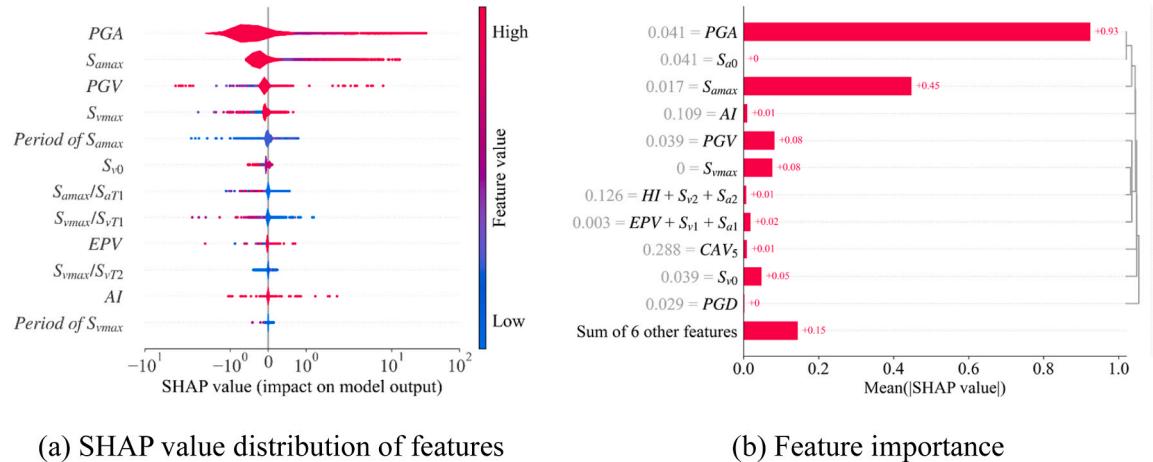


Fig. 9. Explanatory results of XGBoost by SHAP.

extracting the SHAP values, XGBoost is shown to be more credible than other interpretable ML models. It combines multiple seismic intensity features to predict the maximum inter-storey deformation of structures and exhibits better robustness than RF.

5. Development of ML-based models

Various ML methods are used to predict the maximum inter-storey displacement of a damped structure. Analysing the parameter sensitivity of each model can improve the prediction accuracy of the model to develop a more suitable ML model for damped structures.

5.1. Sensitivity analysis of model hyperparameters

Regarding the DT model, the minimum number of samples required to split a node (N_{split}), number of leaf nodes (N_{leaf}), and tree depth (N_{depth}) are the three hyperparameters used to analyse the accuracy of the model. The sensitivity of the RMSE to the three hyperparameters is

shown in Fig. 10(a). Results show that the prediction accuracy of RF is negatively correlated with an increase in N_{split} , whereas N_{leaf} and N_{depth} are positively correlated. The correlation between the unbalanced variable and RMSE can be obtained by calculating the Mahalanobis distance (D_M) between two variables. The Mahalanobis distance is defined as follows:

$$D_M(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}, \quad (16)$$

where Σ is the covariance matrix of a multi-dimensional random variable, and μ represents the sample mean. Moreover, D_M shows that N_{depth} has a significant impact on the RMSE of the model. When N_{depth} is less than 4 or more than 12, the accuracy of the model rapidly improves with increasing N_{depth} ; in other cases, the improvement is gradual.

The influence of the two hyperparameters, N_{tree} and N_{depth} , on the RMSE of the two interpretable ML models (RF and XGBoost) in the form of 3D and 2D contour maps is shown in Fig. 10(b) and (c), respectively. For the RF model, when N_{tree} is less than 20, the prediction error

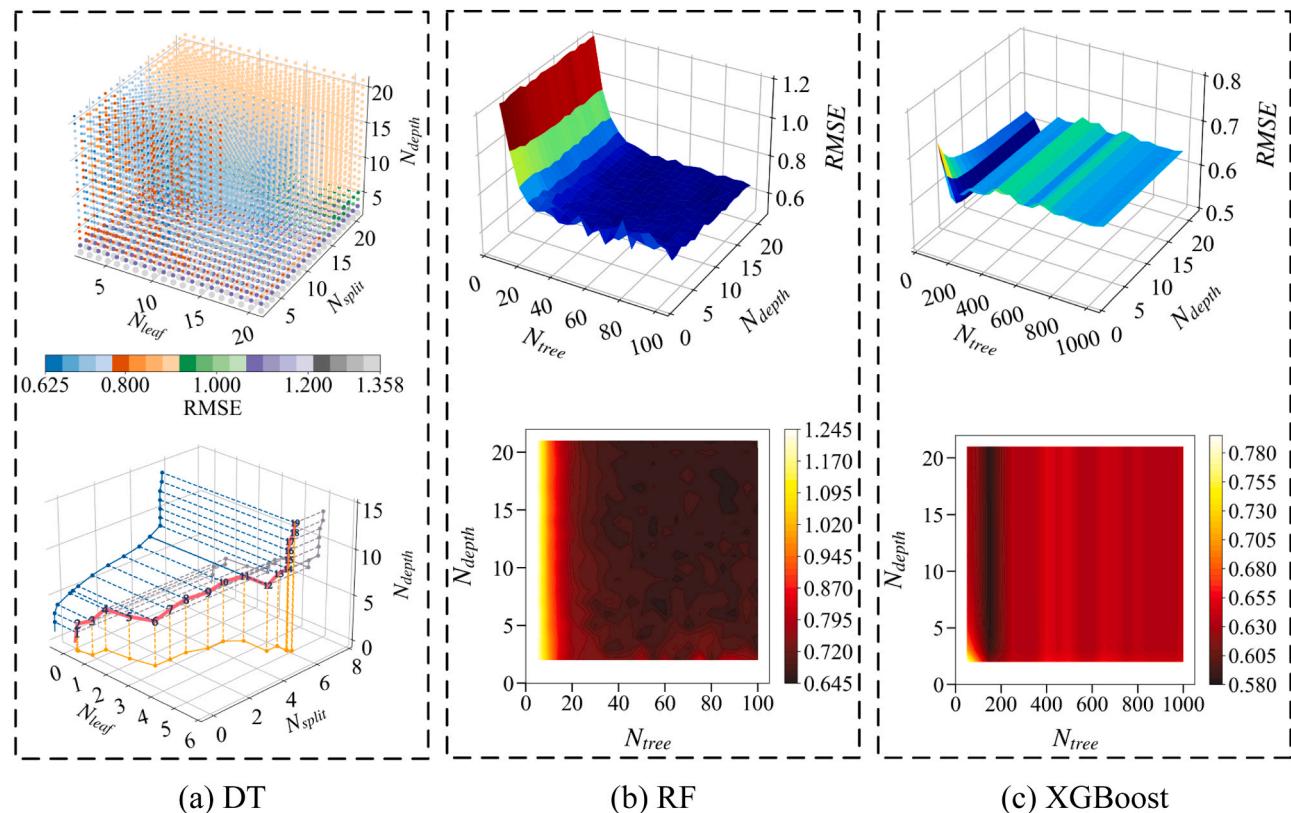


Fig. 10. Parameter sensitivity analysis.

decreases by 30.62% with increasing N_{tree} . As N_{tree} further increases, the prediction accuracy of the model tends to be constant. When N_{depth} is less than 5, the accuracy of the model sharply oscillates. The accuracy of the XGBoost model decreases when the number of N_{tree} exceeds 200. Even an extremely small N_{depth} does not affect the output of the model. The XGBoost model has better accuracy and is more efficient than RF.

5.2. Model optimisation and aggregation

As a boosting algorithm in ensemble learning, XGBoost performs better than RF, which is a bagging algorithm, in predicting the seismic responses of damped structures. Compared with the SWT whose RMSE is minimum, XGBoost exhibits higher prediction accuracy when inter-storey deformations are small. However, in the case of large

displacements, the SWT exhibits extremely high prediction accuracy, with the regression curve approximating a straight line, as shown in Fig. 11(a). Through the forward sequencing of all predicted results, the prediction error of XGBoost is found to exceed that of the SWT when the maximum interlayer displacement is greater than 6.78 mm, and the RMSE of the SWT is relatively gentle, as shown in Fig. 11(b).

The absolute percent error (APE) is used to calculate the prediction error of XGBoost algorithm and SWT algorithm in each *PGA* segment. The definition of APE is as follows:

$$APE = \frac{P_i - O_i}{O_i} \times 100 \quad (17)$$

In the boxplot in Fig. 12, the horizontal line in the middle is the median of the predicted results, and the red dot in the box is the mean

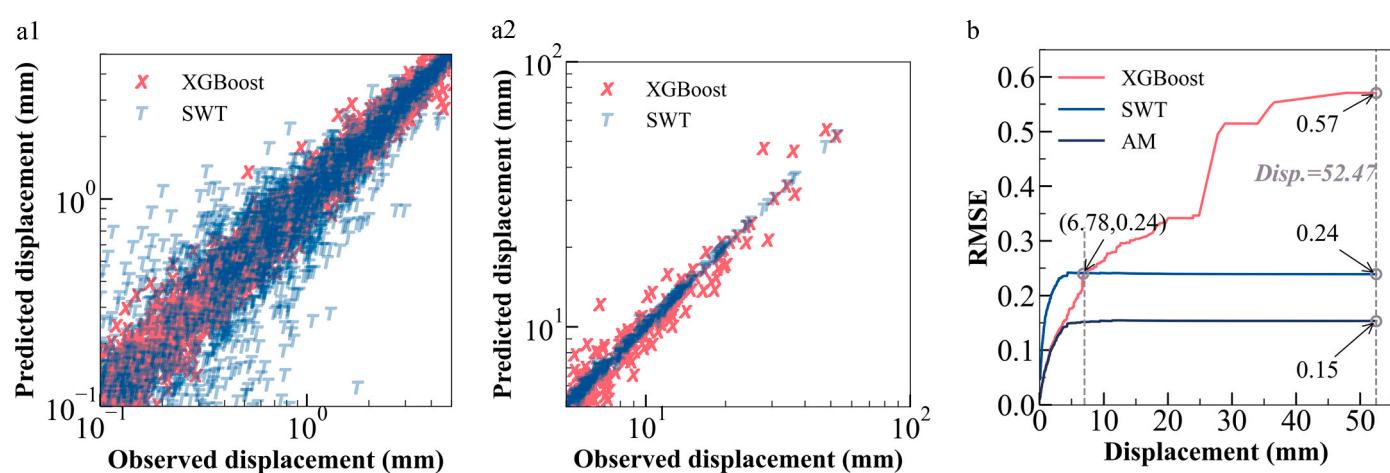


Fig. 11. Comparison between XGBoost and SWT.

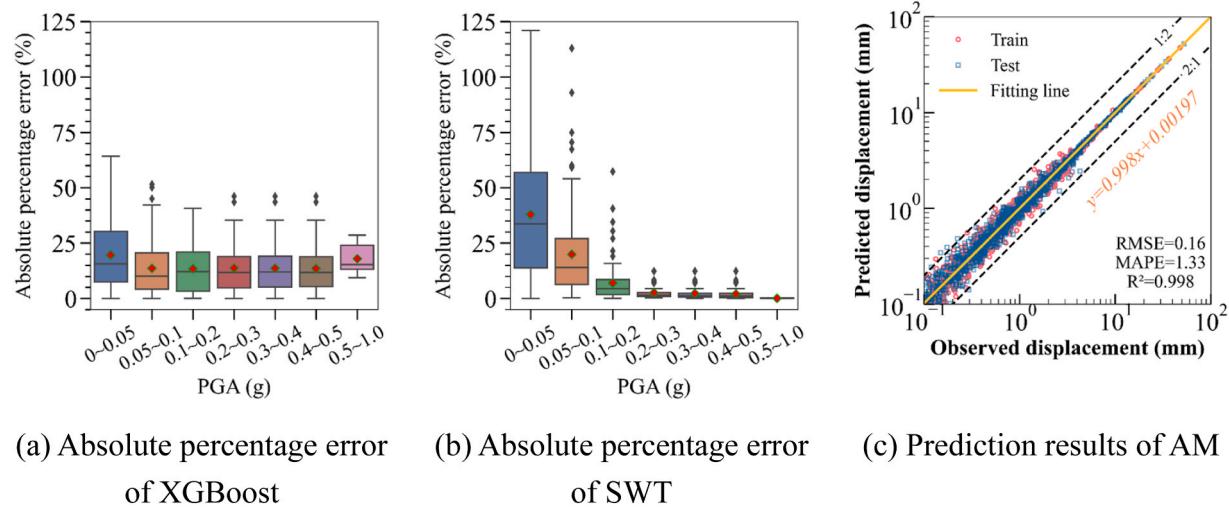


Fig. 12. Performance of AM.

absolute percent error (MAPE), defined as follows:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{P_i - O_i}{O_i} \right| \times 100 \quad (18)$$

When the PGA is less than 0.05 g, the average error rate of XGBoost is 19.65%, whereas the error rate of the SWT is 37.88%. When the PGA of the input seismic wave is in the range 0.05–0.1 g, the prediction errors of XGBoost and SWT are 13.71% and 19.91%, respectively. When the PGA is greater than 0.1 g, the prediction accuracy of the SWT surpasses that of XGBoost. The error of XGBoost gradually increases with the PGA , and the error of the SWT tends to be stable, as shown in Fig. 12(a) and (b), respectively. The aggregation model (AM) can obtain more accurate prediction results through the segmentation of seismic waves, as shown in Fig. 12(c). The RMSE of the AM was close to that of XGBoost under small structural deformations, and its accuracy region was flat under large displacements. Compared with those of XGBoost and SWT, the RMSE of the AM decreased by 73.68% and 37.50%, respectively, on the test set, as shown in Fig. 11(b).

When choosing among multiple models with different structures, hyperparameters, etc., k-fold cross-validation can be used to identify the best model for a specific dataset. The dataset containing N samples is divided into K subsets; one subset is used as the test set, and the remaining $K - 1$ subsets are used as the training set, as shown in Fig. 13;

the cross-validation results are presented in Fig. 14. Compared with other interpretable ML models, XGBoost has the best prediction accuracy. The AM proposed in this study achieves the best results. The average RMSE of cross-validation is considerably lower than that of XGBoost; this is also a significant improvement over the CNN and SWT models. In the ten-fold cross-validation, the prediction accuracy of CNN and SWT based on DL algorithm was 50.90% higher than that of XGBoost. After AM algorithm was adopted, the overall prediction accuracy was further improved, its RMSE decreased by 72.91% compared with XGBoost. The ten-fold cross-validation results of AM algorithm were stable, which proved that it had good generalisation ability.

6. Discussion

The change in structural stiffness reflects the non-linear condition of an entire structure. The non-linear degree of the structure is characterised by its dynamic secant stiffness (K_s) under maximum inter-storey displacement. The lateral stiffness calculation formula for the structure in the elastic stage is

$$K_s = \frac{48EI_c}{h^3} \frac{12\rho + 1}{12\rho + 4} \quad (19)$$

where E is the elastic modulus of steel; I_c is the section moment of inertia

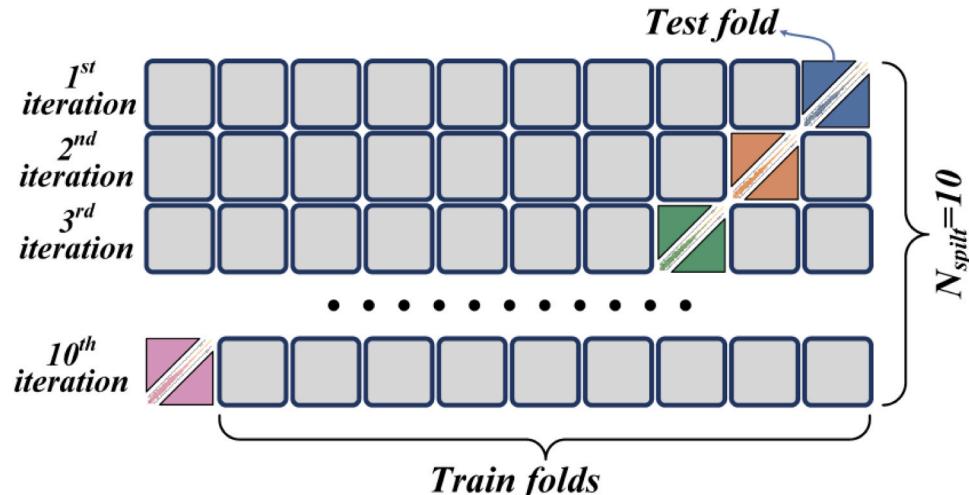


Fig. 13. K-fold cross-validation.

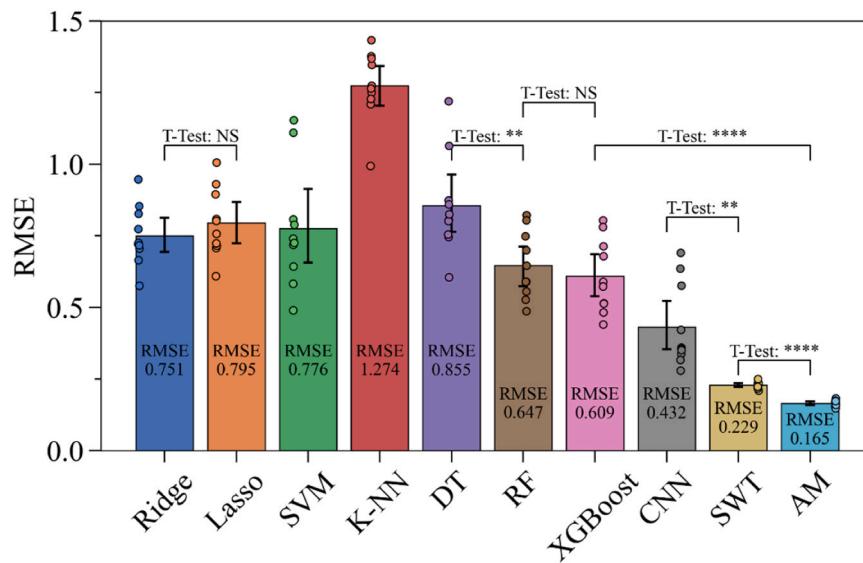


Fig. 14. Cross-validation result.

of the column; h is the storey height; and ρ is the beam–column stiffness ratio (defined as $I_b/4I_c$ in which I_b is the section moment of inertia of the beam). According to Eq. (19), the elastic stiffness of the second storey is 3944.27 N/mm, and the median dynamic stiffness of the structure is 3946.99 N/mm when the PGA is in the range 0.01–0.02 g; the values considerably approximate each other.

When the PGA gradually increases to 0.2 g, the median value of K_s of the first storey has a certain degree of degradation. When the PGA values are greater than 0.3 g and 0.4 g, the median value of K_s degrades by 13.89% and 41.59%, respectively. The stiffness of the second storey with viscous dampers also decreases with increasing PGA . When the PGA is greater than 0.1 g, the median value of K_s is 1592.44 N/mm; it decreases by 59.65% compared with that in the elastic stage, as shown in Fig. 15. The distribution of the maximum inter-storey deformation and force of the metal damper is shown in Fig. 16. When the PGA is less than 0.1 g,

the metal damper only provides stiffness without energy dissipation. As the PGA gradually increases, the metal damper begins to function, exhibiting a distinct bilinear distribution. After the PGA exceeds 0.5 g, the maximum force and displacement distributions of the damper are no longer linear. As a speed-dependent viscous damper, the overall distribution of the maximum inter-storey velocity and force exhibits an approximately linear distribution when the PGA is less than 0.1 g. After exceeding 0.1 g, an exponential distribution follows, as shown in Fig. 17.

The contribution of the damper to energy dissipation in the structure is minimal when the PGA of the seismic wave is less than 0.1 g, indicating low structural non-linearity. Thus, interpretable ML models, such as XGBoost, can yield relatively accurate results. With an increase in PGA , the structural non-linearity intensifies, and the manual extraction of features from the seismic wave fails to capture complete information. Owing to the robust non-linear regression capabilities of DL,

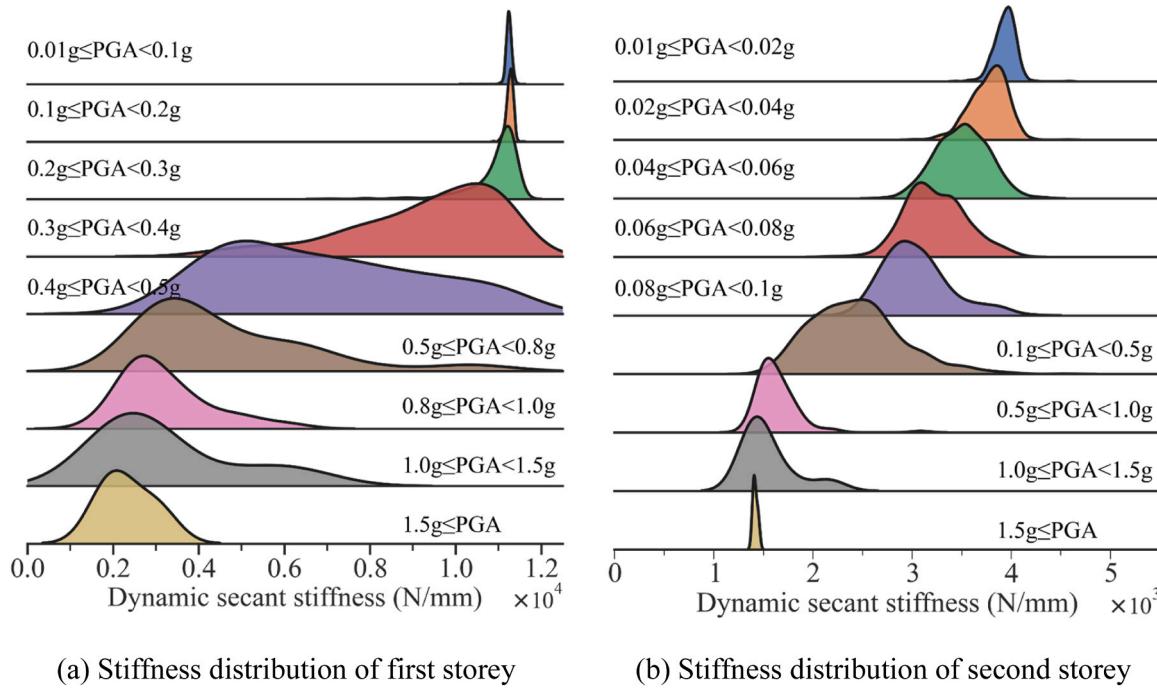


Fig. 15. Dynamic secant stiffness distribution.

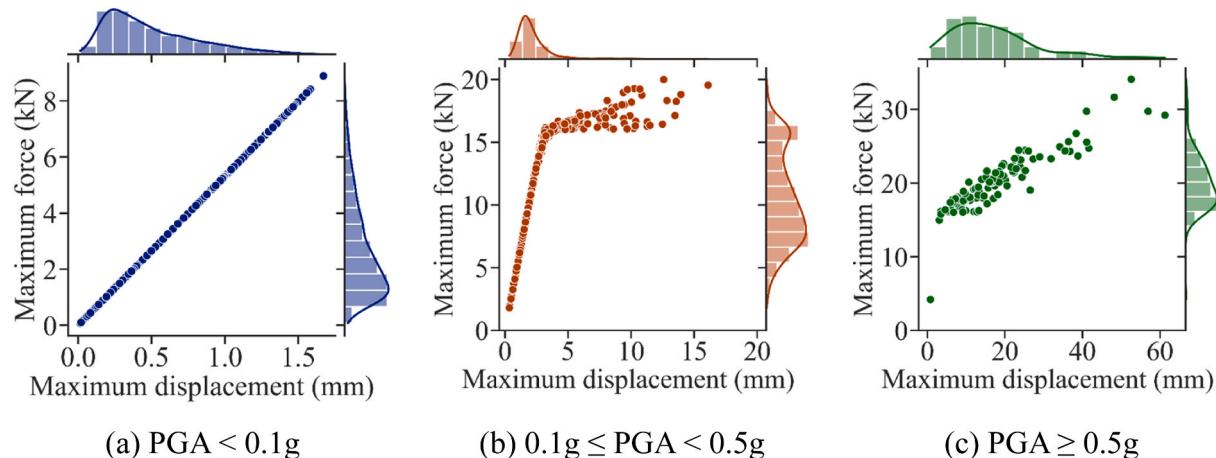


Fig. 16. Inter-storey deformation vs. force distributions of metal damper.

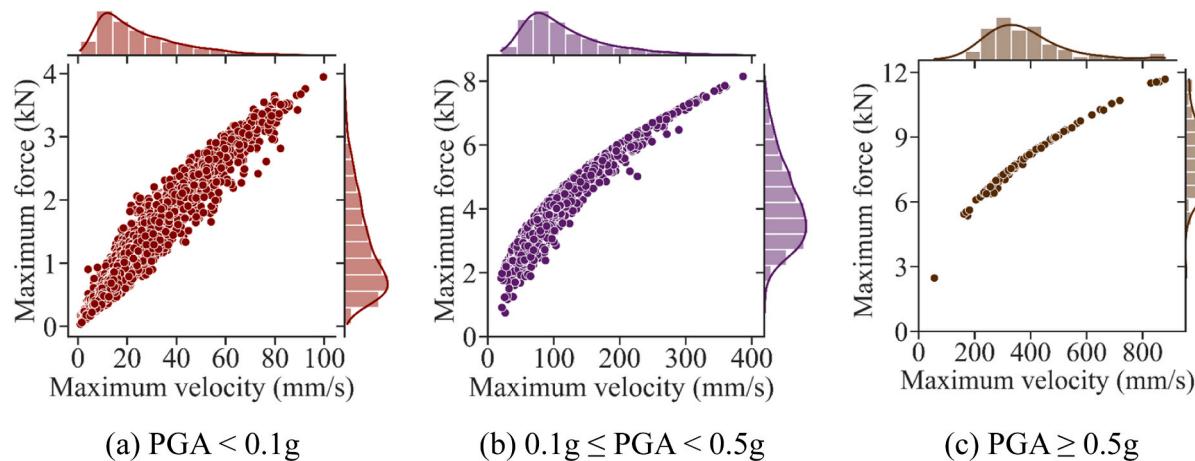


Fig. 17. Inter-storey velocity vs. force distributions of viscous damper.

unsupervised feature extraction strategies based on this approach tend to yield improved results. Consequently, the SWT excels in predicting the structural response under high PGA levels. Hence, the AM approach proposed in this study leverages interpretable ML techniques for lower levels of PGA , whereas the SWT is utilised for higher levels of PGA . This increases the accuracy of predictions and improves prediction efficiency.

7. Conclusion

This paper proposes a seismic response prediction method for damped structures based on an AM. By expanding the database based on shaking table tests, interpretable ML and DL models are used to train and test the data. In addition, by considering the applicability of the models and the non-linear characteristics of structures, a more accurate prediction model is developed. The main conclusions of this study are as follows.

1. Among traditional interpretable ML methods, RF and XGBoost are the recommended ensemble learning methods. Specifically, XGBoost has faster computation speed, more optimisable hyperparameters, better robustness, and more interpretable feature contributions during the prediction process. The prediction accuracy of XGBoost improves by 7.45% compared with that of RF. Moreover, its accuracy is considerably higher than those of other interpretable ML methods.
2. In this study, the CNN with a large receptive field and the SWT network based on the transformer model both achieve better

predictive performance than traditional ML methods, such as XGBoost. In the ten-fold cross-validation, the RMSEs of the two models (CNN and SWT networks) decrease by 29.06% and 50.90% compared with that of XGBoost, and the R^2 of the models are also effectively improved. The DL-based response prediction models do not require manual feature extraction and are more practical than other interpretable ML models, which require IMs.

3. Although the overall accuracy of the SWT model is higher than that of XGBoost, both models exhibit varying errors across different intervals of the seismic PGA indicators. By integrating an ensemble AM model that combines both the XGBoost and SWT models, the XGBoost model can be utilised in situations where the structural nonlinearity is weak, while the SWT model can be applied in cases where nonlinearity is strong. This aggregation leads to improved prediction accuracy compared with using a single model.

It should be noted that the proposed AM model algorithm proposed in this study can directly use the acceleration time history records of ground motions as feature input, which has the potential to predict the response of structures under strong nonlinearity. However, the prototype structure in this paper is single and cannot reflect all the characteristics of seismic structures, and the prediction algorithm used has a large number of parameters, leaving room for further simplification. Therefore, a more generic DL model based on AM algorithm to adapt to different structures remains to be optimised.

CRediT authorship contribution statement

Tianyang Zhang: Investigation, Software, Formal analysis curation, Data curation, Writing – original draft, Writing – Review & Editing. **Weizhi Xu:** Conceptualization, Methodology, Funding acquisition, Writing – original draft, Writing – review & editing. **Shuguang Wang:** Software, Validation, Formal analysis, Investigation, Writing – review & editing. **Dongsheng Du:** Investigation, Writing – review & editing. **Jun Tang:** Funding acquisition, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work is supported by China-Pakistan Belt and Road Joint Laboratory on Smart Disaster Prevention of Major Infrastructures (2022CPBRJL-09), the authors are also greatly thankful for the financial support from the National Natural Science Foundation of China (52208173) and Fellowship of China Postdoctoral Science Foundation (2022M721591).

References

- [1] Zhao B, Taucer F, Rossetto T. Field investigation on the performance of building structures during the 12 May 2008 Wenchuan earthquake in China. *Eng Struct* 2009;31:1707–23.
- [2] Yang ZG, Dai DQ, Zhang Y, Zhang XM, Liu J. Rupture process and aftershock focal mechanisms of the 2022 M6.8 Luding earthquake in Sichuan. *Earthq Sci* 2022;35: 474–84.
- [3] Jiang XY, Song XD, Li T, Wu KX. Moment magnitudes of two large Turkish earthquakes on February 6, 2023 from long-period coda. *Earthq Sci* 2023;36: 169–74.
- [4] Guo L, Wang JF, Wang HL. Seismic design and hybrid simulation test of existing concrete frames upgraded by metallic damper. *Eng Struct* 2023;291:116337.
- [5] Ouyang XY, Zhang YT, Ou XY, Shi YG, Liu SY, Fan JX. Seismic fragility analysis of buckling-restrained brace-strengthened reinforced concrete frames using a performance-based plastic design method. *Structures* 2022;43:338–50.
- [6] Ruiz SE, Santos-Santiago MA, Bojórquez E, Orellana MA, Valenzuela-Beltrán F, Bojórquez J, et al. BRB retrofit of mid-rise soft-first-story RC moment-frame buildings with masonry infill in upper stories. *J Build Eng* 2021;38:101783.
- [7] Lee KS, Lee BG, Jung JS. Nonlinear dynamic response of R/C buildings strengthened with novel stud-typed seismic control system using non-buckling slit damper. *Eng Struct* 2021;244:112749.
- [8] Zhao JC, Sun J, Qiu HX. Rapid seismic retrofit of damaged RC frames using gapped eccentric steel brace system equipped with dampers. *J Build Eng* 2022;53:104532.
- [9] Mehandousti AMA, Jalaeefar A. Effect of seismic sequences on behavior of mid-rise steel moment-resisting frames equipped with fluid viscous dampers. *Structures* 2023;54:657–68.
- [10] Zhang WX, Zhang C, Su LJ, Zheng YR, Du XL. Experimental study on the dynamic performance of a winding rope fluid viscous damper. *Eng Struct* 2023;281:115786.
- [11] Sharma KV, Parmar V, Gautam L, Choudhary S, Gohil J. Modelling efficiency of fluid viscous dampers positioning for increasing tall buildings' Resilience to earthquakes induced structural vibrations. *Soil Dyn Earthq Eng* 2023;173:108108.
- [12] Li LX, Liang Y, Chen GH, Yang DX. Simultaneous layout and size optimization of nonlinear viscous dampers for frame buildings under stochastic seismic excitation. *Eng Struct* 2022;273:115067.
- [13] Núñez E, Aguayo C, Mata R. Incremental dynamic analysis of steel storage racks subjected to Chilean earthquakes. *Thin-Walled Struct* 2023;182:110288.
- [14] Yan JL, Liang Y, Du X. Analysis of the importance coefficient of offshore bridges under earthquakes based on seismic fragility and incremental dynamic analysis. *Soil Dyn Earthq Eng* 2023;171:115067.
- [15] Pandikkadavath MS, Shajjal KM, Mangalathu S, Davis R. Seismic robustness assessment of steel moment resisting frames employing material uncertainty incorporated incremental dynamic analysis. *J Constr Steel Res* 2022;191:107200.
- [16] Miar M, Jankowski R. Incremental dynamic analysis and fragility assessment of buildings founded on different soil types experiencing structural pounding during earthquakes. *Eng Struct* 2022;252:107200.
- [17] Berto L, Di Sarno L, Fragiadakis M, Rocca I, Saetta A. Seismic assessment of free-standing artifacts: Full-scale tests on large shake table. *Earthq Eng Struct Dyn* 2023; 52:2708–30.
- [18] Fang C, Ping YW, Gao Y, Zheng Y, Chen Y. Machine learning-aided multi-objective optimization of structures with hybrid braces – framework and case study. *Eng Struct* 2022;269:114808.
- [19] Hu SL, Zhu SY, Shahria Alam M, Wang W. Machine learning-aided peak and residual displacement-based design method for enhancing seismic performance of steel moment-resisting frames by installing self-centering braces. *Eng Struct* 2022; 271:114935.
- [20] Chou JY, Liu CY, Chang CM. Story drift and damage level estimation of buildings using relative acceleration responses with multi-target deep learning models under seismic excitation. *Earthq Eng Struct Dyn* 2023;52:2554–71.
- [21] Bijelić N, Lignos DG, Alahi A. The automated collapse data constructor technique and the data-driven methodology for seismic collapse risk assessment. *Earthq Eng Struct Dyn* 2023;52:2452–79.
- [22] Ji K, Zhu CB, Yaghmaei-Sabegh S, Lu JQ, Ren YF, Wen RZ. Site classification using deep-learning-based image recognition techniques. *Earthq Eng Struct Dyn* 2022; 52:2323–38.
- [23] Kim T, Kwon OS, Song J. Deep learning based seismic response prediction of hysteretic systems having degradation and pinching. *Earthq Eng Struct Dyn* 2022; 52:2384–406.
- [24] Miao ZH, Ji XD, Wu MH, Gao X. Deep learning-based evaluation for mechanical property degradation of seismically damaged RC columns. *Earthq Eng Struc Dyn* 2022;52:2498–519.
- [25] Xu JG, Feng DC, Mangalathu S, Jeon JS. Data-driven rapid damage evaluation for life-cycle seismic assessment of regional reinforced concrete bridges. *Earthq Eng Struct Dyn* 2022;51:2730–51.
- [26] Xu JG, Hong W, Zhang J, Hou ST, Wu G. Seismic performance assessment of corroded RC columns based on data-driven machine-learning approach. *Eng Struct* 2022;255:113936.
- [27] Wang XW, Li ZQ, Shafieezadeh A. Seismic response prediction and variable importance analysis of extended pile-shaft-supported bridges against lateral spreading: exploring optimized machine learning models. *Eng Struct* 2021;236: 112142.
- [28] Hou C, Zhou XG. Strength prediction of circular CFST columns through advanced machine learning methods. *J Build Eng* 2022;51:104289.
- [29] Zhou XG, Hou C, Feng WQ. Optimized data-driven machine learning models for axial strength prediction of rectangular CFST columns. *Structures* 2023;47:760–80.
- [30] Kazemi F, Asgarkhani N, Jankowski R. Machine learning-based seismic response and performance assessment of reinforced concrete buildings. *Arch Civ Mech Eng* 2023;23:94.
- [31] Kazemi F, Asgarkhani N, Jankowski R. Machine learning-based seismic fragility and seismic vulnerability assessment of reinforced concrete structures. *Soil Dyn Earthq Eng* 2023;166:107761.
- [32] Kazemi F, Asgarkhani N, Jankowski R. Predicting seismic response of SMRFs founded on different soil types using machine learning techniques. *Eng Struct* 2023;274:114953.
- [33] Kazemi F, Jankowski R. Machine learning-based prediction of seismic limit-state capacity of steel moment-resisting frames considering soil-structure interaction. *Comput Struct* 2023;274:106886.
- [34] Asgarkhani N, Kazemi F, Jankowski R. Machine learning-based prediction of residual drift and seismic risk assessment of steel moment-resisting frames considering soil-structure interaction. *Comput Struct* 2023;289:107181.
- [35] Krizhevsky A, Sutskever I, Hinton G.E. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems. 2012, p. 1097–1105.
- [36] Li CX, Li H, Chen X. Fast seismic response estimation of tall pier bridges based on deep learning techniques. *Eng Struct* 2022;266:114566.
- [37] Soleimani-Babakamali MH, Zaker Esteghamati M. Estimating seismic demand models of a building inventory from nonlinear static analysis using deep learning methods. *Eng Struct* 2022;266:114576.
- [38] Zhang D, Chen Y, Zhang C, Xue GX, Zhang JF, Zhang M, et al. Prediction of seismic acceleration response of precast segmental self-centering concrete filled steel tube single-span bridges based on machine learning method. *Eng Struct* 2023;279: 115574.
- [39] Kundu A, Ghosh S, Chakraborty S. A long short-term memory based deep learning algorithm for seismic response uncertainty quantification. *Probabilist Eng Mech* 2022;67:103189.
- [40] Huang PF, Chen ZY. Deep learning for nonlinear seismic responses prediction of subway station. *Eng Struct* 2021;244:112735.
- [41] Vaswani A, Shazeer N, Parmar N., Uszkoreit J., Jones L., Gomez A.N. et al. Attention is all you need. 2017, arXiv preprint arXiv:1706.03762.
- [42] Brown T.B., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P. Language models are few-shot learners. 2020, arXiv preprint arXiv:2005.14165.
- [43] Chen YJ, Sun ZY, Zhang RY, Yao LZ, Wu G. Attention mechanism based neural networks for structural post-earthquake damage state prediction and rapid fragility analysis. *Comput Struct* 2023;281:107038.
- [44] Xu YJ, Lu XZ, Fei YF, Huang YL. Hysteretic behavior simulation based on pyramid neural network: Principle, network architecture, case study and explanation. *Adv Struct Eng* 2023. <https://doi.org/10.1177/13694332231184322>.
- [45] Xu WZ, Wang Y, Guo HR, Du DS, Wang SG. Theoretical and experimental investigation on the seismic performance of a novel variable-damping viscous fluid damper. *J Build Eng* 2022;53:104537.
- [46] Kramer SL, Mitchell RA. Ground motion intensity measures for liquefaction hazard evaluation. *Earthq Spectra* 2006;22(2):413–38.

- [47] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. 2nd ed. New York, NY: Springer-Verlag New York,; 2009. <https://doi.org/10.1007/978-0387-84858-7>.
- [48] Hou C, Zhou XG, Shen LM. Intelligent prediction methods for N-M interaction of CFST under eccentric compression. *Arch Civ Mech Eng* 2023;23:197.
- [49] Ho T.K. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. 1995, p. 278–282.
- [50] Chen T.Q., Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016, p. 785–94.
- [51] Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge: MIT Press; 2016.
- [52] Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., et al. An image is worth 16×16 words: Transformers for image recognition at scale. 2020, arXiv preprint arXiv:2010.11929.
- [53] Kingma D.P., Ba J.L. Adam: A Method for Stochastic Optimization. In: Proceedings of the 3rd International Conference on Learning Representations. 2015.