# Explainable artificial intelligence study on bolt loosening detection using Lamb waves

Hu, Muping; Salmani Pour Avval, Sasan; He, Jian; Yue, Nan; Groves, Roger M.

**DOI**
10.1016/j.ymssp.2024.112285

**Publication date**
2025

**Document Version**
Final published version

**Published in**
Mechanical Systems and Signal Processing

Mechanical Systems and Signal Processing 225 (2025) 112285

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Full Length Article

# Explainable artificial intelligence study on bolt loosening detection using Lamb waves

Muping Hu [a,b,*], Sasan Salmani Pour Avval [b], Jian He [a], Nan Yue [b], Roger M. Groves [b]

[a] College of Aerospace and Civil Engineering, Harbin Engineering University, Harbin 150001, PR China
[b] Aerospace Structures & Materials Department, Delft University of Technology, Delft 2629 HS, the Netherlands

ARTICLE INFO

*Keywords:*
Lamb waves
Deep learning
Explainable AI (XAI)
One-dimensional convolutional neural network
Bolt loosening detection

ABSTRACT

With the rapid development of artificial intelligence (AI) technologies, deep learning-based structural health monitoring (DeepSHM) methods have gained significant attention. However, their *black box* nature often limits interpretability and trust. The field of Explainable AI (XAI) aims to address this by enhancing model transparency and reliability through human-comprehensible explanations. This study investigates the use of XAI algorithms in interpreting a 1D convolutional neural network (1D CNN) developed for Lamb wave monitoring of bolt-loosening detection in multi-bolted double-layer aluminum plates under varying temperatures. Four existing XAI algorithms were employed, including Sensitivity Analysis, Deep Taylor, Gradient-weighted Class Activation Mapping (Grad CAM) and Guided Grad CAM. In addition, this paper introduces two new XAI methods, Smooth Simple Taylor and Deep Grad CAM as an enhancement of the Simple Taylor and Grad CAM methods, respectively. These six XAI algorithms were used to establish the relation between the 1D CNN model parameters and the input vector. The results were evaluated for their effectiveness in comparison to the physical insights of the input vector using two proposed methods, namely the Correlation Coefficient with Residual Signal and the Residual Signal Weighted Importance Score Ratio. The results of the evaluation methods, in conjunction with Infidelity, Sense sum, and Sanity check, were utilized to rank the performance of the six XAI algorithms. The rankings were consistent in both simulation and experiment data sets, and the newly proposed XAI algorithm, Smooth Simple Taylor, appeared to be the best in both data sets. Overall, this research establishes a novel approach to using XAI algorithms to enhance the explainability of AI in practical engineering applications.

## 1. Introduction

As a common mechanical connection method, bolted joints are widely used in fields such as aerospace, civil engineering, shipbuilding, and construction due to their advantages of easy installation and disassembly, low cost, and reusability [1]. As ubiquitous components that bear heavy loads in structures, bolted joint connection status has a significant impact on the safety and reliability of the structures. However, threaded fasteners are prone to loosening when exposed to harsh working environments such as cyclic loading, mechanical attack, chemical corrosion, and improper operation [2,3]. This can lead to a reduction in preload and cause

* Corresponding author.
*E-mail addresses:* humuping@hrbeu.edu.cn (M. Hu), s.salmanipouravval@tudelft.nl (S. Salmani Pour Avval), hejian@hrbeu.edu.cn (J. He), n.yue@tudelft.nl (N. Yue), r.m.groves@tudelft.nl (R.M. Groves).

https://doi.org/10.1016/j.ymssp.2024.112285
Received 21 June 2023; Received in revised form 21 November 2024; Accepted 27 December 2024
Available online 3 January 2025

fatigue fracture of the bolts, which may result in serious consequences such as unstable connections, component detachment, and structural collapse [4–7]. Therefore, continuous monitoring of bolt loosening plays a crucial role in ensuring the safety and applicability of bolted structures and preventing catastrophic failures.

Up to now, various promising structural health monitoring (SHM) methods for detecting bolt loosening have been proposed, including vibration-based methods [8,9], percussion-based methods [10,11], electro-mechanical impedance (EMI) methods [12,13], optical inspection methods [14,15], and non-contact laser excitation methods [16,17]. Although these techniques have their advantages, they also have certain limitations. Percussion-based methods may cause certain impact forces on the structure itself and have lower accuracy [18]. Vibration-based methods are only suitable for detecting the overall vibration characteristics of the structure and cannot locate the position of loosened bolts. EMI-based methods are severely affected by environmental conditions and are only suitable for rigid structures such as bridges and buildings [19]. Optical inspection methods and laser excitation equipment are expensive, have poor adaptability, and are easily affected by environmental factors such as temperature and humidity [20].

Among various attempts, Guided wave-based SHM techniques have gained significant attention due to their excellent sensitivity to different damages, wide detection range, and enormous potential for continuous or periodic monitoring of in-service structures [21]. Yang et al. [22] studied the contact nonlinearity-induced second harmonic generated by bolt looseness by experiment and three-dimensional explicit finite element simulation, proposing an indicator for bolt connection integrity assessment based on guided wave nonlinear features. Tola et al. [23] proposed a bolt looseness detection method based on the ultrasonic field energy in the reflected Lamb wave from the target bolt. Du et al. [24] developed a semi-analytical method combining finite element models and wave superposition to predict the propagation of guided waves in the bolt interface, and discussed how to use the semi-analytical method combining power transmission coefficient for bolt torque monitoring. However, current research has mostly focused on the issue of loosening in single bolt connections of simple structures, which is not commonly found in practical industry application. In actual engineering, bolted structures are typically complex [25] and the transmitted guided waves experience complicated propagation processes and mode conversions, making it challenging to extract stable acoustic features from collected signals, thus hindering precise damage detection.

The combination of SHM and deep learning technology may help overcome these challenges. In recent years, with the rapid development of AI technology, its application in bolt loosening detection has also made significant progress [26–30]. Nguyen et al. [31] developed a novel method for bolt-looseness assessment based on the integration of the impedance-based technique and 1D CNN, and experimentally validated the proposed method by detecting bolt loosening in a girder connection. Wang et al. [32] proposed a multi-bolt loosening detection method based on a newly developed one-dimensional memory augmented convolutional long short-term memory (1D-MACLSTM) networks. Huynh [33] presented an innovative autonomous visual bolt-loosening detection method that employs Faster regional CNN (Faster RCNN) and applied it in the detection of a realistic joint of the Dragon Bridge in Danang, Vietnam. The results demonstrated the great potential of AI-based method for in-situ autonomous monitoring.

AI models typically involve complex algorithms. A common concept in such models is their powerful non-linear analysis ability, which enable them to easily and highly accurately classify damage. However, AI models on their own are unable to provide any fundamental insights on how the specific decision is made [34]. If the underlying principles behind the prediction can be comprehended, the transparency and trustworthiness of the model can be significantly enhanced, also making it possible to extract new knowledge from it. To understand the analyzing logic of the AI model, an additional technique is needed to delve into the *black box* and reveal how the model operates effectively.

Such a technique is known as explainable AI (XAI). According to [35], "XAI is a field of artificial intelligence (AI) that promotes a set of tools, techniques, algorithms to generate high-quality interpretable, intuitive, human-understandable explanations of AI decision." Using appropriate visualizations to explain the metadata of neural networks can help achieve transparency between engineers and AI models [36]. Currently, some studies have applied XAI techniques in inspection and monitoring applications [37–41]. Meister et al. [42] presented an approach for analyzing the classification procedure of fiber layup defects based on Smoothed Integrated Gradients, Guided Gradient Class Activation Mapping and Deep SHAP. Ewald et al. [43] adopted a neuroscientific perspective to mimic the perception of Deep learning based SHM (DeepSHM) method and proposed two interpretable analysis theories to mathematically explain why CNN are effective in detecting damage. Brito et al. [44] proposed an unsupervised fault detection method based on Shapley Additive Explanations (SHAP), and performed the fault diagnosis for rotating machinery through the feature importance ranking obtained by the model's explanation.

Driven by the concept of XAI, this study endeavors to understand the decision-making logic of the 1D CNN models for bolt loosening detection from a physical perspective. The detection of bolt loosening was conducted on a double-layered aluminum plate with sixteen bolted connections. To simulate signal variation that could occur in real-world conditions, phase and amplitude variations induced by temperature changes were introduced into the signals. Under such circumstances, traditional methods that rely on residual signals for damage feature extraction become ineffective. In the light of these challenges, this research aims to address the following questions:

1. How to achieve the precise detection of bolt loosening in multi-bolted structures and unstable service environments using 1D CNN and Lamb wave monitoring techniques?

2. How can a better understanding of the fundamental principles behind the precise predictions of 1D CNN be obtained by utilizing XAI techniques?

3. From the standpoint of SHM, how can the performance of XAI algorithms be evaluated? Specifically, how to select the most suitable XAI algorithm to obtain the most easily understandable model explanation for the signal-feeding 1D CNN model.

To address these issues, the propagation characteristics of Lamb waves in the multi-bolted plate were investigated using numerical simulation and experiment. The 1D CNN models were trained using Lamb wave signals to detect bolt loosening damage. Subsequently, six XAI algorithms, including Saliency map, Smooth Simple Taylor, Deep Taylor, Grad CAM, Deep Grad CAM, and Guided Grad CAM,

were employed to explain the well-trained 1D CNN model, analyzing its classification rules for different damage scenarios. Furthermore, the performance of these XAI algorithms was evaluated using five different assessment methods: Correlation Coefficient with Residual Signal (CCRS), Residual Signal Weighted Importance Score Ratio (RWIR), Infidelity (INFD), Sense sum (SENS) and Sanity check (SACH). Finally, the ranking scores of XAI algorithms based on these evaluation methods were compared between numerical simulation and experiment.

The article is organized as follows: Section 2 introduces the theoretical background of 1D CNN, XAI and evaluation methods for XAI. Section 3 presents numerical simulation of Lamb waves in multi-bolt connection plates and explainable analysis of the 1D CNN model with six XAI algorithms. Section 4 details the experimental validation, while Section 5 compares the performance of the six XAI algorithms in simulation and experiment. Section 6 highlights the innovations and contributions of this study to the SHM field, and the conclusions are summarized in Section 7.

## 2. Theories

### 2.1. 1D CNN

One-dimensional CNN (1D CNN), which is suitable for the 1D input array analysis, was used to detect the bolt loosening in the aluminum plates in this paper, and the time-series signals collected from the sensors were converted to 1D array to suit the model input shape. As a typical classification network, it consists of two stages: feature extraction and classification [45]. In the feature extraction stage, the 1D convolutional kernel slides along the input vector to convolve and extract feature maps. Then, the size of the feature maps is reduced by max-pooling layers. Therefore, the output of the convolution operation of the $u$-th channel in the $(m+1)$-th layer can be expressed as:

$$y_u^{m+1} = b_u^{m+1} + \Sigma_{k \in N_m} x_k^m * w_u^{m+1} \tag{1}$$

where $y_u^{m+1}$ and $x_k^m$ are the feature maps of the $u$-th and $k$-th channel in the $(m+1)$-th and $m$-th layers, respectively. $N_m$ is the number of convolutional kernels in the $m$-th layer. $w_u^{m+1}$ and $b_u^{m+1}$ represent the weight and bias of the $(m+1)$-th layers, respectively. * denotes a 1D convolutional operation. Then, a max-pooling layer is placed after every convolutional layer.

After the feature extraction, the feature map $x^d \in \mathbb{R}^{1 \times N \times D}$ obtained from the last convolutional layer is flattened into a feature vector $x^n \in \mathbb{R}^{1 \times ND}$ and fed into the first fully connected layer, where $N$ is the number of output channels of the last convolution layer and $D$ is the length of the feature map. Then, the classification task is performed:

$$y_q^{n+1} = \mathrm{ReLU}\left(w_q^{n+1} x_p^n + b_q^{n+1}\right) = \max\left[0, \left(w_q^{n+1} x_p^n + b_q^{n+1}\right)\right] \tag{2}$$

where $y_q^{n+1}$ is the $q$-th value in the $(n+1)$-th layer, $ReLU$ is the linear rectification activation function, $w_q^{n+1}$ and $b_q^{n+1}$ represent the weights and bias of $(n+1)$-th layer.

For the training of a 1D CNN, cross-entropy is used to calculate the error between the predicted and true label:

$$\mathscr{L}(\Theta) = -[\widehat{Y}_C \log(\mathrm{F}(V_C)) + (1 - \widehat{Y}_C)\log(1 - \mathrm{F}(V_C))] \tag{3}$$

where $\mathscr{L}$ is the loss function, F represents 1D CNN model, $\Theta$ contains all the parameters (weights and bias) of F, $V_C$ is an input vector from class $C$, $\widehat{Y}_C$ is the label of $V_C$. The 1D CNN is updated using stochastic gradient descent (SGD) algorithm with Adam optimizer [46].

### 2.2. XAI algorithms

The 1D CNN model is interpreted in the form of a saliency map, where the color of each input data point represents its importance score, which could indicate the dependence degree of the 1D CNN on that input data point during the decision-making process. There are various XAI algorithms for extracting the importance score for a saliency map. In this study, six of them were employed, which belonged to three categories: Gradient-based methods, Layer-wise relevance propagation methods and Class activation maps. Gradient-based methods (Sensitivity analysis, Guided Grad CAM) use the gradients of the AI model with respect to the input data to understand their importance. Layer-wise relevance propagation methods (Smooth Simple Taylor, Deep Taylor, Deep Grad CAM) propagate the importance score from the output layer back to the input layer by using the structure of the AI model. Class activation maps (Grad CAM, Deep Grad CAM and Guided Grad CAM) utilize the activations of the convolutional layers to calculate the importance score.

#### 2.2.1. Sensitivity analysis

Sensitivity Analysis uses the square of the gradients of each data point in the input vector to characterize its importance score. A high importance score means a change of that point has a greater impact on the 1D CNN model's decision during classification, making it more crucial in the process [47]. The importance score for the $s$-th data point of the input vector $v$ can be expressed as:

$$I_{SA}(v_s) = \left(\frac{\partial \mathrm{F}}{\partial v_s}\right)^2 \tag{4}$$

*2.2.2. Smooth Simple Taylor*

In this study, a smoothed version of Simple Taylor Decomposition is proposed. In the Simple Taylor Decomposition [48], the neural network function is Taylor-expanded at the root point, and the first-order term is taken as the importance score for each data point:

$$I_{ST}(v_s) = \frac{\partial \mathrm{F}}{\partial v_s}(\widetilde{v}) \cdot \left(v_s - \widetilde{v}_s\right) \tag{5}$$

where $\widetilde{v}$ refers to the root point vector that represents the neutral points on the decision boundary of the neural network and has no impact on the decision-making process. Adding noise to the input signal and taking its average to smooth the Simple Taylor result allows the Smooth Simple Taylor to be determined:

$$I_{SST}(v_s) = \frac{1}{N_{noise}} \sum_{1}^{N_{noise}} I_{ST}(v_s') \tag{6}$$

where $N_{noise}$ is the number of samples, and $v_s'$ represents $s$-th data point int the signal with the noise.

*2.2.3. Deep Taylor*

Deep Taylor Decomposition is an improved version of the Simple Taylor decomposition [49]. It takes into account the structure of the neural network and applies the Taylor expansion layer-by-layer from the output layer to the input layer. The transmission of the importance score between two layers can be expressed as:

$$I_{DT}^{(n)}(x_p) = \sum_q \left.\frac{\partial I_{DT}^{(n+1)}(x_p)}{\partial x_p}\right|_{\left\{\widetilde{x}_p^{(q)}\right\}} \cdot \left(x_p - \widetilde{x}_p^{(q)}\right) \tag{7}$$

where $\widetilde{x}_p^{(q)}$ represents the root point vector in the $n$-th layer.

*2.2.4. Grad CAM*

The Grad CAM method assigns importance values to each neuron using gradient information flowing into the last convolutional layer to obtain specific interest decisions [50]. The importance score can be represented as:

$$I_{GC}(v_s) = ReLU\left(\sum_{k=1}^{N_m} \alpha_k^{m,C} \mathrm{A}_k^m\right) \tag{8}$$

where $N_m$ is the number of convolutional kernels, $\mathrm{A}_k^m$ represents the activation in the $k$-th channel of the $m$-th layer, $C$ represents the class, and $\alpha_k^{m,C}$ is the weight of $\mathrm{A}_k^m$ which can be calculated as:

$$\alpha_k^{m,C} = \mathbb{E}\ominus\left[\frac{\partial \mathrm{F}(V_C)}{\partial A_k^m}\right] \tag{9}$$

where $\mathrm{F}(V_C)$ represents the score predicted by the neural network for class $C$, and $\mathbb{E}$ represents taking the average along the length direction of the activation $\mathrm{A}_k^m$.

*2.2.5. Guided Grad CAM*

Fusing Guided backpropagation (GBP) [51] and Grad CAM via element-wise multiplication yields Guided Grad CAM which can generate saliency maps with more fine-grained details at the pixel level:

$$I_{GGC}(v_s) = I_{GC}(v_s) \cdot GBP(v_s) \tag{10}$$

*2.2.6. Deep Grad CAM*

Deep Grad CAM was proposed in our recent conference paper [52]. It incorporates the hierarchical structure of 1D CNN convolutional layers to propagate the importance vector using its backpropagation mechanism instead of linear mapping. Specifically, the $\alpha$-$\beta$ rule [48] is applied to propagate the importance score layer by layer:

$$I_{DGC}^m(x_{k,i}^m) = a_{k,i}^m \sum_j \frac{w_{ij}^+}{\sum_i a_{k,i}^m w_{ij}^+ + b_j^+} I_{DGC}^{m+1}(x_{u,j}^{m+1}) \tag{11}$$

where $a_{k,i}^m$ represents the activation value at the $i$-th point in the $k$-th channel of the $m$-th layer. $w_{ij}^+$ and $b_j^+$ are the positive parts of the network weights and biases, respectively.

### 2.3. Evaluation of the XAI

Different XAI algorithms can have varying focuses when explaining AI models, leading to divergent explanations. Consequently, researchers have increasingly been paying attention to the evaluation of XAI algorithms, as summarized in [53]. In this study, by leveraging prior knowledge and analytical logic derived from SHM, two evaluation methods, Correlation coefficient with residual signal (CCRS) and Residual signal weighted importance score ratio (RWIR) are proposed with the objective of better comprehending the explanations of XAI from the physical point of view. It is worth noting that these two evaluation methods are not generically applicable to all AI application domains, but are specifically designed for the domain of damage detection based on Lamb waves and 1D CNN. Additionally, evaluation methods including Infidelity, Sense sum, and Sanity check are also considered in this study.

#### 2.3.1. Correlation coefficient with residual signal (CCRS)

Subtracting the baseline signal acquired from a structure in its healthy state from the received signal collected from the damaged structure yields the residual signal. Theoretically, the residual signal can filter out the effects of boundary reflections, non-target point reflections, and highlight the damage reflection waves. Therefore, if the importance vector of a signal generated by an XAI algorithm exhibits a high correlation with the residual signal, it indicates a strong agreement between the algorithm's identified important regions and the locations of damage waves. The CCRS can be calculated as:

$$CCRS = \frac{\sum (\Phi(\mathrm{F}, v_s) - \overline{\Phi}(\mathrm{F}, V))(r_s - \overline{R})}{\sqrt{\sum (\Phi(\mathrm{F}, v_s) - \overline{\Phi}(\mathrm{F}, V))^2 \sum (r_s - \overline{R})^2}} \tag{12}$$

where $\Phi$ is the XAI function, $V$ denotes the input signal, $R$ denotes the corresponding residual signal, and $R = V - V^{Base}$, where $V^{Base}$ represents the baseline signal. $v_s$ and $r_s$ represent the $s$-th point of the vectors $V$ and $R$, respectively. $\Phi(\mathrm{F}, v_s)$ denotes the importance score for the $s$-th point of the signal, $\overline{\Phi}(\mathrm{F}, V)$ represents the mean importance score of the signal, and $\overline{R}$ represents the mean value of the residual signal. The higher the CCRS, the higher the similarity between the importance vector and the residual signal, which indicates that the XAI algorithm believes the neural network relies more on information near the damage wave package when making decisions. This alignment is more in line with prior knowledge in SHM that the information carried by the damage package is considered highly useful and informative. Therefore, under the explanation of this XAI algorithm, the decisions of the neural network are more likely to be trusted by SHM professionals.

#### 2.3.2. Residual signal weighted importance score ratio (RWIR)

The importance vector obtained by XAI is weighted using the residual signal, giving greater weight to the regions with waveform signals. The RWIR can be obtained by dividing the sum of the weighted importance vector by the sum of the original importance vector:

$$RWIR = \frac{\sum r_s \cdot \Phi(\mathrm{F}, v_s)}{\sum \Phi(\mathrm{F}, v_s)} \tag{13}$$

The higher the RWIR, the greater the proportion of scores in the importance vector at the location of the damage wave package. Consequently, a higher RWIR implies a greater proportion of coloring at the location of the damage wave package in the saliency map. By calculating the proportion of coloring at the damage wave package location in the overall saliency map, the absolute value differences in the overall importance vector calculations across different XAI algorithms can be mitigated. Therefore, a larger RWIR indicates a higher proportion of coloring at the damage wave package location in the saliency map, making the results more easily understandable to human observers compared to color in the regions without fluctuations in the signal that do not carry informative content.

#### 2.3.3. Infidelity (INFD)

Infidelity is used to measures the consistency between model predictions and model explanations, and can be used to evaluate the reliability of XAI algorithms [54].

$$INFD = \mathbb{E}\left[\left((V - V')^T \Phi(\mathrm{F}, V) - (\mathrm{F}(V) - \mathrm{F}(V'))\right)^2\right] \tag{14}$$

where $V'$ represents the signal with the noise. The smaller the INFD, the closer the results of XAI align with the AI model's predictions.

#### 2.3.4. Sense sum (SENS)

Sense sum can be used to evaluate the noise robustness of the XAI algorithm, by measuring the impact of perturbations on the XAI [54]:

$$SENS_{sum} = \sum \left|\Phi(\mathrm{F}, v_s) - \Phi(\mathrm{F}, v_s')\right| \tag{15}$$

The smaller the Sense Sum value, the less the explanation result changes under the influence of noise, indicating that XAI algorithm has a stronger ability to resist noise.

#### 2.3.5. Sanity check (SACH)

Sanity check can be used to assess the sensitivity of XAI to the input–output relationship [55]. By randomly shuffling the labels of the input signals, if the explanation result with randomized labels remains the same as the original one, it indicates the algorithm is not sensitive to the relationship between input signals and labels. The SACH can be calculated as:

$$SACH = \frac{\sum (\Phi(\mathrm{F}, v_s) - \overline{\Phi}(\mathrm{F}, V))(\Phi(\mathrm{F}', v_s) - \overline{\Phi}(\mathrm{F}', V))}{\sqrt{\sum (\Phi(\mathrm{F}, v_s) - \overline{\Phi}(\mathrm{F}, V))^2 \sum (\Phi(\mathrm{F}', v_s) - \overline{\Phi}(\mathrm{F}', V))^2}} \tag{16}$$

where F′ is the AI model function trained on dataset with randomized labels. A smaller value of SACH indicates that the algorithm is more sensitive to the input–output relationship.

## 3. Numerical simulation

### 3.1. Simulation model

The propagation of Lamb waves in a multi-bolted double-layer aluminum plate was simulated using ABAQUS software. The Young's modulus of the aluminum plate (Al-7075) is 71 GPa, density is 2800 kg / $m^3$, and Poisson's ratio is 0.33 [56]. Both layers of the aluminum plate are 2 mm thick and were in contact by using a Tie. The dimensions of the plates are 400 × 400 mm. A schematic of the aluminum plate and transducer array is shown in Fig. 1. (a), with four 8 mm diameter piezoelectric sensors (PZT) placed on the aluminum plate, each of them was used in turn as the excitation source, while the other three sensors acted as receivers. Sixteen steel bolts with a diameter of 6 mm bolted the two plates together, and the bolt hole has a diameter of 6.6 mm [57]. The Young's modulus of the bolts is 206 GPa, density is 7800 kg / $m^3$, and Poisson's ratio is 0.3 [58]. The propagation of Lamb waves in the plate was studied under both tightly connected and loosened conditions of the bolt. In the tightly connected case (Connected plate), the bolt and the aluminum plate were in contact by using Tie. In the loosened case (Damaged plate), the interactions between one of the bolts and the aluminum plate were removed.

The numerical simulation model of the bolted aluminum plates is shown in Fig. 1. (b). A concentrated force was applied on the peripheral nodes of the PZT to excite the Lamb wave, and the excitation signal was a tone burst of 3 cycles with a Hanning − window centered at 200 kHz, which was loaded along the thickness direction of the plate. As the shortest wavelength of the Lamb wave with 200 kHz frequency is 15.55 mm, the mesh size was set to 1 mm. The element type is the eight-node brick element with reduced integration (C3D8R), with a total of 666,048 elements. The time increment step was set to 0.01 μs, the sampling frequency was 10 MHz, and the recorded time length was 0.15 ms. The explicit dynamics solver was used.

### 3.2. Wave propagation

In our recent conference paper [52], a detailed analysis was conducted on the influence of bolt connection status on the propagation of Lamb waves. It was found that the bolt reflection is more complex than the hole reflection. Based on the foundation of the conference paper, the propagation of Lamb waves in the single-bolt Connected and Damaged plate was recalculated for this study, the results of which are shown in the Fig. 2 and Fig. 3. It can be observed that obvious reflection waves are produced at both the bolt and hole locations. However, in the Connected plate, some of the Lamb waves become trapped in the bolt and continue to reflect inside it before spreading outwards, causing the bolt to become a weak secondary excitation source where a sustained excitation phenomenon

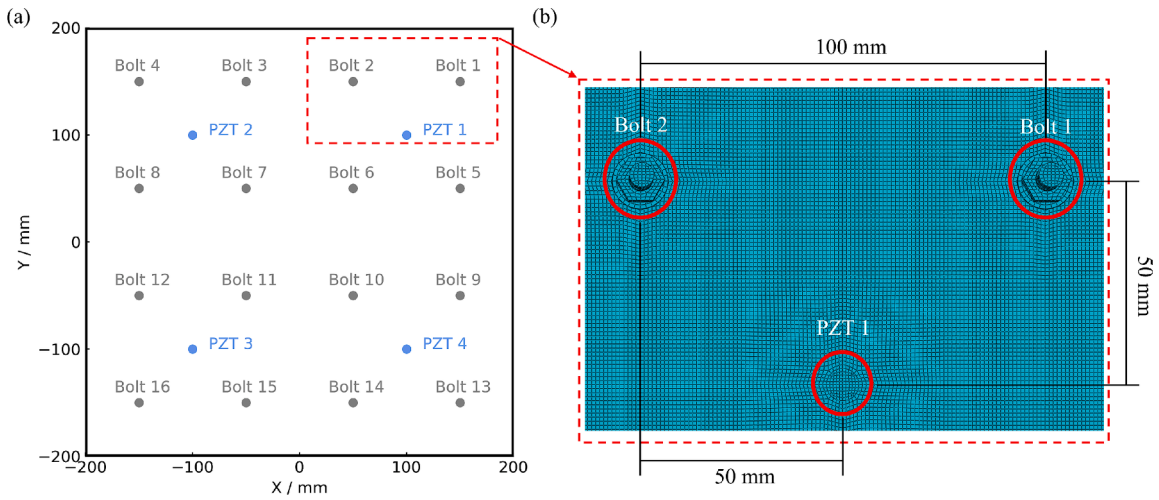


**Fig. 1.** (a) Schematic of the aluminum plate and transducer array; (b) Simulation model of the bolted plate.

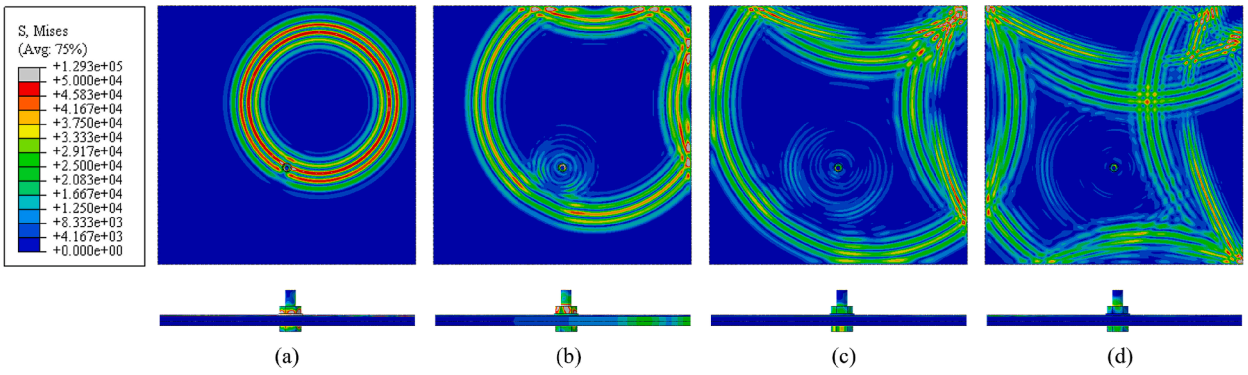


**Fig. 2.** Propagation of Lamb waves in the Connected plate with a single bolt: (a) 3.0 $\times 10^{-5}$ s; (b) 4.2 $\times 10^{-5}$ s; (c) 5.4 $\times 10^{-5}$ s; (d) 6.6 $\times 10^{-5}$ s.

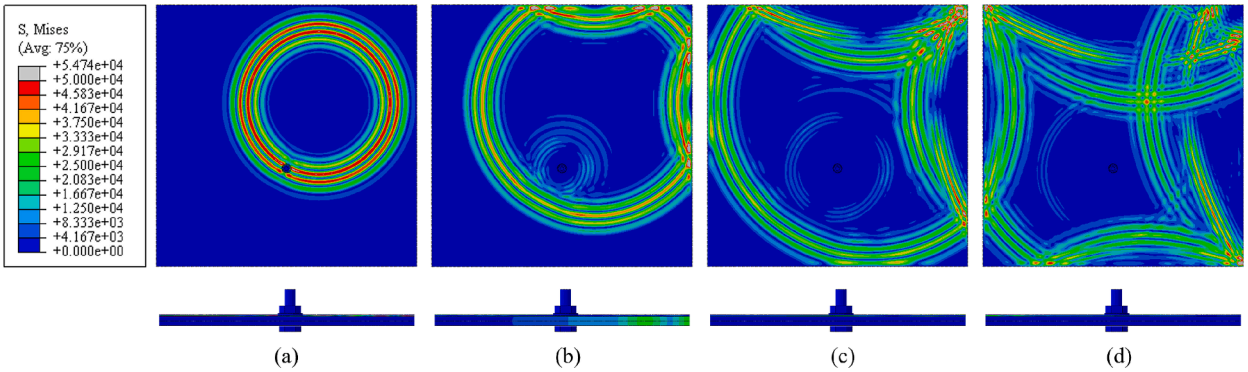


**Fig. 3.** Propagation of Lamb waves in the Damaged plate with a single bolt: (a) 3.0 $\times 10^{-5}$ *s*; (b) 4.2 $\times 10^{-5}$ *s*; (c) 5.4 $\times 10^{-5}$ *s*; (d) 6.6 $\times 10^{-5}$ *s*.

can be observed. In the Damaged plate, this sustained secondary excitation phenomenon no longer occurs. Therefore, there are more wave packets and more complex modes in the reflection waves of the Connected plate.

The propagation of Lamb waves in a multi-bolted Connected and Damaged plate with loosened Bolt 1 is shown in Fig. 4 and Fig. 5, respectively. Consistent with the observations in Fig. 2 and Fig. 3, each bolt serves as a reflection point and a secondary excitation source during the propagation of Lamb waves. When Bolt 1 was loosened, Lamb waves cannot travel into the bolt, so the sustained secondary excitation no longer occurs at that location. However, unlike the single-bolted case, the propagation path and modes of Lamb waves in multi-bolted plates are highly complex due to the generation of reflection waves and mode conversion by each bolt. Furthermore, Lamb waves generated by secondary excitation from the bolts mix with the initial waves. Therefore, the reflected signals generated by bolt loosening can easily be overwhelmed by these complex reflected signals, making it difficult to extract the damage features.

Although using residual signals can eliminate the influence of reflection waves, this may not work if the extraction states of the baseline and damaged signals do not match perfectly. While ensuring complete consistency between the reference and damaged plates

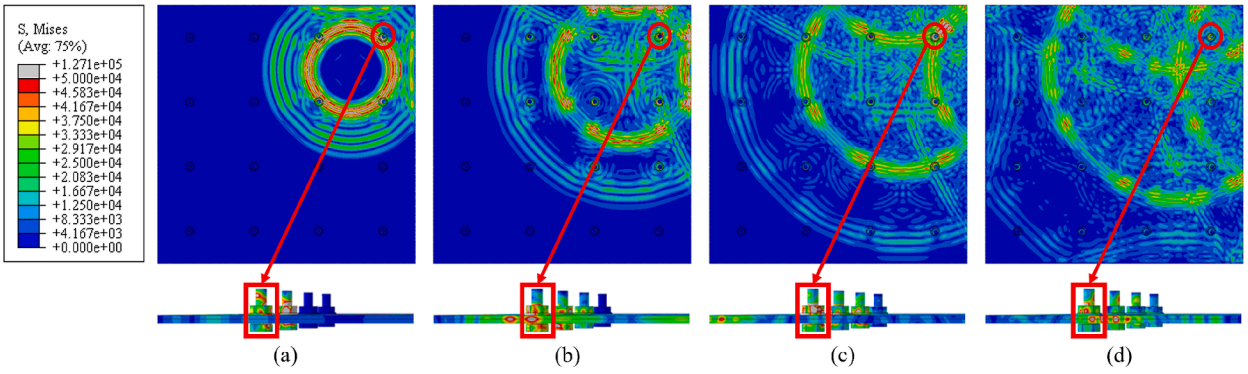


**Fig. 4.** Propagation of Lamb waves in the Connected plate with sixteen bolts: (a) 3.0 $\times 10^{-5}$ *s*; (b) 4.5 $\times 10^{-5}$ *s*; (c) 6.0 $\times 10^{-5}$ *s*; (d) 7.5 $\times 10^{-5}$ *s*.

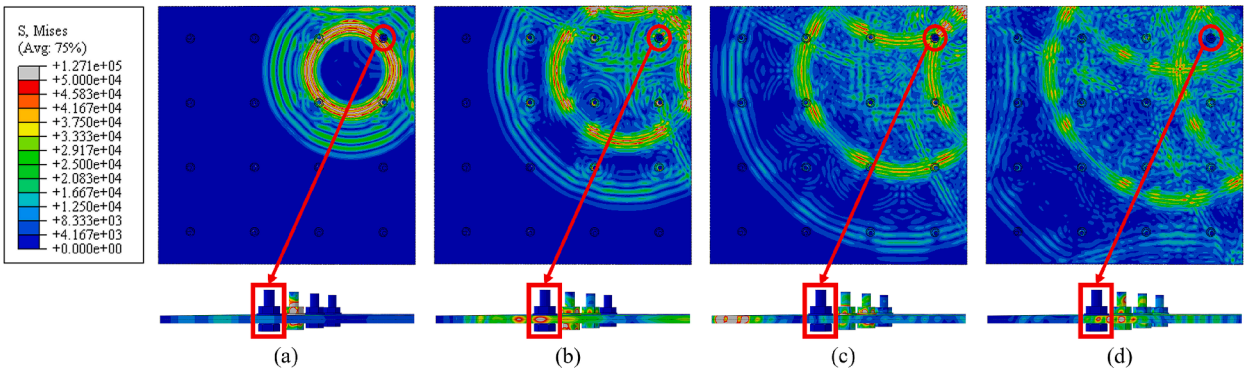


**Fig. 5.** Propagation of Lamb waves in the Damaged plate with sixteen bolts: (a) 3.0 $\times 10^{-5}$ *s*; (b) 4.5 $\times 10^{-5}$ *s*; (c) 6.0 $\times 10^{-5}$ *s*; (d) 7.5 $\times 10^{-5}$ *s*.

is achievable in numerical simulations, guaranteeing such high consistency in practical engineering application is difficult. Various factors such as temperature, external loads, and material property changes can affect the plate during service, leading to a mismatch between the baseline and damaged signals, which makes the use of residual signals for analysis inaccurate. Although the effects of these factors can be filtered out using signal processing techniques, precise compensation or correction is challenging and time-consuming. Therefore, this paper utilized a 1D CNN for bolt loosening detection in the multi-bolted plates. This algorithm is renowned for its powerful non-linear analysis capabilities, making it possible to capture target features under multiple influencing factors without the need of baseline signals.

### 3.3. Simulation database and 1D CNN architecture

In this paper, a 17-class classification was performed using a 1D CNN model, with the first class was labeled as "Healthy", representing a safe bolt-plate connection. The second to seventeenth classes were labeled as "Damage" 1 to 16, representing the loosening of Bolts 1 to 16 respectively. In each simulation, excitation was applied to one of the PZTs, while signals were received from the other three PZTs, with 1500 sampling points.

Temperature and white noise effects were added to the received signals. According to [59,60], the signal phase and amplitude change when temperature varies. In this paper, the phase change range of the signal was set to 7 %, and the amplitude change range was set to 10 % to simulate signals collected over a temperature variation of 25 degrees. In addition, Gaussian white noise with a signal-to-noise ratio of 30 dB was added to the signals. Taking the example of PZT1 exciting and PZT2 receiving, five sets of received signals with temperature and white noise effects and their residual signals are shown in Fig. 6. It can be seen that it is very difficult to visually determine the location of the damage wave based on the residual signals in this situation.

To construct the training and testing databases for the 1D CNN model, received signals under different PZT excitations were concatenated end-to-end to form a new vector of length 18000, which served as the input to the 1D CNN model. Examples of signals for Healthy class and Damage 1 class are shown in Fig. 7. Each set of signals was expanded to 100 and 50 samples, respectively, by adding temperature and white noise effects. So, the size of the training dataset is 1700 × 18000, and the testing dataset is 850 × 18000. In addition, both the training and testing dataset were normalized before being fed into the 1D CNN model.

As shown in Fig. 8, the 1D CNN model used in this study consists of a total of seven layers, including one input layer, three

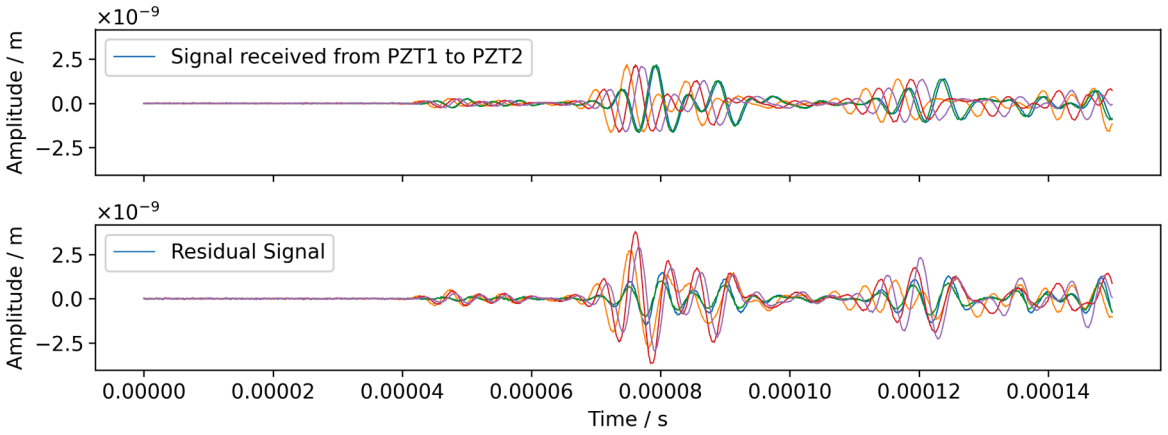


**Fig. 6.** Signals received from PZT1 to PZT2 and their residual signals.


**Fig. 7.** Input vector for 1D CNN (a) from Health Class in which all the bolts and plate were tightly connected, and (b) from Class 1 in which Bolt 1 was loosed.

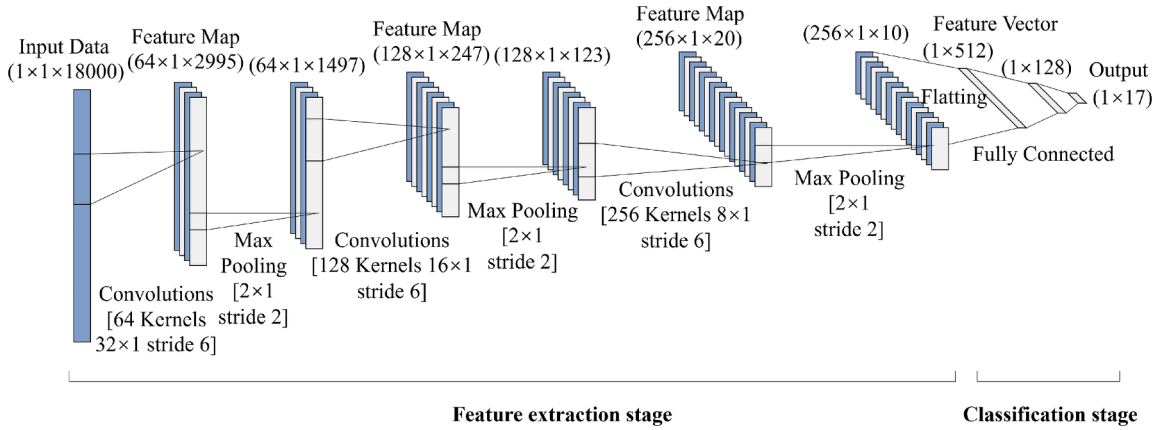


**Fig. 8.** The architecture of the 1D CNN model in simulation and experiment.

convolutional layers, two fully connected layers, and one output layer. The input layer has a size of 18000. The output channels of the convolutional layers were set to 64, 128, and 256, with kernel sizes of 32, 16 and 8, respectively. The stride was set to 6, and the pooling layer has size of 2. The two fully connected layers have 512 and 128 neurons, respectively, and both used the *ReLU* activation function. The length of the output layer is 17.

### *3.4. XAI result*

The 1D CNN model was trained on the training dataset with 1080 epochs and 0.0001 learning rate, and achieved 96.706 % accuracy on the testing database. XAI algorithms were then used to interpret the well-trained model. Taking Damage 3 as an example, the importance-score saliency maps obtained by Sensitivity Analysis (SA), Smooth Simple Taylor (SST), Deep Taylor (DT), Grad CAM (GC), Deep Grad CAM (DGC), and Guided Grad CAM (GGC) are shown in Fig. 9. It is evident that the saliency maps obtained by different XAI algorithms have significant differences, indicating their explanations for the classification of the 1D CNN model are very different.

The details of the Smooth Simple Taylor's results on Damage 1 and Damage 3 class are shown in Fig. 10. For the Damage 1 class, Smooth Simple Taylor considers that the 1D CNN model mainly relies on information from three monitoring paths: PZT1 – PZT2 (PZT1 to PZT2 and PZT2 to PZT1), PZT1 – PZT3 (PZT1 to PZT3 and PZT3 to PZT1), and PZT4 to PZT1 to determine the signal belongs do Damage 1 class; For the Damage 3 class, Smooth Simple Taylor considers that the 1D CNN model mainly relies on information from two monitoring paths: PZT1 – PZT2 (PZT1 to PZT2 and PZT2 to PZT1) and PZT2 – PZT3 (PZT2 to PZT3 and PZT3 to PZT2), to determine the signal belongs to Damage 3 class, and the importance scores of these two paths show a high symmetry. Both of the classification patterns are consistent with human expert knowledge in SHM. As in the case of Damage 1, the monitoring paths PZT1 – PZT2, PZT1 – PZT3 and PZT4 to PZT1 are very close to the location of the damage. And in the case of Damage 3, the monitoring paths PZT1 – PZT2 and PZT2 – PZT3 are the closest to the location of the damage. Theoretically, the strongest damage reflection waves can be captured on these nearest paths. Therefore, it is logical that the 1D CNN model considers these monitoring paths as the most important for the Damage 1 and Damage 3 classification. This also demonstrates that the 1D CNN model effectively filtered the influence of temperature

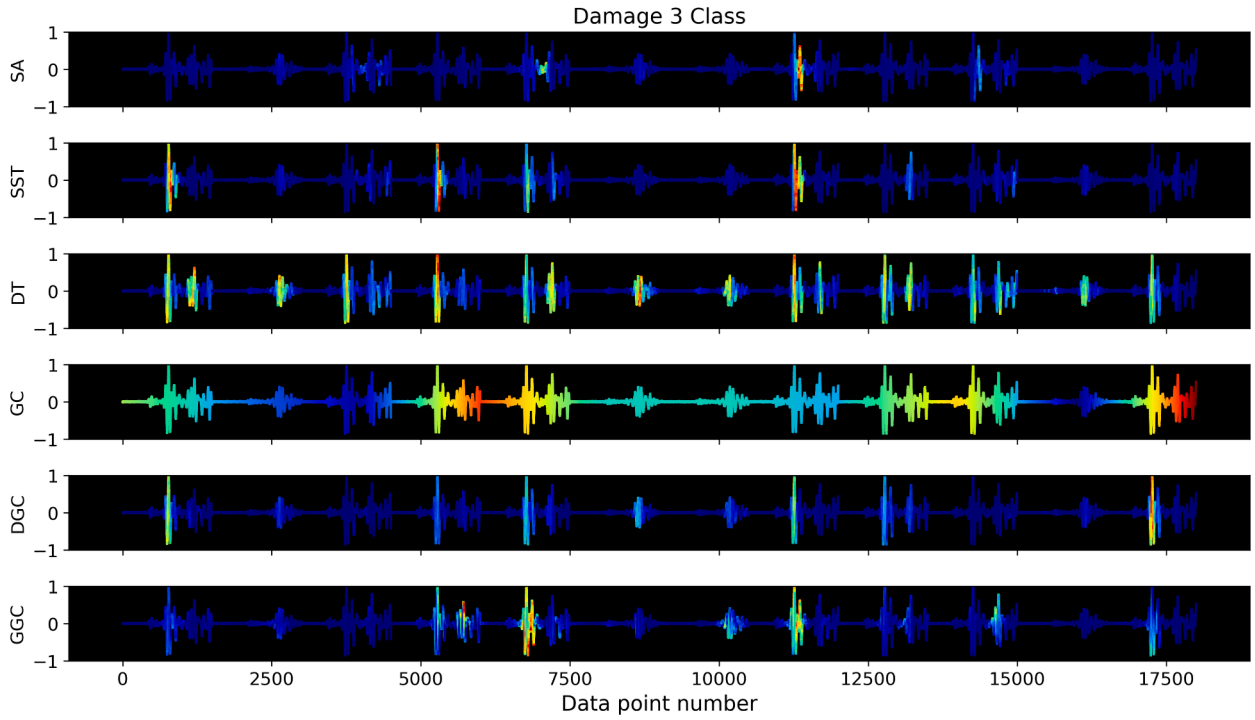


**Fig. 9.** Saliency maps for the Damage 3 classification from Sensitivity analysis (SA), Smooth Simple Taylor (SST), Deep Taylor (DT), Grad CAM (GC), Deep Grad CAM (DGC) and Guided Grad CAM (GGC).

and white noise.

### 3.5. Classification patterns

To better observe the classification patterns of the 1D CNN model, the dependence scores of the monitoring paths were calculated for sixteen damage scenarios using the XAI algorithm, residual signal (Residual) algorithm, and the residual signal without temperature influence (Raw Residual) algorithm. For the XAI algorithm, the dependency score of each monitoring path is the sum of the importance scores of the signals on that path calculated by the XAI algorithms:

$$D_P = \sum_{s=1}^{18000} \Phi(\mathrm{F}, v_s^P) \tag{17}$$

where $D_P$ represents the dependency score of the $P$-th monitoring path, and $v_s^P$ is the $s$-th data point in the received signals on $P$-th path.

For the Residual algorithm, the sum of the absolute values of the residual signals on the monitoring path is used as the dependency score:

$$D_P = \sum_{s=1}^{18000} \left| v_s^P - v_s^{P,base} \right| \tag{18}$$

where $v_s^{P,base}$ is the $s$-th data point in the baseline signals on $P$-th path. In the case of the Raw Residual algorithm, $v_s^P$ and $v_s^{P,base}$ are the signals without changing temperature.

The comparison between the dependency-score saliency maps of the Smooth Simple Taylor and Residual algorithm is shown in Fig. 11. The results of the Smooth Simple Taylor reflect very strong regularity and symmetry: the monitoring paths closest to the damage are considered to be more important in all the damage scenarios. This indicates that Smooth Simple Taylor believes the 1D CNN model considers monitoring paths closer to the damage can provide more critical damage features during the decision-making process, which is very similar to human expert cognition.

On the other hand, as shown in Fig. 11 (b), the result of the Residual algorithm reflects almost identical classification patterns in all the damage scenarios. Therefore, it is known that under the influence of temperature and white noise, it is difficult to effectively determine the location of the damage by referring to the residual signals. However, the 1D CNN model is capable of intelligently filtering out these influences.

Furthermore, observing the distribution of the sixteen damage scenarios, they can be divided into four damage patterns: DP 1

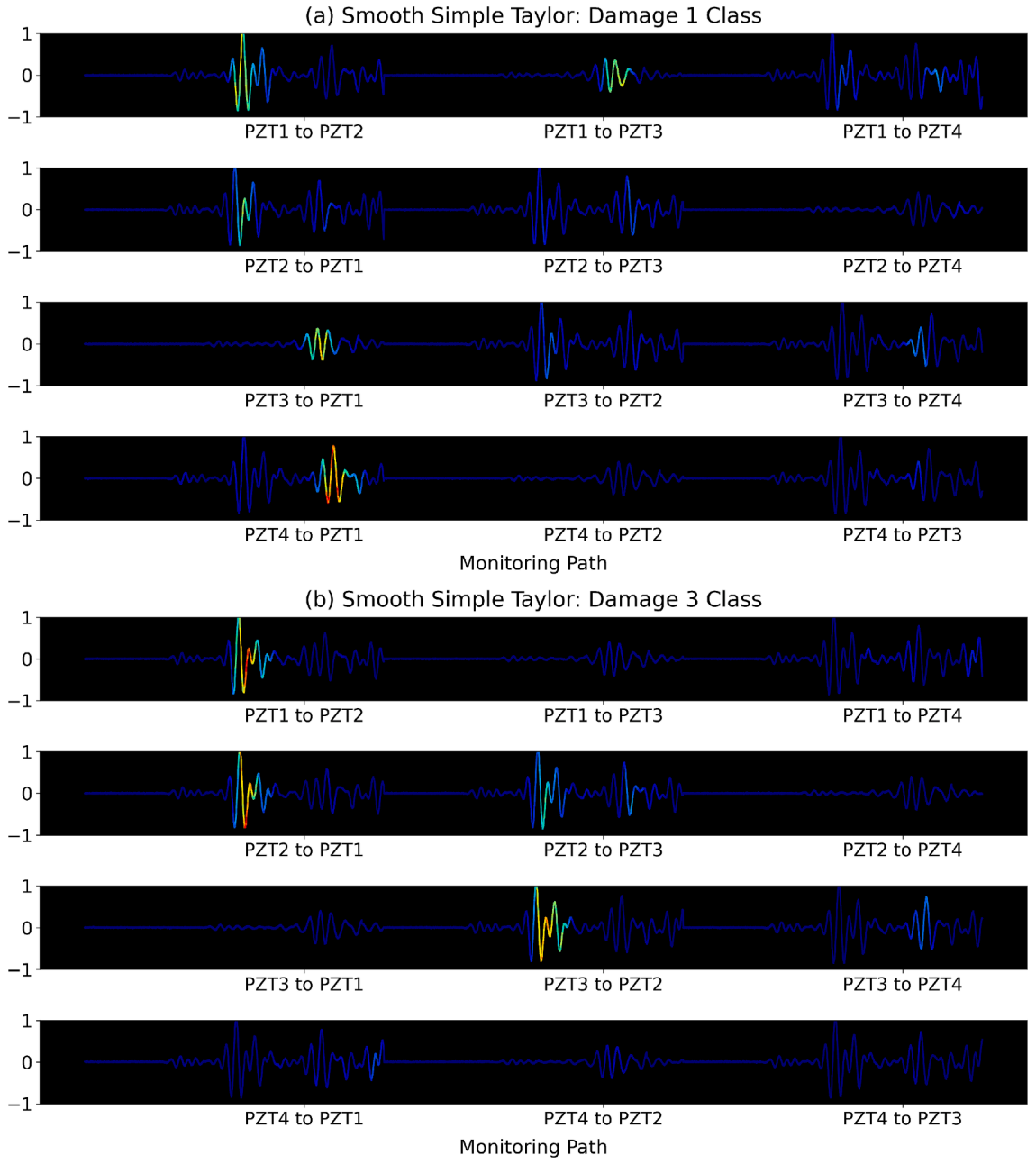


**Fig. 10.** Details of the importance-score Saliency map from Smooth Simple Taylor: (a) Damage 1 Class (b) Damage 3 Class.

consists of Damage 1, 4, 13, and 16, DP 2 consists of Damage 2, 8, 15, and 9, DP3 consists of 3, 12, 14, and 5, DP4 consists of 6, 7, 10, and 11. The damages from the same damage pattern can be obtained by rotating each other, for example, rotating Damage 4 by 90 degrees results in Damage 1, and so on. And the damages from the same damage pattern are plotted with the same shape in the figure. By adding up the dependency-score saliency maps of the same damage pattern, the classification rules for these four damage patterns can be analyzed.

Fig. 12 shows the classification rules for the four damage patterns calculated with the Residual, Raw Residual and six XAI algorithms. It can be seen that the classification rules of the Residual algorithm are almost identical on all damage scenarios, so it cannot provide effective information for damage classification. The results from the Raw Residual algorithm shows that in all four damage patterns, monitoring paths closer to the damage are given higher dependence scores. Furthermore, DP 1 and 4, and DP 2 and 3 showed close classification rules. Therefore, it can be inferred that the use of residual signals can effectively distinguish different damage patterns when the reference and damaged plates are in a good match.

For the results of XAI algorithms, the result from Smooth Simple Taylor is closest to the Raw Residual algorithm, and with a higher

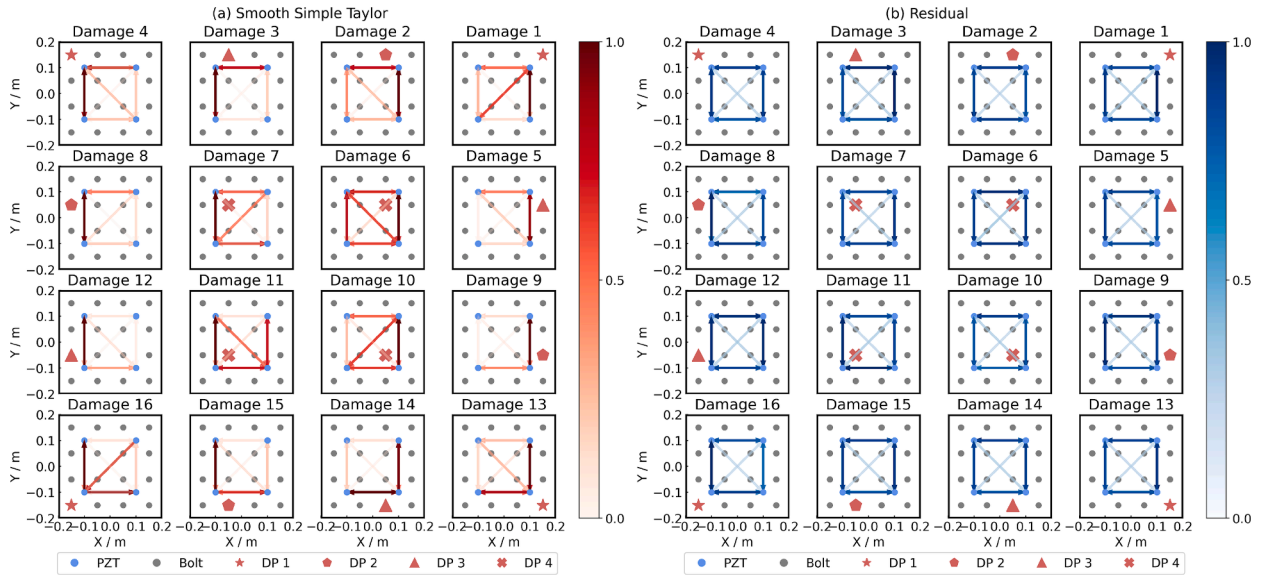


**Fig. 11.** The dependency-score saliency maps for sixteen damage scenarios calculated by (a) Smooth Simple Taylor and (b) Residual algorithm.

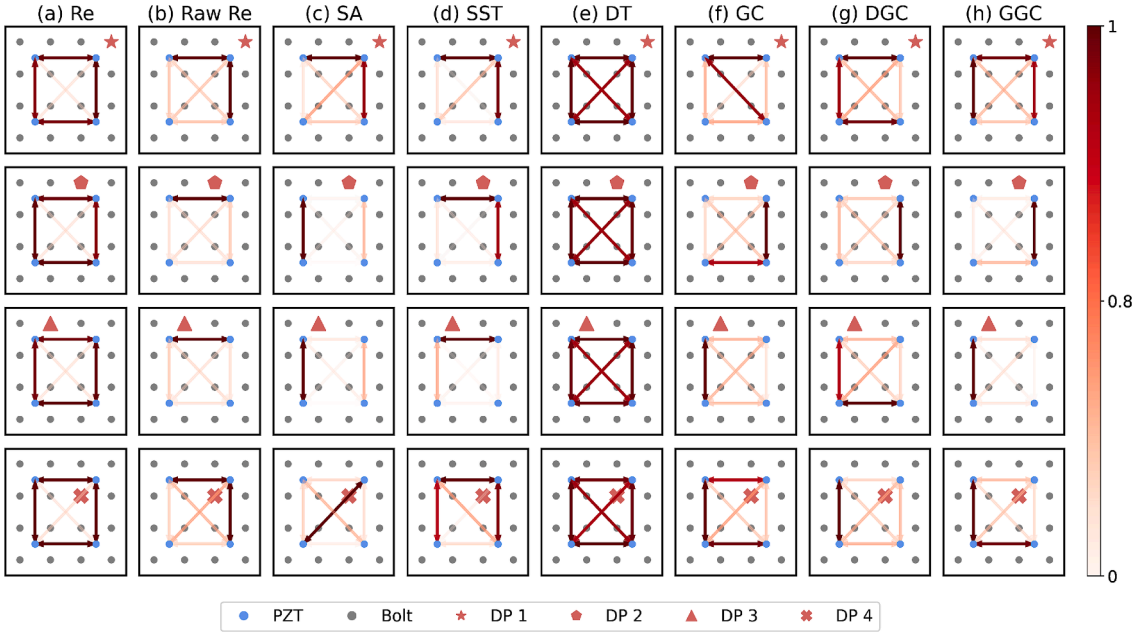


**Fig. 12.** The classification rules for four damage patterns calculated by (a) Residual algorithm (Re) (b) Raw Residual algorithm (Raw Re) (c) Sensitivity analysis (SA) (d) Smooth Simple Taylor (SST) (e) Deep Taylor (DT) (f) Grad CAM (GC) (g) Deep Grad CAM (DGC) (h) Guided Grad CAM (GGC).

distinction for the four damage patterns. This means its classification rules are more targeted. The result obtained from Deep Taylor shows the lowest distinction, suggesting the 1D CNN model almost equally referenced information from all monitoring paths during the decision-making process. The classification rule of DP 1 from the Sensitivity analysis is very similar to that of the Raw Residual algorithm. DP 2 and 3 have the same classification rules, while in DP 4, monitoring paths passing through the damage are considered the most important. The results from Grad CAM, Deep Grad CAM, and Guided Grad CAM are very close compared to other algorithms, with some differences in the details.

## 4. Experiment

### 4.1. Experimental setup

A bolt loosening detection experiment was conducted to construct the experimental database. The experimental setup, shown in Fig. 13, consists of an arbitrary waveform generator (Agilent 33502A), a digital oscilloscope (Pico Scope 6402A), an amplifier (Agilent 33502A), and the aluminum plate specimen. A torque wrench was used to control the bolt torque. The dimensions, material type, PZTs arrangement, and bolt layout of the specimen are fully consistent with those in the numerical simulation.

In the experiment, a 1D CNN model was also utilized for the 17-class classification. The first class was labelled as “Healthy”, where the bolts were tightened to a torque of 4 N•m using a torque wrench, ensuring a secure connection between the bolts and the plates. For the sixteen damage classes labelled as “Damage” 1 to 16, the torque was removed from one of the bolts each time to simulate bolt loosening.

In each experiment, excitation was applied to one of the PZTs, while the signals were received by the remaining three PZTs. The received signal has 1500 data points, and a sampling frequency of 10 MHz, and a recording time of 0.15 ms. A total of 150 signal samples were recorded for each case, with 100 samples used for training and 50 samples for testing. Temperature effects were also introduced to the signals, with a phase variation range of 7 % and an amplitude variation range of 10 % to simulate signals collected over a temperature variation of 25 degrees [59,60]. Taking the example of PZT1 exciting and PZT2 receiving, five sets of received signals with temperature and white noise effects and their residual signals are shown in Fig. 14. Consequently, the training dataset consisted of 1700 samples with a size of 18000, while the testing dataset consisted of 850 samples with a size of 18000. Example signals for the Healthy class and the Damage 1 class are shown in Fig. 15. The same 1D CNN model architecture as the numerical simulation was employed in the experiment.

### 4.2. XAI result and classification rules

In the experiment, the 1D CNN model achieved an accuracy of 97.647 % on the testing dataset. Fig. 16 presents the details of the importance-score saliency map obtained by Smooth Simple Taylor for the classification of Damage 1 and Damage 3. Consistent with the results from the numerical simulation, Smooth Simple Taylor indicates that the 1D CNN model primarily relies on information from the monitoring paths closest to the damage for classification, which aligns with the logic of SHM methods. In the case of Damage 1 class, PZT1-PZT2 and PZT1-PZT4 are considered important. In the case of Damage 3 class, PZT2-PZT3, PZT2-PZT4 and PZT2 to PZT1 are considered important.

Fig. 17 displays the dependency-score saliency maps for the four damage patterns obtained through Residual, Raw Residual, and six different XAI algorithms in the experiment. It can be seen that the results from the Residual algorithm show nearly identical classification patterns across the four different damages patterns, which indicates that it is challenging to classify the damages using the residual signal under the influence of temperature variations. The results from the Raw Residual algorithm reveal that monitoring paths closer to the damage are assigned higher dependence scores in all four damage patterns. Hence, it can be concluded that if the temperature remains constant, the residual signal can still effectively differentiate between damage patterns. As for the results of the six XAI algorithms, Smooth Simple Taylor shows the closest similarity to the results from the Raw Residual algorithm. In all cases, the monitoring paths closest to the damage are considered more important. Conversely, Deep Taylor exhibits the lowest distinction among the four damage patterns, suggesting that the 1D CNN almost equally references information from all monitoring paths during the decision-making process.

## 5. Evaluation of the explanation result of XAI

Based on the research findings from the numerical simulations and experiments, it can be seen that XAI can provide information

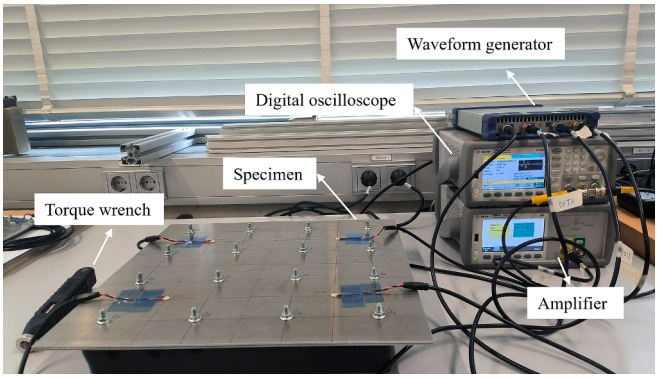


**Fig. 13.** Experimental setup for bolt loosening detection.

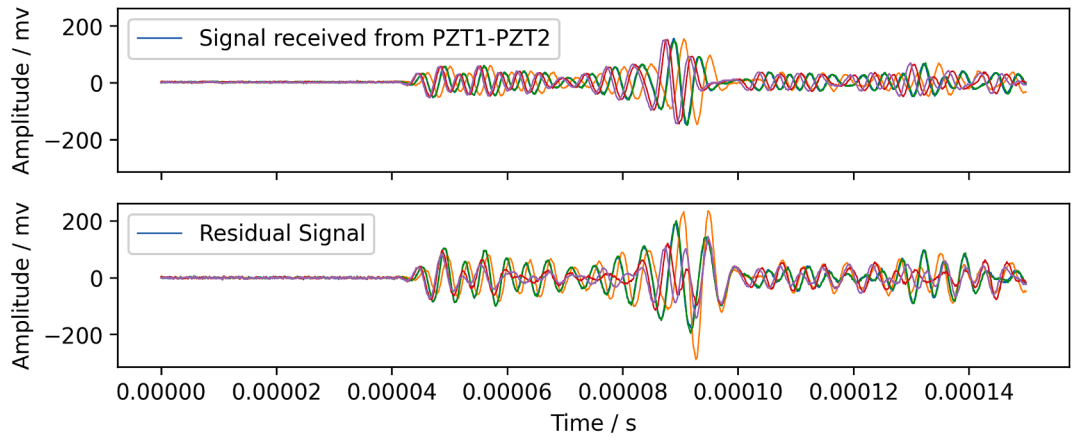


**Fig. 14.** Signals received from PZT1 to PZT2 and their residual signals in the experiment.

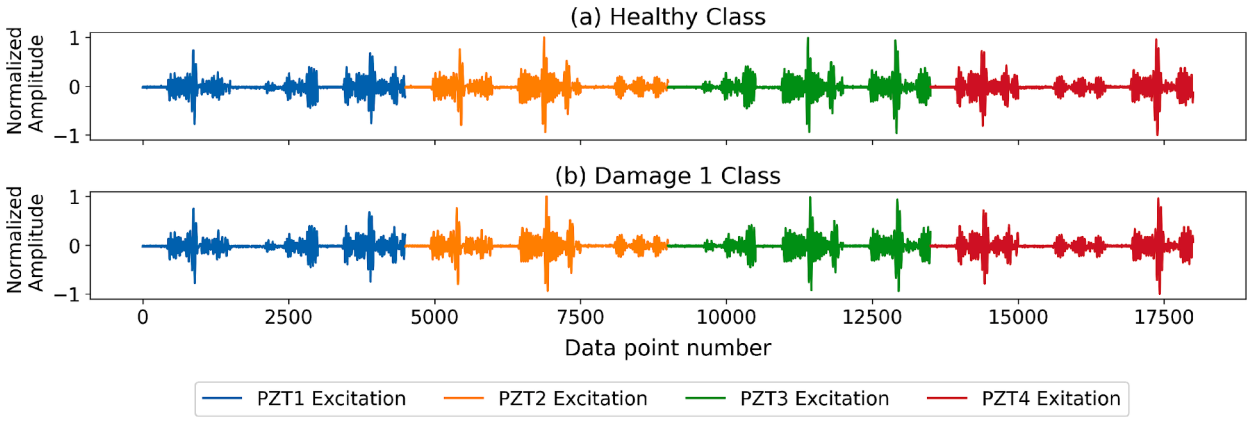


**Fig. 15.** Input vector for 1D CNN in the experiment (a) from Healthy class (b) from Damage 1 class.

about the specific segments of the signal that the 1D CNN model predominantly leverages during its decision-making process. This can enhance the transparency of the model and contribute to the human understanding of AI's analyzing mechanisms from a physical perspective. However, it can also be observed that due to different perspectives, the explanations obtained from various XAI algorithms exhibit significant differences. This naturally raises the question: How can we select the most suitable XAI algorithm among the numerous options?

In this study, five different evaluation methods, including CCRS, RWIR, INFD, SENS, and SACH were used to assess the results of the XAI algorithms. The average values of the evaluation methods for these XAI algorithms were calculated across 17 classes in both numerical simulations and experiments, as shown in Fig. 18. To facilitate a convenient comparison of algorithm performance, the reciprocal values of Infidelity, Sense Sum, and Sanity Check results were adopted. Therefore, in all bar charts, the taller the bar, the better the performance of the algorithm under that evaluation method.

By comparing the evaluation results from numerical simulations and experiments, it can be observed that the algorithm rankings for CCRS, RWIR, and SACH are completely identical. In terms of Sense Sum, except for the reversed rankings of Sensitivity Analysis and Deep Grad CAM, the rankings of the other algorithms are entirely consistent. INFD exhibits the greatest variation among these five evaluation methods. Consequently, it can be inferred that, overall, the evaluation results of the algorithms in numerical simulations and experiments are in good agreement.

Based on the results of CCRS and RWIR, it can be inferred that from the perspective of SHM that Smooth Simple Taylor is the optimal choice for comprehending the decisions made by the 1D CNN model. Because according to equations (12) and (13), higher values of CCRS indicates a better alignment between the XAI's explanation and the Raw Residual, higher values of RWIR indicates a higher proportion of wave-package regions are considered important. Consequently, the explanations of Smooth Simple Taylor are more consistent with human analyzing patterns in SHM and are easier to comprehend.

Then the performance of each XAI algorithm was scored based on the results of the five evaluation methods. For instance, in the CCRS result, Smooth Simple Taylor received a score of 6 as the top-ranked algorithm, while GC received a score of 1 as the lowest-ranked algorithm, and so on. After summing up the scores from all the evaluation methods, the results are shown in Table 1. It can be observed that there is a high level of consistency in the score rankings between numerical simulations and experiments. Except for

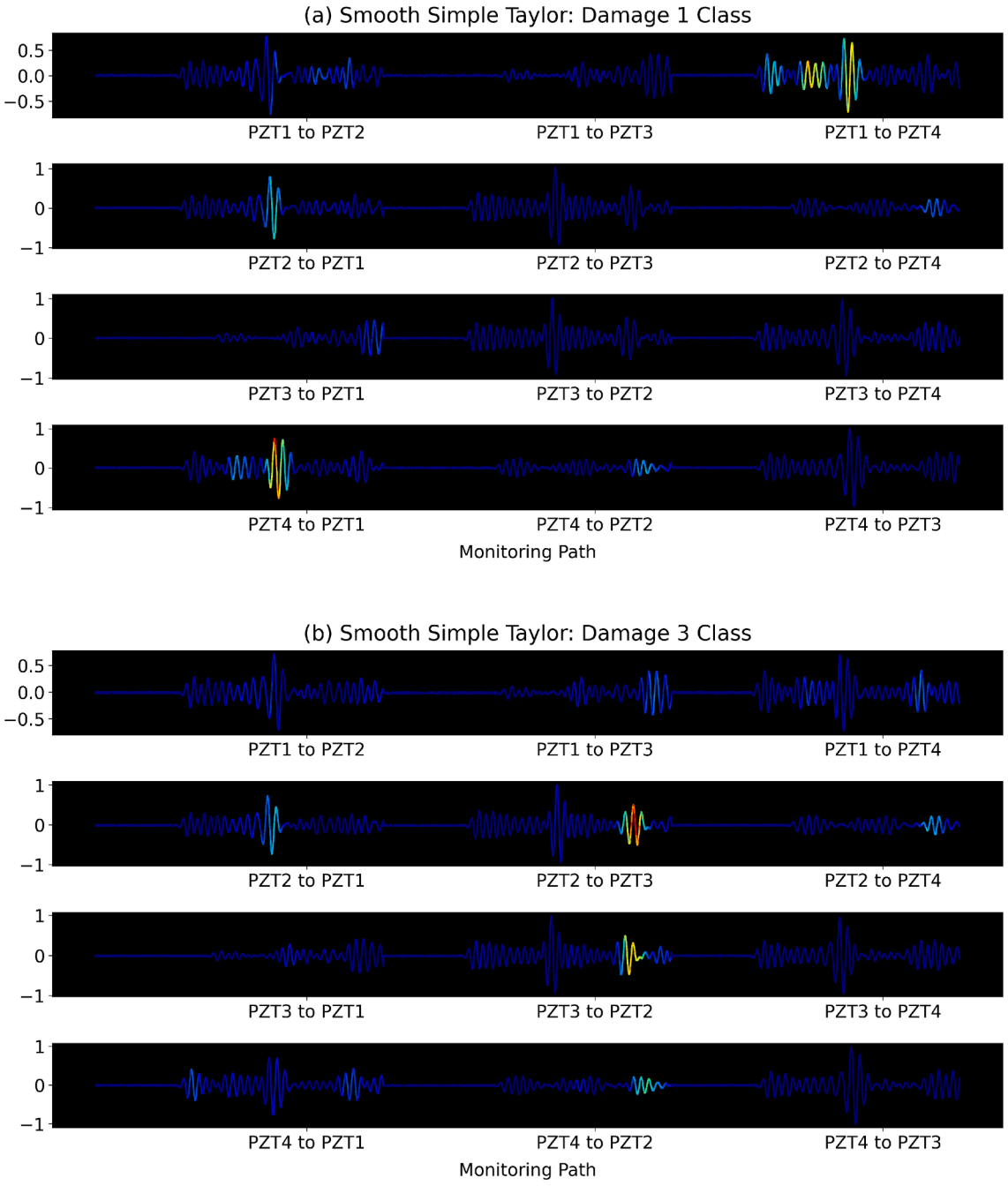


**Fig. 16.** Details of the importance-score saliency map from Smooth Simple Taylor: (a) Damage 1 Class (b) Damage 3 Class.

the swapped positions of Grad CAM and Guided Grad CAM, all other algorithms have identical rankings. Additionally, there is only a one-point difference between Grad CAM and Guided Grad CAM, and the score differences for other algorithms are within 2 points. Therefore, it can be concluded that the ranking results of the XAI algorithms are relatively consistent between numerical simulations and experiments.

## 6. Discussion

To address the *black box* effect of deep learning in DeepSHM method, the XAI algorithms were used to interpret the decision-making process of the 1D CNN model. The physical basis of the 1D CNN model for bolt loosening identification was analyzed, enhancing the transparency and reliability of the method. To achieve more accurate and SHM-principle-aligned explanations, two improved

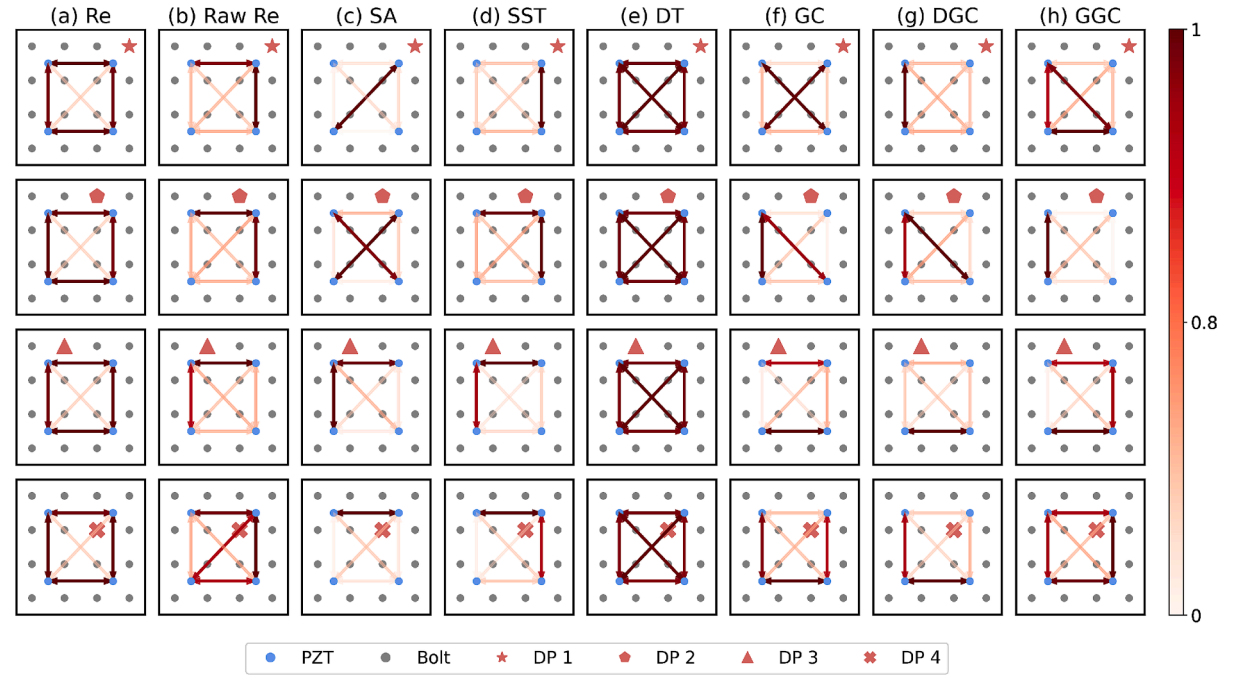


**Fig. 17.** The classification rules for four damage patterns in the experiment calculated by (a) Residual algorithm (Re) (b) Raw Residual algorithm (Raw Re) (c) Sensitivity analysis (SA) (d) Smooth Simple Taylor (SST) (e) Deep Taylor (DT) (f) Grad CAM (GC) (g) Deep Grad CAM (DGC) (h) Guided Grad CAM (GGC).

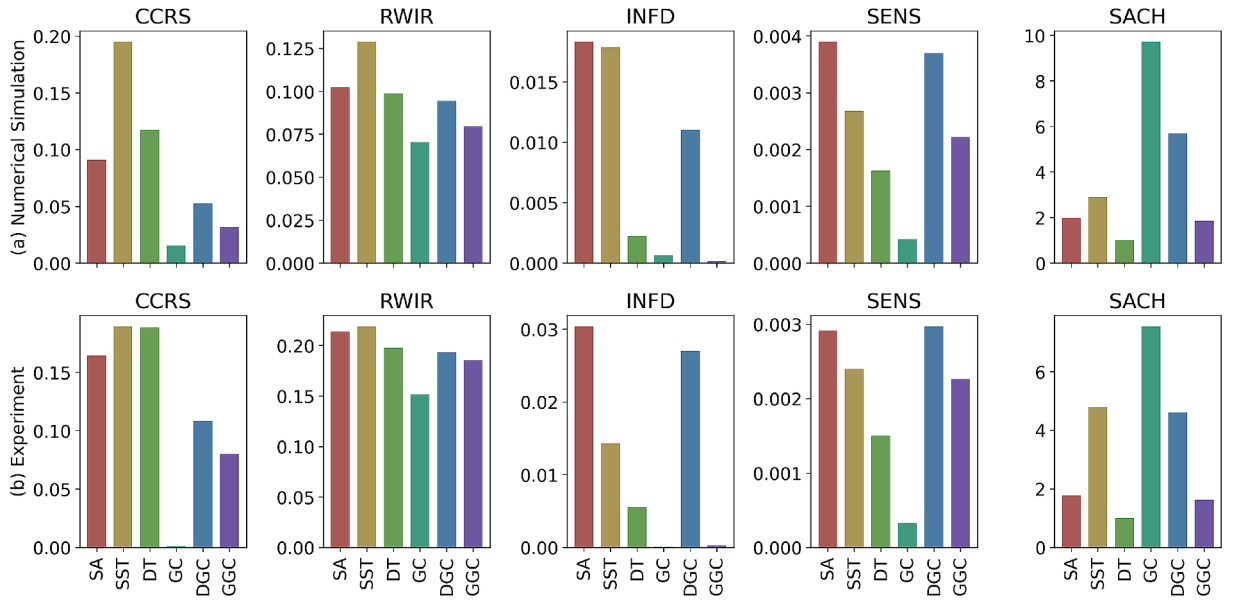


**Fig. 18.** The average values of the evaluation methods for XAI algorithms across 17 classes from (a) numerical simulation (b) experiments.

**Table 1**
The sum of ranking scores for XAI algorithms across all evaluation methods.

| | Sensitivity Analysis | Smooth Simple Taylor | Deep Taylor | Grad CAM | Deep Grad CAM | Guided Grad CAM |
|---|---|---|---|---|---|---|
| Simulation | 24 | 25 | 15 | 11 | 20 | 10 |
| Experiment | 23 | 24 | 15 | 10 | 22 | 11 |

algorithms—Smooth Simple Taylor and Deep Grad CAM—were introduced, both of which demonstrate significant performance enhancements. The introduction of such physical interpretations offers stronger acceptability for the application of deep learning in complex engineering structures, particularly in SHM scenarios where high reliability and stringent scrutiny are required.

From the perspective of SHM applications, two novel evaluation methods (CCRS and RWIR) for XAI techniques were established. These metrics not only assist AI researchers in systematically evaluating algorithm performance but also provide a theoretical foundation and practical evaluation tools for applying XAI in real-world SHM engineering. Moreover, under the consideration of five distinct evaluation criteria, the rankings of six different XAI algorithms exhibit remarkable consistency between numerical simulations and experimental results. This consistency suggests the potential for algorithms performing well in simulations to achieve similarly high performance in experimental settings. Thus, it is anticipated that numerical simulations could guide the selection of XAI algorithms for experimental applications in the future, significantly reducing both the economic and time costs associated with experimental studies.

## 7. Conclusion

The propagation of Lamb waves in a double-layered aluminum plate with 16 bolt connections were investigated in this study through numerical simulation and experiment. The influence of bolt loosening on the propagation of Lamb waves was analyzed. Subsequently, 1D CNN models were trained using Lamb wave signals and employed for the detection of bolt loosening. The well-trained 1D CNN models were analyzed using XAI algorithms, including Sensitivity Analysis, Smooth Simple Taylor, Deep Taylor, Grad CAM, Deep Grad CAM, and Guided Grad CAM. The performance of which was evaluated using CCRS, RWIR, INFD, SENS, and SACH evaluation methods, and the following conclusions are drawn:

1. The 1D CNN has significant advantages in detecting bolt loosening in multi-bolted structures and challenging operating conditions. In the presence of temperature variations, where traditional methods relying on residual signals failed to provide accurate damage information, the 1D CNN maintained detection accuracies of 96.706 % and 97.647 % in numerical simulation and experiment respectively
2. XAI can provide a physical interpretation for the classification mechanism of the 1D CNN model. By identifying the specific segments of the input signal that are essential for the decision-making process of the 1D CNN, XAI can assist human observers in better comprehending AI's operation from a physical perspective, which contributes to the improvement of the transparency and trustworthiness of the 1D CNN model
3. From the perspective of SHM, Smooth Simple Taylor is the optimal choice for explaining the 1D CNN model. It consistently ranks first in both numerical simulations and experiments in the results of CCRS and RWIR, which confirms its strong alignment with SHM's analytical logic. Additionally, it secures the first position in the overall rankings among the five evaluation methods, further supporting its superiority
4. The ranking scores of XAI algorithms, calculated based on five different evaluation methods, exhibit a high level of consistency between numerical simulations and experiments. Smooth Simple Taylor, Sensitivity Analysis, Deep Grad CAM, and Deep Taylor consistently maintain an identical order, while Grad CAM and Guided Grad CAM exchange positions with only a difference of one point

## Statement

During the preparation of this manuscript, the authors used ChatGPT to improve the readability of the original draft. The authors critically reviewed the output of ChatGPT and edited the manuscript as needed. The authors take full responsibility for the content of the publication.

## CRediT authorship contribution statement

**Muping Hu:** Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Sasan Salmani Pour Avval:** Writing – review & editing, Validation, Software, Methodology. **Jian He:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Nan Yue:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Roger M. Groves:** Writing – review & editing, Resources, Methodology, Funding acquisition, Conceptualization.

## 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.

## Acknowledgment

The authors gratefully acknowledge financial support from the China Scholarship Council (No. 202206680031) and the European Union Horizon Europe OVERLEAF project (No. 101056818).

## Data availability

Dataset will be made avaliable on the 4TU.ResearchData repository (data.4TU.nl)

## References


[1] C. Li, R. Qiao, Q. Tang, X. Miao, Investigation on the Vibration and Interface State of a Thin-Walled Cylindrical Shell with Bolted Joints Considering Its Bilinear Stiffness, Appl. Acoust. 172 (2021).
[2] H. Gong, J. Liu, X. Ding, Thorough Understanding on the Mechanism of Vibration-Induced Loosening of Threaded Fasteners Based on Modified Iwan Model, J. Sound Vib. 473 (2020).
[3] N.G. Pai, D.P. Hess, Three-Dimensional Finite Element Analysis of Threaded Fastener Loosening Due to Dynamic Shear Load, Eng. Fail. Anal. 9 (2002) 383–402.
[4] K. He, W.D. Zhu, Detecting Loosening of Bolted Connections in a Pipeline Using Changes in Natural Frequencies, J. Vib. Acoust. 136 (2014).
[5] R. Lacayo, L. Pesaresi, J. Groß, D. Fochler, J. Armand, L. Salles, C. Schwingshackl, M. Allen, M. Brake, Nonlinear Modeling of Structures with Bolted Joints: A Comparison of Two Approaches Based on a Time-Domain and Frequency-Domain Solver, Mech. Syst. Sig. Process. 114 (2019) 413–438.
[6] Y. Liu, J. Zhi, E. Liu, Y. Chen, X. Wang, C. Sun, C. zifei, H. Ma, J. Tan, Influence of Different Ultrasonic Transducers on the Precision of Fastening Force Measurement, Appl. Acoust. 185 (2022).
[7] J. Huang, J. Liu, H. Gong, X. Deng, A Comprehensive Review of Loosening Detection Methods for Threaded Fasteners, Mech. Syst. Sig. Process. 168 (2022).
[8] F. Amerini, E. Barbieri, M. Meo, U. Polimeno, Detecting Loosening/Tightening of Clamped Structures Using Nonlinear Vibration Techniques, Smart Mater. Struct. 19 (2010).
[9] Q.K. Li, X.J. Jing, Fault Diagnosis of Bolt Loosening in Structures with a Novel Second-Order Output Spectrum-Based Method, Struct. Health Monit. 19 (2020) 123–141.
[10] J.J. Meyer, D.E. Adams, Using Impact Modulation to Quantify Nonlinearities Associated with Bolt Loosening with Applications to Satellite Structures, Mech. Syst. Sig. Process. 116 (2019) 787–795.
[11] F. Wang, S.C.M. Ho, G. Song, Modeling and Analysis of an Impact-Acoustic Method for Bolt Looseness Identification, Mech. Syst. Sig. Process. 133 (2019).
[12] F. Wang, S.C.M. Ho, L. Huo, G. Song, A Novel Fractal Contact-Electromechanical Impedance Model for Quantitative Monitoring of Bolted Joint Looseness, IEEE Access 6 (2018) 40212–40220.
[13] Y. Zhuang, F. Kopsaftopoulos, R. Dugnani, F.K. Chang, Integrity Monitoring of Adhesively Bonded Joints Via an Electromechanical Impedance-Based Approach, Struct. Health Monit. 17 (2018) 1031–1045.
[14] M. Yeager, A. Whitaker, M. Todd, A Method for Monitoring Bolt Torque in a Composite Connection Using an Embedded Fiber Bragg Grating Sensor, J. Intell. Mater. Syst. Struct. 29 (2018) 335–344.
[15] D.D. Chen, L.S. Huo, H.N. Li, G.B. Song, A Fiber Bragg Grating (FBG)-Enabled Smart Washer for Bolt Pre-Load Measurement: Design, Analysis, Calibration, and Experimental Validation, Sensors (switzerland) 18 (2018).
[16] N. Hosoya, I. Kajiwara, T. Hosokawa, Vibration Testing Based on Impulse Response Excited by Pulsed-Laser Ablation: Measurement of Frequency Response Function with Detection-Free Input, J. Sound Vib. 331 (2012) 1355–1365.
[17] F. Huda, I. Kajiwara, N. Hosoya, S. Kawamura, Bolt Loosening Analysis and Diagnosis by Non-Contact Laser Excitation Vibration Tests, Mech. Syst. Sig. Process. 40 (2013) 589–604.
[18] F.R. Wang, Z. Chen, G.B. Song, Smart Crawfish: A Concept of Underwater Multi-Bolt Looseness Identification Using Entropy-Enhanced Active Sensing and Ensemble Learning, Mech. Syst. Sig. Process. 149 (2021).
[19] F. Wang, Z. Chen, G. Song, Monitoring of Multi-Bolt Connection Looseness Using Entropy-Based Active Sensing and Genetic Algorithm-Based Least Square Support Vector Machine, Mech. Syst. Sig. Process. 136 (2020) 106507.
[20] H. Gong, X. Deng, J. Liu, J. Huang, Quantitative Loosening Detection of Threaded Fasteners Using Vision-Based Deep Learning and Geometric Imaging Theory, Autom. Constr. 133 (2022) 104009.
[21] X. Qin, C. Peng, G. Zhao, Z. Ju, S. Lv, M. Jiang, Q. Sui, L. Jia, Full Life-Cycle Monitoring and Earlier Warning for Bolt Joint Loosening Using Modified Vibro-Acoustic Modulation, Mech. Syst. Sig. Process. 162 (2022) 108054.
[22] Y. Yang, C.T. Ng, A. Kotousov, Bolted Joint Integrity Monitoring with Second Harmonic Generated by Guided Waves, Struct. Health Monit. 18 (2019) 193–204.
[23] K.D. Tola, C. Lee, J. Park, J.W. Kim, S. Park, Bolt Looseness Detection Based on Ultrasonic Wavefield Energy Analysis Using an Nd:Yag Pulsed Laser Scanning System, Struct. Control Health Monit. 27 (2020).
[24] F. Du, S.W. Wu, R.Z. Sheng, C. Xu, H. Gong, J. Zhang, Investigation into the Transmission of Guided Waves across Bolt Jointed Plates, Appl. Acoust. 196 (2022).
[25] E.M. Hassan, H. Mahmoud, G. Riveros, S. Lopez, Multi-Axial Fatigue Behavior of High-Strength Structural Bolts, J. Constr. Steel Res. 205 (2023) 107912.
[26] S. Zakir Sarothi, K. Sakil Ahmed, N. Imtiaz Khan, A. Ahmed, M.L. Nehdi, Machine Learning-Based Failure Mode Identification of Double Shear Bolted Connections in Structural Steel, Eng. Fail. Anal. 139 (2022) 106471.
[27] X. Deng, J. Liu, H. Gong, J. Huang, Detection of Loosening Angle for Mark Bolted Joints with Computer Vision and Geometric Imaging, Autom. Constr. 142 (2022) 104517.
[28] X. Yang, Y. Gao, C. Fang, Y. Zheng, W. Wang, Deep Learning-Based Bolt Loosening Detection for Wind Turbine Towers, Struct. Control Health Monit. 29 (2022).
[29] W.S. Na, Bolt Loosening Detection Using Impedance Based Non-Destructive Method and Probabilistic Neural Network Technique with Minimal Training Data, Eng. Struct. 226 (2021) 111228.
[30] Q.B. Ta, J.T. Kim, Monitoring of Corroded and Loosened Bolts in Steel Structures Via Deep Learning and Hough Transforms, Sensors 20 (2020).
[31] T.T. Nguyen, Q.B. Ta, D.D. Ho, J.T. Kim, T.C. Huynh, A Method for Automated Bolt-Loosening Monitoring and Assessment Using Impedance Technique and Deep Learning, Dev. Built Environ. 14 (2023).
[32] F. Wang, G. Song, A Novel Percussion-Based Method for Multi-Bolt Looseness Detection Using One-Dimensional Memory Augmented Convolutional Long Short-Term Memory Networks, Mech. Syst. Sig. Process. 161 (2021).
[33] T.C. Huynh, Vision-Based Autonomous Bolt-Looseness Detection Method for Splice Connections: Design, Lab-Scale Evaluation, and Field Application, Autom. Constr. 124 (2021).
[34] M.K. al-Bashiti, M.Z. Naser, Verifying Domain Knowledge and Theories on Fire-Induced Spalling of Concrete through Explainable Artificial Intelligence, Constr. Build. Mater. 348 (2022).
[35] A. Das, P. Rad, Opportunities and Challenges in Explainable Artificial Intelligence (XAI), A Survey, arXiv Prepr. arXiv2006.11371 (2020).
[36] S.L. Brunton, J.N. Kutz, Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control, Cambridge University Press, 2022.
[37] S. Meister, M. Wermes, J. Stuve, R.M. Groves, Cross-Evaluation of a Parallel Operating SVM-CNN Classifier for Reliable Internal Decision-Making Processes in Composite Inspection, J. Manuf. Syst. 60 (2021) 620–639.
[38] R. Miorelli, C. Fisher, A. Kulakovskyi, B. Chapuis, O. Mesnil, O. D'Almeida, Defect Sizing in Guided Wave Imaging Structural Health Monitoring Using Convolutional Neural Networks, NDT and E Int. 122 (2021).
[39] D. Kim, J. Lee, Predictive Evaluation of Spectrogram-Based Vehicle Sound Quality Via Data Augmentation and Explainable Artificial Intelligence: Image Color Adjustment with Brightness and Contrast, Mech. Syst. Sig. Process. 179 (2022).
[40] P. Pandey, A. Rai, M. Mitra, Explainable 1-D Convolutional Neural Network for Damage Detection Using Lamb Wave, Mech. Syst. Sig. Process. 164 (2022).
[41] H. Zhang, J. Lin, J.D. Hua, T. Tong, Interpretable Convolutional Sparse Coding Method of Lamb Waves for Damage Identification and Localization, Structural Health Monitoring-An, Int. J. (2021).
[42] S. Meister, M. Wermes, J. Stuve, R.M. Groves, Investigations on Explainable Artificial Intelligence Methods for the Deep Learning Classification of Fibre Layup Defect in the Automated Composite Manufacturing, Composites Part B-Engineering 224 (2021).

[43] V. Ewald, R.S. Venkat, A. Asokkumar, R. Benedictus, C. Boller, R.M. Grvoes, Perception Modelling by Invariant Representation of Deep Learning for Automated Structural Diagnostic in Aircraft Maintenance: A Study Case Using DeepSHM, Mech. Syst. Sig. Process. 165 (2022) 108153.
[44] L.C. Brito, G.A. Susto, J.N. Brito, M.A.V. Duarte, An Explainable Artificial Intelligence Approach for Unsupervised Fault Detection and Diagnosis in Rotating Machinery, Mech. Syst. Sig. Process. 163 (2022).
[45] J. Patterson, A. Gibson, Deep Learning: A Practitioner's Approach, O'Reilly, 2017.
[46] D.P. Kingma, L.J. Ba, Adam, A Method for Stochastic Optimization, International Conference on Learning Representations (ICLR) (2015) 1–15.
[47] K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, arXiv preprint arXiv: 1312.6034, (2013).
[48] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.R. Muller, W. Samek, On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation, PLoS One 10 (2015).
[49] G. Montavon, S. Lapuschkin, A. Binder, W. Samek, K.-R. Müller, Explaining Nonlinear Classification Decisions with Deep Taylor Decomposition, Pattern Recogn. 65 (2017) 211–222.
[50] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-Cam: Visual Explanations from Deep Networks Via Gradient-Based Localization, Int. J. Comput. Vis. 128 (2019) 336–359.
[51] J.T. Springenberg A. Dosovitskiy T. Brox M. Riedmiller Striving for Simplicity: The All Convolutional Net 2014 arXiv preprint arXiv:1412.6806.
[52] M.P. Hu N. Yue R.M. Groves Damage Classification of a Bolted Connection Using Guided Waves and Explainable Artificial Intelligence 2023 London.
[53] G. Vilone, L. Longo, Notions of Explainability and Evaluation Approaches for Explainable Artificial Intelligence, Inf. Fusion 76 (2021) 89–106.
[54] C.K. Yeh, C.Y. Hsieh, A. Suggala, D.I. Inouye, P.K. Ravikumar, On the (in)Fidelity and Sensitivity of Explanations, Adv. Neural Inf. Proces. Syst. 32 (2019).
[55] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, B. Kim, Sanity Checks for Saliency Maps, Adv. Neural Inf. Proces. Syst. 31 (2018).
[56] Aluminum Association (2000).
[57] E. Oberg, F.D. Jones, H.L. Horton, H.H. Ryffel, Machinery's Handbook, 20th Edition,, Industrial Press Inc, 1976, pp. 1446–1447.
[58] C. Humer, S. Holl, C. Kralovec, M. Schagerl, Damage Identification Using Wave Damage Interaction Coefficients Predicted by Deep Neural Networks, Ultrasonics 124 (2022).
[59] N. Yue, M.H. Aliabadi, A Scalable Data-Driven Approach to Temperature Baseline Reconstruction for Guided Wave Structural Health Monitoring of Anisotropic Carbon-Fibre-Reinforced Polymer Structures, Struct. Health Monit. 19 (2020) 1487–1506.
[60] B. Lee, G. Manson, W. Staszewski, Environmental Effects on Lamb Wave Responses from Piezoceramic Sensors, Materials Science Forum, Trans Tech Publ 440 (2003) 195–202.