Damage identification using ultrasonic Lamb waves with multi-scale adaptive attention Transformer-based unsupervised domain adaptation

Houyu Lu[a,*], Sergio Cantero Chinchilla[b], Brecht Rotsaert[c], Anthony Croxford[b], Konstantinos Gryllias[d,e], Dimitrios Chronopoulos[a]

[a]*Department of Mechanical Engineering & Division of Mechatronic System Dynamics (LMSD), KU Leuven, Ghent Campus, 9000, Belgium*
[b]*School of Electrical, Electronic, and Mechanical Engineering, University of Bristol, Bristol, BS8 1TR, UK*
[c]*Department of Mechanical Engineering, KU Leuven, Ghent Campus, 9000, Belgium*
[d]*Department of Mechanical Engineering, Division LMSD, KU Leuven, Celestijnenlaan 300, Box 2420, Leuven, 3001, Belgium*
[e]*Flanders Make@KU Leuven, Belgium*

Abstract

In this paper, a novel unsupervised transfer learning method – the Multi-scale Adaptive Attention Transformer Domain Adaptation Network (MAAT-DAN) is proposed for structural damage identification in regression and classification tasks. MAAT-DAN comprises three main components: a feature extractor, a domain adapter, and a label predictor. The feature extractor, for the first time, introduces a novel dynamic adaptive weighting module that enhances multi-head attention by adjusting attention head weights based on the input. Additionally, a multi-scale feature extraction technique is employed to simultaneously extract damage features at three different scales from Lamb waves, a popular means in structural health monitoring. The label predictor is composed of multiple fully connected layers. MAAT-DAN's domain adapter integrates a gradient reversal layer for managing high-dimensional data, maximum mean discrepancy for improved stability, and feature reconstruction techniques to support effective domain adaptation. MAAT-DAN's performance is validated in both regression and classification tasks using two specially designed ultrasonic experiments: (1) for regression, an experiment on a reinforced aluminum plate using test beams of two different materials and Blu-Tack to simulate repeatable damage; (2) for classification, an experiment on a steel water tank deploying two pairs of sensors to collect data over two years. Its feature extractors and domain adapters are compared across customized metrics against five other methods using these datasets. Results show that MAAT-DAN achieved the highest prediction accuracy and consistent prediction stability in both tasks, demonstrating its effectiveness in unsupervised domain adaptation for damage identification.

Keywords: Unsupervised transfer learning, Domain adaptation, Ultrasonic Lamb waves, Multi-scale adaptive attention Transformer, Structural health monitoring

1. Introduction

Monitoring and interpreting changes in engineering structure performance has long been a critical research field. From airplanes, pipes, and bridges to tidal turbines, industries place great importance on Structural Health Monitoring (SHM), often involving permanently attached sensors and systems to continuously assess its integrity [1, 2]. These measures aim to prevent or mitigate unexpected structural failures that could arise from manufacturing flaws, harsh environmental conditions, fatigue, or overload degradation [3]. The significance of SHM is multifaceted. It not only ensures structural safety but also substantially reduces maintenance expenses. Furthermore, by identifying and addressing potential issues early, SHM effectively extends the lifespan of engineered structures. This proactive approach to management has made SHM an indispensable component of modern engineering maintenance practices [4, 5].

Among various monitoring techniques (such as acoustic emission, vibration analysis, thermal imaging, etc.), damage detection guided wave-based methods have emerged as particularly promising solutions for SHM. Lamb waves are a type of ultrasonic guided wave, and they have been proven effective for SHM in thin-walled structures [6, 7]. Their ability to travel long distances, coupled with minimal equipment requirements and high

*Corresponding author
Email address: houyu.lu@kuleuven.be (Houyu Lu)

sensitivity to defects, makes them a cost-efficient choice for SHM applications [8]. The propagation characteristics of guided waves in a structural waveguide can reveal valuable information about the state of the structure, including the presence and location of potential defects. In guided wave-based SHM, wave propagation is highly sensitive to the physical properties of the material. Structural damage—such as cracks, delamination, corrosion, and holes—changes the local properties like elastic modulus, density, and thickness, thereby affecting wave velocity and propagation behavior. The effects of damage in guided waves can be summarised in: (1) Reflection and scattering effects: When Lamb waves interact with damage, a part of the signal is reflected or scattered. These effects cause amplitude reduction, phase shifts, or the generation of additional reflected signals. Reflection and scattering intensity are influenced by the damage size, shape, and location, making them key indicators for damage identification [2]. (2) Group and phase velocity effects: Damage changes local stiffness and density, which consequently reduces Lamb wave propagation velocity. By measuring wave travel time or phase variations at different detection points, quantitative analysis of velocity changes enables damage localization [9]. (3) Energy attenuation effects: The damaged area leads to wave energy absorption, resulting in significant attenuation in the received signal. This energy loss can serve as an indicator for assessing damage severity or evaluating overall structural health [7].

Some traditional physics-based methods can be used to detect and assess damage presence [6, 10, 11]. However, traditional wave-physics damage detection methods face challenges when applied to complex structures in the real world [12]. Multiple wave reflections and mode conversions create intricate patterns that are difficult to interpret accurately. Overlapping and interfering wave modes in received signals make damage detection and localization less reliable [12, 13]. To complement the capabilities of purely physics-based approaches, researchers have increasingly adopted data-driven methods to enhance the robustness and generalizability of models in damage detection [12, 14, 15]. These methods leverage the power of large-scale sensor data to capture structural degradation patterns. By employing sophisticated feature extraction techniques and machine learning algorithms, data-driven approaches can establish correlations between extracted features and damage, irrespective of their immediate physical interpretability [16]. Unlike traditional physics-based methods, which rely on explicit mathematical models of structural behavior [17], data-driven approaches offer the advantage of learning complex relationships directly from observed data. This capability allows them to potentially uncover subtle damage indicators that might be overlooked by conventional analysis [18]. It's worth mentioning that the attention mechanism, an emerging data-driven approach in SHM, has recently gained traction due to its impressive results in natural language processing [19–23]. The Attention mechanism emulates human cognitive focus, allowing models to dynamically prioritize task relevant input features. This approach enhances task-specific sensitive feature processing, potentially improving model performance in complex scenarios [24, 25]. The Transformer-based architecture in the attention mechanism has shown notable feature extraction capabilities and holds significant potential for extracting damage-sensitive features in SHM applications [19]. However, data-driven methods in structural health monitoring, while promising, face challenges in practical applications. The effectiveness of these approaches is often limited by the availability of comprehensive datasets that represent various damage scenarios [26]. A critical limitation lies in the substantial data requirements for building robust and reliable models. Collecting experimental data across diverse damage conditions is typically a time-consuming and resource-intensive process and potentially limits the use to a specific structure.

To overcome the challenges of limited datasets, transfer learning has gained popularity in recent years [27–30]. Transfer learning facilitates the transfer of knowledge acquired in the source domain to address challenges in a related but distinct target domain. Using similarities between domains enables the application of learned concepts, features, models, or techniques to novel contexts, potentially improving problem-solving efficiency and effectiveness in the target domain. This cross-domain knowledge utilization can be particularly valuable when direct experience or data in the target domain are limited or difficult to obtain [31]. However, the labeled data are time-consuming and expensive to obtain in industrial applications. The newly collected data or the data under different working conditions are usually unlabeled, and it is sometimes very difficult or even impossible to label these data [3]. Unsupervised Transfer Learning (UTL) has been developed to further address the challenge of limited labeled data in real-world application scenarios.

UTL is the case where the source domain has large labeled data available, while the target domain lacks or has no labeled data. For instance, in the SHM task, the source domain may have a comprehensive database of ultrasonic signal patterns corresponding to different types of damage in metal aircraft components. However, the target domain could consist of composite wind turbine blades, for which labeled ultrasonic inspection data is insufficient. In the case of an unlabeled target domain, the main strategy has been to guide feature learning by minimizing the difference between the source and target feature distributions [32]. Maximum Mean Discrepancy (MMD) is one of the popular statistical feature transformation methods [33]. MMD quantifies the difference between the means of two probability distributions by computing the norm of the maximum difference between the expected values of functions sampled from the two distributions [32]. In UTL, MMD measures how

different the average features are between the source and target domains. It provides a way to align feature spaces across domains. Deep domain confusion [34] incorporated MMD into the network's loss function, combining it with the standard classification loss on source data. This approach aimed to learn representations that are both discriminative and domain-invariant. Central moment discrepancy [35] expanded the idea to match higher-order moments between domains. This method offered a more comprehensive comparison of distributions beyond just their means. The Deep Adaptation Network (DAN) [36] took a different approach by applying MMD to layers embedded in a reproducing kernel Hilbert space(a type of high-dimensional feature space that allows efficient computation of distances between distributions). This technique effectively matched higher-order statistics of the source and target distributions, potentially capturing more nuanced differences between domains. Domain-adversarial transfer learning is one of the most popular transfer learning methods to minimize domain shift [33, 37, 38]. This method leverages an adversarial loss function to learn a feature representation that is both discriminative for the source domain labels and invariant across domains. By simultaneously optimizing for task performance and domain indistinguishability, this technique aims to improve the generalization of models to new, unseen target domains [32, 39]. Domain-Adversarial Neural Network (DANN) is a famous domain adversarial method [40]. It trains neural networks to handle data from different distributions by learning a domain-invariant feature representation. DANN consists of three key components: a feature extractor, a label predictor, and a domain classifier. It uses a Gradient Reversal Layer (GRL) during backpropagation to encourage the network to produce features that cannot differentiate between the source and target domains.

Most Lamb wave-based transfer learning SHM approaches currently focus on supervised or semi-supervised methods on classification problems [1, 41, 42]. Shao et al. [1] presents a multi-level damage classification technique for aircraft plate structures, using a Lamb wave-based supervised deep transfer learning network. Zhang et al. [42] proposes a supervised cross-SHM method using joint distribution adaptation and convolutional long short-term memory deep transfer learning and guided waves, aiming to generalize SHM models across different structures. Hong et al. [43] used convolutional neural networks and feature Domain Adaptation (DA) to introduce a deep cross-domain adaptive semi-supervised damage identification method for SHM of stranded wires under varying vibration conditions. And most of the UTL SHM classification algorithms use vibration and acoustic emission signals [38, 44, 45], the application of UTL methods in SHM using Lamb waves remains limited. Numerous researchers have made appreciable progress in addressing the DA regression problem in the field of computer science, contributing both theoretical insights and novel algorithmic approaches [46–49]. Taghiyarrenani et al. [47] developed a supervised multi-domain adaptation method that establishes a common feature space, enabling linear regression models to generalize effectively across multiple domains. De Mathelin et al. [48] present an instance-based approach by adequately reweighting the source instances to handle regression tasks in the context of supervised domain adaptation under an assumption of covariate shift. Cortes et al. [50] introduced theoretical foundations, algorithmic approaches, and empirical findings in the areas of domain adaptation and sample bias correction for regression problems. Although these advancements have gradually begun to find applications in engineering contexts where SHM is widely implemented, the use of Lamb waves for UTL in both classification and regression analyses within the SHM field remains relatively underexplored.

In this paper, to reliably extract damage features from ultrasonic signals, address detection challenges associated with insufficient or unlabeled data, and enable rapid adaptation to new situations, tasks, or environments, a novel Multi-scale Adaptive Attention Transformer Domain Adaptation Network (MAAT-DAN) is proposed for the first time. This method is based on the transformer architecture, introduces adaptive weight heads and multi-scale feature extraction techniques, and combines two DA techniques: GRL and MMD. This network employs two types of experimental Lamb wave datasets for both UTL regression and classification tasks in SHM. Specifically, it focuses on damage localization through regression and damage identification through classification. The multi-scale mechanism enables ultrasonic-based damage diagnosis to extract more comprehensive signal features across different scales. Additionally, in the context of SHM, the proposed MAAT-DAN UTL methods can greatly benefit from the powerful feature extraction capabilities of the Transformer architecture-based attention mechanism. The adaptive mechanism further enhances this by dynamically adjusting the weight of each attention head based on the input data. For DA, the proposed network employs a combination of MMD and the GRL to reduce distributional differences between the source and target domains, while simultaneously learning domain-invariant features. This improves the ability of the model to generalize across different domains. The overview of the proposed MAAT-DAN is shown in Fig. 1. The main contributions of this paper can be summarized as follows:

- Proposed is a novel UTL method MAAT-DAN for structural damage identification in both regression and classification tasks based on Lamb wave signals.

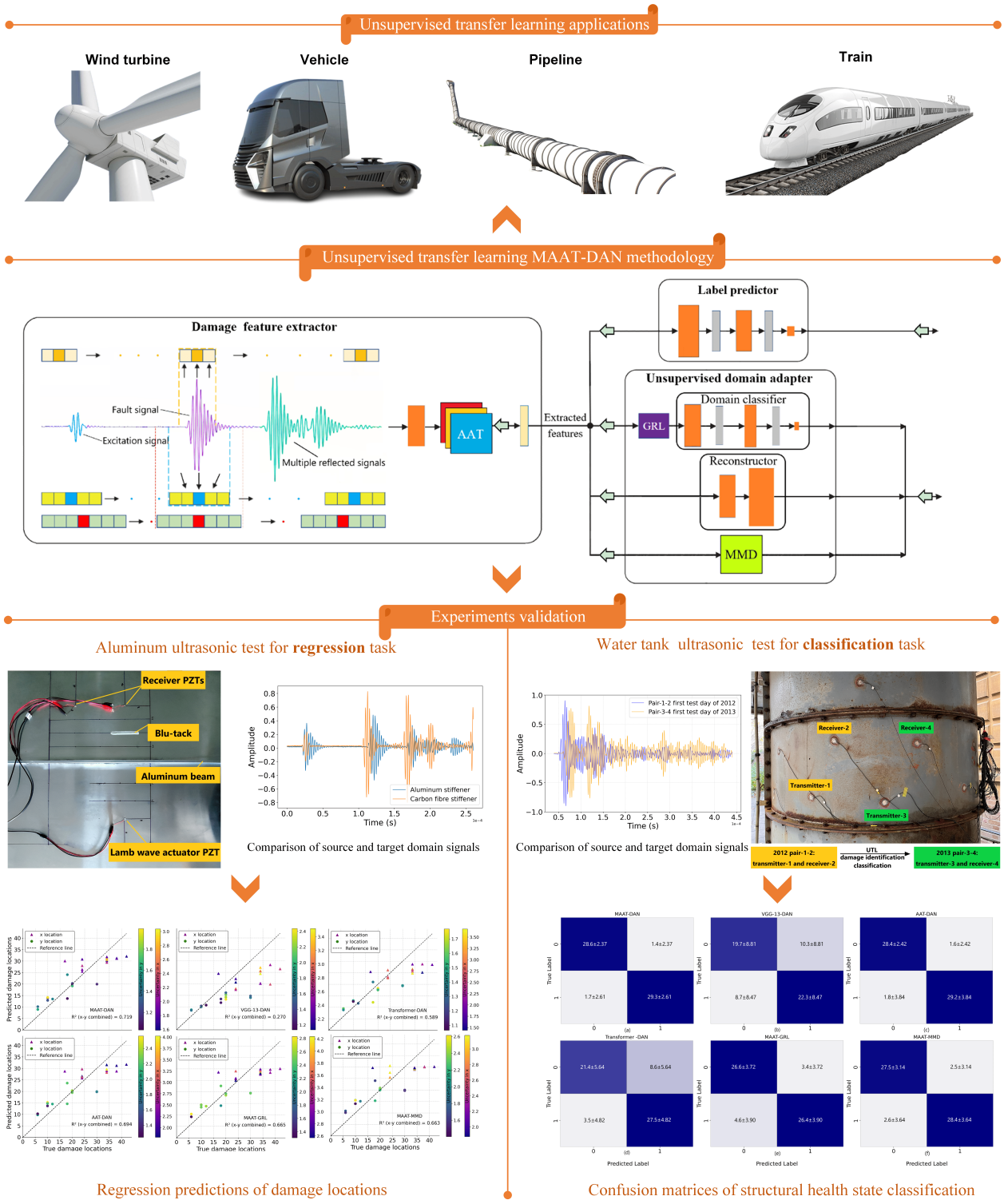


Figure 1: Overview of the proposed methodology for damage identification based on two ultrasonic tests: comparing five DA methods in both regression and classification applications.

- Developed multi-scale and adaptive weight heads attention mechanism within the Transformer architecture as a feature extractor to extract features effectively.
- Integrated GRL and MMD as a domain adapter enable the model to simultaneously learn domain-invariant features and align domain distributions.
- Implemented cross-material (from aluminum to carbon-fibre-reinforced plates) UTL regression and cross-year/sensor-pair (from 2012 to 2013 sensor pairs) UTL classification for damage identification, using two

ultrasonic experimental datasets on reinforced aluminum plates and steel water tanks.

The remainder of the paper is organized as follows: Section 2 provides a detailed description of the proposed MAAT-DAN model, including its key components, training optimization strategies, and overall architecture. Section 3 outlines the two specially designed ultrasonic experiments for both regression and classification task validation. Section 4 compares the proposed method performance with five other approaches using ultrasonic experimental data in regression and classification tasks. Finally, Section 5 presents concluding remarks and discusses future development directions.

2. Proposed Methodology

The proposed MAAT-DAN framework aims to address the unsupervised DA problem using Lamb wave data, in order to improve the damage identification performance on the target domain. The MAAT-DAN framework consists of three main components: damage feature extractor, domain adapter, and label predictor. The label predictor's architecture, consisting of three fully connected layers interspersed with two dropout layers, offers a streamlined approach to label prediction. The MAAT-DAN framework emphasizes precise feature extraction and DA, which are crucial to accurate damage identification. Its architecture supports rapid inference by leveraging robust, domain-invariant damage-sensitive features and aligning domain distributions, accomplished through the feature extractor and domain adapter components. The following subsections will detail the feature extractor and domain adapter components.

2.1. Damage feature extractor of MAAT-DAN: multi-scale adaptive attention Transformer

Feature extraction is crucial for the effectiveness of training in UTL. Effective feature extraction is the foundation upon which the DA algorithm can fully function, ultimately enabling successful task output in the target domain. In UTL, the absence of labels in the target domain makes the abstract features learned by the network from raw data at various levels critical to the success of transfer learning.

2.1.1. Adaptive attention Transformer

The Transformer architecture, introduced by Vaswani et al. [19], represents a significant departure from traditional recurrent and Convolutional Neural Networks (CNN) and marks a pivotal moment in the evolution of deep learning models [51]. This model relies exclusively on self-attention mechanisms to process input and output representations, making it highly effective at handling long-range dependencies within the data. At its core, the Transformer consists of two key components: a position-wise Feed-Forward Network (FFN) and a multi-head attention layer. The FFN processes each input element independently while preserving their sequential relationships. Meanwhile, the Transformer's multi-head attention mechanism allows the model to focus on different parts of the input sequence simultaneously, enhancing its ability to capture complex relationships between sequence elements from different perspectives and abstraction levels [52]. Building on this, the Transformer's attention mechanism employs three crucial components—Query (Q), Key (K), and Value (V)—derived from the input representations using learned weight matrices. This architecture not only facilitates greater parallelization, but can also, through its attention mechanism, efficiently establish long-range dependencies. This addresses a key limitation of traditional deep learning architectures which, due to their sequential processing nature or local receptive field constraints, struggle to capture relationships between distant elements in the input data [51]. In the Transformer's multi-head attention mechanism, each head learns and attends to distinct information by performing attention over different learned projections of the Q, K, and V. However, no explicit weights are applied to differentiate the heads before they are concatenated into the linear projection layer. Each attention head is processed in parallel and treated equally in terms of computation steps. Even though the subsequent linear layer can implicitly adjust the influence of each head through static weights, the design does not dynamically adjust the weights of each head based on input data during inference. In other words, before concatenation, the outputs of each head are not explicitly assigned different weights. The relative influence between the heads is not explicitly differentiated at this stage. While the weight matrix of the linear layer can implicitly impact the final influence of each head, these weights are not dynamically adjusted based on the specific content of each head's output. To address this limitation, a dynamic weighting mechanism is introduced. This mechanism would involve a trainable module that adaptively assigns weights to each head based on the current input context. This adaptive weighting allows the model to dynamically adjust the output of each attention head, making it more flexible in capturing relevant features. By controlling the influence (weights) of each head, the model highlights the important ones while downplaying those that are less relevant. This enables the model to theoretically capture key features more accurately, potentially improving performance in specific tasks.

Specifically, the MAAT-DAN adjusts the attention head weights using the Q, which provides a summary representation of the input sequence features. This method generates adaptive weights based on the global information from the Q vector, determining which heads to emphasize or downplay, thereby dynamically adjusting their influence and allowing for adaptive feature adjustment. In the specific implementation process, queries are averaged over all sequence positions (the entire sequence length) and the embedding dimensions (d_{model}/number of heads). This results in an average value for each attention head across all input samples, reflecting the overall "attention level" or "activity level" of the attention head. These averages are then used to compute adaptive weights through a fully connected layer, dynamically adjusting the influence of each attention head on the input dataset. These adaptive weights theoretically help the model prioritize the most important attention heads for the current query more effectively, enhancing the model's interpretability, and flexibility.

Fig. 2 illustrates the position and method of implementation of the adaptive weights module within the original Transformer architecture used in MAAT-DAN. In this paper, the dimensions of the MAAT's embeddings and hidden representations are $d_{model} = 64$, and the inner layer in the FFN has a dimension of $d_{ff} = 256$ [19]. Each multi-head adaptive attention mechanism consists of four heads (h=4), and the Adaptive Attention Transformer (AAT) comprises three layers (N=3) (note that these above four hyperparameters may require slight adjustments depending on the specific dataset). These layers are progressively connected, with the output of each layer serving as the input for the next, allowing the model to gradually learn more complex feature representations. Note that, unlike the traditional transformer architecture, the training data in this paper consists of ultrasonic data with uniform lengths, eliminating the need for masking during the training process. Furthermore, since ultrasonic signals inherently encode sequential information due to the data's natural ordering, the signals' relative positional relationships are implicitly captured through the sampling frequency. MAAT-DAN does not incorporate positional encoding for the signal data to reduce computational overhead and improve training efficiency.

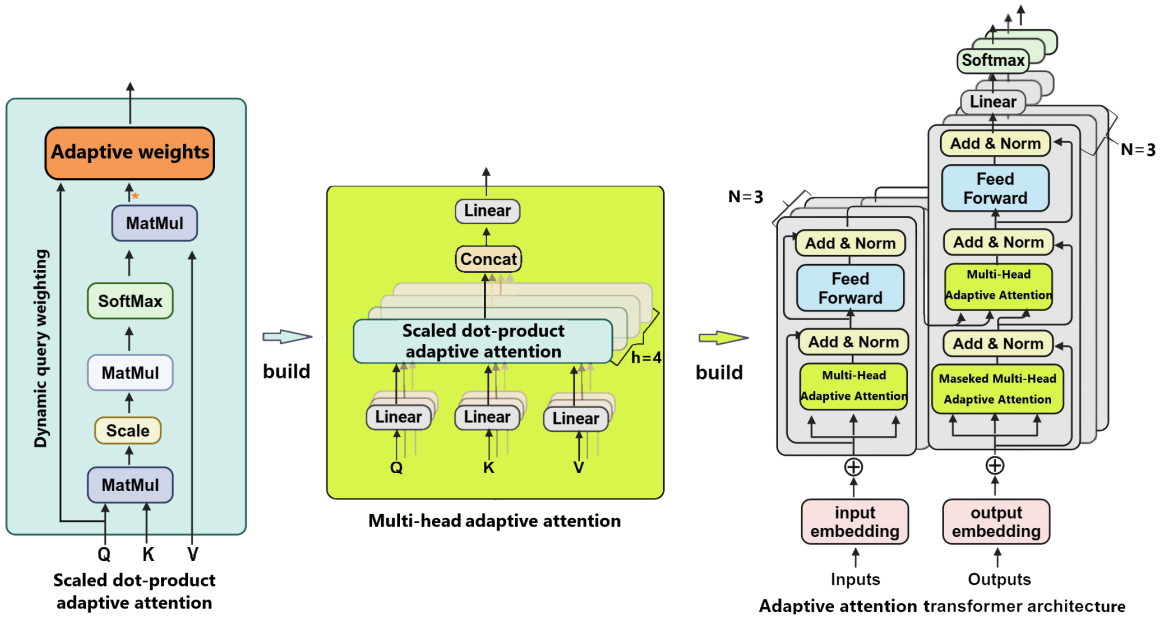


Figure 2: The AAT architecture in detail: Scaled dot-product adaptive attention (left); Multi-head adaptive attention with 4 heads (center); Transformer architecture with stack of 3 identical layers (right).

2.1.2. Multi-scale approach

In most engineering applications of Lamb waves for SHM, the multi-scale properties of Lamb waves are often not fully utilized. The multi-scale character of ultrasonic waves refers to the signal containing distinct features and information across different time or space scales. This multi-scale approach to signal analysis is highly effective for peak detection, especially in noisy environments [53, 54]. In deep learning and signal processing, the multi-scale approach can simultaneously capture the signal's global structure and local details, improving feature extraction and pattern recognition [55–57]. The propagation of Lamb waves within structures spans various spatial scales, from small local feature changes to the overall waveform structure. In terms of time scales, rapid signal variations (such as peaks and troughs) provide high-resolution details, while slower changes reveal global trends at lower time resolutions. A multi-scale signal contains multi-level information, ranging from rough to fine and from global to local. Additionally, these patterns carry different information across

various frequency scales. By applying attention mechanisms at multi-scales, the model can extract features at varying levels of detail. For instance, they capture local details on high-resolution scales and global structural features on low-resolution scales. This approach enhances the model's ability to adapt to defects across different scales, and reduces reliance on any specific scale, theoretically improving the accuracy of SHM [58].

Based on this, the paper introduces the multi-scale concept into the adaptive multi-head attention mechanism in the Transformer architecture. By applying attention layers across different scales, the model captures feature relationships at multiple levels, potentially enhancing the representation and performance of ultrasonic wave-based models. This paper employs three different scales for feature extraction: down-sampling, original, and up-sampling. In application in MAAT-DAN, the down-sampled feature dimension is half of the original scale, allowing the model to capture broader, more abstract features of the Lamb wave signals, reduce computational costs, and potentially focus on broader patterns [59, 60]. Conversely, the up-sampled feature dimension is increased to twice that of the original scale, allowing the model to better capture fine local features and strengthen important features of the Lamb wave signals [61, 62]. The concept of multi-scale sampling, along with its application in Lamb wave feature extraction and AAT are shown in Fig. 3.

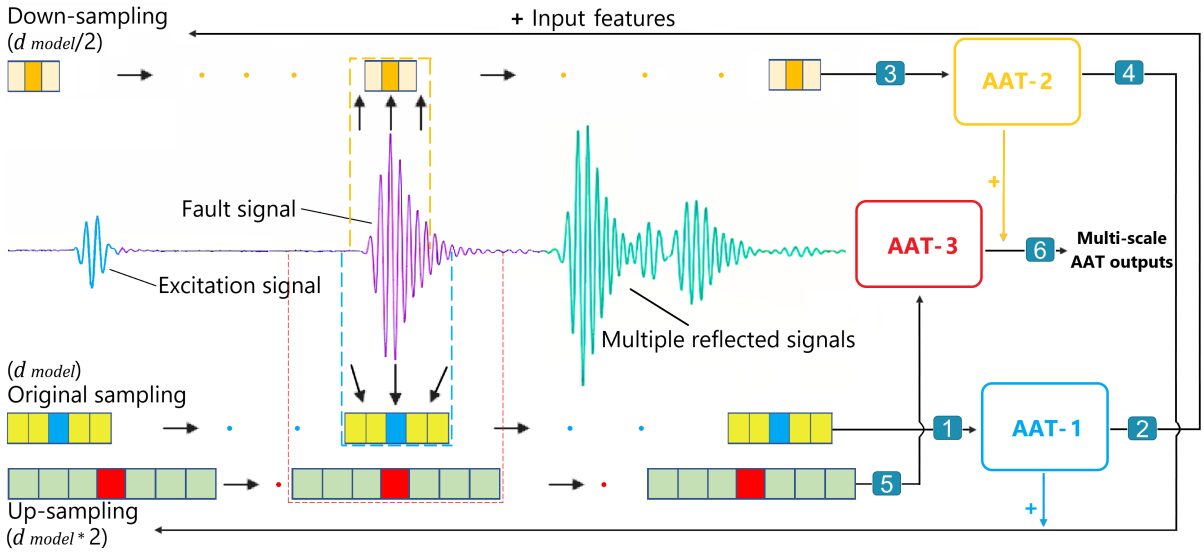


Figure 3: The schematic diagram of multi-scale sampling of Lamb wave signal in AAT: original (blue), down-sampling (orange), and up-sampling (red) techniques. The three colored dots represent the omission of sampling at different scales.

In particular, the MAAT feature extraction procedures based on Fig. 3 are described as follows:

Step 1: the raw input dataset initially undergoes feature extraction through a complete AAT-1 architecture, using the original sampling method. The resulting features are denoted as AAT-O;

Step 2: AAT-O is then subject to dropout and combined with the input dataset via the residual connection. This merger ensures comprehensive information transfer and mitigates the vanishing gradient problem [63]. Layer normalization is subsequently applied to stabilize the merged features, standardizing the input features to have a mean of zero and a variance of one, resulting in the original scale feature output, named Out-1;

Step 3: Out-1 is then downsampled for dimensionality reduction before being processed by the AAT-2 for further feature extraction, producing downsampled feature representation, named AAT-Down;

Step 4: AAT-Down undergoes dropout, residual connection with Out-1, and normalization, resulting in feature output after downsampling, named Out-2;

Step 5: Out-2 is upsampled to expand its dimensionality and fed into the AAT-3 for additional feature extraction, generating upsampled feature representation, named AAT-Up;

Step 6: AAT-Up is subjected to dropout, residual connection with Out-2, and layer normalization, resulting in multi-scale AAT outputs after up-sampling, named Out-3;

After the multi-scale feature extraction, as shown in Fig. 4, Out-3 is fed into an FFN consisting of two dense layers, using the computationally efficient ReLU activation to introduce non-linearity. This step enables further nonlinear transformations and deep feature fusion of the multi-scale features, resulting in Out-FFN. This output then goes through dropout, a residual connection with Out-3, and normalization. Note that this research employed four dropout layers in the multi-scale feature extraction process, each set to a ratio of 0.1. This value, which is empirically obtained, achieves a balance between feature learning capacity and generalization ability, mitigating the risk of critical feature loss while enhancing the model's adaptability to varying data distributions. However, the optimal dropout ratio depends on the specific dataset characteristics

and task requirements and should be determined through hyperparameter tuning and practical validation. This paper recommends selecting an optimal dropout ratio of 0.1 to 0.3 for the proposed model based on practical evaluation to optimize performance under the given experimental conditions [64]. In addition, during the multi-scale feature extraction process, four residual connections are incorporated at different steps to enhance feature consistency and gradient propagation. Residual connections are implemented at each scale by applying element-wise addition between the output of the previous layer and the AAT-processed features at the current scale. This facilitates effective information retention and mitigates potential feature degradation during multi-scale feature transformations [65]. The result is the final set of features extracted by the multi-scale AAT. Fig. 4 provides a detailed overview of the entire feature extraction workflow carried out by the feature extractor MAAT.

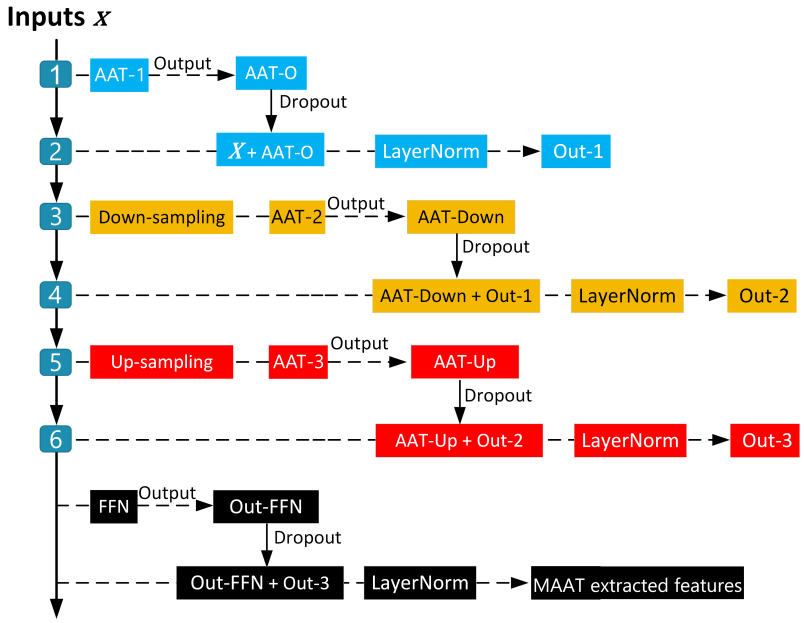


Figure 4: Overall feature extraction process by feature extractor MAAT

2.2. Domain adapter of MAAT-DAN: Joint domain-adversarial and discrepancy-based methods

DA addresses the challenge of training a classifier or predictor when there is a discrepancy between the distributions of the training dataset and the test dataset [40]. DA approaches establish a correspondence between source and target domains, enabling the transfer of discriminative features learned from the source domain to new, related domains. This approach is particularly valuable in scenarios with unlabeled or sparsely labeled target data, facilitating both unsupervised and semi-supervised adaptation. The MAAT-DAN framework integrates two DA methods: MMD and GRL. MMD provides an explicit measure to align the feature distributions between the source and target domains, while GRL implicitly forces the feature extractor to learn domain-invariant features by reversing gradients, theoretically ensuring more robust DA. This dual-DA mechanism not only directly aligns the distributions of the source and target domains but also strengthens the learning of domain-invariant features through adversarial training.

The dual-DA mechanism addresses the limitations of using GRL alone, such as instability in adversarial learning or difficulty in capturing subtle distribution differences [66–68]. GRL induces adversarial training between the feature extractor and the domain classifier by reversing the gradient. The former learns to produce domain-invariant features through gradient reversal, while the latter aims to correctly identify both domains' data. This adversarial process may result in parameter oscillation or training instability [66]. Moreover, when cross-domain discrepancies are subtle and complex, relying only on gradient reversal may not produce sufficiently strong adversarial signals, thus limiting domain alignment effectiveness [32]. A more direct statistical discrepancy measure MMD is needed for stable domain alignment. Meanwhile, the GRL method complements MMD by addressing the challenges of MMD-based DA, particularly when kernel methods struggle to capture all distribution differences. This issue becomes more pronounced with complex distributions or high-dimensional data, especially when dealing with higher-order statistical information [69, 70]. A detailed description of the two DA techniques is provided as follows.

(1) **GRL**: The GRL, originally introduced in DANN [40], a classic DA method, was previously used to enable adversarial training by reversing the gradients of the domain classifier during backpropagation. This allows the

feature extractor to learn domain-invariant feature representations, thereby facilitating DA. GRL plays a critical role in learning domain-invariant features, enabling effective DA. It is a parameter-free component that functions as an identity transformation during forward propagation. During forward propagation, the GRL has no effect on the input features. However, in the backpropagation phase, it inverts the gradient by multiplying it by -1 before passing it to the previous layer [40], effectively reversing the direction of the update. The GRL in this paper is represented by the function $f(X)$, which can be defined below:

$$f(\mathbf{X}) = \mathbf{X}, \tag{1}$$

$$\frac{\mathrm{d}f}{\mathrm{d}\mathbf{X}} = -\mathbf{I}, \tag{2}$$

where $\mathbf{I}$ is an identity matrix, $\mathbf{X}$ is the input vector.

(2) **MMD**: Another classic DA method is the statistical discrepancy-based MMD. The MMD approach aligns the source and target domain distributions by minimizing the statistical distance between their feature distributions [71]. In statistical terms, this refers to calculating the difference in the means of data points from two probability distributions after they are mapped into another space [33]. Specifically, for a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}_k$ associated with a kernel k, the population statistic approach MMD can be expressed as an integral probability metric as follows:

$$\mathrm{MMD}_{\mathcal{H}_\mathrm{k}}(p, q) = \sup_{f \in \mathcal{H}_k : \|f\|_{\mathcal{H}_k} \leq 1} \left| \mathbb{E}_{X \sim p}[f(X)] - \mathbb{E}_{Y \sim q}[f(Y)] \right|, \tag{3}$$

where the $\sup \cdot$ is the supremum of the input aggregate, and X and Y represent samples drawn from distributions p and q, respectively. The terms $\mathbb{E}_{X \sim p}[f(X)]$ and $\mathbb{E}_{Y \sim q}[f(Y)]$, often referred to as "kernel mean maps", represent the kernel mean embeddings of the distributions p and q into the RKHS corresponding to the positive-definite kernel k. The focus lies on a specific kernel k, which ensures that $\mathrm{MMD}_{\mathcal{H}_\mathrm{k}}(p, q) = 0$ if and only if the two distributions $p = q$ [72, 73].

2.3. Loss function and optimization strategy of MAAT-DAN

The MAAT-DAN framework utilizes multiple loss functions for joint optimization, primarily divided into task loss and total domain loss. They are used to monitor the performance of task label prediction and DA, respectively.

Task loss: To verify the performance of MAAT-DAN on both classification and regression tasks, MAAT-DAN employs two separate task-specific loss functions.

For the regression task loss, the Huber loss function is used to measure the discrepancy between the predicted values and the true labels for the regression task loss. The Huber loss function is a combination of Mean Squared Error (MSE) and Mean Absolute Error (MAE) loss function, selecting the loss type based on a threshold that controls sensitivity to outliers. By combining the strengths of MSE and MAE, the Huber loss is more robust to outliers [74, 75]. The regression task loss $L_{\text{task-R}}$ using the Huber can be defined as follows:

$$L_{\text{task-R}} = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{if } |y - \hat{y}| \leq \delta \\ \delta \cdot \left(|y - \hat{y}| - \frac{1}{2}\delta\right) & \text{otherwise} \end{cases}, \tag{4}$$

where y is the true value, $\hat{y}$ is the predicted value, and δ is the threshold that controls outlier sensitivity. The Huber loss function behaves like MSE when $|y - \hat{y}|$ is less than or equal to δ, providing a smooth quadratic nature for small errors. However, when $|y - \hat{y}|$ exceeds δ, it transitions to a linear loss, similar to MAE, meaning it is less influenced by extreme values compared to MSE.

For the classification task loss, sparse categorical cross-entropy is used to calculate the negative logarithm of the probability that each sample is predicted as its true class. This value is averaged over the entire dataset to quantify the discrepancy between model predictions and true labels. The classification task loss $L_{\text{task-C}}$ by using the sparse categorical cross-entropy can be defined as follows:

$$L_{\text{task-C}} = -\frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \log p_{i,y_i}, \tag{5}$$

where $\mathcal{N}$ represents the total number of samples, and p_{i,y_i} denotes the probability that the model predicts the i sample as its true class y_i.

Total domain loss: The total domain loss of MAAT-DAN consists of three key loss terms: domain loss, MMD loss, and reconstruction loss, all aimed at improving the model's DA capability and feature extraction effectiveness:

(1) Domain loss: A binary cross-entropy loss function is used to classify the source and target domain data. Source domain data is assigned label 0, while target domain data is assigned label 1, training the model to differentiate domain features. During backpropagation, using GRL, the model learns generalized, domain-invariant features, enhancing its ability to adapt across domains. In the domain classifier, each source domain sample is labeled 1, and each target domain sample is labeled 0. Based on this, the domain loss can be defined as follows [32, 40]:

$$L_{\text{domain}} = -\frac{1}{n_s + n_t}\left(\sum_{i=1}^{n_s}\left[\log\left(D\left(f_s^i\right)\right)\right] + \sum_{j=1}^{n_t}\left[\log\left(1 - D\left(f_t^j\right)\right)\right]\right), \tag{6}$$

where, $D\left(f_s^i\right)$ represents the predicted probability of domain classification for the i-th sample in the source domain with n_s labeled examples. $D\left(f_t^j\right)$ represents the predicted probability for the j-th sample in the target domain with n_t unlabeled examples. f_s^i and f_t^j are the feature representations of the i-th and j-th samples from the source and target domains, respectively.

(2) MMD loss: As mentioned in Section 2.2, MMD loss is used to minimize the statistical differences between the feature distributions of the source and target domains, promoting the alignment of their feature spaces. If the MMD value is close to 0, it indicates that the two distributions are highly similar; otherwise, it suggests a difference between them. The approach employs a Gaussian kernel for its non-linear smoothing properties in calculating the MMD loss within the MAAT-DAN framework. The formula for calculating MMD in the MAAT-DAN framework is as follows [76]:

$$L_{\text{MMD}}^2 = \frac{1}{n_s^2}\sum_{i,j} k\left(f_s^i, f_s^j\right) - \frac{2}{n_s n_t}\sum_{i,j} k\left(f_s^i, f_t^j\right) + \frac{1}{n_t^2}\sum_{i,j} k\left(f_t^i, f_t^j\right), \tag{7}$$

where the first term represents the similarity between source domain samples; the second term represents the similarity between source and target domain samples; and the third term represents the similarity between target domain samples. $k(\cdot,\cdot)$ denotes the Gaussian kernel function with a bandwidth parameter set to 1, which measures the similarity between feature representations.

(3) Reconstruction loss: The MSE between the original data and the reconstructed data (i.e., the data after passing through the model's feature extraction) from both domains is computed. Theoretically, if the data undergo effective feature extraction, the data containing key features should have minimal differences from the original data. By minimizing the difference between the original and reconstructed data, the model is encouraged to learn feature representations that work effectively for both domains, ensuring that these features represent the input data well in both the source and target domains. This helps reduce inter-domain differences while preserving critical feature information, improving the effectiveness of cross-domain transfer [77]. The reconstruction loss can be defined as follows:

$$L_{\text{recon}} = \frac{1}{n_s + n_t}\left(\sum_{i=1}^{n_s}\|x_s^i - \hat{x}_s^i\|^2 + \sum_{j=1}^{n_t}\|x_t^j - \hat{x}_t^j\|^2\right), \tag{8}$$

Where $\hat{x}_s^i$ is the MAAT-DAN's reconstructed output for the i-th source domain sample x_s^i, and $\hat{x}_t^j$ is the reconstructed output for the j-th target domain sample x_t^j. $\|\cdot\|^2$ represents the squared Euclidean norm, which calculates the sum of the squares of the differences between the vector components.

Total loss: The total loss is a combination of task loss and total domain loss. To improve the feature extractor's robustness to domain shift, it is essential to initially reduce the influence of negative gradients via GRL during the early learning stages. This procedure is necessary because the substantial domain classification errors resulting from insufficient training at early learning stages can hinder the development of effective label prediction capabilities [78]. To make the model focus more on task accuracy during early-stage training, MAAT-DAN employs a total domain loss consisting of three components, with λ_d, λ_{mmd}, and λ_{recon} controlling the contributions of domain loss, MMD loss, and reconstruction loss, respectively. This loss function design emphasizes task performance in the early stages of training, while progressively increasing the weight of DA losses over time, allowing the model to strike a balance between task performance and DA. All three weighting parameters gradually increase from 0 to specific maximum values with the progress of the training epochs. When selecting these particular values, to improve the task training accuracy, λ_d and λ_{mmd} should be adjusted within a smaller range to minimize their impact on the training task during the early stages of model training. This paper recommends testing and choosing the maximum values of these two weighting parameters within the range of 0.1 to 10. The other weighting parameter, λ_{recon}, influences whether the features of the source and

target domains retain more original information, thereby reducing potential information loss during the transfer learning process. This paper suggests testing and choosing the maximum value of λ_{recon} within the range of 100 to 10,000. Configuring the loss function in this way potentially enhances the model's generalization ability across both the source and target domains. The total loss function is defined as follows:

$$L_{\text{total}} = L_{\text{task}} + \lambda_d \cdot L_{\text{domain}} + \lambda_{\text{mmd}} \cdot L^2_{\text{MMD}} + \lambda_{\text{recon}} \cdot L_{\text{recon}} \tag{9}$$

Note that L_{task} should be replaced with $L_{\text{task-R}}$ or $L_{\text{task-C}}$ based on the task type, to accommodate regression and classification tasks, respectively. By integrating the feature extraction capabilities of multi-scale adaptive multi-head attention based on the Transformer architecture, along with adversarial domain discrimination, statistical distribution alignment, and data reconstruction, the model has the prospect of learning more robust and transferable feature representations. It has the potential to enhance performance and reliability in practical engineering applications.

2.4. The learning rate setting and architecture of MAAT-DAN

In this research, the exponential decay approach is adopted for learning rate scheduling to optimize the model's training process. Specifically, the initial learning rate is set to 0.0001, with decay steps of 300 and a decay rate of 0.96. This strategy is implemented using TensorFlow's ExponentialDecay function, with the learning rate decreasing in a staircase fashion. This means that the learning rate drops by a fixed proportion after a set number of training steps (decay steps). This approach maintains a higher learning rate in the early training stages to accelerate convergence, while gradually lowering it later to avoid local optima. Finally, this learning rate scheduler is applied to the Adam optimizer, dynamically adjusting the learning rate to enhance the model's performance and stability.

The architecture with forwardprop and backprop, detailed layer parameters, and the four loss functions of MAAT-DAN are illustrated in Fig. 5. During the training phase, the feature extractor extracts features from the labeled source domain and the unlabeled target domain, fed into the regressor label predictor and unsupervised domain adapter for training. In the testing or application phase, the data will only use the damage feature extractor and regressor label predictor for label prediction in the target domain.

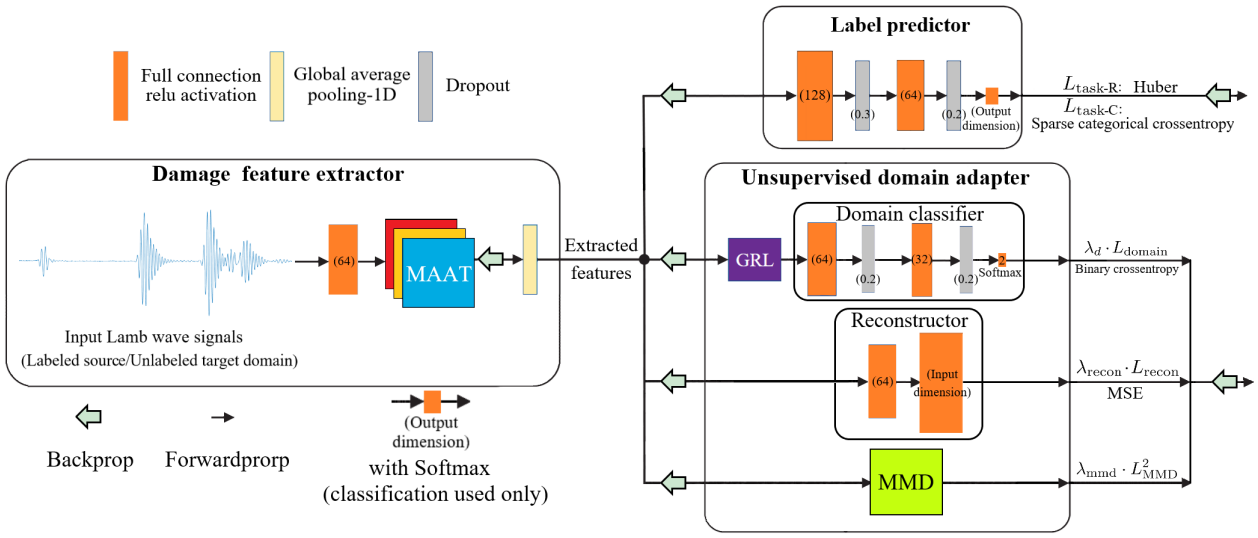


Figure 5: The architecture of the proposed MAAT-DAN framework with three components: damage feature extractor, regressor label predictor, and unsupervised domain adapter. The Lamb wave damage-sensitive features extracted by the MAAT feature extractor, are used to predict regression labels through a feedforward neural network; unsupervised DA, which effectively combines the GRL and MMD, and is further enhanced by feature reconstructor.

3. Ultrasound experiments validation

This section will conduct two ultrasonic experiments on a reinforced aluminum plate and a steel water tank, using the experimental data to validate MAAT-DAN's UTL performance for regression and classification tasks, respectively. Employing two experimental setups serves two key objectives. It examines the model's generalization in different transfer scenarios, including material-based transfer learning and temporal-sensor adaptation. Additionally, it assesses the model's capability in both regression and classification tasks, demonstrating its robustness in cross-domain transfer learning.

3.1. Regression task: ultrasonic experiment on reinforced aluminum plate

The aim of this experiment is to collect ultrasonic response signals from various damage locations on two different materials, treating them as source and target domain datasets. These signals will be used to evaluate the effectiveness of the proposed MAAT-DAN in localizing damage on the x and y coordinates within the target domain.

This experiment involves an aluminum plate (EN AW-1050A H14/H24, 1000 × 1000 × 1.5 mm) reinforced with two types of stiffeners: one made of aluminum (EN AW-6082 T6, 1000 × 10 × 10 mm) and the other made of solid carbon composite (unidirectional carbon fibre, 1000 × 10 × 10 mm). Both stiffeners are individually glued to the aluminum plate using Loctite 480 adhesive and left to dry for 24 hours. In this experiment, three Steminc piezoelectric discs (5 × 0.4 mm, 450 kHz, wire leads), with part number SMD05T04R111WL, designed for radial mode vibration applications, are used. The PZTs are affixed to the aluminum structure using Loctite 480 adhesive, with sensor locations depicted in Fig. 6. One side of the plate, before the stiffener, contains a PZT actuator that will generate the ultrasonic waves. The waves will pass through the stiffener and be received by the two PZT sensors.

Blu-tack, a clay-like substance conventionally used for adhering objects, simulates damage in this experiment. It facilitates the easy replication of experimental conditions while offering reusability, making it a widely accepted tool in ultrasonic guided wave-based SHM tests [79]. These unique properties establish Blu-tack as an ideal choice for achieving repeatable and relatively consistent damage simulations, with its feasibility validated in the literature [80–82]. This experiment examines 15 different damage locations on an aluminum plate, with each location assessed individually. The 15 damage locations in the aluminum plate and their corresponding x and y coordinates are shown in Fig.6 (b) and Table 1. For each damage location, three levels of severity are assumed. The severity is represented by three weights of Blu-tack samples, each with almost the same area: 16g, 32g, and 48g (shown in Fig. 7 (b)), in ascending order of damage severity.

The reinforced aluminum plate ultrasonic testing protocol, as depicted in Fig. 7, follows a carefully structured procedure. Initially, the excitation signal is determined through careful observation of waveform response characteristics—including resolution, clarity, and noise levels—across various frequencies for the aluminum plate with stiffener. Among multiple trials, this protocol achieved an improved signal-to-noise ratio and noticeably reduced waveform distortion. This inspection leads to the selection of a Hanning windowed sinusoid with a 300 kHz frequency and 4-cycle duration, chosen to optimize the experimental signal quality. The subsequent experimental procedure is executed as follows: A laptop running MATLAB controls the excitation signal generation. This signal, produced by a wave generator (KEYSIGHT 33500B), is visualized on an oscilloscope (InfiniiVision DSOX2014A with sampling rate 2 GS/s and vertical scaling) before being transmitted to the aluminum structure via a PZT transducer. Concurrently, two receiving PZT sensors, each connected to separate oscilloscope channels using crocodile clip cables, capture the response signal.

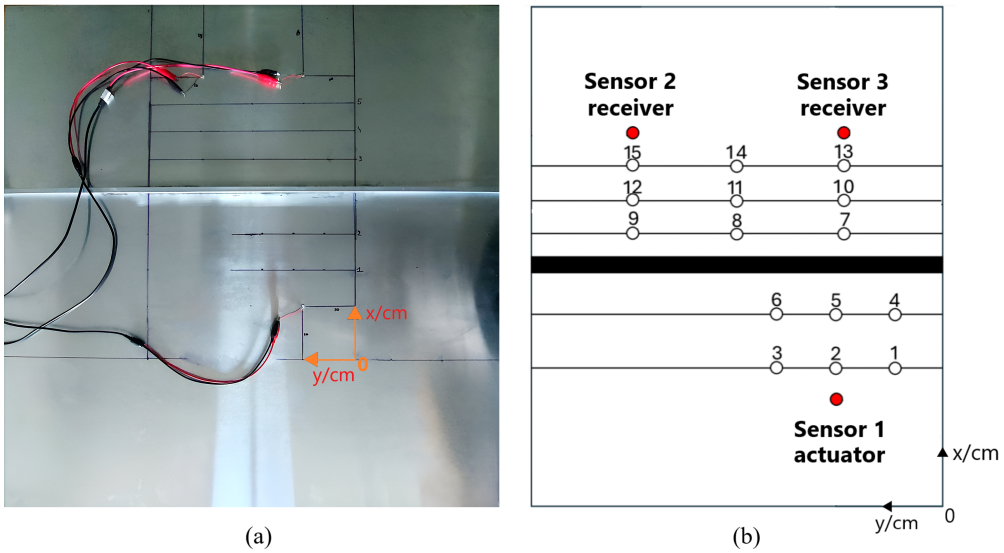


Figure 6: Ultrasonic experiment sample and damage locations: (a) Aluminum plate with aluminum stiffener PZT sensors; (b) The positions of the excitation and receiving PZT sensors, as well as the 15 different damage locations.

Fig. 8 presents a comparison of signals across three scenarios, received by sensor 3. In particular, Fig. 8 (a) shows a comparison of signals at damage location 1, with 16 g of Blu-tack placed under two different stiffeners: aluminum and carbon fibre. The difference is especially noticeable between 1.25×10^{-4} and 1.75×10^{-4} seconds,

Table 1: Coordinates of 15 independently existing damage locations.

Damage location indices	x-coordinate (cm)	y-coordinate (cm)
1	17	6
2	17	12
3	17	18
4	24	6
5	24	12
6	24	18
7	34	10
8	34	20
9	34	30
10	38	10
11	38	20
12	38	30
13	42	10
14	42	20
15	42	30

where the carbon fibre signal exhibits higher peaks and significantly lower dispersion compared to the aluminum signal. This suggests that the carbon fibre stiffener exhibits higher amplitudes, likely due to its lower density, greater stiffness, and slower damping compared to the aluminum stiffener. The ultrasonic response signals for three different types of damage are shown in Fig. 8 (b) and (c). These damages are simulated by varying the weight of Blu-tack at location 1 with the carbon fibre stiffener, and by changing the damage locations. From the comparison, the ultrasonic response signals for different Blu-tack weights and positions have similar overall patterns. Although there are some amplitude fluctuations when the Blu-tack weight increases from 16 g to 48 g or when the damage location changes, these fluctuations do not significantly affect the overall ultrasonic response of the structure.

3.1.1. Validating and evaluating strategy

This section evaluates MAAT-DAN's performance in feature extraction and unsupervised DA through comparative analyses of its key components. The assessment investigates how different feature extractors and domain adapters impact the proposed framework. The contributions of both MAAT and DAN are examined by comparing:

(1) The effectiveness of different feature extraction strategies: Comparing MAAT-DAN against feature extraction approaches CNN (The VGG-13 architecture, as a classic CNN-based feature extraction framework, is used [14]) with DAN (VGG-13-DAN), AAT with DAN (AAT-DAN), and standard Transformer with DAN (Transformer-DAN).

(2) The effectiveness of different DA strategies: Comparing MAAT-DAN against DA approaches using only GRL (MAAT-GRL) and MMD (MAAT-MMD) in unsupervised DA for damage localization.

This comparative analysis highlights the relative contributions of each component within MAAT-DAN. It also sheds light on the interdependencies between these components and their collective impact on improving damage localization accuracy.

3.1.2. Dataset description of regression task

Ultrasonic experimental data from the described structures in Section 3.1 are used to validate the efficacy of the proposed MAAT-DAN methodology. These data, collected from 15 locations and with three damage severities, serve to evaluate MAAT-DAN's capabilities in UTL for structural damage localization.

During training we are given the source domain $\mathcal{D}_{sr} = \left\{\left(x_{sr}^{i}, y_{sr}^{i}\right)\right\}_{i=1}^{n_{sr}}$ of n_{sr} labeled examples and a target domain $\mathcal{D}_{tr} = \{x_{tr}^{j}\}_{j=1}^{n_{tr}}$ of n_{tr} unlabeled examples for the regression task. To better reflect the practical engineering scenario where the target data volume is smaller than the source data volume, we set $n_{tr} < n_{sr}$. Both raw input data $x_{sr}^{i}, x_{tr}^{j} \in \mathbb{R}^{dr}$, where dr is the dimensionality. The distributions between two domains are different, i.e. $p_{sr}(x_{sr}) \neq q_{tr}(x_{tr})$.

The input data consists of raw ultrasonic signals from an aluminum plate with the aluminum stiffener. These signals are captured by PZT receivers 2 and 3, with minor reflected components removed. This ensures that the model focuses on learning meaningful features from the primary signal components and enhances training efficiency. This serves as the source domain features, with the damage location's x and y coordinates used as

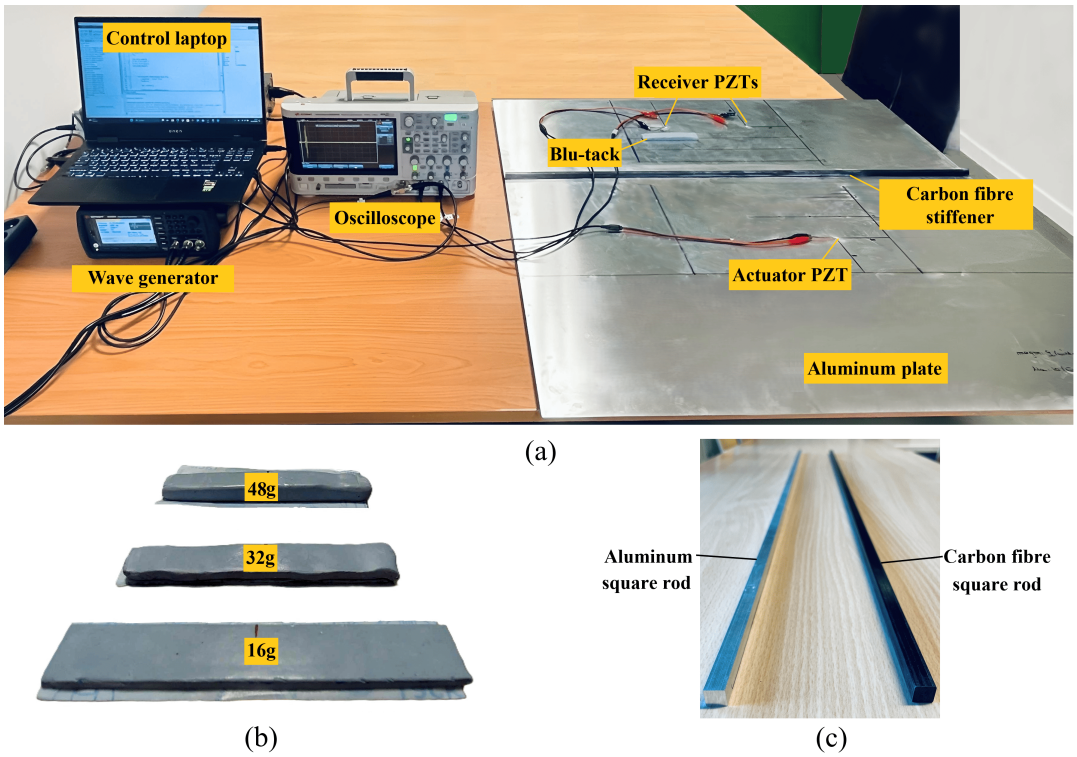


Figure 7: Ultrasonic experiments on aluminum plates: (a) Experiment equipment with damage location index 15 and carbon fibre stiffener; (b) Three weights of Blu-tack samples; (c) Aluminum and carbon fibre stiffeners.

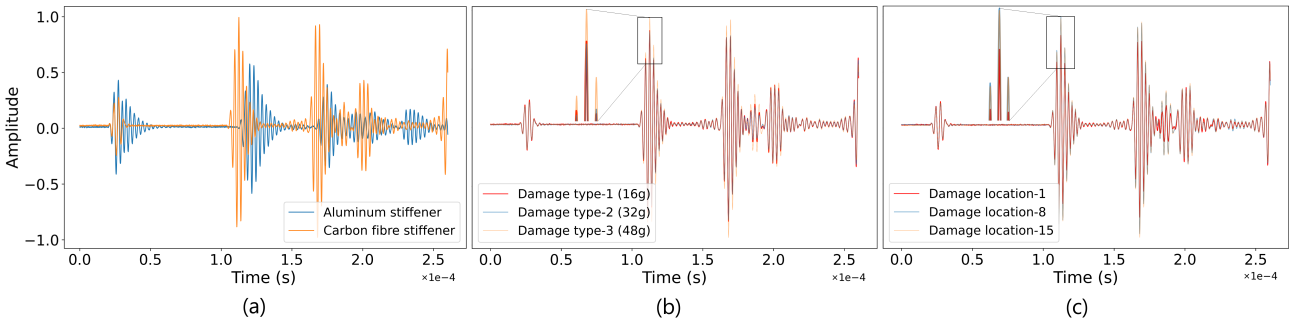


Figure 8: A comparison of signals received by Sensor 3 under 300 kHz excitation in three scenarios: (a) Signals for 16g Blu-tack at damage location 1 under two different stiffeners; (b) Signals for three damage severities (weights) at damage location 1 under the carbon fibre stiffener; (c) Signals for 16g Blu-tack at three different damage locations under the carbon fibre stiffener.

the source domain labels. The source domain dataset consists of 90 ultrasonic signals, each one having 2,500 feature points and two location labels. For the target domain, the input data consists of unlabeled raw ultrasonic signals from an aluminum plate with a carbon fibre stiffener. These signals are collected only by PZT receiver 3 to simulate the practical scenario of having limited target domain data, with minor reflected components also removed. The target domain dataset consists of 45 ultrasonic signals with each 2500 feature points.

Note that we employed a different label design strategy, involving many identical labels, to evaluate the model's feature extraction ability and regression performance. Although this design increases the difficulty of regression analysis, it also creates a more rigorous testing environment for the model's robustness and adaptability. Specifically, the repeated labels arise from three factors: (1) the use of three different Blu-tack weights to simulate varying damage severities at the same damage location; (2) the x-coordinate remains the same across three consecutive locations in the experimental setup (as shown in Table 1); (3) signals from both PZT 2 and PZT 3 receivers are used in the source domain. These factors result in many equal position labels across both domains, forcing the model to extract meaningful features from subtle signal differences. For the source domain (90 signal samples), categorizing by identical x and y coordinates results in 15 groups, each containing 6 signal samples with the same x and y coordinates. If categorized by identical x-coordinates, the data can be divided into 5 groups, each consisting of 18 signal samples sharing the same x-coordinate. The target domain, utilizing only one PZT receiver, contains half the number of the same coordinates compared to the source domain. It's

worth noting that target domain labels are only used for testing purposes. While the large number of identical labels increases the difficulty of accurately predicting damage locations in regression analysis, it also provides a more stringent test of the model's performance.

All waveform signal amplitudes are normalized to the range $[-1, 1]$ through a two-step process. First, the maximum amplitude among all waveforms is identified separately for the source domain and target domain. This maximum amplitude in each domain is set as the reference and normalized to the range $[-1, 1]$. Subsequently, all other signals within the same domain are scaled proportionally based on their amplitude ratio to this reference waveform. This normalization approach ensures that the relative amplitude relationships within each domain are preserved while standardizing the overall amplitude range, thereby avoiding the effects of scale differences between domains. To provide better visualization of the damage location prediction and a clearer and more intuitive display of the results, after randomly dividing the dataset, both domain datasets are reorganized based on their corresponding damage locations, arranged from smallest to largest.

3.1.3. Validation of regression task

This research examines all six UTL methods mentioned above, implemented using Python 3.10.12 and the Keras deep learning framework. Training is conducted on TensorFlow 2.17.0. The models are run on a Google Colaboratory Jupyter notebook platform, with T4 GPU acceleration. This platform, hosted on Google Cloud, provides computational resources for efficiently executing resource-intensive tasks without requiring local hardware.

To ensure a fair comparison, all methods undergo identical evaluation conditions: 2000 epochs (converged around 1500 epochs) of training with a batch size of 16, repeated over 10 iterations. The performance assessment utilizes four key metrics:

Metric-1: Mean and Standard Deviation (STD) of predicted each x and y damage location in the target domain, computed across the 10 trials. The metric-1 can be defined as follows:

$$\bar{\hat{c}} = \frac{1}{N}\sum_{k=1}^{N}\hat{c}_k, \quad \sigma_{\hat{c}} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\hat{c}_k - \bar{\hat{c}}\right)^2}, \tag{10}$$

where $\hat{c}$ is a generic symbol representing a coordinate, which can be either x or y. $\hat{c}_k$ represent the predicted x or y coordinates of the damage location for each trial k; N denotes the total number of trials conducted (in this specific case, $N = 10$), and $\bar{\hat{c}}$ signify the mean values of the predicted x or y coordinates across all trials; The $\sigma_{\hat{c}}$ indicate STD in the x or y predictions.

Metric-2: Combined R^2 value for x and y coordinates, derived from the mean predicted x and y values versus true values over 10 trials. The metric-2 can be defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{2n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{2n}\left(Y_i - \bar{Y}\right)^2}, \ \bar{Y} = \frac{1}{2n}\sum_{i=1}^{2n}Y_i, \tag{11}$$

where the combined true values $Y = \{x_i,\ y_i\}_{i=1}^{n}$ and the combined predicted values $\hat{Y} = \{\bar{\hat{x}}_i,\ \bar{\hat{y}}_i\}_{i=1}^{n}$. $x_i,\ y_i$ are the true x and y values for sample i within the total number of data points n. $\bar{\hat{x}}_i,\ \bar{\hat{y}}_i$ are the mean predicted x and y values over the 10 trials for sample i.

Metric-3: Combined mean and STD of MAE and RMSE across x and y coordinates for damage location predictions in the target domain, calculated from 10 trials. The metric-3 can be defined as follows:

$$\text{Mean of } \mathcal{C}: \quad \overline{\mathcal{C}} = \frac{1}{N}\sum_{k=1}^{N}\mathcal{C}_k, \quad \text{STD of } \mathcal{C}: \quad \sigma_{\mathcal{C}} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\mathcal{C}_k - \overline{\mathcal{C}}\right)^2}, \tag{12}$$

where $\mathcal{C}$ serves as a placeholder for both MAE and RMSE. The combined MAE and RMSE across x and y coordinates are defined below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left(|x_i - \hat{x}_i| + |y_i - \hat{y}_i|\right), \ RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left((x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2\right)}. \tag{13}$$

Metric-4: Combined x and y error distribution with Gaussian Kernel Density Estimation (KDE), based on the mean predicted x and y values compared to the true values over 10 trials. The distribution of the prediction

errors is estimated using KDE, providing a non-parametric view of the error distribution. The metric-4 can be defined as follows:

$$\begin{gathered}\text{Error}_{c,i} = \bar{\hat{c}}_i - c_i, \quad c \in \{x, y\}, \quad i = 1, 2, \ldots, n, \quad \text{Error}_{\text{combined}} = \{\text{Error}_{x,i}, \text{Error}_{y,i}\}_{i=1}^{n}, \\ \hat{f}(E_{\text{combined}}) = \text{KDE}(\text{Error}_{\text{combined}}), \\ \mu = \frac{1}{2n}\sum_{i=1}^{2n} E_{\text{combined},i}, \ \sigma = \sqrt{\frac{1}{2n}\sum_{i=1}^{2n}(E_{\text{combined},i} - \mu)^2},\end{gathered} \tag{14}$$

where $\bar{\hat{c}}_i$ represents the mean predicted value for coordinate c (either x or y) over the 10 trials for sample i; The c_i is true value for coordinate c for sample i. The combined error, Errorcombined, is the collection of errors for both x and y coordinates. The error distribution $\hat{f}(E_{\text{combined}})$ is estimated via KDE, which provides a smooth, non-parametric Probability Density Function (PDF) that captures the underlying distribution of the prediction errors.

(1) Regression performance on different feature extractors

To verify the effectiveness of the feature extractor proposed in this paper for regression prediction in the target domain, this section will evaluate the model's performance in predicting damage locations in the target domain without multi-scale feature extraction and Transformer adaptive weighting module. Additionally, this performance will be compared with that of the commonly used CNN feature extractor. The detailed results analysis of the different methods will be presented in Section 4.1.

MAAT-DAN: The MAAT-DAN method proposed in this paper predicts the damage locations in the target domain. Based on the results of 10 iterations of training and prediction for the x and y coordinates, the mean and STD for each damage location in the target domain are calculated. The mean of the x and y predictions at each location over 10 iterations serves as the final prediction, while the STD at each location is used as an uncertainty measure. To evaluate the overall regression performance in the target domain, the mean x and y data are combined into a vector, and both x and y mean prediction results are treated as a whole for calculating metric-2. The computed combined R^2 value reflects the model's overall fit when simultaneously predicting both x and y coordinates, providing a comprehensive measure of goodness of fit. Based on the current input data, the training time for MAAT-DAN is approximately 4-5 minutes. The overall prediction performance of MAAT-DAN in the target domain is shown in Fig. 9 (a), where $R^2 = 0.719$. In the x and y plot, the color intensity of each point increases from light to dark, indicating a rise in uncertainty, which corresponds to the uncertainty bar on the right side of Fig. 9 (a), derived from the STD of the 10 predictions. The reference line in Fig. 9 along the diagonal represents the ideal prediction, with points closer to this line indicating a closer match to the true values.

VGG-13-DAN: CNNs are among the most commonly used deep learning methods for feature extraction. To display the advanced feature extraction capability of MAAT, this section evaluates the performance of a CNN-based feature extractor in the target domain, with a comparison of the two methods presented in Section 4.1. The VGG neural network, widely recognized for its strong feature extraction capabilities [7, 83], is employed in this research. Specifically, we use a VGG-13-based feature extractor with an added Dropout layer, combined with the Domain adapter proposed in this paper, to perform unsupervised DA training and test its performance in the target domain. For the current input data, the training time for VGG-13-DAN is approximately 5-6 minutes. The prediction results for x and y damage locations in the target domain using VGG-13-DAN are shown in Fig. 9 (b), where the combined $R^2 = 0.270$. These results indicate that the feature extraction and prediction performance of VGG-13 is suboptimal for an ultrasonic dataset characterized by numerous identical labels and similar waveforms. The prediction STD for each damage location is represented in the uncertainty bar on the right side of Fig. 9 (b).

AAT-DAN: The AAT-DAN is designed to assess the impact of multi-scale feature extraction on regression task. It is derived from the MAAT-DAN method by excluding the multi-scale feature extraction component. This approach allows for a direct comparison to assess the impact of this specific technique on the overall performance. Based on the current input data, the training time for AAT-DAN is around 2.5 minutes. The mean results of AAT-DAN's unsupervised DA for 10 predictions of damage locations in the target domain are shown in Fig. 9 (c), with a combined $R^2 = 0.694$, slightly lower than that of MAAT-DAN. This suggests that multi-scale feature extraction contributes positively to the final UDAR. The STD in the predictions for each damage location is illustrated by the uncertainty bar on the right side of Fig. 9 (c). From the uncertainty bar, it can be observed that the range of variability in AAT-DAN's x and y predictions over 10 iterations has slightly increased than MAAT-DAN.

Transformer-DAN: To further assess the impact of weight adaptation on feature extraction and DA, the weight adaptation technique is removed from AAT-DAN, using only the original Transformer architecture as the

feature extractor. Based on this, the DA technique proposed in this paper is integrated to form Transformer-DAN, which was then applied for DA training on ultrasonic damage data. With the given input data, the training time for Transformer-DAN is approximately 2 minutes. The mean results of Transformer-DAN's unsupervised DA over 10 predictions of damage locations in the target domain are shown in Fig. 9 (d), where the combined R^2 = 0.589, which is noticeably lower than that of AAT-DAN. This indicates that the weight adaptation technique plays a positive role in enhancing DA regression. The prediction STD for each damage location is displayed in the uncertainty bar on the right side of Fig. 9 (d).

(2) Regression performance on different domain adapters

To assess the effectiveness of the proposed method that combines GRL and MMD for DA, this section will individually apply the GRL and MMD techniques, in conjunction with the feature extraction method MAAT introduced in this paper, for DA training and testing. The relative efficacy of these techniques is examined both when applied individually and in combination. The comprehensive analysis comparing the various methods is presented in Section 4.1.

MAAT-GRL: The GRL technique, initially introduced for DA in DANN [40], is employed in this study. We integrate this technique with our proposed MAAT feature extractor, creating MAAT-GRL. This combined approach is then trained and evaluated for its performance in predicting damage locations within the target domain, using ultrasonic damage data. With the current input data, the training time for MAAT-GRL is approximately 5.5 minutes. The mean results of MAAT-GRL's unsupervised DA for 10 predictions of damage locations in the target domain are shown in Fig.9 (e), where the combined $R^2 = 0.665$. The prediction variance for each damage location is displayed in the uncertainty bar on the right. It can be observed that the STD of MAAT-GRL's x and y predictions over 10 iterations has significantly increased, consistent with the instability of GRL training discussed in Section 2.2.

MAAT-MMD: MMD is a commonly used DA technique. In this section, the effectiveness of applying MMD alone for unsupervised DA is evaluated. MMD is similarly combined with the MAAT feature extractor proposed in this paper. Based on the current input data, the training time for MAAT-MMD is approximately 4.5 minutes. The mean results of MAAT-MMD's unsupervised DA for 10 predictions of damage locations in the target domain are shown in Fig. 9 (f), where the combined $R^2 = 0.663$, indicating the goodness of fit in the target domain similar to that of using GRL alone. The prediction STD for each damage location is displayed in the uncertainty bar on the right side of Fig. 9 (f). Based on the STD from the uncertainty bar, the MMD-based DA technique shows relatively small variability in STD than the GRL-based DA technique.

3.2. Classification task: ultrasonic experiment on steel water tank

In the UTL classification task, the purpose of this task is to perform transfer learning for damage identification across time (1 year) and across two sensor pairs to identify structural integrity. The validation dataset is derived from ultrasonic experiments conducted over 8 years (2012-2020), utilizing a sparse PZT sensor array positioned on a steel water tank to capture comprehensive structural monitoring data [84]. In the experimental setup depicted within Fig. 10. The experimental water tank investigation utilized a 2 m diameter cylindrical reservoir constructed from 5 mm thick semi-cylindrical steel panels, interconnected via bolted flanges with an unspecified sealant. The monitoring system incorporated eight strategically positioned guided wave sensors, each comprising a 20 mm diameter, 1 mm thick piezoelectric disk (NCE51, Noliac A/S, Denmark) with wraparound electrodes, securely bonded to the tank's external surface using cyanoacrylate adhesive [85]. For signal transmission, the piezoelectric disk's electrodes are connected to a coaxial cable through soldered conductors. A protective silicone sealant encapsulation shields the assembly from environmental factors, ensuring operational reliability.

In this experimental setup, sensor wire connections utilize parallel arrangements with a pair of single-pole, 10-channel radio frequency relays functioning as multiplexing interfaces. One multiplexor connects to a Handyscope HS3 (an integrated digital oscilloscope and signal generator by TiePie Engineering), while the other connects to the output of the device's signal generator [85]. The multiplexing configuration enables selective sensor interconnectivity, allowing specific sensors to alternate between transmitter and receiver roles. For signal excitation, a specialized waveform is employed: 1000 transmission cycles mimicking a 5-cycle Hanning-windowed toneburst centered at 250 kHz (through a de-chirping procedure to reconstruct the S-cycle response), intentionally designed to maximize signal energy transmission.

The monitoring process involves systematically recording time domain responses from sensor array pairs during each measurement cycle. A multiplexing system sequentially switches between different sensor configurations, ensuring comprehensive data collection across all possible transmit-receiver combinations. For analytical clarity, this study focuses on data derived from four selected sensors, specifically the transmit-receiver pairs of

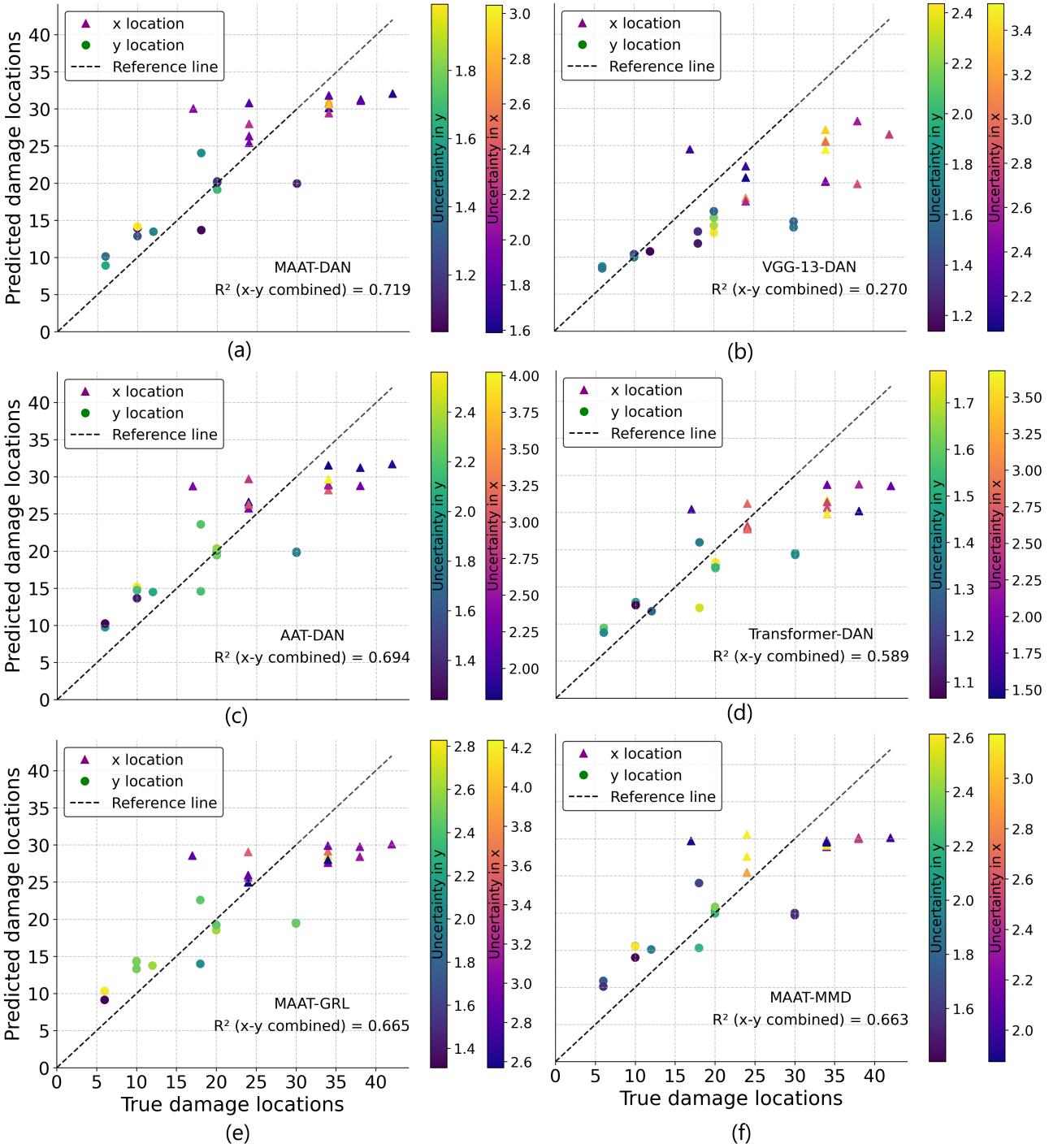


Figure 9: Regression predictions and uncertainty results of damage locations in the target domain: mean and STD predictions from six methods across 10 training iterations.

sensors 1-2 and 3-4, whose spatial arrangement is illustrated in Fig. 10. Throughout the monitoring duration, the water tank remained empty, serving as a controlled experimental medium for the SHM investigation.

The experimental setup is situated in an exposed exterior environment, subjecting the structural monitoring apparatus to natural weathering over an extended period. While the monitoring protocol originally prescribed daily measurements, practical constraints such as infrastructure modifications and equipment malfunctions resulted in intermittent data acquisition, creating some temporal gaps in the recorded dataset.

3.2.1. Dataset description of classification task

To evaluate the performance of the proposed MAAT-DAN methodology on the classification task, we use ultrasonic data from the first two years (2012 and 2013), which have higher data integrity compared to the

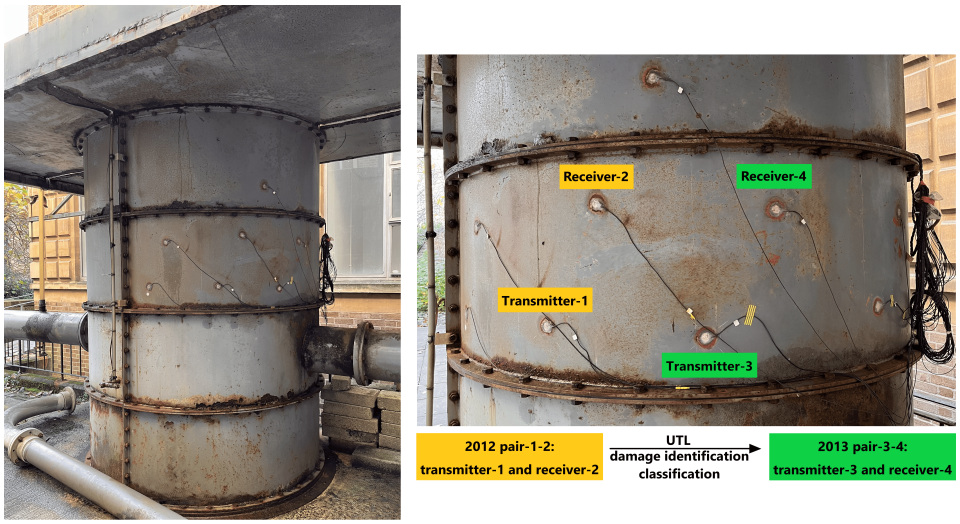


Figure 10: Steel water tank ultrasonic experiment and UTL classification task transfer direction.

following years (from 2014 to 2020). For the classification task, a larger set of labeled source data and a smaller set of unlabeled target data with different distributions are used. This reflects a practical scenario where the target data volume is less than the source data volume

Specifically, the 2012 test data from sensor pair-1-2, serving as the source domain, was recorded starting on January 4, 2012, with signals captured every 20 minutes until December 31, 2012. The 2013 test data from sensor pair-3-4, acting as the target domain, was recorded starting on January 1, 2013, with signals captured every 20 minutes until August 20, 2013. Both sensor pairs experienced signal loss on certain days and months due to uncontrollable factors. In total, pair-1-2 recorded 18,455 ultrasonic guided wave signals in 2012, while pair-3-4 recorded 14,240 signals in 2013. Given the high similarity of signals recorded on the same day and the computational resource constraints, this research averaged all signals received on the same day, retaining a single representative ultrasonic signal per day for model validation.

The data collection included two datasets. The source domain from sensor pair 1-2 comprises 263 ultrasonic signals, each consisting of 2000 feature points. The identical excitation waveform and the final minor reflection components, which showed minimal amplitude variation, are removed. The signals are labeled as 0 (healthy) or 1 (damaged), with the damaged class simulated by adding localized reflection signals. The target domain from sensor pair 3-4 contains 203 waveforms divided into unlabeled training data and labeled test data using the same binary classification scheme. All signals are normalized using the same method described in Section 3.1.2. Fig. 11 (a) and (c) compare the signals from pair-1-2 in 2012 and pair-3-4 in 2013 on their first and last test days, respectively. The results indicate amplitude and phase shifts in both sensor pairs, with more pronounced changes observed in pair-3-4 compared to pair-1-2. These variations may result from factors such as sensor quality, adhesion, environmental conditions, and differing levels of structural surface corrosion. Fig. 11 (b) and (d) show the signal comparisons for the two sensor pairs on their first and last test days within their respective years. Clear differences in the signals are evident due to recording time changes and variations in sensor placement.

The joint Principal Component Analysis (PCA) is used to project data from the two domains into the same PCA space, enabling a comparison of their relative positions and distributions. PCA is a widely used dimensionality reduction technique that maps high-dimensional data to a lower-dimensional space while preserving the maximum variance [86]. In this study, source and target domain data are combined for PCA, and the samples are mapped onto the two most significant principal components for visualization purposes. Fig. 11 (e) illustrates the data distributions of the two classes in each domain by PCA. The source and target domain data shown in Fig. 11 (e) are almost entirely separated in the PCA space, reflecting significant differences in their feature distributions. Source domain data points are tightly clustered on the left side of the coordinate plane near (-4, 0), indicating low variability in its feature space. In contrast, target domain data are more widely dispersed, spanning a larger range of principal components, which suggests greater diversity in feature representation. Moreover, the two classes within each domain (Class 0 and Class 1) show minor partial internal separation in the projected space. The pronounced differences in feature distributions between the source and target domains increase the difficulty of transfer learning classification tasks, serving to evaluate the model's robustness under conditions of substantial data distribution discrepancies.

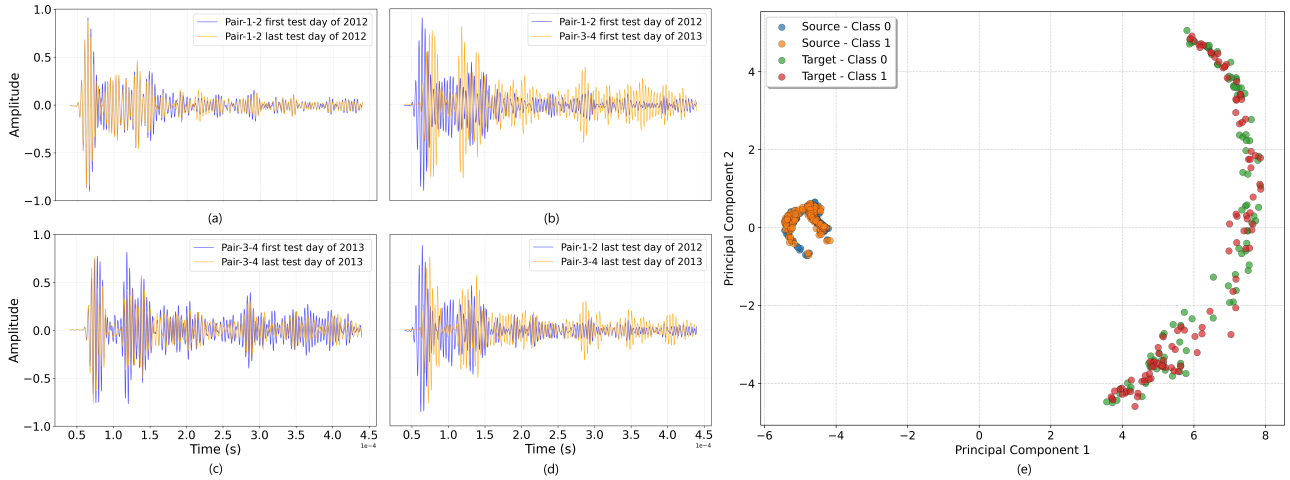


Figure 11: Comparison of both domain ultrasonic signals: waveforms (a-d) and PCA dimensionality reduction distribution characteristics (e).

The data from 2012 and 2013 were measured in the early stages following sensor installation, when the structural integrity of the water tank was relatively intact. Therefore, all signals received during these two years can be considered healthy. To evaluate the model's performance in classification tasks, a simulated defect response is added to a randomly selected half of the signals received by pair-1-2 and pair-3-4. The simulated defect response is generated by modulating a sine wave with a Hanning window, resulting in a Hanning-windowed tone burst with a center frequency of 250 kHz. The formula for the localized defect signal (100 valid points) is defined as follows:

$$y_{\text{defect}}\ (t) = A_{\text{defect}}\ \cdot \sin(2\pi f t) \cdot w(t), \tag{15}$$

where A_{defect} is the amplitude adjustment coefficient for the defect signal, f represents the frequency of the defect signal, which is set to 250 kHz. $w(t)$ represents the Hanning window function, expressed as $w(t) = 0.5\left(1 - \cos\left(\frac{2\pi t}{N-1}\right)\right)$ with a window length of $N = 100$ (equivalent to 0.2×10^{-4} s). The variable t denotes the defect signal duration. Assuming the defect location causes receiver 4 detecting half the reflected signal amplitude compared to receiver 2. Consequently, the amplitude coefficients are calibrated to $A_{\text{defect}} = 0.1$ and 0.2 for receivers 4 and 2, respectively. The Hanning window is implemented starting from the 600th valid data point, with time delays of 10 and 20 valid points assigned to receivers 2 and 4, respectively, based on the assumed distance of the damage relative to each receiver. Using the signals from the first test days of 2012 and 2013 as examples, Fig. 12 presents the original signals alongside the defect-synthesized signals. At the defect location, the synthesized waveforms show subtle differences from the original signals in both Fig. 12 (a) and (b). The signals used for the source domain exhibit smaller changes.

Note that the defect response consider only the primary defect reflections, occurring at the initial point of wave interaction with the defect. This approach is thought to be more conservative than the real measurements, as it excludes both multiple reflections and shadowing effects on structural features, which typically produce stronger responses than the initial defect reflection.

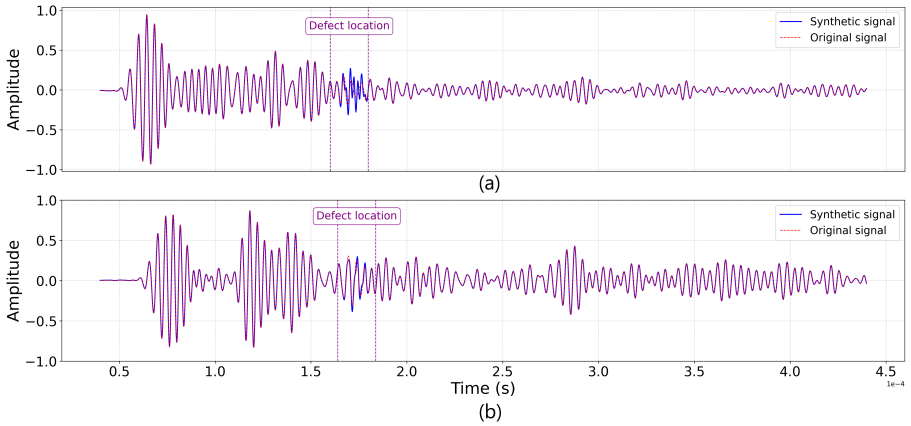


Figure 12: Comparison of original and defect-synthesized signals from the first test day: (a) year 2012 with sensor pair 1-2, and (b) year 2013 with sensor pair 3-4.

3.2.2. Validation of classification task

The classification task is conducted in the same computational environment as the regression task. To ensure a fair comparison, all methods undergo identical evaluation conditions: 500 epochs (converged around 300 epochs) of training with a batch size of 16, repeated over 10 iterations. The performance assessment utilizes three metrics—Accuracy, Recall, F1—as described in [87], ensuring a clear and robust assessment of its effectiveness.

To validate the effectiveness of MAAT-DAN for the classification task, the same approach from regression is employed to evaluate the performance of feature extractors and domain adapters across six methods on the target domain test dataset. We conducted 10 independent classification trials for each method to evaluate model performance in distinguishing structural health states. The architecture of each method used for the classification task is largely consistent with that used for the regression task. The only differences are the addition of a softmax activation function to the task label predictor and the modification of the label predictor's loss function from Huber to sparse categorical cross-entropy (as shown in Fig. 5). In the target domain tests, the dataset contains 30 and 31 samples labeled as 0 and 1, respectively. The detailed results analysis of the different methods will be presented in Section 4.2.

(1) Classification performance on different feature extractors

The validation of feature extractors will be performed using MAAT-DAN, VGG-13-DAN, AAT-DAN, and Transformer-DAN for the classification task. This incremental validation will assess the improvements in classification performance due to the multi-scale feature extraction and adaptive weights technologies in the feature extractor module MAAT. Additionally, the performance of MAAT will be compared with that of traditional Transformer and VGG-13 feature extractors. The confusion matrices illustrating the mean and STD results for both categories (0 and 1) in the target domain are presented in Fig. 13 (a) to (d). These mean and STD values for each category are derived from classification performance over 10 trials. Using the water tank input data and the specified computation environment, the training times for different models are approximately as follows: MAAT-DAN takes 4–5 minutes, VGG-13-DAN requires 5–6 minutes, AAT-DAN completes in 3–4 minutes, and Transformer-DAN finishes in 2–3 minutes.

The classification results of MAAT-DAN and AAT-DAN demonstrate high accuracy and stability in classifying labels 0 and 1, with both achieving comparable performance. Transformer-DAN, which removes the adaptive weight technique from AAT-DAN, shows a notable decrease in classification performance compared to AAT-DAN. This highlights the positive impact of the adaptive weight technique in enhancing classification performance. VGG-13, as a traditional CNN-based feature extractor, produces more frequent FP and FN in its classification results. Additionally, the standard deviation across 10 predictions indicates that the VGG-13-DAN model exhibits higher classification instability.

(2) Classification performance on different domain adapters

MAAT-DAN's domain adapters combine GRL and MMD techniques. The domain adapters will be validated by comparing the classification performance of individually applied GRL and MMD's DA techniques with the combined approach, MAAT-DAN. The mean and STD of classification results across 10 training trials in the target domain are shown in Fig. 13 (e) and (f). The results indicate that, in the classification task, MAAT-GRL demonstrates a higher level of uncertainty relative to MAAT-MMD and MAAT-DAN. Based on the water tank input data and the calculation environment, the training time for both MAAT-GRL and MAAT-MMD are approximately 5-6 and 4-5 minutes, respectively.

4. Results comparison and analysis

4.1. Results comparison and analysis of regression task

4.1.1. Feature extractors comparative analysis of regression task

Fig.9 (a) to Fig.9 (d) illustrate the regression results from unsupervised DA using different feature extractors based on the reinforced aluminum plate ultrasonic data. Overall, the prediction accuracy for the y location is better than that for the x location. This is primarily due to the higher number of the same x labels in the ultrasonic data (as mentioned in Section 3.1.2), which makes accurate prediction more challenging. In contrast, for the y location, where the number of the same labels is one-third that of the x labels, prediction accuracy has improved.

The comparison based on metric-1 and metric-2 demonstrates that MAAT-DAN considerably outperforms the other three UTL methods constructed with different feature extractors. The transformer-based feature extractor performs notably better than the VGG-13 architecture. Specifically, as the MAAT technology modules

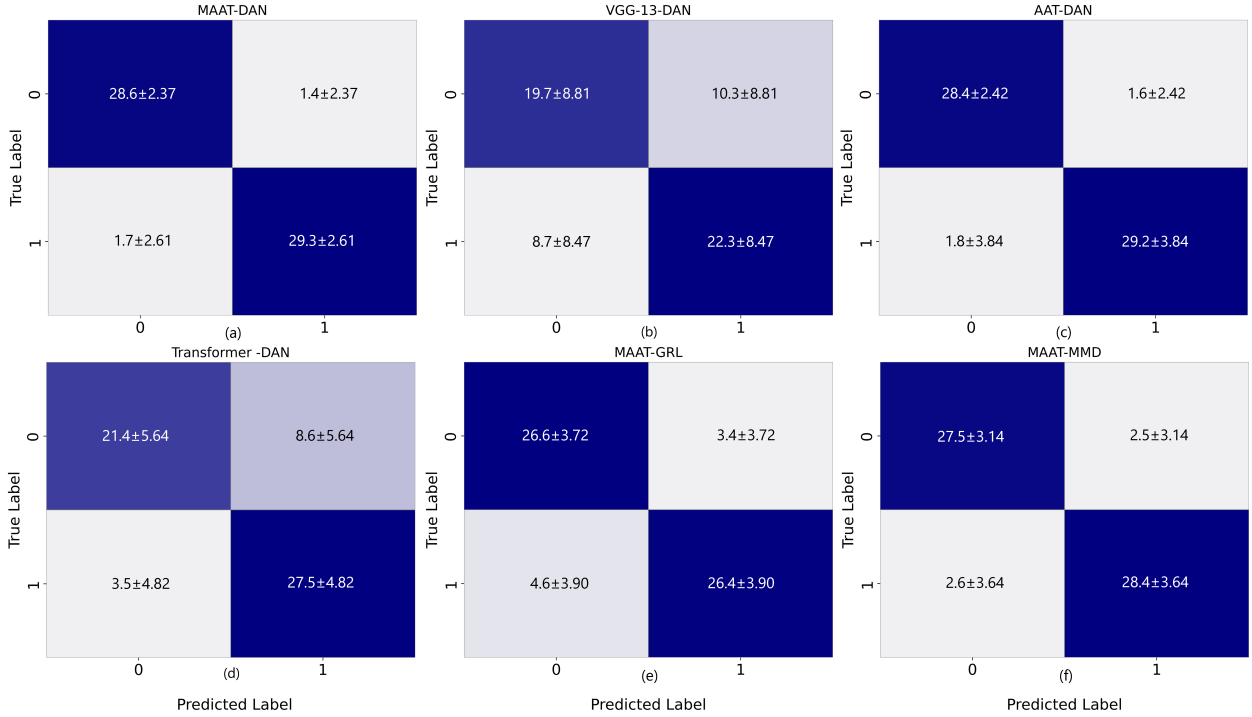


Figure 13: Averaged mean and standard deviation of confusion matrices for structural health state classification across 10 trials for six models.

are reduced (from MAAT to AAT to Transformer), the combined R^2 values gradually decrease (from 0.719 to 0.694 to 0.589), highlighting the effectiveness of multi-scale feature extraction and adaptive weight techniques in enhancing the model's overall predictive performance.

The comparison results based on metric-3 are shown in Fig. 14. The MAAT-DAN method exhibits the lowest MAE and RMSE values compared to the other methods, while VGG-13-DAN shows the highest. These results confirm the excellent performance of the MAAT feature extractor in domain adaptation regression. Regarding the STD of MAE and RMSE, both MAAT-DAN and AAT-DAN demonstrate lower prediction error variability compared to Transformer-DNA. In contrast, using VGG-13 as the UTL feature extractor results in the highest STD for both MAE and RMSE. This highlights the positive impact of multi-scale feature extraction and adaptive weight techniques on improving prediction stability.

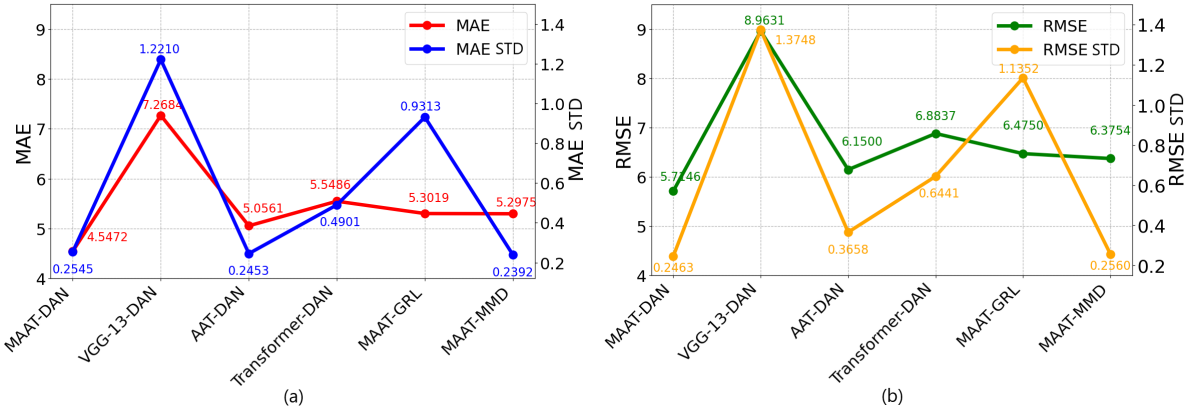


Figure 14: Comparative analysis of damage location prediction errors in the target domain: MAE (a), RMSE (b), and their STD from six methods across 10 training iterations.

The comparison results based on metric-4 are shown in Fig.15, where the error distribution is divided into seven bins for visualization. The mean and standard deviation of the error distribution are calculated directly from the combined coordinate errors. The results indicate that when the mean approaches 0, the model produces more frequent small errors. The STD reflects the variability of the error in the average predicted values across multiple training and testing iterations. A smaller STD suggests that most errors are close

to the mean, even if the mean represents a relatively large error. When both the mean and STD of the error distribution are small, it indicates that the errors are consistently near the true values, implying better prediction performance. Conversely, a large mean with a small STD indicates that errors are consistently far from the true values, reflecting poorer performance. By observing the error distribution, the occurrence frequency of various prediction errors can be visually observed. The highlighted $\pm$ 1σ region visually shows the range in which most prediction errors deviate from the true values. Fig.15 shows that MAAT-DAN has the mean error closest to 0 and the smallest error variability. In contrast, VGG-13-DAN exhibits the largest prediction bias and variability, indicating that VGG-13 is not ideal as a UTL feature extractor in datasets with many same labels and similar ultrasonic signal waveforms. Furthermore, after removing the multi-scale feature extraction and adaptive weight techniques, the changes in error mean and STD across different methods follow the same trend observed in metric-2 and metric-3, further demonstrating the advantages of the proposed MAAT feature extractor in terms of prediction accuracy and stability.

4.1.2. Domain adapters comparative analysis of regression task

Fig.9 (a), (e), and (f) illustrate the regression results from unsupervised DA using different domain adapters on the ultrasonic data presented in this paper. The comparison based on metric-1 and metric-2 demonstrates that MAAT-DAN significantly outperforms the other two UTL methods using different feature extractors. Specifically, the R^2 values of MAAT-GRL and MAAT-MMD are very similar (0.665 and 0.663, respectively), but MAAT-GRL exhibits much higher prediction variability compared to the other two methods, which aligns with the prediction instability of GRL discussed in Section 2.2.

The comparison results based on metric-3 are shown in Fig. 14. The MAAT-DAN method exhibits the lowest MAE and RMSE values compared to the other methods, with MAAT-GRL and MAAT-MMD showing similar results for both metrics. However, MAAT-GRL shows a significant increase in STD values. In contrast, MAAT-MMD demonstrates noticeably smaller STD values, indicating more stable error performance across multiple training and testing iterations. This further confirms the differences in stability between GRL and MMD. After combining GRL and MMD, MAAT-DAN achieves both lower errors and greater stability. These results validate that the proposed DA method combining GRL and MMD outperforms using either method alone, reducing GRL's training uncertainty and improving prediction accuracy.

The comparison results based on metric-4 are shown in Fig.15 (a), (e), and (f). MAAT-DAN exhibits the mean error closest to 0 and the smallest error variability among the three domain adapters. In contrast, the error distributions for MAAT-GRL and MAAT-MMD show that the absolute value of GRL's mean error (-1.57) is larger than MMD's (-0.65), indicating that MAAT-MMD produces smaller prediction errors than MAAT-GRL. This observation aligns with the trend seen in Fig.14 for MAE and RMSE. It is also worth noting that MAAT-GRL's error STD is smaller than MAAT-MMD's, which contrasts with the STD trends for MAE and RMSE. This difference arises because the STD values for MAE and RMSE reflect the variability of errors within each individual experiment, summarizing the overall variability across 10 trials. In contrast, the mean error in Fig.15 represents the variability in the averaged predictions across the 10 experiments, smoothing out any larger fluctuations from individual trials. Overall, MAAT-DAN, which combines the GRL and MMD domain adapters, demonstrates better performance compared to the other methods.

The comparison of the PDF curves constructed by Gaussian KDE for the 6 methods is shown in Fig. 16. VGG-13-DAN's error distribution is the furthest from 0 overall, while MAAT-DAN's error distribution is the closest to 0. MAAT-DAN's PDF encompasses the majority of the area covered by the methods using MAAT and AAT as feature extractors.

4.2. Results comparison and analysis of classification task

4.2.1. Feature extractors comparative analysis of classification task

Fig. 17 depicts the classification performance of UTL using different feature extractors and domain adapters based on water tank ultrasonic data. The results fall within the range of 55% to 100%, with the mean and STD of three performance metrics over 10 trials. Since the number of data items for the two categories (0 and 1) is relatively balanced, the values of the three evaluation metrics (accuracy, recall, and F1-score) are close, indicating that the model performs fairly across both categories without bias toward any specific one.

Specifically, the MAAT-DAN model outperforms the other models in terms of accuracy, recall, and F1-score, achieving performance close to 95%. The low STD across these three evaluation metrics indicates the model's classification stability. This suggests that MAAT-DAN has superior feature representation capabilities, especially showing significant advantages in distinguishing complex patterns damage identification.

In contrast, the performance of the VGG-13-DAN model is noticeably lower than that of the other models, with mean accuracy, recall, and F1-score around 68% and relatively larger STD. This indicates that the CNN-based feature extractor is insufficient for feature extraction in the transfer learning data used in this study,

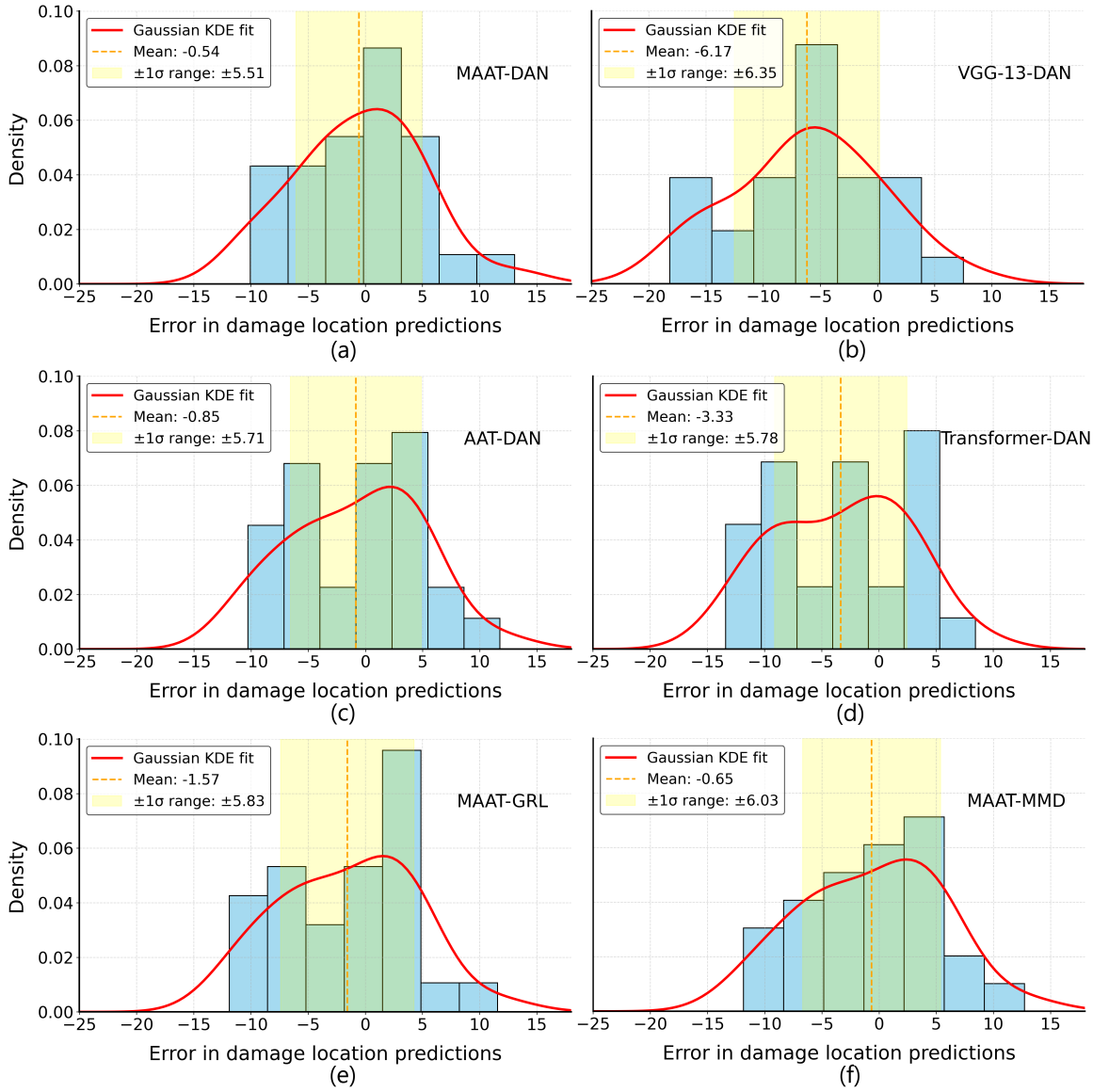


Figure 15: The error distribution of the combined x and y coordinates, based on the mean of 10 training predictions, is estimated using Gaussian KDE.

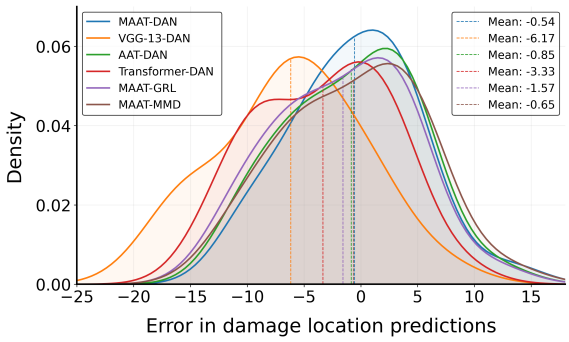


Figure 16: The PDF curves of the Gaussian KDE for the 6 methods are derived from the mean predicted values for the combined coordinates x and y over the 10 trials. The mean of the error distribution is calculated directly from the errors in the combined coordinates.

which spans different years and sensor pairs, and where PCA analysis reveals significant differences in data distribution.

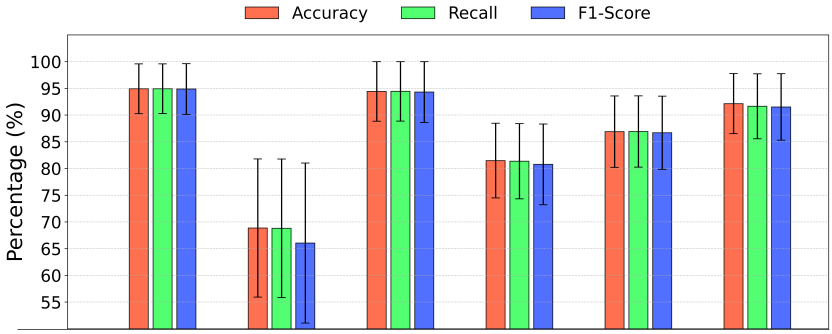
The AAT-DAN model, which removed multi-scale feature extraction, has mean evaluation metrics close to

those of MAAT-DAN. This suggests that multi-scale feature extraction provides relatively little benefit for the classification task. This may be because classification tasks typically focus on key features that distinguish different categories (such as signal peaks) or on the overall patterns in the data, and may not be highly sensitive to subtle changes in scale.

Furthermore, the mean values of the three evaluation metrics for the Transformer-DAN model—without the adaptive weighted multi-head attention mechanism—decrease significantly. This reflects that the adaptive weighted multi-head attention mechanism has a significant positive effect on the classification task.

4.2.2. Domain adapters comparative analysis of classification task

We conducted classification tasks using MAAT-GRL and MAAT-MMD separately and compared their results with MAAT-DAN to assess the performance of domain adapters that combine MMD and GRL. As shown in Fig. 17, both methods achieved average scores exceeding 86% across three evaluation metrics. Notably, MAAT-MMD had slightly better mean and STD than MAAT-GRL, indicating improved model stability and generalization. This outcome is similar to what we observed in regression tasks with these methods. Comparing these classic domain adaptation methods to MAAT-DAN, which combines both approaches revealed that MAAT-DAN enhanced classification performance and stability. This suggests that integrating GRL and MMD for DA helps improve the model's robustness and classification accuracy.



Metrics	MAAT-DAN	VGG-13-DAN	AAT-DAN	Transformer-DAN	MAAT-GRL	MAAT-MMD
Accuracy	**94.92**±4.66	68.80±12.93	94.43±6.44	80.16±6.45	86.89±6.68	92.13±5.62
Recall	**94.92**±4.65	68.80±12.96	94.43±6.38	80.02±6.52	86.91±6.67	91.64±6.09
F1-Score	**94.86**±4.74	66.04±14.97	94.31±6.67	79.32±7.00	86.69±6.85	91.52±6.23

Figure 17: Performance comparison of six structural health state classification models: Accuracy, Recall, and F1-Score across 10 trials.

4.3. Results comparison and analysis of computational complexity

Although MAAT-DAN demonstrates the highest performance metrics in both regression and classification tasks and confirms the beneficial contributions of different modules, it is also at the cost of high computational complexity for transfer learning model pre-training. Such complexity, arising from the integration of multiple modules, typically leads to longer training times, particularly for large-scale data.

In most cases, minimizing training time is an important objective in data-driven methods, as it directly reflects improved efficiency. This validation procedure indicates that MAAT-DAN (4-5 minutes) requires less training time than the traditional CNN-based VGG-13-DAN (5-6 minutes) in both tasks, suggesting that MAAT-DAN offers a potential computational advantage. Moreover, MAAT-DAN's training time is comparable to that of models employing a single domain adaptation technique (MAAT-GRL or MAAT-MMD). The AAT-DAN and Transformer-DAN, which utilize fewer feature extraction modules, exhibit notably lower training times (2.5 and 2 minutes, respectively). This indicates that the majority of the training time is spent on effective feature extraction.

The multi-scale feature extraction enhances overall model performance in both tasks; however, it also introduces higher computational complexity and longer training time, as it extracts three times more features from the same dataset. In practical applications, it is necessary to balance computation time and performance based on specific requirements. Since multi-scale feature extraction techniques require extended training time, we propose the following optimization strategies for different scenarios:

(1) For scenarios that prioritize high performance and have sufficient time, full multi-scale feature extraction—which includes down-sampling, original sampling, and up-sampling—is suggested. This strategy is particularly suitable for SHM of valuable historical buildings and artifacts, pre-launch integrity assessments of rockets and launch vehicles, and routine inspections of nuclear reactor pressure vessels, where longer training times are acceptable to achieve higher precision.

(2) For scenarios requiring both high performance and less training time, either down-sampling or up-sampling in combination with original sampling is suggested. This strategy is particularly advantageous in environments where frequent retraining is necessary due to continuously evolving data distributions or operational conditions. For instance, in aerospace system monitoring during flight, sensor data can exhibit rapid shifts that necessitate timely model recalibration, while in the SHM of large-scale infrastructures like dams, the ability to quickly update the model is critical for preventing potential failures. In these cases, minor performance trade-offs may be necessary in specific scenarios to achieve an optimal balance between precision and efficiency.

(3) For applications where minimizing training time is the primary constraint and a slight reduction in performance is acceptable, the AAT-DAN model is suggested. For instance, in systems such as automated fire alarm systems and smoke detectors—where models may require frequent retraining to adapt to evolving conditions and where occasional false alarms are an acceptable trade-off for rapid update cycles—AAT-DAN offers a compelling solution. AAT-DAN achieves the second-best performance after MAAT-DAN in both tasks while considerably reducing training time, making it an optimal choice for time-sensitive applications and large-scale data.

4.4. Results comparison and analysis: Baseline without transfer learning and high noise validation

4.4.1. Results comparison and analysis without transfer learning

After analyzing the performance of different modules in MAAT-DAN on regression and classification tasks, we also evaluated the model's performance on the target domain data without the domain adapter module (without transfer learning) to further highlight the necessity of transfer learning. Without the domain adapter, the model is trained using the same source domain data and tested on the same target domain test dataset. All remaining aspects of the model setup—including the network architecture (comprising the MAAT feature extractor and label predictor), hyperparameters, data preprocessing workflow, optimizer settings, and training environment—are kept consistent for MAAT-DAN. This uniformity ensures that any observed differences in performance are solely attributable to the presence or absence of the domain adapter module, thereby guaranteeing a fair comparison between the scenarios with and without transfer learning. The averaged performance metrics over 10 training runs for both tasks are summarized in Table 2. For the regression task, the R^2 value on the target domain test is negative ($R^2 = -0.1477$), significantly worse than any method employing the domain adapter module. This indicates that the model's predictions are even less effective than using the averaged target value as a prediction. Additionally, the averaged MAE and RMSE over 10 runs are considerably higher than those of any other method discussed in this paper, suggesting that without transfer learning, the model fails to accurately predict damage locations in reinforced aluminum plates.

A similar result is observed in the classification task. Across 10 training trials, the accuracy is only slightly above 50%, approaching the random guessing baseline for binary classification (50%). The recall and F1-scores are the lowest among all methods, with relatively high STD for all three classification metrics. These results indicate that without the domain adapter module, the model is unable to classify the target domain data effectively. This further emphasizes the necessity of transfer learning for both regression and classification datasets used in this paper.

Table 2: Averaged performance metrics of regression and classification tasks without transfer learning across 10 training iterations.

Regression task			**Classification task**		
R^2	MAE	RMSE	Accuracy	Recall	F1-Score
-0.1477	8.6593 ± 0.4122	11.6233 ± 0.4472	53.11 ± 7.45	52.58 ± 7.74	37.66 ± 12.13

4.4.2. Results comparison and analysis with high noise validation

Due to the continuous nature of regression outputs, noise interference can directly induce deviations in predictions, making regression tasks inherently more sensitive to noise than classification tasks [88, 89]. Therefore, this section evaluates MAAT-DAN's regression performance under high-noise conditions to simulate potential real-world noisy environments and assess the model's robustness. White Gaussian noise at a Signal-to-Noise Ratio (SNR) of 5 dB and 10 dB is introduced into the target domain ultrasonic signals of a reinforced aluminum plate with a carbon fiber stiffener. Fig.18 presents a comparative analysis of signals across three noise levels.

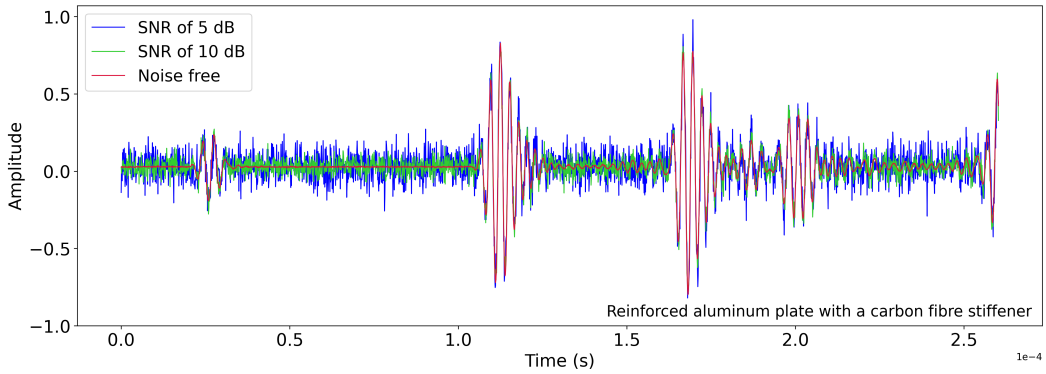


Figure 18: Ultrasonic signals at three noise levels for a reinforced aluminum plate with a carbon fiber stiffener.

Note that introducing noise in the target domain is consistent with the principles of transfer learning, which aims to enable the model to generalize effectively from the source to the target domain. Target domain data typically characterize real-world environments, requiring models to adapt to varying noise conditions without retraining on the source dataset. In contrast, high-quality source domain data can typically be collected under controlled experimental conditions with minimal noise interference. For each noise level, ten independent training runs are conducted for the regression task, and the mean performance metrics are evaluated. The results averaged over these runs are summarized in Table 3.

Table 3: Comparison of R^2, MAE, and RMSE across three noise levels of MAAT-DAN.

SNR	R^2	MAE	RMSE
Noise-free	0.719	4.5472 ± 0.2545	5.7146 ± 0.2463
10dB	0.645	4.9999 ± 0.2715	6.8529 ± 0.2433
5dB	0.598	5.5432 ± 0.2650	7.2881 ± 0.2894

The validation results of MAAT-DAN under high-noise conditions indicate that while model errors exhibit a slight increase under 5 dB and 10 dB noise conditions, the R^2 values remain relatively high. Especially in comparison to methods that do not apply any TF, such as VGG-13-DAN ($R^2 = 0.270$). This suggests the model's capability to extract meaningful signal features despite high noise interference. These results further validate the model's reliability and robustness for deployment in noisy environments. Specifically, comparing Fig. 9 and Fig. 14, at a noise level of 10 dB SNR, the three performance metrics—R^2, MAE, and RMSE—remain close to those of AAT-DAN under noise-free conditions, although the MAE is slightly higher. Notably, the MAAT-DAN still exhibits better performance compared to the Transformer-DAN under noise-free conditions. Similarly, at 5 dB noise levels, the model maintains better performance over the Transformer-DAN under noise-free conditions, except in RMSE. This further validates the effectiveness of the multi-scale feature extraction technique for processing high-noise environments and also reveals the overall robustness of MAAT-DAN under noisy conditions.

5. Conclusions and discussion

This paper proposes an advanced UTL method named MAAT-DAN, designed for damage identification in both regression and classification tasks based on two ultrasonic data types. The method introduces a novel transfer learning architecture by innovatively optimizing two key components of transfer learning: the feature extractor and the domain adapter. Specifically, to address the issue that the multi-head attention mechanism in transformer architectures cannot dynamically adjust the influence of each head based on specific input conditions, a dynamic adaptive weighting module is introduced to form the AAT. This enables the model to adjust the weights of each attention head in real-time according to the current input data, more flexibly enhancing important attention heads and suppressing less significant ones. Building on this, multi-scale feature extraction techniques are employed to perform multi-dimensional feature extraction of ultrasonic signals. These multi-scale feature extraction techniques are then combined with AAT to form a new feature extractor called MAAT.

On the other hand, in terms of the domain adapter, the advantages of the GRL in handling high-dimensional data and the stability of MMD are combined. A scheme is proposed where both GRL and MMD DA mech-

anisms work together to align distributions and enhance domain-invariant features. During the DA process, reconstruction loss is also utilized to learn feature representations that are effective for both domains.

The proposed MAAT-DAN method is validated in both UTL regression and classification tasks using two specially designed ultrasonic experiments: a reinforced aluminum plate for regression validation and a steel water tank for classification validation. Compared with five other methods, MAAT-DAN demonstrates superiority in both its feature extractor and domain adapter modules across both tasks. The main contributions and conclusions of this work are as follows:

- A novel unsupervised DA transfer learning method, MAAT-DAN, is proposed. This method introduces a new DA scheme that combines multi-scale feature extraction and weight-adaptive transformers as feature extractors. Additionally, it integrates GRL and MMD as domain adapters, supported by feature reconstruction assistance.

- To validate the regression performance of the MAAT-DAN ultrasonic experiment, the damage is simulated at various locations using Blue-tack on a structure. The structure comprises an aluminum plate with beam stiffeners individually bonded using two different materials. Regression UTL employs ultrasonic signals from the structure with aluminum-bonded beams as the source domain. Transfer learning is applied to predict damage locations (X and Y) in the target domain with carbon-bonded stiffeners.

- Regression performance is compared with five other UTL methods and one method without transfer learning from feature extractors' perspectives and domain adapters' perspectives, while additional evaluations under high-noise conditions assess the model's robustness. Each method individually demonstrated advantages in multi-scale feature extraction techniques, weight-adaptive techniques, and the combined use of GRL and MMD techniques. The feature extractor based on the Transformer architecture significantly outperformed the VGG-13-based extractor. In domain adapters, Among domain adapters, GRL exhibited prediction errors similar to MMD but with higher STD. Furthermore, under high-noise conditions, MAAT-DAN demonstrated its robustness to noise interference. As a result of these advantages, MAAT-DAN achieved the highest damage location prediction R^2 value of 0.719, the lowest MAE of 4.5472, the lowest RMSE of 5.7146, and low STD for both MAE and RMSE.

- Another ultrasonic experiment on a steel water tank is conducted for classification tasks across different years (2012 to 2013) and sensor pairs (pair-1-2 to pair-3-4). Damage reflection signals at local signal positions are simulated using a sine wave with a Hanning window. Transfer learning is applied from the source domain data of 2012 pair-1-2 to the target domain data of 2013 pair-3-4 for damage identification classification.

- Classification performance validation uses the same comparison strategy as regression validation. The comparison results confirmed the necessity of transfer learning for this dataset. The weight-adaptive technique provided more significant gains for the classification task than the multi-scale feature extraction approach. Additionally, GRL techniques also exhibited higher STD in classification tasks. Ultimately, MAAT-DAN achieved the highest classification accuracy value of around 95%, and the most consistent prediction stability compared to the other methods.

This paper demonstrates the viability and efficacy of the innovative MAAT-DAN framework for UTL in both regression and classification tasks. The evaluation process involves two custom-designed ultrasonic experiments and a comprehensive comparison of various feature extractors and domain adapters. Performance assessment is conducted using four regression metrics and three classification metrics, providing a thorough validation of the framework's capabilities.

Although the performance of MAAT-DAN has been demonstrated, the flexibility and scalability of the proposed UTL framework suggest promising future development directions that warrant further discussion.

Levels of SHM: While the current MAAT-DAN framework has been successfully applied to both damage localization and classification, its highly scalable architecture offers even broader potential SHM applications. By customizing the task label predictor to match specific data requirements, MAAT-DAN can efficiently address additional challenges such as the classification of multiple types of damage, severity assessment, and lifespan prediction. This flexibility underscores MAAT-DAN's capability to support diverse SHM objectives, including continuous monitoring and optimizing data needed to maintain acceptable accuracy, paving the way for more comprehensive and robust SHM solutions.

Further Optimisation: Although MAAT-DAN has demonstrated excellent performance across different modules, there remains room for further optimization in future research. Three potential areas for improvement are summarized as follows:

(1) Adaptive weight allocation: While the transformer architecture proposed in this paper, incorporating adaptive weights (averaged over all sequence positions and embedding dimensions), has effectively reduced prediction errors compared to the original transformer model, further research could explore more advanced adaptive weight allocation strategies. For instance, exploring techniques like weighted summation or multilayer perceptrons has the potential to further refine the adaptive weight allocation process, helping to retain more valuable information and minimize information loss.

(2) Fine-tuning adapters: When partial labels are available in the target domain data, small adapter modules can be inserted between the layers of the pre-trained model. Two potential adapter locations are worth further investigation: a. A fine-tuning adapter is inserted after the residual connection of the multi-head attention layer (before layer normalization). This enables adapter modules of task-specific features while effectively integrating the attention output and original information, thereby potentially improving the model's adaptability to the target task while retaining pre-trained knowledge. b. A fine-tuning adapter is inserted after the gradient reversal layer (before the domain classifier). This potentially helps better capture and correct distribution differences between the source and target domains, further enhancing domain adaptation performance. Fine-tuning these adapters with the labeled data could improve the model's adaptability to different tasks or real-world engineering applications across domains.

(3) Hyperparameter optimization: In application scenarios that require high-precision predictions, optimizing the hyperparameters of the MAAT-DAN model proposed in this paper based on input data can help further improve the model's prediction accuracy, enabling it to meet the demands of higher-precision applications.

Acknowledgement

The authors acknowledge KU Leuven (STG/21/008) for their financial support.

References

1. Shao, W., Sun, H., Wang, Y., Qing, X.. A multi-level damage classification technique of aircraft plate structures using lamb wave-based deep transfer learning network. *Smart Materials and Structures* 2022;31(7):075019.
2. Lu, H., Chandran, B., Wu, W., Ninic, J., Gryllias, K., Chronopoulos, D.. Damage features for structural health monitoring based on ultrasonic lamb waves: Evaluation criteria, survey of recent work and outlook. *Measurement* 2024;:114666.
3. Farrar, C.R., Worden, K.. Structural health monitoring: a machine learning perspective. John Wiley & Sons; 2012.
4. Balageas, D., Fritzen, C.P., Güemes, A.. Structural health monitoring; vol. 90. John Wiley & Sons; 2010.
5. Yazdi, M.. Reliability-centered design and system resilience. In: *Advances in Computational Mathematics for Industrial System Reliability and Maintainability*. Springer; 2024:79–103.
6. Malik, M.K., Chronopoulos, D., Tanner, G.. Transient ultrasonic guided wave simulation in layered composite structures using a hybrid wave and finite element scheme. *Composite Structures* 2020;246:112376.
7. Lu, H., Farrokhabadi, A., Mardanshahi, A., Rauf, A., Talemi, R., Gryllias, K., Chronopoulos, D.. Uncertainty quantification for damage detection in 3d-printed auxetic structures using ultrasonic guided waves and a probabilistic neural network. *Thin-Walled Structures* 2024;:112466.
8. Aranguren, G., Monje, P., Cokonaj, V., Barrera, E., Ruiz, M.. Ultrasonic wave-based structural health monitoring embedded instrument. *Review of Scientific Instruments* 2013;84(12).
9. Mitra, M., Gopalakrishnan, S.. Guided wave based structural health monitoring: A review. *Smart Materials and Structures* 2016;25(5):053001.
10. Watkins, R., Jha, R.. A modified time reversal method for lamb wave based diagnostics of composite structures. *Mechanical Systems and Signal Processing* 2012;31:345–354.
11. Tanveer, M., Elahi, M.U., Jung, J., Azad, M.M., Khalid, S., Kim, H.S.. Recent advancements in guided ultrasonic waves for structural health monitoring of composite structures. *Applied Sciences* 2024;14(23):11091.
12. Liu, C., Xu, X., Wu, J., Zhu, H., Wang, C.. Deep transfer learning-based damage detection of composite structures by fusing monitoring data with physical mechanism. *Engineering Applications of Artificial Intelligence* 2023;123:106245.
13. Xu, K., Ta, D., Su, Z., Wang, W.. Transmission analysis of ultrasonic lamb mode conversion in a plate with partial-thickness notch. *Ultrasonics* 2014;54(1):395–401.
14. Lu, H., Cantero-Chinchilla, S., Yang, X., Gryllias, K., Chronopoulos, D.. Deep learning uncertainty quantification for ultrasonic damage identification in composite structures. *Composite Structures* 2024;338:118087.
15. Peng, D., Liu, C., Desmet, W., Gryllias, K.. Attention-guided cross-layer feature fusion convolutional neural network for vibration signal denoising. In: *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON NOISE AND VIBRATION ENGINEERING (ISMA2020)/INTERNATIONAL CONFERENCE ON UNCERTAINTY IN STRUCTURAL DYNAMICS (USD2020)*. KATHOLIEKE UNIV LEUVEN, DEPT WERKTUIGKUNDE; 2020:3563–3577.
16. Hassani, S., Dackermann, U., Mousavi, M., Li, J.. A systematic review of data fusion techniques for optimized structural health monitoring. *Information Fusion* 2023;:102136.
17. Cantero-Chinchilla, S., Chiachío, J., Chiachío, M., Chronopoulos, D., Jones, A.. A robust bayesian methodology for damage localization in plate-like structures using ultrasonic guided-waves. *Mechanical Systems and Signal Processing* 2019;122:192–205.
18. Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., Inman, D.J.. A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications. *Mechanical systems and signal processing* 2021;147:107077.
19. Vaswani, A.. Attention is all you need. *Advances in Neural Information Processing Systems* 2017;.

20. Zhang, H., Lin, J., Hua, J., Zhang, T., Tong, T.. Attention-based interpretable prototypical network towards small-sample damage identification using ultrasonic guided waves. *Mechanical Systems and Signal Processing* 2023;188:109990.
21. Ma, J., Hu, M., Yang, Z., Yang, H., Ma, S., Xu, H., Yang, L., Wu, Z.. An efficient lightweight deep-learning approach for guided lamb wave-based damage detection in composite structures. *Applied Sciences* 2023;13(8):5022.
22. Liao, S., Liu, H., Yang, J., Ge, Y.. A channel-spatial-temporal attention-based network for vibration-based damage detection. *Information Sciences* 2022;606:213–229.
23. Wang, X., Chen, Z., Sun, W., Shao, N., You, Z., Xu, J., Yan, R.. A small sample piezoelectric impedance-based structural damage identification using signal reshaping-based enhance attention transformer. *Mechanical Systems and Signal Processing* 2024;208:111067.
24. Hu, J., Shen, L., Sun, G.. Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018:7132–7141.
25. Wang, H., Liu, Z., Peng, D., Yang, M., Qin, Y.. Feature-level attention-guided multitask cnn for fault diagnosis and working conditions identification of rolling bearing. *IEEE transactions on neural networks and learning systems* 2021;33(9):4757–4769.
26. Cantero-Chinchilla, S., Wilcox, P.D., Croxford, A.J.. Deep learning in automated ultrasonic nde–developments, axioms and opportunities. *NDT & E International* 2022;131:102703.
27. Giglioni, V., Poole, J., Mills, R., Venanzi, I., Ubertini, F., Worden, K.. Transfer learning in bridge monitoring: Laboratory study on domain adaptation for population-based shm of multispan continuous girder bridges. *Mechanical Systems and Signal Processing* 2025;224:112151.
28. Minotti, D.. Exploring transfer learning for enhanced performance in ultrasonic guided waves-based damage diagnosis frameworks 2022;.
29. Song, Y., Li, J., Gao, P., Li, L., Tian, T., Tian, J.. Two-stage cross-modality transfer learning method for military-civilian sar ship recognition. *IEEE Geoscience and Remote Sensing Letters* 2022;19:1–5.
30. Azad, M.M., Kim, S., Cheon, Y.B., Kim, H.S.. Intelligent structural health monitoring of composite structures using machine learning, deep learning, and transfer learning: a review. *Advanced Composite Materials* 2024;33(2):162–188.
31. Weiss, K., Khoshgoftaar, T.M., Wang, D.. A survey of transfer learning. *Journal of Big data* 2016;3:1–40.
32. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.. Adversarial discriminative domain adaptation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017:7167–7176.
33. Yang, Q.. An introduction to transfer learning. In: *ADMA*. Springer; 2008: 1.
34. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:14123474* 2014;.
35. Zellinger, W., Grubinger, T., Lughofer, E., Natschläger, T., Saminger-Platz, S.. Central moment discrepancy (cmd) for domain-invariant representation learning. *arXiv preprint arXiv:170208811* 2017;.
36. Long, M., Cao, Y., Wang, J., Jordan, M.. Learning transferable features with deep adaptation networks. In: *International conference on machine learning*. PMLR; 2015:97–105.
37. Chi, J., Mao, Z.. Deep domain-adversarial anomaly detection with robust one-class transfer learning. *Knowledge-Based Systems* 2024;300:112225.
38. Mao, W., Wang, G., Kou, L., Liang, X.. Deep domain-adversarial anomaly detection with one-class transfer learning. *IEEE/CAA Journal of Automatica Sinica* 2023;10(2):524–546.
39. Tzeng, E., Hoffman, J., Darrell, T., Saenko, K.. Simultaneous deep transfer across domains and tasks. In: *Proceedings of the IEEE international conference on computer vision*. 2015:4068–4076.
40. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V.. Domain-adversarial training of neural networks. *Journal of machine learning research* 2016;17(59):1–35.
41. Almahakeri, M., Al-Mahasneh, A.J., Mallouh, M.A., Jouda, B.. Deep neural network-based intelligent health monitoring system for oil and gas pipelines. *Applied Soft Computing* 2025;:112827.
42. Zhang, B., Hong, X., Liu, Y.. Distribution adaptation deep transfer learning method for cross-structure health monitoring using guided waves. *Structural Health Monitoring* 2022;21(3):853–871.
43. Hong, X., Yang, D., Huang, L., Zhang, B., Jin, G.. Vibration-adaption deep convolutional transfer learning method for stranded wire structural health monitoring using guided wave. *IEEE Transactions on Instrumentation and Measurement* 2022;72:1–10.
44. Mao, J., Wang, H., Spencer Jr, B.F.. Toward data anomaly detection for automated structural health monitoring: Exploiting generative adversarial nets and autoencoders. *Structural Health Monitoring* 2021;20(4):1609–1626.
45. Doan, D.D., Ramasso, E., Placet, V., Zhang, S., Boubakar, L., Zerhouni, N.. An unsupervised pattern recognition approach for ae data originating from fatigue tests on polymer–composite materials. *Mechanical Systems and Signal Processing* 2015;64:465–478.
46. Chen, X., Wang, S., Wang, J., Long, M.. Representation subspace distance for domain adaptation regression. In: *ICML*. 2021:1749–1759.
47. Taghiyarrenani, Z., Nowaczyk, S., Pashami, S., Bouguelia, M.R.. Multi-domain adaptation for regression under conditional distribution shift. *Expert systems with applications* 2023;224:119907.
48. De Mathelin, A., Richard, G., Deheeger, F., Mougeot, M., Vayatis, N.. Adversarial weighting for domain adaptation in regression. In: *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE; 2021:49–56.
49. Singh, A., Chakraborty, S.. Deep domain adaptation for regression. *Development and Analysis of Deep Learning Architectures* 2020;:91–115.
50. Cortes, C., Mohri, M.. Domain adaptation and sample bias correction theory and algorithm for regression. *Theoretical Computer Science* 2014;519:103–126.
51. Niu, Z., Zhong, G., Yu, H.. A review on the attention mechanism of deep learning. *Neurocomputing* 2021;452:48–62.
52. Choi, S.R., Lee, M.. Transformer architecture and attention mechanisms in genome data analysis: a comprehensive review. *Biology* 2023;12(7):1033.
53. Legendre, S., Massicotte, D., Goyette, J., Bose, T.K.. Wavelet-transform-based method of analysis for lamb-wave ultrasonic nde signals. *IEEE Transactions on Instrumentation and Measurement* 2000;49(3):524–530.
54. Liao, Y., Huang, Y., Wang, Y., Wu, M., Zhang, F., Qing, X.. Complex cfrp damage detection with dual multiscale 1dcnns and self-selective fusion technique of guided waves. *Composite Structures* 2024;335:117993.
55. Du, B., Liu, Z., Luo, F.. Deep multi-scale attention network for rna-binding proteins prediction. *Information Sciences* 2022;582:287–301.

56. Li, X., Zhang, W., Ding, Q.. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliability engineering & system safety* 2019;182:208–218.
57. Neverova, N., Wolf, C., Taylor, G.W., Nebout, F.. Multi-scale deep learning for gesture detection and localization. In: *Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13*. Springer; 2015:474–490.
58. Hu, C., Xu, J., Li, Y.. Slight looseness detection of reinforcing bar's threaded sleeve connections using convolutional neural network trained by magnetostrictive guided wave signals. *Journal of Nondestructive Evaluation* 2021;40:1–8.
59. Yao, T., Pan, Y., Li, Y., Ngo, C.W., Mei, T.. Wave-vit: Unifying wavelet and transformers for visual representation learning. In: *European Conference on Computer Vision*. Springer; 2022:328–345.
60. Hesse, R., Schaub-Meyer, S., Roth, S.. Content-adaptive downsampling in convolutional neural networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023:4544–4553.
61. Lee, H., Lim, S.. Pu-mfa: Point cloud up-sampling via multi-scale features attention. *Sensors* 2022;22(23):9308.
62. Hu, K., Li, M., Xia, M., Lin, H.. Multi-scale feature aggregation network for water area segmentation. *Remote Sensing* 2022;14(1):206.
63. He, K., Zhang, X., Ren, S., Sun, J.. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016:770–778.
64. Gal, Y., Ghahramani, Z.. A theoretically grounded application of dropout in recurrent neural networks. *Advances in neural information processing systems* 2016;29.
65. Lee, C.Y., Zhuo, G.L.. Identifying bearing faults using multiscale residual attention and multichannel neural network. *IEEE Access* 2023;11:26953–26963.
66. Yang, J., Zou, H., Zhou, Y., Xie, L.. Towards stable and comprehensive domain alignment: Max-margin domain-adversarial training. *arXiv preprint arXiv:200313249* 2020;.
67. Zhang, Y., Wang, X., Liang, J., Zhang, Z., Wang, L., Jin, R., Tan, T.. Free lunch for domain adversarial training: Environment label smoothing. *arXiv preprint arXiv:230200194* 2023;.
68. Dai, Z., Yang, J., Fan, A., Jia, J., Chen, J.. Sadd: Generative adversarial networks via self-attention and dual discriminator in unsupervised domain adaptation. In: *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. Springer; 2023:473–484.
69. Reddi, S., Ramdas, A., Póczos, B., Singh, A., Wasserman, L.. Kernel mmd the median heuristic and distance correlation in high dimensions. *stat* 2014;1050:9.
70. Ramdas, A., Reddi, S.J., Poczos, B., Singh, A., Wasserman, L.. On the high-dimensional power of linear-time kernel two-sample testing under mean-difference alternatives. *arXiv preprint arXiv:14116314* 2014;.
71. Borgwardt, K.M., Gretton, A., Rasch, M.J., Kriegel, H.P., Schölkopf, B., Smola, A.J.. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics* 2006;22(14):e49–e57.
72. Schrab, A., Kim, I., Albert, M., Laurent, B., Guedj, B., Gretton, A.. Mmd aggregated two-sample test. *Journal of Machine Learning Research* 2023;24(194):1–81.
73. Yang, B., Lei, Y., Jia, F., Li, N., Du, Z.. A polynomial kernel induced distance metric to improve deep transfer learning for fault diagnosis of machines. *IEEE Transactions on Industrial Electronics* 2019;67(11):9747–9757.
74. Wang, Y., Hu, S., Wu, S.. Object tracking based on huber loss function. *The Visual Computer* 2019;35:1641–1654.
75. Meyer, G.P.. An alternative probabilistic interpretation of the huber loss. In: *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*. 2021:5261–5269.
76. Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.. A kernel two-sample test. *The Journal of Machine Learning Research* 2012;13(1):723–773.
77. Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., Erhan, D.. Domain separation networks. *Advances in neural information processing systems* 2016;29.
78. Osumi, K., Yamashita, T., Fujiyoshi, H.. Domain adaptation using a gradient reversal layer with instance weighting. In: *2019 16th International Conference on Machine Vision Applications (MVA)*. IEEE; 2019:1–5.
79. Cui, R., Azuara, G., Lanza di Scalea, F., Barrera, E.. Damage imaging in skin-stringer composite aircraft panel by ultrasonic-guided waves using deep learning with convolutional neural network. *Structural Health Monitoring* 2022;21(3):1123–1138.
80. Feng, T., Sharif Khodaei, Z., Aliabadi, M.F.. Influence of composite thickness on ultrasonic guided wave propagation for damage detection. *Sensors* 2022;22(20):7799.
81. Azuara, G., Ruiz, M., Barrera, E.. Damage localization in composite plates using wavelet transform and 2-d convolutional neural networks. *Sensors* 2021;21(17):5825.
82. Dang, D.Z., Lai, C.C., Ni, Y.Q., Zhao, Q., Su, B., Zhou, Q.F.. Image classification-based defect detection of railway tracks using fiber bragg grating ultrasonic sensors. *Applied Sciences* 2022;13(1):384.
83. Simonyan, K., Zisserman, A.. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:14091556* 2014;.
84. Tu, X.L., Pyle, R.J., Croxford, A.J., Wilcox, P.D.. Potential and limitations of narx for defect detection in guided wave signals. *Structural Health Monitoring* 2023;22(3):1863–1875.
85. Wang, K., Zhang, J., Shen, Y., Karkera, B., Croxford, A.J., Wilcox, P.D.. Defect detection in guided wave signals using nonlinear autoregressive exogenous method. *Structural Health Monitoring* 2022;21(3):1012–1030.
86. Zou, H., Hastie, T., Tibshirani, R.. Sparse principal component analysis. *Journal of computational and graphical statistics* 2006;15(2):265–286.
87. Azadi, S., Okabe, Y., Carvelli, V.. Bayesian-optimized 1d-cnn for delamination classification in cfrp laminates using raw ultrasonic guided waves. *Composites Science and Technology* 2025;:111101.
88. Tang, H., Liu, L., Li, Z., Jia, Z., Zhang, Q.. Reducing influence of noise in regression problems on memristor computing-in-memory chips. *IEEE Access* 2025;.
89. Muthukumar, V., Narang, A., Subramanian, V., Belkin, M., Hsu, D., Sahai, A.. Classification vs regression in overparameterized regimes: Does the loss function matter? *Journal of Machine Learning Research* 2021;22(222):1–69.