A Dual-Contrastive-Attention Transformer for Unsupervised Lamb Wave Defect Detection

Jiawei Guo[1], Boshi Chen[1], Sen Zheng[2], Nikta Amiri[3], Ge Song[1], Lingyu Yu[1], Yi Wang[1]*

[1] *Department of Mechanical Engineering, University of South Carolina, Columbia, SC, 29208*

jiaweig@email.sc.edu; boshi@email.sc.edu; gsong@email.sc.edu; yu3@cec.sc.edu
*yiwang@cec.sc.edu (*corresponding author)*

[2]*Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, 21205*

szhan236@jh.edu

[3]*Department of Mechanical Engineering, Alfred University, NY, 14802*

amiri@alfred.edu

ABSTRACT

Lamb wave-based Structural Health Monitoring (SHM) is a promising technique for detecting defects in materials and structures. However, traditional methods often rely on computationally intensive signal processing and struggle to detect subtle anomalies wave patterns. In this work, we propose a novel transformer-based framework, called Dual-Contrastive-Attention Transformer (DCAT), for unsupervised anomaly detection in Lamb wave data. DCAT uses two attention branches during training: a Global-Context Attention (GCA) branch that captures long-range patterns, and a Local-Context Attention (LCA) branch that serves as a constraint. A contrastive loss is used to prevent the global branch from over-learning local features, encouraging it to focus on the overall structure. Both branches are trained to reconstruct the input, using a structural similarity (SSIM) loss that better reflects waveform patterns than traditional mean squared error. After training, only the global branch is retained for inference. Anomalies are detected by comparing the input and reconstructed output. Since the global branch cannot easily reproduce local defects, it produces a higher SSIM loss when anomalies are present. We test our model on a Lamb wave dataset with multiple types of defects. DCAT achieves 97.8% accuracy and a precision of 98.6%, outperforming other SOTA baselines. These results show that DCAT is well-suited for accurate Lamb wave-based SHM without the need for labeled data.

1. INTRODUCTION

The SHM plays a crucial role in ensuring the safety and longevity of aerospace and civil infrastructures by enabling early detection of structural defects. Among various nondestructive evaluation (NDE) techniques, ultrasonic guided waves, particularly Lamb waves, are widely used due to their ability to propagate over long distances and sensitivity to both surface and internal defects (Zhao et al., 2011). Lamb waves propagate within structural boundaries and are particularly responsive to thickness discontinuities, allowing efficient inspection of large-scale structures (Alleyne & Cawley, 1992). When interacting with defects, Lamb waves generate unique spatiotemporal patterns that serve as "wave fingerprints" for identifying the presence and characteristics of damage (Giurgiutiu, 2005). Recent studies have applied Lamb wave methods in practical SHM systems, particularly within the PHM Society community. For example, Cantero-Chinchilla et al. (2018) developed a highly efficient Lamb wave–based damage indicator for plate-like structures, relying on baseline comparisons and cumulative damage factors to detect and track delamination. Additionally, Mishra et al. (2015) proposed a multivariate cumulative sum (CUSUM) technique for Lamb-wave sensor data, enabling online detection of damage progression in composite structures. These efforts highlight the increasing viability of Lamb wave–based SHM for practical engineering applications. A typical Lamb wave inspection system comprises an actuator and a receiver; damage is inferred by comparing received signals to defect-free baselines. To automate this comparison, machine learning techniques have increasingly been employed.

Convolutional Neural Networks (CNNs) have proven highly effective for extracting spatial features from wave images,

improving SHM and NDT capabilities. Recent work has improved CNNs further by incorporating wavelet transforms and attention mechanisms (Zhao, 2022). For instance, the Wavelet-Attention CNN (WA-CNN) applies Discrete Wavelet Transform (DWT) to decompose feature maps into low- and high-frequency components, then directs attention mechanisms to the high-frequency parts to enhance detail sensitivity while maintaining low-frequency structure. This technique has improved classification performance on benchmark datasets such as CIFAR-10 and CIFAR-100. Similarly, the Multi-level Wavelet CNN (MWCNN) uses wavelet decomposition within a U-Net structure (Liu et al., 2018), effectively balancing receptive field size and computational complexity in image restoration tasks like denoising and super-resolution. In other domains, WaDeNet directly integrates wavelet decomposition into CNNs for speech signal analysis, capturing both temporal and spectral patterns and improving non-invasive emotion recognition (Suresh & Ragav, 2020).

Recurrent Neural Networks (RNNs), including Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) networks, have also been successfully applied to Lamb wave–based SHM because of their ability to model temporal dynamics. These models are adept at capturing complex wave propagation effects, which are critical for detecting time-dependent anomalies. For example, Azad et al. (2024) developed an LSTM-based model for localizing and quantifying damage severity in composite structures using Lamb wave data, demonstrating high accuracy across various operational conditions. Likewise, Zhang et al. (2020) reported that GRU models outperformed traditional techniques in identifying subtle damage signatures, confirming the robustness of RNNs for modeling dispersive waveforms.

More recently, Transformer architectures have been adopted in SHM due to their superior global feature modeling capabilities. Self-attention mechanisms in Transformers effectively capture long-range dependencies in time-series data, which traditional RNNs may miss. Ding et al. (2022) proposed a time-frequency Transformer architecture that achieved significant accuracy improvements in rolling bearing fault diagnosis. However, most Transformer-based models require large amounts of labeled anomaly data for training, limiting their application in real-world SHM scenarios where defect data is scarce. To address this, Wang et al. (2023) introduced the Defect Transformer (DefT), which combines CNNs and Transformers to detect surface defects in complex environments more efficiently. Despite these advances, reducing reliance on labeled data remains a major challenge.

Unsupervised learning, especially with autoencoders (AEs), provides a promising solution to this issue. AEs are typically trained on normal Lamb wave data and used to reconstruct those signals. Anomalies manifest as reconstruction errors when the model fails to recreate signals that deviate from the learned normal patterns. For instance, Rizvi et al. (2024) proposed a Bi-LSTM autoencoder augmented with Maximal Overlap Discrete Wavelet Transform (MODWT) to improve detection and localization of structural anomalies in composites. Similarly, Lee et al. (2022) used a deep autoencoder to automatically classify fatigue damage in composite materials via Lamb wave inputs. These studies demonstrate that integrating components such as LSTM and wavelet transforms into AEs architectures can significantly enhance anomaly discrimination.

Despite these advances, existing AEs models for Lamb wave anomaly detection often struggle to detect subtle or structurally similar defects. These models typically rely on reconstruction loss, which only becomes significant when the anomaly deviates clearly from normal patterns, a condition not always met in practice. Moreover, they are often overfit to local features and underutilize global information, allowing even anomalous signals to be reconstructed accurately and thus reducing detection effectiveness.

To address these limitations, this study proposes a self-supervised framework named Dual-Contrastive-Attention Transformer (DCAT). The model integrates Global-Context Attention (GCA) and Local-Context Attention (LCA) mechanisms and introduces a contrastive learning strategy to guide the encoder toward learning globally consistent representations. By enforcing structural similarity-based reconstruction and suppressing over-reliance on local context, the model enhances its sensitivity to detect anomaly while requiring no labeled data during training.

The rest of this paper is organized as follows: Section 2 introduces transformer-based anomaly detection methods and discusses their limitations. Section 3 describes the proposed DCAT framework in detail. Section 4 outlines the implementation, experimental setup, and results. Section 5 concludes the paper.

2. PRELIMINARIES

The section describes the preliminaries of the Transformer-based anomaly detection. Section 2.1 outlines the standard architecture and reconstruction-based detection paradigm. Section 2.2 discusses the limitations of these methods, particularly in the context of guided wave signals for SHM.

2.1. Transformer-based Anomaly Detection

Transformer-based models have gained popularity in anomaly detection due to their strong capacity to capture long-range dependencies. In reconstruction-based settings, these models are trained on normal data and tasked with reconstructing the input. Anomalies are identified by comparing the input and output signals, with reconstruction errors serving as indicators of abnormality. This type of methods have been applied in various domains including

industrial inspection, video surveillance, time-series analysis, and more recently, structural health monitoring (SHM) using guided waves.

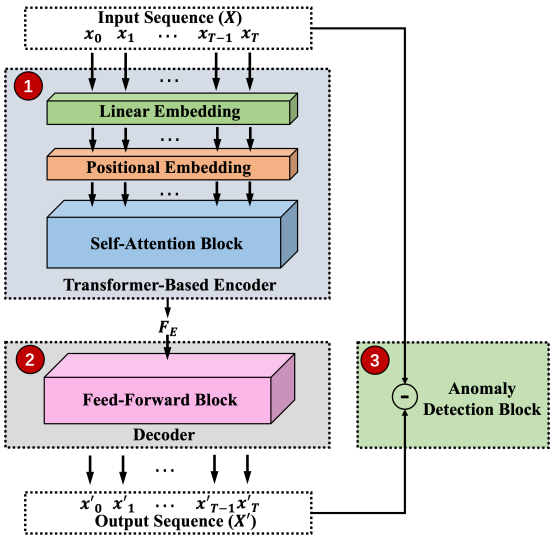


Figure 1. The pipeline of Transformer-based Anomaly Detection

1. The first part of the architecture, as illustrated in Figure 1, is an encoder composed of three stages. The input waveform $X = \{x_0, x_1, \dots, x_T\}$ is first passed through a linear embedding layer to project it into a high-dimensional latent space. Positional encoding is then added to preserve the sequential order and relative distance of signal components, which is critical for modeling spatiotemporal dependencies. The resulting sequence is processed by multi-head self-attention layers, enabling the model to capture global context across the entire input. This produces a feature representation F_T, which is then supplied to the decoder for reconstruction. The encoder process can be formally represented as the following equation:

$$F_E = f_{en}(X) \quad (1)$$

2. The second part is the decoder, typically implemented as a feedforward block that maps the latent feature representation F_E back to the original signal space and outputs a reconstruction waveform $X' = \{x'_0, x'_2, \dots, x'_T\}$. This process can be formulated as:

$$X' = f_{de}(F_E) \quad (2)$$

   Since the model is trained exclusively on normal data, it becomes proficient at reconstructing normal signal patterns. However, when the input contains anomalous components that deviate from the training distribution, the decoder struggles to reproduce them accurately.

3. The third part is the anomaly detection module, which leverages the difference between the input and the reconstructed output. The reconstruction errors — quantified as the difference between the original input X and the reconstructed signal X'. These errors are then aggregated into an anomaly score map that highlights regions where the reconstruction fails to align with the input. High reconstruction errors are interpreted as potential indicators of structural defects.

This approach enables unsupervised detection, as the model does not require labeled anomalies during training, avoids the need for manually labeled defects.

2.2. Issues in Transformer-based Method

While Transformer-based reconstruction models show strong performance on many anomaly detection tasks, they also suffer from several limitations, particularly in the context of guided wave-based structural inspection.

First, the self-attention mechanism is designed to capture global relationships across the entire input. While effective for modeling long-range dependencies, it tends to suppress localized variations. As a result, subtle or spatially small anomalies may be overlooked, as the model learns to enforce global consistency and smooth out local disruptions.

Second, the high model capacity of Transformer decoders can lead to over-generalization. When trained exclusively on normal data, the model may still be capable of reconstructing abnormal inputs with low error due to its strong representation power. This reduces the gap in reconstruction error between normal and anomalous cases, which compromises the sensitivity and reliability of detection.

Lastly, in low signal-to-noise environments common to Lamb wave applications, Transformers may misinterpret noise as structural variation, or vice versa. Without additional constraints or anomaly-aware learning mechanisms, their ability to separate signal from defect-related patterns can be compromised.

These limitations highlight the need for architectures that can better capture local anomaly patterns while maintaining global context understanding, motivating the design of our proposed DCAT framework.

3. PROPOSED DUAL-CONTRASTIVE-ATTENTION TRANSFORMER (DCAT)

To address the limitations of existing Transformer-based anomaly detection methods, this work proposes a Dual-Contrastive-Attention Transformer (DCAT) tailored for unsupervised Lamb wave defect detection, and the architecture is designed to enhance sensitivity to local anomalies. DCAT builds upon the reconstruction-based detection framework but introduces novel attention mechanisms and training strategies to improve anomaly localization.

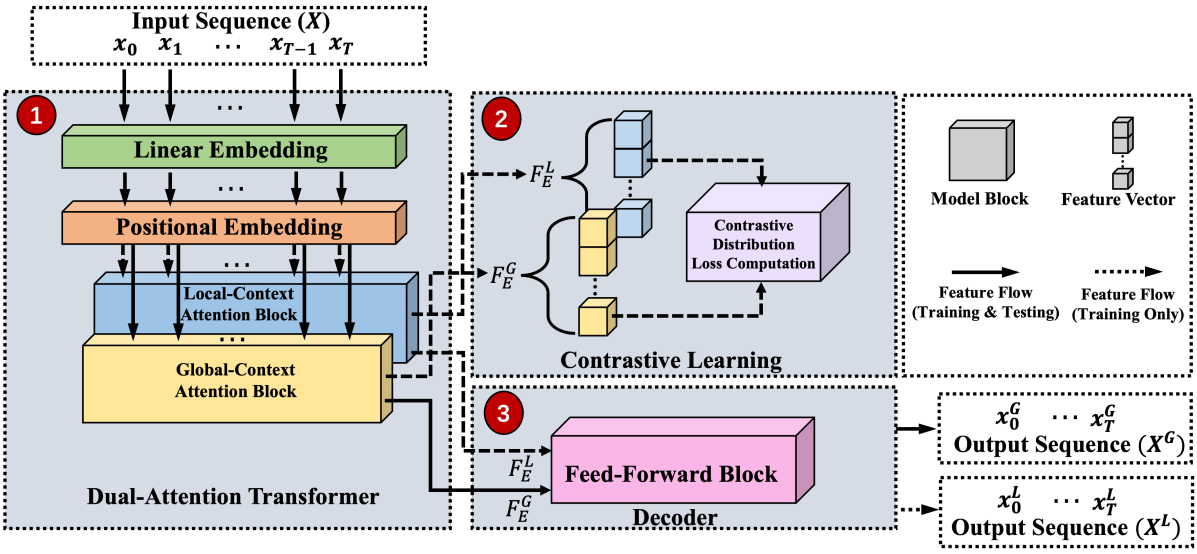


Figure 2. The Proposed Dual-Attention Contrastive Transformer Framework

As shown in Figure 2, the proposed framework DCAT consists of a dual-attention Transformer encoder, contrastive learning, and a decoder, forming a complete pipeline for unsupervised anomaly detection. During training, the encoder adopts a dual-attention structure, consisting of a primary Global-Context Attention (GCA) branch and an auxiliary Local-Context Attention (LCA) branch. The global branch uses the long-range global features to reconstruct the spatial continuity of normal wavefields, which typically exhibit globally consistent patterns. In contrast, anomalies usually appear only in small areas and don't follow global patterns. If the model focuses too much on local details, it may reconstruct both normal and abnormal signals well—making it hard to distinguish them. To address this, the LCA branch is introduced as a training-time constraint: it extracts local features, and a contrastive loss is imposed to penalize the GCA branch when its output features are too close to the LCA's features. This encourages the GCA branch to focus on global structural patterns and avoid overfitting of local details, ensuring that anomalies are poorly reconstructed, while normal patterns are well reconstructed. This amplifies the distinction between normal and abnormal inputs. Each main components in this architecture will be detailed in the following section.

3.1. Dual-Attention Transformer

In this section, the dual-attention transformer structure will be detailed. A LCA branch is introduced during training to constrain the GCA branch, and then the GCA is used for reconstruction.

3.1.1. Local-Context Attention (LCA)

The LCA module is designed to restrict the receptive field of the self-attention mechanism to a small neighborhood around each token. Unlike native self-attention, which allows each query to access the entire sequence information, our proposed method limits attention to a local window based on Euclidean distance. The restriction forces the model to extract only localized patterns and prevent it from incorporating long-range context. The LCA module serves as an auxiliary component for regularizing GCA branch and is only enabled during training. To achieve this, it applies a locally constrained self-attention mechanism, where each token attends only to its temporal neighbors. The attention is computed as:

$$\begin{aligned} Q(t) &= f_t W^Q \\ K(t) &= f_t W^K \\ V(t) &= f_t W^V \\ Attention\big(Q(t), K(t), V(t)\big) &\\ = \left(softmax\left(\frac{Q(t)K(t)^T}{\sqrt{d_k}} \odot M(t)\right) V(t)\right) & \end{aligned} \tag{3}$$

Here, the input feature f_t represents the embedded signal at the t^{th} time instance, derived from the positional embedding as shown in Figure 2. And W^Q, W^K, W^V, similar to the original self-attention, are learnable projection matrices, and d_k is the key dimension. The $\odot$ denotes element-wise multiplication. The attention mask $M(i) \in [0,1]$ introduces soft local constraints based on Euclidean distance. For each query t, the mask element $M(t)$ is defined as:

$$M(t) = \exp\left(-\frac{\|p_t - p_j\|_2}{L}\right) \tag{4}$$

where p_t and p_j are the temporal position of token t and j, L is the total length of the input sequence. This exponential formulation ensures that closer tokens receive higher attention weights, while more distant tokens are

exponentially suppressed. It smoothly encourages the model to focus on nearby regions, without using any hard or sudden cutoffs.

3.1.2. Global-Context Attention (GCA)

The GCA serves as the primary encoder for signal reconstruction in the DCAT framework. Unlike the LCA branch, which is limited to neighborhood-level information, GCA attends to the full input sequence, allowing it to capture long-range dependencies and structural continuity across the entire wavefield. This global perspective is particularly well-suited for reconstructing normal waveforms, which typically follow consistent and structured propagation patterns. As a result, the global attention mechanism can effectively reconstruct normal inputs without relying on too much localized information.

The self-attention in GCA follows the standard transformer formulation. Given the embedded feature f_t at time step t, the global self-attention is computed as:

$$\begin{gathered} Q(t) = f_t W^Q \\ K(t) = f_t W^K \\ V(t) = f_t W^V \\ Attention\big(Q(t), K(t), V(t)\big) \\ = (softmax\left(\frac{Q(t)K(t)^T}{\sqrt{d_k}}\right)V(t)) \end{gathered} \tag{5}$$

The formulation follows the standard transformer attention structure, where the projection weights and dimensional terms are as defined in **Section 3.1.1**.

To avoid overfitting to local anomalies, the global attention branch is regularized during training using a contrastive constraint with the LCA branch. The detailed contrastive formulation is described in the following section.

3.2. Contrastive Learning

To prevent the GCA branch from overusing local information during training, we introduce a contrastive regularization based on distribution-level dissimilarity. Specifically, we apply a Kullback-Leibler (KL) divergence loss between the feature maps generated by the GCA and LCA branches:

$$\mathcal{L}_{KL} = D_{KL}(F_E^G || F_E^L) = F_E^G \log\left(\frac{F_E^G}{F_E^L}\right) \tag{6}$$

Here, F_E^G and F_E^L represent the feature maps (treated as distributions) from the GCA and LCA, as Figure 2 shows, respectively. Unlike typical loss terms that are minimized, our objective is to maximize this divergence. A higher KL value reflects a greater divergence between global and local representations, indicating that the GCA branch emphasizes long-range structural features instead of relying on local information.

This constraint is only active during training. After training, both the LCA branch and this regularization are removed, leaving the GCA encoder for inference.

3.3. Decoder

To ensure the feature maps extracted by both the GCA and LCA branches are meaningful and reconstruction-relevant, we pass their outputs F_E^G and F_E^L to the decoder. The decoder is a feedforward network that transforms the feature maps back into the original waveform. By applying the reconstruction loss to both the global and local feature maps during training, it is ensured that both branches learn useful and meaningful features instead of random or low-quality outputs.

To measure reconstruction quality, we adopt the Structural Similarity Index Measure (SSIM) instead of the commonly used Mean Squared Error (MSE). SSIM is more appropriate for waveform data, as it evaluates structural similarity in terms of wave pattern, rather than point-wise numerical accuracy. The SSIM-based reconstruction loss is defined as:

$$\mathcal{L}_{rec} = 1 - SSIM(X, X') \tag{7}$$

The SSIM index ranges from 0 to 1, where 1 indicates perfect structural similarity between the input and reconstructed signal. Since we aim to maximize similarity, we use the loss in the form of $1 - SSIM$, so that lower values of the loss correspond to better reconstructions.

After training, the LCA branch is discarded, and only the GCA encoder and decoder are used during inference. Anomaly detection is then performed by comparing the SSIM between the input and the reconstructed output.

Because the GCA branch is trained only on normal data and penalized for using local patterns, it learns to focus on global structure. But anomalies usually do not follow any global patterns, therefore the trained model fails to reconstruct them well. Thus, the SSIM loss is much higher when an anomaly is present.

4. EXPERIMENTAL SETUP AND RESULT

All experiments were implemented using PyTorch 1.13.1 and conducted on a workstation equipped with an NVIDIA RTX A6000 GPU, which features 10,752 CUDA cores and 48 GB of GDDR6 memory. Each input sample consists of a temporal sequence of 100 steps ($T = 100$). The DCAT architecture used for training includes 3 transformer layers, each with 8 self-attention heads. The dimensionality of the query, key, and value vectors is fixed at 512. A batch size of 32 and a learning rate of 0.005 are used for all training runs. Anomaly detection was based on (1 – SSIM) reconstruction loss, with a threshold of 0.42 selected for DCAT based on validation. For all baselines, the best-performing thresholds were used.

4.1. Dataset collection & Pre-process

To generate data for model development and evaluation, a non-invasive Lamb wave inspection system, as shown in Figure 3, was established using a Scanning Laser Doppler Vibrometer (SLDV). The SLDV excites and measures wave signals on the plate surface. When defects such as notches or attachments are present, they cause noticeable distortions in the wave propagation patterns. And the system was applied to a stainless-steel plate with the dimensions $310\text{mm} \times 310\text{mm} \times 1\text{mm}$. The guided Lamb waves were excited at 120 kHz and measured along radial scan lines.

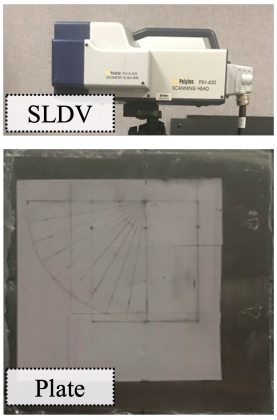
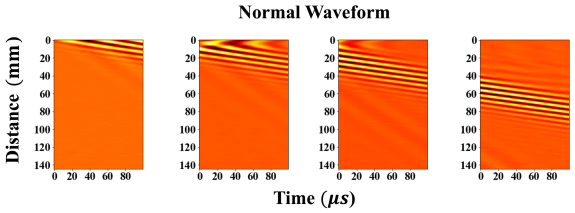
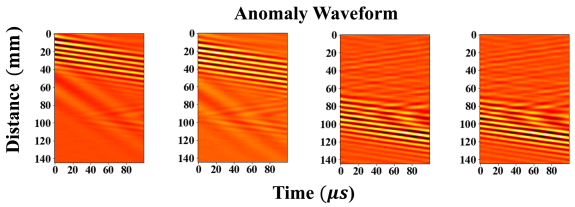


Figure 3. Inspection System

Defects were introduced in the form of a surface notch to produce abnormal wave propagation patterns as shown in Figure 4. It compares normal and anomalous wavefields. Subfigure (a) shows a typical wave pattern from an undamaged plate, while subfigure (b) shows the altered pattern caused by structural defects.



(a)



(b)

Figure 4. Example of Normal & Anomaly Waveform

The collected spatiotemporal wavefield data form two-dimensional representations of Lamb wave propagation over both time and space. To effectively train and evaluate the proposed model, each full measurement is further divided into smaller patches, as visualized in Figure 4, using the sliding window technique. Each patch captures a short temporal evolution of wave propagation over the spatial scan line. Specifically, the x-axis corresponds to time (i.e., the evolution of the wave signal), while the y-axis represents spatial locations along the scan line (i.e., distances from the wave source). The length of each path is 100 along the x-axis and 145 along the y-axis, matching the number of spatial sampling points from the SLDV scan. The temporal window size of 100 is chosen to balance the need to capture meaningful wave patterns while maintaining computational efficiency.

The preprocessed patches were subsequently split into two subsets: the training dataset and the testing dataset. The training dataset consists of 2,000 patches that reflect only normal wave samples for model training purposes. On the other hand, the testing dataset includes 600 samples in total, where the number of wave patches representing normal and abnormal cases is 450 and 150, respectively.

4.2. Performance Metrics

The detection performance of the proposed model is quantitatively measured by four main evaluation metrics: Accuracy, Precision, Recall, and F_1-score. The value of accuracy indicates the overall rate at which the model produces correct predictions, including both normal and abnormal outcomes. Precision is the percentage of predicted anomalies that are real defects. Recall represents the ratio of detected anomalies to all the abnormal samples in the ground truth. F_1-score provides a balanced evaluation by combining precision and recall into a single metric. It is especially useful when the data is imbalanced, as it highlights the trade-off between missing anomalies and incorrectly predicted normal instances. The associated equations are defined below:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$F_1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{11}$$

where TP, FP, TN, and FN represents the value of true positive, false positive, true negative, and false negative, respectively. In this work, anomalies are defined as positive instances, while normal samples are considered negative. A larger value in each of these metrics indicates stronger detection performance.

4.3. Result

This section presents the experimental results of our proposed DCAT model. We first compare its performance with other state-of-the-art (SOTA) baselines to evaluate its effectiveness in detecting Lamb wave anomalies. And the ablation study is also presented to demonstrate the contribution of key component in DCAT framework.

4.3.1. Comparison with SOTA

In the following section, the performance of the proposed DCAT model is evaluated and compared against a range of representative anomaly detection baselines. These include LSTM-AE (Wong et al., 2022), DACAD (Darban et al., 2025), ACR-DSVDD(Li et al., 2024), LUNAR (Goodge et al., 2022), and DADA(Shentu et al., 2024), each representing a different modeling strategy widely used in Lamb wave or sequential anomaly detection. All methods are trained and tested under the same preprocessing and evaluation setup.

As shown in Table 1, DCAT achieves the best overall performance across the board, with 97.8% accuracy, 98.6% precision, 92.7% recall, and an F_1-score of 95.6%. These results highlight the effectiveness of DCAT's dual-attention design and contrastive regularization, which together improve sensitivity to anomalies.

Compared to DADA, which achieves a competitive F_1-score of 91.8% and recall 92.6%, DCAT shows much higher precision. DADA's dual-decoder structure lacks explicit attention control, making it less reliable in isolating anomaly-relevant features. LSTM-AE performs the worst in recall (35.3%) due to its limited temporal modeling and lack of spatial awareness. ACR-DSVDD and DACAD rely on feature compactness and adversarial learning, but being non-reconstruction-based, they struggle to capture full complex spatiotemporal patterns. LUNAR integrates channel and temporal attention but lacks DCAT's contrastive design, resulting in weaker anomaly separation and lower recall.

Model	*Accuracy (%)*↑	*Precision (%)*↑	*Recall (%)*↑	F_1 *Score (%)*↑
DCAT (Proposed)	**97.8**	**98.6**	**92.7**	**95.6**
LSTM-AE	78.5	62.4	35.3	45.1
DACAD	93.3	90.4	82.0	86.0
ACR-DSVDD	92.1	86.5	81.3	83.9
LUNAR	90.9	84.2	78.0	80.0
DADA	95.8	90.9	92.6	91.8

Table 1. Comparison Result with Benchmark Models

4.3.2. Ablation Analysis

To assess the contribution of each component in the DCAT framework, we conduct an ablation study by selectively removing the LCA and GCA branches. The results are summarized in Table 2.

When the GCA is removed and only the LCA is used, the model shows a drastic drop in recall (20.0%). This is because the local branch only sees a narrow spatial window and lacks global awareness, allowing it to reconstruct both normal and anomalous signals well, thus failing to distinguish anomalies. A low recall indicates that the model is not sensitive to defects.

On the other hand, when the LCA is removed and only the GCA is retained, the model becomes a plain Transformer autoencoder without any constraint. Although the GCA captures global features, without contrastive regularization it also tends to inordinately use local information, resulting in high-quality reconstructions even for anomalous inputs, which leads to poor recall (28.0%).

When both attention branches are removed, the model degrades into a feedforward convolutional autoencoder (FCN-AE). In this case, the overall reconstruction ability is weaker, which makes it slightly more sensitive to anomalies (recall improves to 37.3%), but still significantly worse than the full DCAT model with structured attention and contrastive learning.

These results confirm that both global and local attention, as well as the interaction between them, are critical to achieving strong anomaly detection performance.

Component		*Accuracy (%)*↑	*Precision (%)*↑	*Recall (%)*↑	F_1 *Score* ↑
LCA	GCA				
✓	✓	**97.8**	**98.6**	**92.7**	**95.6**
✓	✗	79.2	85.7	20.0	32.4
✗	✓	80.7	84.0	28.0	42.0
✗	✗	83.7	76.7	37.3	50.2

Table 2. Ablation Results

5. CONCLUSION

This study introduces DCAT, a transformer-based unsupervised framework developed for Lamb wave anomaly detection. DCAT integrates GCA and LCA mechanisms with contrastive learning and SSIM-guided reconstruction to mitigate drawbacks of traditional autoencoder and transformer models, that is, anomalies are also reconstructed well. DCAT demonstrates salient performance in identifying subtle and structurally similar anomalies. Future work will explore adaptation to multi-sensor configurations, real-time deployment, and improve the robustness under noisy conditions.

ACKNOWLEDGEMENT

The authors are grateful for the financial support from the Department of Energy Nuclear Energy University Program under the grant numbers DE-NE0008959 and EPSCoR grant number DE-SC0025538.

REFERENCES

Alleyne, D. N., & Cawley, P. (1992). The interaction of Lamb waves with defects. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, *39*(3), 381–397.

Azad, M. M., Munyaneza, O., Jung, J., Sohn, J. W., Han, J.-W., & Kim, H. S. (2024). Damage Localization and Severity Assessment in Composite Structures Using Deep Learning Based on Lamb Waves. *Sensors*, *24*(24), 8057.

Cantero-Chinchilla, S., Chiachío-Ruano, J., Chiachío-Ruano, M., Etxaniz, J., Aranguren, G., Jones, A., Essa, Y., & Martin De La Escalera, F. (2018). Lamb wave-based damage indicator for plate-like structures. *European Conference of the PHM Society*, *4*(1).

Darban, Z. Z., Yang, Y., Webb, G. I., Aggarwal, C. C., Wen, Q., Pan, S., & Salehi, M. (2025). DACAD: Domain adaptation contrastive learning for anomaly detection in multivariate time series. *IEEE Transactions on Knowledge and Data Engineering*.

Ding, Y., Jia, M., Miao, Q., & Cao, Y. (2022). A novel time–frequency Transformer based on self–attention mechanism and its application in fault diagnosis of rolling bearings. *Mechanical Systems and Signal Processing*, *168*, 108616.

Giurgiutiu, V. (2005). Tuned Lamb wave excitation and detection with piezoelectric wafer active sensors for structural health monitoring. *Journal of Intelligent Material Systems and Structures*, *16*(4), 291–305.

Goodge, A., Hooi, B., Ng, S.-K., & Ng, W. S. (2022). Lunar: Unifying local outlier detection methods via graph neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(6), 6737–6745.

Lee, H., Lim, H. J., Skinner, T., Chattopadhyay, A., & Hall, A. (2022). Automated fatigue damage detection and classification technique for composite structures using Lamb waves and deep autoencoder. *Mechanical Systems and Signal Processing*, *163*, 108148.

Li, A., Qiu, C., Kloft, M., Smyth, P., Rudolph, M., & Mandt, S. (2024). Zero-shot anomaly detection via batch normalization. *Advances in Neural Information Processing Systems*, *36*.

Liu, P., Zhang, H., Zhang, K., Lin, L., & Zuo, W. (2018). Multi-level wavelet-CNN for image restoration. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 773–782.

Mishra, S., Vanli, O. A., & Park, C. (2015). A multivariate cumulative sum method for continuous damage monitoring with lamb-wave sensors. *International Journal of Prognostics and Health Management*, *6*(4).

Rizvi, S. H. M., Abbas, M., Zaidi, S. S. H., Tayyab, M., & Malik, A. (2024). LSTM-Based Autoencoder with Maximal Overlap Discrete Wavelet Transforms Using Lamb Wave for Anomaly Detection in Composites. *Applied Sciences*, *14*(7), 2925.

Shentu, Q., Li, B., Zhao, K., Shu, Y., Rao, Z., Pan, L., Yang, B., & Guo, C. (2024). Towards a General Time Series Anomaly Detector with Adaptive Bottlenecks and Dual Adversarial Decoders. *ArXiv Preprint ArXiv:2405.15273*.

Suresh, P., & Ragav, A. (2020). WaDeNet: Wavelet Decomposition based CNN for Speech Processing. *ArXiv Preprint ArXiv:2011.05594*.

Wang, J., Xu, G., Yan, F., Wang, J., & Wang, Z. (2023). Defect transformer: An efficient hybrid transformer

architecture for surface defect detection. *Measurement*, *211*, 112614.

Wong, L., Liu, D., Berti-Equille, L., Alnegheimish, S., & Veeramachaneni, K. (2022). AER: Auto-encoder with regression for time series anomaly detection. *2022 IEEE International Conference on Big Data (Big Data)*, 1152–1161.

Zhang, Z., Pan, H., Wang, X., & Lin, Z. (2020). Machine learning-enriched lamb wave approaches for automated damage detection. *Sensors*, *20*(6), 1790.

Zhao, X. (2022). Wavelet-attention CNN for image classification. *ArXiv Preprint ArXiv:2201.09271*.

Zhao, X., Royer, R. L., Owens, S. E., & Rose, J. L. (2011). Ultrasonic Lamb wave tomography in structural health monitoring. *Smart Materials and Structures*, *20*(10), 105002.

BIOGRAPHIES

Jiawei Guo earned his B.S. from Tianjin University of Technology and Education (2019), Tianjin, China, and his M.S. from the University of Southern California, CA, USA. He is now pursuing his Ph.D. of mechanical engineering with the University of South Carolina. His research interests are in computer vision and machine learning for engineering applications.

Boshi Chen received his B.S. degree in Biomedical Engineering from Hefei University of Technology, Hefei, China, in 2023. He is currently pursuing a Ph.D. in Mechanical Engineering at the University of South Carolina. His research interests include non-destructive evaluation, robotics, and autonomous systems.

Sen Zhang received his B.S. degree in computer science with the University of South Carolina in 2023, and his M.S from Johns Hopkins University in 2025. His research interests include deep learning and machine learning algorithms and its application to various engineering fields

Nikta Amiri is an Assistant Professor of Mechanical Engineering at Alfred University. She earned her Ph.D. in Mechanical Engineering from the University at Buffalo in 2022, where her dissertation focused on Numerical and Analytical Modeling of Piezoelectric Transduction in Biomedical Applications. Following her doctoral studies, she completed a postdoctoral fellowship at the University of South Carolina (2022–2023), conducting experimental research in ultrasonics-based material evaluation and damage detection. Her research interests include smart materials, nondestructive evaluation, and transducer modeling for biomedical and structural applications.

Ge Song received his B.S. degree in Mechanical Engineering from the Nanjing University of Science and Technology, Nanjing, China in 2019, and M.S. degree from the Boston University in 2021. He is currently working towards the Ph. D. degree in the department of Mechanical Engineering with the University of South Carolina, Columbia, SC, USA. His research interests include computer vision and robotics and autonomous system.

Lingyu Yu 's research interests include structural health monitoring (SHM) and nondestructive evaluation (NDE) using ultrasonic guided waves supported by advanced sensor technology, theoretical modeling, signal processing and data analysis. Applications of this type of research include condition monitoring of the aircraft structures and systems, integrity monitoring of large composite structures, detecting and quantifying cracks in metal structures and their subsequent propagation, corrosion growth monitoring in piping structures, and many more.

Yi Wang earned his B.S. and M.S. from Shanghai Jiao Tong University (1998, 2000), and his Ph.D. from Carnegie Mellon University (2005). Currently, he is a Professor at the University of South Carolina. His research focuses on computational and data-enabled science and engineering, multi-fidelity surrogate modeling, machine learning, computer vision, and autonomous systems.