A Masked Autoencoder-Based Novel Deep Learning Framework of Delamination Detection Using Improved HHT-Based Signal Representations

SUMIT SAHA[1], STEVEN LINFORTH[1], NANDAKISHOR DESAI[2] and TUAN NGO[1]

ABSTRACT

Ensuring the longevity and operational efficiency of composite structures demands highly effective structural health monitoring (SHM) strategies, particularly for detecting delamination, a critical defect that severely compromises structural integrity. Guided ultrasonic wave (GUW) methods have proven particularly adept at identifying the presence, location, and extent of delamination. Despite these advantages, implementing conventional, supervised deep learning (DL) approaches for GUW-based SHM has been constrained by the scarcity of high-quality, labelled training data and the high computational expense of generating representative damage scenarios. To overcome these limitations, this study introduces an unsupervised DL framework for delamination detection that incorporates a masked autoencoder (MAE) architecture. Firstly, one-dimensional time-series signals from GUW experiments are converted into two-dimensional time-frequency representations using an enhanced Hilbert-Huang Transform, which is chosen for its superior time-frequency resolution and computational efficiency over wavelet-based approaches. In the subsequent phase, the MAE framework is trained exclusively on these wavefield representations corresponding to undamaged conditions, allowing it to excel at reconstructing masked segments. The model is then presented with new, unlabeled data that includes both undamaged and damaged signals. If the input data significantly diverges from the patterns learned during training, the model's reconstruction of masked regions deteriorates, resulting in higher reconstruction error that serves as an indicator of potential damage. A benchmark of an experimental dataset from the open guided waves platform was used to validate this framework. This dataset encompasses carbon fibre-reinforced polymer plates with 28 distinct damage scenarios. The framework identifies delamination damage through root mean squared reconstruction errors and adaptive thresholds and has been found to outperform state-of-the-art DL architectures, including conventional autoencoders.

[1]Department of Infrastructure Engineering, University of Melbourne, Australia
[2]Department of Electrical Engineering, University of Melbourne, Australia

INTRODUCTION

Composite materials, such as carbon fiber-reinforced polymers (CFRPs) [1], are extensively used across the civil, aerospace, wind energy, and automotive sectors due to their exceptional strength-to-weight ratios and outstanding fatigue resistance. However, these structures are susceptible to various failure mechanisms, including delamination [2] and fatigue [3], with delamination representing a particularly critical concern. If undetected, delamination can severely compromise structural integrity, potentially resulting in catastrophic failure. Delamination in structural components often progresses in extent and severity with aging, particularly accelerating toward the end of the service life. Owing to its typical subsurface nature, it remains undetectable through conventional visual inspection. Structural health monitoring (SHM) offers a promising solution to mitigate the associated loss of residual strength, thereby preventing catastrophic failures and safeguarding both human life and infrastructure.

SHM systems utilise distributed sensor networks and advanced signal processing techniques to enable real-time condition assessment while optimising maintenance strategies and reducing operational costs. Among various SHM approaches, ultrasonic guided waves (UGWs) [4,5] are particularly advantageous due to their capability for rapid, large-area inspection, high sensitivity to internal defects, and cost-efficient deployment. The efficacy of any UGW-based SHM system relies on the integration of accurate wave propagation modelling, high-resolution sensing, sophisticated signal interpretation, and robust damage diagnostics.

Two modelling approaches have been explored to enhance the effectiveness of the UGW-based SHM systems: physics-based and data-driven. Physics-based methods such as FEM and analytical models simulate wave propagation but struggle with complex geometries, environmental changes, and material damping. As an alternative, data-driven approaches leverage experimental or simulated datasets to build predictive models without requiring explicit knowledge of the underlying physics. Among data-driven techniques, machine learning methods have gained significant attention, with deep learning (DL) approaches [6] proving particularly promising due to their ability to extract features and learn patterns automatically without explicit physical modeling. Although previously constrained by limited data availability, the rise of big data and Industry 4.0 [7] has opened new opportunities for DL-enabled SHM.

However, current DL approaches for SHM face significant challenges. Supervised deep learning models [8] require large, labeled datasets for delamination detection, while unsupervised autoencoders (AEs) [9] rely on healthy data but struggle to capture long-range dependencies in time-frequency data. To address these limitations, self-supervised learning (SSL) has emerged as a promising alternative. In particular, vision transformer-based Masked Autoencoders (MAEs) [10–12] overcome these limitations by learning robust, high-level representations from unlabeled data through masked reconstruction tasks. This offers superior capacity to model in a global context compared to standard AEs, making MAEs well-suited for data-scarce scenarios common in structural health monitoring applications.

However, the effectiveness of these SSL approaches depends on the appropriate representation of input sensor data. Guided Lamb wave signals, captured using

piezoelectric transducers or ultrasonic sensors, require advanced time-frequency analysis techniques such as the Hilbert-Huang Transform (HHT) [13,14] to effectively characterise their inherently non-stationary and nonlinear behaviour. While HHT is well-suited for processing large data volumes in wideband damage detection and provides a detailed representation of Lamb wave responses, conventional HHT spectrograms face several limitations, including spurious low-frequency components, broad initial modes, and inadequate separation of low-energy signals. To overcome these challenges, Peng et al. [15] proposed an enhanced HHT approach that integrates Wavelet Packet Transform (WPT) for narrowband decomposition before Empirical Mode Decomposition, significantly improving spectral resolution and signal separation.

This study introduces a novel framework for detecting delamination damage in composite structures, combining enhanced signal processing technique with a novel SSL-based DL approach. Our work makes three key contributions to SHM. First, we introduce a self-supervised learning framework tailored for practical SHM scenarios, where healthy-state data is abundant while damaged data is limited due to the cost and complexity of experimental acquisition. Second, we employ an improved HHT technique, integrating Wavelet Packet Transform (WPT) and Empirical Mode Decomposition (EMD), to convert one-dimensional guided wave signals into informative two-dimensional time-frequency representations. Third, we utilize these spectrograms as input to a Vision Transformer (ViT)-based MAE, which is trained to reconstruct masked regions of the input, enabling the model to learn robust spatial features from unlabeled baseline data. The proposed MAE-based self-supervised framework, as shown in Figure 1, is validated on a guided wave dataset containing 28 delamination scenarios in CFRP laminates, and it outperforms existing autoencoder-based methods in damage detection accuracy.

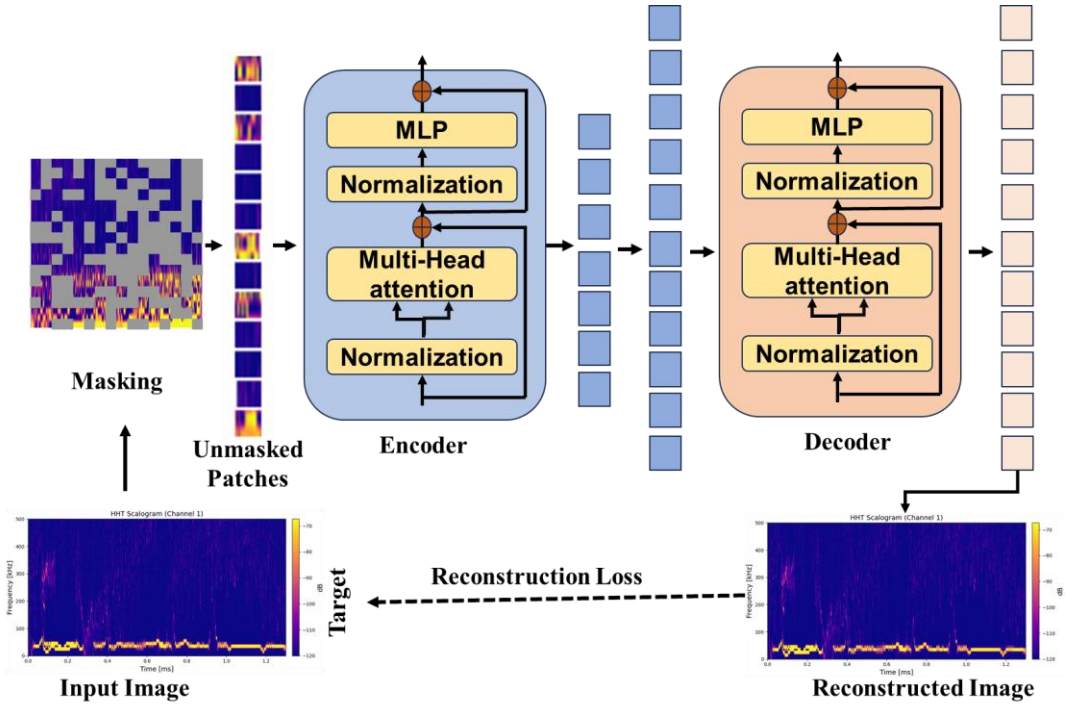

Figure 1. Overview of the MAE framework using HHT spectrograms.

BACKGROUND

Self-supervised Framework

Our self-supervised framework leverages the abundance of healthy structural data while addressing the scarcity of labelled damage data. During training, a masked autoencoder is trained exclusively on undamaged structural data, where random patches of the input spectrograms are masked, and the model learns to reconstruct the missing regions from the visible context. This allows the model to learn the intrinsic features and distribution of healthy signals without requiring labelled data.

During testing, unseen spectrograms of guided signals from both undamaged and damaged conditions are passed through the trained model without masking. Since the model is unfamiliar with anomalous patterns, its reconstruction performance deteriorates in damaged regions. Anomalies are detected by computing the pixel-wise reconstruction error in terms of MSE between the input and the reconstructed output. To distinguish normal from anomalous samples, a percentile-based thresholding strategy is adopted. The distribution of reconstruction errors from the validation set (comprising only undamaged data) is used to estimate a statistical threshold. Test samples with reconstruction errors exceeding 99 percentiles are flagged as anomalies, indicating delamination structural damage.

Masked Autoencoder (MAE)

Autoencoders (AEs) trained exclusively on healthy structural data aim to reconstruct nominal signal patterns, where deviations in reconstruction indicate potential anomalies. However, traditional models like variational and convolutional AEs often fail to capture global dependencies due to their localised receptive fields. MAEs overcome these limitations by encoding only visible patches and reconstructing masked regions, enabling efficient self-supervised learning of rich spatial features from unlabeled time-frequency data with reduced computational cost.

MAE Encoder

The encoder of the MAE [11] adopts a ViT backbone, omitting its classification head. Instead of processing the full input, it operates only on a uniformly random 50% subset of non-overlapping image patches, masking the remaining 50% of the HHT spectrogram. This uniform sampling eliminates spatial bias (e.g., centre clustering) and compels the model to leverage global context, beyond local pixel neighbourhoods or adjacent patches, to reconstruct missing content. Each visible patch is linearly projected into an embedding space, augmented with positional encodings, and passed through a sequence of transformer encoder layers, each comprising multi-head self-attention followed by a position-wise feed-forward network. This encoder processes the following key steps:

Patch segmentation: Convert 2D HHT Spectrograms into 16×16 flattened patches.

Masking: Randomly mask 50% of the patches.

Embedding: Apply a linear projection to each visible patch and add positional embeddings.

Transformer encoding: Process embedded patches through multiple transformer blocks featuring multi-head self-attention and feed-forward networks.

MAE Decoder

The MAE decoder uses a stacked Transformer architecture to reconstruct masked regions of HHT spectrograms by processing both encoded visible patches and learned mask tokens. Positional embeddings preserve spatial structure, while multi-head self-attention and feedforward layers capture global dependencies and refine feature representations. A final linear layer projects these features to the pixel space, and reconstruction loss is computed as the mean squared error (MSE) between predicted and original masked regions:

$$MSE = \frac{\sum_{i=1}^{N}(PPV_i - OPV_i)^2}{N} \quad (1)$$

where N is the number of samples, PPV is the predicted pixel value, and OPV is the original pixel value for the i^{th} sample.

Improved Hilbert Huang Transform (HHT)

The HHT is a powerful time-frequency analysis method, fundamentally based on EMD [16] and the Hilbert Transform. In the application of HHT, EMD adaptively decomposes a nonlinear and non-stationary signal into a set of Intrinsic Mode Functions (IMFs) as shown in Fig. 2. These IMFs offer a complete, data-driven, and nearly orthogonal representation of the signal, with each function approximating a monocomponent signal. This property allows for the extraction of instantaneous frequencies from complex signals. Following decomposition, the Hilbert Transform is applied to each IMF to derive instantaneous frequency and corresponding local energy, resulting in the HHT spectrum, a comprehensive energy-frequency-time representation that enables accurate localisation of transient events in both time and frequency domains.

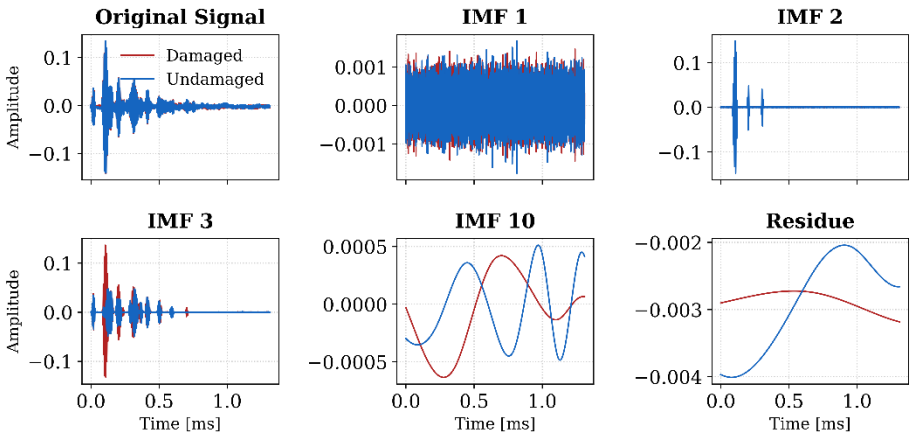

Figure 2. Variation of IMFs and residue for 1D-guided wave signals under different conditions.

Compared to wavelet-based methods, EMD, though computationally intensive, is convolution-free and thus offers improved efficiency in large-scale signal analysis.

Owing to these advantages, HHT has been widely employed in various domains for vibration signal analysis and transient event detection. Despite its utility, standard HHT suffers from several limitations. First, EMD may produce spurious IMFs in the low-frequency range, potentially leading to misinterpretation. Second, the first IMF may cover an overly broad frequency range, violating the assumption of monocomponent behaviour. Third, EMD often fails to resolve signal components with low energy, making subtle features difficult to detect.

To address these issues, an improved HHT framework has been adopted. The enhancement incorporates WPT as a preprocessing step, effectively decomposing the input signal into a set of narrowband components. This facilitates the identification of low-energy frequency components that would otherwise be obscured. EMD is subsequently applied to each narrowband signal, yielding IMFs that more reliably satisfy the monocomponent condition. To further refine the analysis, a screening mechanism based on correlation coefficients between individual IMFs, and the original signal is employed. This step suppresses irrelevant or noise-like IMFs, particularly in the low-frequency range, thereby mitigating the effects of spurious modes and enhancing interpretability. Fig. 3 shows the improved HHT spectrogram of guided waves under damaged and undamaged conditions, demonstrating superior resolution and accuracy in detecting and characterising nonlinear, non-stationary events compared to conventional HHT.

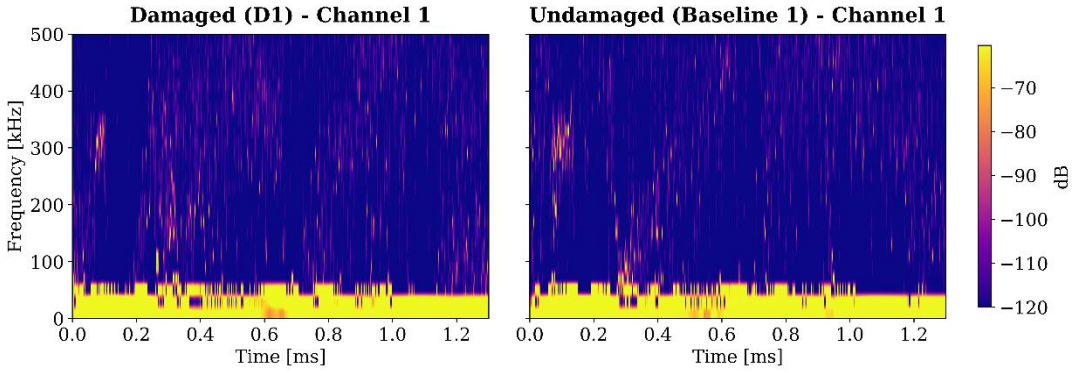

Figure 3. Improved HHT Spectrogram of 1D Guided Wave signals corresponding to channel 1 (a) undamaged (b) damaged.

DATABASE CONSTRUCTION

We utilise an experimental guided-wave benchmark dataset (http://openguidedwaves.de/downloads/) from the Open Guided Waves (OGW) repository. Specifically, we employ a constant-temperature dataset [17] that provides controlled conditions ideal for validating our self-supervised damage detection framework. The dataset was acquired from a 500 mm × 500 mm, 2 mm-thick carbon fibre-reinforced polymer (CFRP) panel fabricated using Hexply M21/34/UD134/T700/300 prepreg in a quasi-isotropic layup configuration: [45/0/45/90/45/0/45/90]. The specimen was instrumented with twelve surface-mounted piezoelectric (PZT) transducers arranged in a pitch-and-catch configuration. All measurements were conducted inside an environmental chamber maintained at a constant temperature of 23°C and 50% relative humidity.

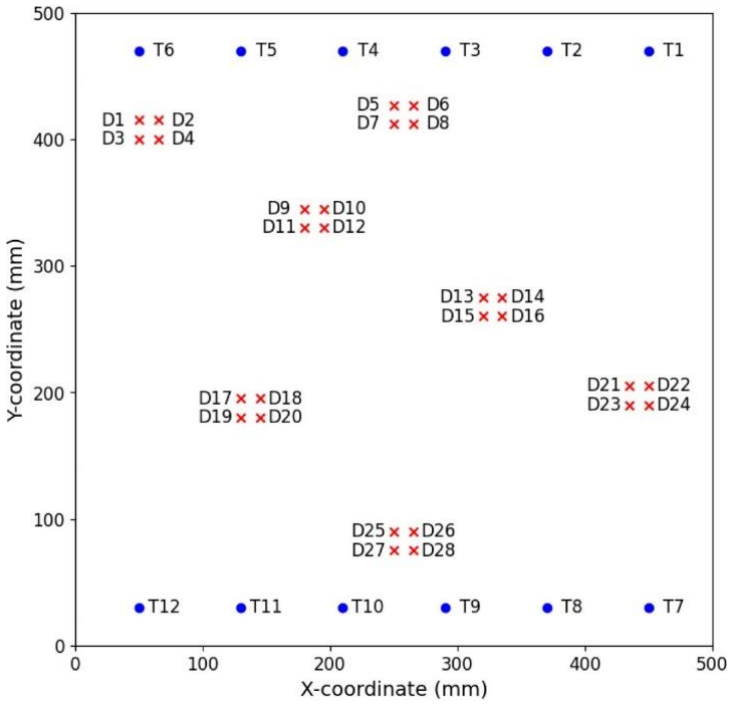

Figure 4. Schematic diagram of transducer and damage locations.

To simulate delamination, a reversible damage model was adopted, wherein a 10 mm-diameter, 2.35 mm-thick aluminium disc (mass: 0.5 g) was temporarily affixed at 28 predefined locations (D1-D28, see Fig. 4). The structure was interrogated at twelve discrete excitation frequencies ranging from 40 to 260 kHz. For each frequency, wavefield measurements were captured across 66 actuator-sensor paths (covering all round-robin permutations of the 12 transducers), with each path generating a signal comprising 13,108 time-series data points, resulting in a data matrix of dimensions 13,108 × 66 per frequency scan. Signal excitation was delivered using a five-cycle Hann-windowed sine burst with a peak amplitude of ±100 V.

HHT spectrograms constructed from the acquired signals reveal that the simulated reversible damage mimics the dynamic characteristics of actual delamination. The complete dataset comprises 47,520 baseline signals and 22,176 damage signals. For model development, a training subset of 10,000 signals randomly selected from the baseline signals is used, while the test set includes 2,000 signals from both baseline and damage cases. After selection, 1D raw guided wave signals are converted into 2D time-frequency-based HHT representations, which are then used in the self-supervised framework for anomaly detection.

TRAINING SET-UP

In this study, we develop a self-supervised anomaly detection framework based on MAEs, trained using undamaged (healthy) structural data. The objective is to enable the model to learn the intrinsic distribution of normal HHT images through a reconstruction-based learning task. By masking random patches of the input and reconstructing the missing regions, the MAE learns to capture spatial dependencies and semantic structure without requiring labeled data. Anomalies are later identified

based on deviations in reconstruction error, as the model fails to accurately reconstruct unseen damaged patterns.

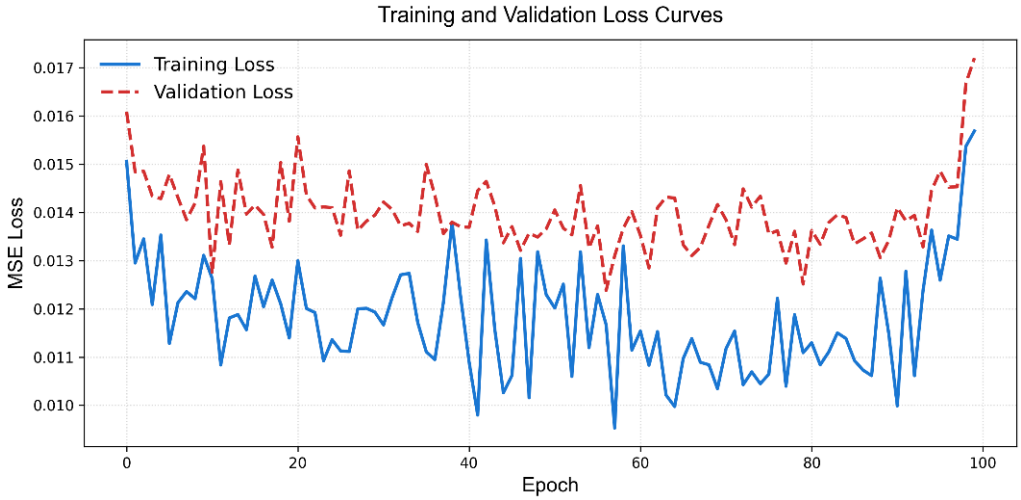

Figure 5. Learning curves of the self-supervised MAE model.

The training dataset consists of 10,000 HHT spectrogram images derived from healthy 1D guided wave signals. These are split into an 80%-20% ratio for training and validation. To reduce computational load and ensure compatibility with the model architecture, the original 500 × 1000 RGB images are resized to 128 × 128. Robustness is further enhanced through standard data augmentation techniques, including random rotations and horizontal flips.

The MAE model includes an encoder containing 6 Transformer blocks with a hidden width of 128, and a lightweight decoder comprising 2 Transformer blocks with a width of 64. The reconstruction target is defined at the pixel level without normalisation, allowing the model to capture the raw spatial variations. Empirical tuning of hyperparameters is carried out to ensure training stability, with a learning rate of 5e-5 using the AdamW optimiser. Regularisation techniques such as weight decay and dropout are incorporated to prevent overfitting and enhance generalisation. Details of the training configuration are presented in Table 1, and the training progression is visualised through the learning curve in Fig. 5.

TABLE I. TRAINING SETUP AND HYPERPARAMETERS FOR MAE MODEL

Input dimensions	Patch size	Learning rate	Dropout rate	Weight decay rate	Optimizer	Batch size	Epochs
500 x 1000	16	5e-5	0.1	1e-4	AdamW	16	100

The training was conducted on an NVIDIA GeForce RTX 3070 GPU (8 GB RAM), requiring approximately 20 hours due to the high computational demand of patch-wise image reconstruction. Once trained, the model is evaluated for anomaly detection by calculating the pixel-wise reconstruction error on test samples. HHT spectrograms with errors exceeding a statistically derived 99% percentile threshold (based on the train error distribution) are flagged as anomalous, indicating potential delamination defects, inconsistencies, or structural damage.

RESULTS & DISCUSSION

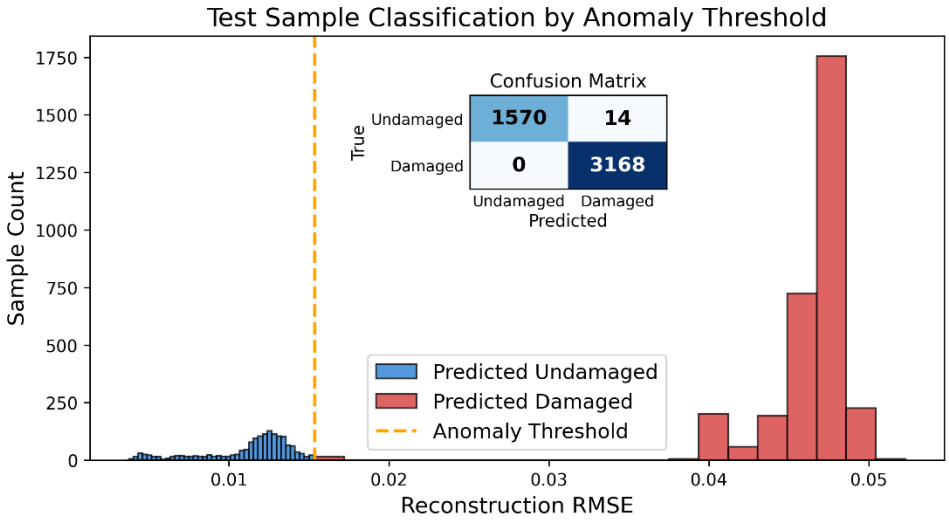

Figure 6. Classification of test samples based on anomaly threshold.

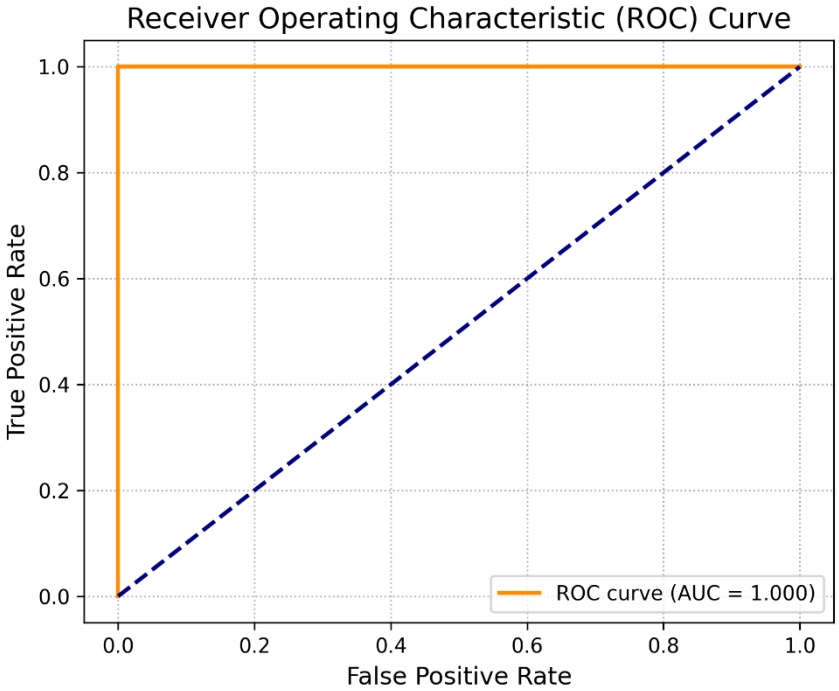

Figure 7. Classification of test samples based on anomaly threshold.

To enhance time-frequency analysis of one-dimensional guided wave signals in SHM, an improved HHT representation is proposed. EMD is first applied to extract IMFs, with only the physically meaningful ones retained based on frequency relevance. These selected IMFs are then used to generate high-resolution HHT spectrograms that effectively reduce energy leakage and localise transient damage-related features. The MAE-based framework is employed for unsupervised feature learning and anomaly detection. This MAE framework is trained based on HHT spectrograms derived using signals under undamaged conditions and learns to capture the distribution of healthy structural responses. Its training objective

minimises the MSE between the original and reconstructed spectrograms, but only on randomly masked input regions, following the self-supervised learning paradigm.

We validated our framework by applying the trained MAE to new HHT spectrograms from both healthy and delaminated composite samples. As shown in Fig. 6, the error histograms form two distinct peaks: one at low error values for intact specimens and another at much higher values for damaged ones. This clear split demonstrates that the MAE, aided by the richer time–frequency details of the enhanced HHT, can reliably detect subtle deviations from normal structural behaviour.

Next, we assessed detection performance using receiver operating characteristic (ROC) and precision–recall (PR) analyses. The ROC curve in Fig. 7 has an area under the curve (AUC) of about 0.99, indicating almost perfect separation between undamaged and delaminated states. The PR curve further shows precision exceeding 95 % even when recall surpasses 90 %. In practice, for monitoring safety-critical composites like CFRP and GFRP, it is preferable to keep false alarms low, even if it means accepting a few missed detections. Minimising false positives avoids unnecessary inspections and false shutdowns, thereby improving system reliability under real-world conditions. In future, we will work on multi-class damage detection (other than delamination) and the damage quantification process by extending this work.

CONCLUSIONS

This study presents a novel unsupervised framework for delamination detection in CFRP plates, combining a ViT-based MAE with an enhanced HHT representation of UGWs. Validated on a benchmark dataset, the approach demonstrates superior detection accuracy and computational efficiency. The global self-attention of the MAE effectively captures long-range dependencies in wavefields, enabling precise anomaly localisation without requiring labelled damage data, enhancing scalability for SHM deployment.

Future work will explore real-time implementation, adaptation to complex composite geometries, and integration with edge computing for on-site monitoring. Enhancements may include temporal encoding to track damage progression, multimodal sensor fusion via cross-modal pretraining, and multi-scale MAE architectures to boost robustness across diverse SHM conditions.

REFERENCES

[1] S. Han, Q. Li, Z. Cui, P. Xiao, Y. Miao, L. Chen, Y. Li, Non-destructive testing and structural health monitoring technologies for carbon fiber reinforced polymers: a review, Nondestructive Testing and Evaluation 39 (2024) 725–761. https://doi.org/10.1080/10589759.2024.2324149.

[2] T. Huang, M. Bobyr, A Review of Delamination Damage of Composite Materials, J. Compos. Sci. 7 (2023) 468. https://doi.org/10.3390/jcs7110468.

[3] K. Senthil, A. Arockiarajan, R. Palaninathan, B. Santhosh, K.M. Usha, Defects in composite structures: Its effects and prediction methods – A comprehensive review, Composite Structures 106 (2013) 139–149. https://doi.org/10.1016/j.compstruct.2013.06.008.

[4] M. Mitra, S. Gopalakrishnan, Guided wave based structural health monitoring: A review, Smart Mater. Struct. 25 (2016) 053001. https://doi.org/10.1088/0964-1726/25/5/053001.
[5] M. Rautela, S. Jayavelu, J. Moll, S. Gopalakrishnan, Temperature compensation for guided waves using convolutional denoising autoencoders, in: P. Fromme, Z. Su (Eds.), Health Monitoring of Structural and Biological Systems XV, SPIE, Online Only, United States, 2021: p. 40. https://doi.org/10.1117/12.2582986.
[6] A. Sattarifar, T. Nestorović, Emergence of Machine Learning Techniques in Ultrasonic Guided Wave-based Structural Health Monitoring: A Narrative Review, IJPHM 13 (2022). https://doi.org/10.36001/ijphm.2022.v13i1.3107.
[7] S. Cantero-Chinchilla, P.D. Wilcox, A.J. Croxford, Deep learning in automated ultrasonic NDE – Developments, axioms and opportunities, NDT & E International 131 (2022) 102703. https://doi.org/10.1016/j.ndteint.2022.102703.
[8] H. Liu, S. Liu, Z. Liu, N. Mrad, A.S. Milani, Data-Driven Approaches for Characterization of Delamination Damage in Composite Materials, IEEE Trans. Ind. Electron. 68 (2021) 2532–2542. https://doi.org/10.1109/TIE.2020.2973877.
[9] Y. Yu, X. Liu, Y. Wang, Y. Wang, X. Qing, Lamb wave-based damage imaging of CFRP composite structures using autoencoder and delay-and-sum, Composite Structures 303 (2023) 116263. https://doi.org/10.1016/j.compstruct.2022.116263.
[10] P. Dutta, K. Kanti Podder, J. Zhang, C. Hecht, S. Swarna, P. Bhavsar, A Self-Supervised Learning Approach to Road Anomaly Detection Using Masked Autoencoders, in: International Conference on Transportation and Development 2024, American Society of Civil Engineers, Atlanta, Georgia, 2024: pp. 536–547. https://doi.org/10.1061/9780784485538.047.
[11] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, R. Girshick, Masked Autoencoders Are Scalable Vision Learners, (2021). https://doi.org/10.48550/ARXIV.2111.06377.
[12] W. Luo, C. Lin, H. Zhou, Underwater Target Detection by Residual Spatial Cooperative Attention Module–Based Self-Supervised Learning, IEEE J. Oceanic Eng. (2025) 1–14. https://doi.org/10.1109/JOE.2025.3556153.
[13] Y. Zhang, S. Wang, S. Huang, W. Zhao, Mode Recognition of Lamb Wave Detecting Signals in Metal Plate Using the Hilbert-Huang Transform Method, JST 05 (2015) 7–14. https://doi.org/10.4236/jst.2015.51002.
[14] R. Gangadharan, C.R.L. Murthy, S. Gopalakrishnan, M.R. Bhat, D.R. Mahapatra, Characterization Of Cracks And Delaminations Using Pwas Ad Lamb Wave Based Time-Frequency Methods, International Journal on Smart Sensing and Intelligent Systems 3 (2010) 703–735. https://doi.org/10.21307/ijssis-2017-417.
[15] Z.K. Peng, P.W. Tse, F.L. Chu, An improved Hilbert–Huang transform and its application in vibration signal analysis, Journal of Sound and Vibration 286 (2005) 187–205. https://doi.org/10.1016/j.jsv.2004.10.005.
[16] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. A 454 (1998) 903–995. https://doi.org/10.1098/rspa.1998.0193.
[17] J. Moll, J. Kathol, C.-P. Fritzen, M. Moix-Bonet, M. Rennoch, M. Koerdt, A.S. Herrmann, M.G. Sause, M. Bach, Open Guided Waves: online platform for ultrasonic guided wave measurements, Structural Health Monitoring 18 (2019) 1903–1914. https://doi.org/10.1177/1475921718817169.