

Fracture, Damage and Structural Health Monitoring

Deep learning approaches for impact identification on composite structures under environmental and operational variabilities: a comparative study

Dong Xiao^{ID*}, Zahra Sharif-Khodaei^{ID}, M. H. Aliabadi^{ID}*Department of Aeronautics, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom.*

Abstract

This study presents a comprehensive evaluation of deep learning approaches for impact identification in composite structures under environmental and operational variabilities (EOVs). Five representative architectures—Convolutional Neural Networks (CNNs), Temporal Convolutional Networks (TCNs), Recurrent Neural Networks (RNNs), Graph Neural Networks (GNNs), and Transformers (XFMRs)—are compared across two key tasks: impact localisation (predicting spatial coordinates) and impact force reconstruction (estimating time-varying force histories). Particular emphasis is placed on model robustness when testing conditions deviate from those used in training, including temperature changes and impact mass variation. Additionally, the effects of critical data acquisition parameters—such as sampling frequency, signal window length, and sensor density—on model performance and generalisability are systematically investigated. Experimental validation is conducted using controlled impact tests on composite panels, providing insight into the strengths and limitations of each model architecture in realistic structural health monitoring scenarios.

© 2025 The Authors. Published by ELSEVIER B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of Ferri Aliabadi

Keywords: Structural health monitoring; Composite structures; Deep learning; Environmental and operational variabilities (EOVs); Impact localisation; Force reconstruction;

1. Introduction

Deep learning techniques have garnered increasing attention in structural health monitoring (SHM) due to their powerful capabilities in learning hierarchical representations for both classification tasks (e.g., damage detection [1, 2]) and regression tasks (e.g., damage localisation [3] and remaining useful life prediction [4]). By leveraging large datasets and automatic feature extraction, deep learning models have demonstrated remarkable success in capturing complex, nonlinear patterns in structural response data that are often difficult to represent explicitly using traditional physics-based models.

Among the many SHM challenges, impact identification [5]—including both impact localisation (IL) [6] and impact force reconstruction (IFR) [7]—has emerged as a critical problem, especially for composite structures widely

* Corresponding author.

E-mail address: d.xiao21@imperial.ac.uk (D. Xiao); z.sharif-khodaei@imperial.ac.uk (Z. Sharif-Khodaei); m.h.aliabadi@imperial.ac.uk (M.H. Aliabadi)

used in aerospace, automotive, and civil infrastructure. These structures are particularly susceptible to barely visible impact damage (BVID) from low-velocity impacts, which can lead to internal delaminations without obvious surface signatures. Timely and accurate identification of such impacts is crucial for assessing structural integrity and informing maintenance strategies.

Recent progress in deep learning has led to the development of various model architectures tailored for impact localisation and force reconstruction in composites, as summarised in Table 1. Convolutional Neural Networks (CNNs) have been widely applied to extract spatial features from sensor signals, particularly when time–frequency representations such as the Continuous Wavelet Transform (CWT) or Short-Time Fourier Transform (STFT) are used as inputs [8–11]. Temporal Convolutional Networks (TCNs) leverage dilated causal convolutions to model long-range temporal dependencies in sensor signals, offering advantages in training stability and inference speed [12].

Table 1: An summary of the deep learning approaches for impact identification.

Tasks	Model	References	Characteristics
IL	CNNs	[8–11]	Learns spatial features from time–frequency input; robust to noise and moderate variation
	TCNs	[12]	Captures long-term temporal features; efficient and stable for time series
	GNNs	[13]	Models spatial relationships among irregular sensor layouts; topology-aware
	XFMR	[14]	Captures global dependencies via attention; suitable for variable-length inputs
IFR	TCNs	[12]	Efficient modeling of long-range temporal dynamics; low memory footprint
	RNNs	[15, 16]	Learns sequential patterns; effective for time-varying force signals
	GNNs	[17, 18]	Incorporates spatial context into temporal prediction; flexible for structural variations

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have also been employed for impact force reconstruction due to their ability to learn time-varying dependencies across multiple sensor channels [15, 16]. Hybrid CNN-RNN architectures have shown enhanced performance by capturing both spatial and temporal features simultaneously, proving effective under noisy or sparse sensing conditions. More recently, Graph Neural Networks (GNNs) have been applied to model structural connectivity and sensor relationships explicitly, allowing more accurate localisation and force estimation across irregular sensor layouts [13, 17, 18].

Transformers (XFMRs), originally proposed for natural language processing, have recently gained traction in SHM due to their attention-based mechanism that flexibly captures global dependencies in both space and time. Their ability to model variable-length inputs and maintain performance under noisy or complex conditions makes them a promising candidate for impact localisation and reconstruction tasks under variable environmental and operational conditions [14].

Despite these advancements, the robustness and generalisation of deep learning models remain key challenges when transitioning from laboratory settings to real-world SHM applications. Environmental and operational variabilities (EOVs)—such as temperature changes, varying impactor mass, and sensor noise—can significantly alter the structural response signals and impair model performance. Systematic benchmarking of model robustness under such variations is therefore necessary to guide practical deployment of deep learning models for SHM.

To this end, this study conducts a comprehensive evaluation of deep learning models for impact identification in composite structures under realistic EOVs. Specifically, the objectives are:

- To compare the performance of representative deep learning architectures—CNN, TCN, CNN-LSTM, GNN, and Transformer—in both impact localisation and force reconstruction;
- To investigate the influence of impactor mass and temperature variation on model generalisation;
- To evaluate the sensitivity of each model to acquisition parameters including sampling frequency, time window, and sensor count;
- To provide practical insights into model selection, training strategies, and limitations for SHM under uncertain operating conditions.

The remainder of the paper is organised as follows: Section 2 outlines the architectures and training strategies of the deep learning models for impact localisation and force reconstruction. Section 3 describes the experimental setup,

sensor layout, and data acquisition under controlled variations in temperature and impact mass. [Section 4](#) presents a comparative analysis of model performance under EOVI and examines the effects of acquisition parameters on model accuracy and generalisation. Finally, [Section 5](#) summarises the key findings and suggests directions for future research on robust, adaptive impact identification methods.

2. Deep learning models for impact identification

2.1. Impact localisation

To evaluate the localisation performance of deep learning models, four representative architectures are considered: a Convolutional Neural Network (CNN), a Temporal Convolutional Network (TCN), a Transformer (XFMR), and a Graph Neural Network (GNN), as illustrated in [Fig. 1](#).

The CNN model ([Fig. 1\(a\)](#)) adopts a compact structure composed of stacked convolutional blocks with kernel strides and max-pooling layers for progressive downsampling. This hierarchical design enables efficient capture of local spatial-temporal features while reducing computational cost. A global average pooling layer replaces traditional fully connected layers, significantly reducing parameter count without sacrificing performance. The final dense layers map the compressed features to impact coordinates.

The TCN model ([Fig. 1\(c\)](#)) comprises residual blocks with dilated causal convolutions, where dilation factors grow exponentially (e.g., 1, 2, 4, ...). This architecture captures long-range temporal dependencies in the sensor data. Residual connections enhance gradient flow and training stability, while dropout and adaptive average pooling improve generalisation and reduce dimensionality before the output layer.

In contrast to CNNs and TCNs, the Transformer model ([Fig. 1\(d\)](#)) employs self-attention to model global dependencies across the input sequence. A preliminary convolutional block extracts local features, followed by positional encoding to reintroduce temporal information lost during pooling. The Transformer encoder then adaptively focuses on signal segments most informative for localisation, with the final decoder outputting predicted impact coordinates.

The GNN model ([Fig. 1\(e\)](#)) takes a graph-based approach, representing the sensor network as nodes and edges that encode spatial or functional relationships. This topology-aware structure is well suited to irregular sensor layouts and non-uniform geometries. Node features—extracted signal embeddings—are updated through message-passing layers, capturing spatial correlations and supporting robustness to sensor noise and missing data. The learned graph representation is pooled and mapped to the predicted impact location through a regression head.

2.2. Impact force estimation

For reconstructing the time history of impact force, four deep learning models are evaluated: a hybrid CNN–Long Short-Term Memory (CNN–LSTM) network, a TCN, a Transformer, and a GNN. These architectures are selected for their ability to capture the temporal dynamics and spatial relationships embedded in sensor signals.

The CNN–LSTM model ([Fig. 1\(b\)](#)) combines convolutional layers for local feature extraction with LSTM units to learn long-term temporal dependencies. The CNN layers first extract hierarchical temporal patterns relevant to the contact event. These features are then processed by stacked LSTM layers, which preserve memory over time and enable sequential prediction of the impact force. Fully connected layers map the LSTM outputs to the reconstructed force history. This hybrid architecture is particularly effective for modelling force profiles with distinct phases such as rapid rise and decay.

The TCN model ([Fig. 1\(c\)](#)) retains its causal structure with stacked dilated convolutions, allowing efficient modelling of long-range temporal dependencies without recurrence. The use of residual connections and dropout layers promotes stable and generalisable training. Compared to recurrent models, TCNs offer superior training speed and lower memory consumption, making them attractive for time-series regression.

The Transformer model ([Fig. 1\(d\)](#)) applies self-attention to encode dependencies across the entire signal sequence. A convolutional front-end captures local features, followed by positional encoding and Transformer encoder layers to learn context-aware representations of the sequence. A dedicated decoder then reconstructs the time-varying force signal. This architecture is particularly suited for handling variable-length signals and modelling global patterns influenced by structural dynamics.

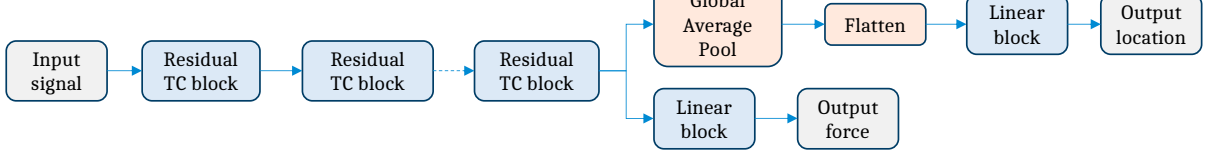
(a) CNN model for impact localisation



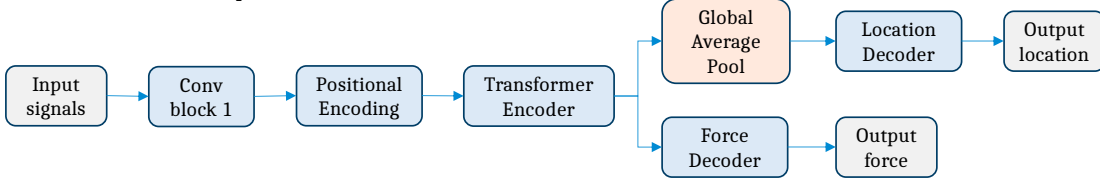
(b) CNN-LSTM model for impact force estimation



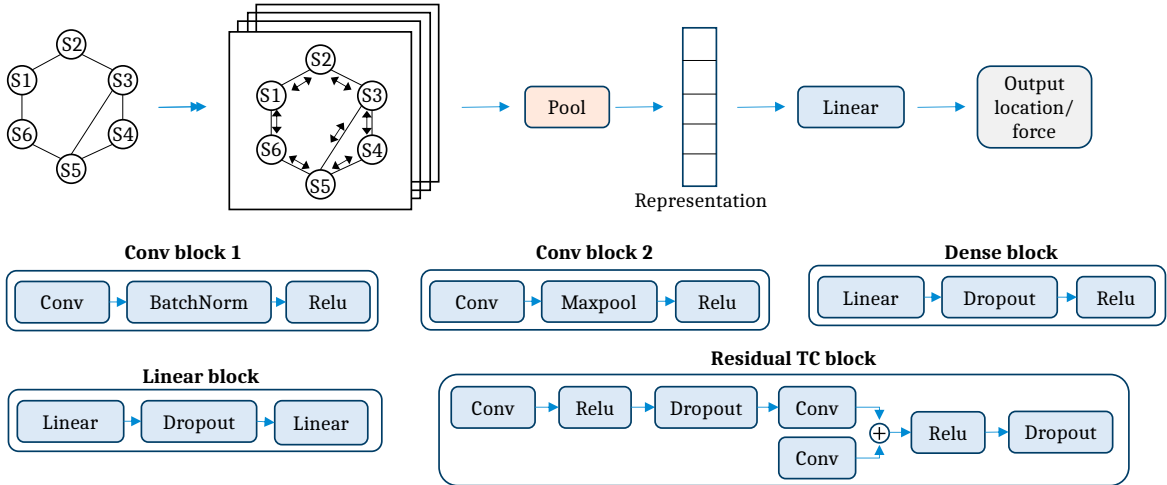
(c) TCN model for impact localisation and force estimation



(d) Transformer for impact localisation and force estimation



(e) GNN for impact localisation and force estimation

**Fig. 1:** Diagram of the deep learning architectures for impact identification.

The GNN model (Fig. 1(e)) extends its spatial modelling capabilities to dynamic force estimation. Here, each sensor is treated as a node, and graph-based message passing allows temporal features to propagate across the network. Time-varying node inputs are either encoded using temporal slices or processed through integrated temporal modules. The final output, either at the graph or node level, is decoded into the impact force sequence. The GNN's structural inductive bias supports generalisation across varying sensor configurations and enhances resilience to environmental variability.

In summary, each architecture offers complementary strengths: CNN-LSTM and TCN excel in capturing fine-grained and long-range temporal features; the Transformer leverages global attention mechanisms; and the GNN integrates spatial topology with temporal dynamics. Together, these models provide a diverse foundation for benchmarking impact force estimation under diverse test conditions and environmental uncertainties.

3. Experimental validation on composite panel

To experimentally validate the proposed methodologies for impact localisation and force reconstruction, a comprehensive set of controlled impact experiments was conducted on a sensorised carbon fibre-reinforced polymer (CFRP) panel. These experiments systematically varied impactor mass and ambient temperature to emulate operational diversity and assess the generalisation capabilities of the developed learning-based framework under environmental and operational variability (EOV).

The test specimen was a quasi-isotropic laminate composed of M21/T800 carbon/epoxy, with nominal dimensions of $290 \times 200 \times 4$ mm and a stacking sequence of $[0/+45/-45/90]_{2s}$. Six piezoelectric (PZT) sensors were surface-mounted at carefully selected locations to capture the structural responses resulting from the transient impact events. The panel layout, sensor configuration, and acquisition hardware follow standardised procedures detailed in prior work [19, 20]. Signal acquisition was carried out using a National Instruments PXI-5105 digitiser, ensuring high-fidelity data capture suitable for downstream signal processing and model training.

A summary of all test conditions—including variations in impact tools, impactor mass, ambient temperature, and the presence of ground-truth force measurements—is presented in Table 2. This test matrix provides a diverse, multi-condition dataset suitable for both model training and cross-domain generalisation assessment.

3.1. Impact testing using drop mass under temperature fluctuations

To evaluate model robustness to temperature variation, a set of impact experiments was performed using a guided drop-mass system, as shown in Fig. 2(b). A 100 g cylindrical steel mass was dropped vertically along a guide rail to achieve repeatable and perpendicular impacts on the panel surface. The boundary conditions matched those used in the hammer tests, with the longer edges of the panel clamped and the shorter edges left free to mimic practical structural constraints.

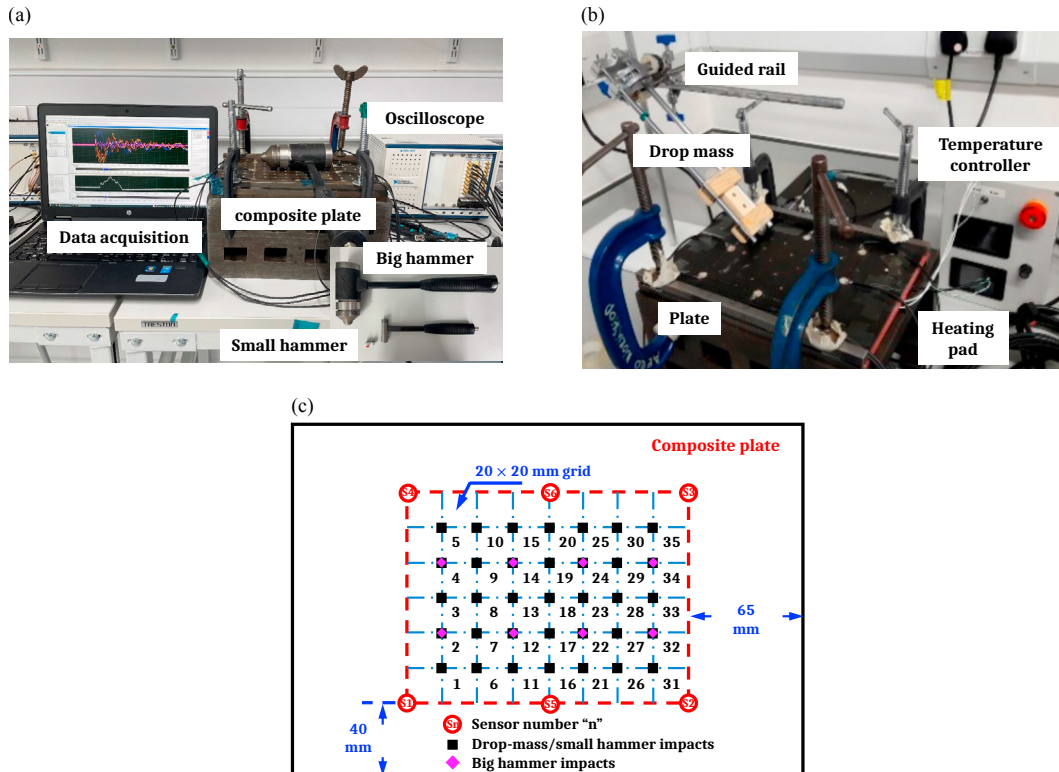


Fig. 2: Diagram of the reference CNN architectures for impact localisation.

To introduce controlled thermal variation, a flexible heating pad was affixed to the back side of the laminate, raising the surface temperature to an elevated state of 70°C. Two thermal scenarios were tested: a baseline room temperature (24°C) and the elevated temperature (70°C) [21]. Temperature was monitored using surface thermocouples and maintained uniformly throughout each testing session. Six PZT sensors (numbered 1–6, see Fig. 2(c)) were used to capture structural responses, and signals were sampled (or downsampled) to 400 kHz.

For both temperature cases—denoted REF and TEM in Table 2—impacts were performed at 35 evenly spaced locations on a 5×7 grid (identical to the SH case). These tests did not include direct force measurements due to the lack of instrumentation in the drop mass, making them ideal for evaluating the generalisability of impact localisation models. Additionally, the TEM condition serves to assess model performance under out-of-distribution thermal environments, providing insight into model adaptability and reliability in real-world applications.

3.2. Impact testing using hammers under impact mass variations

To investigate the effect of impactor mass on localisation and force prediction performance, additional impact experiments were conducted using two types of instrumented hammers with integrated force sensors: a light hammer (PCB Piezotronics 086C03, 160 g) and a heavy hammer (PCB Piezotronics 086D20, 1100 g), as shown in Fig. 2(a). The composite panel was fixed along its two longer edges to simulate a simply supported boundary condition, and four PZT sensors (numbered 1–4) were deployed for response measurement at a sampling rate of 200 kHz.

Impacts were delivered at predefined locations on a 20×20 mm regular grid spanning the panel surface, as depicted in Fig. 2(c). Small-mass hammer impacts (SH) were conducted at 35 unique locations, distributed over a 5×7 grid, and are marked by black squares in the figure. Large-mass hammer impacts (BH) were conducted at 8 selected locations twice (marked in magenta) that form a spatial subset of the small-mass grid. This overlapping spatial design enables direct comparison of sensor responses across different impactor masses at common locations, facilitating the investigation of amplitude- and waveform-dependent variations in both localisation and force estimation.

The use of instrumented hammers provided time-resolved ground-truth force histories for each impact, enabling supervised learning and validation of the proposed force reconstruction models. The inclusion of both low-energy (SH) and high-energy (BH) impacts supports model robustness across a wide range of excitation magnitudes.

3.3. Overview of experimental test matrix

Table 2 provides an overview of the four experimental scenarios comprising the test matrix:

- SH (Small Hammer): 160 g impacts at 35 locations at 24°C with available force measurements.
- BH (Big Hammer): 1100 g impacts at 8 locations at 24°C with available force measurements, 2 repetitions.
- REF (Drop Mass, Room Temp): 100 g impacts at 35 locations at 24°C without force measurements.
- TEM (Drop Mass, Elevated Temp): 100 g impacts at 35 locations at 70°C without force measurements.

These four distinct experimental conditions span variations in impact energy, mass distribution, and thermal environment, providing a comprehensive dataset for assessing the generalisability, robustness, and uncertainty quantification capabilities of the proposed deep learning models.

Table 2: Summary of experimental impact tests on the composite panel.

Case	Impact tool	Impact Mass (g)	Temperature (°C)	Force measurement (Y/N)	Number of impact
SH	Small hammer	160	24	Y	35
BH	Big hammer	1100	24	Y	8*2
REF	Drop mass	100	24	N	35
TEM	Drop mass	100	70	N	35

This comprehensive and well-structured experimental design enables targeted evaluation of the candidate DL framework under various sources of environmental and operational variability:

- **Thermal generalisation for localisation:** Assesses the impact of temperature variations on localisation performance by training models on room-temperature drop-mass data (REF) and testing on elevated-temperature data (TEM).
- **Mass variation for localisation** Examines the sensitivity of localisation accuracy to changes in impactor mass by training on small hammer data (SH) and evaluating on large hammer data (BH).
- **Mass variation for force reconstruction:** Investigates the robustness of force estimation models across different impact energies by training on SH data and testing on BH data.

The diversity and structure of the dataset are essential for assessing the robustness, generalisability, and uncertainty quantification capabilities of the proposed deep learning approach in real-world composite aerostructure scenarios subjected to varying impact conditions and thermal environments.

4. Comparative analysis of deep learning models on impact identification

Based on the assumption of linear elastic behaviour, the sensor signals and corresponding impact force histories for each impact event are normalised by the maximum absolute amplitude observed across all sensors. This normalisation ensures data comparability across varying impact energies and enhances the consistency and generalisability of the deep learning models.

4.1. Thermal generalisation for localisation

Fig. 3 presents the empirical cumulative distribution functions (CDFs) of localisation errors—defined as the root-sum-square (RSS) of the spatial coordinate errors—for four deep learning models trained on room-temperature data (REF) and tested on elevated-temperature data (TEM), representing a thermal shift of 46°C. These results assess each model's ability to generalise under thermal variability, a critical challenge for reliable deployment in real-world applications.

With six sensors ($N_s = 6$), a sampling frequency of $f_s = 400$ kHz, and a temporal window of 10 ms ($n = 4000$ samples), as shown in Fig. 3(a), both the Transformer-based (XFMR) and CNN models exhibited the highest robustness and precision. Each achieved a 90% probability of detection (PoD) within 12.8 mm, well below the training grid resolution (20 mm). This demonstrates their ability to retain spatial generalisation under thermal perturbation, likely due to their strong feature learning and spatial-temporal encoding capabilities.

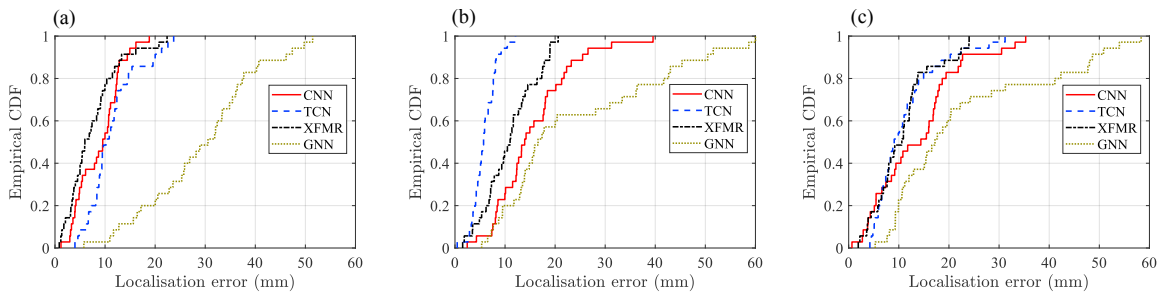


Fig. 3: Impact localisation under temperature shift: (a) $N_s = 6$, $f_s = 400$ kHz, $n = 4000$, (b) $N_s = 6$, $f_s = 200$ kHz, $n = 2000$, (c) $N_s = 4$, $f_s = 400$ kHz, $n = 4000$.

In contrast, the TCN and GNN models underperformed, with PoDs of 19.6 mm and 40.7 mm, respectively. The GNN model's markedly degraded accuracy indicates poor resilience to temperature-induced waveform distortions, possibly due to its reliance on spatial graph topology without sufficiently encoding temporal dynamics or thermal variability.

As shown in Fig. 3(b), reducing the sampling frequency to $f_s = 200$ kHz (with $N_s = 6$ and $n = 2000$) generally decreased model accuracy for CNN, XFMR, and GNN. This degradation can be attributed to the loss of high-frequency temporal information essential for characterising early wave arrival features and impact-induced transients.

Interestingly, the TCN model showed improved localisation accuracy at the lower sampling rate, suggesting that its hierarchical temporal convolutions may benefit from reduced input dimensionality, making learning more stable under domain shift. This observation warrants further investigation into optimal signal resolution for TCN-based models under variable conditions.

Fig. 3(c) investigates the effect of sensor sparsity by reducing the number of PZT sensors from six to four (while maintaining $f_s = 400$ kHz and $n = 4000$). As expected, all models exhibited decreased localisation accuracy due to reduced spatial coverage. The best-performing models, XFMR and TCN, achieved PoDs of 19.1 mm and 18.5 mm, respectively—marginally below the grid resolution. These results confirm that while sensor redundancy enhances robustness, high-capacity models like XFMR can still operate near the resolution limit with limited input.

Among all sensor-sampling configurations, the XFMR model consistently demonstrated the highest robustness, maintaining a PoD below 20 mm across all cases. This consistent performance highlights the model's strong generalisation capability and adaptability to reduced signal fidelity or sensor density. It suggests that Transformer-based architectures are especially suited for impact localisation in operational environments where both thermal variability and sensor constraints are present.

4.2. Mass variation for localisation

Fig. 4 presents the empirical CDFs of localisation errors for large-mass hammer (BH) impacts, using models trained exclusively on small-mass hammer (SH) impact data. This evaluation targets the models' generalisation capacity under impact mass variations, where both signal amplitude and frequency content can differ substantially.

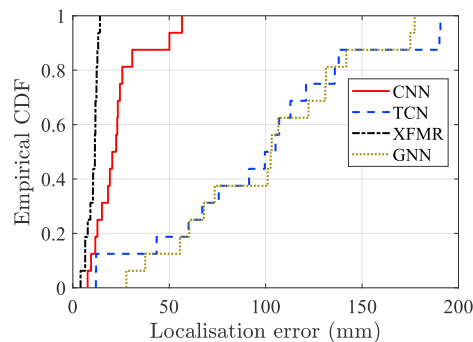


Fig. 4: Impact localisation under mass variation.

In this context, BH impacts introduce a clear domain shift: the heavier impactor produces signals with longer contact duration and energy concentrated in lower-frequency components. Spectral analysis reveals that BH-induced signals are dominated by low-frequency content, in contrast to the broader and higher-frequency spectra typical of SH-induced responses. This frequency shift poses a challenge for models trained on SH data, which have learned to localise impacts based on higher-frequency signal features.

Among the four evaluated models, the Transformer-based XFMR model exhibits the most robust generalisation to these spectral and amplitude variations. It achieves a 90% PoD within 12.7 mm—well below the 20 mm spatial grid resolution—indicating high localisation accuracy despite the spectral mismatch. The model's capacity to capture long-range temporal dependencies and integrate information across the full signal duration likely underpins its resilience to such domain shifts.

In contrast, the CNN, TCN, and GNN models perform markedly worse, with 90% PoD values of 30.7 mm, 137.9 mm, and 141.8 mm, respectively. These degraded performances reflect an overreliance on localised time–frequency features learned from SH signals, which do not generalise well to the lower-frequency BH signal domain. The CNN's fixed-size convolutional kernels and the TCN's finite receptive field limit their ability to adapt to the temporal spreading and frequency downshift introduced by the heavier impacts.

The GNN model's poor performance is particularly notable. While graph-based models are designed to capture structural relationships across sensors, the fixed topology and lack of explicit frequency-awareness in the GNN archi-

texture hinder its capacity to discriminate between impacts with different frequency signatures. As a result, it fails to extract transferable localisation features when the spectral content deviates from the training distribution.

These findings underscore the importance of selecting model architectures that are not only expressive but also resilient to changes in signal frequency content and dynamics. The XFMR model emerges as a strong candidate for real-world scenarios where impact conditions vary unpredictably—offering robust localisation even under significant variations in impact energy, contact duration, and signal frequency characteristics.

4.3. Mass variation for force reconstruction

To highlight the influence of impactor mass on dynamic response characteristics, Fig. 5 presents a comparison of measured impact forces and corresponding sensor signals at location 2 under small-mass hammer (SH) and big-mass hammer (BH) impacts. As shown, BH impacts generate sensor signals with more pronounced low-frequency components compared to the broader and higher-frequency content observed in SH impacts. This observation aligns with prior findings [22], which indicate that impactor mass plays a dominant role in shaping the frequency content and temporal characteristics of impact responses, particularly by altering the contact duration and energy transfer dynamics.

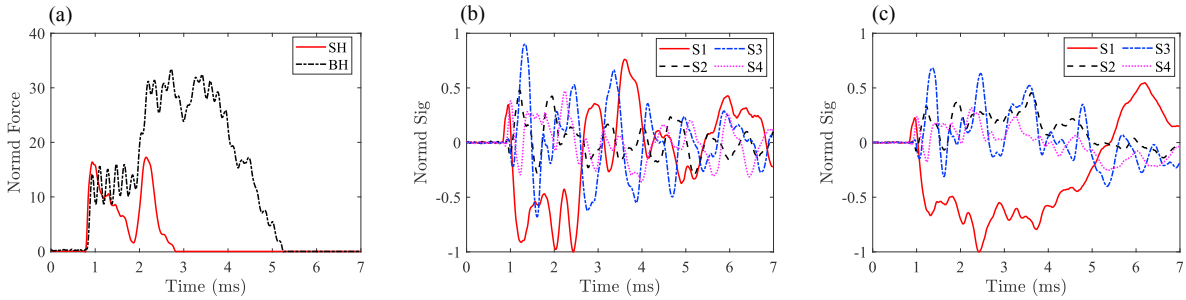


Fig. 5: Impact forces and sensor signals under mass variation: (a) impact forces of SH and BH impacts at location 2, (b) sensor signals of SH impact, (c) sensor signals of BH impact.

Beyond frequency content, impactor mass also affects the scaling relationship between the measured impact force and the corresponding sensor response. This relationship can be quantified using the force-to-signal ratio, R , is defined as:

$$R = \frac{F_{peak}}{S_{peak}}, \quad (1)$$

where F_{peak} denotes the peak impact force and S_{peak} represents the maximum absolute amplitude of the sensor signal. It is consistently observed that BH impacts yield a higher R than SH impacts. This implies a nonlinear relationship between input force and sensor output amplitude, challenging the common assumption of a fixed linear mapping often adopted in force reconstruction models.

This nonlinearity does not arise solely from the intrinsic characteristics of the piezoelectric sensor, but from the complex interaction between contact mechanics, wave propagation, and sensor-structure coupling. SH impacts, characterised by shorter contact durations and broader high-frequency content, tend to excite structural modes more effectively within the sensor's operational bandwidth, leading to relatively stronger sensor responses for a given force. In contrast, BH impacts involve longer contact durations and predominantly low-frequency content, which may couple less efficiently into the structure's dynamic response or fall outside the frequency range to which the sensor-structure system is most responsive. Additionally, differences in modal density and energy distribution between SH and BH impacts contribute further to the observed force-to-signal scaling nonlinearity.

To assess the generalisation ability of data-driven models across impact mass variations, Fig. 6 compares the performance of four deep learning models—CNN-LSTM, TCN, Transformer (XFMR), and GNN—in reconstructing impact forces for both SH (training) and BH (testing) datasets at multiple locations.

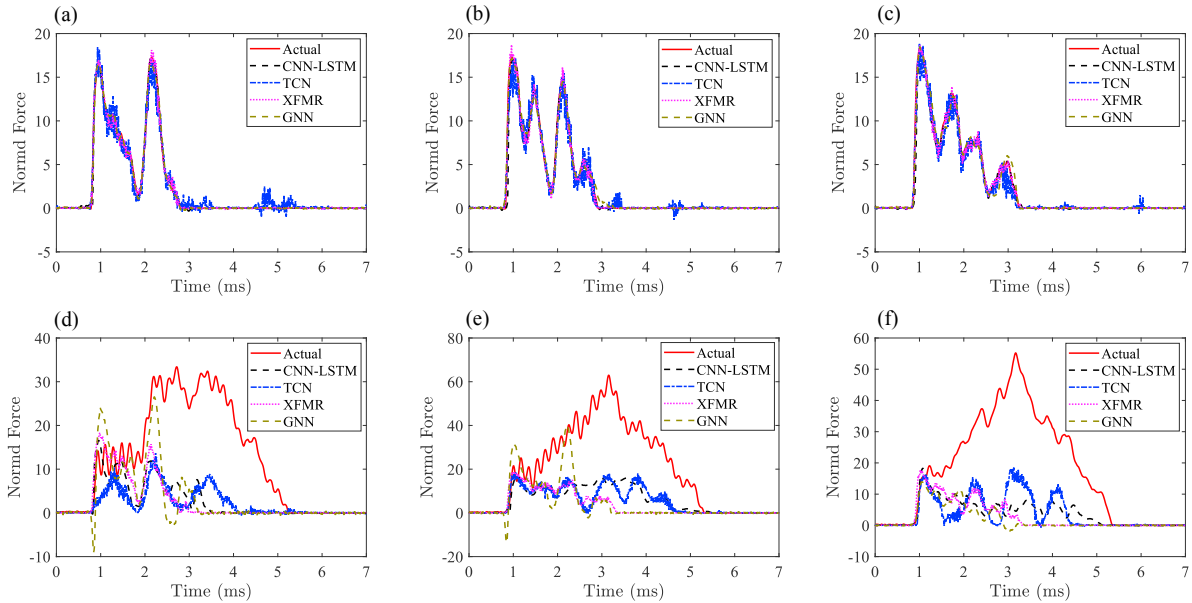


Fig. 6: Impact force reconstruction by deep learning models: (a)-(c) force reconstruction on SH training impacts at locations 2, 7, and 12, respectively; (d)-(f) force reconstruction on BH testing impacts at the same locations.

As shown in Fig. 6(a)-(c), all models except TCN successfully learned from SH data, with CNN-LSTM, XFMR, and GNN producing accurate force reconstructions. The TCN model, however, exhibited signs of overfitting, manifesting as high-frequency artifacts in its predictions, likely due to its limited receptive field and sensitivity to sampling frequency.

When applied to BH impacts, as shown in Fig. 6(d)-(f), all models failed to generalise effectively. The reconstructed force profiles suffered from significant underestimation of both amplitude and shape, with CNN-LSTM and TCN performing marginally better than XFMR and GNN in capturing impact duration. This systematic underestimation is attributable to the mismatch in force-to-signal ratio between the SH training and BH testing data. The models, having learned mappings biased toward the lower force-to-signal ratios of SH impacts, could not extrapolate to the higher ratios observed in BH cases.

Additionally, differences in predicted impact durations highlight the models' varying abilities to generalise temporal characteristics. XFMR and GNN models tended to reproduce impact durations similar to those seen in the training set, significantly underestimating the actual longer durations of BH impacts. In contrast, CNN-LSTM and TCN models showed improved generalisation of duration but still struggled with amplitude accuracy.

In summary, all evaluated deep learning models demonstrated limited extrapolation capability for force reconstruction under varying impact masses. These results underscore the critical influence of impactor mass on both the dynamic response and the sensor measurement characteristics. Accurate force reconstruction in such scenarios requires models capable of capturing nonlinear relationships and generalising across a broad spectrum of impact dynamics—a challenge for conventional supervised deep learning approaches trained under narrow operating conditions.

5. Conclusion and outlook

This study systematically investigated the performance and generalisation capabilities of various deep learning models for impact localisation and force reconstruction in composite structures, particularly under environmental and operational variability (EOV). Specifically, the effects of temperature shifts and impact mass variations were examined

using a diverse experimental dataset comprising impacts from small-mass and large-mass hammers under multiple conditions. The models evaluated included CNN, TCN, CNN-LSTM, GNN, and Transformer (XFMR), enabling a comprehensive comparison of architectures with differing temporal, spatial, and attention-based capabilities.

The main conclusions of this work are summarised as follows:

- The Transformer-based model (XFMR) consistently achieved the best localisation performance and robustness to EOV, especially under unseen impact mass and temperature shift scenarios.
- Impact mass significantly affects both the spectral content of sensor signals and the relationship between force and sensor response. BH (large-mass) impacts exhibited more low-frequency dominant responses and a higher force-to-signal ratio than SH (small-mass) impacts.
- The force-to-signal ratio was found to be nonlinear across impact conditions, indicating that a fixed linear mapping between sensor signals and force magnitude does not generalise well. This challenges conventional assumptions in data-driven force reconstruction.
- Deep learning force reconstruction models trained on SH data struggled to extrapolate to BH impacts. All models severely underestimated impact force amplitude and duration under mass variation, highlighting the sensitivity of learned models to training data distribution.
- CNN-LSTM and TCN showed slightly better generalisation of impact duration than XFMR and GNN; however, none of the models could accurately reconstruct force profiles for unseen mass conditions, indicating limited robustness in extrapolation tasks.

Building upon the findings of this study, several promising research directions are proposed:

- **Physics-Informed Learning:** Incorporate physics-based constraints, such as sensor frequency response models, dispersion relationships, and contact mechanics, into the training of neural networks to improve extrapolation across EOVs.
- **Domain Adaptation and Transfer Learning:** Explore transfer learning techniques or unsupervised domain adaptation strategies to improve model robustness under EOV by aligning latent feature distributions between source (training) and target (testing) domains.
- **Uncertainty Quantification:** Integrate Bayesian deep learning or ensemble methods to capture epistemic and aleatoric uncertainties, enabling more reliable predictions in real-world SHM applications.
- **Multi-modal and Hybrid Input Representations:** Investigate the use of combined time–frequency representations (e.g., CWT) and raw signals to enrich model input and enhance feature extraction under complex impact dynamics.
- **Adaptive Digital Twin Integration:** Incorporate the deep learning framework into a closed-loop, EOV-adaptive digital twin system for continuous structural monitoring, diagnosis, and prognosis.
- **Generalisation across Geometry and Boundary Conditions:** Extend the experimental validation to different composite geometries, layouts, and support conditions to test the scalability and universality of the proposed models.

This study lays the groundwork for building resilient, adaptive, and interpretable data-driven impact identification systems for composite aerostructures operating under complex real-world conditions.

Declaration of competing interest

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

Acknowledgements

The first author acknowledges the financial support received from the China Scholarship Council (CSC) for his doctoral studies under Scholarship No. [2021]339. The authors also extend their gratitude to Dr. Aldyandra Hami

Seno^{1D}, a former Ph.D. member of the research group, for generating the guided drop mass experimental data under temperature variations (TEM), which were utilised in this study.

Data prior usage and availability

Data will be made available on request.

References

- [1] Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., Inman, D.J., 2021. 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* 147, 107077. doi:[10.1016/j.ymssp.2020.107077](https://doi.org/10.1016/j.ymssp.2020.107077).
- [2] Azimi, M., Eslamlou, A.D., Pekcan, G., 2020. Data-Driven Structural Health Monitoring and Damage Detection through Deep Learning: State-of-the-Art Review. *Sensors* 20, 2778. doi:[10.3390/s20102778](https://doi.org/10.3390/s20102778).
- [3] Sun, H., Song, L., Yu, Z., 2023. A deep learning-based bridge damage detection and localization method. *Mechanical Systems and Signal Processing* 193, 110277. doi:[10.1016/j.ymssp.2023.110277](https://doi.org/10.1016/j.ymssp.2023.110277).
- [4] Wang, Y., Zhao, Y., Addepalli, S., 2020. Remaining Useful Life Prediction using Deep Learning Approaches: A Review. *Procedia Manufacturing* 49, 81–88. doi:[10.1016/j.promfg.2020.06.015](https://doi.org/10.1016/j.promfg.2020.06.015).
- [5] Xiao, D., Khodaei, Z.S., Aliabadi, M.H.F., 2024. Impact Identification Based on Surrogate-assisted Efficient Global Optimisation. *Procedia Structural Integrity* 52, 667–678. doi:[10.1016/j.prostr.2023.12.067](https://doi.org/10.1016/j.prostr.2023.12.067).
- [6] Xiao, D., Sharif-Khodaei, Z., Aliabadi, M.H., 2025. Robust impact localisation on composite aerostructures using kernel design and Bayesian fusion under environmental and operational uncertainties. doi:[10.48550/arXiv.2501.18393](https://doi.org/10.48550/arXiv.2501.18393). arXiv:2501.18393 [stat].
- [7] Xiao, D., Sharif-Khodaei, Z., Aliabadi, M.H., 2024. Impact force identification for composite structures using adaptive wavelet-regularised deconvolution. *Mechanical Systems and Signal Processing* 220, 111608. doi:[10.1016/j.ymssp.2024.111608](https://doi.org/10.1016/j.ymssp.2024.111608).
- [8] Tabian, I., Fu, H., Sharif Khodaei, Z., 2019. A Convolutional Neural Network for Impact Detection and Characterization of Complex Composite Structures. *Sensors* 19, 4933. doi:[10.3390/s19224933](https://doi.org/10.3390/s19224933).
- [9] Zhao, B., Zhang, Y., Liu, Q., Qing, X., 2024. Impact monitoring of large size complex metal structures based on sparse sensor array and transfer learning. *Ultrasonics* 140, 107305. doi:[10.1016/j.ultras.2024.107305](https://doi.org/10.1016/j.ultras.2024.107305).
- [10] Hamadneh, M., Mustapha, S., Fakihi, M.A., 2024. Impact-force identification using deep learning and Bayesian inference with application on pipeline structures. *Structural Health Monitoring*, 14759217241297572doi:[10.1177/14759217241297572](https://doi.org/10.1177/14759217241297572).
- [11] Hesser, D.F., Mostafavi, S., Kocur, G.K., Markert, B., 2021. Identification of acoustic emission sources for structural health monitoring applications based on convolutional neural networks and deep transfer learning. *Neurocomputing* 453, 1–12. doi:[10.1016/j.neucom.2021.04.108](https://doi.org/10.1016/j.neucom.2021.04.108).
- [12] Zhou, R., Qiao, B., Liu, J., Cheng, W., Chen, X., 2024. Impact force localization and reconstruction via gated temporal convolutional network. *Aerospace Science and Technology* 144, 108819. doi:[10.1016/j.ast.2023.108819](https://doi.org/10.1016/j.ast.2023.108819).
- [13] Zhao, Z., Chen, N.Z., 2023. Spatial-temporal graph convolutional networks (STGCN) based method for localizing acoustic emission sources in composite panels. *Composite Structures* 323, 117496. doi:[10.1016/j.compstruct.2023.117496](https://doi.org/10.1016/j.compstruct.2023.117496).
- [14] Yan, H., Xie, W., Gao, B., Yang, F., Meng, S., 2025. A deep learning approach to impact localization and uncertainty assessment in CFRP composites using sparse PZTs: Integrating experiments and simulations. *Thin-Walled Structures* 212, 113143. doi:[10.1016/j.tws.2025.113143](https://doi.org/10.1016/j.tws.2025.113143).
- [15] Huang, C., Liao, W., Sun, H., Wang, Y., Qing, X., 2023. A hybrid FCN-BiGRU with transfer learning for low-velocity impact identification on aircraft structure. *Smart Materials and Structures* 32, 055012. doi:[10.1088/1361-665X/acc623](https://doi.org/10.1088/1361-665X/acc623).
- [16] Zhou, J., Cai, Y., Dong, L., Zhang, B., Peng, Z., 2024. Data-physics hybrid-driven deep learning method for impact force identification. *Mechanical Systems and Signal Processing* 211, 111238. doi:[10.1016/j.ymssp.2024.111238](https://doi.org/10.1016/j.ymssp.2024.111238).
- [17] Huang, C., Tao, C., Ji, H., Qiu, J., 2023. Impact force reconstruction and localization using Distance-assisted Graph Neural Network. *Mechanical Systems and Signal Processing* 200, 110606. doi:[10.1016/j.ymssp.2023.110606](https://doi.org/10.1016/j.ymssp.2023.110606).
- [18] Huang, C., Tao, C., Ji, H., Qiu, J., 2025. Impact time history reconstruction using dynamic weighted graph neural network under sensor fault situations. *Structural Health Monitoring*, 14759217241313177doi:[10.1177/14759217241313177](https://doi.org/10.1177/14759217241313177).
- [19] Xiao, D., Sharif-Khodaei, Z., Aliabadi, M.H., 2024. Hybrid physics-based and data-driven impact localisation for composite laminates. *International Journal of Mechanical Sciences* 274, 109222. doi:[10.1016/j.ijmecsci.2024.109222](https://doi.org/10.1016/j.ijmecsci.2024.109222).
- [20] Xiao, D., Rodrigues, F.d.S., Sharif-Khodaei, Z., Aliabadi, M.H., 2025. A general probabilistic framework for impact localisation based on flexural wave propagation. *Mechanical Systems and Signal Processing* 226, 112320. doi:[10.1016/j.ymssp.2025.112320](https://doi.org/10.1016/j.ymssp.2025.112320).
- [21] Seno, A.H., Sharif Khodaei, Z., Aliabadi, M.H.F., 2019. Passive sensing method for impact localisation in composite plates under simulated environmental and operational conditions. *Mechanical Systems and Signal Processing* 129, 20–36. doi:[10.1016/j.ymssp.2019.04.023](https://doi.org/10.1016/j.ymssp.2019.04.023).
- [22] Olsson, R., 2000. Mass criterion for wave controlled impact response of composite plates. *Composites Part A: Applied Science and Manufacturing* 31, 879–887. doi:[10.1016/S1359-835X\(00\)00020-8](https://doi.org/10.1016/S1359-835X(00)00020-8).