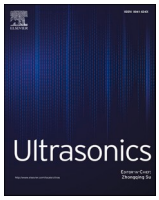
Ultrasonics 140 (2024) 107313

Contents lists available at ScienceDirect

Ultrasonics

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



Unsupervised machine learning for flaw detection in automated ultrasonic testing of carbon fibre reinforced plastic composites

Vedran Tunukovic [a,b,*], Shaun McKnight [a], Richard Pyle [a], Zhiming Wang [c], Ehsan Mohseni [a], S. Gareth Pierce [a], Randika K. W. Vithanage [a], Gordon Dobie [a], Charles N. MacLeod [a], Sandy Cochran [b], Tom O'Hare [d]

[a] Sensor Enabled Automation, Robotics, and Control Hub (SEARCH), Centre for Ultrasonic Engineering (CUE), Electronic and Electrical Engineering Department, University of Strathclyde, Glasgow, UK
[b] Future Ultrasonic Engineering, FUSE CDT, Glasgow, UK
[c] University of Strathclyde, Glasgow, UK
[d] Spirit AeroSystems, Belfast, UK

ARTICLE INFO

Keywords:
Automated defect detection
Machine learning
Deep learning in ultrasonic testing
Non-destructive evaluation
Carbon fibre reinforced polymers
Aerospace composites

ABSTRACT

The use of Carbon Fibre Reinforced Plastic (CFRP) composite materials for critical components has significantly surged within the energy and aerospace industry. With this rapid increase in deployment, reliable post-manufacturing Non-Destructive Evaluation (NDE) is critical for verifying the mechanical integrity of manufactured components. To this end, an automated Ultrasonic Testing (UT) NDE process delivered by an industrial manipulator was developed, greatly increasing the measurement speed, repeatability, and locational precision, while increasing the throughput of data generated by the selected NDE modality. Data interpretation of UT signals presents a current bottleneck, as it is still predominantly performed manually in industrial settings. To reduce the interpretation time and minimise human error, this paper presents a two-stage automated NDE evaluation pipeline consisting of a) an intelligent gating process and b) an autoencoder (AE) defect detector. Both stages are based on an unsupervised method, leveraging density-based spatial clustering of applications with noise clustering method for robust automated gating and undefective UT data for the training of the AE architecture. The AE network trained on ultrasonic B-scan data was tested for performance on a set of reference CFRP samples with embedded and manufactured defects. The developed model is rapid during inference, processing over 2000 ultrasonic B-scans in 1.26 s with the area under the receiver operating characteristic curve of 0.922 in simple and 0.879 in complex geometry samples. The benefits and shortcomings of the presented methods are discussed, and uncertainties associated with the reported results are evaluated.

1. Introduction

Composite materials are manufactured by combining two or more materials with differing characteristics. Various industries, including aerospace, biomedical, and renewable energy, heavily utilise composite materials owing to their superior fatigue resistance and design flexibility. Among these composites, Carbon Fibre Reinforced Plastics (CFRPs) stand out due to their high corrosion resistance and strength-to-weight ratio, while maintaining a lightweight structure [1]. Consequently, there has been a notable uptake in the adoption of CFRPs within the energy sector, particularly for the fabrication of critical components such as wind turbine blades. In the present context, it is estimated that the production of 1 kW of energy requires approximately 10 kg of CFRPs [2]. This estimate implies a substantial growth in CFRP production as the United Kingdom aims to deploy up to 50 GW of new offshore wind capacity by 2030 [3], a major increase compared to the current 13.8 GW of installed capacity. This surge in usage significantly extends the time required to conduct Non-Destructive Evaluation (NDE) procedures, both for verification of structural integrity of critical components and to facilitate preventive maintenance [4]. Similar challenges were identified in the aerospace industry, where flagship aircraft models such as the Airbus A350XWB and Boeing 787 allocate approximately 50 % of their

* Corresponding author at: Sensor Enabled Automation, Robotics, and Control Hub (SEARCH), Centre for Ultrasonic Engineering (CUE), Electronic and Electrical Engineering Department, University of Strathclyde, Glasgow, UK.
E-mail address: vedran.tunukovic@strath.ac.uk (V. Tunukovic).

https://doi.org/10.1016/j.ultras.2024.107313
Received 3 November 2023; Received in revised form 5 March 2024; Accepted 2 April 2024
Available online 6 April 2024

structural weight to CFRPs [5,6,7], a substantial increase from previous models like the A380, where composite materials represented only 20 % of the structural weight [8].

Non-Destructive Evaluation describes a collection of physical measurement methods employed to evaluate the integrity of components and materials of importance without causing any damage. Among these methods, Ultrasonic Testing (UT) is the most frequently used bulk inspection method due to its application flexibility and safety [9,10]. During UT, the ultrasonic transducer is coupled to the target material before producing an ultrasonic pulse. This pulse travels throughout the material while interacting with the material's structure and potential discontinuities, scattering the acoustic wave with some returning to the transducer, which are then recorded and analysed. Phased Array Ultrasonic Testing (PAUT) is an improvement over the traditional UT methods, offering the implementation of advanced imaging techniques like the total focusing method, and the advanced electronic beamforming such as linear scanning, and beam steering [11,12,13]. These methods are facilitated by introducing delays into the transmission/reception channels of the PAUT arrays. Currently, PAUT is increasingly employed in automated NDE inspections within the renewable energy [14] and aerospace [15] industries. The data captured through such systems can be presented in various formats, including B-scans or C-scans, each format offering a distinct way of visualizing the outcomes of the inspection.

At the same time, advancements in the field of robotics profoundly influenced the NDE field, especially through the development of setups based on industrial robotic manipulators. These setups have greatly improved the precision and repeatability of the NDE processes, while reducing the need for certified manual inspectors [16]. However, while automated NDE pipelines accelerated the data collection, data interpretation is still commonly performed manually, increasing the risks of human error and causing a current bottleneck in the manufacturing process due to the long processing times [17,18]. Therefore, automated data interpretation tools designed to aid the decision-making process used in conjunction with automated robotic delivery systems would be a key enabler to realise the full potential of industrial automation.

Machine Learning (ML) involves algorithms that improve in performance with exposure to data. Deep Learning (DL) is a subset of ML, centred around the advancement of complex models that use layers of processing to extract features from the input data. Although these methods have found their way into the NDE field with a wave of research studies concerning the application of ML to automated defect detection and characterisation in the past decade, the research concerning UT data analysis for composite materials remains relatively limited in comparison to the extensive work conducted on the inspection of metallic bulk materials and welds.

In a comprehensive comparative study [19], a range of DL and traditional methods were investigated to extend and enhance UT training datasets through the synthetic generation of UT data. The study showcased that a generative adversarial network brought significant improvements to the datasets created with simulation software, effectively bridging the gap between the artificial and real UT responses in CFRPs. In [20], authors have analysed different ML models and feature extractors on A-scan data acquired from CFRP samples. The samples had embedded artificial defects, enabling the collection of both undefective and defective ultrasonic signals. 3D braided composites were used in [21], where researchers employed convolutional networks to classify A-scan data to identify debonding. Data produced with ultrasonic guided waves was used in [22], with the task of estimating fatigue damage. Lastly, researchers in [23] focused their efforts on improvements in feature extractions by adopting alternative signal decomposition methods.

Autoencoders (AE) represent a category of DL networks that have found applications in the detection of potential security threats [24], denoising of data [25], and undertaking various NDE tasks [26,27,28,29], among many others. AEs can be divided into three distinct components: the encoder, the latent space, and the decoder. The encoder's task is to process the original input data with a series of layers into a representation of features known as the latent space. The decoder part of the network aims to approximate the inverse of the encoder, taking the latent space as an input, and attempting to reconstruct the original inputs. AE as a concept has seen multiple iterations and improvements over the years, including: a) GANomaly [24] where additional encoder and discriminator structures are introduced, b) Variational Autoencoders (VAEs) [30] where the latent space is a statistical distribution, and c) U-NET models that adopt encoder-decoder structure and are used for segmentation of medical images [31]. In the scope of this study, attention is directed toward the utilisation of AEs as unsupervised defect detectors. The fundamental structure of a basic autoencoder is illustrated in Fig. 1.

Anomaly detection is a method used in data science, statistics, and ML, primarily focused on identifying abnormal patterns or outliers within a dataset. Oftentimes data follows regular patterns or can be approximated by specific statistical distributions. Anomaly detectors aim to capture these regularities while extracting outliers or data points that fall outside the probable distribution. The research of ML in the domain of NDE faces significant challenges due to data scarcity. Specifically, acquiring genuine defective data is often accompanied by costly manufacturing and testing procedures. AE-based anomaly detectors offer the advantage of being trainable and deployable in an unsupervised manner, diminishing the need for extensive acquisition and labelling of data containing defective samples. Instead, training can be based on undefective data, allowing the AEs to learn distributions and representative features of undefective samples. Their performance can then be evaluated with the reconstruction error observed between the input and output data, as undefective data should exhibit successful reconstruction with lower errors in comparison to defective data.

In the field of NDE research, AEs have been previously used as anomaly (defect) detectors and denoising mechanisms. In [26], authors have developed a VAE model to detect defects within ultrasonic B-scans of bulk metallic materials. To enhance training and performance, taking inspiration from GANomaly, another encoder was added to the AE structure after the decoder. The study revealed that the model successfully identified larger defects, but struggled with smaller defects that cause minor reconstruction errors. Ultrasonic scans of rails were explored in [32] and [27]. In [32], AEs were deployed on a dataset comprising ultrasonic guided waves A-scans, achieving promising results. In [27], the authors have demonstrated that AE structures work well in the identification of different flaws that are visible in ultrasonic B-scans. The consistent geometry of inspected material made the application of signal gating easier, greatly simplifying the dataset and positively impacting the final performance. In [29], authors focused on the ultrasonic dead zone, an area of an ultrasonic pulse close to the transmission source that can heavily mask reflections from near-field features. Positive results were achieved in the identification of near-surface defects, with the recommendation that the method be further validated on different material specimens and alternative defect types. In [28], through-transmission UT was used for the task of identification of defective adhesive bonds. Explainable anomaly scores were demonstrated as the sigmoid activation function was added to the calculated Mean Squared Error (MSE) between inputs and outputs of the model, presenting the anomaly score as a percentage. Overall, the authors have reported valuable quantitative results, but the scope of the study was limited. Lastly, in [25], AEs were utilised as a tool to denoise A-scans before classification. This method yielded great success as the classification performance was improved.

During the NDE of CFRP components, various signal processing techniques are used to improve the interpretability and quality of visualisation. The most frequently used processing techniques include the Hilbert transform [33], a mathematical method of capturing the envelope of the signal, and signal gating, an application of the windowing technique in the time domain to extract areas of interest from the signal.

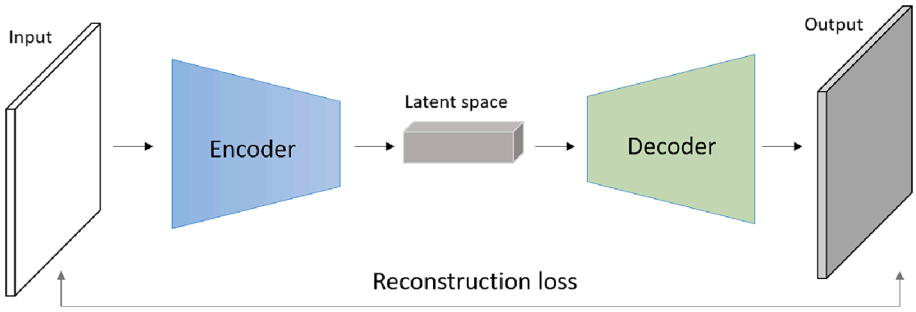


Fig. 1. Basic autoencoder structure.

This stage is crucial in preparing C-scan and B-scan images because, with appropriate gating, the surface and backwall echoes, which often have significantly higher amplitudes compared to potential volumetric defects, can be excluded from the selected time window. This results in images that provide better visibility for lower amplitude features. The gating process is usually performed by a human operator while automatic gating is rarely discussed in the academic literature. In [34], authors developed a back wall echo filter method based on the computation of gradients of Time-Of-Flight (TOF) variations and thickness tolerances. However, this method is incompatible with complex and stepped materials. Authors in [35] have introduced an automated UT analysis software that performs automated gating in two steps. The proposed approach is relatively complex and was tested only on samples with smooth and slight curvatures. Authors expanded the previous work in [36] by incorporating the Amplituden Laufzeit Ortskurven algorithm, characterising the front wall responses as echoes with the overall smallest TOF while other echoes correspond to defects, back wall, or repetitions. While yielding positive results, this method is not an all-round solution for gating complex samples. In the study [37], authors achieved automatic gating on a per A-scan basis, by identifying amplitude peaks and assigning the leftmost peak label of the initial pulse and the rightmost peak the label of the back wall echo. However, this approach can misclassify echoes in the middle of the scan as defects in cases involving thinner samples with multiple back wall reflections. Lastly, in [38], the authors illustrated two adaptive gating processes for detecting defects: the first relies on the CAD model, and the second is grounded in back wall echo tracking. Both approaches draw upon external knowledge, either through possessing a comprehensive CAD model of the inspected specimen or, in the latter method, by manually establishing a back-wall gate width that is larger than the maximum thickness of the component.

Given the limited CFRP-focused research and the demonstrated potential of AE models, this study presents a two-step process to achieve defect detection of ultrasonic B-scan data. The initial step presents an alternative approach to the automatic gating process based on the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm. This method offers a robust and flexible automatic gating that streamlines the next phase of the process. Subsequently, a convolutional-based AE is employed to rapidly identify defective data within the UT dataset, which was collected through an automated robotic system. This system ensures precise and repeatable sensor delivery, mirroring the technology currently employed within the industry. Lastly, a study of uncertainties was conducted to identify potential shortcomings of human labelling and the repeatability of ultrasonic scans. The combination of outlined methods provides insight into the potential of DL to tackle current data interpretation bottlenecks in the industry. To this end, proposed methods are evaluated across a diverse range of CFRP samples with varying characteristics. The developed model successfully identified 38 out of 40 embedded defects in the simple geometry samples, and 22 out of 24 defects present in the complex geometry sample. The inference of 2070B-scans was performed in 1.26 ± 0.09 s on a GPU accelerated desktop machine.

In practical industrial applications, the manual analysis and interpretation of UT scans typically start with a focused examination of sectioned C-scans. If these analyses reveal defects surpassing predefined area limits established by industrial guidelines, subsequent analysis of individual B-scans is performed to further examine such areas. In an industrially representative example, considering a pristine sample comprising approximately 4500 individual B-scans, an NDE inspector expends around 1.5 h to complete the analysis. In contrast, an additional hour is added to the process when defective areas are found. This additional time is allocated to the inspection of individual B-scans and generation of quality report. It is crucial to note that the individual inspection of every B-scan is practically unattainable within a reasonable timeframe. Therefore, the underlying idea behind the proposed method is to serve as a supplementary tool for NDE inspectors, enabling processing of all B-scans without incurring significant computational costs.

The rest of the paper is organised as follows: Section 2 covers the materials and methods that were used, and Section 3 provides results and a discussion of all evaluated methods. Lastly, Section 4 provides the conclusions and the prospects for future work.

2. Materials and methods

2.1. Acquisition of experimental data

The scope of this study encompassed the examination of 11 CFRP samples of varying characteristics. These samples were manufactured in accordance with the BAPS 260 standard and supplied by Spirit AeroSystems, UK and are listed in Table 1. The process involved the utilisation of Cycom 890 resin in conjunction with woven carbon fibre fabric sheets. All but three samples were undefective, enabling the acquisition of a relatively large number of healthy ultrasonic scans that were used for training the DL models. The three samples with manufactured defects were used to generate three different testing datasets. Drilled Flat Bottom Holes (FBHs) along with embedded Teflon inserts and bagging film inserts enabled the acquisition of ultrasonic signals with defect indications that partially mimic naturally occurring defects during the manufacturing process [39]. In defective sample A, a range of

Table 1
The range of defective/undefective CFRP samples used in this study.

Sample	Dimensions [mm]	Thickness [mm]	B-scans total [-]	Defective B-scans [-]	Used in
1	254.0 x 254.0	2.20	1000	N/A	Training
2	254.0 x 254.0	2.14	1000	N/A	Training
3	254.0 x 254.0	2.75	750	N/A	Training
4	254.0 x 254.0	2.75	1000	N/A	Training
5	254.0 x 254.0	4.25	1000	N/A	Validation
6	254.0 x 254.0	4.25	1000	N/A	Training
7	254.0 x 254.0	6.00	1000	N/A	Training
8	254.0 x 254.0	6.00	1250	N/A	Training
A	254.0 x 254.0	8.60	750	153	Testing
B	254.0 x 254.0	8.60	1150	239	Testing
C	780.0 x 197.0	7.50 – 16.0	2070	215	Testing

15 FBHs were introduced with diameters ranging from 3.0, 6.0, and 9.0 mm with a +/- 0.2 mm tolerance at depths of 1.5, 3.0, 4.5, 6.0, and 7.5 mm with +/- 0.3 mm tolerance. Sample B was treated similarly, with the addition of 4.0 and 7.0 mm FBHs, totalling 25 defects. Lastly, sample C was a large, stepped specimen with dimensions 780 x 200 mm and thicknesses ranging from 7.5 to 16.0 mm. At each step, three 6.0 x 6.0 mm Teflon and three 6.0 x 6.0 mm bagging film inserts were used, positioned close to the front surface, in the middle of the sample, and close to the back wall of the sample. Overall, 64 manufactured defects were examined in this study.

The automated acquisition setup was based on an industrial manipulator KUKA KR90 R3100 extra HA [40], a Schunk GmbH & Co. FTN-GAMMA-IP65 SI-130–10 Force Torque (FT) sensor [41], and an Inspection Solutions RollerFORM-5L64 [42] phased array. The assembly used for this study is illustrated in Fig. 2.

The usage of industrial manipulators enabled precise probe delivery, with pose repeatability of +/- 0.06 mm. The KUKA KR90 model provided a maximum reach of 3095 mm, accompanied by a payload capacity of 90 kg. The movements, path planning, and real-time end tool pose corrections based on the FT sensor feedback were programmed and controlled within a LabVIEW Virtual Instrument (VI) program on a desktop PC linked with the robotic controller via the ethernet cable. The velocity of the industrial manipulator was set at 10 mm/s, ensuring smooth movement and a consistent contact quality of the mounted phased array probe on the inspected sample. The important benefit of using the robotic setup is encoding the captured UT data through the recorded robotic positions, ultimately creating a 3D scan of the volume of the scanned part.

Continuous vertical movement monitoring and control of the end effector was established using the FT sensor, capturing the values of forces normal to the inspection surface of test samples. The used FT sensor boasts the capability to capture forces in three dimensions, with a maximum load of 400 N and 130 N in vertical and horizontal directions respectively. FT sensing served several purposes; firstly, by providing continuous readings to the central VI, real-time pose corrections were dictated to the robotic controller ensuring the constant force between the phased array and the surface of the sample; secondly, it served as a fail-safe system to protect the phased array in case that high force readings occurred during the scan. Stable contact force together with the application of sufficient liquid couplant on the surface is crucial for successful UT, as these ensure that no air pockets are present between the tyre of the phased array roller-probe and the sample. Air acts as a strong reflector, disallowing the propagation of the ultrasonic beam from the roller probe into the test specimen. To this end, besides the adaptive end-effector behaviour facilitated by FT sensing, water was also sprayed as a thin film over the sample surface as a couplant.

For UT inspection, the 5 MHz phased array roller probe comprised 64 elements, with 0.8 mm pitch and 6.4 mm elevation. The array's transmission/reception was controlled with Peak NDT Ltd. MicroPulse 6 unit [43], a controller with 128 transmission and reception channels, allowing for a design of custom focal laws to drive the phased array. A linear electronic scanning mode was used, with an excitation voltage of 80 V and a sub-aperture of 4 elements providing an active aperture of 48.8 mm. The pulse width was set at 100 ns, with a 760 Hz pulse repetition rate. To filter out the unwanted higher-frequency signals, a digital 6 MHz lowpass filter in the unit was activated. The digitiser was set to acquire data with a 100 MHz sample rate and 32-bit precision. Upon reception, a gain of 22.5 dB was applied to the signal with the addition of Time Compensated Gain (TCG). TCG was used to account for losses of amplitude that occur due to high attenuation, scattering, and other physical interactions between the transmitted ultrasonic wave and the inspected CFRP material. The shape of TCG was a linear ramp function, starting at the depth of 1.5 mm, and ending with 23.75 dBs of additional gain at the depth of 15 mm. The complete automated NDE setup is illustrated as a block diagram in Fig. 3.

All the collected data were stored within a three-dimensional array structure, assuming the format of [B-scans, Time samples, A-scans] as illustrated in Fig. 4. The pulse repetition frequency and robotic speed was set to achieve B-Scan capture every 0.8 mm. Each B-scan consisted of 61 A-scans resulting from an electronic sweep with a sub-aperture of 4 elements. Values of each scan were normalised, with respect to the maximum occurring amplitude in the conducted scan. Afterwards, a Hilbert transform was performed and both normalised and Hilbert-processed data of the same scan were stored.

2.2. Machine Learning

Before conducting experiments, a preliminary small grid-based search was performed to assess the range of hyperparameters and potential architectures. To ensure compatibility between the encoder and decoder, a reflective padding technique was applied before passing the B-scans to the model. This adjustment expanded the input dimensions [batch size, 61, 1000] to the nearest multiple of 32, allowing for flexibility in the input sizes the model could accommodate.

The encoder part of the network consists of four convolutional layers, each coupled with a batch normalisation layer [44] and a hyperbolic tangent activation function. Convolution was performed with square kernels of size 7 and stride of 2, resulting in dimensionality reduction as each convolutional layer reduced the input size by a factor of 2. The integration of batch normalisation served to mitigate the risk of overfitting, whereas the hyperbolic tangent activation function was selected due to its alignment with the amplitude extremities of raw B-scan data

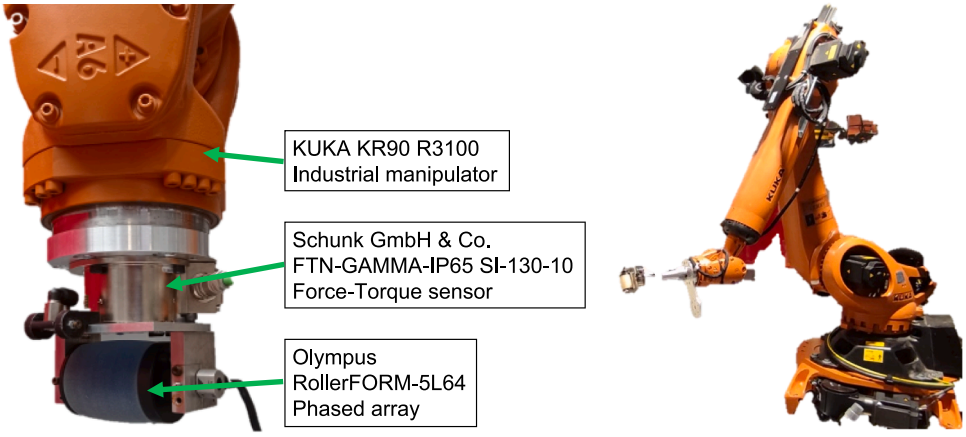


Fig. 2. Assembly used for automated NDE inspection (left) and KUKA KR90 industrial manipulator (right).

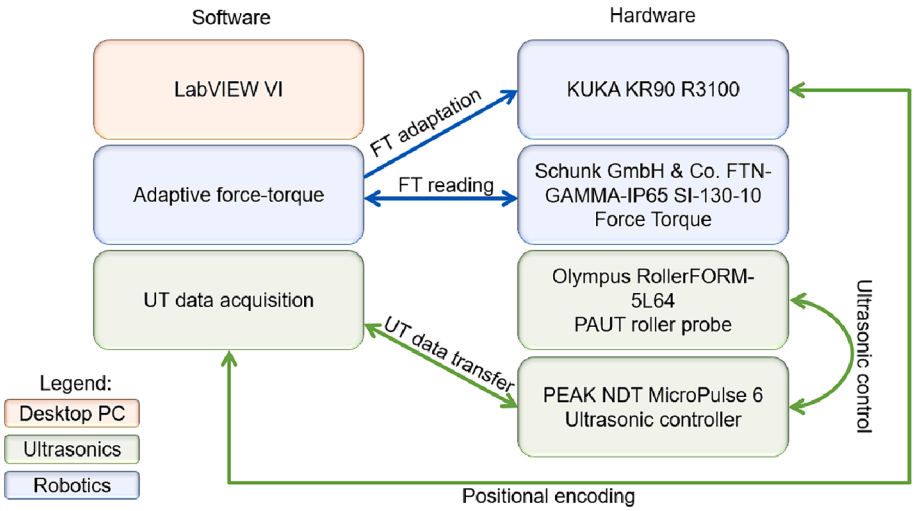


Fig. 3. System design of sensor-enabled robotic scanning with ultrasonic phased array roller probe.

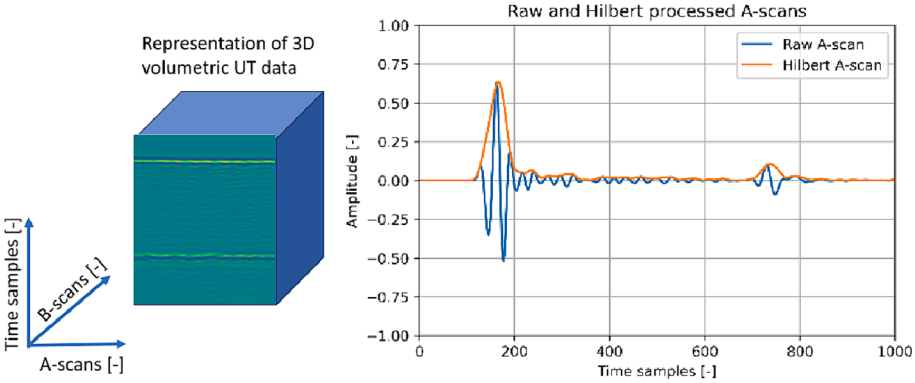


Fig. 4. Illustration of the collected data (left), raw and Hilbert-processed A-scans (right).

after normalisation to front wall response, which ranged between −1 and 1.

For the decoder architecture, bilinear interpolation upsampling layers with a factor of 2 were integrated. This upsampling was followed by the inclusion of a convolutional layer. Notably, this convolutional layer was configured with a kernel size of 7, a stride of 1, and padding to preserve the spatial dimensions of the feature maps. This approach deviates from the utilisation of transpose convolution blocks that are often used. The shift was induced due to the periodic artefacts that were observed in reconstructed images. These artefacts are a consequence of the overlap inherent in striding transpose convolutions, resulting in certain pixels being subjected to multiple passes by the kernel, while others receive only one pass. Such artefacts have been observed in various works such as [45,46]. The decision to adopt the upsampling approach, as suggested in [47], has proven effective in mitigating these artefacts, thus resulting in a clearer reconstruction.

The training process employed MSE as the chosen loss function, measuring the difference between the input and output images. During training, only undefective data was used for both training and validation. Measures to avoid overfitting also included the use of L2 regularisation in the form of weight decay which was set at 0.0001. For all trained models, ADAM optimiser was used [48], with β_1 and β_2 values of 0.9 and 0.999 respectively. To determine the length of training, early stopping with patience of 10 epochs was used on calculated MSE losses on the validation dataset. Lastly, a batch size of 64 individual B-scans was used, with a learning rate of 0.0005. All DL models in this study were trained and tested using a desktop Windows 11 PC with Nvidia RTX 3090 Ti GPU, 128 GB RAM, and two Intel® Xeon® Gold 6428 2.50 GHz CPUs. PyTorch library and Python 3.10.8 were used for coding. Full network architecture is presented in Fig. 5. For a comprehensive overview of ML vocabulary and respective definitions refer to [49].

A key benefit of the unsupervised method outlined in this work is its lack of reliance on positive examples of defects during training. Evaluating B-scans against an expected defect free reconstructing loss turns defect detection from a positive classification problem to an outlier identification/anomaly detection problem. While this method is unsuitable for defect classification, its strength lies in its robustness for defect detection. This is attributed to its training process, which did not involve the use of examples of defects, resulting in increased generalisability. To demonstrate the method's effectiveness, testing was conducted on various manufactured defects. While naturally occurring defect examples are limited, the generalisability of the method is expected to be robust for most defect types if the defect response yields an anomalous B-scan.

2.3. Automatic gating method

Following the data acquisition, prominent geometric patterns of higher amplitude were identified in front and back wall responses. The internal structure predominantly consists of much lower amplitude levels. When such data was used in conjunction with the AEs, it was difficult to distinguish defective and undefective B-scans in the dataset based on reconstruction errors, especially for B-scans containing smaller defects (these observations are further discussed in section 3). This

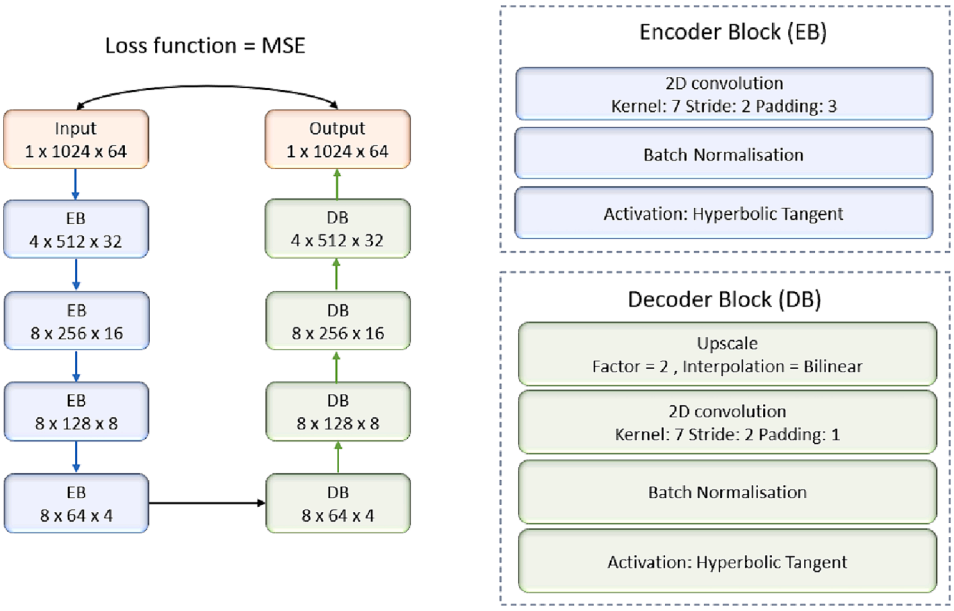


Fig. 5. Autoencoder architecture used in this study.

complexity arose from the fact that defects which occupy small areas of a B-scan, whilst locally producing large MSE around the defect, would often be lost in the global reconstruction error which considers differences from all individual pixels. Similar observations were reported in [26] where authors have developed a VAE that provides good defect detection when encountering large defects, while sometimes failing to reconstruct undefective images or identify smaller defects. Furthermore, the authors state that geometrical signals usually have large reconstruction errors which in turn cause false positive indications. An approach that provided the solution to this was presented in [27], where constant geometry of the scanned specimens allowed researchers to effectively apply gating of the captured ultrasonic signals to remove the front and back wall indications, thereby reducing the complexity of the data. An example of the undefective B-scan and the defective B-scan are presented in Fig. 6.

Although manual gating was possible, an automated gating setup would be beneficial as a range of samples of various thicknesses was used to generate training and validation datasets. In the aerospace industry, CFRPs are used for critical components such as fuselage, wing covers, spoilers, and stabilisers and they are manufactured in various geometries and material thicknesses to serve the functional purpose and meet the required performance criteria for different components. For instance, CFRPs are used for wing components with thickness variation, having thinner measurements towards the wingtip and thicker measurements near the root of the wing. For this specific application, thicknesses of around 24 mm are used [50] with wing spars reaching thicknesses of 40 mm or more [51]. An example of a complex geometry wing cover with stringers is illustrated in Fig. 7.

To this end, an automated gating approach that leverages the DBSCAN algorithm [52] was introduced. DBSCAN is a robust unsupervised ML clustering algorithm, characterised by two adjustable parameters: ε, which defines the maximum distance between a pair of data points for them to be considered neighbours, and *minimum_points*, a parameter specifying the minimum count of neighbouring data points necessary to form a distinct cluster. The proposed pipeline is initiated by the definition of minimum amplitude threshold and minimum distance between the peaks used in the peak-finding algorithm. The utilisation of

Fig. 7. Wing cover component with complex geometry and varying thickness.

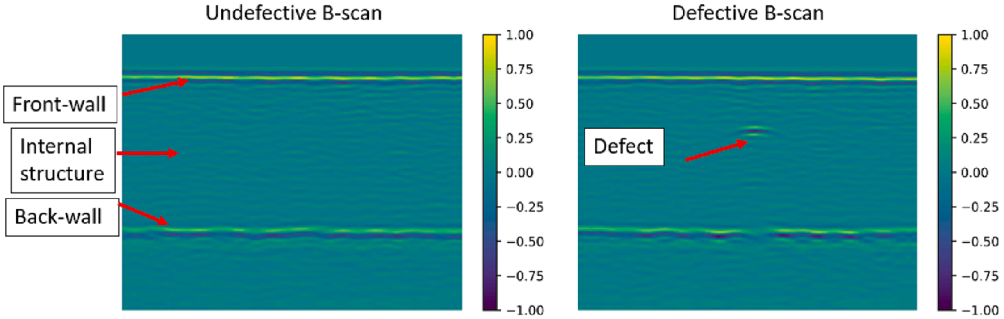


Fig. 6. Undefective B-scan (left) and defective B-scan (right).

peak finding algorithm was crucial to induce dimensionality reduction as clustering of the raw data is impractical due to the large number of data points, which results in poor outcomes. For this method, a full 3D data set [B-scans, Time samples, A-scans] was Hilbert transformed and processed at once. The minimum threshold amplitude used in peak finding was defined as:

$$Minimumthresholdamplitude = RMS(noise)*\alpha \quad (1)$$

where α signifies the scaling factor that adjusts the threshold in relation to the Root Mean Square (RMS) of the noise level. Next, the minimum peak distance is mathematically presented as:

$$Minimumpeakdistance = \frac{f_{sampling}}{f_{operating}}*\beta \quad (2)$$

where $f_{sampling}$ and $f_{operating}$ correspond to the sampling and operating frequency of the ultrasonic setup, respectively. The quotient of these values represents the wave packet, which is scaled by a tuneable parameter β. The rationale underlying this formulation is to decrease the overall number of identified peaks. This strategy is useful to produce a separation between the identified peaks in the front wall and subsequent peaks produced by stronger subsurface reverberations. An example of such separation between the clusters is presented in Fig. 8, where front wall peaks are coloured in blue while subsurface reverberations are indicated in red.

The specified parameters were implemented in the peak finding algorithm sourced from the Python Library SciPy [53], whose output served as input to the DBSCAN model. For flat samples, the two largest identified clusters corresponding to the front and back walls were chosen for the automatic gating process. However, in cases where samples exhibited complex stepped geometries, an additional user input variable n was introduced, representing the anticipated number of steps within the sample. This variable was used to extract produced indices of n number of clusters, starting with the one with most points.

In both scenarios, a loss of back wall response might occur due to the presence of reflective defects, which in turn could result in no detections from the peak finding algorithm in the back wall areas. Therefore, clusters that belong to the back wall were checked for such occurrences and interpolated, similar to the work presented in [31]. To finish the process, the identified clusters generated indices indicative of areas for exclusion during the gating process. This outcome yielded a scan focused exclusively on the internal portion of the material.

In experiments, a range of parameter values were tested to determine the optimal ones for the dataset used in the study. These parameters are listed in Table 2. Parameter α showed the largest influence on the overall results discovering that for thinner CFRP samples, all tested α values yielded satisfactory results. However, this was not true for the thicker samples, as higher α values generated a much higher threshold for peak detection and the back wall was not identified correctly. This was attributed to the high attenuation of CFRP materials (around 1.5 dB/mm), causing significant amplitude loss over longer acoustic paths. For β, higher values led to better separation between the front wall and subsurface reflections. For DBSCAN, smaller ε values produced better results, while higher values sometimes resulted in unwanted merging of distinctive clusters. *The minimum_points* parameter had very little to no impact on the overall performance.

Table 2
Tuneable parameters in the automatic gating process.

Parameters	Used in	Tested values	Selected value	Influence
α	Peak finding algorithm	3 – 10	4	High
β	Peak finding algorithm	1.5 – 6	6	Medium
ε	DBSCAN	7 – 12	7	Medium
minimum_points	DBSCAN	3 – 10	3	Low

This process offers a twofold advantage. Firstly, it serves as a valuable component within the NDE inspection pipeline, facilitating the automated generation of C-scans by excluding the front and backwall echoes. Secondly, the process contributes to a reduction in data complexity in the interest of the DL process, focusing exclusively on the material's internal structure. This enhances the effectiveness of subsequent approaches in detecting defects with greater ease and accuracy.

2.4. Performance metrics

In this study, a set of performance metrics to evaluate trained models were considered. Among these metrics, the False Positive Rate (FPR) and True Positive Rate (TPR) were used for assessing the accuracy of binary classification tasks and are given by:

$$FPR = \frac{FalsePositives(FP)}{TrueNegative(TN) + FalsePositives(FP)} \quad (3)$$

$$TPR = \frac{TruePositives(TP)}{TrueNegatives(TN) + FalsePositives(FP)} \quad (4)$$

FPR, denotes the ratio of undefective scans incorrectly identified as defective. On the other hand, TPR quantifies the rate at which models correctly identify defective B-scans as such. Together, FPR and TPR form the foundation for constructing the Receiver Operating Characteristic (ROC) curve, a graphical representation for binary classification tasks. Mathematically, the coordinates of ROC curve can be presented as:

$$ROCcurve = \{(FPR_1, TPR_1), (FPR_2, TPR_2), \cdots, (FPR_n, TPR_n), \} \quad (5)$$

The ROC curve enables the visualisation of the trade-off between TPR and FPR at various thresholds. Furthermore, it enables the comparison between the presented methods with Area Under the Curve (AUC) metric. AUC condenses the model's overall performance into a single

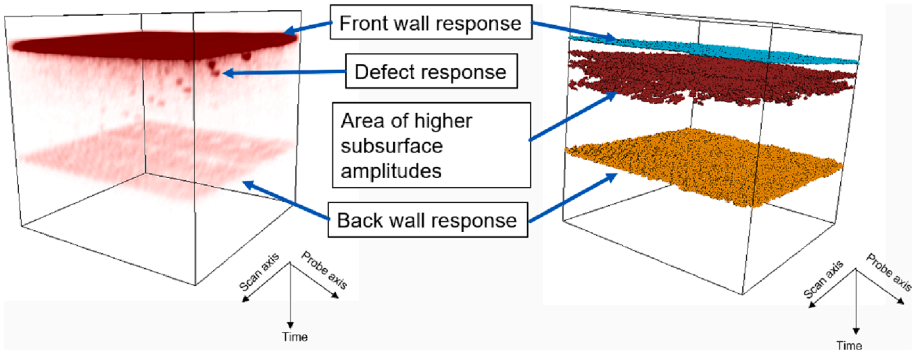


Fig. 8. Example of the automatic gating progress; Hilbert processed 3D data (left), DBSCAN formed clusters (right).

scalar value and is given through:

$$ROCAUC = \int ROCCurve(FPR, TPR)dFPR \tag{6}$$

3. Results and discussion

3.1. Training and deployment on simple geometry samples

Training on both ungated and gated datasets yielded consistent outcomes in terms of evaluation metrics during inference, although the point of training convergence exhibited variability. To investigate the effect of the stochastic nature of the training process and the random initialisation of weights and biases on performance, ten training repetitions on both datasets were conducted. In the case of training on ungated datasets, convergence was typically achieved after an average of 118 epochs, with a standard deviation of 62 epochs. Similarly, for gated datasets, the training process converged on average at 122 epochs, with a standard deviation of 77 epochs. The relatively high standard deviation highlights the neural network training's sensitivity to the initial weight and bias values. Nonetheless, it is important to note that while the random initialisation may impact the time taken for convergence, the performance of the models after convergence was not substantially influenced, showing good convergence to a global minimum. An example of training convergence for both datasets is presented in Fig. 9, alongside schematics of samples A and B.

When assessing the ungated data reconstruction losses for samples A and B during the inference, several observations were noted. For sample A, both defective and undefective B-scans were reconstructed with comparable reconstruction loss levels, giving rise to interpretation difficulty as the separation between data was not discernible. For sample B, slightly improved outcomes were observed as spikes in reconstruction losses for shallow defects, such as the 3.0, 4.0, 6.0, 7.0, and 9.0 mm diameter defects were identifiable. However, the deeper defects remained undetected and were masked with relatively higher reconstruction errors of undefective B-scans. ROC AUC scores of 0.763 and 0.863 were achieved for samples A and B, respectively. From an NDE application viewpoint, this performance would be deemed unsatisfactory as important defects would remain undetected. The reason for the poor performance was that the reconstruction error associated with the front wall and back wall echoes outweighed the error corresponding to the acoustic responses of defects in the internal part of the material. Reconstruction losses for samples A and B are presented in Fig. 10.

When using gated datasets during training, as suggested by the results presented in Fig. 11, major detection improvements were observed. In both samples, all 6.0, 7.0, and 9-mm diameter defects produced evident spikes in the reconstruction loss making them easily identifiable. Two 4.0 mm diameter defects located close to the front wall also produced elevated reconstruction error values, but three deeper defects produced very small deviations from the undefective B-scans. Lastly, 3.0 mm diameter defects produced the smallest reconstruction losses, with three defects closest to the surface producing small but visible spikes. The last two 3.0 mm diameter defects in both samples were not identified. ROC AUC scores were improved, with 0.920 and 0.922 for sample A and B respectively as compared to the ungated data, indicating a clear improvement. ROC curves are presented in Fig. 12.

In industrial settings, the typical practice involves establishing a threshold to distinguish between defective and undefective samples. This approach is flexible and can be adjusted based on the specific requirements of the industry. For example, in NDE of critical components where defect detection is an imperative, the threshold may be set more aggressively, even if it allows for some false positive indications.

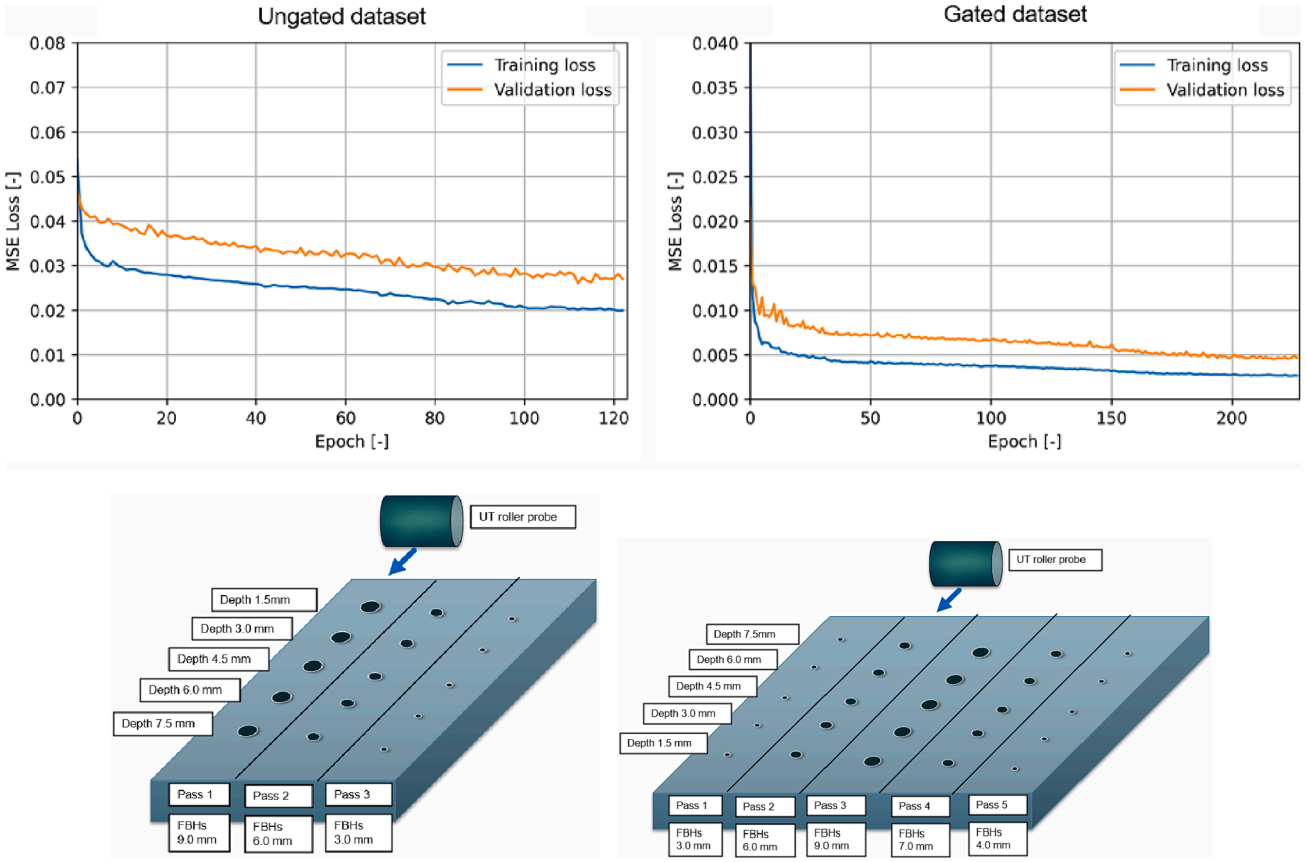


Fig. 9. Training and validation losses for ungated dataset (top left), gated dataset (top right), schematics for sample A (bottom left), and sample B (bottom right).

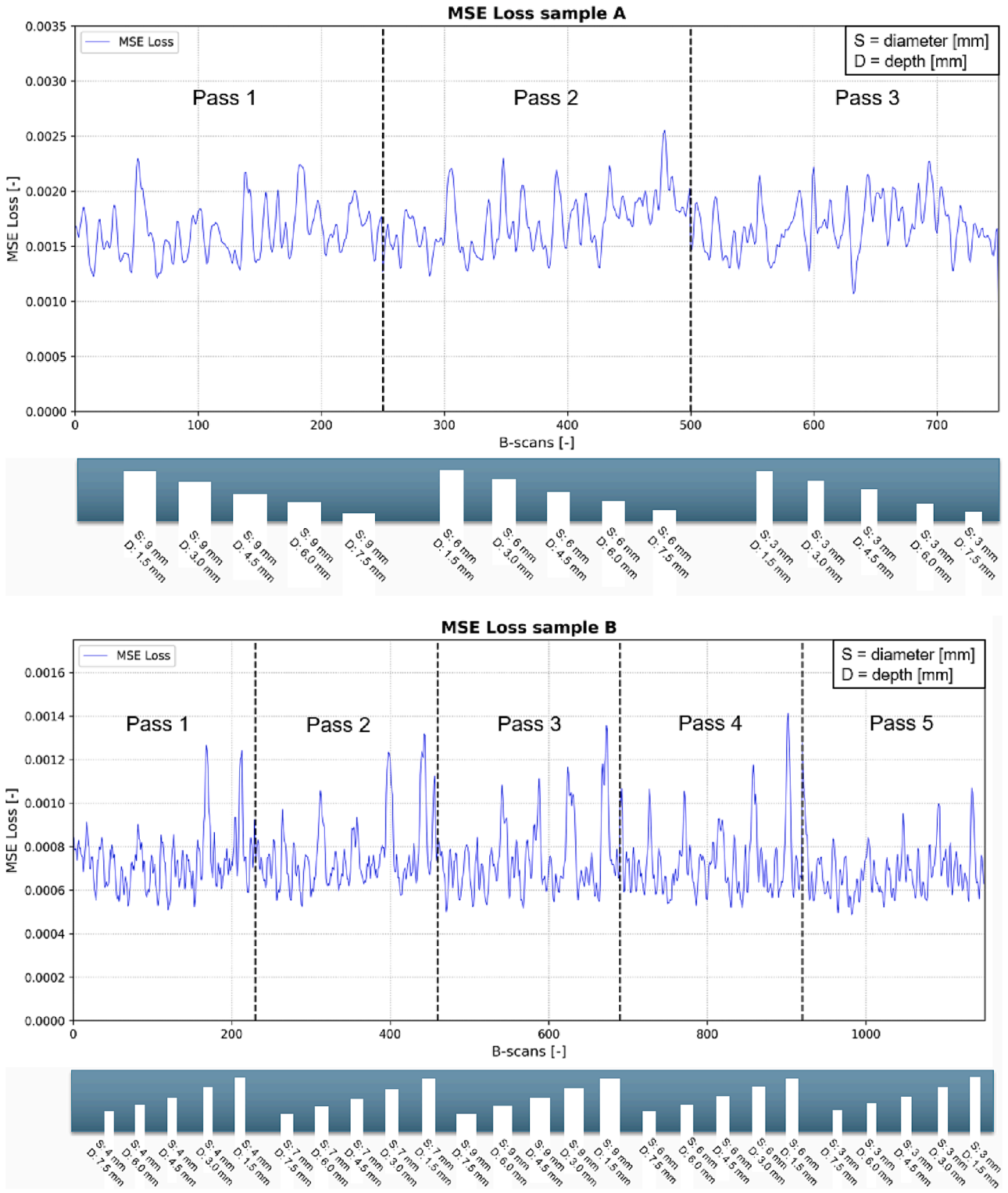


Fig. 10. Reconstruction losses and side view schematics for sample A (top) and for sample B (bottom) for ungated dataset.

3.2. *Uncertainties associated with the repeatability of ultrasonic scans*

When employing gated datasets, two distinctive characteristics were observed within the reconstruction losses. First, sudden and large spikes in the reconstruction loss, which have already been discussed as they pertain to defects. The second feature is the underlying reconstruction value of undefective B-scans, which tends to average around specific value with relatively minimal variances. This was expected since undefective B-scans do not contain features that were not observed during training; therefore, these are reconstructed well with consistency. However, upon further analysis, it was found that the scan-to-scan error level for undefective scans is varied. There are two main reasons for this, the inference on the data that falls outside the distribution of observed training data, and mathematical implications due to the loss function. The former is represented with several factors:

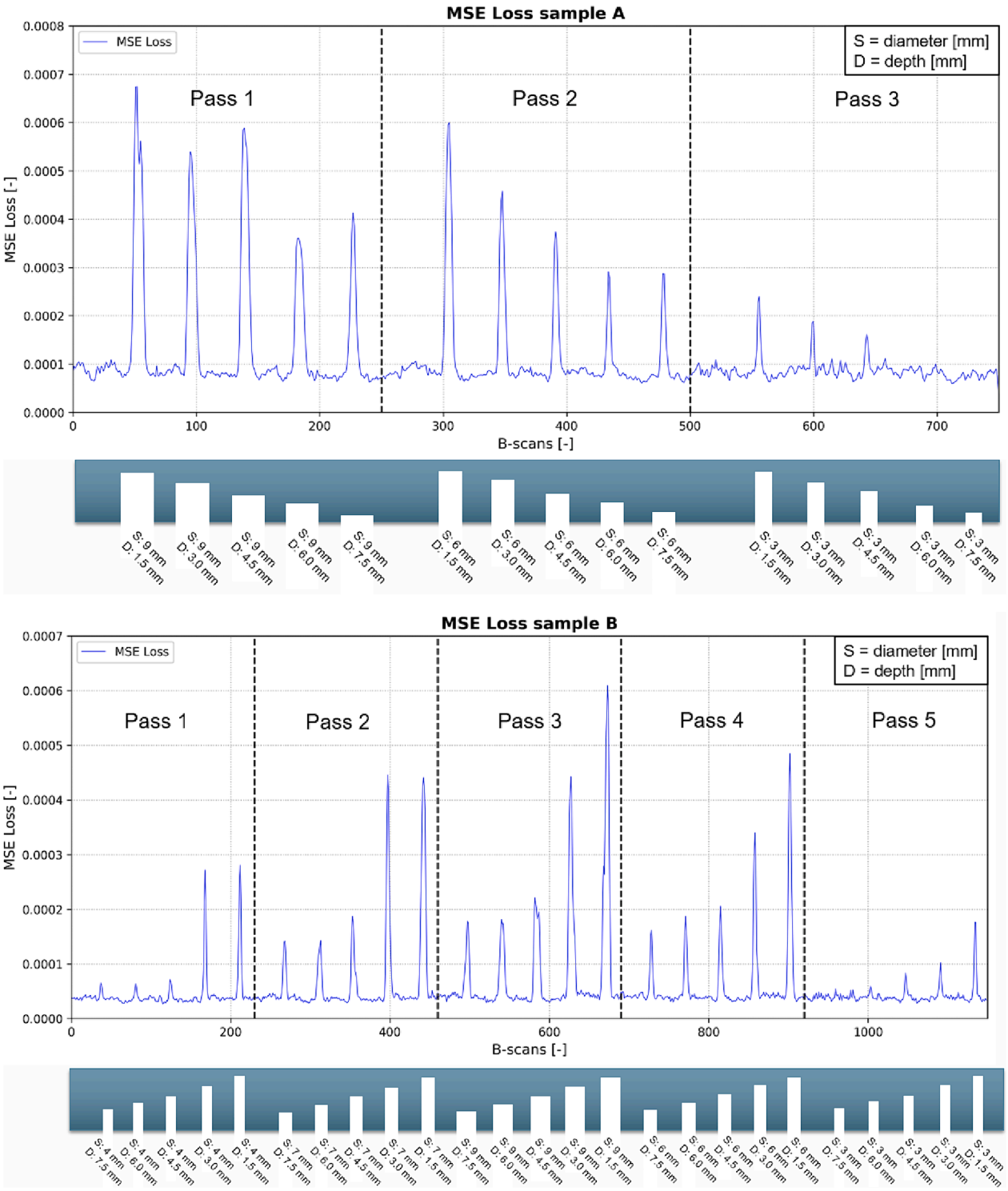


Fig. 11. Reconstruction losses and side view schematics for sample A (top) and for sample B (bottom) for gated dataset.

- **Sample finish quality:** The training dataset contains high quality scans performed on the samples with the similar surface finish. While some variance is surface finish is covered in the training data, significant deviations during testing can have a substantial impact on reconstruction errors.
- **Material anisotropy:** CFRPs have complex, anisotropic structures that can interact unpredictably with UT. Therefore, test samples may feature different macro and microscale properties not observed in the training data.
- **Variability within the equipment**: An inherent variance between the performance of individual array elements is present, as arrays are manufactured to operate within certain tolerances to pass the quality assessment by the manufacturer.
- **Variance in coupling:** Performed scans vary from each other due to changes in coupling conditions during scans. While the process is automated, the coupling dynamics during the scanning is unpredictable due to the manual application and contributes to fluctuations in reconstruction errors across different scans. This has a

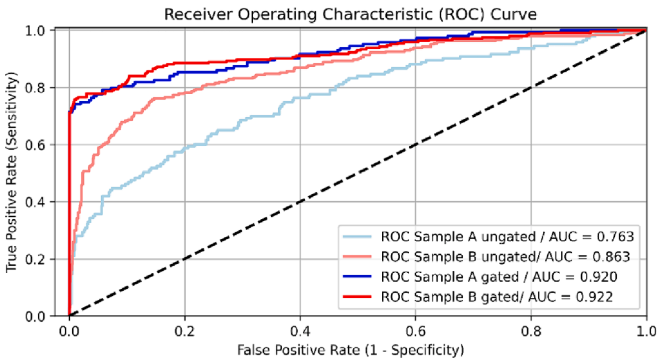


Fig. 12. Receiver Operating Characteristic curve for samples A and B.

significant impact on the energy transfer which is discussed in the later example.

Mathematical variances are exhibited due to:

- **Type of reconstruction loss**: The chosen MSE loss calculates a summation of differences between all pixels. As a result, areas with higher amplitudes exhibit larger absolute errors compared to areas with lower amplitudes. This relationship is directly linked to coupling variance, as it is a significant factor in energy transfer during the scanning process.
- **Sample thickness**: Scans of thicker samples result in larger B-scans in terms of time samples, leading to greater reconstruction error. With a larger number of data points to be reconstructed, the likelihood of errors occurring in the reconstructed image increases. Consequently, the global sum of errors in thicker samples tends to be larger.

All the aforementioned occurrences lead to challenges in the repeatability of PAUT scans which in turn greatly influences the performance of ML models.

In the experiments, coupling inconsistency exhibited the highest influence on the reconstruction errors. As the roller-probe tyre moves over the sample surface under a fixed contact force, it pushes out the water film which can cause inconsistencies in the coupling. The tyre's surface is slightly roughened by the manufacturer to retain water; however, this can sometimes lead to uneven wetting and improper coupling which would impede the path of the ultrasound beam. Such variations become apparent if an amplitude C-scan of the front wall is created, as displayed in Fig. 13, where brighter areas indicate portions of the scan where higher amplitudes were recorded, compared to dimmer/ darker areas of lower amplitudes corresponding to good and poor coupling, respectively. In the same figure, an example where the coupling between the PAUT and the sample was inconsistent is shown, where inconsistency creates a large variance in the reconstruction errors when it comes to undefective B-scans. The underlying reason for this issue was the excessive application of couplant onto the sample. In this specific scenario, reconstruction losses of certain undefective B-scans exceeds the reconstruction losses of defective responses, which negatively influences the final performance of the deployed model.

3.3. Uncertainties associated with human annotations

The process of labelling the test datasets involved three operators who were presented with visualised B-scans accompanied by corresponding scan indexes (positions where the B-scans were acquired). As each defect has breadth and depth parallel to the inspection surface, it can provide indications captured across consecutive B-scan frames. Therefore, the operators were tasked with identifying and marking the beginning and concluding indexes at which defects were believed to appear within the dataset. However, this task proved to be quite challenging, as the achievement of consensus among the operators was infrequent. Notably, disparities in the identified starting and ending indexes exhibited variances of up to 3 indexes. This discrepancy carries significant implications, as frames were captured at a robotic displacement of 0.8 mm.

The underlying cause for this challenge stemmed from the operators' approach, as they relied on observing sequences of B-scans to pinpoint defects. By navigating back and forth within these sequences, they searched to identify the precise starting and ending points of defects. This deviation from the way the AE model processes B-scans introduces a discrepancy, as the model doesn't operate sequentially or retrospectively review datasets to arrive at conclusions. An example of B-scan sequence where the individual defect is observable over several instances and a corresponding C-scan image with ground truths generated by operators is shown in Fig. 14.

Given the lack of consensus for most defects, a pragmatic approach was adopted: labels from all observers were averaged, resulting in a composite ground truth. Outlined uncertainties associated with the labelling process, result in several observations. Firstly, it validates the significance of developing a robust automated approach for defect detection in the context of NDE, as ground truth was challenging to produce since the process heavily relied on the manual human interpretation of data. Secondly, it directly influences the reported performance of deployed model (while data and model remain unchanged) as ROC AUC varies up to 10.2 % between different operators. To shed light on how the network's reported performance would be impacted by the labelling discrepancies of different operators, the ROC curves

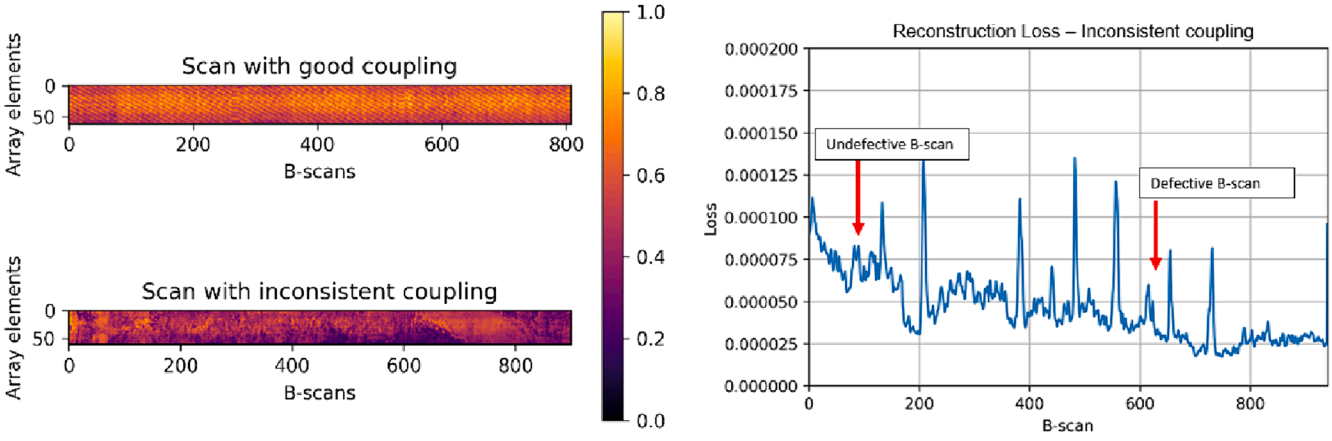


Fig. 13. Amplitude C-scan comparison of good and poor scans (left) and resulting reconstruction loss from poor scan (right).

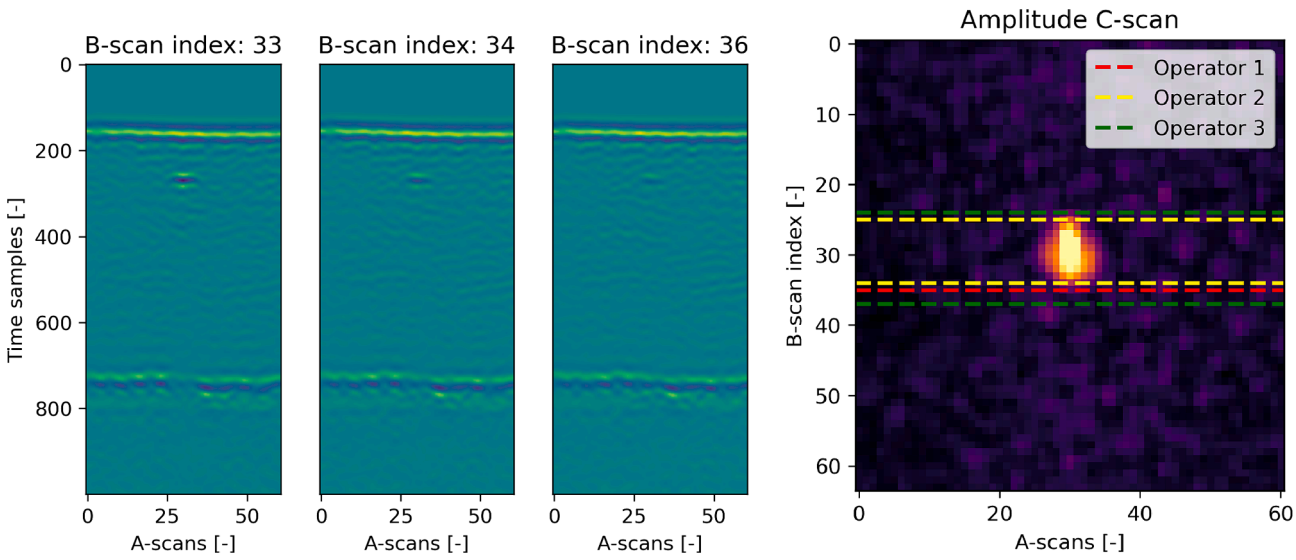


Fig. 14. A sequence of B-scans with an observable defect (left) and a C-scan of the same defect with marked labels from each operator (right).

constructed by using different ground truths are presented in Fig. 15.

3.4. Deployment on complex geometry sample

To evaluate the automated gating method and the developed AE model, an additional testing dataset was created from sample C. The range of defects positioned close to the front wall, in the middle of the sample, and near the back wall serves as a challenging example to scrutinize the performance of the model. Furthermore, as the sample is stepped, it demonstrates the ability of the automated gating process to cope with sudden changes in the geometry. Sample's geometry and the output from the gating process is presented in Fig. 16.

After the data acquisition process, it was observed that the scan performed on the thickest part of the sample exhibited a visible repetition of the tyre reflection. This occurred due to the disparities in velocities of CFRPs (~3000 m/s) and glycol/tyre (~1638 m/s), causing the reception of the second tyre reflection before the first backwall echo from the thickest composite material section. This ultrasonic indication was a limiting factor for the successful extraction of unique clusters during the automated gating process. To this end, that part of the scan was excluded in this case study and altering of the ultrasonic setup to eliminate such reflections is left for future work. Potential solutions to this problem include using a larger roller probe diameter to extend the acoustic path inside the tyre or using an alternative liquid filler material in the roller probe with a lower speed of sound, which weren't explored in this manuscript.

The scan was performed with three robotic passes over the sample with the results presented in Fig. 17. All defects in the first two passes were identified successfully. In the third pass, containing defects close to the back wall, two defects in the thickest section were missed. These missed defects present a current limitation of the proposed NDE pipeline and the performance on these types of defects could be improved with a better ultrasonic setup and a more in-depth analysis of CFRP attenuation properties which in turn would result in a more effective TCG procedures. Furthermore, these defects were challenging to observe both due to their position and relatively low acoustic response. An example of such missed defect is also included in Fig. 17. Achieved ROC AUC on gated dataset was 0.879, an improvement from 0.815 when ungated dataset was employed. Overall, this case study presents a practical application with a realistic scan conducted on a complex geometry sample. The model's deployment results in fast inference, processing 2070 samples in 1.26 ± 0.09 s on a GPU accelerated machine.

The rapid inference achieved can be attributed to the model's lightweight architecture. To further enhance inference times, exploring serialisation and saving models in environments better suited for production deployment, such as Open Neural Network Exchange (ONNX) or similar formats, rather than running them directly from Python scripts, could contribute to additional improvements in efficiency; however, this is left for future work.

4. Conclusion

In this paper, a two-stage defect detection method based on automated gating and unsupervised ML was developed for analysing

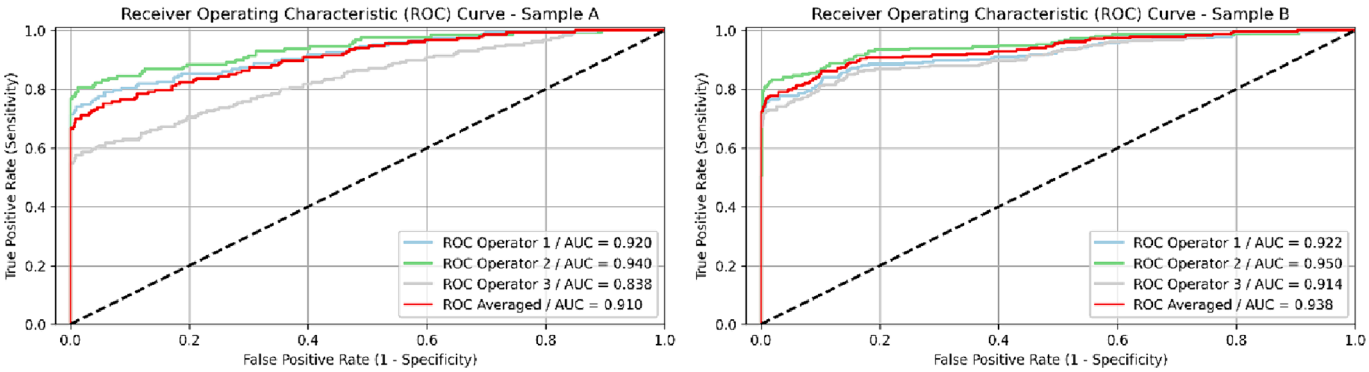


Fig. 15. ROC Comparison with respect to ground truth produced by different operators.

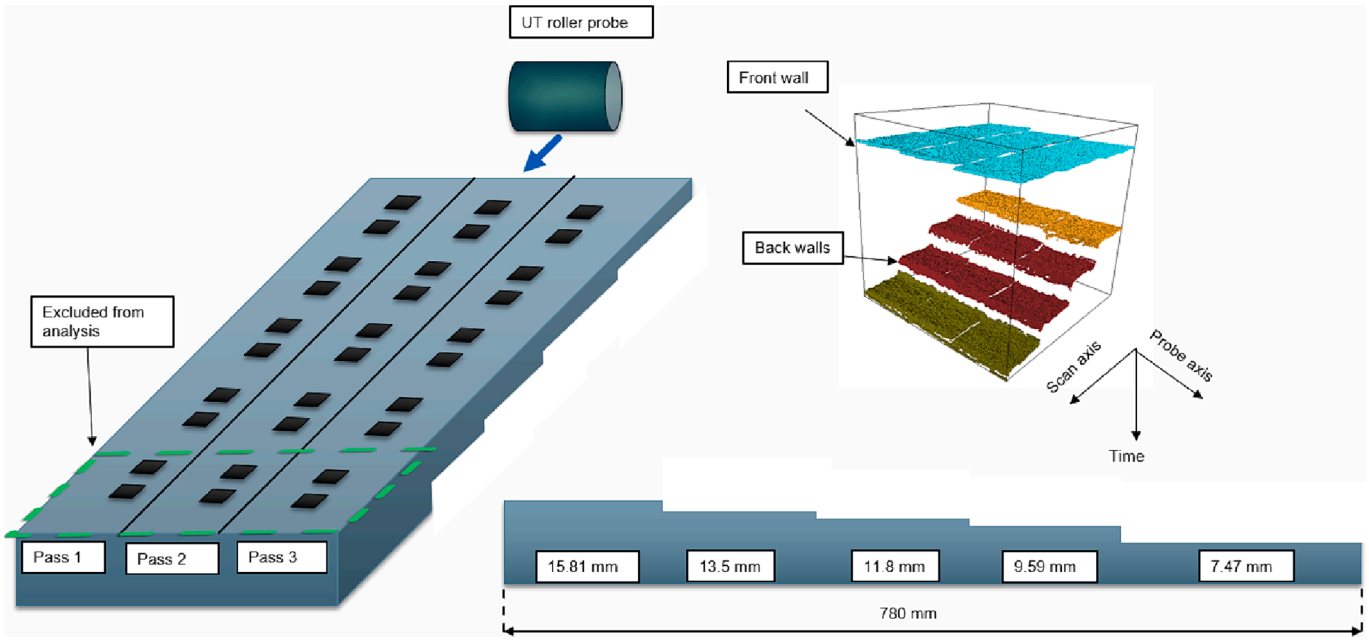


Fig. 16. Schematic of sample C (left), DBSCAN output for automated gating (top right), side view of the sample (bottom right).

ultrasonic B-scans images of CFRP components. Unlike the past efforts at automated gating, the proposed method is agnostic to geometry of the scanned sample, resolving limitations and heavy reliance of currently available methods on operators' fine-tuning. For this stage, unsupervised clustering through the DBSCAN algorithm was employed to isolate the front wall and backwall echoes of the scanned component to prepare the data in a way to maximise the ML performance in the next stage.

The subsequent stage featured an AE-based ML model, tasked with processing ultrasonic datasets captured using an automated NDE setup that emulates an industrial environment. The approach proposed here stands out among other ML methods, in that it is fully unsupervised removing the need for labelling of the B-scans. This saves expert NDE operator time in preparing training datasets while still achieving good detection performance. The study yielded several key findings:

- The DBSCAN-based automated gating has proved to be a practical technique, effectively extracting front and back wall indications from 3D datasets without geometric constraints, making it a strong candidate towards true automation of the interpretation process.
- The AE performance on the ungated data resulted in unsatisfactory results as the separation between undefective and defective B-scans in terms of reconstruction error was not achieved.
- Implementing the automated gating process significantly increased the performance of the AE-based defect detector, with ROC AUC improvements of 15.7 %, 5.9 %, and 6.4 % for different testing datasets.
- Overall, 36 out of 40 defects produced visible reconstruction error spikes in the simple geometry samples, and 22 out of 24 defects in the complex geometry sample. Inference on a GPU accelerated machine was rapid, processing 2070B-scans in 1.26 $\pm$ 0.09 s.
- The overall performance of AE models was significantly influenced by the consistency of conducted scans. This was largely controlled by stability of energy transmission into the sample, a factor greatly influenced by coupling quality.
- Uncertainties stemming from variations in producing ground truth had a direct impact on the reported results, highlighting the potential advantages associated with robust automated systems within the NDE pipeline.

Limitations of this work include missed detections of smallest defects closest to the back wall of the samples. Due to the nature of analysing B-scans, in industrial situations this could extend to missed detection of thin defects oriented parallel to ultrasound beam propagation. Furthermore, in the present study, controlled scans were performed where the entirety of the defects were captured within the active aperture of the ultrasonic setup. More challenging scans where defects are not captured in their entirety is left for future work. Lastly, while the automated gating method improved the overall results of the AE model, by removing the back wall of the scan a valuable information that pertains to the loss of back wall is lost, which is often used in the analysis performed by an expert NDE operator.

As a trajectory for future work, detection of smaller 3.0 mm defects is planned, alongside testing on UT signals captured from naturally occurring complex defects in CFRPs. For improving the autoencoder model, several architectural changes could be implemented and tested. For example, exploring the addition of another encoder, drawing inspiration from approaches like GANomaly. Alternatively, incorporating and computing feature reconstruction errors, as demonstrated in previous works [54], could be another avenue for enhancement. Additionally, the assessment of ultrasonic scan quality remains an unexplored domain within the field of ML. Consequently, an attempt to formulate or correlate a scan quality evaluation metric is deemed promising.

CRediT authorship contribution statement

Vedran Tunukovic: Conceptualization, Data curation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Shaun McKnight:** Methodology, Writing – review & editing. **Richard Pyle:** Conceptualization, Methodology, Writing – review & editing. **Zhiming Wang:** Methodology. **Ehsan Mohseni:** Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review & editing. **S. Gareth Pierce:** Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing. **Randika K. W. Vithanage:** Resources, Supervision, Writing – review & editing. **Gordon Dobie:** Supervision, Writing – review & editing. **Charles N. MacLeod:** Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review & editing. **Sandy Cochran:** Funding acquisition, Project administration, Supervision. **Tom O'Hare:**

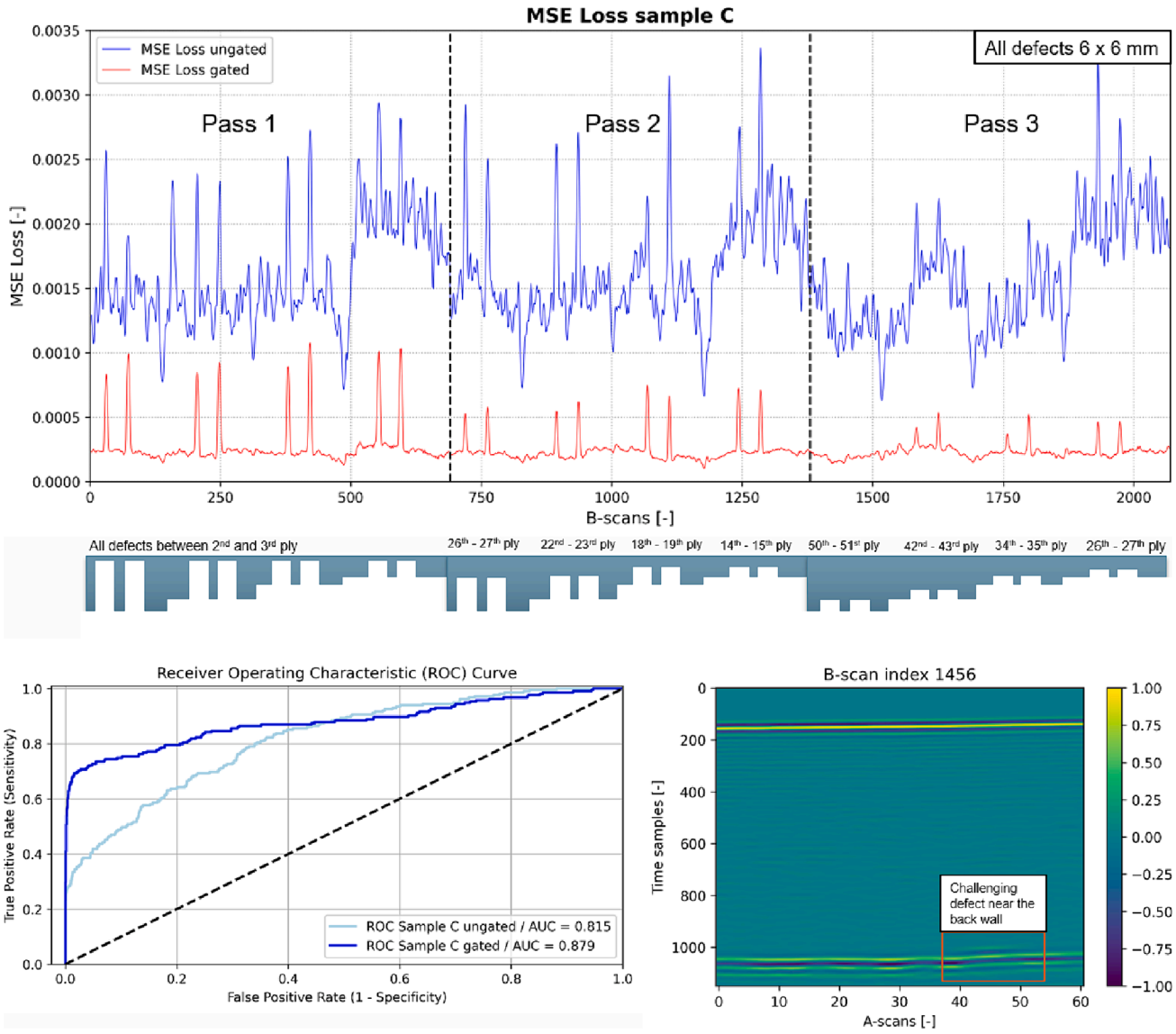


Fig. 17. Reconstruction losses for gated and ungated sample C (top), ROC comparison (bottom left), and an example of a challenging defect (bottom right).

Investigation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [Gareth Pierce reports financial support was provided by Spirit AeroSystems Inc. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.].

Data availability

The data that has been used is confidential.

Acknowledgement

This work was supported through EPSRC Centre for Doctoral Training in Future Ultrasonic Engineering (FUSE CDT) EP/S023879/1, and Spirit AeroSystems/Royal Academy of Engineering Research Chair for In-Process Non-Destructive Testing of Composites, RCSRF 1920/10/32.

References

[1] P.D. Mangalgiri, Composite materials for aerospace applications, Bull Mater Sci 22 (3) (1999) 657–664.
[2] J. Chen, J. Wang, A. Ni, Recycling and reuse of composite materials for wind turbine blades: an overview, J. Reinf. Plast. Compos. 38 (12) (Jun. 2019) 567–577, https://doi.org/10.1177/0731684419833470.
[3] 'Offshore wind net zero investment roadmap', GOV.UK. Accessed: Sep. 21, 2023. [Online]. Available: https://www.gov.uk/government/publications/offshore-wind-net-zero-investment-roadmap.
[4] N.P. Avdelidis and T.-H. Gan, '24 - Non-destructive evaluation (NDE) of Composites: infrared (IR) thermography of wind turbine blades', in *Non-Destructive Evaluation (NDE) of Polymer Matrix Composites*, V.M. Karbhari, Ed., in Woodhead Publishing Series in Composites Science and Engineering. , Woodhead Publishing, 2013, pp. 634–650e. doi: 10.1533/9780857093554.4.634.
[5] R. Slayton, G. Spinardi, Radical innovation in scaling up: Boeing's dreamliner and the challenge of socio-technical transitions, Technovation 47 (Jan. 2016) 47–58, https://doi.org/10.1016/j.technovation.2015.08.004.
[6] V. Giurgiutiu, *Structural health monitoring of aerospace composites*. in structural health monitoring of aerospace composites, Elsevier (2016) 23, https://doi.org/10.1016/B978-0-12-409605-9.00001-5.
[7] J. Bachmann, C. Hidalgo, S. Bricout, Environmental analysis of innovative sustainable composites with potential use in aviation sector—A life cycle assessment review, Sci. China Technol. Sci. 60 (9) (Sep. 2017) 1301–1317, https://doi.org/10.1007/S11431-016-9094-Y.
[8] A. Wilson, *Advances in Technical Nonwovens*. in Advances in Technical Nonwovens. Elsevier Inc., 2016, p. 271. doi: 10.1016/B978-0-08-100575-0.00009-7.

[9] U. Schnars, R. Henrich, Applications of NDT methods on composite structures in aerospace industry. Presented at the Conference on Damage in Composite Materials, 2006.

[10] A. Kapadia, 'National Composites Network Best Practice Guide non Destructive Testing of, Compos. Mater. (2007). http://www.twi.co.uk/j32k/index.xtp.

[11] S.C. Wooh, Y. Shi, Optimum beam steering of linear phased arrays, Wave Motion 29 (3) (1999) 245–265, https://doi.org/10.1016/S0165-2125(98)00039-0.

[12] C. Holmes, B.W. Drinkwater, P.D. Wilcox, Post-processing of the full matrix of ultrasonic transmit–receive array data for non-destructive evaluation, NDT E Int. 38 (8) (Dec. 2005) 701–711, https://doi.org/10.1016/J.NDTEINT.2005.04.002.

[13] P.D. Wilcox, Ultrasonic arrays in NDE: beyond the B-scan, AIP Conference Proceedings 1511 (1) (Jan. 2013) 33, https://doi.org/10.1063/1.4789029.

[14] E. Duernberger, C. MacLeod, D. Lines, C. Loukas, M. Vasilev, Adaptive optimisation of multi-aperture ultrasonic phased array imaging for increased inspection speeds of wind turbine blade composite panels, NDT E Int. 132 (Dec. 2022) 102725, https://doi.org/10.1016/j.ndteint.2022.102725.

[15] C. Mineo, 'Automated NDT inspection for large and complex geometries of composite materials', 2015, doi: 10.48730/GXQ8-WA04.

[16] C. Mineo, et al., Flexible integration of robotics, ultrasonics and metrology for the inspection of aerospace components, AIP Conference Proceedings 1806 (1) (Feb. 2017) 020026, https://doi.org/10.1063/1.4974567.

[17] M. Bertovic, I. Virkkunen, NDE 4.0: new Paradigm for the NDE inspection personnel, Handb. Nondestruct. Eval. 40 (2021) 1–31, https://doi.org/10.1007/978-3-030-48200-8_9-1.

[18] I. Virkkunen, T. Koskinen, O. Jessen-Juhler, J. Rinta-aho, Augmented ultrasonic data for machine Learning, J. Nondestruct. Eval. 40 (1) (Mar. 2021) 1–11, https://doi.org/10.1007/S10921-020-00739-5/TABLES/1.

[19] S. McKnight, et al., A comparison of methods for generating synthetic training data for domain adaption of deep learning models in ultrasonic non-destructive evaluation, NDT E Int. 141 (Jan. 2024) 102978, https://doi.org/10.1016/j.ndteint.2023.102978.

[20] M. Meng, Y.J. Chua, E. Wouterson, C.P.K. Ong, Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks, Neurocomputing 257 (Sep. 2017) 128–135, https://doi.org/10.1016/J.NEUCOM.2016.11.066.

[21] Y. Guo, et al., Fully convolutional neural network with GRU for 3D braided composite material flaw detection, IEEE Access 7 (2019) 151180–151188, https://doi.org/10.1109/ACCESS.2019.2946447.

[22] C. Tao, C. Zhang, H. Ji, J. Qiu, Fatigue damage characterization for composite laminates using deep learning and laser ultrasonic, Compos. Part B Eng. 216 (Jul. 2021), https://doi.org/10.1016/J.COMPOSITESB.2021.108816.

[23] J.C. Aldrin, D.S. Forsyth, Demonstration of using signal feature extraction and deep learning neural networks with ultrasonic data for detecting challenging discontinuities in composite panels, AIP Conference Proceedings 2102 (May 2019) 230004, https://doi.org/10.1063/1.5099716/FORMAT/PDF.

[24] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, 'GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training'. arXiv, Nov. 13, 2018. doi: 10.48550/arXiv.1805.06725.

[25] N. Munir, J. Park, H.J. Kim, S.J. Song, S.S. Kang, Performance enhancement of convolutional neural network for ultrasonic flaw classification by adopting autoencoder, NDT E Int. 111 (Apr. 2020) 102218, https://doi.org/10.1016/J.NDTEINT.2020.102218.

[26] F. Milković, B. Filipović, M. Subašić, T. Petković, S. Lončarić, and M. Budimir, 'Ultrasound Anomaly Detection Based on Variational Autoencoders', in *2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*, Sep. 2021, pp. 225–229. doi: 10.1109/ISPA52656.2021.9552041.

[27] Y. Wu and X. Zhu, 'Rail Defect Detection Using Ultrasonic A-Scan Data and Deep Autoencoder', *https://doi.org/10.1177/03611981221150923*, pp. 036119812211509–036119812211509, Jan. 2023, doi: 10.1177/03611981221150923.

[28] I. Kraljevski, F. Duckhorn, M. Barth, C. Tschoepe, F. Schubert, and M. Wolff, 'Autoencoder-based Ultrasonic NDT of Adhesive Bonds', Proc. IEEE Sens., vol. 2021-October, 2021, doi: 10.1109/SENSORS47087.2021.9639864.

[29] J.M. Ha, H.M. Seung, W. Choi, Autoencoder-based detection of near-surface defects in ultrasonic testing, Ultrasonics 119 (Feb. 2022) 106637, https://doi.org/10.1016/J.ULTRAS.2021.106637.

[30] D. P. Kingma and M. Welling, 'Auto-Encoding Variational Bayes'. arXiv, Dec. 10, 2022. doi: 10.48550/arXiv.1312.6114.

[31] O. Ronneberger, P. Fischer, and T. Brox, 'U-Net: Convolutional Networks for Biomedical Image Segmentation'. arXiv, May 18, 2015. doi: 10.48550/arXiv.1505.04597.

[32] Y. Wu, X. Zhu, and J. Baillargeon, 'Deep Autoencoder for Ultrasound-Based Rail Flaw Detection', *Proc. 2022 Jt. Rail Conf. JRC 2022*, Jun. 2022, doi: 10.1115/JRC2022-79554.

[33] R. Drai, F. Sellidj, M. Khelil, A. Benchaala, Elaboration of some signal processing algorithms in ultrasonic techniques: application to materials NDT, Ultrasonics 38 (2000) 503–507.

[34] S. Barut and N. Dominguez, 'NDT Diagnosis Automation: a Key to Efficient Production in the Aeronautic Industry', *E-J. Nondestruct. Test.*, vol. 21, no. 07, Jul. 2016, Accessed: Aug. 05, 2023. [Online]. Available: https://www.ndt.net/search/docs.php3?id=19184&msgID=0&rootID=0.

[35] J.C. Aldrin, C. Coughlin, D.S. Forsyth, J.T. Welter, Progress on the development of automated data analysis algorithms and software for ultrasonic inspection of composites, AIP Conference Proceedings 1581 (1) (Feb. 2014) 1920–1927, https://doi.org/10.1063/1.4865058.

[36] L. Séguin-Charbonneau, J. Walter, L.-D. Théroux, L. Scheed, A. Beausoleil, B. Masson, Automated defect detection for ultrasonic inspection of CFRP aircraft components, NDT E Int. 122 (Sep. 2021) 102478, https://doi.org/10.1016/j.ndteint.2021.102478.

[37] K. Lee, V. Estivill-Castro, Feature extraction and gating techniques for ultrasonic shaft signal classification, Appl. Soft Comput. 7 (1) (Jan. 2007) 156–165, https://doi.org/10.1016/j.asoc.2005.05.003.

[38] D. Guo, G. Jiang, X. Lin, and Y. Wu, 'Automated ultrasonic testing for 3D laser-rapid prototyping blisk blades', in *2016 7th International Conference on Mechanical and Aerospace Engineering (ICMAE)*, Jul. 2016, pp. 214–218. doi: 10.1109/ICMAE.2016.7549537.

[39] P. Blain *et al.*, 'Artificial defects in CFRP composite structure for thermography and shearography nondestructive inspection', *https://doi.org/10.1117/12.2271701*, vol. 10449, no. 13, pp. 562–571, Jun. 2017, doi: 10.1117/12.2271701.

[40] KUKA Robotics, 'KUKA KR90 R3100 extra HA specification manual', 2023, [Online]. Available: https://www.kuka.com/-/media/kuka-downloads/imported/8350ff3ca11642998dbdc81dcc2ed44c/0000208694_en.pdf.

[41] Schunk, 'SCHUNK Force Torque sensors manual', 2023, [Online]. Available: https://schunk.com/us/en/automation-technology/force/torque-sensors/ft/ftn-gamma-si-130-10/p/EPIM_ID-30865.

[42] Olympus-ims, 'RollerFORM: Phased Array Wheel Probe manual', 2023, [Online]. Available: https://www.olympus-ims.com/en/rollerform/.

[43] 'MicoPulse 6PA | Phased Array Ultrasonic Technology | Peak NDT', [Online]. Available: https://www.peakndt.com/products/micropulse-6pa/.

[44] S. Ioffe and C. Szegedy, 'Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift', *32nd Int. Conf. Mach. Learn. ICML 2015*, vol. 1, pp. 448–456, Feb. 2015, doi: 10.48550/arxiv.1502.03167.

[45] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, 'Improved Techniques for Training GANs'. arXiv, Jun. 10, 2016. doi: 10.48550/arXiv.1606.03498.

[46] A. Radford, L. Metz, and S. Chintala, 'Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks', *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.*, Nov. 2015, doi: 10.48550/arxiv.1511.06434.

[47] A. Odena, V. Dumoulin, C. Olah, Deconvolution and Checkerboard Artifacts, Distill 1 (10) (Oct. 2016) e3.

[48] D. P. Kingma and J. Ba, 'Adam: A Method for Stochastic Optimization', 2014, doi: 10.48550/ARXIV.1412.6980.

[49] Accessed: Oct. 02 (2023) [Online]. Available: https://developers.google.com/machine-learning/glossary.

[50] C. Breen, F. Guild, M. Pavier, Impact of thick CFRP laminates: the effect of impact velocity, Compos. Part Appl. Sci. Manuf. 36 (2) (Feb. 2005) 205–211, https://doi.org/10.1016/j.compositesa.2004.06.005.

[51] S. Nilsson, A. Bredberg, L.E. Asp, Effects of CFRP laminate thickness on bending after impact strength, Plast. Rubber Compos. 38 (2–4) (May 2009) 61–66, https://doi.org/10.1179/174328909X387801.

[52] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, 'A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise'.

[53] P. Virtanen, et al., SciPy 1.0: fundamental algorithms for scientific computing in python, Nat. Methods 17 (3) (Mar. 2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2.

[54] I. Ndiour, N. Ahuja, U. Genc, and O. Tickoo, 'FRE: A Fast Method For Anomaly Detection And Segmentation'. arXiv, Nov. 22, 2022. doi: 10.48550/arXiv.2211.12650.