Artificial Intelligence-Based SHM Approach for Fatigue Damage Assessment

ANASTASIIA VOLOVIKOVA, DANIELLE STEPHENS,
PAUL SWINDELL and STEFFEN FREITAG

ABSTRACT

Periodic inspections are common to find flaws in aircraft structures. Continuous monitoring of the structural integrity using Structural Health Monitoring (SHM) is desirable, as undetectable damage can occur at any time during a structure's lifespan. Numerous SHM technologies have been developed to identify and to evaluate defects in structures. Using data-driven analysis methods, such as artificial intelligence (AI), information about the condition of structures can be extracted from large amounts of complex measured data. However, current AI-based damage assessment methods are often limited to simple, well-defined structures. This paper presents an AI-based approach for damage assessment on a full-scale fuselage panel with a complex substructure. An aluminum panel tested by the Federal Aviation Administration (FAA) at the William J Hughes Technical Center is used to explore new applications of artificial neural networks (ANN) by different input scenarios and classification concepts, for damage detection and damage level prediction. Initial damage is created by a sawcut. The panel is fatigued through cyclic loading, causing the crack to propagate, providing comprehensive sensor data responses due to the crack growth. The experimental setup involves Acellent Technologies piezoelectric (PZT) sensors symmetrically placed on opposite sides of the sawcut, capturing pitch-catch signals. The ANN model utilizes damage indices derived from the sensor signals to detect and assess the damage. This research highlights the potential of AI-based approaches in enhancing the capabilities of SHM systems, paving the way for more reliable and efficient structural monitoring.

INTRODUCTION

Structural Health Monitoring (SHM) extends non-destructive testing by enabling in-service damage detection, assessment, and prognosis, with guided waves offering high sensitivity to early-stage damage. However, practical deployment is hindered

Anastasiia Volovikova, Karlsruhe Institute of Technology, Karlsruhe, Germany
Danielle Stephens, FAA, William J Hughes Technical Center, Atlantic City, United States
Paul Swindell, Diakon Solutions LLC, 110 W. Beaver Dr, Cape May Court House, United States
Steffen Freitag, Karlsruhe Institute of Technology, Karlsruhe, Germany

by complex wave phenomena such as multimodal propagation, reflections, and mode conversions at structural discontinuities [1]. To address this, data-driven approaches using artificial intelligence (AI)-based assessment methods are explored for solving the inverse problem of damage identification from measured wave responses. Current damage assessment methods, however, are often restricted to simple, well-defined structures. This research extends AI-based approaches for damage assessment in stiffened structures — developed in preliminary work [2–5] — particularly using artificial neural networks (ANNs), to a new experimental setup and dataset to ensure the robustness and applicability of the approaches across different use cases.

The use case for this paper focuses on a full-scale fuselage panel with a complex substructure. In partnership with Arconic and Embraer, the Federal Aviation Administration (FAA) is assessing emerging metallic structures technologies (EMST) using the FAA's Full-Scale Aircraft Structural Test Evaluation and Research (FASTER) facility. Full-scale fuselage panel test data is obtained to assess the effect of EMST concepts on damage tolerance performance of a baseline aluminum fuselage structure, which is designed as located on the crown of a typical single-aisle aircraft, forward of the wing [6]. This first baseline panel (Figure 1) test was done in a phased approach to study three scenarios simulating damage that represents cracks in the aircraft structure. The focus of this paper is on SHM data gathered in Phase III, where a

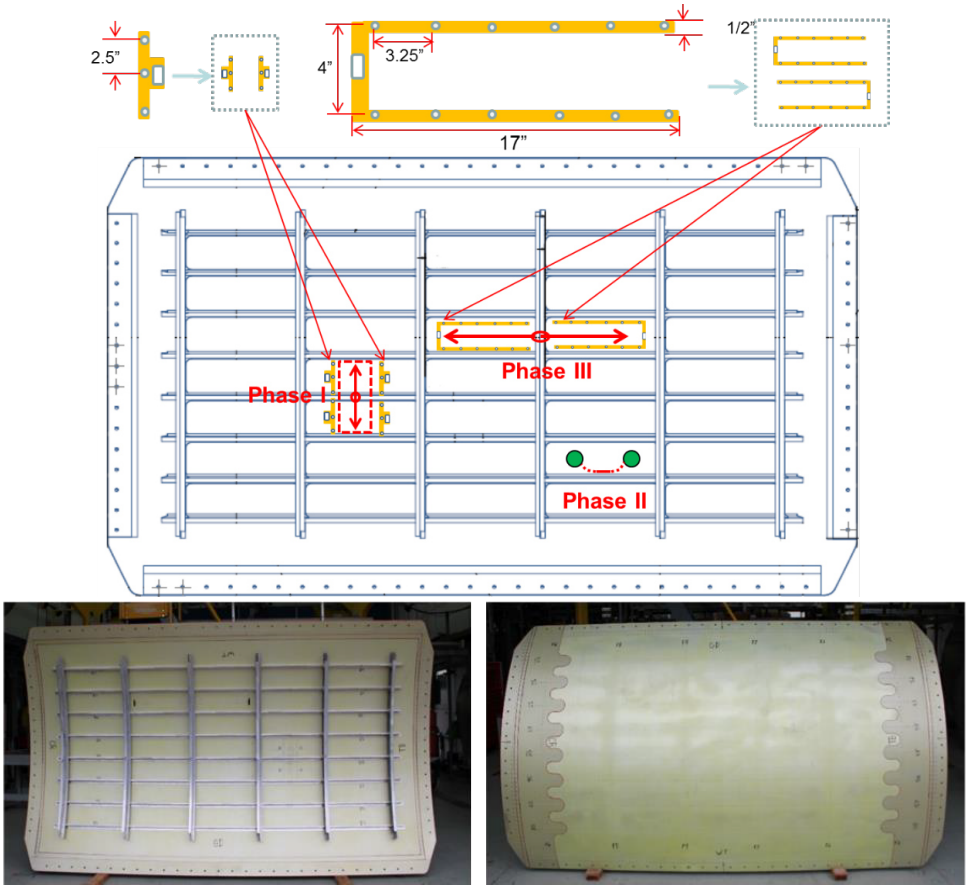

Figure 1. (top) Panel configuration with actuator and receiver layouts for Phases I - III and (bottom) inner and outer surfaces of the panel.

two-bay crack-like notch in the longitudinal direction is considered.

EXPERIMENTAL SETUP

In Phase III, a 1.5 inch crack-like notch was sawcut in the skin between two bays resulting in the severing of the central frame. The panel was fatigue tested to 43600 cycles. During the fatigue test, the crack extended across the two bays to a final total length of 16 inches.

For SHM data collection, the FAA used an Acellent sensor system consisting of their Scan Genie data acquisition hardware, a laptop with SHM Patch software, and custom "smart layer" sensors specifically designed for this test. The Acellent smart layer sensors were installed in two horseshoe-like configurations on the internal surface of the panel on either side of the sawcut-severed frame. The sensors were placed in the two bays around the path of the crack extension (see Figure 1) to monitor and record damage formation during Phase III testing. Each smart layer contains 12 individual sensors, 6 above and 6 below the expected path of crack propagation.

DESCRIPTION OF DATASET

The panel cycle count and crack growth truth data in inches are shown in Table I. Nine baseline SHM inspections were performed to gather initial data on the panel condition, post sawcut but pre-crack initiation. The SHM data collection was performed at 200, 250, 300 and 350 kHz frequencies approximately every 3000 cycles. During data collection, each of the Acellent sensors was cycled as the pulse generator, with the other sensors being receivers. For example, pulsing starts with sensor 1 and ends with sensor 23 with no repeated paths (i.e. 1-2, 1-3, 1-4, to 23-24). There were 11 data collection events, which are referred to as damage levels, since the crack length is increasing with each cycle loading event, as can be seen in Table I.

The Acellent algorithm performs a damage detection by subtracting the baseline data and the current state data, the resulting value is called a damage indicator or index (DI). This damage indicator from Acellent (A) is referred to as DI_A, and its calculation formula can be found in [7, 8]. DI_As are generated for each path, each frequency and each test. The DI_A data can be graphed against the flaw size, see

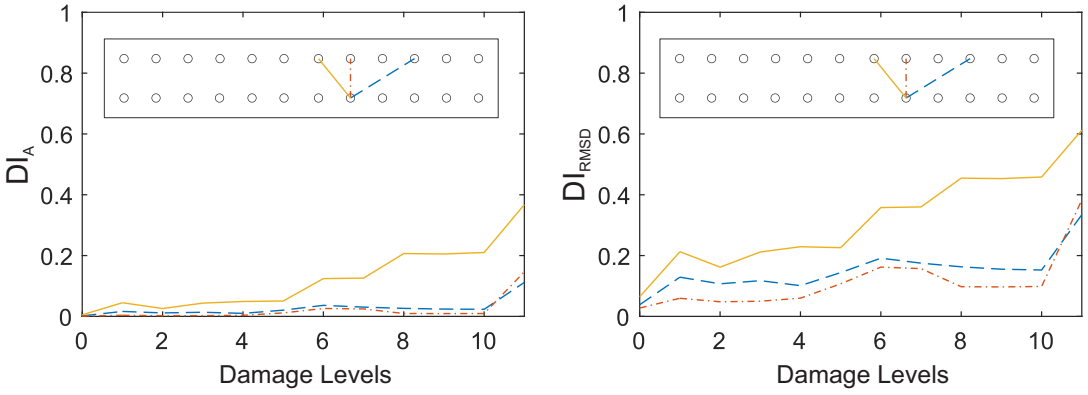

Figure 2. DI_A and DI_{RMSD} vs damage levels, presented for selected paths at 200 kHz.

Figure 2. There should be a monotonic relationship as DI_A increases with crack growth.

The performance of different types of damage indicators from previous work [2,4] was tested, and an alternative DI is chosen. This DI_{RMSD} is based on the root-mean-square deviation from the baseline signal and its calculation formula can be found in [2,4]. Similarly to DI_A, DI_{RMSD} is calculated for each signal path, each frequency and each test. DI_{RMSD} is also plotted against the flaw size, see Figure 2.

DESCRIPTION OF AI-BASED METHOD

In order to assess the condition of the structure, the damage indicators DI_{RMSD} and DI_A calculated from the received signals are given to an ANN, which evaluates the input and learns to sort it into predefined classes. First, several scenarios are elaborated for the ANN input. Then, different classification concepts are determined according to the required damage assessment, which can include damage detection or damage size evaluation.

Input Scenarios

The total number of signals that can be considered for the ANN training and testing consists of 96 signal paths. Each signal path is repeated for four frequencies and acquired in nine repetitions for undamaged panel states (baseline signals) and 11 stages of damaged panel states (damage signals), resulting in a total of 7680 signals.

In the **single-path input scenario**, a single value input is considered. The value is a damage indicator for one actuator-receiver path. In this way, different types of damage indicators are tested separately. Here, the total number of input samples (damage indicators) is the same as the number of signals, namely 7680.

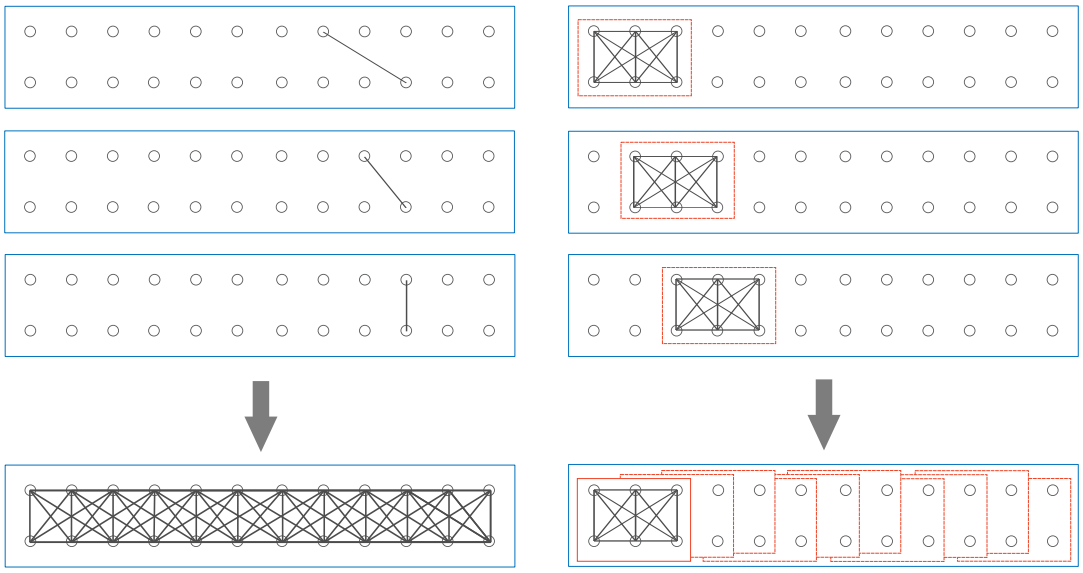

Figure 3. (left) Schematic representation of single-path input and (right) multiple-path input, i.e., 6-sensor input, resulting in 10 path-groups.

TABLE I. CLASSIFICATION CONCEPTS BASED ON THE CRACK LENGTH IN INCHES RESULTING FROM THE GIVEN NUMBER OF LOAD CYCLES

Cycle	Left Crack	Right Crack	2 Classes	3 Classes	4 Classes
0	0.750	0.750	undamaged	undamaged	undamaged
3000	0.862	0.836	damaged	small	small
4500	0.902	0.845			
6000	1.000	1.000			
9000	1.177	1.174			
12000	1.625	1.625			medium
15000	1.704	1.723			
24500	2.127	2.112		large	
32500	2.698	2.671			large
34500	2.890	2.861			
36500	3.103	3.086			
43600	7.550	7.875			

In the **multiple-paths input scenario**, a multiple-values input is considered. Instead of just one, several damage indicators from selected actuator-receiver paths are grouped into one input. The paths are grouped on the basis of their location on the plate and their position relative to each other. For example, a 6-sensor input consists of 6 PZT sensors, three on top and three at the bottom of the panel, positioned oppositely or adjacent to each other (see Figure 3). Consequently, 15 signal paths originating between these 6 sensors can be evaluated, which form one input of the ANN. In this case, 800 samples in total can be considered for network training and testing, consisting of 10 of such path-groups $\times$ 4 frequencies $\times$ (9 baseline $+$ 11 damaged states). In this work, different sensor groups were built and tested, including: a 4-sensor group with 6 signal paths and 880 samples in total, a 6-sensor group with 15 signal paths and 800 samples in total, and an 8-sensor group with 28 paths and 720 samples.

Description of Classification Concepts

The ANN classifies the input values of the damage indicators into different crack length classes. The number of defined classes is chosen according to the required damage assessment: damage detection or crack length estimation.

In the case of **binary classification**, damage detection can be performed. The classes are defined as 'undamaged' for panel states with baseline signals from 324 samples, and as 'damaged' for signals obtained when the crack appears and propagates (396 samples).

For **multiclass** damage assessment, size evaluation can be achieved through the definition of flaw extension classes. Two variants were investigated: three classes (undamaged, small, and large crack lengths) and four classes (undamaged, small, medium, and large crack lengths). The definitions of the class boundaries are listed in Table I.

ANN CLASSIFICATION MODEL

The fully connected feed-forward ANN consists of input neurons (the number is

defined according to the input scenario), two hidden layers with 25 neurons per layer, and output neurons, as many as there are classes defined. It is important to note that the chosen network topology has been selected out of different investigated ANN architectures. The hidden neurons are activated by the hyperbolic tangent sigmoid function (tansig), and the softmax activation function is used in the output neuron. The scaled conjugate gradient backpropagation algorithm is applied to train the ANN, where 70% of the samples are used for the identification of the unknown network parameters (synaptic weights and bias values), 15% of the samples are employed for validation, and 15% for testing the training performance.

To evaluate the ANN performance, the detection probability is calculated by dividing the correctly classified number of samples with the total number of samples of each class. This enables an accuracy comparison of the network results for different input scenarios as well as output classes, see results in Tables II - V. Examples of ANN outputs for one variant of a multipath input and different classification concepts are shown in the confusion charts in Figure 4. Each chart shows the total count and percentage of correct (green) and incorrect (red) predictions. The detection probability is shown in the gray boxes per class (right and bottom) and overall (bottom-right corner).

RESULTS

The results are listed in Tables II - V, where the calculated detection probability shows the accuracy of the ANN, compared for different input scenarios. All values represent mean values of 20 ANN training repetitions.

First, the performance of the ANN models is compared on the basis of the predefined input scenarios, see Table II. Here, ANN receives single-paths and various multipath inputs, and the classification with 2, 3 and 4 classes is analyzed. In addition, the difference between the damage indicators DI_{RMSD} (black) and DI_A (green) is presented. The detection probability of each variant is provided using all data and just the test data in percentage. Here, all paths and all frequencies are included in the input.

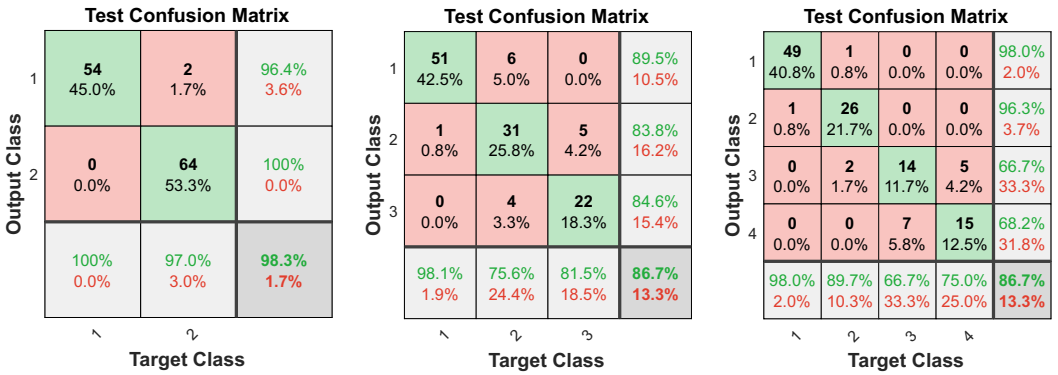

Figure 4. Confusion charts of the test data for the three classification scenarios of the 6-sensor input scenario.

TABLE II. DETECTION PROBABILITY FOR DIFFERENT INPUT SCENARIOS AND CLASSIFICATION CONCEPTS FOR ALL DATA / ONLY TEST DATA

Input	DI-Type	2 Classes	3 Classes	4 Classes
Single-path	DI_{RMSD}	85 / 85 %	68 / 67 %	61 / 60 %
	DI_A	85 / 85 %	68 / 68 %	61 / 60 %
4-sensor group	DI_{RMSD}	95 / 95 %	83 / 81 %	77 / 75 %
	DI_A	94 / 94%	80 / 79 %	75 / 74 %
6-sensor group	DI_{RMSD}	97 / 97 %	89 / 87 %	88 / 83%
	DI_A	96 / 95 %	85 / 82 %	83 / 79 %
8-sensor group	DI_{RMSD}	97 / 97%	89 / 87 %	88 / 83%
	DI_A	98 / 97 %	85 / 81 %	81 / 77 %

TABLE III. DETECTION PROBABILITY FOR MULTIPATH INPUT BY VARIOUS FREQUENCIES

Input	Frequencies in kHz	DI-Type	2 Classes
4-sensor group	all	DI_{RMSD}	95 %
		DI_A	94 %
	200; 250; 300; 350	DI_{RMSD}	97; 97; 95; 93 %
		DI_A	96; 96; 94; 91 %
6-sensor group	all	DI_{RMSD}	97 %
		DI_A	96 %
	200; 250; 300; 350	DI_{RMSD}	98; 99; 98; 97 %
		DI_A	98; 97; 97; 94 %

TABLE IV. DETECTION PROBABILITY FOR SINGLE-PATH INPUT BY VARIOUS FREQUENCIES WITH/WITHOUT PATHS CROSSING MIDDLE FRAME

Input	Paths crossing frame	Frequencies in kHz	2 Classes	3 Classes
Single DI_{RMSD}	with	all	85 %	68 %
		200; 250; 300; 350	91; 87; 82; 83 %	73; 69; 66; 67%
	without	all	85%	69 %
		200; 250; 300; 350	91; 89; 83; 82 %	75; 71; 65; 67 %

In Table III, a comparison of the input that includes all frequencies is made with the single-frequency inputs for all four frequencies separately. The detection probability (using all data) in binary classification is presented for the input of multiple sensors. Here, DI_{RMSD} and DI_A are also compared, and all paths are included.

Furthermore, the same frequency comparison is made for single-path input scenarios using all data, as shown in Table IV, including only DI_{RMSD}. Here, 2 and 3 classes are considered. In addition, the results achieved through the input of all paths and through the input without paths crossing the middle frame are presented.

Finally, the detection probability with respect to the length of the crack and the location the of path-group is presented in Table V using all data. Taking into account all paths and all frequencies, the input is formed by DI_{RMSD}, and a binary classification is carried out. Here, the path-group numbers range from 1 to 10, where the first group is located on the far left side and the tenth group on the far right side of the panel. Groups five and six include paths in the middle of the panel that cross the frame. The crack length is represented through 12 rows (crack level), where the first row corresponds to the undamaged state, and ascending numbers express the expansion of the crack.

TABLE V. COMPARISON OF DETECTION PROBABILITY WITH 6-SENSOR INPUT OF DI_{RMSD} FOR BINARY CLASSIFICATION

Crack level	Path-groups									
	1	2	3	4	5	6	7	8	9	10
1	94 %	100 %	100 %	99 %	93 %	94 %	99 %	99 %	92 %	96 %
2	89 %	41 %	95 %	86 %	100 %	100 %	100 %	96 %	76 %	40 %
3	81 %	49 %	100 %	100 %	100 %	100 %	100 %	100 %	99 %	89 %
4	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	99 %	89 %
5	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	96 %	98 %
6	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
7	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
8	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
9	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
10	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
11	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
12	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %

Discussion

These results present the performance accuracy of the ANN classification model in predicting the damage level, expressed as the percentage detection probability. The influence of different input and target output is compared. Additionally, the type of damage indicator, signal frequencies, and the locations of the paths are examined.

Tables II - III indicate that DI_{RMSD} performs slightly better than DI_A. The multipath scenario significantly outperforms the single-path scenario, with performance improving as the number of paths in the path-groups increases. When comparing different frequencies, it is evident that signals at a lower frequency (200 kHz) yield better performance compared to higher frequencies (350 kHz) or a mix of all frequencies.

The performance of the ANN model declines as the number of classes increases. However, the ANN provides reliable results for damage detection in binary classification, achieving over 90% detection probability with multipath input. Even for damage size classification across four classes, the ANN achieves over 80% detection probability with multipath input.

The confusion charts in Figure 4 indicate that false predictions are more frequent when the target class is closer to another class. For instance, samples targeted as class 3 but located near the boundary between class 3 and class 2 are more likely to be misclassified as class 2 rather than class 1.

From Table V, it is evident that the performance is worse with increasing distance to the crack, e.g. on the edge of the panel. Predictions improve when the path-groups are located closer to the flaw. However, predictions deteriorate slightly in the middle of the panel, likely due to the presence of the frame, which disturbs the signals. Paths crossing the frame in the middle of the panel have minimal impact on the overall performance of damage assessment. Additionally, undamaged states are better predicted than early damaged states, as the flaw may be too small to be detected confidently.

Overall, the ANN is capable of classifying damage levels up to 99% accuracy (in the case of binary classification with multipath input, see Table III). Furthermore, the

performance of the model can be improved by adjusting the input.

CONCLUDING REMARKS

This paper includes the investigation of a full-scale fuselage panel with a complex sub-structure, featuring a crack-like notch that extended to 16 inches during fatigue testing. Using a custom SHM system, signals were acquired and damage indices were calculated for various sensor paths and frequencies. The condition of the structure was assessed using an ANN, which received the damage indicators and classified them into predefined damage categories to enable damage detection and damage size evaluation. The study compared the performance of ANN models across different input scenarios, classification concepts, frequencies, and damage index types. Significant results showed that multipath scenarios improved detection probabilities, with reliable binary classification achieving more than 90% accuracy.

For future work, it would be beneficial to use a larger dataset that allows holding back certain damage states for a generalization testing phase of the ANN and exploring finer classification concepts. Furthermore, using raw signals instead of damage indicators can further improve the ANN performance. When working with raw signals, convolutional neural networks could offer a valuable alternative.

ACKNOWLEDGMENT

The authors thank Hannah Felber for her help with the technical implementation of the ANN training.

REFERENCES

1. Mitra, M. and S. Gopalakrishnan. 2016. “Guided wave based structural health monitoring: A review,” *Smart Materials and Structures*, 25(5):053001.
2. Mueller, I., S. Freitag, V. Memmolo, M. Moix-Bonet, K. Möllenhoff, M. Golub, R. S. Venkat, Y. Lugovtsova, A. Eremin, J. Moll, and K. Tschöke. 2023. “Performance Assessment for Artificial Intelligence-Based Data Analysis in Ultrasonic Guided Wave-Based Inspection: A Comparison to Classic Path-Based Probability of Detection,” *European Workshop on Structural Health Monitoring*:953–961.
3. Memmolo, V., J. Moll, O. Schackmann, S. Freitag, A. Volovikova, K. Tschoke, E. Savli, Y. Lugovtsova, M. Moix-Bonet, A. Bayoumi, et al. 2023. “Promoting novel strategies for the reliability assessment of guided wave based SHM systems,” *STRUCTURAL HEALTH MONITORING 2023*.
4. Volovikova, A., S. Freitag, O. Schackmann, V. Memmolo, A. Bayoumi, I. Mueller, and J. Moll. 2024. “Artificial intelligence-based approach for damage localization in ultrasonic guided wave-based structural health monitoring,” *Structural health monitoring in the light of climate impact and data science. Res Rev J Nondestruct Test*, 2(2):31–44.
5. Stähle, J., A. Volovikova, S. Freitag, and A. Stark. 2025. “Artificial Intelligence-Based Deformation Analysis for Damage Identification in Structural Health Monitoring,” in *Proceedings of the 6th Joint International Symposium on Deformation Monitoring*.

6. Tian, Y., D. Stanley, J. J. Bakuckas, M. Kulak, and E. Fulton. 2021. “Assessment of Emerging Metallic Structures Technologies through Test and Analysis of Fuselage Structure: Test Panel 1,” Tech. Rep. DOT/FAA/TC-21/25, United States. Department of Transportation. Federal Aviation Administration. William J. Hughes Technical Center.
7. Yadav, S. K., S. Mishra, F. Kopsaftopoulos, and F.-K. Chang. 2021. “Reliability of crack quantification via acousto-ultrasound active-sensing structural health monitoring using surface-mounted PZT actuators/sensors,” *Structural Health Monitoring*, 20(1):219–239.
8. Janapati, V., F. Kopsaftopoulos, F. Li, S. J. Lee, and F.-K. Chang. 2016. “Damage detection sensitivity characterization of acousto-ultrasound-based structural health monitoring techniques,” *Structural Health Monitoring*, 15(2):143–161, doi:10.1177/1475921715627490.