Multiple Tapping Method for Non-Destructive Testing of Defects in Metal Components

Dima Bykhovsky[a,*], Shalev Neuman[b], Oz Golan[b], Moshe Dror Kobo[b], Yael Ashkenaz[b], Michael Zolotih[b], Strokin Evgeny[c], Shai Essel[d] and Oshrit Hoffer[b]

[a]SCE - Shamoon College of Engineering, Be'er Sheva, Israel
[b]Afeka Academic College of Engineering, Tel Aviv, Israel
[c]Technion - Israel Institute of Technology, Haifa, Israel
[d]IAI - Israel Aerospace Industries, Israel

ARTICLE INFO

Keywords:
3D printing
machine learning
acoustic signal
tapping
sensor fusion
additive manufacturing

ABSTRACT

Non-destructive testing (NDT) of additively manufactured (AM) metal components typically relies on costly imaging or ultrasonic systems. We introduce a low-cost mechanical tapping device combined with a machine learning (ML)-based acoustic measurement workflow, and demonstrate its superiority as a measurement system compared to standard fundamental frequency analysis approaches. We treat the classification pipeline as a calibrated "soft sensor" that outputs a defect probability. While maintaining a very simple mechanical tapping system, we hypothesize that introducing Type A measurement uncertainty and fusing multiple probabilistic outputs significantly improves decision accuracy. Furthermore, we demonstrate that varying the tap location constitutes a distinct measurement modality, offering validation beyond the benefits accrued from mere repetition. Using a dataset of 80 Ti-6Al-4V specimens with defects validated by computed tomography (CT), we experimentally show that the proposed multiple mechanical tapping fusion method significantly reduces the probability of error.

Accepted manuscript of: D. Bykhovsky, S. Neuman, O. Golan, M. D. Kobo, Y. Ashkenaz, M. Zolotih, S. Evgeny, S. Essel, and O. Hoffer, "Multiple Tapping Method for Non-Destructive Testing of Defects in Metal Components," *Measurement*, vol. 269, art. no. 120751, Apr. 2026, https://doi.org/10.1016/j.measurement.2026.120751.

1. Introduction

The rise of additive manufacturing (AM) techniques, notably laser powder bed fusion (LPBF), has greatly expanded the ability to fabricate intricate metal components. However, this advancement has also introduced new challenges in identifying hidden and hard-to-detect internal defects, such as incomplete fusion zones, porosity, and residual powder Ramírez, Márquez and Papaelias (2023); Guillen et al. (2024). The identification of these defects requires inspection methods that balance sensitivity, speed, and cost Silva, Malitckii, Santos and Vilaça (2023). Traditional non-destructive testing (NDT) methods, such as ultrasonic testing and X-ray imaging, provide high sensitivity but require expensive equipment and skilled operators, and have limited portability Liu, Hu, Xiao, Yin, Zhou, Li and Shen (2024).

Acoustic emission (AE) techniques are widely employed within NDT to inspect metallic components for defects, internal stresses, and mechanical characteristics through the analysis of sound and ultrasound vibrations Pollock (1969); Devenport, Rolfe, Pereira and Griffin (2022). These acoustic methods utilize various sources of mechanical excitation and employ different types of receiving devices, ranging from single microphones to microphone arrays, to capture detailed vibrational responses. They are also popular in structural health monitoring applications Cha, Ali, Lewis and

*Corresponding author

dmitrby@ac.sce.ac.il (D. Bykhovsky); shneuman42@gmail.com (S. Neuman); golano@afeka.ac.il (O. Golan); drorkobo@gmail.com (M.D. Kobo); yaelashkenaz@gmail.com (Y. Ashkenaz); michaelz@afeka.ac.il (M. Zolotih); strokin@trdf.technion.ac.il (S. Evgeny); sessel@iai.co.il (S. Essel); oshrith@afeka.ac.il (O. Hoffer)

ORCID(s):

Büyüköztürk (2024). Recently, higher sensitivity and accuracy AE-related methods were proposed for defect detection tasks Ciaburro and Iannace (2022). An optimized AE-based approach was used to distinguish between defect-related events and environmental disturbances, enabling the identification of process-induced cracks during AM processes Ansari, Arcondoulis, Roccisano, Schulz, Schlaefer and Hall (2024). Li, Liu, Shi, Wang and Zhang (2025) proposed an acoustic characteristics analysis method for stress and defect detection in X80 pipes, and demonstrated the potential of acoustic features for structural integrity assessment. In parallel, optical techniques have been integrated with acoustic detection to enhance sensitivity in limited-access scenarios Vangi, Bruzzi, Caron and Gulino (2022). While these advanced methods, ranging from physics-guided deep learning Li, Nie, Wang and Ren (2024) to laser-based inspection, offer high precision, they often carry the costs of significant hardware complexity.

Recently, much research has been devoted to the study of a variety of techniques to characterize AM parts with high precision throughout the manufacturing lifecycle, rather than focusing on inspection of the final components. At the raw material level, the geometric properties and internal porosity of metal powders significantly influence the structural integrity of LPBF parts. A recent work by Xue, Li, Jin, Li, Zhang, Dong, Shang, Bao and Yu (2025) examined these properties by utilizing Nano-computed tomography (Nano-CT) and multi-level supervised U-shape deep learning (DL) networks to achieve high-precision 3D segmentation of internal pores in metal powders, thus overcoming traditional limitations related to particle adhesion. At the other end of the manufacturing lifecycle, AM-produced structural prototypes are increasingly increasingly employed as a platform for geometry-centric NDT research. For instance, Çeçen (2025) utilized scaled 3D-printed prototypes to systematically isolate the modal behavior of distinct geometric designs, demonstrating how resonance frequencies and damping ratios can serve as effective benchmarks for vibrational resilience. These advancements highlight the use of high-resolution imaging and modal screening to identify structural configurations with improved performance. However, while these methods offer high precision, they underscore the need for complementary, low-cost acoustic workflows that can bridge the gap between complex 3D characterization, and rapid and mobile inspection.

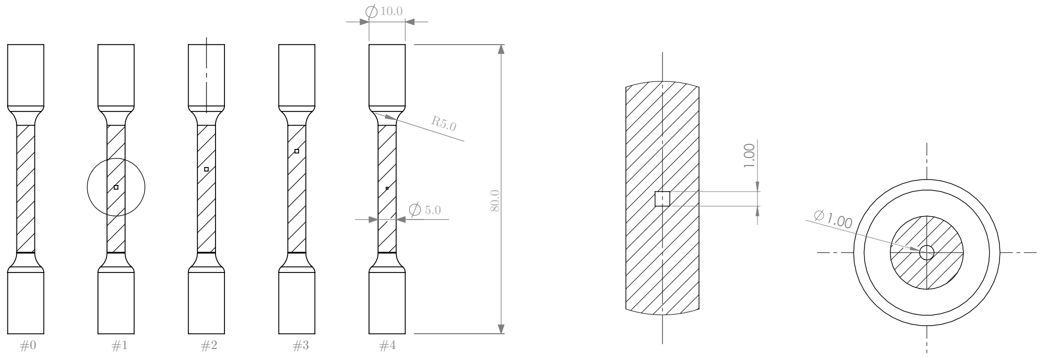
The simplest and most cost-effective AE-based method is the tap test (also termed coin tap testing or impulse excitation testing), which involves gently striking the surface of a metallic component and analyzing the resulting sound to detect internal defects or anomalies. The tap test uses acoustic signatures to identify structural issues Queiroz, Santos, da Silva and Farias (2021). The most common analysis of the resulting signatures is resonant acoustic inspection, where an impact excites the sample's natural resonant frequencies that are used to identify anomalies or distinguish between different internal conditions ASTM International (2023). This technique complements continuous AE monitoring that captures stress-induced emissions during loading Lin, Yang, Liu, Yao, Ding, Zhang, Yang, Xiao and Wang (2025).

In recent years, the scope of the tap test has expanded through the integration of machine learning (ML) and DL, transforming it from a manual qualitative check into an automated diagnostic tool Li, Jiang and Wang (2022); Wu, Li, Jiang and Wang (2025). Other recent work has treated tap-based NDT as a pattern-recognition task, utilizing complex classifier architectures to achieve high accuracy with diverse materials and structures Ge, Qin, Xie and Lu (2025); Yang, Xiong, Wang and Lu (2022). Specifically for AM metal parts, recent work recognizes that geometric complexity and process-induced variations significantly complicate the interpretation of resonance spectra Kolb, Zier, Grager et al. (2021); Boyadzhieva, Gutmann and Fischer (2024).

However, most recent studies that combine AE with ML for defect detection focus either on high-end sensor suites (e.g., multi-channel piezoelectric arrays) or on in-process monitoring in industrial AM systems. These solutions involve substantial hardware cost, complex data handling, and expertise in both ultrasonic instrumentation and algorithm design. Moreover, the majority of current AE/ML research treats the classification model as the central contribution, while the AE acquisition chain is assumed as given and rarely analyzed in terms of repeatability, measurement uncertainty, or the influence of excitation and sensor placement Guillen et al. (2024).

By contrast, the present work starts from a deliberately simple and low-cost experimental setup: an EV3-based mechanical tapping device and a single condenser microphone. With this minimal hardware, we carry out a systematic investigation of: (i) how the acoustical response of LPBF titanium "dog-bone" samples varies with tap location and thermal post-processing; (ii) how stable the measured signatures are under repeated impacts at the same nominal location; and (iii) how different feature representations and decision-level fusion strategies affect the reliability of defect detection. The experimental dataset is based on 80 Ti64 samples with intentionally introduced internal defects; the presence and characteristics of these defects were verified by industrial CT scans, providing a strong ground truth.

From a metrology perspective, we treat the tap test and subsequent processing as a measurement system. This provides measurement-driven insights into how a simple acoustic NDT setup should be optimally configured (multiple tap locations, repetitions, and outcome fusion) in order to achieve robust detection with minimal hardware. Unlike many



(a) Full set. (b) Close-up of group #4 defect.

Figure 1: Set of 80 samples, 10 of them defect-free and 70 with internal defects (Table 1).

AE/ML applications, where the algorithm is the main focus, our emphasis is on understanding and characterizing the measurement chain as a soft sensor Curreri, Patanè and Xibilia (2021); Cao, Wang, Su, Luo, Cao, Braatz and Gopaluni (2025). We identify the conditions under which a low-cost mechanical tapping system can approximate the defect classification capability of more complex NDT techniques such as CT scanning and ultrasonic testing.

Our key contributions can be summarized as follows:

- We demonstrate that a simple and automated tap test is highly effective for detecting internal defects in CT-validated LPBF Ti64 samples.
- We quantify the influence of tap location and repeated impacts on the acoustic response and classification performance, showing that careful location selection and multi-tap fusion significantly reduce misclassification rates.
- We demonstrate that basic, yet robust, MFCC-based feature extraction, combined with a logistic regression classifier, provides very good detection performance while remaining interpretable and easy to implement as a measurement model.

The rest of the paper is organized as follows. Section 2 describes the experimental design and dataset. Exploratory signal analysis of the recordings is presented in Sec 3. Classification methodology is detailed in Sec. 4. Section 5 provides a discussion of the implications, limitations, and future research directions. Section 6 concludes.

2. Experimental Design

2.1. Sample and Defect Design

Samples were produced using an EOS metal 3D printer with Ti-6Al-4V powder, commonly used in aerospace and medical applications. The samples were shaped as rounded "dog bones". A single batch of 80 samples (Fig. 1) was divided into eight series, each consisting of ten samples. In each series, five samples underwent thermal post-processing, while the other five did not. The goal of the thermal post-processing was to simulate post-print stress relief treatment and evaluate its influence.

The first 10 samples served as a defect-free control group, while the remaining 70 samples were created with deliberate internal defects by selectively disabling the laser during printing. These uncured regions retained residual powder and affected the structural and acoustic response of the samples.

All defects (Table 1), except those in group #4, share the same dimensions: diameter of 2 mm, height of 1 mm, and width of 1 mm. The defects in group #4 have a reduced diameter of 1 mm while maintaining the same height and width, facilitating valuation of the sensitivity of the model to smaller anomalies.

To validate the presence, size, and location of the internal defects in the printed titanium samples, industrial CT scans were performed using an RX Solutions EasyTom 150 machine. The samples were scanned in groups of ten, arranged in a circular fixture to maximize scanning efficiency and ensure consistent orientation. The CT scans (Fig. 2) confirmed the presence of these intentionally introduced defects, which appear as clearly distinguishable regions of lower density within the cylindrical volume of each sample.

Table 1
Defects in samples

Group #	Number of samples	Defect location
0	10	no defect
1	20	in the center
2	20	closer to the outer edge
3	20	between the center and outer edge
4	10	small defect in the center
Total	80	

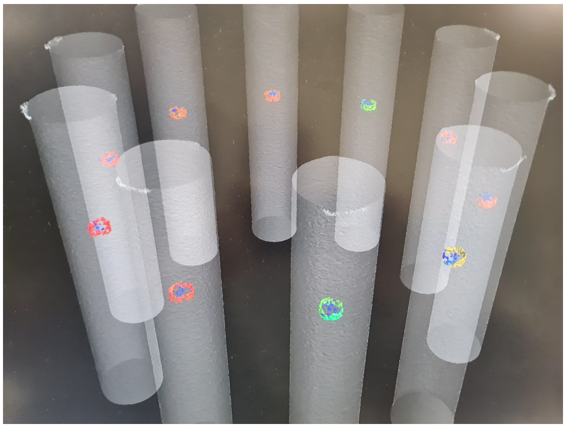


Figure 2: CT post-validation of the defects.

2.2. Tapping and Recording Setup

The custom-built tapping device was constructed from an EV3 Lego kit with servo motors (Fig. 3) to ensure consistent impact energy across all samples and locations. The striker consisted of a long metallic strip terminating in a carbon steel (AISI 1015) ball with a diameter of 6.35mm (1/4"). The sample was placed on a foam base in order to acoustically decouple it from the measurement stand.

All experiments were conducted in a small acoustic room lined with sound-absorbing panels to minimize background noise and reflection artifacts. Acoustic measurements of each sample's vibrational response were captured using a Peavey PM 18S podium condenser microphone, which was selected for its wide frequency response, directional pickup pattern, and low noise floor. The microphone was connected through a PV6BT mixer, which ensured clean signal amplification and consistent input levels across recordings. The microphone was kept at a fixed distance and orientation relative to the samples throughout all tests to maintain measurement consistency. Audio recordings were made using Audacity software, which allows manual control and standardized acquisition. All audio files were exported in high-resolution `.wav` format with a uniform sampling rate of 48 kHz.

Each sample was tapped in three predefined regions: far (F), middle (M), and near (N) from the label printed on the same side of each sample (Fig. 4). Five repetitions per region were recorded, resulting in 15 recordings per sample.

2.3. Dataset

The final dataset comprised 1,199 labeled acoustic recordings: 80 samples $\times$ 3 regions $\times$ 5 repetitions, following the rejection of one record (group #1, (F) location) that was found to have been incorrectly recorded and was therefore omitted from further processing. Each recording included a transient acoustic waveform of 5-10 seconds in length, as illustrated in Fig. 5. For subsequent analyses (with the exception of defect classification, described in Sec. 5.4), we reduced the task to a binary target with defect/non-defect labels. The dataset is inherently strongly imbalanced, as only 10 of the 80 samples are defect-free, resulting in a $1:7$ ratio between the two classes.

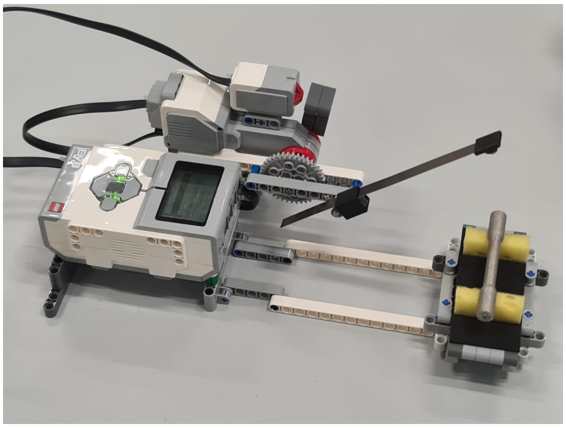


Figure 3: Tapping device based on EV3 Lego kit with servo motors with steel ball striker.

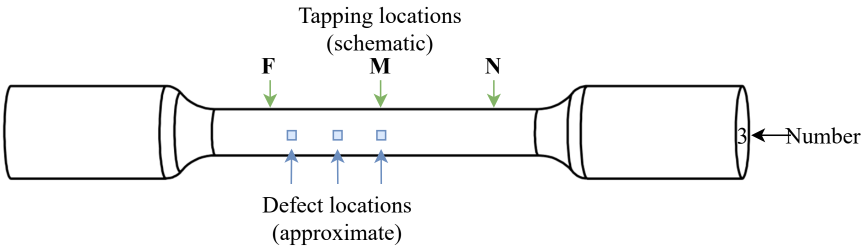


Figure 4: Tapping locations: F - far, M - middle, N - near.

3. Exploratory Signal Analysis

The goal of this section is to explore the acoustical signal properties and their physical interpretation.

3.1. Natural Frequency

The natural (or fundamental) frequency of a system is the vibration frequency at which the response to a disturbance has the highest amplitude. By defining the recorded signal $\{x_t\}_{t=0}^{L-1}$, this frequency can be found by maximization of a periodogram

$$F_0 = \underset{F}{\operatorname{argmax}} \frac{1}{L} \left| \sum_{t=0}^{L-1} x_t \exp\left(-j2\pi \frac{t}{F_s} F \right) \right|^2, \tag{1}$$

where F_s is the sampling frequency and F_0 is the resulting natural frequency. Natural frequency analysis is a straightforward method for the applied measurement setup ASTM International (2023).

The initial analysis comprised an evaluation of the fundamental frequencies for all the signals. The results (Fig. 6) show that these frequencies depend on the tapping location. At the middle tap location, the frequencies are approximately half of the frequencies at the outer ((N) and (F)) locations. The frequencies at (N) and (F) are similar, probably due to the symmetry between these locations.

The presence or absence of thermal post-processing results in higher values of fundamental frequencies (Fig. 6b). However, no clear relationship between natural frequency and defect condition was found (Fig. 6a).

Multiple taps at the same location on a single sample occasionally yielded small variations in the measured fundamental frequency. Out of 240 tap-location groups, 200 exhibited a standard deviation (std) of 0.1 Hz or less, while the remainder showed variations up to 5.3 Hz. We conjecture that tapping slightly changes the sample location. Nevertheless, of those 40 high-variability groups, 15 contained a single outlier tap, which occurred randomly as the first, middle, or last impact.

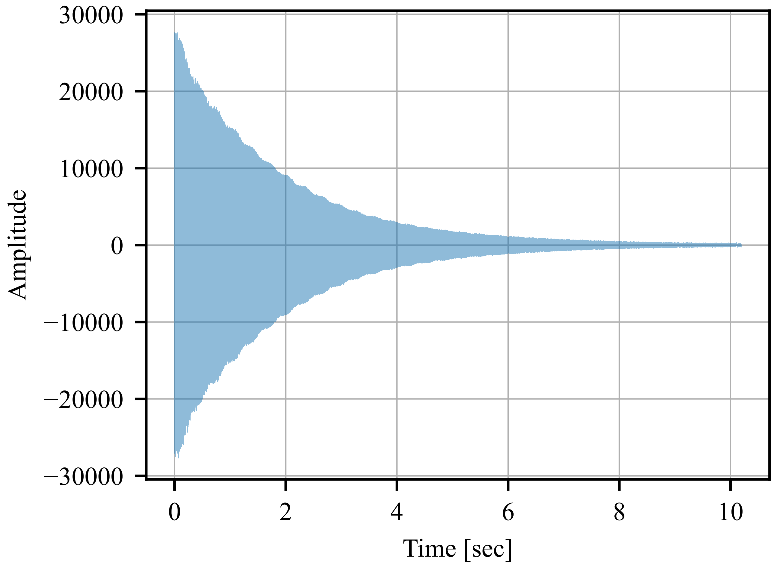


Figure 5: An example of the recorded acoustic signal.

To summarize the outcomes of the natural frequency analysis:

- Fundamental frequency on its own is an unreliable indicator of hidden voids or porosity.
- Parts that underwent heat treatment exhibit a measurable upward shift in resonance.
- Repeated taps at (nominally) the same spot still show frequency scatter, reflecting slight repositioning.
- Resonant peaks change between different tap locations, suggesting that each location may require its own ML classification model.
- A sample periodogram (see Eq. (1) and Fig. 7) exhibits multiple prominent peaks, revealing several underlying periodic components and indicating that a richer modal or time–frequency representation is needed for ML analysis.

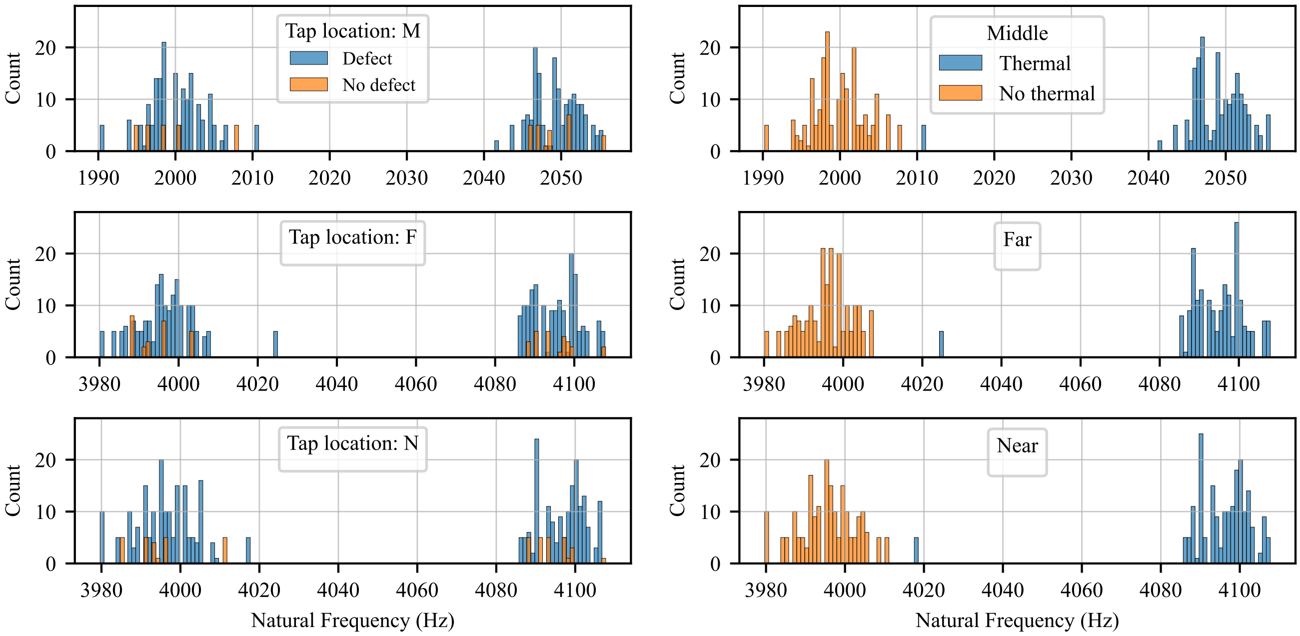
3.2. Mode Modeling

Building upon the natural frequency analysis presented in the previous section, mode modeling offers a more comprehensive modeling framework. It decomposes the recorded acoustic signals into a sum of damped sinusoidal components, with each component representing a distinct vibration mode governed by the material's properties, such as geometry, stiffness, and damping characteristics. The corresponding signal model is of the form of a sum of damped sinusoidal signals,

$$x_t = \sum_{k=1}^{K} A_k e^{-\alpha_k t} \cos\left(2\pi \frac{F_k}{F_s} t + \varphi_k\right) + \eta_t, \tag{2}$$

where K is the number of signal components, α_i is the damping factor, F_k is a resonance frequency, φ_k is the phase and η_t is noise. The value F_k that corresponds to the highest A_k/α_k ratio is the natural frequency from Eq. (1).

The modal parameters are expected to provide physically meaningful insights into the mechanical response, where each parameter corresponds to a distinct structural property. The damping factor (α_k) reflects internal energy dissipation, the modal frequency (f_k) is influenced by the stiffness and mass distribution, and amplitude (A_k) captures the strength of the vibration response. Collectively, these parameters are expected to form a signature of the structural state of each sample.



(a) Defect condition

(b) Thermal post-processing

Figure 6: The natural frequency histogram for each tapping location shows that while the defect condition in (a) does not significantly influence the frequency, thermal post-processing in (b) does.

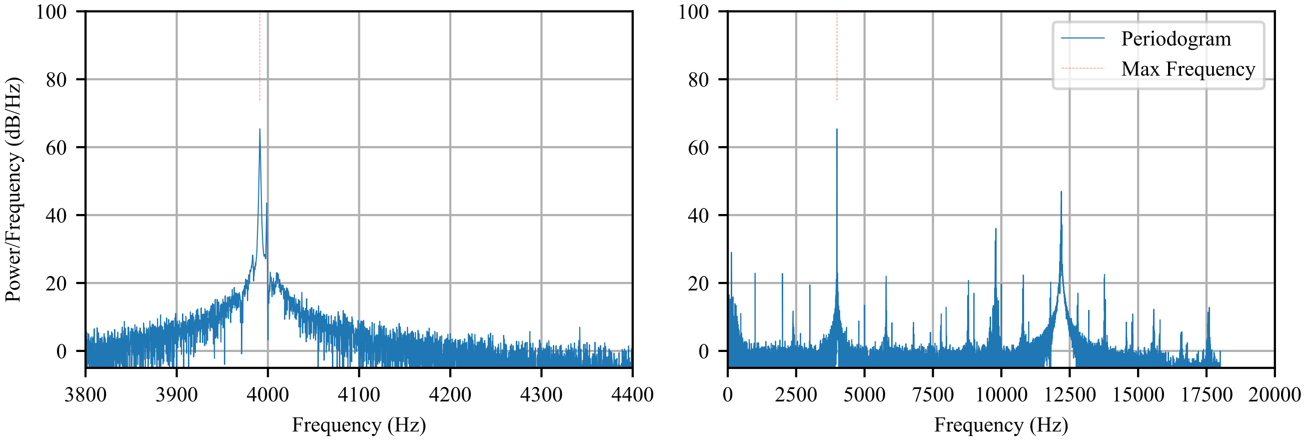


Figure 7: Example of the signal periodogram (Group #1, tap location 'N'). Note multiple additional frequency peaks.

In the context of the proposed research, several challenges are associated with this mode modeling approach. First, the relatively high signal length of $\sim 10^6$ samples. Second, a significant dynamic range of the amplitudes A_k and damping factors α_k. Finally, the expected number of signal components, K, is unknown, and inaccuracies in estimating K can alter estimation accuracy and overall performance.

Several established methods exist for estimating modal signal parameters from time-domain vibration data, including the Prony, matrix pencil, MUSIC (Multiple Signal Classification), and ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) methods Brown and Wells (2024). Each method possesses unique advantages and limitations. In this study, the ESPRIT method is employed due to its balanced performance in terms of

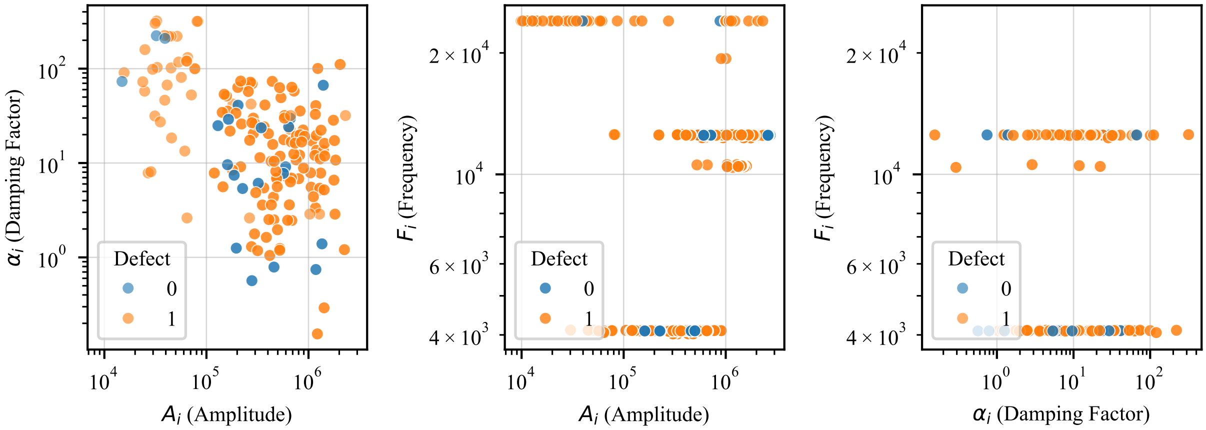


Figure 8: Visualization of ESPRIT-derived modal parameters (A_k, α_k, F_k) for $K = 5$ at the (F) tap location and with thermal post-processing; no obvious separation between the defective condition is evident.

resolution, robustness to a wide dynamic range of amplitude (A_k) and damping factor (α_k) values, and computational complexity.

An illustration of estimated parameters is presented in Fig. 8 for $K = 5$, (F) tap location and with thermal post-processing. The results show a few dominant frequencies but do not show a notable pattern for defect presence. Further analysis for additional K values also did not show this pattern.

3.3. STFT & MFCC

The primary advantage of the modeling in the previous section is its direct interpretability through physically meaningful parameters such as frequency, damping, and amplitude. However, modal modeling typically lacks the robustness associated with non-parametric methods like the Short-Time Fourier Transform (STFT) and Mel-Frequency Cepstral Coefficients (MFCCs). While not directly linked to physical properties like stiffness or damping, these are highly informative in characterizing the acoustic signal's spectral content.

To illustrate the STFT and MFCC time-frequency feature representations, we first zero-padded each signal to ensure the same signal length. Then, we used analysis of variance (ANOVA)-based feature selection that scored how strongly each time-frequency bin distinguishes defective from non-defective parts. Finally, the 100 highest-scoring features were selected.

STFT provides explicit frequency content at each point in time by computing localized Fourier transforms, retaining detailed frequency information and delivering more direct insights into how frequency components evolve over time. The STFT results for an FFT length of 8192 are presented in Fig. 9 and clearly show the differences between tapping location features. In the presented results, outer tapping locations correspond to modes with higher decay rates in the damped sinusoidal model (Eq. (2)).

The MFCC result in Fig. 10 also emphasizes the difference between central and outer taping locations. This observation also aligns with the physical intuition that defects alter the weaker initial impact response more noticeably than the long-term decay behavior.

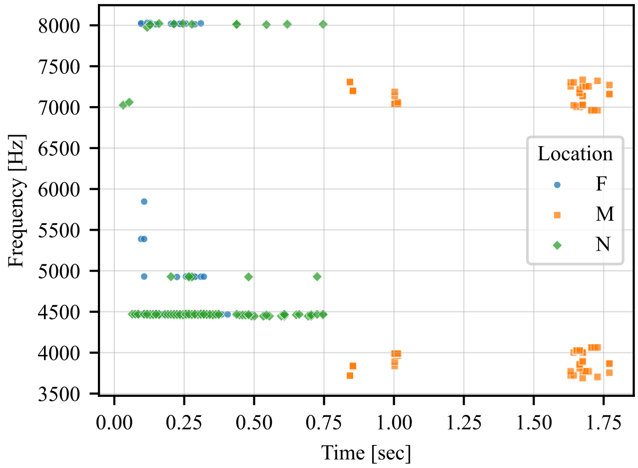


Figure 9: ANOVA-selected STFT time-frequency features, emphasizing the distinct spectral modes most sensitive to internal defects.

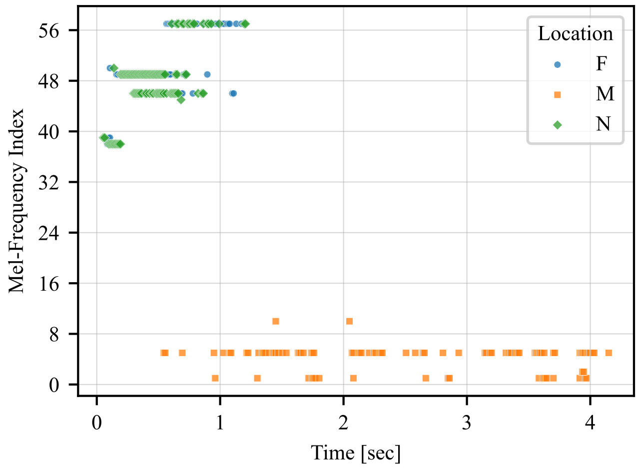


Figure 10: ANOVA-selected MFCC features highlighting the frequency bands that significantly discriminate between defect conditions for tapping locations.

4. Classification

4.1. Single Tap Classification Pipeline

Standard acoustic inspection often relies on fundamental frequency analysis (resonance testing). However, as shown in Section 3, the defect-induced shift in natural frequency for complex AM geometries is often masked by process variability. To overcome this, we propose an ML-based workflow that functions as a sophisticated measurement instrument.

The classification pipeline is based on MFCC-based feature extraction. First, ANOVA is used for feature selection to retain the most discriminative features, resulting in 1000 selected features. Since ANOVA outputs are correlated, principal component analysis (PCA) is used to decorrelate the resulting features and further reduce the dimensionality to 50 principal components. These 50 PCA components were standardized. Finally, a logistic-regression (LR) classifier was trained; it was used as the simplest linear regressor that minimizes the risk of overfitting on a small and imbalanced dataset (further discussed in Sec. 5.1). For performance evaluation, we used stratified grouped 10-fold cross-validation,

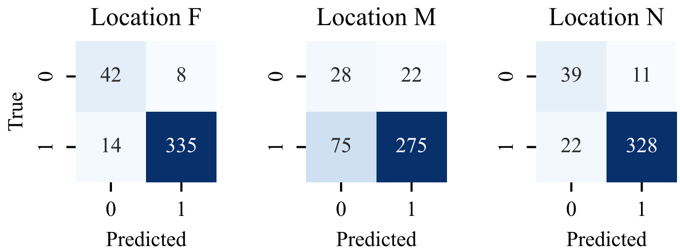


Figure 11: Confusion matrix of classifications at each location.

which ensures that taps from the same part are not included in both the training and the testing sets. The results were accumulated across all 10 splits.

We used a distinct classifier for each location, as follows from the exploratory analysis. The per-tap confusion matrix for all tap locations is presented in Fig. 11. Note the difference in classification performance (false positives) between different tapping locations, with outer locations having significantly higher performance, and the (F) location having the highest.

4.2. Fusion Analysis of Multiple Taps

4.2.1. Measurement Uncertainty and Fusion Strategies

The tapping measurements are subject to random variations caused by imperfect actuator, sample positioning errors, and environmental noise. As shown in Sec. 4.1, the acoustic response is also highly sensitive to impact location; slight positional deviations between repetitions contribute to Type A measurement uncertainty. This uncertainty manifests as variance in the classifier's probability estimation across repeated taps on the same sample. We hypothesize that fusing these classifier outputs reduces this uncertainty by treating multiple taps as independent measurements of the same physical state.

To aggregate the information from the five taps recorded at each location on a single part, we evaluated two fusion strategies. The straightforward approach involves a majority vote for each part and location, where we determine the mode from the five individual tap predictions. If all five predictions agree, the vote is unanimous; otherwise, the most frequent label is chosen as the part-level decision.

The probabilistic approach relies on LR classifier outputs, which can be interpreted as calibrated probabilities. These probabilities can be further applied for the fusion of multiple taps by the log-likelihood-ratio (LLR) test. Let $\mathcal{H}_1$ and $\mathcal{H}_0$ be the hypotheses for "defective" and "non-defective", respectively. The LR classifier output is $p(x_i|\mathcal{H}_1)$, representing the likelihood of a defect given the i-th tap feature vector x_i. Using the probabilities $p\left(x_i|\mathcal{H}_1\right)$ and $p\left(x_i|\mathcal{H}_0\right) = 1 - p\left(x_i|\mathcal{H}_1\right)$, the combined LLR over N taps is given by

$$LLR = \sum_{i=1}^{N} \log\left(\frac{p\left(x_i|\mathcal{H}_1\right)}{p\left(x_i|\mathcal{H}_0\right)}\right). \tag{3}$$

This scheme naturally accumulates evidence from individual taps into a single probabilistic decision criterion, $LLR \gtrless \gamma$, where γ is the decision threshold. Due to the inherent imbalance of the dataset, the applied threshold is based on *a priori* probabilities,

$$\gamma = \log\left(\frac{p\left(\mathcal{H}_0\right)}{p\left(\mathcal{H}_1\right)}\right). \tag{4}$$

4.2.2. Experimental Results

Our experimental results (Fig. 12) confirm the fusion improvement hypothesis. The combined results show a significant improvement in classification performance when compared to per-tap results. This validates the claim that the fusion process effectively mitigates the inherent Type A uncertainty of the simple tapping mechanism. Further discussion of this point is presented in Sec. 5.5.

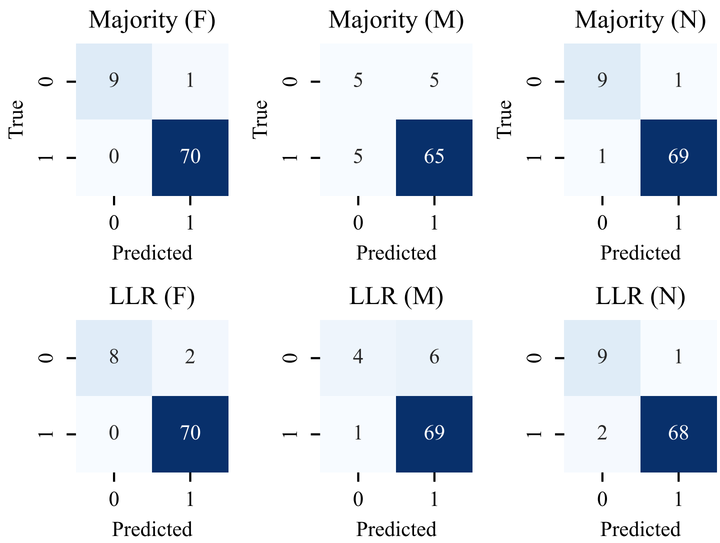


Figure 12: Confusion matrices; combining taps per location with majority vote and LLR.

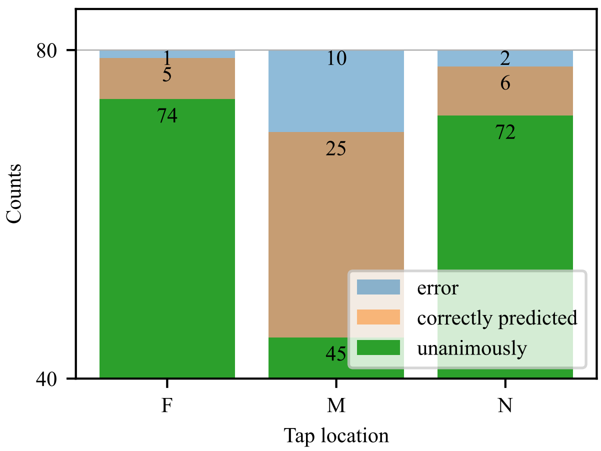


Figure 13: Majority vote combining analysis. While some of the defects are detected unanimously, the majority vote can improve the performance.

The majority vote combining influence is outlined in Fig. 13. The results show that while most decisions are unanimous, the majority vote has a significant influence on classification performance, with the greatest effect in the case of the (M) location.

The LLR combining method did not demonstrate any performance improvement compared to the majority vote. The distribution of LLR values is presented in Fig. 14, which illustrates the lower performance for the (M) location and highlights the challenge of dataset imbalance due to the small number of non-defective parts. In general, the decision threshold value γ can be further optimized. However, in the particular case presented, the number of non-defective parts in the dataset is insufficient for such an optimization.

As for location dependency, the results are significantly better for outer locations, with only one to two errors per dataset. The fusion of all or of only outer locations does not reveal additional performance gain.

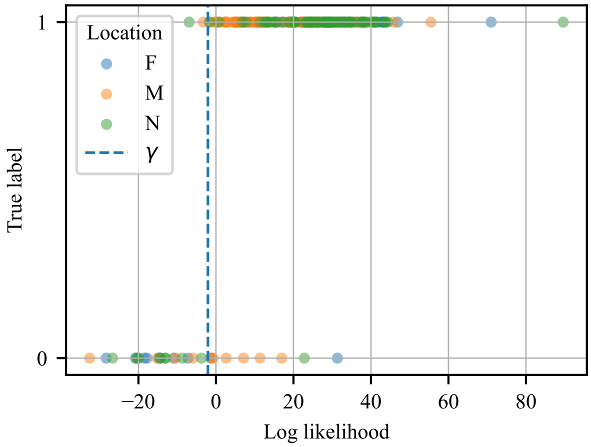


Figure 14: LLR scores defined in Eq. (3) for every part-and-location result in the bottom of Fig. 12. Points that fall to the right of the decision threshold γ are labeled positive ("defective"), while those to the left are labeled negative.

5. Discussion

5.1. Limitations of Results

The applied dataset is small and has inherent imbalance due to the low number of non-defect samples. While showing promising results, the selected classification pipeline (Sec. 4.1) is based on one of the basic linear classifiers. Nevertheless, further classification pipeline improvements (i.e., random forests or neural networks) and the corresponding hyperparameter optimization are expected to improve the performance on larger and more balanced datasets.

5.2. Tap Location

Our results show that classification accuracy is strongly influenced by tap position. As illustrated in Figs. 11 and 12, the middle (M) location has a significantly higher misclassification rate compared to the near (N) and far (F) positions. This indicates that each location couples to distinct vibrational modes and interference patterns, justifying the use of separate, location-specific models. Moreover, fusing multiple taps at the same location consistently improves classification performance, as is evident from a comparison between Figs. 11 and 12.

5.3. Feature Extraction

We have utilized all the methods outlined in Section 3, including natural frequencies, a damped sinusoidal model with various K hyperparameters, and STFT coefficients with different time-frequency parameters. Additionally, we employed the classical *catch22* package, which includes 22 lightweight time-series statistics Lubba, Sethi, Knaute, Schultz, Fulcher and Jones (2019), for each recording. Unfortunately, all these methods showed insufficient performance with a specificity (the percentage of correctly classified non-defect samples) of less than 50%. In the interest of brevity, these findings are not presented here.

5.4. Defect Classification

We extended the task to a five-class problem, with the aim of classifying the specific defect group (or location) listed in Table 1 rather than merely distinguishing defective from non-defective parts. However, the classification results were consistently poor. This suggests that the current experimental setup and classification pipeline lack the spatial resolution and discriminative power needed to distinguish between defect categories. Future work should investigate enhanced measurement configurations (e.g., additional microphone arrays), richer feature sets, and more advanced classification models to reliably localize and characterize defect types.

5.5. Measurement-Science Framing

In this measurement chain, the mechanical tapping process introduces random effects that we attempt to characterize. Despite the use of a robotic actuator, slight variations in the strike point and contact duration contribute to Type

A measurement uncertainty. Standard acoustic NDT studies often assume a single impact is representative; however, our characterization of the scatter in fundamental frequency and modal amplitudes confirms that a single tap provides an incomplete measurement of the system state.

The core novelty of this work from a measurement perspective is the treatment of sensor fusion as a formal method for uncertainty reduction. By applying the LLR (Eq. (3)), we aggregate independent observations to refine the measurand. Mathematically, if each individual tap carries a probability of error due to noise or poor coupling, the fusion of N taps acts as a weighted accumulation in the log-probability space. This reduces the effective uncertainty of the final soft-sensor measurement. The transition from per-tap results (Fig. 11) to fused results (Fig. 12) demonstrates a measurable decrease in the probability of erroneous decisions (Type I errors).

We distinguish between replication (repeated taps at one location) and variation of influence quantities (changing tap locations). Our findings indicate that changing the tap location does not merely provide more data but couples the measurement to distinct vibrational modes. In metrological terms, different tap locations constitute different measurement modalities. This justifies the use of location-specific models and multi-location fusion for AM, where process-induced variability can often mask defect signatures. Finally, the proposed workflow transforms a low-cost acoustic test into a traceable measurement system.

6. Summary & Conclusions

In this paper, we have presented a low-cost and simple NDT method for identifying defects in AM-produced Ti-6Al-4V specimens. The proposed acoustic tapping method exploits ML tools to achieve good classification results with the help of fusion techniques.

Well-established methods, such as natural frequency analysis, also provided physically tractable results but were less successful in classification. On the contrary, the robust and non-parametric MFCC-based pipeline showed excellent classification performance. Our results show a clear distinction between outcomes captured by tapping different locations. Moreover, recording the results from multiple taps at the same location may significantly improve the classification performance.

The primary limitation of the proposed method is its dependence on having labeled datasets of both defect-free ("known good") and defective samples to establish baseline relationships. Future research could address this constraint by employing unsupervised anomaly detection approaches. Additionally, while the current ML solution is constrained by the relatively small dataset, future work should investigate advanced machine and deep learning architectures to potentially improve detection performance and generalization.

Data Statement

Data will be made available on request.

References

Ansari, M.J., Arcondoulis, E.J., Roccisano, A., Schulz, C., Schlaefer, T., Hall, C., 2024. Optimized analytical approach for the detection of process-induced defects using acoustic emission during directed energy deposition process. Additive Manufacturing 86, 104218. doi:10.1016/j.addma.2024.104218.

ASTM International, 2023. Standard practice for resonance testing using the impulse excitation method. ASTM Standard E3397-23. URL: https://www.astm.org/e3397-23.html, doi:10.1520/E3397-23. accessed: 2025-05-31.

Boyadzhieva, S.M., Gutmann, F., Fischer, S.C.L., 2024. Simulation-based approach to estimate influencing factors on acoustic resonance spectra of additively manufactured mechanical metamaterials. Journal of Sound and Vibration , 118436doi:10.1016/j.jsv.2024.118436.

Brown, K.I., Wells, J.J., 2024. Comparison of modelled pursuits with esprit and the matrix pencil method in the modelling of medical percussion signals. Biomedical signal processing and control 89, 105777. doi:10.1016/j.bspc.2023.105777.

Cao, L., Wang, J., Su, J., Luo, Y., Cao, Y., Braatz, R.D., Gopaluni, B., 2025. Comprehensive analysis on machine learning approaches for interpretable and stable soft sensors. IEEE Transactions on Instrumentation and Measurement doi:10.1109/TIM.2025.3556830. early Access.

Çeçen, F., 2025. Scaled additive manufacturing for geometry-centric nondestructive evaluation: a case study on five distinct railway sleeper models. Nondestructive Testing and Evaluation 40, 1–21. doi:10.1080/10589759.2025.2599485.

Cha, Y.J., Ali, R., Lewis, J., Büyüköztürk, O., 2024. Deep learning-based structural health monitoring. Automation in Construction 161, 105328. doi:10.1016/j.autcon.2024.105328.

Ciaburro, G., Iannace, G., 2022. Machine-learning-based methods for acoustic emission testing: A review. Applied Sciences 12, 10476. doi:10.3390/app122010476.

Curreri, F., Patanè, L., Xibilia, M.G., 2021. Soft sensor transferability: A survey. Applied Sciences 11, 7710. doi:10.3390/app11167710.

Devenport, T., Rolfe, B., Pereira, M., Griffin, J.M., 2022. Analysis of acoustic emissions for determination of the mechanical effects of scratch tests. Applied Sciences 12, 6724. doi:10.3390/app12136724.

Ge, C., Qin, Y., Xie, K., Lu, Z., 2025. Fast and intelligent detection of concrete cracks based on sound signals and convolutional neural network. Frontiers in Built Environment 11, 1627643. doi:10.3389/fbuil.2025.1627643.

Guillen, D., et al., 2024. Critical review of LPBF metal print defects detection. Applied Sciences 14, 6718. doi:10.3390/app14156718.

Kolb, C.G., Zier, K., Grager, J.C., et al., 2021. An investigation on the suitability of modern nondestructive testing methods for the inspection of specimens manufactured by laser powder bed fusion. SN Applied Sciences doi:10.1007/s42452-021-04685-3.

Li, D., Nie, J.H., Wang, H., Ren, W.X., 2024. Loading condition monitoring of high-strength bolt connections based on physics-guided deep learning of acoustic emission data. Mechanical systems and signal processing 206, 110908. doi:10.1016/j.ymssp.2023.110908.

Li, H., Jiang, P., Wang, T., 2022. Deep learning based intelligent coin-tap test for defect recognition. arXiv preprint arXiv:2203.12594 doi:10.48550/arXiv.2203.12594, arXiv:2203.12594.

Li, J., Liu, X., Shi, T., Wang, H., Zhang, H., 2025. Stress and defect detection of x80 pipes based on acoustic characteristics analysis. Nondestructive Testing and Evaluation , 1–28doi:10.1080/10589759.2025.2459307.

Lin, Q., Yang, B., Liu, L., Yao, X., Ding, H., Zhang, S., Yang, M., Xiao, Q., Wang, W., 2025. Acoustic emission signal characteristics of crack and porosity defects formation during laser cladding wc/fe313 coatings. NDT & E International 155, 103404. doi:10.1016/j.ndteint.2025.103404.

Liu, W., Hu, P., Xiao, J., Yin, Q., Zhou, X., Li, H., Shen, S., 2024. High precision detection of artificial defects in additively manufactured ti6al4v alloy via laser ultrasonic testing. Journal of Materials Research and Technology 30, 8740–8748. doi:10.1016/j.jmrt.2024.05.140.

Lubba, C.H., Sethi, S.S., Knaute, P., Schultz, S.R., Fulcher, B.D., Jones, N.S., 2019. catch22: Canonical time-series characteristics: Selected through highly comparative time-series analysis. Data mining and knowledge discovery 33, 1821–1852. doi:10.1007/s10618-019-00647-x.

Pollock, A., 1969. Stress-wave emission in ndt. Non-destructive testing 2, 178–182. doi:10.1016/0029-1021(69)90109-1.

Queiroz, J.C., Santos, Y.T., da Silva, I.C., Farias, C.T., 2021. Damage detection in composite materials using tap test technique and neural networks. Journal of Nondestructive Evaluation 40, 27. doi:10.1007/s10921-021-00759-9.

Ramírez, I.S., Márquez, F.P.G., Papaelias, M., 2023. Review on additive manufacturing and non-destructive testing. Journal of Manufacturing Systems 66, 260–286. doi:10.1016/j.jmsy.2022.12.005.

Silva, M.I., Malitckii, E., Santos, T.G., Vilaça, P., 2023. Review of conventional and advanced non-destructive testing techniques for detection and characterization of small-scale defects. Progress in Materials Science 138, 101155. doi:10.1016/j.pmatsci.2023.101155.

Vangi, D., Bruzzi, M., Caron, J., Gulino, M., 2022. Compact probe for non-contact ultrasonic inspection with the gas-coupled laser acoustic detection (gclad) technique. Experimental Mechanics 62, 403–415. doi:10.1007/s11340-021-00786-7.

Wu, Y., Li, H., Jiang, P., Wang, T., 2025. Data-driven deep representation of acoustic signals amplifies the accuracy of coin-tap test for non-destructive detection of defects. NDT & E International 156, 103488. doi:10.1016/j.ndteint.2025.103488.

Xue, H., Li, X., Jin, Z., Li, X., Zhang, Y., Dong, W., Shang, E., Bao, C., Yu, H., 2025. 3D characterisation of metal powders for additive manufacturing using Nano-CT and deep learning. Nondestructive Testing and Evaluation 40, 4828–4845. doi:10.1080/10589759.2024.2431146.

Yang, D., Xiong, M., Wang, T., Lu, G., 2022. Percussion-based pipeline ponding detection using a convolutional neural network. Applied Sciences 12, 2127. doi:10.3390/app12042127.