

Unsupervised novelty detection for EMAT-based inspection on embedded hardware.

Eduardo A. Rodrigues and Moises A. Oliveira

Abstract—The interpretation of Electromagnetic Acoustic Transducer (EMAT) signals for assessing corrosion-induced wall loss in metallic plates is constrained by the scarcity of labeled defects in industrial operation, motivating one-class formulations. To address this, the 20,000-sample A-scan undergoes reduction to a 23-dimensional RMS envelope using band-pass filtering, dead-zone suppression, and sequential windowing. Next, an autoencoder, an Isolation Forest, and a centroid-distance K-Means are trained on normal specimens, and thresholded under a common three-sigma rule applied to their respective training scores. Finally, embedded inference on a Raspberry Pi 4 yields 99% specificity and a near-unity AUC-ROC for both through-wall holes and three graded levels of thickness loss. Five-fold cross-validation confirms low dispersion (standard deviation of 0.0% for sensitivity and specificity), supporting stability under limited training data.

Keywords—inspection, unsupervised novelty detection, feature extraction, embedded inference.

I. INTRODUCTION

The structural integrity of critical components in industrial sectors such as oil and gas, petrochemical, and power generation is essential for operational safety and the prevention of catastrophic failures. Among non-destructive testing (NDT) techniques, EMATs have gained prominence because they enable inspection of high-temperature surfaces and moving parts without physical contact, eliminating the need for couplants and facilitating automated inspection [1]. EMATs generate and receive ultrasonic waves via Lorentz force and magnetostriction, making them particularly suitable for corrosion monitoring under industrial conditions [2], [3]. Recent advances in coil design and signal processing have expanded their application to defect detection in aluminum and steel structures [4], [5].

Despite these advantages, EMAT signals in industrial environments are highly susceptible to electromagnetic noise, lift-off variations, and temperature drifts [6]. Moreover, continuous inspection generates massive volumes of A-scan data. Traditional analysis relies on a limited number of qualified inspectors, which restricts response time and introduces subjectivity. Automation using machine learning has been explored for ultrasonic and guided-wave inspection [7], [8], but most approaches require large labeled defects datasets.

In practice, labeled defect data are extremely scarce because structural failures are rare under normal operating conditions [9], [10]. This constraint motivates one-class or unsupervised

novelty detection, where models are trained exclusively on signals from intact structures and flag deviations as potential anomalies. Recent studies have applied autoencoders, Isolation Forests, and cluster-based methods to ultrasonic testing [11], [12], [13], [14], [15]. However, the majority of these works focus on offline analysis or high-performance computing, leaving a gap in low-cost embedded deployment.

This study addresses this gap by proposing a lightweight pipeline that compresses raw EMAT A-scans into a 23-dimensional RMS envelope. Three one-class detectors, an autoencoder, K-Means, and Isolation Forest, are trained exclusively on normal data under two static thresholds, executing on a Raspberry Pi 4 without accelerators. The main contributions are: demonstrating that front-end feature extraction dominates performance, making detector choice secondary; showing that RMS windowing yields near-perfect separability for thickness loss and through-wall holes; validating estimator stability via cross-validation and proving that scalar RMS aggregation achieves identical performance, reinforcing the sufficiency of physical features.

II. METHODOLOGY

A. Experimental data

The dataset comprises ultrasonic A-scans acquired with an EMAT transducer on metallic plates under varying structural integrity. Amplitudes are digitized on an 8-bit scale from 0 to 255. Each acquisition has a record length of 20,000 samples at a sampling rate of 500 MHz. This rate is consistent with the 5 MHz transducer center frequency and the subsequent 1–5 MHz band-pass filtering. Five physical conditions were sampled: one normal condition and four anomalous conditions.

Aluminum specimens were machined with three graded levels of thickness loss (1 mm, 3 mm, and 5 mm) and one condition with through-wall holes, all acquired at room temperature with a fixed lift-off of 2 mm. After preprocessing, the corpus contains 100 normal, 152 thickness-loss, and 48 through-wall-hole signals. The normal signals are split into 80 for training and 20 for validation used in threshold calibration. All defective acquisitions are reserved exclusively for testing. Given the limited scale and severity of this dataset, the results should be interpreted as a feasibility study under controlled laboratory conditions rather than as a demonstration of field readiness. To assess stability, five-fold cross-validation was performed on the normal training set consisting of 80% of normal samples. The remaining 20% of normal samples were held out as validation samples for threshold calibration after each fold.

Eduardo A. Rodrigues, Computer Engineering Undergraduate, SENAI CIMATEC University, Salvador-Bahia, e-mail: eduardo.rodrigues@aln.senaicimatec.edu.br; Moisés A. Oliveira, Assistant Professor, Federal University of Recôncavo da Bahia, Cruz das Almas-Bahia, e-mail: moises.oliveira@ufrb.edu.br.

The preprocessing workflow is shown in Figure 1. Raw amplitudes are centered and rescaled to $[-1, 1]$.

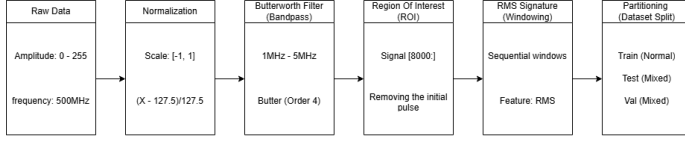


Fig. 1: Initial stages of the EMAT signal preprocessing pipeline.

A 4th-order digital Butterworth band-pass filter with pass-band $[1, 5]$ MHz is applied (zero-phase forward-backward implementation to avoid phase distortion near the region of interest). The later 8,000 samples are discarded to suppress the dead-zone from transducer excitation. The remaining 12,000-sample segment is partitioned into non-overlapping windows of 512 samples, yielding exactly 23 full windows; the remaining 224 samples are discarded. The RMS value of each window is computed according to Equation 1, producing a 23-dimensional feature vector.

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \quad (1)$$

B. One-class detectors

The three unsupervised models were trained on the 800 normal samples using the following parameters: the *Autoencoder* uses symmetric 23-16-8-16-23 architecture [14] trained for 200 epochs with Adam optimizer, learning rate 10^{-3} , batch size 16 with ReLU hidden layers, linear output, and MSE reconstruction error as novelty score; *Isolation Forest* uses an ensemble of 100 trees [15] with a subsample size of 256 and a contamination rate of 1%, where scores are inverted so that larger values indicate anomalies; and *K-Means* uses a single centroid μ fitted to the training vectors [13], using the Euclidean distance $d(z) = \|z - \mu\|$ as novelty score.

For the Autoencoder and K-Means detectors, the anomaly threshold is established as $T^* = \mu + 3\sigma$, where μ and σ are derived from the distribution of their respective training scores with MSE and Euclidean distance. Samples exceeding T^* are flagged as anomalies. Instead of this statistical rule, the Isolation Forest employs its inherent zero-boundary decision function, classifying any negative score as anomalous. Crucially, the parameters μ and σ are established using only the training set; the 100 normal samples reserved for testing are strictly held out to prevent any information leakage during the threshold definition.

III. RESULTS AND DISCUSSION

The ROC curves (Figure 2) and score distributions (Figure 3) demonstrate optimal class separation with near-unity AUCs across all detectors, confirming the robustness of the physical feature extraction stage.

Detector thresholds were set to maximize class separation: $T^* = 3.5 \times 10^{-5}$ for the autoencoder, $T^* = 1.94 \times 10^{-2}$ for K-Means, and native $T^* = 0.0$ for the Isolation Forest,

all calculated using the methods discussed previously. Under these criteria, all defects were successfully detected (100% sensitivity), while a single false positive in all models reduced specificity to 99.0%. The corresponding Wilson 95% confidence intervals are $[98.1\% - 100.0\%]$ and $[94.6\% - 99.8\%]$, respectively. Consequently, all ROC curves approach the unity AUC (≈ 1.000), showing robust diagnostic performance.

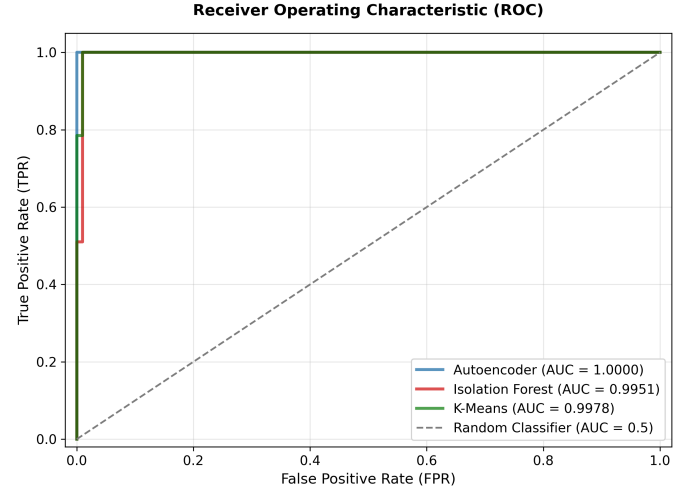


Fig. 2: ROC curves for the three evaluated unsupervised detectors on the independent test set.

To assess estimator stability, a five-fold cross-validation was conducted on the 800 normal training samples (640 train/160 validation per fold). For each partition, thresholds were recomputed from training scores to evaluate a fixed test set of (200 defective/100 normal) holdout samples. Table I reports the resulting cross-fold mean and standard deviation for sensitivity, specificity, and AUC.

TABLE I: Five-fold cross-validation results (mean $\pm$ std across folds). Test set fixed and independent.

Detector	Sensitivity (%)	Specificity (%)	AUC
Autoencoder	100.0 ± 0.0	99.0 ± 0.0	0.998 ± 0.000
Isolation Forest	100.0 ± 0.0	99.0 ± 0.0	0.995 ± 0.009
K-Means	100.0 ± 0.0	99.0 ± 0.0	0.998 ± 0.000

The lack of variance (standard deviation of 0.0) in both sensitivity and specificity, alongside negligible AUC fluctuations, demonstrates that the proposed feature extraction and thresholding policies are highly robust to training set partitions. This stability is critical for real-world industrial deployments, where algorithms must be commissioned using only a limited set of normal reference signals.

To test the claim that the front end dominates performance, we evaluated two scalar baselines on the same 23-dimensional RMS features: the sum of all window RMS values, and maximum window RMS. Both baselines were thresholded using the same $\mu + 3\sigma$ rule applied to their training scores. Both achieved 100% sensitivity and 99% specificity, with an AUC ≈ 1.0 , identical to the three detectors. This confirms that a simple threshold on aggregated or peak RMS suffices, reinforcing that the physical feature extraction, not the detector complexity,

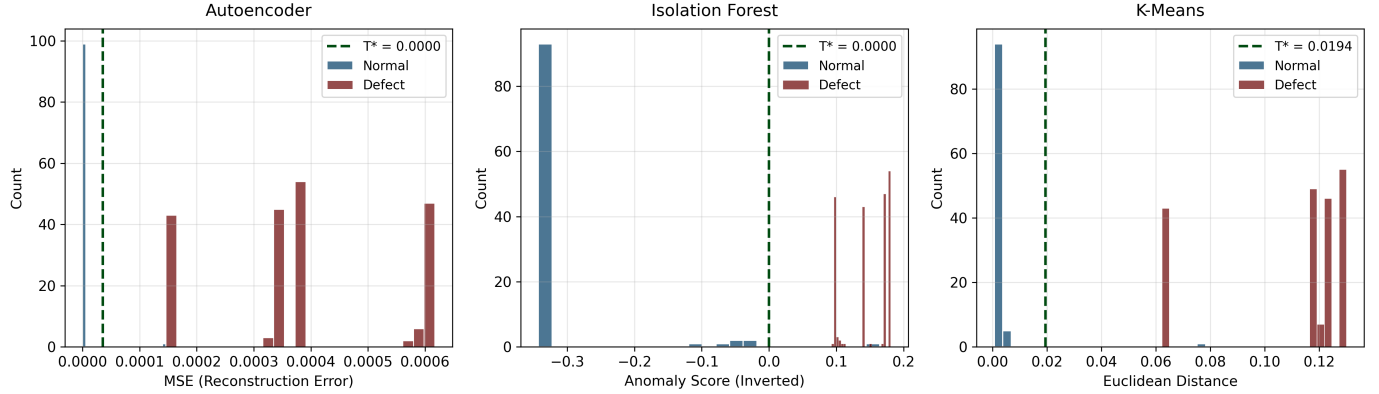


Fig. 3: Score distributions on the test set. The dashed vertical line indicates the anomaly decision threshold (T^*) optimized for each detector.

drives the separability. In an industrial setting, such scalar baselines could be implemented even on microcontroller-class devices without floating-point units, further lowering the barrier to deployment.

The autoencoder learns a low-dimensional reconstruction manifold, the Isolation Forest exploits partitioning depth, and K-Means relies on Euclidean proximity to a nominal centroid. Their performance is statistically indistinguishable, consistent with the unifying account of authors in [9]: the gap between deep and shallow anomaly detectors narrows under sufficiently informative feature spaces. The compactness of the 23-dimensional descriptor allows the complete pipeline preprocessing, feature extraction, and inference to execute on a Raspberry Pi 4 without dedicated accelerators or quantization, in contrast to the FPGA implementation reported in [10].

IV. CONCLUSIONS

A unified pipeline combining RMS windowing of band-pass-filtered EMAT A-scans with three one-class detectors was evaluated on a Raspberry Pi 4 for screening corrosion-induced thickness loss and through-wall holes. All detectors achieved 99.0% specificity, 100% sensitivity, and close to unity AUC under algorithm-specific thresholds, with scalar baselines performing identically, confirming that the physically motivated front end, rather than detector sophistication, provides the discriminative power. Five-fold cross-validation yielded zero standard deviation for specificity and sensitivity, indicating exceptional stability under limited normal data.

Future work should address real-world variabilities like temperature drift, lift-off, transducer aging, electromagnetic interference, and incorporate more features to help detect incipient corrosion and micro-cracking. Percentile-based or extreme-value thresholds can replace the static rule and the native zero boundary under heavy-tailed or drifting conditions, while standardized benchmarks will support the quantitative assessment of these embedded deployments.

ACKNOWLEDGMENTS

The authors thank SENAI CIMATEC for the infrastructure and execution of the work.

REFERENCES

- [1] T. Yin, L. Zheng, C. Ng, L. R. Francis Rose, A. Kotousov, "Advances in Rayleigh and shear-horizontal EMAT systems for NDT and SHM: A review," *NDT & E International*, v. 160, 103677, 2026.
- [2] J. Wang, X. Wu, Y. Song, L. Sun, "Study of the influence of the backplate position on EMAT thickness-measurement signals," *Sensors*, v. 22, n. 22, 2022.
- [3] P. Zuo, "Underwater quantitative thickness mapping through marine growth for corrosion measurement using shear wave EMAT with high lift-off performance," *Ultrasonics*, v. 143, 107426, 2024.
- [4] D. Shi, G. Sun, H. Liu, "Improvement of EMAT butterfly coil for defect detection in aluminum alloy plate," *Materials*, v. 18, n. 13, 2025.
- [5] Y. Shi, S. Tian, J. Jiang, T. Lei, S. Wang, Z. Lin, K. Xu, "Thickness measurements with EMAT based on fuzzy logic," *Sensors*, v. 24, n. 13, 2024.
- [6] I. Rossteutscher, O. Blaschke, F. Dötzer, T. Uphues, K. Stefan Drese, "Improved EMAT sensor design for enhanced ultrasonic signal detection in steel wire ropes," *Sensors*, v. 24, n. 22, 2024.
- [7] S. C. Hespeler, H. Nemati, N. Masurkar, F. Alvidrez, H. Marvi, E. Dehghan-Niri, "Deep learning-based time-series classification for robotic inspection of pipe condition using non-contact ultrasonic testing," *ASME Journal of Nondestructive Evaluation*, v. 7, n. 1, 2024.
- [8] J. Moon Ha, H. M. Seung, W. Chai, "Autoencoder-based detection of near-surface defects in ultrasonic testing," *Ultrasonics*, v. 119, art. 106637, 2022.
- [9] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, K.-R. Müller, "A unifying review of deep and shallow anomaly detection," *Proceedings of the IEEE*, v. 109, n. 5, pp. 756–795, May 2021.
- [10] P. Kashyap, K. Shivgan, S. Patil, B. Ramana Raja, S. Mahajan, S. Banerjee, S. Tallur, "Unsupervised deep learning framework for temperature-compensated damage assessment using ultrasonic guided waves on edge device," *Scientific Reports*, v. 14, art. 18887, 2024.
- [11] V. Tunukovic, S. McKnight, R. Pyle, Z. Wang, E. Mohseni, S. Gareth Pierce, R. K. W. Vithanage, G. Dobie, C. N. MacLeod, S. Cochran, T. O'Hare, "Unsupervised machine learning for flaw detection in automated ultrasonic testing of carbon fibre reinforced plastic composites," *Ultrasonics*, v. 140, art. 107313, 2024.
- [12] J. Wu, Y. Yang, Z. Lin, Y. Lin, Y. Wang, W. Zhang, H. Ma, "Weak ultrasonic guided wave signal recognition based on one-dimensional convolutional neural network denoising autoencoder and its application to small defect detection in pipelines," *Measurement*, v. 242, art. 116234, 2025.
- [13] A. E. Bouzenad, M. El Mountassir, S. Yaacoubi, F. Dahmene, M. Koabaz, L. Buchheit, and W. Ke, "A semi-supervised based K-means algorithm for optimal guided waves structural health monitoring: a case study," *Inventions*, v. 4, n. 1, 2019.
- [14] A. Abhaya, B. K. Patra, "An efficient method for autoencoder based outlier detection," *Expert Systems With Applications*, v. 213, 118904, 2023.
- [15] V. Toufigh, I. Ranjbar, "Unsupervised deep learning framework for ultrasonic-based distributed damage detection in concrete: integration of a deep auto-encoder and Isolation Forest for anomaly detection," *Structural Health Monitoring*, v. 23, n. 3, 2023.