electronics

Article

One Class Fault Detection in Rotating Machinery Using Distributional Features from Triggered Acoustic Emission Data

Nikolaos Angelopoulos [1,*], George Georgoulas [2] and Vassilios Kappatos [1]

[1] Hellenic Institute of Transport (HIT), Centre for Research and Technology Hellas (CERTH), 57001 Thessaloniki, Greece; vkappatos@certh.gr
[2] Robotics Group, Department of Mechanical Engineering and Aeronautics, University of Patras, 26504 Rio, Greece; georgoul@gmail.com
* Correspondence: naggel@certh.gr

Abstract

Acoustic Emission (AE) monitoring offers high sensitivity in detecting faults and defects in rotating machinery. Hit-based AE acquisition systems produce only short, intermittent waveform signals rather than continuous recordings. This work addresses the challenge of fault detection from such fragmented sensor data. It is demonstrated that individual AE hits are insufficient for reliable fault detection, as waveforms from different fault conditions overlap substantially in feature spaces. To overcome this, a distributional feature aggregation approach is proposed. AE hit features are extracted from each waveform, grouped into non-overlapping sequential bags, and summarised through order statistics and distribution moments. Four one-class classifiers namely: (i) Isolation Forest, (ii) PCA one class classifier, (iii) K-nearest neighbour one class classifier, and (iv) Local Outlier Factor were evaluated on a Drivetrain Dimulator (DTS) test rig that simulates faulty gear and bearing conditions. AE signals coming from the healthy condition were used to train the classifiers. Signal deviations due to the presence of gear and bearing faults were subsequently identified. Fault detection results show that bag-level distributional features substantially outperform per-hit features: four of five faults achieve 100% single-bag detection under five-fold cross-validation, while per-hit detection rates remain below 50% for four of five faults. The PCA one class classifier achieves 100% detection for all faults individually but is excluded from the final ensemble due to a 99% false alarm rate caused by the limited healthy training sample in a high-dimensional feature space. A CUSUM sequential detector applied to a three-method ensemble (Isolation Forest, KNN, and LOF) is evaluated on a chronological 80/20 split of the healthy data, triggering alarms for all five fault conditions with a false alarm rate of 0% at the recommended setting. Four faults are detected with sustained alarm patterns (54–100% alarm rates), while bearing outer race, the most challenging fault, is detected intermittently (10.5% alarm rate), demonstrating that sequential evidence accumulation can identify faults that are invisible to single-bag thresholding.

Keywords: acoustic emission; condition monitoring; one-class classification; CUSUM sequential detection; fault detection; distributional feature aggregation; anomaly detection; hit-based acquisition; gearbox monitoring; bearing fault

Academic Editors: Marcin Witczak, Leonidas Akritidis and Panayiotis Bozanis

Received: 7 April 2026
Revised: 6 May 2026
Accepted: 11 May 2026
Published: 14 May 2026

1. Introduction

The condition of gearboxes and bearing assemblies is a critical factor for the reliable and safe operation of rotating machinery. Defects in transmission systems such as faulty

 https://doi.org/10.3390/electronics15102105

gears and bearings can lead to malfunctions and mechanical failures in rotating machinery, wind turbines, and industrial gearboxes. An overheating bearing can increase the temperature and ignite the oil in a gearbox, resulting in complete destruction of the asset [1]. Unexpected gearbox failure can result in considerable expenditure due to loss of operation and the associated maintenance and repair costs, with gearbox repairs requiring the longest downtime compared to other systems [2]. Early detection of underlying faults can prevent further damage to sensitive drivetrain components and prevent unexpected failures. Furthermore, it can improve the maintenance schedule and increase asset operational efficiency, reliability, and safety.

Condition Monitoring (CM) techniques can lead to the detection of faults such as defective gears and defects on the rolling surfaces of bearings. Acoustic Emission (AE) is a CM technique with strong potential for fault detection and diagnosis in gearboxes and rotating machinery. AE is based on the detection of elastic waves that are generated inside the material when cracking, deformation, or other permanent changes occur. These ultrasonic waves are within the frequency range of 20–1000 kHz [3]. In the case of gearboxes and rotating parts, AE waves are generated due to interactions at contact surfaces of defective bearings and gears during rotation. The elastic AE waves propagate on the surface of the material and excite piezoelectric sensors that convert the elastic waves into electrical signals, which are then recorded using a data acquisition system.

Conventional AE analysis relies on the characterisation of individual AE hits using signal features such as amplitude, duration, rise time, and energy, augmented by both time domain and frequency domain analysis. These signal features have been used to detect and identify gear and bearing faults, and to correlate the AE activity with the underlying defects [4,5]. In addition, advanced signal processing techniques such as envelope analysis, spectral kurtosis, and wavelet transforms have been applied to determine the characteristic frequencies associated with bearing and gear meshing interactions from AE signals [6]. Wavelet transform [7] and Empirical Mode Decomposition (EMD) [8] have been further utilised for gearbox fault detection and diagnosis. In general, AE has been shown to be a more sensitive method for monitoring gearbox faults compared to vibration analysis methods, due to the higher sensitivity and signal-to-noise ratio of AE sensors [9,10]. For bearing fault detection, AE has been shown to detect subsurface defects such as spalling and raceway damage at earlier stages compared to conventional vibration monitoring [11,12]. For faulty gear detection, the generated AE activity has been related to tooth surface pitting, root cracks, and meshing irregularities [13]. A threshold-based methodology has been presented to detect AE bursts caused by faults in a planetary gearbox under varying operational conditions [14]. Fixed threshold values can lead to burst detection at certain operational conditions, but they can lead to poor detectability results at varying conditions. Using signal factors such as the Root Mean Square (RMS) value can lead to improved consistency across different conditions, but low burst levels might be missed. Optimum burst detection results under variable conditions can be achieved using the Edge Detection (EDG) method that is based on the demodulation and subsequent differentiation of the resulting envelope of the AE signal [14]. In another study, Leaman et al. [15] proposed a pattern recognition burst detection methodology using an ANN. Amplitude demodulation is applied on the AE signals, followed by signal segmentation that is transformed into a feature vector. The signal vectors are classified through a multilayer perceptron neural network to detect the presence or absence of bursts. The method demonstrated clear burst detection in signals related to the presence of gearbox faults under varying operational conditions of rotational speed and load [15]. These studies typically employ continuous AE acquisition, where the signal is sampled without interruption, producing a complete time-domain record of the AE signal.

https://doi.org/10.3390/electronics15102105

However, many AE systems do not acquire data continuously. Instead, they operate in a hit-based mode, where the system records a short waveform only when the signal amplitude exceeds a pre-defined threshold. Between AE hits, no data is recorded. Hit-based acquisition reduces data volume and storage requirements, making it the standard operating mode in many commercial AE systems. The resulting AE dataset is a stream of short waveform segments lasting a few milliseconds each, followed by gaps during which no signal information is available. Individual AE hit recording poses a fundamental challenge for fault diagnosis. Each individual hit captures only a small window of the emission activity, and its features are influenced by the nature of AE source mechanisms, wave propagation, signal reflections, as well as constructive and destructive interference. As a consequence, the feature distributions of individual AE hits from different machine conditions may overlap substantially, making it challenging to classify an individual hit to a healthy or faulty state.

To automate the detection and identification of gear and bearing faults, Machine Learning (ML) techniques can be applied on the recorded AE signals. Several ML techniques that detect wear and fatigue damage in bearings from AE signals have been presented [16]. Other approaches such as Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) have been applied on AE signals from different bearing fault patterns [17,18]. Ensemble anomaly detection strategies, which combine multiple detectors to improve robustness, have also been applied to bearing fault diagnosis from vibration and AE data [19]. More recently, deep learning methods have been applied to AE-based structural health monitoring in operational environments [20]. Despite these advancements, the limitations of hit-based acquisition persist. AE hit signal features are sensitive to rotational speed, load, and lubrication state [21], as well as to operational background noise. To address these limitations and overcome the per-hit fault identification challenge, a bag-level signal representation can be adopted, where AE hits acquired over a specific time window are treated as a bag of signal instances, drawing on concepts from Multiple Instance Learning [22]. Analysis of AE features across such bags can identify signal characteristics that are not observable at the individual AE hit level, improving the fault discrimination capabilities of AE monitoring.

In this work, a fault detection methodology for hit-based AE data is proposed, which is usually the first line of defence in every CM system. First, N groups of consecutive AE hits are formed into non-overlapping bags and processed through order statistics and distribution moments, yielding bag-level distributional features. Second, multiple one-class anomaly scorers are combined to produce a single anomaly score per bag. Third, the Cumulative Sum (CUSUM) procedure [23] is applied to accumulate these scores over time and trigger an alarm when a shift in the overall distribution is observed. Four one-class techniques are evaluated: Isolation Forest [24], Principal Component Analysis (PCA) one class classifier, K-nearest neighbour one class classifier (KNN) [25], and Local Outlier Factor (LOF) [26]. The evaluation indicates that while PCA detector achieves high individual detection rates, it produces high levels of false alarms due to the limited availability of healthy training data relative to the feature dimensionality. Therefore, the final one-class ensemble comprises three methods: Isolation Forest, KNN, and LOF, whose scores are monitored using CUSUM sequential detection.

This approach is validated on AE hits recorded using a DTS test rig operated under controlled gear and bearing fault conditions. Three bearing faults (ball, inner race, outer race) and two gear faults (chipped tooth, missing tooth) were tested, with promising results. A sensitivity analysis across CUSUM parameters demonstrates the robustness of the detection framework. The proposed methodology is the first step towards the development of a decision support system that could assist maintenance engineers in interpreting hit-based AE data and detect faults in rotating machinery.

The remainder of this paper is organised as follows. Section 2 describes the experimental setup, the per-hit feature extraction, the bag-level distributional representation, the one-class methods, the CUSUM sequential detector, and the evaluation protocol. Section 3 presents the results including a sensitivity analysis across CUSUM parameters and bag sizes. Section 4 discusses the findings, their practical implications for a decision support system, and the limitations of this study. Section 5 concludes the paper suggesting directions for future work.

2. Materials and Methods

2.1. Test Rig and Data Acquisition

Experiments were conducted on a DTS test rig provided by SpectraQuest (Richmond, VA, USA) to replicate common gearbox faults such as faulty gears and axle bearings in a laboratory environment. The DTS that was used is presented in Figure 1 with the position of the gears and bearings shown in (Figure 2). Six operating conditions were tested, including the following defect types and health configuration:

1. Healthy—baseline configuration with healthy gear and healthy bearing.
2. Bearing ball—a seeded defect on a rolling element.
3. Bearing inner race—a seeded defect on the inner raceway.
4. Bearing outer race—a seeded defect on the outer raceway.
5. Chipped gear—a localised chip on a gear tooth.
6. Missing tooth—complete removal of one gear tooth.

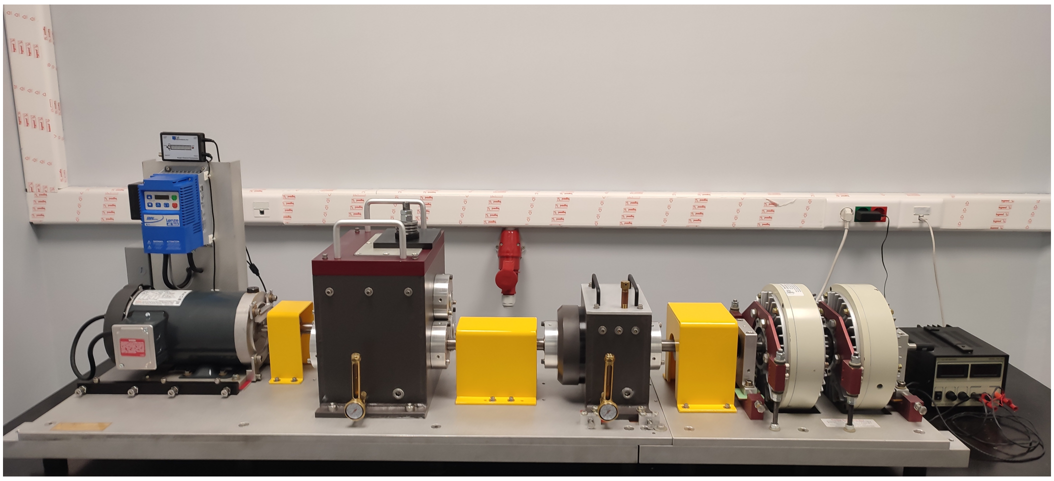

Figure 1. Drivetrain Simulator (DTS) test rig.

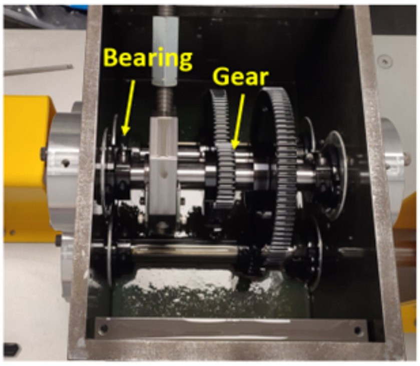

Figure 2. Gearbox interior showing the positions of the tested gear and bearing.

For each fault condition, AE measurements were carried to record the signals associated with the specific fault conditions tested. After AE sensor installation (Figure 3), test runs on the DTS were carried out at 800 RPM for 5 min of continuous operation, corresponding to more than 5000 shaft revolutions per condition.

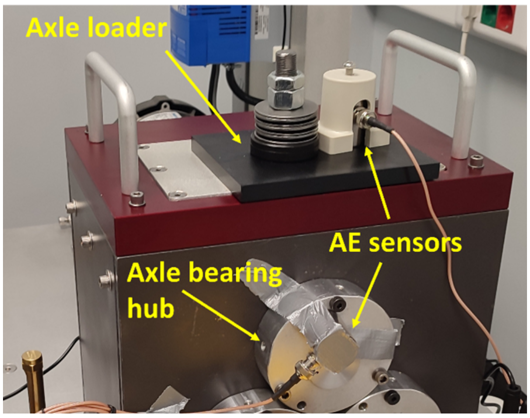

Figure 3. AE sensor positions on drivetrain simulator.

The AE measurements (hit based) were carried out using the Micro-SHM system provided by Mistras Group (Princeton Junction, NJ, USA) Two piezoelectric PK15I sensors were used provided by Mistras Group, that operate at the frequency range of 100–450 kHz. The recording and storing of the AE signals were performed using the AE-Win V1.30 software with the parameters shown in Table 1.

Table 1. AE acquisition parameters.

Parameter	Value
Threshold	45 dB
Analogue filter	20 kHz
Sample rate	5 MSPS (Mega Samples per Second)
Length	15k samples
PDT (Peak Definition Time)	800 µs
HDT (Hit Definition Time)	1600 µs
HLT (Hit Lockout Time)	1600 µs
Max. duration	25 ms

The acquisition parameters were determined after preliminary testing. Proper acquisition parameter configuration aims to minimise background noise recording, maintain high data acquisition rates, maximise the acquisition duration per AE hit and ensure minimal signal loss. In addition, appropriate timing parameter settings reject the recording of insignificant signal echoes and prevent premature termination of AE hit acquisition at low amplitudes.

The overall detection pipeline is illustrated in Figure 4. Raw AE hits are first transformed into 24-dimensional feature vectors, then grouped into non overlapping bags and summarised into 203-dimensional distributional descriptors. Three one class anomaly scorers are applied independently, their outputs are z-normalised and averaged into an ensemble score, and the CUSUM sequential detector accumulates evidence over consecutive bags to trigger alarms. Note: the PCA one class classifier is not included in the final ensemble due to its high False Alarm Rate (FAR), as discussed in Section 3. The pipeline is meant to operate in an online fashion, with each bag processed sequentially as data are acquired, enabling real-time fault detection.

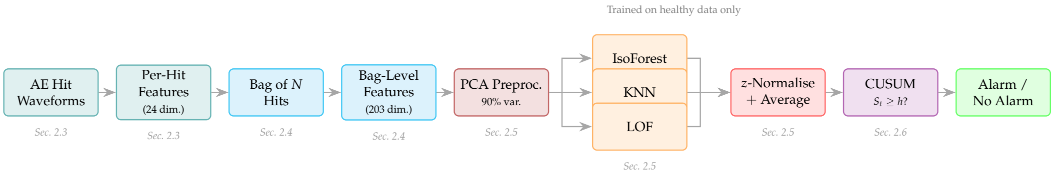

Figure 4. Overview of the detection pipeline. AE hit waveforms are transformed into 24 per-hit features (Section 2.3), grouped into bags and summarised as 203-dimensional distributional descriptors (Section 2.4), reduced via PCA preprocessing to 24 dimensions for bag-50 or 33–34 dimensions for bag-20 (Section 2.5), scored by three one-class methods (Section 2.5), combined via z-normalised averaging, and monitored by a CUSUM sequential detector (Section 2.6). Colour coding: teal/cyan—data processing stages; orange—one-class detectors; red—ensemble scoring; violet—sequential detection; green—alarm output.

2.2. Data Characteristics

Table 2 summarises the hit-level statistics for each condition. The total number of hits ranges from 1832 (missing tooth) to 5514 (bearing ball), with recording durations between 53 and 218 s. Hit rates vary between 25 and 49 Hz. The median waveform snippet duration ranges from 1.3 ms (missing tooth) to 3.1 ms (most conditions). Given that one shaft revolution takes approximately 259 ms, each individual hit covers at most 1.2% of a single revolution.

Table 2. Hit-level statistics per operating condition (Channel 1 only).

Condition	Hits	Dur. (s)	Rev.	Hits/rev	Med. Gap (ms)	Med. Snip. (ms)	Rate (Hz)
Healthy	4664	179.4	694	6.72	10.7	3.1	26.0
Bearing ball	5514	218.1	843	6.54	13.4	3.1	25.3
Bearing inner	5186	105.9	410	12.66	9.7	2.4	49.0
Bearing outer	3822	103.4	400	9.56	10.2	3.1	37.0
Chipped gear	3517	139.6	540	6.52	16.1	3.1	25.2
Missing tooth	1832	53.4	207	8.87	39.8	1.3	34.3

Figure 5 shows representative waveform snippets and their power spectral densities (PSDs) for each condition. Figure 6 overlays the median PSDs, which depict the high spectral similarity across conditions.

2.3. Per-Hit Feature Extraction

A set of 24 features is extracted from each AE hit, organised into five groups (Table 3).

Spectral features (8). The Welch PSD estimate is computed for each waveform using a Hann window. The PSD is integrated over four frequency bands: 20–100, 100–300, 300–600, and 600–1000 kHz, yielding into four band power values. Each band power is also divided by the total spectral power to obtain four band power ratios that characterise the relative spectral energy distribution.

Waveform features (3). The spectral centroid, computed as the energy-weighted mean frequency from the PSD. The waveform kurtosis, quantifying the peakedness of the amplitude distribution [27]. The crest factor (peak-to-RMS ratio) measuring the severity of transient content within the waveform [28].

Envelope features (4). The waveform is bandpass-filtered between 100 kHz and 1 MHz using a fourth-order Butterworth filter, and the analytic signal envelope is obtained via the Hilbert transform. Four statistics of the envelope amplitude are computed: the mean, standard deviation, kurtosis, and coefficient of variation.

AE parametric features (6). The hit-based acquisition system automatically extracts standard parametric descriptors for each hit [21,29]: amplitude (AMP, the peak signal level in dB), duration (DURATION, the time in microseconds from the first to the last threshold crossing), counts (COUN, the number of positive threshold crossings), rise time (RISE, the time from the first threshold crossing to the peak amplitude), energy (ENER, the integral of the rectified signal), and absolute energy (ABS-ENERGY, the integral of the squared signal in attojoules).

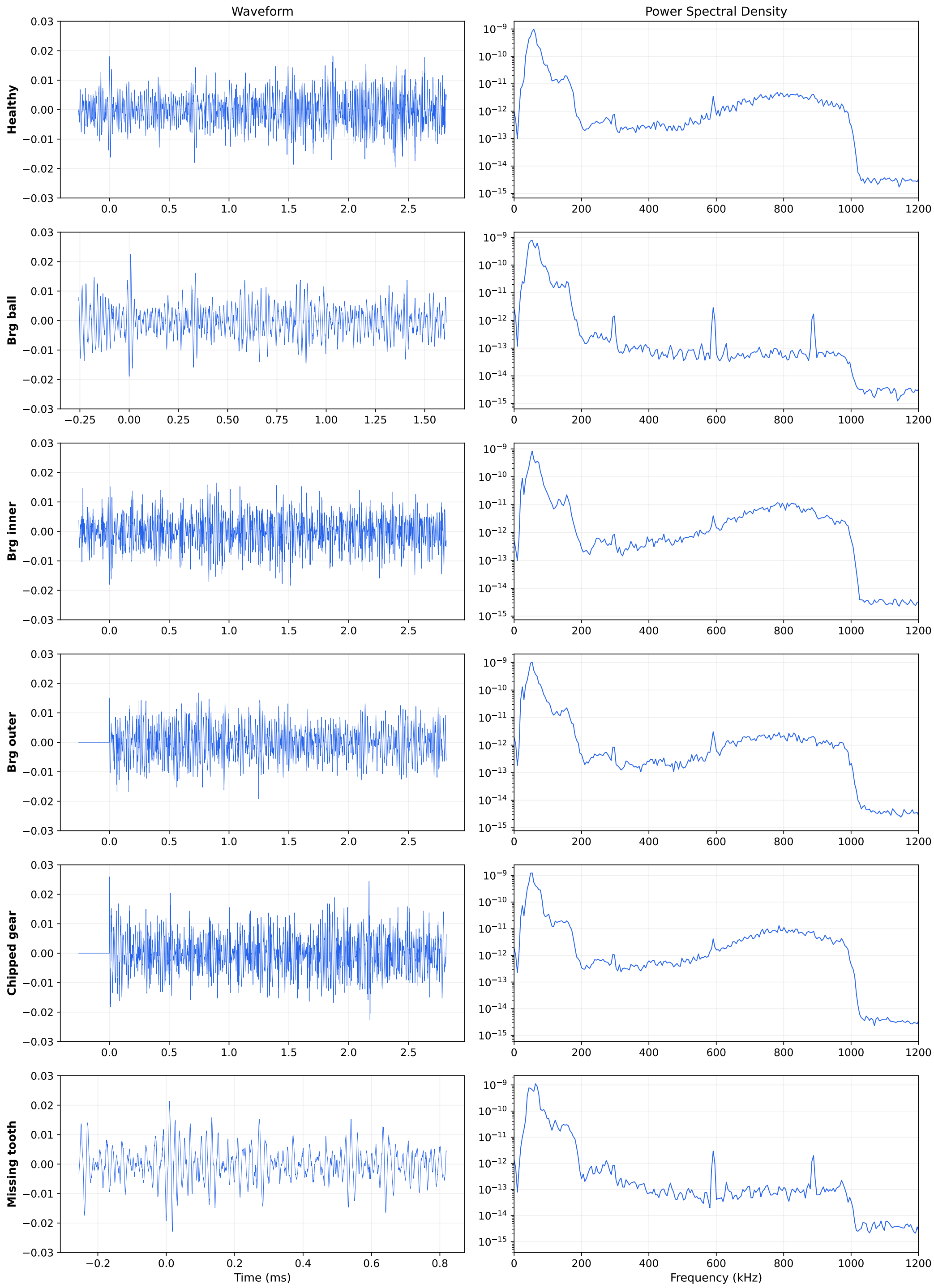

Figure 5. Representative AE waveform snippets (**left**) and corresponding PSDs (**right**) for each condition, (**top** to **bottom**) Healthy, Bearing ball, Bearing inner, Bearing outer, Chipped gear, Missing tooth.

Timing and size features (3). The inter hit gap is defined as the time elapsed between the current hit and the immediately preceding hit, recorded both in seconds (`gap_before`) and as a fraction of one shaft revolution (`gap_before_as_rev`). The waveform sample count (n_{samples}) records the number of digitised samples in the hit.

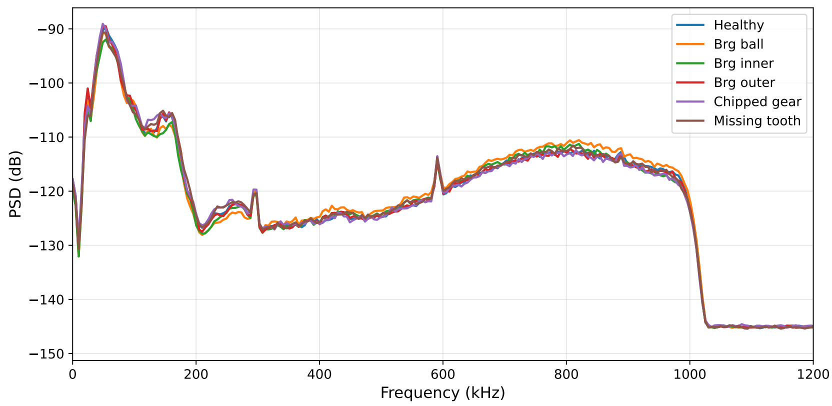

Figure 6. Median PSDs overlaid for all six conditions.

Table 3. Per-hit feature set (24 features).

#	Category	Feature	Computation
1–4	Spectral	Band power (4 bands)	Welch PSD, trapezoidal integration
5–8	Spectral	Band power ratio (4 bands)	Band power/total power
9	Waveform	Spectral centroid	Energy-weighted mean frequency
10	Waveform	Kurtosis	4th standardised moment
11	Waveform	Crest factor	Peak/RMS
12	Envelope	Envelope mean	Hilbert envelope mean amplitude
13	Envelope	Envelope std	Hilbert envelope standard deviation
14	Envelope	Envelope kurtosis	Hilbert envelope peakedness
15	Envelope	Envelope CoV	Envelope std/mean
16	AE param.	AMP	Peak amplitude (dB), from hardware
17	AE param.	DURATION	First-to-last threshold crossing (μs)
18	AE param.	COUN	Threshold crossing count
19	AE param.	RISE	Time to peak amplitude (μs)
20	AE param.	ENER	Rectified signal integral
21	AE param.	ABS-ENERGY	Squared signal integral (aJ)
22	Timing	gap_before	Time since previous hit (s)
23	Timing	gap_before_as_rev	Gap as fraction of shaft revolution
24	Size	n_{samples}	Waveform sample count

2.4. Bag-Level Distributional Representation

Hits are ordered chronologically and partitioned into non-overlapping bags of size N. For each of the 24 per hit features and each bag, eight distributional summary statistics are computed: mean, standard deviation, 5th percentile, median (50th percentile), 95th percentile, interquartile range, skewness, and kurtosis. This yields $24 \times 8 = 192$ distributional features.

Additionally, 11 temporal bag-level features are computed directly from the hit timestamps within each bag:

- Hit rate: the number of hits divided by the time span of the bag (hits/s), providing a direct measure of AE activity.
- Inter-hit gap statistics (8 features): the mean, standard deviation, coefficient of variation, 5th and 95th percentiles, skewness, kurtosis, and maximum of the inter-hit time gaps within the bag.

- Long-gap indicators (2 features): the count and fraction of inter-hit gaps exceeding one shaft revolution period T_{rev}, capturing periods of AE silence that may indicate intermittent fault activity.

The complete bag descriptor thus comprises 203 features. Two bag sizes are evaluated: $N = 50$ ($\approx$2–4 s at the observed hit rates; 93 healthy and 395 fault bags) and $N = 20$ ($\approx$0.8–1.5 s; 233 healthy and 991 fault bags). The smaller bag size produces 2.5$\times$ more healthy training samples, which improves threshold calibration stability at the cost of weaker distributional summaries. The derived datasets are summarised in Table 4.

Table 4. Bag-level dataset summary.

Condition	Total Hits	Bags (N = 50)	Bags (N = 20)
Healthy	4664	93	233
Bearing ball	5514	110	275
Bearing inner	5186	103	259
Bearing outer	3822	76	191
Chipped gear	3517	70	175
Missing tooth	1832	36	91
Total	24,535	488	1224

2.5. One-Class Anomaly Detection

Four one class methods are evaluated, all trained exclusively on healthy condition data. Each method produces a scalar anomaly score $s_m(\mathbf{x})$ for a test sample $\mathbf{x}$, with higher values indicating greater deviation from the healthy baseline:

Isolation Forest (IsoForest) [24]: An ensemble of 100 random isolation trees that recursively partition the feature space along random axes. Anomalous samples require fewer random splits to isolate and thus have shorter average path lengths. The anomaly score is derived from the normalised mean path length across all trees; higher values correspond to greater anomaly. The contamination parameter is set to 0.05.

Principal Component Analysis one-class classifier (PCA): PCA is fit on healthy training data, retaining components explaining at least 90% of the variance (typically 25 components for bag-50). The anomaly score is the mean squared reconstruction error between the original sample and its PCA reconstructed version.

K-Nearest Neighbour one-class classifier (KNN) [25]: The anomaly score for a test sample is the mean Euclidean distance to its K nearest healthy neighbours, where the neighbourhood size is set to $K = \lceil \sqrt{n_{\text{train}}} \rceil$ (9 for bag-50, 14 for bag-20).

Local Outlier Factor (LOF) [26]: LOF compares the local reachability density of a test sample to the densities of its neighbours. Samples residing in sparser regions receive higher LOF scores. The algorithm is trained on healthy data, with MinPts $= 20$.

KNN and LOF are distance-based methods and are therefore susceptible to the curse of dimensionality in the 203-dimensional bag feature space. To address this, Principal Component Analysis is applied as a preprocessing step prior to the distance-based detectors, with the number of retained components determined by a 90% variance retention threshold computed exclusively from the healthy training data. This reduces the bag-50 feature space from 203 to 24 dimensions and the bag-20 space from 203 to 33–34 dimensions. It is important to note that this use of PCA differs from the PCA one class classifier described above: here PCA is only used as a dimensionality reduction preprocessor, not as a detector. The effect of this preprocessing on detection performance is evaluated in Section 3.3.

Ensemble scoring. Each method's raw scores are first z-normalised using the mean $\hat{\mu}_{0,m}$ and standard deviation $\hat{\sigma}_{0,m}$ of the method's scores computed on the healthy training set, to reside in a common scale. The ensemble score is then the unweighted average across

M methods (simple averaging is a robust combination strategy for anomaly detectors [19]). The initial evaluation uses all four methods ($M = 4$). As shown in Section 3.2, PCA produces an unacceptably high FAR, leading to a final ensemble of $M = 3$ (IsoForest, KNN, LOF).

2.6. Sequential Detection with CUSUM

While bag-level anomaly scores improve feature separability, setting a reliable per bag detection threshold is better addressed by aggregating sequential information. The CUSUM (Cumulative Sum) procedure [23,30] addresses this by accumulating evidence over consecutive bags.

The ensemble scores are first standardised using healthy training data:

$$\tilde{z}_t = \frac{s_{\text{ens}}(\mathbf{x}_t) - \hat{\mu}_0}{\hat{\sigma}_0} \tag{1}$$

Because larger ensemble scores indicate greater deviation from the healthy baseline, a one-sided upper CUSUM is adopted:

$$S_t = \max\{0,\ S_{t-1} + \tilde{z}_t - k\}, \quad S_0 = 0 \tag{2}$$

An alarm is raised when $S_t \geq h$, after which S_t is reset to zero. The allowance parameter k controls sensitivity: the suggested k for detecting a shift of δ standard deviations is $k = \delta/2$ [31]. Setting $k = 0.5$ targets 1σ shifts, appropriate for moderate fault signatures. The decision threshold h controls the FAR vs. detection speed trade-off [32]. A sensitivity analysis across $k \in \{0.25, 0.5, 1.0\}$ and $h \in \{3, 5, 8\}$ is presented in Section 3.6.

2.7. Evaluation Protocol

Two complementary evaluation protocols are used, the first targeting a static setting and the other one aiming for a sequential setting.

Single-bag detection (Sections 3.2 and 3.4) is evaluated via five-fold chronological cross-validation on the healthy data. The healthy bags are divided chronologically into five approximately equal folds. In each fold, four folds serve as training data and classifiers are fit with the 95 percentile threshold computed on these bags only (FAR of 5%).

CUSUM sequential detection (Sections 3.5 and 3.6) is evaluated using a chronological 80/20 split of the healthy data. The first 80% of healthy bags (74 bags for $N = 50$; 186 bags for $N = 20$) are used to train the three-method ensemble and calibrate the score standardisation parameters. The remaining 20% of healthy bags (19 bags for $N = 50$; 47 bags for $N = 20$) form a held-out healthy test stream, and all fault bags are scored by the trained model. Therefore, the CUSUM statistic is then run as a single continuous sequence on each condition's score stream. This protocol preserves strict chronological ordering, simulating the intended deployment scenario.

Reproducibility

All experiments were implemented in Python 3.10 using scikit-learn 1.3 for the one-class classifiers, SciPy 1.11 for signal processing and statistical functions, and NumPy 1.25 for numerical computation. All features were standardised (zero mean, unit variance) using statistics computed exclusively from the healthy training data, prior to fitting PCA, KNN, LOF, and IsoForest. Where PCA preprocessing is applied, the PCA transform is also fit exclusively on the healthy training data and applied to all test and fault bags. The IsoForest random seed was fixed to 42 for reproducibility. Bootstrap confidence intervals used 2000 resamples with a fixed seed. Only data from Channel 1 (the sensor mounted on the gearbox top lid) were used throughout.

3. Results

3.1. Per-Hit Classification

Table 5 presents the anomaly detection rates when individual AE hits are used as classification samples. The total dataset comprises 24,535 hits across the six conditions. At the hit level, a large number of healthy samples are available and the P95 threshold is stable under resampling, so in-sample thresholding is acceptable for this baseline comparison. The per-hit results serve to motivate the bag-level representation rather than as a deployable detection protocol. Importantly, the proposed method does not attempt to classify individual AE hits as normal or fault-related, individual AE hits from different conditions exhibit significant overlap in feature space, as confirmed by the results below. Fault discrimination is achieved at the distributional level across bags of consecutive hits, not at the individual hit level. The results confirm that individual hits are insufficient for reliable fault detection. Bearing outer race is effectively indistinguishable from healthy (ensemble detection rate of 3.9%), and chipped gear is only marginally above the false alarm floor (7.2%). Only missing tooth (LOF: 81.6%) and bearing inner (LOF: 49.0%) achieve meaningful per-hit detection.

Table 5. Per-hit DRs (%). Healthy row: FAR (in-sample P95 threshold); fault rows: DR.

Condition	IsoForest	PCA	KNN	LOF	3-Ens.	4-Ens.
Healthy (FAR)	5.0	5.0	5.0	5.0	5.0	5.0
Bearing ball	22.9	10.5	9.4	22.8	21.5	19.3
Bearing inner	48.0	26.2	26.6	49.0	48.9	44.9
Bearing outer	4.2	3.0	3.3	5.5	3.9	3.7
Chipped gear	6.1	6.3	7.1	9.3	7.2	7.2
Missing tooth	55.3	29.9	61.2	81.6	74.2	66.0

3.2. Static Bag Level Detection

Tables 6 and 7 present the bag-level detection rates.

Table 6. Bag-level ($N = 50$) detection rates (%) from 5-fold CV, with 95% bootstrap CI for the 3-method ensemble.

Condition	IsoF.	PCA	KNN	LOF	4-Ens.	3-Ens.	95% CI
Healthy (FAR)	7.5	98.9	20.4	12.9	67.7	16.1	[8.6, 24.7]
Bearing ball	100.0	100.0	100.0	100.0	100.0	100.0	[100.0, 100.0]
Bearing inner	100.0	100.0	100.0	100.0	100.0	100.0	[100.0, 100.0]
Bearing outer	12.1	100.0	28.4	16.3	98.4	23.2	[18.9, 27.4]
Chipped gear	8.6	100.0	100.0	100.0	100.0	100.0	[100.0, 100.0]
Missing tooth	100.0	100.0	100.0	100.0	100.0	100.0	[100.0, 100.0]

Table 7. Bag-level ($N = 20$) detection rates (%) from 5-fold CV, with 95% bootstrap CI for the 3-method ensemble.

Condition	IsoF.	PCA	KNN	LOF	4-Ens.	3-Ens.	95% CI
Healthy (FAR)	7.3	70.8	10.3	6.9	28.3	8.6	[5.2, 12.4]
Bearing ball	93.2	100.0	96.2	98.9	100.0	98.5	[97.9, 99.1]
Bearing inner	100.0	100.0	100.0	100.0	100.0	100.0	[100.0, 100.0]
Bearing outer	5.3	67.3	9.2	9.4	30.7	8.3	[6.5, 10.0]
Chipped gear	4.2	99.2	48.7	42.4	90.9	30.6	[27.5, 33.6]
Missing tooth	100.0	100.0	100.0	100.0	100.0	100.0	[100.0, 100.0]

PCA achieves 100% detection for all faults at bag-50 and almost perfect detection for bag-20, but its FAR is unacceptably high for both bag sizes: 98.9% (bag-50) and 70.8% (bag-20). This contrasts sharply with the well-behaved per-hit results, which is a manifestation of the curse of dimensionality: even with 187 training bags, the 203-dimensional feature space remains too poorly sampled for reliable PCA reconstruction error estimation. Therefore, the three-method ensemble (IsoForest, KNN, LOF) is adopted for all subsequent analyses, excluding PCA from the detection ensemble.

At bag-50, chipped gear, which is almost undetectable at the per-hit level (7.2AUC values (Figure 7) are: bearing ball 0.98, bearing inner 1.00, bearing outer 0.68, chipped gear 0.95, and missing tooth 1.00.

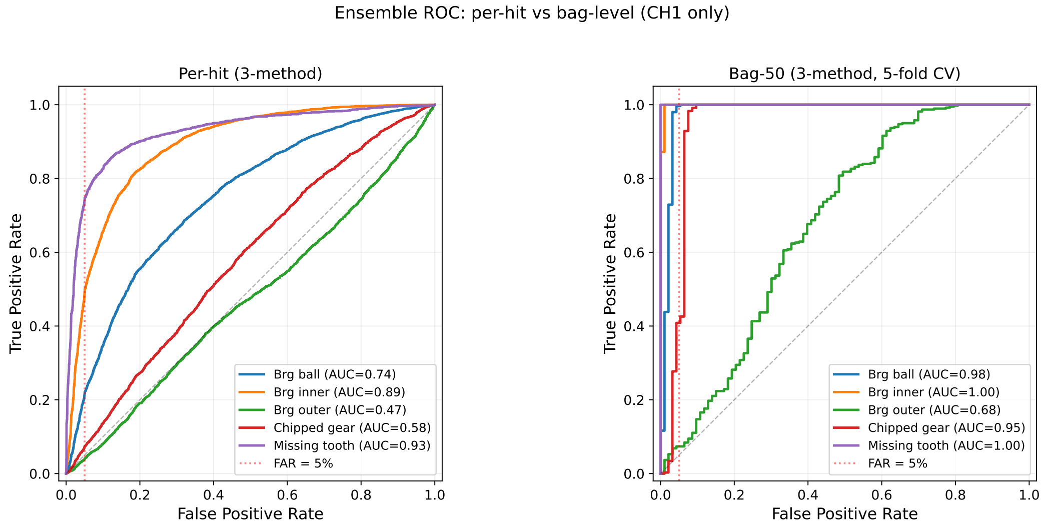

Figure 7. Three-method ensemble ROC curves. **Left:** per-hit. **Right:** bag-50, 5-fold CV. Bearing outer race (AUC = 0.68) is the only condition with modest score separation; chipped gear (AUC = 0.95) is now well separated at the bag level despite being near-random at the per-hit level.

3.3. Effect of PCA Preprocessing on Single-Bag Detection

Table 8 presents the effect of applying PCA preprocessing (90% variance threshold, fit on healthy training data only) on the three-method ensemble detection rates, for both bag sizes. For bag-50, PCA preprocessing reduces the dimensionality from 203 to 24 components; for bag-20, from 203 to 33–34 components.

Table 8. Effect of PCA preprocessing (90% variance threshold) on 3-method ensemble detection rates (%), with 95% bootstrap CI. Values: detection rate [CI].

	Bag-50		**Bag-20**	
Condition	**No PCA**	**PCA 90%**	**No PCA**	**PCA 90%**
Healthy (FAR)	16.1 [9.7, 24.7]	9.7 [4.3, 16.1]	8.6 [5.2, 12.0]	6.0 [3.0, 9.4]
Bearing ball	100.0	100.0	98.5 [97.9, 99.1]	89.1 [87.4, 90.6]
Bearing inner	100.0	100.0	100.0	100.0
Bearing outer	23.2 [18.9, 27.4]	11.8 [8.7, 15.3]	8.3 [6.5, 10.1]	6.8 [5.2, 8.4]
Chipped gear	100.0	100.0	30.6 [27.5, 33.6]	28.8 [25.8, 31.8]
Missing tooth	100.0	100.0	100.0	100.0
PCA components	—	24	—	33–34
Variance explained	—	90.0%	—	90.0–90.4%

For bag-50, PCA preprocessing reduces the single-bag FAR from 16.1% to 9.7%, which is a positive outcome. However, bearing outer race detection rate drops from 23.2% to

11.8%, the fault that most needed improvement becomes even harder to detect. All other fault detection rates remain at 100% and are unaffected. For bag-20, a similar pattern holds: FAR improves modestly (8.6% to 6.0%), but bearing ball drops from 98.5% to 89.1%, bearing outer from 8.3% to 6.8%, and chipped gear from 30.6% to 28.8%.

The net effect of PCA preprocessing is therefore ambiguous: the FAR reduction is real but comes at the cost of reduced detection rates for the already-difficult faults. The improvement is more pronounced at bag-50 because the ratio of training samples to feature dimensions is lower (74 bags in 203 dimensions), making the distance metrics more unstable, yet PCA does not recover detection for bearing outer. This suggests that the discriminative information for bearing outer race fault is distributed across the discarded low-variance components, and removing them hurts more than the noise reduction helps.

At the CUSUM level, PCA preprocessing also reduces alarm rates for the easily detected faults (bag-50 bearing ball: 100% without PCA vs. 62.7% with PCA; chipped gear: 54.3% vs. 48.6%). Therefore, for this particular setting, did not provide the expected benefits and were not furhter pursued, i.e., the full 203-dimensional feature space is used for all results reported in this paper.

3.4. Feature Discriminability

At the per-hit level, no feature exceeds AUC = 0.80 (best: band power 20–100 kHz, AUC = 0.75), and 13 of 24 features fall below 0.60. At the bag-50 level, 47 features exceed 0.80, but none exceeds 0.95 against the pooled fault class (maximum: 0.87). This confirms that distributional aggregation substantially increases feature discriminability, yet no single feature is sufficient on its own. On the other hand, the per-fault ROC curves in Figure 7 show the merit of using all of them through the one-class classification scheme.

3.5. CUSUM Sequential Detection

The CUSUM sequential detector is evaluated using a chronological 80/20 split of the healthy data: The results are presented for the recommended parameter setting (k = 0.5, $h = 5.0$), selected from the sensitivity analysis in conjunction with standard one-sided CUSUM design guidance. Tables 9 and 10 present the CUSUM results for bag-50 and bag-20 respectively.

Table 9. CUSUM detection results ($k = 0.5$, $h = 5.0$, bag-50, 80/20 split). Alarm rate is the fraction of bags triggering an alarm (alarm density over the monitored stream, not a per-stream detection probability).

Condition	Bags	Alarms	Rate (%)	Approx. Detection Time	Med. Gap	Pattern
Healthy (FAR)	19	0	0.0		–	–
Bearing ball	110	110	100.0	<2 s	1	Sustained
Bearing inner	103	103	100.0	<2 s	1	Sustained
Bearing outer	76	8	10.5	≈6–12 s	8	Intermittent
Chipped gear	70	38	54.3	≈2–4 s	2	Sustained
Missing tooth	36	36	100.0	<2 s	1	Sustained

All five fault conditions trigger CUSUM alarms at both bag sizes. The detection patterns at bag-50 are illustrated in Figure 8. Four faults produce *sustained* detection at bag-50: bearing ball (100% alarm rate), bearing inner and missing tooth (both 100%) show almost continuous alarming. The chipped gear, which was almost undetectable per-hit (DR equal to 7.2%) turns into a 54.3% alarm rate, showing that distributional aggregation combined with sequential accumulation can reveal fault signatures that are completely invisible in individual AE hits. Bearing outer race is the only condition producing *intermittent* detection at bag-50: it triggers its first alarm at bag 3 (≈6–12 s), with a 10.5% alarm rate and a median inter-alarm gap of 8 bags, yet the CUSUM statistic remains above zero for 90.8% of the monitoring period.

Table 10. CUSUM detection results ($k = 0.5$, $h = 5.0$, bag-20, 80/20 split). Alarm rate is the fraction of bags triggering an alarm (alarm density over the monitored stream, not a per-stream detection probability).

Condition	Bags	Alarms	Rate (%)	Approx. Detection Time	Med. Gap	Pattern
Healthy (FAR)	47	1	2.1		–	–
Bearing ball	275	181	65.8	<1 s	1	Sustained
Bearing inner	259	256	98.8	<1 s	1	Sustained
Bearing outer	191	9	4.7	≈4–7 s	15	Intermittent
Chipped gear	175	17	9.7	≈2–3 s	8	Intermittent
Missing tooth	91	91	100.0	<1 s	1	Sustained

3.6. CUSUM Sensitivity Analysis

Tables 11 and 12 present the three-method ensemble CUSUM alarm rates across nine parameter combinations for bag-50 and bag-20 respectively.

Table 11. CUSUM sensitivity analysis (bag-50, 3-method ensemble, 80/20 split). Values: alarm rate (%).

k	h	FAR (%)	Brg Ball	Brg Inner	Brg Outer	Chp. Gear	Miss. Tooth
0.25	3	5.3	100.0	100.0	19.7	90.0	100.0
0.25	5	0.0	100.0	100.0	13.2	54.3	100.0
0.25	8	0.0	67.3	100.0	10.5	40.0	100.0
0.50	3	0.0	100.0	100.0	14.5	80.0	100.0
0.50	**5**	**0.0**	**100.0**	**100.0**	**10.5**	**54.3**	**100.0**
0.50	8	0.0	67.3	100.0	6.6	37.1	100.0
1.00	3	0.0	100.0	100.0	7.9	67.1	100.0
1.00	5	0.0	95.5	100.0	6.6	48.6	100.0
1.00	8	0.0	66.4	100.0	5.3	32.9	100.0

Recommended configuration in bold.

Table 12. CUSUM sensitivity analysis (bag-20, 3-method ensemble, 80/20 split). Values: alarm rate (%). The FAR is 2.1% (1 alarm in 47 bags) across all configurations, reflecting the coarse FAR granularity at this sample size; parameter selection is therefore based on the bag-50 analysis.

k	h	FAR (%)	Brg Ball	Brg Inner	Brg Outer	Chp. Gear	Miss. Tooth
0.25	3	2.1	80.0	100.0	5.8	19.4	100.0
0.25	5	2.1	66.5	99.2	5.2	14.3	100.0
0.25	8	2.1	58.9	94.6	4.2	9.1	100.0
0.50	3	2.1	77.5	100.0	5.8	16.0	100.0
0.50	5	2.1	65.8	98.8	4.7	9.7	100.0
0.50	8	2.1	57.8	94.6	3.7	6.3	100.0
1.00	3	2.1	71.6	100.0	4.7	6.9	100.0
1.00	5	2.1	62.2	97.7	4.7	2.9	100.0
1.00	8	2.1	55.6	94.6	2.6	1.1	100.0

For bag-50, the recommended setting was selected according to three criteria: (i) zero observed alarms on the held-out healthy stream, (ii) consistency with standard one-sided

CUSUM design guidance for detecting an approximately 1σ upward shift, and (iii) selection of a balanced operating point rather than the most aggressive or most conservative configuration. Because several parameter combinations satisfy criterion (i), the pair $(k, h) = (0.5, 5.0)$ was therefore selected as a balanced choice, preserving reliable detection across all fault conditions.

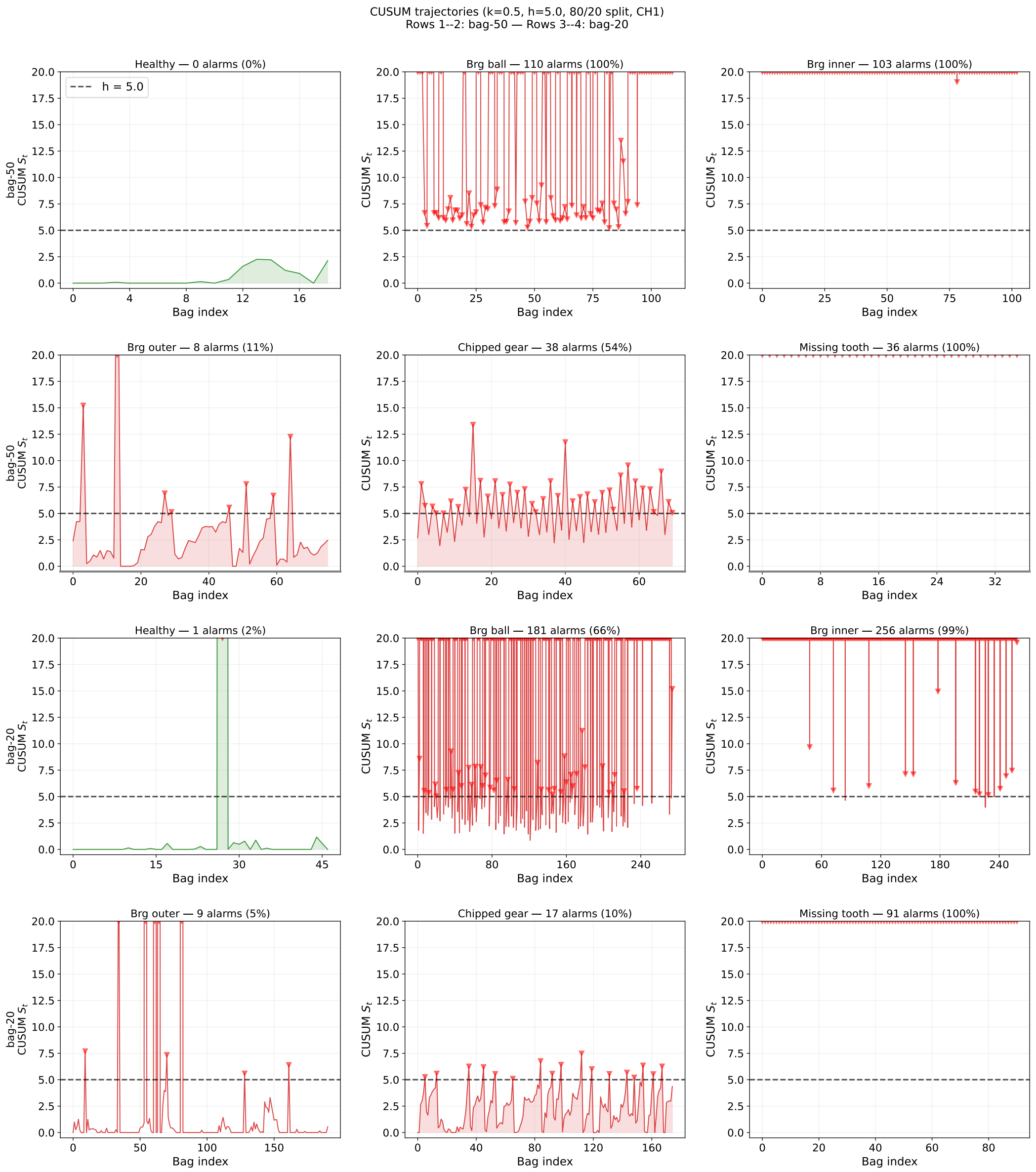

Figure 8. CUSUM trajectories ($k = 0.5$, $h = 5.0$, 80/20 split, CH1). Rows 1–2: bag-50. Rows 3–4: bag-20. The y-axis is truncated at $4h = 20$ for clarity; sustained faults routinely exceed this value between resets. The dashed line indicates the decision threshold $h = 5$. At bag-50, bearing ball, bearing inner, chipped gear, and missing tooth show sustained alarm patterns; bearing outer shows intermittent alarms with the statistic persistently above zero. At bag-20, chipped gear transitions to intermittent detection, reflecting the reduced distributional resolution of smaller bags.

4. Discussion

4.1. Evaluation Protocol and Sample-Size Considerations

The five-fold cross-validation with held-out FAR estimation revealed that single-bag FARs range from 7% to 20%, substantially above the targeted 5%. This arises because 75 training bags (bag-50) in a 203-dimensional space provide an insufficiently stable estimate of the healthy distribution tail. The bootstrap confidence intervals quantify this instability: the three-method ensemble FAR of 16.1% has a 95% CI of [9.7%, 23.7%]. For bearing outer race fault, the DR of 23.2% has a CI of [18.9%, 27.4%], confirming that single-bag thresholding cannot reliably distinguish this fault from the healthy baseline. These uncertainty estimates reinforce the necessity of using CUSUM sequential detection rather than static per-bag decision making.

4.2. The Role of CUSUM

CUSUM reframes anomaly detection by focusing on the accumulation of deviations to identify persistent distributional changes. Instead of assessing whether a single bag contains deviating signals, it determines whether a sustained shift in the underlying distribution has occurred. Rather than evaluating whether any individual bag exceeds a binary threshold, which is unreliable when the healthy and fault score distributions overlap, CUSUM accumulates small but consistent deviations over consecutive bags, triggering an alarm only when the evidence for a persistent change in the process has built up sufficiently. This distinction is critical for example for the bearing outer race, whose per-bag anomaly scores are only mildly elevated above chance (AUC = 0.68) and cannot be reliably separated from healthy variability in any single bag.

The chronological 80/20 evaluation reveals two distinct detection regimes. Four faults (bearing ball, bearing inner, chipped gear, and missing tooth) which produce sustained alarm patterns with alarm rates of 54–100% with a CUSUM statistic above zero for 100% of the monitoring period. The transformation of chipped gear from near-undetectable per-hit (7.2%) to sustainedly detected by CUSUM (54% alarm rate) is particularly noteworthy. It demonstrates that distributional aggregation combined with sequential accumulation can reveal fault signatures that are hidden in individual AE hits.

Bearing outer race is the only fault producing ”intermittent detection”: 10.5% alarm rate, with the CUSUM statistic above zero for 90.8% of the monitoring period. Even for this most challenging condition, the first alarm arrives at the third bag (≈6–12 s). The bag-size comparison further illustrates the trade-off between distributional resolution and training stability. Bag-50 provides stronger distributional summaries and more consistent CUSUM detection. Bag-20 provides more training data and a lower single-bag FAR (8.6% vs. 16.1%), but produces substantially weaker CUSUM detection for the more difficult faults.

4.3. Toward a Decision Support System

The proposed framework is intended as the first component of a decision support system for incipient fault detection. The CUSUM evidence index $E_t = S_t/h$ provides a continuous, interpretable measure of deviation from the healthy baseline. An E_t value approaching 1.0 signals accumulating evidence that a distributional shift has occurred, prompting the engineer to investigate further (human in the loop). The detection pattern contains two alarm types: intermittent for the bearing outer race fault, and sustained for the four others. This detection pattern can be mapped into different alert levels in a monitoring dashboard. Sustained alarms indicate the need for immediate investigation, while intermittent alarms suggest increased inspection intervals, or cross-referencing with other condition monitoring systems.

4.4. Limitations

The principal limitations are: (i) validation on a single laboratory test rig with seeded faults under fixed speed and load; (ii) a relatively small healthy dataset (93 bags at bag-50), of which only 19 bags are available for CUSUM evaluation under the 80/20 split, limiting the precision of the FAR estimate; (iii) fault conditions are evaluated independently rather than as a continuous mixed stream; (iv) the procedure was not tesed under varying operating conditions such as speed, load, and temperature; and (v) a systematic sensitivity analysis of the AE acquisition parameters (threshold, PDT, HDT) was not performed, as these were set following standard practice and preliminary testing.

4.5. Varying Operating Conditions and Future Directions

In real-world rotating machinery, varying loads and speeds create substantial distribution shifts between the training baseline and operational measurements. Under such conditions, the healthy score distribution would shift, potentially inflating the false alarm rate or masking genuine fault signatures. Operating condition normalisation—standardising AE features with respect to known operating parameters before bag construction—has been shown to reduce condition-induced variability in AE-based monitoring [14,15], and represents a promising direction for future work. Validating this approach would however require additional experiments across multiple speed and load conditions, which is beyond the scope of the current proof-of-concept study.

The current framework deliberately adopts traditional one-class methods (IsoForest, KNN, LOF), prioritising interpretability and minimal data requirements. Modern semi-supervised frameworks such as contrastive learning [33] could potentially learn richer feature representations than the 24 engineered parametric features used here, particularly for subtle faults such as bearing outer race. However, such approaches typically require labelled fault examples during training, which contradicts the one-class assumption that fault data are unavailable at training time. Unsupervised contrastive learning applied directly to AE hit waveforms is a promising direction for future investigation.

5. Conclusions

This work addressed fault detection in rotating machinery using hit-based AE data. The following conclusions are drawn from the laboratory validation:

1. Individual AE hits are insufficient for reliable fault discrimination. Per-hit detection rates were below 50% for four of five faults, with bearing outer race (AUC = 0.47) being effectively indistinguishable from healthy.
2. Distributional feature aggregation improves score separability (bag-50 ensemble AUC: 0.68–1.00). Four of five faults achieve 100% single-bag detection under five-fold cross-validation. However, bearing outer race remains at 23% single-bag detection, and the held-out FAR (16%) exceeds the nominal 5% target due to the limited healthy training sample relative to the 203-dimensional feature space.
3. Among the four classifiers evaluated, PCA behaves differently across the two representation levels, illustrating the curse of dimensionality. At the per-hit level (24 dimensions, 4664 healthy training samples), PCA produces a well-controlled false alarm rate of 5% and modest but reasonable detection rates. At the bag level (203 dimensions, 75 healthy training bags), the principal subspace becomes unreliable and PCA produces a false alarm rate of 99%, making it unsuitable for the ensemble despite achieving 100% fault detection. It is therefore excluded from the final ensemble.
4. PCA preprocessing (90% variance threshold, fit on healthy training data only) was evaluated as a dimensionality reduction step prior to the distance-based detectors. While it reduces the single-bag FAR (16.1% to 9.7% for bag-50), it simultaneously reduces bearing outer race detection from 23.2% to 11.8% and also reduces CUSUM alarm rates for the easily detected faults. Therefore, the full 203-dimensional feature space is therefore adopted, as the detectors implicitly exploit the low-dimensional structure of the data without explicit preprocessing.
5. The CUSUM sequential detector, applied to a three-method ensemble (IsoForest, KNN, LOF) and evaluated on a chronological 80/20 split of the healthy data, triggers alarms for all five fault conditions with FAR = 0% at the recommended setting ($k = 0.5$,

$h = 5.0$, bag-50). Four faults are detected with sustained alarm patterns (54–100% alarm rates). The bearing outer race fault, as the most challenging condition to detect, is identified through the CUSUM statistic (10.5% alarm rate), with the statistic remaining above zero for 90.8% of the monitoring period.

6. The CUSUM results are robust across all nine parameter combinations tested ($k \in \{0.25, 0.5, 1.0\}$, $h \in \{3, 5, 8\}$): all five faults trigger alarms in every configuration, and FAR = 0% in eight of nine combinations.
7. The bag-size comparison indicates bag-50 as the recommended configuration: it provides stronger distributional summaries and more consistent CUSUM detection than bag-20, at the cost of fewer training samples per fold for single-bag threshold calibration.

Future work should validate the approach on field data from operational machinery under varying conditions, investigate and create a richer fature bank from AE hit waveforms, and conduct a systematic sensitivity analysis of AE acquisition parameters across different machine types.

Author Contributions: Conceptualization, N.A. and G.G.; methodology, N.A. and G.G.; software, N.A.; validation, N.A. and G.G.; formal analysis, G.G.; investigation, N.A.; data curation, N.A. and G.G.; writing—original draft preparation, N.A. and G.G.; writing—review and editing, N.A., G.G. and V.K.; visualization, G.G.; supervision, G.G. and V.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been funded by the European Union, under Grant Agreement N° 101146936 TWINVEST. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or RIA. Neither the European Union nor the granting authority can be held responsible for them.

Data Availability Statement: The data presented in this study are not publicly available due to confidentiality restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AE	Acoustic Emission
ANN	Artificial Neural Network
AUC	Area Under the ROC Curve
CI	Confidence Interval
CM	Condition Monitoring
CUSUM	Cumulative Sum
DTS	Drivetrain Simulator
DR	Detection Rate
EDG	Edge Detection
EMD	Empirical Mode Decomposition
FAR	False Alarm Rate
HDT	Hit Definition Time
HLT	Hit Lockout Time
IsoForest	Isolation Forest
KNN	K-Nearest Neighbour
LOF	Local Outlier Factor
ML	Machine Learning
MSPS	Mega Samples Per Second
PDT	Peak Definition Time
PCA	Principal Component Analysis

PSD	Power Spectral Density
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine

References

1. Papaelias, M.; Márquez, F. Wind Turbine Inspection and Condition Monitoring. In *Non-Destructive Testing and Condition Monitoring Techniques for Renewable Energy Industrial Assets*; Butterworth-Heinemann: Waltham, MA, USA, 2020; pp. 19–29.
2. Jantara, V., Jr.; Papaelias, M. Wind Turbine Gearboxes: Failures, Surface Treatments and Condition Monitoring. In *Non-Destructive Testing and Condition Monitoring Techniques for Renewable Energy Industrial Assets*; Butterworth-Heinemann: Waltham, MA, USA, 2020; pp. 69–90.
3. Pollock, A. Acoustic Emission Inspection. In *Nondestructive Evaluation of Materials*; ASM International: Almere, The Netherlands, 2018; pp. 360–379.
4. Niknam, S.; Thomas, T.; Hines, J.; Sawhney, R. Analysis of Acoustic Emission Data for Bearings Subject to Unbalance. *Int. J. Progn. Health Manag.* **2013**, *4*, 80–89. [CrossRef]
5. Bote-Garcia, J.; Mokhtari, N.; Gühmann, C. Wear Monitoring of Journal Bearings with Acoustic Emission Under Different Operating Conditions. *PHM Soc. Eur. Conf.* **2020**, *5*, 8. [CrossRef]
6. Hemmati, F.; Alqaradawi, M.; Gadala, M. Rolling Element Bearing Fault Diagnostics Using Acoustic Emission Technique and Advanced Signal Processing. *Proc. Inst. Mech. Eng. Part J J. Eng. Tribol.* **2016**, *230*, 64–77. [CrossRef]
7. Leaman, F. A Review on Acoustic Emissions of Gear Transmissions: Source, Influencing Parameters, Applications and Modeling. *J. Vib. Eng. Technol.* **2024**, *12*, 7835–7846. [CrossRef]
8. Amarnath, M.; Praveen Krishna, I. Empirical Mode Decomposition of Acoustic Signals for Diagnosis of Faults in Gears and Rolling Element Bearings. *IET Sci. Meas. Technol.* **2012**, *6*, 279–287. [CrossRef]
9. Qu, Y.; He, D.; Yoon, J.; Van Hecke, B.; Bechhoefer, E.; Zhu, J. Gearbox Tooth Cut Fault Diagnostics Using Acoustic Emission and Vibration Sensors—A Comparative Study. *Sensors* **2014**, *14*, 1372–1393. [CrossRef]
10. Guo, Y.; Thomson, A.; Bergua, R.; Bankestrom, O.; Erskine, J.; Keller, J. *Acoustic Emission Measurement of a Wind Turbine Main Bearing*; Technical Report; National Renewable Energy Laboratory (NREL): Golden, CO, USA, 2022.
11. Elforjani, M.; Mba, D. Accelerated Natural Fault Diagnosis in Slow Speed Bearings with Acoustic Emission. *Eng. Fract. Mech.* **2010**, *77*, 112–127. [CrossRef]
12. He, Y.; Zhang, X.; Friswell, M. Defect Diagnosis for Rolling Element Bearings Using Acoustic Emission. *J. Vib. Acoust.* **2009**, *131*, 061012. [CrossRef]
13. Tan, C.; Irving, P.; Mba, D. A Comparative Experimental Study on the Diagnostic and Prognostic Capabilities of Acoustics Emission, Vibration and Spectrometric Oil Analysis for Spur Gears. *Mech. Syst. Signal Process.* **2007**, *21*, 208–233. [CrossRef]
14. Leaman, F.; Niedringhaus, C.; Hinderer, S.; Nienhaus, K. Evaluation of acoustic emission burst detection methods in a gearbox under different operating conditions. *J. Vib. Control* **2019**, *25*, 895–906. [CrossRef]
15. Leaman, F.; Vicuña, C.; Baltes, R.; Clausen, E. Pattern recognition approach for acoustic emission burst detection in a gearbox under different operating conditions. *J. Vib. Control* **2019**, *25*, 2295–2304. [CrossRef]
16. König, F.; Jacobs, G.; Stratmann, A.; Cornel, D. Fault Detection for Sliding Bearings Using Acoustic Emission Signals and Machine Learning Methods. In *Proceedings of the IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2021; Volume 1097, p. 012013.
17. Jiang, P.; Xia, J.; Li, W.; Xu, C.; Sun, W. Innovative Bearing Fault Diagnosis Method: Combining Swin Transformer Deep Learning and Acoustic Emission Technology. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part B Mech. Eng.* **2025**, *11*, 011102. [CrossRef]
18. Omoregbee, H.; Heyns, P. Fault Classification of Low-Speed Bearings Based on Support Vector Machine for Regression and Genetic Algorithms Using Acoustic Emission. *J. Vib. Eng. Technol.* **2019**, *7*, 455–464. [CrossRef]
19. Georgoulas, G.; Loutas, T.; Stylios, C.; Kostopoulos, V. Bearing Fault Detection Based on Hybrid Ensemble Detector and Empirical Mode Decomposition. *Mech. Syst. Signal Process.* **2013**, *41*, 510–525. [CrossRef]
20. Karvelis, P.; Georgoulas, G.; Kappatos, V.; Stylios, C. Deep Machine Learning for Structural Health Monitoring on Ship Hulls Using Acoustic Emission Method. *Ships Offshore Struct.* **2021**, *16*, 440–448. [CrossRef]
21. Mba, D.; Rao, R. Development of Acoustic Emission Technology for Condition Monitoring and Diagnosis of Rotating Machines: Bearings, Pumps, Gearboxes, Engines, and Rotating Structures. *Shock Vib. Dig.* **2006**, *38*, 3–16. [CrossRef]
22. Carbonneau, M.; Cheplygina, V.; Granger, E.; Gagnon, G. Multiple Instance Learning: A Survey of Problem Characteristics and Applications. *Pattern Recognit.* **2018**, *77*, 329–353. [CrossRef]
23. Page, E. Continuous Inspection Schemes. *Biometrika* **1954**, *41*, 100–115. [CrossRef]
24. Liu, F.; Ting, K.; Zhou, Z.H. Isolation Forest. In *Proceedings of the IEEE International Conference on Data Mining (ICDM), Pisa, Italy, 15–19 December 2008*; IEEE: Piscataway, NJ, USA, 2008; pp. 413–422.

25. Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]
26. Breunig, M.; Kriegel, H.P.; Ng, R.; Sander, J. LOF: Identifying Density-Based Local Outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, TX, USA, 15–18 May 2000*; ACM: New York, NY, USA, 2000; pp. 93–104.
27. Randall, R.; Antoni, J. Rolling Element Bearing Diagnostics—A Tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520. [CrossRef]
28. Randall, R. *Vibration-Based Condition Monitoring: Industrial, Automotive and Aerospace Applications*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2021.
29. *E1316*; Standard Terminology for Nondestructive Examinations. ASTM International: West Conshohocken, PA, USA, 2021.
30. Basseville, M.; Nikiforov, I. *Detection of Abrupt Changes: Theory and Application*; Prentice Hall: Englewood Cliffs, NJ, USA, 1993.
31. Montgomery, D. *Statistical Quality Control*, 6th ed.; Wiley: Hoboken, NJ, USA, 2009.
32. Hawkins, D.; Olwell, D. *Cumulative Sum Charts and Charting for Quality Improvement*; Springer: Berlin/Heidelberg, Germany, 1998.
33. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of the 37th International Conference on Machine Learning, PMLR, Virtual, 13–18 July 2020; pp. 1597–1607.