

An End-to-End Predictive Maintenance System for Autonomous Inspection based on Health Index

Ho-Min Park*, Dongkoo Shon*, Yeonggwang Oh*, Tae Hyun Yoon*,

Woo-Sung Jung*, Jeong-Ho Park**, Dae Seung Yoo*

*Ulsan Intelligent Convergence Research Section, Electronics and Telecommunications Research Institute,
Ulsan, Republic of Korea

**Field Robotics Research Section, Electronics and Telecommunications Research Institute,
Daejeon, Republic of Korea

{hominpark, sdk, oykyoung, thyoon0820, woosung, parkjh, ooseyds}@etri.re.kr

Abstract— This paper presents WFSPM, a transferable end-to-end predictive maintenance system for autonomous inspection centered on a Mahalanobis-distance (MD) health index (HI). We demonstrate the approach on a percussion-acoustics case study (stator wedge fastener strength) while keeping the architecture sensor- and asset-agnostic. The standards-compliant pipeline (data flow, features, decision rules, UI/reporting) includes adapter endpoints for cell controllers and line dispatchers and maps the MD-HI to a 0–100 scale for field interpretability. WFSPM shows strong static performance and robustness to injected noise, supporting in-cell mobile-manipulator rechecks without retuning thresholds. Scenario-based prognostics and Health Management (PHM) key performance indicators (KPIs) and lead-time analyses quantify an actionable cycle of early warning, action and recovery, with HI/dual-threshold events mappable to manufacturing-cell KPIs. Overall, WFSPM is explainable, robust, and policy-tunable, and transferable to mixed-model manufacturing cells for autonomous inspection.

Keywords— Autonomous Inspection, Predictive Maintenance, Health Index, Acoustic Signal Processing

I. INTRODUCTION

Large rotating generators still inspect stator wedge fastener strength mainly by percussion- and auscultation-based NDT, which makes results subjective, sensitive to environmental noise, and hard to accumulate as traceable history [1]. End-to-end (E2E) diagnostic/PHM software designed for field operation should be aligned with standards in data processing, presentation, and reporting [2–4]. Acoustic-signal-based diagnostics have recently been applied to bolt loosening, joints, and structural defects, spanning microphones, acoustic emission, and deep learning [5–8]. On the representation side, mel-spectral features (e.g., mel spectrogram, MFCCs) have repeatedly shown effectiveness for industrial anomaly detection, and recent surveys systematize the landscape of acoustics-ML integration [6, 9]. In parallel, autonomous inspection in manufacturing cells increasingly demands

standards-compliant, sensor-agnostic PHM stacks that can transfer across assets without redesign.

Beyond mere accuracy, the temporal validity of a HI is evaluated by monotonicity, trendability, and prognosability, and established methods assess HI with these metrics [10]. We propose Wedge Fastener Strength Predictive Maintenance (WFSPM) as a transferable E2E PHM stack for autonomous inspection and instantiate it on percussion-acoustic inspection of generator wedges as a case study. WFSPM forms a single pipeline that includes optional denoising, mel-spectrogram/spectral/statistical features, MD-based HI against the normal distribution, two thresholds (τ_1, τ_2) for three-class labeling (Normal, Caution, Fault), and UI for live/comparative/trend views with auto-reporting. The architecture remains sensor- and asset-agnostic, and an adapter layer (e.g., OPC UA/MQTT/REST) exposes alarms and metadata to cell controllers/line dispatchers without retuning thresholds. Along with static three-class metrics, we reproducibly compute monotonicity, trendability, prognosability, and lead time on synthetic timelines to compensate for the lack of long continuous field logs.

The rest of this paper is organized as follows. Section II reviews related work; Section III outlines the WFSPM system; Section IV details signal processing, HI, and rules; Section V reports static classification, scenario-based PHM KPIs, and robustness; Section VI concludes and discusses future work.

II. RELATED WORK

This section surveys prior work directly related to predictive maintenance of wedge fastener strength along three axes: (i) condition monitoring and diagnostics (CM&D) standards and operational case studies; (ii) PHM-oriented HI evaluation metrics; and (iii) acoustic-signal-based detection of fastener loosening and structural defects. From this review, we derive the rationale for the WFSPM architecture, the design of the MD-based HI and thresholds, and the operating choices adopted in this paper, with an emphasis on transferability to autonomous inspection in manufacturing cells.

A. CM&D Standards and Operational Case Studies

CM&D software should integrate data acquisition, processing, presentation, and reporting in a consistent structure, and it is recommended to refer to guidelines such as ISO 13374 and ISO 13379-1 [2–4]. Recent requirement-driven designs reaffirm that ISO-13374-type standards remain a valid reference axis for predictive/preventive/prescriptive (PMx) systems and systematize functional and presentation requirements [11]. Furthermore, field-oriented E2E implementations have been reported, including a cloud-edge-device collaborative framework for smart-grid inspection [12] and a digital-twin - based CBM application on real axle bearings [13]. These studies lay out an operational software loop — sensing → preprocessing/model → decision → visualization/reporting — and share concrete implementation points for dashboards, reports, and data pipelines. In this context, autonomous inspection in manufacturing cells benefits from standards-driven stacks that expose alarms/metadata to controllers and dispatchers; accordingly, the WFSPM architecture in this paper reflects these principles and implementation practices, with adapter endpoints suitable for cell integration, and the UI/report follow an information hierarchy akin to the display guidelines of ISO 13374-4.

B. PHM-Oriented HI Evaluation Metrics and Trends

In prognostics and health management, the temporal validity of a HI is evaluated by monotonicity, trendability, and prognosability [10]. Recent studies have reorganized this classical triad into a systematic framework, clarifying the temporal properties an HI should possess and the procedures to verify them [14]. Furthermore, some works formulate HI design itself as a multi-objective, metric-driven optimization problem, e.g., maximizing several metrics simultaneously [15]. This perspective reduces over-reliance on any single metric and allows designing HI that are more trade-off-aware for specific operational goals.

In data-driven approaches, empirical studies continue to select and validate candidate HI with the classical metrics and then connect the chosen HI to remaining useful life (RUL) estimation. For bearings, for example, researchers construct/evaluate HI and perform RUL regression [16]; others employ deep autoencoders and time-series models to build HI in an unsupervised way from raw signals and report significant gains in monotonicity, trendability, and prognosability over traditional baselines [17].

More recent reviews recommend incorporating model uncertainty, robustness, and interpretability into both HI design and evaluation [6, 18], arguing that one must assess whether performance sustains under environmental shifts, how model/threshold uncertainty propagates to alarm timing or RUL dispersion, and whether field engineers can interpret the rationale behind decisions. Accordingly, this paper computes HI via MD with respect to the normal data distribution, and reports monotonicity, trendability, and prognosability on synthetic scenarios. When deployed in manufacturing cells, these HI properties enable policy-tunable alarms (Caution/Fault) that can be mapped to OEE-linked KPIs, and, when needed, we

adopt cost-sensitive operating choices such as weighted thresholds and a stage-2 decision in the caution band.

C. Acoustic Signal-based Detection of Structural Defects

Acoustic and acoustic-emission (AE) approaches have been actively adopted for fastener loosening and structural defect detection. Representative examples include a framework that combines AE with deep learning for automatic assessment of bolt tightness [7], the use of acoustic features as inputs to a GA-SVM model to identify loosening [5], a two-stream CNN with hybrid dimensions to classify joint states [19], and a method for detecting micro-looseness via deep learning [20].

More recently, fusion of visual and acoustic modalities has gained traction: visual-based detection of loosening structures [17] and an attempt to judge generator slot-wedge tightness from acoustic signals using an improved KNN model [21]. This naturally leads to deriving a single health index (HI) from fused information or acoustic-derived features. For example, a study applied Euclidean distance to a normal reference as an online monitoring HI [22], illustrating simple and explainable distance-based schemes. In parallel with this line of work, we extract statistical/acoustic features from percussive signals and employ an MD-HI with dual thresholds to judge wedge fastener strength. Among diverse methods, our choice emphasizes explainability and operational simplicity, and the standards-compliant UI/reporting facilitate integration with autonomous re-inspection (e.g., mobile manipulators) in manufacturing cells.

III. PROPOSED WFSPM SYSTEM

This section presents the overall architecture of WFSPM from an implementation-and-operation viewpoint, translating the standards, HI metrics, and acoustic diagnostics summarized in Section II into a working system. We outline the end-to-end workflow—from field acquisition and optional denoising to conditioning, feature engineering, MD-HI, dual-threshold decisions, and UI/reporting—and specify inputs/outputs at each block so that the pipeline is reproducible and field-deployable. In addition, we highlight adapter endpoints that enable integration with manufacturing-cell controllers and line dispatchers for autonomous inspection.

A. Design Principles and Overall Workflow

WFSPM is designed as a standards-compliant, field-operable software system, accepting the condition-monitoring and diagnostics standards summarized in Section II as structural constraints. Percussive data are acquired in the field, and from the moment of acquisition, metadata—site, slot ID, wedge ID, hit point ID, and timestamp—are linked under a consistent key. To support transferability to manufacturing cells, robot/cell-side sensing is treated as a pluggable source (acoustics/vision/vibration), while the downstream HI/thresholding and reporting remain unchanged. Figure 1 illustrates the proposed WFSPM architecture, showing both the flow of acoustic data and the operator's workflow through the system.

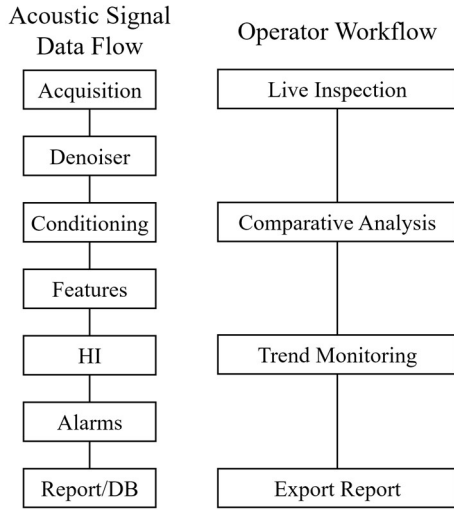


Figure 1. The proposed WFSPM system architecture

The pipeline—Acquisition → Denoiser → Conditioning → Features → HI → Alarms → Report/DB—implements ISO 13374 recommendations for data processing/presentation/communication and ISO 13379-1 procedures for diagnostic interpretation, with documented I/O specifications for each block. In addition, we adopt dashboard, auto-reporting, and DB-ingestion patterns observed in recent end-to-end implementations so that field users can accumulate diagnostic records as repeatable operational practice. Adapter endpoints (e.g., OPC UA/MQTT/REST) expose HI and alarm events (Caution/Fault) to cell controllers and line-level dispatchers without retuning thresholds, enabling autonomous re-inspection and policy-tunable actions.

B. Operating Context and Data Handling

Percussive acoustic signals combine repeatability with sensitivity to the environment. WFSPM supports multiple hit points on a single wedge and manages data in a three-layer hierarchy of sample–session–asset. For labeling, we use the distribution collected in our prior work [8]—320 Normal, 320 Caution, and 440 Fault samples, totaling 1,080—and adopt a synthetic-timeline protocol to evaluate prognostic KPIs such as monotonicity, trendability, prognosability, and lead time, compensating for the lack of long continuous field logs. Acquired percussive signals are stored with metadata in a database for reproducibility. Such data standardization directly supports the standards-based operability emphasized in system perspective of Section II. For cell integration, alarms are serialized with HI, τ_1 , τ_2 , timestamps, and context fields in an event schema suitable for dispatcher/MES consumption and OEE/KPI accounting.

C. Signal Pathway and the HI/Decision Logic

This section explains signal pathway and decision logic of WFSPM. Considering field noise and operator variability, the pipeline is organized as Denoiser → Conditioning → Features → MD-HI → dual thresholds, in order to reduce variance while keeping interpretability. The metric-level HI design is sensor-

agnostic; replacing the acquisition modality does not alter the MD-HI or dual-threshold policy. Table 1 summarizes how each WFSPM module maps to the ISO 13374/13379 categories.

TABLE 1. MAPPING OF WFSPM MODULES TO STANDARDS

Module	ISO 13374/13379 Category	The WFSPM System
Acquisition	[2] §5.2, §5.3	Synchronized trigger/sensor capture with metadata
Denoiser, Conditioning	[2] §5.3; [3] cl.5-6	Filtering, alignment, length normalization, RMS scaling
Features	[2] §5.4; [4] §6	Mel spectrogram, spectrum, statistics, zmeta management
HI, Scaling	[2] §5.5; [4] §7	MD-HI using μ/Σ of the normal reference; 0–100 health-scale mapping
Decision, Alarms	[2] §5.6–§5.7, §6; [4] §8	Dual thresholds τ_1/τ_2 ; cost-sensitive operating point
Report	[2] §6; [3] cl.5-6	One-click PDF report with results/plots/thresholds and operational metadata
DB	[2] §5.1; [3] cl.4-5	Keying scheme and artifact schema

1) Acquisition & Denoising: Percussive acquisition must manage field realism and repeatability together. WFSPM records metadata including slot, wedge, hit point IDs, and timestamp, and treats “one wedge with multiple hit points” as the basic unit. Immediately after capture, a denoiser is applied to suppress environmental noise. The module prioritizes preserving the impact portion of the waveform and prevents over-smoothing by constraining gain and bandwidth. To verify robustness, Section IV reports two modes—without/with denoiser—so that noise-reduction benefits and inference-latency trade-offs are presented side by side. In cell settings, the same pathway applies to other modalities (e.g., vision/vibration) with the HI/threshold stack unchanged.

2) Conditioning: Conditioning aims at variance reduction and reproducibility. After DC removal, we apply a band-pass filter to the information-dense band of the percussive signal (e.g., 0.3–8 kHz), align to a reference impact point (energy rise/envelope peak), normalize the analysis window length, and perform RMS scaling to control input-level spread. Optional clipping guards and winsorization suppress outliers. All parameters are written to artifact headers so that within-session and cross-session comparisons are reproducible. This step raises scale invariance of downstream mel/spectral features and directly stabilizes the MD-HI distribution.

3) Features: Grounded in Section II-C, WFSPM mixes mel-spectral, spectral-shape, and statistical features. Log-mel energies capture bandwise redistribution of impact energy;

spectral centroid/roll-off/flatness quantify the relative balance of tonal/noise components and high-frequency tails. High-band energy, kurtosis, and ZCR reflect increased impulsiveness with defect progression. Every feature is standardized by $zmeta$ (μ , σ), and the feature rationale is summarized as inference text for module documentation and the report. This explicit, low-dimensional set supports the explainability of the MD-HI.

4) HI: We estimate the mean and covariance of the normal class and compute MD as in equation (1), then linearly map it to a 0–100 health scale for operational interpretability. The parameters used in computation, together with scaling anchors, are stored so that the setup can be ported to other sites. MD-HI is simple, explainable, and has prior applications in online monitoring [22]. Combined with threshold rules, it yields three-grade decisions, and we cross-validate temporal validity by monotonicity, trendability, prognosability, and lead time on synthetic scenarios.

$$d_M(X) = \sqrt{(X - \mu)^T \Sigma^{-1} (X - \mu)} \quad (1)$$

5) Decision & Alarms: Two thresholds τ_1 and τ_2 define the operating point: τ_1 for normal rejection (sensitivity) and τ_2 for fault recognition (specificity), with $\tau_2 > \tau_1$. Algorithm 1 selects them via static optimization (e.g., macro-F1 maximization) with optional lower bounds on fault recall; cost-sensitive operating choices are permitted when missed-fault cost is high. Outputs are Normal/Caution/Fault; Caution acts as a secondary inspection queue that absorbs uncertainty and the cost of false alarms. All thresholds, operating points, and versions are recorded in the report for audit. Algorithm 2 defines online labeling of percussive signals during inspection. In cell deployment, emitted Caution/Fault events include HI and τ_1/τ_2 to trigger line dispatch or mobile-manipulator re-inspection without retuning thresholds.

D. Application Layer: Multi-Resolution UI and Reporting

Field operators move across slot–wedge–hit point while making decisions. To support this workflow, the proposed WFSPM UI provides three complementary views as in Figure 2. The live inspection view presents, with low latency, the current HI of hit point, class label, and signal summaries (cropped waveform/spectrum); when an alarm occurs, markers and operator recommendations appear immediately. The comparative inspection view visualizes differences against a chosen reference (equipment baseline, previous session, etc.) using radar/bar charts, and places past sessions of the same wedge side by side to aid operational judgment. The trend view plots the run–HI curve with τ_1/τ_2 threshold lines and annotates maintenance/replacement timestamps as evidence for subsequent analysis. From any screen, one action generates a report; layouts, colors, and labels follow standard display recommendations to keep the presentation consistent. For manufacturing-cell integration, the UI/report layer also serializes event payloads for controllers/dispatchers (HI, τ_1/τ_2 , label, timestamps, and context), enabling KPI/OEE accounting. The resulting PDF report includes summary tables, representative plots, thresholds (τ_1/τ_2), and operational

metadata (site, time, operator, software version), ensuring auditability.

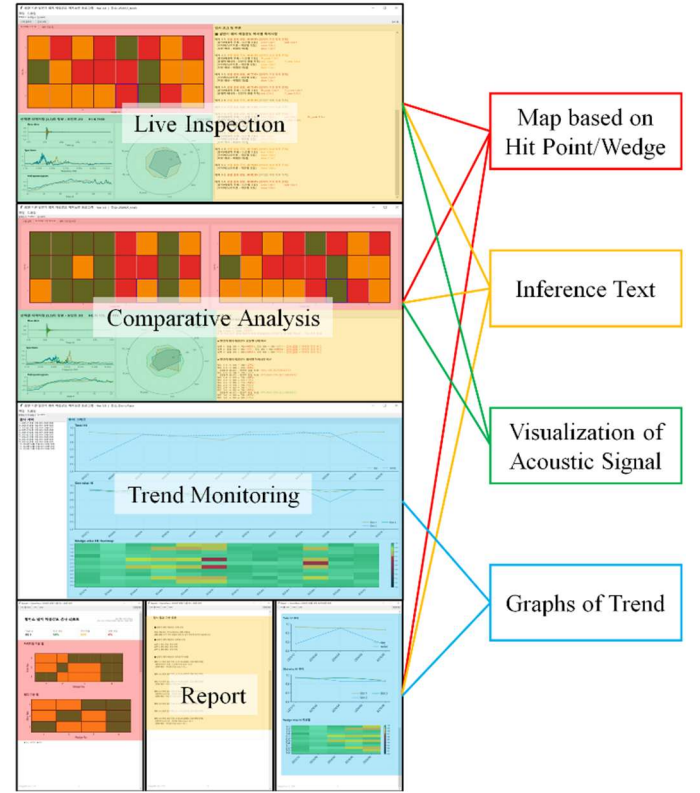


Figure 2. The proposed WFSPM system UI and functional modules

E. Figure Captions and Artifacts

WFSPM outputs are intentionally simplified into two artifact types for auditability and portability. First, recorded_data stores the raw capture (or denoised output) and key intermediates in the session folder, including alignment/normalization parameters, the feature set and its standardization metadata ($zmeta$), and the HI scaling metadata ($himeta$). Second, the report is a PDF that packages class decisions and HI/alarms, representative waveform/spectrum, comparative/trend plots, thresholds (τ_1/τ_2), and operational metadata (version, scenario, seed). The database operates in a standalone, local-only mode by design, considering field deployment, security, and network constraints; this enables offline continuous operation and single-PC auditing. Latency varies with hardware and denoiser usage, but the impact $\rightarrow$ features $\rightarrow$ HI $\rightarrow$ label loop is optimized to update within the UI frame budget. Artifact metadata guarantee reproducibility and site transferability; Section IV reports robustness and latency trade-offs in context. When transferred to manufacturing cells, the same artifacts serve as the event/logging source for line dispatch and OEE/KPI dashboards.

IV. METHODS: SIGNAL PROCESSING, HI, RULES

A. Signal Conditioning

Field percussion signals are sensitive to background noise and operator variability, so we standardize a lightweight conditioning stage to reduce feature variance. For an input $x[t]$, we apply: (i) DC removal; (ii) band-pass filtering (e.g., 0.3–8 kHz); (iii) impact alignment to a reference point (energy rise/envelope peak); (iv) length normalization with a fixed analysis window; and (v) RMS scaling. All parameters (passband, window length, reference definition) are written into the recorded_data metadata to ensure cross-session reproducibility. When site noise is high, the optional denoiser of Section III-C is inserted immediately before this stage; Section V reports results with and without denoising to interpret robustness. For transfer to manufacturing cells, the same conditioning interface can front-end alternative modalities (vision/vibration), while the downstream HI and thresholding remain unchanged.

B. Feature Engineering

The proposed WFSPM secures explainability and portability with an explicit, low-dimensional feature set. It composes by combining (i) mel-band energies, (ii) spectral descriptors: centroid, roll-off, flatness, high-band ratio, and (iii) statistical/temporal: RMS, ZCR, kurtosis, etc. Each dimension is standardized by the normal training set (μ_i, σ_i) as ($z_i = (x_i - \mu_i)/\sigma_i$), and this scaling bundle is stored as *zmeta*. The validity of mel spectrogram/spectrum/statistical features has been widely reported in industrial acoustics/AE literature [5–9], and we confirmed that they respond sensitively to the separation between Caution and Fault in the percussion signals of this task. When instantiated with other sensors in manufacturing cells, the feature block acts as a plug-in (e.g., image- or vibration-derived descriptors) feeding the same HI stack without policy changes. Table 2 lists the features used in WFSPM.

TABLE 2. WFSPM FEATURES FOR PERCUSSIVE ACOUSTIC SIGNAL

No.	Feature	Description
1	Mel spectrogram	Log-mel time–frequency map of the hit points, highlights resonances and high-frequency content
2	Frame spectral contrast	Per-frame peak-normalized mel contrast ($\max_{\text{mel}} - \text{mel}$), emphasizes narrow-band peaks/transients
3	Band-average energy	Mean mel energy per frame, wideband intensity trace of the hit
4	FFT spectrum	Magnitude spectrum over the analysis window, linear-frequency energy distribution
5	Mean	Time-domain average amplitude (DC offset check)
6	RMS	Root-mean-square amplitude, overall signal power
7	Peak	Maximum absolute amplitude, impact severity indicator
8	Crest factor	Peak/RMS ratio, impulsiveness of the percussive signal

9	Skewness	Third central moment, waveform asymmetry
10	Kurtosis	Fourth standardized moment, degree of peakedness/impulsiveness
11	FFT peak	Maximum spectral magnitude, dominant frequency component
12	FFT RMS	RMS of the magnitude spectrum, overall spectral power
13	FFT crest factor	Spectral peak/RMS ratio, narrow-band dominance
14	Normal probability(AI)	Classifier posterior for the “Normal” class (0–1)

C. MD-HI

We learn the mean (μ) and covariance (Σ) of features from the normal reference data and compute the MD as in equation (1). For operational interpretability, the MD is linearly mapped $[p_1, p_{99}]$ to a 0–100 health scale using predefined anchors, which set the upper and lower bounds of HI. MD-HI is structurally simple, explainable, and has prior applications in online condition monitoring [22]. Because MD operates at the metric level, the HI is sensor- and asset-agnostic, which eases transfer to autonomous inspection in mixed-model manufacturing cells. To align with the monotonicity, trendability, and prognosability framework summarized in Section II [10, 14–18], we choose features and scaling with an emphasis on temporal consistency and compactness near the alarm region.

D. Thresholds and Decision Rules

WFSPM uses two thresholds, τ_1 and τ_2 , to define the operating point for three-grade decisions. τ_1 primarily controls normal rejection (sensitivity), and τ_2 governs fault recognition (specificity), with $\tau_2 > \tau_1$. The thresholds are selected offline by Algorithm 1 via static optimization (e.g., macro-F1 maximization) with optional constraints on minimum fault recall and cost-sensitive weighting when missed faults are costly. Outputs are Normal/Caution/Fault. In the caution band, an optional stage-2 lightweight classifier (e.g., LDA or logistic regression) can be applied to raise precision while limiting recall loss. All thresholds, operating points, and software versions are written to the report metadata for auditability. Algorithm 2 specifies the online inference procedure during inspection. In manufacturing cells, emitted Caution/Fault events carry HI, τ_1 , τ_2 , timestamps, and context for line dispatch or mobile-manipulator reinspection without retuning thresholds; these events can be aggregated to cell-level KPI/OEE.

Algorithm 1. Offline selection of τ_1, τ_2

Inputs

- $D = \{(x_i, y_i)\}$: labelled samples ($y \in \{N, C, F\}$)
- $zmeta = \{\mu_i, \sigma_i\}$ of each features
- $himeta = \{\mu, \Sigma^{-1}, p_1, p_{99}\}$ for MD-HI
- grid $\mathcal{G}$: candidate pairs (τ_1, τ_2) with $\tau_1 < \tau_2$

- constraint R_F^{min} (min fault recall), weights $w = (w_N, w_C, w_F)$ (optional)

Outputs

- optimized thresholds (τ_1^*, τ_2^*)

Procedure

1. Standardize with $zmeta \rightarrow \tilde{x}$, compute $HI(\tilde{x})$ using $himeta$
2. For every $(\tau_1, \tau_2) \in \mathcal{G}$:
 - 2.1 Predict $\hat{y}$ by rule in Algorithm 2 (without stage-2)
 - 2.2 Compute per-class *precision*, *recall*, *F1*, *macroF1*
 - 2.3 If $R_F < R_F^{min}$, discard this pair
3. Select $(\tau_1^*, \tau_2^*) = \argmax_{(\tau_1, \tau_2)} macroF1$
4. Store τ_1^*, τ_2^* in report metadata

Algorithm 2. Online inference on inspection

Inputs

- pre-denoised signal $x[t]$
- $zmeta, himeta, \tau_1, \tau_2$
- Stage-2 classifier $g(\cdot)$ for caution band (e.g., LDA, logistic) (optional)

Outputs

- HI score and AI label ($HI, \hat{y} \in \{N, C, F\}$)

Procedure

1. Conditioning (DC removal, band-pass, alignment, RMS scaling)
2. Compute features on Table 2 to vector x
3. x to standardize $(x - \mu)/\sigma \rightarrow \tilde{x}$ using $zmeta$
4. $\tilde{x}$ to $scale(d_M(\tilde{x}; \mu, \Sigma^{-1})) \rightarrow HI$ using $himeta$
5. Primary rule:
 - If $HI < \tau_1 \rightarrow \hat{y} = N$
 - else if $HI \geq \tau_2 \rightarrow \hat{y} = F$
 - else go to step 6 (caution band)
6. Stage-2 in caution band (optional)
 - $g(\tilde{x}) \rightarrow \hat{y}$ with labels restricted to $\{C, F\}$ or $\{N, C\}$
 - If g is not provided $\rightarrow \hat{y} = C$

Note: In cell deployment, step 6 can trigger an autonomous reinspection routine via the line dispatcher when $\hat{y} = C$.

E. Scenario-based PHM KPIs

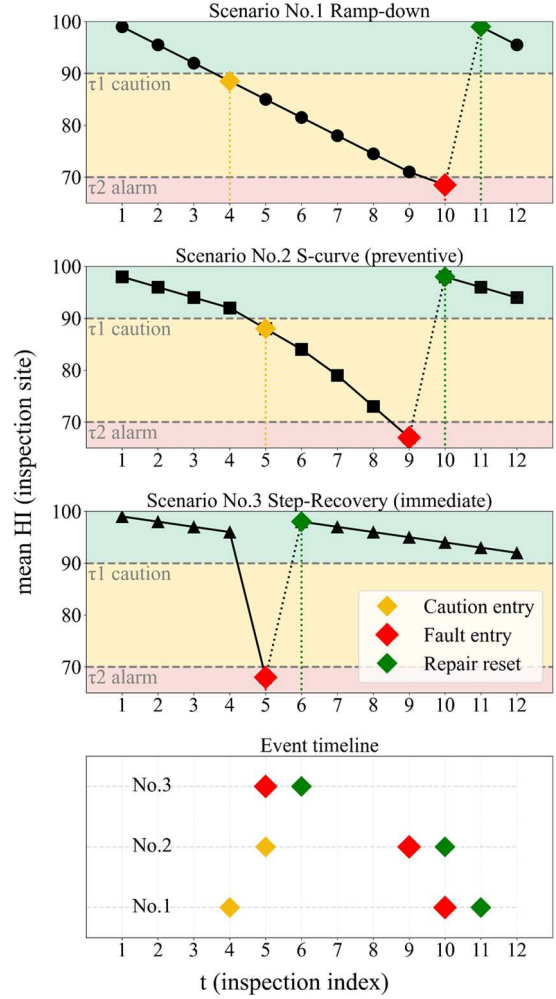


Figure 3. Scenario-based PHM KPI Protocol

Protocol remains identical to compensate for the absence of long continuous field logs, we use a synthetic timeline of length $L = 12$. At each time t , class samples are drawn with replacement, and a sequence $\{h_t\}$ is formed from the HI of each sample. Monotonicity is computed as the sign-consistency of Δh_t , trendability as correlation/alignment agreement across sequences, and prognosability as the dispersion near the alarm; we additionally report the Spearman ρ . Lead time is the average time index at which h_t first meets or exceeds τ_2 . As in Figure 3, evaluation considers three scenarios: ramp-down, S-curve, and step-to-recovery. Each scenario follows an HI-based process of degradation $\rightarrow$ alarm $\rightarrow$ repair reset over a length-12 sequence. In manufacturing-cell terms, lead time quantifies the action window available for dispatch or robotic reinspection before a Caution to Fault transition. Quantitative results (mean $\pm$ standard deviation) are presented in Section V alongside denoiser on/off KPI changes, to interpret the robustness of the proposed WFSPM.

F. Statistical Reporting and Reproducibility

We aggregate static classification performance—accuracy, macro-F1, and the confusion matrix—as internal results, and reflect PHM KPIs from the synthetic metrics in the tables and figures. The final submission artifacts are deliberately simplified to two types: recorded data and a PDF report. The report contains (i) version, (ii) summaries of zmeta/himeta, (iii) thresholds τ_1/τ_2 and the rationale for the operating point, (iv) representative signal/spectrum, (v) comparative/trend plots, and (vi) KPI summary values. This composition follows the presentation/reporting recommendations of ISO 13374/13379 and recent operational PM implementation practices. For cell integration, the same artifacts serve as the event/log source for controllers/dispatchers, enabling KPI/OEE accounting without altering the HI/threshold policy.

V. EXPERIMENTAL RESULTS

This section empirically validates the WFSPM operating point—its thresholds (τ_1, τ_2) and operational suitability in terms of early warning and action window. By consolidating numeric results, we demonstrate the stability of the chosen thresholds through static performance and noise robustness, and we analyze the early-warning to recovery flow using scenario-based PHM KPIs. In the context of manufacturing cells, these results indicate whether alarms can drive dispatcher policies or autonomous reinspection by mobile manipulators without retuning thresholds. These two experiments are arranged so that users can understand, in a continuous narrative of results and analysis, both why these thresholds are selected and what field-level benefits they guarantee.

A. Experiment Design

The experiments follow the pipeline in Sections III and IV. From the provided percussive acoustic data (320 Normal, 320 Caution, 440 Fault), we extract the features listed in Table 2 and compute the MD-HI with respect to the normal-data distribution. Classification is then performed by fixing the thresholds τ_1 and τ_2 selected by Algorithm 1. At this operating point, we first verify static performance using accuracy and macro-F1, and then, keeping the same thresholds, we inject synthetic noise (SNR = 30/20/10/5 dB) to assess robustness. Next, using the protocol in Section IV-E, we synthesize three scenarios—Ramp-down, S-curve (preventive maintenance), and Step-recovery (immediate repair)—with sequence length $L = 12$ and $n = 20$ sequences. We aggregate Monotonicity (M), Trendability (T: median Spearman ρ across sequence pairs), Prognosability (P: inverse of normalized dispersion near τ_2), auxiliary ρ , and Lead time = $t(\text{Fault}) - t(\text{Caution})$. Preventive maintenance in S-curve is triggered by sustained Caution, leading to recovery without entering Fault; Step-recovery represents immediate repair; Ramp-down follows the typical Caution $\rightarrow$ Fault $\rightarrow$ repair progression. For cell deployment, the operating point and KPIs are interpreted as cues for line dispatch decisions and robotic reinspection triggers.

B. Experiment 1: Static Performance and Noise Robustness

Tables 3–5 summarize accuracy, macro-F1, and robustness under SNR degradation (30/20/10/5 dB) at the operating point defined by τ_1 and τ_2 . As Table 3 indicates, in the Baseline condition WFSPM achieves 85.9% accuracy and 86.2% macro-F1, which is sufficiently competitive for a three-grade percussion-based task. By class, Normal is strong with $P=0.944$ and $R=0.950$ ($F1=0.947$). Caution reflects boundary uncertainty with $P=0.751$ and $R=0.850$ ($F1=0.798$). Fault shows a conservative bias—high precision with some recall trade-off—at $P=0.889$ and $R=0.800$ ($F1=0.842$). This pattern is consistent with the confusion matrix in Table 4: most errors occur between Caution and Fault, while leakage into Normal is limited, suggesting the MD-HI dual-threshold rule transparently manages the real overlap between τ_1 and τ_2 . Table 5 fixes thresholds and injects synthetic noise to assess robustness. Without denoising, performance declines gradually from SNR 30 $\rightarrow$ 5 dB (accuracy 0.832 $\rightarrow$ 0.658; macro-F1 0.835 $\rightarrow$ 0.664). With the denoiser, the drop is significantly mitigated; even at low SNR, denoising delivers consistent advantages (e.g., 0.805/0.730 accuracy and 0.802/0.726 macro-F1). Overall, Tables 3–5 jointly show (i) adequate baseline accuracy, (ii) an interpretable error pattern centered in the caution band, and (iii) graded insensitivity to SNR changes plus a buffering effect with denoising. When needed, the cost-sensitive policy or a stage-2 classifier in the caution band (Section IV-D) can further raise Caution precision and Fault recall by redistributing operational risk. Operationally, this means Caution/Fault alarms can be used as stable inputs to dispatcher policies and to trigger mobile manipulator rechecks without changing thresholds, while denoising acts as an engineering knob to preserve performance under adverse SNR.

TABLE 3. STATIC 3-CLASS PERFORMANCE AT THE OPTIMIZED τ_1, τ_2

Class	Precision	Recall	F1
Normal	0.944	0.950	0.947
Caution	0.751	0.850	0.798
Fault	0.889	0.800	0.842

TABLE 4. CONFUSION MATRIX AT THE OPTIMIZED τ_1, τ_2

	Pred-N	Pred-C	Pred-F
True-N	304	12	4
True-C	8	272	40
True-F	10	78	352

TABLE 5. NOISE ROBUSTNESS ACROSS SNR BY EXISTENCE OF DENOISING

Setting	Accuracy (noised/denoised)	Macro-F1 (noised/denoised)
Baseline	0.859 / -	0.862 / -
SNR 30dB	0.832 / 0.980	0.835 / 0.976
SNR 20dB	0.792 / 0.860	0.798 / 0.857
SNR 10dB	0.732 / 0.805	0.739 / 0.802
SNR 5dB	0.658 / 0.730	0.664 / 0.726

C. Experiment 2: Scenario-based PHM KPIs

Figure 4 consolidates monotonicity (M), trendability (T), prognosability (P), auxiliary ρ , and lead time for Ramp-down,

S-curve (preventive), and Step-recovery (immediate). T remains high overall—0.975, 0.952, and 0.904—indicating that different degradation patterns still share a common HI shape template. In contrast, M, P, and ρ lie in the mid range (e.g., Ramp-down $M=0.625\pm0.077$, $P=0.731\pm0.081$, $\rho=0.654\pm0.142$), reflecting realistic variability rather than perfect monotonicity or clustering under field noise and operator spread. For S-curve, despite preventive action triggered by sustained Caution, $M/P/\rho$ are not exaggerated (0.547 ± 0.095 , 0.626 ± 0.105 , 0.564 ± 0.166), showing that early intervention is feasible while the data remain imperfect. Step-recovery mixes abrupt change and repair, yielding the largest variability ($M=0.482\pm0.121$, $P=0.582\pm0.120$, $\rho=0.508\pm0.186$); T is relatively lower at 0.904 but still acceptable. Notably, lead time averages about five frames across scenarios— 5.72 ± 2.06 , 5.42 ± 2.49 , 5.08 ± 2.79 —indicating that WFSPM provides a stable action window before a Caution $\rightarrow$ Fault transition. In manufacturing-cell terms, $\text{Lead}\approx 5$ quantifies a scheduling headroom for dispatch or robotic reinspection between the first sustained Caution and a potential Fault, enabling policy-tunable interventions without threshold retuning. In short, Figure 4 jointly exhibits (i) shape consistency (high T), (ii) realistic variability (mid $M/P/\rho$), and (iii) actionable headroom ($\text{Lead}\approx 5$), supporting a practical early-warning, response and recovery cycle when combined with the operating point justified in Section V-B.

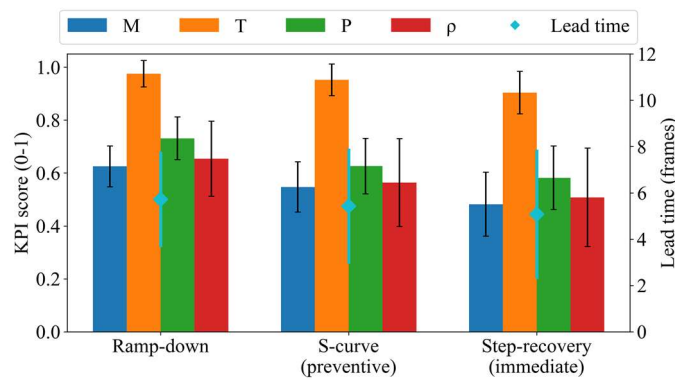


Figure 4. The Proposed WFSPM KPIs across Scenarios

VI. CONCLUSIONS

This paper proposed WFSPM, centered on an MD-HI and dual thresholds, and implemented/validated a standards-compliant end-to-end operating stack covering data flow, features, decisions, and UI/reporting. Static performance and noise robustness show that the thresholding strategy achieves a practical, deployable level without overstatement. With thresholds fixed, degrading SNR causes only gradual performance loss, and applying the denoiser restores performance—especially at 30/20 dB—near the baseline, indicating stable operation under varying noise. Scenario-wise KPIs maintain high T while securing an average lead time of about five frames, quantitatively supporting an early-warning, action-window to recovery cycle. In short, WFSPM simultaneously delivers (i) an explainable HI with transparent thresholds, (ii) robustness to industrial noise/variability, and (iii)

risk control tunable by operating policy. Because the HI operates at the metric level and the stack is standards-driven, alarms and metadata can be exposed to manufacturing-cell controllers/dispatchers without retuning, enabling autonomous reinspection by mobile manipulators and mapping events to OEE/KPI. The artifact design—focused on percussive signal data and reports—also improves reproducibility and auditability.

As future work, we plan to (i) validate with long-term field logs tied to real inspection/maintenance events, and (ii) advance online threshold auto-tuning and uncertainty estimation to further improve practicality, including a transfer to mixed-model manufacturing cells for autonomous inspection with mobile-manipulator acquisitions and cell-to-line KPI integration.

ACKNOWLEDGMENT

This research was supported by RS-2025-25454751 (Development of an AI-Based Reconfigurable Cell Assembly Autonomous Manufacturing System Using a High-Payload Mobile Manipulator for Mixed-Model Manufacturing) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) through the Korea Evaluation Institute of Industrial Technology (KEIT).

REFERENCES

- [1] D. Shon, T. H. Yoon, W.-S. Jung, and D. S. Yoo, "A study on creating a wedge map based on generator stator wedge boundary recognition," *KICS Fall Conference*, vol. 82, pp. 91-92, 2023.
- [2] ISO, "ISO 13374-1 - Condition monitoring and diagnostics of machines - Data processing, communication and presentation - Part 1: General guidelines," Geneva: International Organization for Standardization, 2003.
- [3] ISO, "ISO 13374-3 - Condition monitoring and diagnostics of machines - Data processing, communication and presentation - Part 3: Communication," Geneva: International Organization for Standardization, 2013.
- [4] ISO, "ISO 13379-1 - Condition monitoring and diagnostics of machines - Data interpretation and diagnostics techniques - Part 1: General guidelines," Geneva: International Organization for Standardization, 2013.
- [5] L.-Y. Hou and D.-B. Zhuo, "Research on bolt loosening recognition based on sound signal and GA-SVM-RFE," *Scientific Reports*, vol. 15, pp. 24165, 2025.
- [6] F. Alharbi, S. Luo, H. Zhang, K. Shaukat, G. Yang, C. A. Wheeler and Z. Chen, "A brief review of acoustic and vibration signal-based fault detection for belt conveyor idlers using machine learning models," *Sensors*, vol. 23, no. 4, pp. 1902, 2023.
- [7] W. Fu, R. Zhou, and Z. Guo, "Automatic bolt tightness detection using acoustic emission and deep learning," *Structures*, vol. 55, pp. 1774-1782, 2023.
- [8] H.-M. Park, D. Shon, T. H. Yoon, W.-S. Jung, J.-H. Park, and D. S. Yoo, "AI-based Percussive Acoustic Signal Classification for Fastener Strength Inspection of Stator Wedge," *In Proceedings of the 2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 2180-2185, 2024.
- [9] R. A. McCarthy, Y. Zhang, S. A. Verburg, W. F. Jenkins, and P. Gerstoft, "Machine Learning in Acoustics: A Review and Open-source Repository," *npj Acoustics*, vol. 1, no. 1, pp. 18, 2025.
- [10] J. Coble and J. W. Hines, "Identifying optimal prognostic parameters from data: a genetic algorithms approach," *In Proceedings of the Annual Conference of the PHM Society*, 2009.
- [11] S. Ma, K. A. Flanigan, and M. Berges, "State-of-the-art review and synthesis: A requirement-based roadmap for standardized predictive

- maintenance automation using digital twin technologies,” *Advanced Engineering Informatics*, vol. 62, pp. 102800, 2024.
- [12] ITU, “ITU-T F. 743.28 Functional framework and requirements for an intelligent inspection system based on cloud-edge-device collaboration in smart grids,” *ITU Publications Recommendations*, 2025.
- [13] A. C. Marquez, J. A. M. Alberca, A. J. G. Lopez, and A. D. L. F. Carmona, “Digital twins in condition-based maintenance apps: A case study for train axle bearings,” *Computers in Industry*, vol. 151, pp. 103980, 2023.
- [14] J. Cuesta, U. Leturiondo, Y. Vidal, and F. Pozo, “A review of prognostics and health management techniques in wind energy,” *Reliability Engineering & System Safety*, vol. 2025, pp. 111004, 2025.
- [15] Y. S. Zhao, P. Li, Y. Kang, and Y. B. Zhao, “A health indicator enabling both first predicting time detection and remaining useful life prediction: Application to rotating machinery,” *Measurement*, vol. 235, pp. 114994, 2024.
- [16] B. P. Duong, S. A. Khan, D. Shon, K. Im, J. Park, D.-S. Lim, B. Jang, and J.-M. Kim, “A Reliable Health Indicator for Fault Prognosis of Bearings,” *Sensors*, vol. 18, no. 11, pp. 3740, 2018.
- [17] L. Huang, X. Pan, Y. Liu, and L. Gong, “An unsupervised machine learning approach for monitoring data fusion and health indicator construction,” *Sensors*, vol. 23, no. 16, pp. 7239, 2023.
- [18] M. Salinas-Camus, K. Goebel, and N. Eleftheroglou, “A Comprehensive review and evaluation framework for data-driven prognostics: Uncertainty, robustness, interpretability, and feasibility,” *Mechanical Systems and Signal Processing*, vol. 237, pp. 113015, 2025.
- [19] X. Hou, W. Guo, S. Ren, Y. Li, Y. Si, and L. Su, “Bolt-loosening detection using 1D and 2D input data based on two-stream convolutional neural networks,” *Materials*, vol. 15, no. 19, pp. 6757, 2022.
- [20] M. Qin, Z. Xie, J. Xie, X. Yu, Z. Ma, and J. Wang, “Automated Structural Bolt Micro Looseness Monitoring Method Using Deep Learning,” *Sensors*, vol. 24, no. 22, pp. 7340, 2024.
- [21] L. Dou, W. Li, B. Wang, T. Zhang, and Q. Cao, “Study on Fault Detection of Generator SlotWedge Tightness Based on Improved KNN,” *Machine Learning: Engineering*, vol. 1, no. 1, pp. 15007, 2025.
- [22] Y. Fukushima, and N. Kawada, “Online monitoring method for mold deformation using Mahalanobis distance,” *International Journal of Automation Technology*, vol. 15, no. 5, pp. 689-695, 2021.