Journal of Intelligent Manufacturing
https://doi.org/10.1007/s10845-026-02809-3

Adaptive multi-scale threshold with time-gated evaluation for real-time Fault detection in non-stationary and high-dynamic industrial cyber-physical systems

Joma Aldrini[1] · **Ines Chihi**[1]

Received: 16 July 2025 / Accepted: 27 January 2026

Abstract
Fault detection and diagnosis approaches play a vital role in maintaining reliability and resilience of industrial cyber-physical systems, especially in smart manufacturing settings characterized by complex interdependencies and dynamic operating conditions. Conventional threshold-based fault detection and diagnosis methods, including global or local, often struggle to detect faults reliably under non-stationary conditions, where current amplitudes and cycle durations are varying due to changes in load or machine behavior. To address these limitations, this article proposes a novel multi-scale fault detection approach that integrates global and local detection methods via an adaptive dual-threshold strategy. The proposed approach integrates exponentially weighted moving average for capturing global signal drifts with a peak-to-peak envelope for detecting localized deviations. A key innovation lies in the use of adaptive temporal scaling, where the window size used in local thresholding adjusts dynamically in real time based on signal variance. Allowing robust detection across both abrupt and incipient faults with high temporal accuracy. Additionally, the approach also employs station-specific thresholds and a time-gated evaluation mechanism to suppress false positives and ensure operational relevance and improve interpretability. The proposed approach is validated on a real-world conveyor-based sorting system, demonstrating superior performance across several fault types across multiple stations. Experimental comparative results demonstrate its effectiveness in multi-scale responsiveness while reducing false alarm rates compared with global and local methods. Detection rates exceeding 98%, zero false alarms, and robust accuracy under varying operational conditions. This multi-scale approach offers a scalable and deployable solution for real-time fault detection in next-generation industrial cyber-physical systems.

Keywords Fault detection and diagnosis · Smart manufacturing · Adaptive dual threshold · Exponentially weighted moving average · Adaptive temporal scaling · Time-gated evaluation · Industrial cyber-physical systems

Introduction

With the advent of Industry 4.0, Industrial Cyber-Physical Systems (ICPS) have become central to Smart Manufacturing (SM), which integrates advanced sensors, intelligent machinery, robotics, Radio Frequency Identification (RFID), Manufacturing Execution System (MES), Enterprise Resource Planning (ERP), and more. These advancements enable real-time monitoring, autonomous decision-making in SM processes, and improved efficacy and responsiveness. However, the inherent interconnectivity and complexity of ICPS introduces new Fault Detection and Diagnosis (FDD) challenges. Faults such as actuator malfunctions, sensor dropouts, or unexpected mechanical behaviors can propagate rapidly through interconnected subsystems, leading to system failures, unplanned downtime, and reduced productivity (Aldrini et al., 2023).

In conveyor-based discrete manufacturing systems, common faults include intermittent actuator faults or unrecognized carrier movement, which can trigger looping failures, serial faults, or misrouted components. These faults often evolve in non-stationary environments, where parameters

✉ Joma Aldrini
joma.aldrini@uni.lu

Ines Chihi
ines.chihi@uni.lu

[1] Department of Engineering, Faculty of Science, Technology and Medicine, University of Luxembourg, 6, Rue Richard Coudenhove-Kalergi, L-1359 Esch-Sur-Alzette, Luxembourg

Published online: 18 February 2026

such as motor current amplitude and cycle duration naturally fluctuate due to dynamic loading, part variability, and station (machinery unit)-specific processing behaviors. This variability complicates the establishment of a consistent nominal reference for fault detection (Leite et al., 2024).

A core component of many FDD approaches is the use of threshold-based methods, where a fault is flagged once an operational signal crosses a predefined threshold (Aslansefat et al., 2020; Tang & Li, 2023). Thresholds can be implemented in two principal strategies: Global Threshold Approaches (GTAs), which apply static or dynamic thresholds across the entire system, and Local Threshold Approaches (LTAs), which adjust thresholds for specific system segments or time window. While GTAs are effective in detecting long-term drifts, for instance, using Exponentially Weighted Moving Average (EWMA) and Cumulative Sum (CUMSUM) control charts, they often fail to detect abrupt or transient faults due to their averaging nature (Iqbal et al., 2023). Conversely, LTAs such as peak-to-peak analysis or local outlier factor methods are sensitive to localized faults but often lack robustness to noise and wider system context (Kim et al., 2022; Ma et al., 2023).

Recent studies have proposed adaptive and hybrid approaches that combine global monitoring with localized sensitivity, aiming to overcome the limitations of each strategy (Lee et al., 2024; L. Wang et al., 2025a, 2025b). However, many of these solutions are still constrained by offline adjustment requirements, domain-specific tuning, or inflexibility under real-time operation conditions. Moreover, they often address only one side of the detection spectrum, either gradual evolving or abrupt faults, while overlooking the need for simultaneous detection across multiple fault types in dynamic and non-stationary systems.

To overcome these limitations, this paper proposes a novel Multi-Scale Fault Detection (MSFD) framework customized for real-time ICPS. The approach integrates adaptive global and local thresholds, utilizing EWMA for long-term trend detection and adaptive peak-to-peak analysis within adaptive windows for local fault sensitivity. The MSFD framework introduces station-specific dual thresholds to accommodate mechanical heterogeneity and employs time-gated evaluation to suppress false positives and reduce operator cognitive load. Most critically, the thresholds are adjusted online, adapting to nominal and early fault signals without manual reconfiguration. This unified approach is validated on a conveyor-based sorting system, demonstrating superior performance across a range of fault types and varying operational conditions.

The remainder of this article is structured as follows: Sect. 2 reviews related literature; Sect. 3 explains the proposed methodology. Section 4 describes the experimental case study and fault scenarios; Sect. 5 discusses the results, and Sect. 6 concludes the research work with future research directions.

Related work

Fault detection in ICPS remains a significant challenge due to the dynamic and non-stationary nature of industrial operations. In conveyor-based production and discrete manufacturing systems, process parameters such as motor current amplitude and operational cycle duration naturally vary even under nominal conditions (Eslami et al., 2023). These variations, driven by changes in load, speed, or component characteristics, complicate the establishment of a healthy (fault-free) reference for fault detection (Rudawska et al., 2020). These issues are not only common in practice but also widely reported in the literature, emphasizing the limitations of static thresholding and rigid model assumptions (Dowdeswell et al., 2020; Leite et al., 2024).

GTAs involve statistical control methods such as EWMA, CUMSUM, X- and S-bar methods have traditionally been used for long-term detection (Marais et al., 2022). For instance, Haddar et al., (2024) investigated the univariate statistical control methods for crack fault detection in bevel gears. The results demonstrated that EWMA is an effective method compared to others. Along the same lines, Maras et al., (2021) proposed a wear fault detection in spur gears utilizing the vibration analysis method, X- and S-bar statistical control charts, and statistical parameters. Yet, these methods suffer from latency in identifying short-term or abrupt faults due to their reliance on smoothing and averaging (Li et al., 2023; Tran, 2022).

In contrast, LTAs offer higher sensitivity to context-specific variations by analyzing localized segments of signal data. Such techniques peak-to-peak (Xu et al., 2019), local outlier factor (Kim et al., 2022) and binary segmentation (Fan et al., 2021) have been applied to detect sudden or transient faults. Despite peak-to-peak offers faster responsiveness, but its prone to false alarms in a noisy, discrete manufacturing environment due to contextual unawareness of load and speed variations (Li & Dong, 2025). Whereas the local outlier factor is a powerful method for fault detection, especially in complex systems where data distribution is irregular or multimodal. it relies on calculating the local density of each data point relative to its neighbors, which becomes increasingly computational cost, making it limited in practice for real-world applications (Alghushairy et al., 2020). Although binary segmentation is an effective method for change-point detection, particularly in fault detection of time series data characterized by several sudden changes. However, it is limited in handling closely spaced or small changes, and its performance can be influenced by noise and cost function selection (Yan et al., 2023). In short, LTAs are

vulnerable to noise sensitivity and lack insight into global system behavior, often leading to false positives in dynamic environments (Wang et al., 2025a, 2025b).

Several studies have proposed hybrid approaches to overcome the weaknesses of standalone approaches. For instance, Harrou et al., (2013) presented a hybrid fault detection method based on an EWMA and PCA model. EWMA was applied to the residual to detect faults when the data did not fit the PCA model. The results demonstrated the effectiveness of the proposed method compared to conventional fault detection methods using a simulated continuously stirred tank reactor. However, the EWMA–PCA hybrid approach has notable limitations. Its effectiveness relies heavily on the quality of the residuals produced by the PCA model, which can be compromised in dynamic or time-varying systems, potentially resulting in missed or delayed fault detection. Dong et al., (2025) suggests a hybrid mode based on integrating slow feature analysis and local outlier factor methods for fault detection in the nonstationary process of the Tennessee Eastman process. Harrou et al., (2015) proposed a Partial Least Squares (PLS)-based EWMA fault detection method for process monitoring. The simulated results demonstrated the effectiveness of the proposed method over the conventional PLS, particularly in the presence of faults with small magnitudes. Nevertheless, the method still has certain limitations from PLS, such as reduced effectiveness in handling nonlinear relationships and time-dependent dynamics. Additionally, its sensitivity may decrease in the presence of noise or when model assumptions are not fully met. Although hybrid approaches improve sensitivity to small faults, they are limited by their reliance on fixed parameters, data dependency, or address only one aspect of the global or local detection spectrum and sensitivity to noise. This reduces their adaptability and robustness in dynamic environments with non-stationary behavior systems in real-time applications (Sorostinean et al., 2025).

Recent advancements have introduced adaptive threshold techniques for FD in industrial systems. Such advancements, focusing on AI-driven optimization, for instance, (Veerasamy et al., 2022) integrated a genetic algorithm adjusted adaptive fading memory Kalman filter in a model predictive controller for fault sensor fault detection in cement kiln pyro-process. The results demonstrate a reduction in modelling errors through automated adjustment rather than manual calibration. Other advanced deep reinforcement learning can adaptively learn optimal detection strategies based on the system's dynamics. For example, a deep Q-network was employed to develop a real-time fault detection policy that effectively managed complex interactions between cyber-physical system components (Stanly Jayaprakash et al., 2022). In article (Liu et al., 2023) presented a multi-threshold segmentation by combining gray wolf optimization with symmetric cross entropy to partition time–frequency images of gearbox vibrations under varying speed conditions. The suggested method identifies subdomain-specific thresholds for micro-cracks and misalignment, fostering fault detection accuracy compared to other methods. Ma et al., (2023) introduced an adaptive threshold calculation-based method for interval combining the P2P method for residual generation, demonstrating improved resilience to system uncertainties. However, the method's performance can be limited by the sensitivity of the adaptive threshold to parameter tuning and its dependency on system-specific characteristics. In article (Ahmadini et al., 2025) presented adaptive CUMSUM control chart to minimize the unfavorable effects of measurement errors through combining a linear covariate model and multi-measurement method. The results demonstrated the suggested method to improve industrial process monitoring. Sun et al., (2021) proposed an adaptive fault detection and root cause analysis for dynamic industrial processes utilizing window kernel principal component analysis and information geometric causal interface. The findings show that the proposed methodology has acceptable performance in minimizing false alarms and missed detection rates.

Despite these advancements, most existing approaches remain focused on either global or local aspects exclusively. Few works provide a unified framework capable of detecting several types of faults in real-time, and variable operational conditions in manufacturing operations. Furthermore, many existing systems require offline calibration, manual reconfiguration, or are sensitive to parameter tuning, which hinders practical deployment in SM (Aldrini & Chihi, 2025). This gap is particularly evident in discrete manufacturing systems where processing steps vary by station, and fault manifestations differ significantly in time scale and amplitude signatures.

To address these limitations, this study proposes a novel MSFD framework that integrates adaptive EWMA-based global monitoring with adaptive window localized P2P analysis in a unified architecture for fault detection. The proposed approach incorporates station-specific dual thresholds, a time-gated detection mechanism allowing real-time adaptation. Resulting in a robust and scalable solution for modern SM systems characterized by complex dynamics and nonstationary behaviors.

Methodology: proposed adaptive EWMA-based dual threshold with time-gated fault detection for ICPS

This methodology addresses the critical challenge of detecting incipient faults in ICPS through advanced analysis of current signatures. It leverages dynamic segmentation,

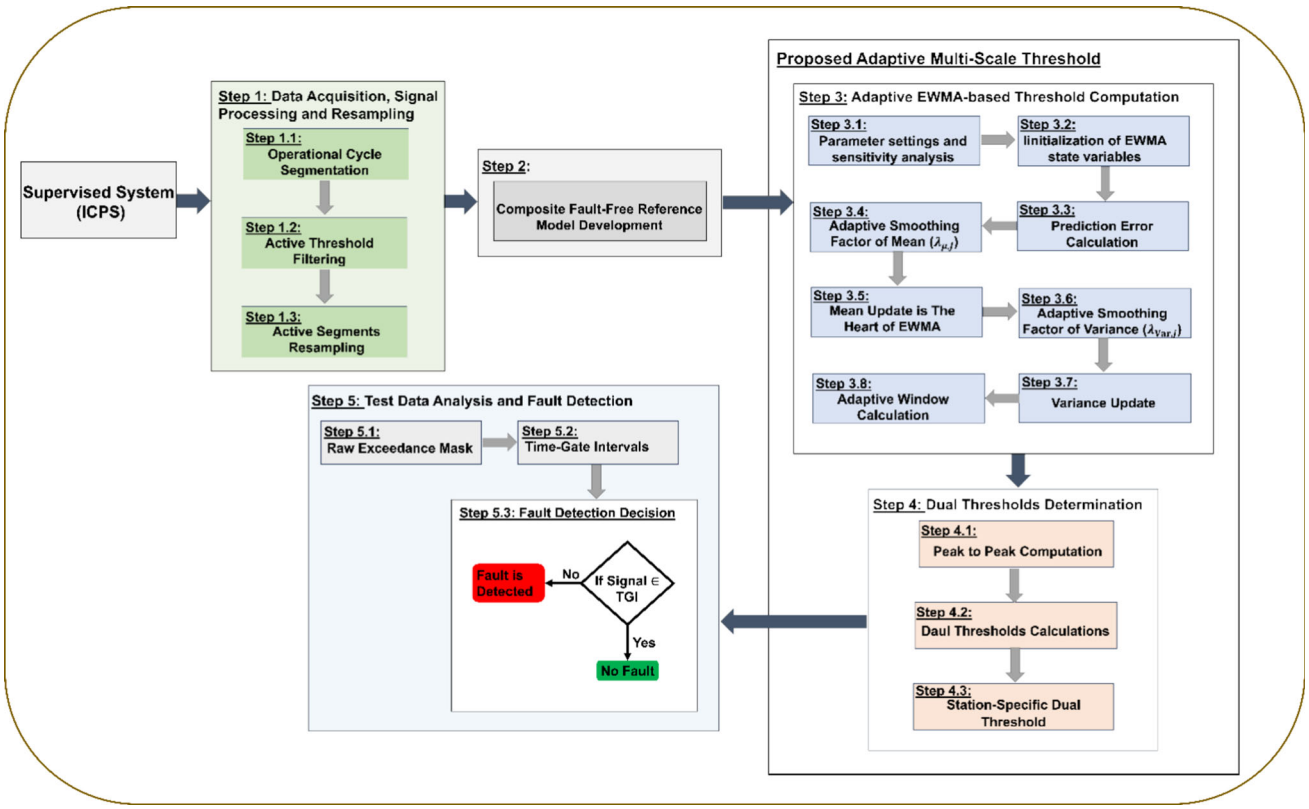

Fig. 1 Workflow of the proposed methodology

signal normalization, and adaptive thresholding to effectively manage non-stationary signals and variable operational cycles. The core innovation centers around the application of EWMA with dual adaptive thresholds that dynamically adjust based on local signal volatility and specific operational characteristics unique to each station. The methodology is structured into five steps as displayed in Fig. 1. Step 1, data acquisition, signal processing, and resampling. Crucially, it incorporates fixed-length resampling of variable-duration working segments, enabling consistent and reliable comparative analysis across cycles. Step 2, Development of fault-free reference model using multiple operational cycles. Step 3. Calculation of adaptive EWMA-based threshold. Step 4, Determination of dual thresholds. Finally, Step 5, wherein operational cycles are monitored and compared against this reference to detect faults. The integration of EWMA, adaptive threshold calculations, and adaptive moving window enable dynamic peak-to-peak signal analysis, ensures accurate and timely fault detection. The faults are detected through time-gated evaluation to ensure robustness against short transient faults. In other words, a time-gate ensures the alarm log is meaningful, indicating that each alarm reflects a true, sustained deviation that requires investigation. Hence, reduce false alarms and minimize operator cognitive loads. A key contribution of the proposed methodology is the implementation of adaptive temporal scaling, wherein the window length used for fault thresholding is dynamically adjusted based on local signal variance ratios. This enables the method to flexibly capture both fast transient faults and gradual faults in real time, a capability critical for non-stationary industrial settings.

These five main steps are formalized in Algorithm 1, which summarizes the unified procedure for station-specific, real-time fault detection using the adaptive dual threshold principle.

Algorithm 1 Adaptive multi-scale fault detection using EWMA and dynamic peak-to-peak thresholds

Input:
- Healthy dataset D_h, Test dataset D_t
- Station set S = {I_sel, I_cam, I_mill, I_hand}
- Resample length N
- Adaptive thresholding parameters θ per station

Initialization (parameter configurations)
- EWMA parameters (station-specific ranges):
  $\lambda\_\mu\_min = 0.05$, $\lambda\_\mu\_max = 0.5$
  $\lambda\_var\_min = 0.05$, $\lambda\_var\_max = 0.5$
  $W\_min = 10$, $W\_max = 100$
  $\alpha_0 = 0.015–0.02$
  $\alpha\_min = 0.01–0.015$, $\alpha\_max = 0.02–0.025$
  τ_i = 10th percentile(D_h) + 0.01 A (station-specific current signal)
  $\Delta t\_min = 0.01$ s (time-gating minimum fault duration)

Output:
- Fault intervals F_i for each station $i \in S$
- Detection metrics (DR, MDR, FPR, Accuracy)

1. For each station $i \in S$:
   a. Extract and resample healthy working segments from D_h (current signal > τ_i)
   b. For each resampled segment:
      i. Initialize EWMA mean μ(0), variance v(0) and W(0)
      ii. For each time step t = 1 to N:
         - Compute Error e(t) = x(t) - μ(t−1)
         - Adapt EWMA learning rates λ_μ(t), λ_var(t) based on e(t)
         - Update μ(t) and var(t) recursively
         - Adapt temporal scaling of window W(t) using variance ratio var(t−1)/var(t)
         - Compute local dynamic peak-to-peak range over W(t)
         - Compute adaptive thresholds:
           $U(t) = \mu(t) + \alpha(t) \cdot P2P(t)$
           $Lw(t) = \mu(t) - \alpha(t) \cdot P2P(t)$
   c. Composite all fault-free segments to establish reference envelopes : μ_ref(t), U_ref(t), Lw_ref(t)

2. For each test segment in D_t:
   a. Resample and computing EWMA profile μ_test(t)
   b. Compare μ_test(t) to U_ref(t), Lw_ref(t) :
      - raw_pred(t) = [μ_test(t) > U_ref(t)] ∨ [μ_test(t) < Lw_ref(t)]
   c. Apply time-gated filter to exclude faults < Δt_min:
      - $F_i = \{(t_s, t_e) \mid duration(t_e - t_s) \geq \Delta t\_min\}$

3. Evaluate detections:
   a. Construct detections vs actual fault masks
   b. Compute detection delay, DR, MDR, FPR, Accuracy

Return: F_i and performance metrics per station.

This algorithm outlines the proposed MSFD framework. It uses real-time EWMA mean, and variance tracking combined with an adaptive moving window to capture local signal dynamics via the peak-to-peak method. Fault regions are determined through time-gated analysis to ensure robustness against short transients and noise.

Step 1: Data acquisition, signal processing, and segmentation.

Data utilized in this methodology is acquired from the ICPS in the lab-scale, which contains operational measurements across multiple operational cycles. This step aims to isolate active working segments of each operational cycle and normalize them to a fixed length for cross-cycle compatibility. The preprocessing phase includes the following:

Step 1.1: Extraction and segmentation of operational cycles.

Industrial operation cycles exhibit variable durations and non-stationary current signatures. Each operational cycle is identified by a unique ID. A given cycle is segmented into a working phase based on a binary state of the position sensor, which indicates when the station is working (active). The working segment starts at the first transition from 1 to 0 and ends at the next transition from 0 to 1. This is critical because, in fact, the active interval carries the mechanical or electrical signatures meaningful for fault detection.

Step 1.2: Active threshold filtering.

Within the working segment, a further filter out idle periods to isolate the most relevant operational data by applying a threshold for the current signal for each station. The threshold is set as the 10th percentile of the current healthy signal plus a small constant (0.01 A) by using the following formula:

$$\theta = Q_{0.10}(x) + 0.01 \tag{1}$$

Step 1.3: Resampling of working (Active) segments.

The filtered active segment of the variable length is resampled to a fixed $N = 100$ using linear interpolation. This normalization ensures that the segments from cycles of varying durations are comparable. The resampling is performed on the normalized time domain (from 0 to 1) to preserve the shape of the segment. By resampling every segment to the same length N, we create a common time-normalized grid ($j = 0 \ldots N-1$).

For a given working segment, the timestamps $t_0, t_1, \ldots, t_{M_c-1}$, Where M_c is the raw length. The normalized time coordinate is computed by:

$$\tau_i = \frac{t_i - t_0}{t_{M_c-1} - t_0} \in [0, 1] \tag{2}$$

This maps the start of work to $\tau = 0$ and end to $\tau = 1$, regardless of absolute duration. Then we measured values x_i at those τ_i to build an interpolation function $f(\tau)$. After that, we evaluate f at N evenly spaced points.

$$\tau_i = \frac{j}{(N-1)},\ j = 0, \ldots, N-1$$

The result is a length-N vector $n_j = f(\tau_i)$.

Equation recap for $j = 0, \ldots, N-1$

$$n_j = x\left(t_0 + \left(t_{M_c-1} - t_0\right)\frac{j}{(N-1)}\right) \tag{3}$$

(In practice, we do this by normalizing to τ and then interpolating)

In short, resampling to a fixed length N turns a collection of irregular-length signals into a neat matrix of size $K \times N$, enabling all the downstream averaging, thresholding, and fault comparison to work properly.

Step 2: Composite fault-free reference model development.

This step aims to create a reference model of normal behavior by aggregating the multiple fault-free cycles. For each station's current signals, multiple fault-free cycles are processed to extract and resample the working segments $\{s^k\}$. These segments form a matrix size $(K \times N)$, where K is the number of fault-free cycles and N is the fixed length.

The element-wise mean (representative reference model), $\bar{s}[j]$ for each $j \in \{0, \ldots, N-1\}$ is computed as the average of all fault-free segments.

$$\bar{s}[j] = \frac{1}{M}\sum_{k=1}^{K} s^k[j] \tag{4}$$

where M refers to the fault-free cycle segments for a given station, $s^k[j]$ denotes the j^{th} sample of k^{th} segment.

Step 3: Adaptive EWMA-based threshold computation.

The core of the novel methodology is the computation of adaptive thresholds for the resampled healthy (fault-free) segment. The EWMA parameters are adjusted dynamically based on the signal's local volatility. This allows the thresholds to tighten during stable periods (better sensitivity) and widen during transition (avoiding false alarms).

Step 3.1: Parameter settings and sensitivity analysis.

The selection of EWMA parameters corresponds to standard control charts theory (Montgomery, 2009). The smoothing factor λ is conventionally selected to balance responsiveness and noise reduction. For fault detection in high-dynamic and non-stationary systems, we employ an adaptive range $\lambda \in [0.05, 0.5]$, where $\lambda = 0.05$ ensures noise smoothing and stability, minimizing false alarms, and $\lambda = 0.5$ enables a fast response during transient events. The adaptive mechanisms (Steps 3.3 and 3.5) automatically interpolate based on the magnitude of the local prediction error.

Table 1 Sensitivity analysis results using OFAT

parameter	Range tested	Effect on DR	Effect on FPR	Selected value
$\boldsymbol{\lambda_{\mu,min}}$	0.02–0.1	↓ at 0.02	↑ at 0.02	0.05
$\boldsymbol{\lambda_{\mu,max}}$	0.3–0.7	↑ at 0.7	↑↑ at 0.7	0.5
$\boldsymbol{W_{min}}$	5–20	≈ stable	≈ stable	10
$\boldsymbol{W_{max}}$	50–150	↓ at 150	↑ at 150	100
α_0	0.01–0.03	↑ at 0.03	↑ at 0.03	0.015–0.02

The window size bounds $W \in [10, 100]$ ($W_{min} = 10$, $W_{max} = 100$) are constrained by the resampling length $N = 100$, ensuring $W_{min} \geq 10\%$ of N for robust peak-to-peak calculation, while $W_{max} \leq$ N prevents over-smoothing to ensure computational efficiency. The ration $W_{max}/W_{min} = 10$ offers one order of magnitude adaptivity. The variance-weighted scaling (Step 3.8) enables both fast transient and gradual faults.

Threshold scaling factors ($\alpha_0 = 0.015$–0.02, $\alpha_{min} = 0.01$–0.015, $\alpha_{max} = 0.02$–0.025): these are derived from the requirement, which U_j, and L_j (upper and lower thresholds) should capture $\geq 95\%$ of healthy observations, adjusted for the peak-to-peak metric rather than standard deviation.

Furthermore, the parameters were initially adjusted using the fault-free dataset through: a) testing multiple parameter combinations on the reference profile, b) ensuring that the reference thresholds do not falsely alarm on nominal cycles, and c) validating that the reference threshold is tight enough to detect significant deviations.

Additionally, a sensitivity analysis is conducted to strengthen the robustness of parameter selections using the One-Factor-At-a-Time (OFAT) method (Hegazy et al., 2024). Our objective is to determine how different parameters of the proposed adaptive multi-scale fault detection approach affect its evaluation metrics, including detection rate and false positive rates. Table 1 describes the parameters and metrics used in the OFAT analysis.

Step 3.2: Initialization of the EWMA state variables.

Let n_k, $k = 1, \ldots, N$, detonate the resampled working segment samples. The EWMA estimates of mean and variance at index j are denoted by μ_j and var_j respectively, and the adaptive window size W_j. To avoid ambiguity at the beginning of the segment, firstly we define a seed window length as follows.

$$N_{seed} = min\{W_{min}, N\} \tag{5}$$

where W_{min} is the minimum adaptive window size. In the methodology, all segments are resampled to $N = 100$ samples and $W_{min} = 10$, so $N_{seed} = 10$. For numerical stability, the initial EWMA mean $\mu(0)$ is calculated as the arithmetic mean of the first $N_\mu = \min\{2, N\}$ samples.

$$\mu(0) = \begin{cases} \frac{1}{N_\mu} \sum_{k=0}^{N_\mu - 1} n_k, \; N \geq 1 \\ 0, \; N = 0 \end{cases} \tag{6}$$

In the methodology setting ($N = 100$), this reduces to

$$\mu(0) = \frac{(n_0 + n_1)}{2} \tag{7}$$

The initial variance estimate $var(0)$ is computed over the seed window using the following equations:

$$\overline{n}_{seed} = \frac{1}{N_{seed}} \sum_{k=0}^{N_{seed}-1} n_k \tag{8}$$

$$var(0) = \begin{cases} \frac{1}{N_{seed}} \sum_{k=0}^{N_{seed}-1} (n_k - \overline{n}_{seed})^2, \; N_{seed} \geq 2 \\ 0, \; N_{seed} < 2 \end{cases} \tag{9}$$

Hence, for the parameters used in this work:

$$var(0) = \frac{1}{10} \sum_{k=0}^{9} (n_k - \overline{n}_{seed})^2 \tag{10}$$

The adaptive window is initialized at its minimum size:

$$W(0) = W_{min} \tag{11}$$

These initial values $\mu(0)$, $var(0)$, and $W(0)$ rely on the first samples of the resampled working segment.

In the methodology, segments are resampled to $N = 100$ samples and $W_{min} = 10$, so $N_{seed} = 10$,, and $var(0)$ is the variance over the first 10 samples.

Step 3.3: Prediction error calculation.

The relation is the instantaneous prediction error at time j is computed by.

$$e_j = n_j - \mu_{j-1} \tag{12}$$

where, n_j is j[th] resampled signal value in the working segment. μ_{j-1} is the EWMA estimate of the underlying healthy

signal mean, as in the previous step $(j-1)$. e_j is the difference between what we observed n_j and what we expected based on past behavior μ_{j-1}.

The importance of computing the prediction error lies in: (1) In EWMA, both the mean estimate (μ_j) and variance estimate (Var_j) are updated on this error, (2) Forming adaptive thresholds (U_j, L_j) as they were designed around the μ_j by adding and subtracting a multiple of the local range. That multiple $(\alpha_j \cdot P_j)$ itself tuned by how big past errors have been (via the variance Var_j, which itself is driven by these same errors).

To sum up, the prediction error is the fundamental innovation term that drives the entire adaptive filtering fault detection algorithm, which enables the algorithm to determine how much to learn from the new data, and how to update confidence in the fault-free state.

Step 3.4: Adaptive smoothing factor of mean $(\boldsymbol{\lambda_{\mu,j}})$.

This sub-step explains the selection of the forgetting factor of the EWMA mean update at each time step j.

$$\lambda_{\mu,j} = clip\Big(\lambda_{\mu,min} + \big(\lambda_{\mu,max} - \lambda_{\mu,min}\big) \cdot \frac{|e_j|}{|e_j| + \sqrt{Var_{j-1}}}, \big[\lambda_{\mu,min}, \lambda_{\mu,max}\big]\Big) \tag{13}$$

where, $\lambda_{\mu,min}$, $\lambda_{\mu,max}$ are lower and upper bounds we allow for the EWMA mean-update rate. Var_{j-1} refers to the EWMA estimate of the variance at the previous step. It measures how 'noisy' the process has been up to $j-1$. Clipping is used to ensure numerical safety.

This dynamic adjustment allows the EWMA to be self-tuning, offering robustness in varying operation conditions and making the fault detection algorithm both sensitive to true faults and resistant to random fluctuations.

Step 3.5: Mean updating is the heart of the EWMA filter

$$\mu_j = \lambda_{\mu,j} \cdot n_j + \big(1 - \lambda_{\mu,j}\big) \cdot \mu_{j-1} \tag{14}$$

where, μ_j is the updated estimate of the faulty-free signal mean at the time step j. $\lambda_{\mu,j}$ refers to adaptive gain (forgetting factor) at the time step j, chosen between $\lambda_{\mu,min}$ and $\lambda_{\mu,max}$. A large $\lambda_{\mu,j}$ makes the filter respond more quickly to the new sample, while a smaller $\lambda_{\mu,j}$ makes it smoother and less reactive. $\big(1-\lambda_{\mu,j}\big)\cdot\mu_{j-1}$ is the 'memory' term that carries forward the previous estimate $\big(\mu_{j-1}\big)$ ensures that past data still influences the estimate, which prevents it from jumping wildly on every new sample.

This EWMA mean update offers a self-tuning compromise between reactivity and robustness, which is necessary for fault detection in real-world applications.

Step 3.6: Adaptive smoothing factor of variance

$$\lambda_{Var,j} = clip\Big(\lambda_{Var,min} + \big(\lambda_{Var,max} - \lambda_{Var,min}\big) \cdot \frac{|e_j|}{|e_j| + \sqrt{Var_{j-1}}}, \big[\lambda_{Var,min}, \lambda_{Var,max}\big]\Big) \tag{15}$$

$\lambda_{Var,j}$: the adaptive variance update rate (forgetting factor) at time j. Like its mean-update counter part $\lambda_{\mu,j}$. It is clipped between $\lambda_{Var,min}$ and $\lambda_{Var,max}$.

Step 3.7: Variance update

$$Var_j = \lambda_{var,j} e_j^2 + \big(1 - \lambda_{Var,j}\big) \cdot Var_{j-1} \tag{16}$$

Is the EWMA's way of tracking how noisy or variability of the signal is over time.

Where, Var_j is the updated estimate of variance at time j. It indicates how large the recent deviation from the mean has been. $\big(1 - \lambda_{Var,j}\big) \cdot Var_{j-1}$ is the 'memory' term carrying forward the previous variance estimate. It ensures pass variability still influences Var_j, preventing it from over-reacting to a single outlier.

In short, this EWMA variance update is what allows the fault detection algorithm to know how much natural fluctuations to expect at any time, information that directly shapes both the adaptive learning rate and the dynamic thresholds that raise a fault alarm.

Step 3.8: Adaptive window calculations.

Adaptive window size W_j improves responsiveness to changes in signal variance. This adaptive temporal scaling enables the fault detection approach to dynamically decide how far back in time to look, instead of using a fixed window.

calculated by

$$W_j = clip\left(W_{j-1} \cdot \sqrt{\frac{Var_{j-1}}{Var_j}}, W_{min}, W_{max}\right) \tag{17}$$

The window size W_j is dynamically constrained within bounds (W_{min}, W_{max}) which allows optimal balance between sensitivity and robustness. It controls how many past points are used to calculate the local peak-to-peak range P_j.

Where, W_{j-1}: the window size that was used at the previous time step j. This provides information about how many samples are considered when computing the last range P_{j-1}.

$\sqrt{\frac{Var_{j-1}}{Var_j}}$ is a volatility adjustment factor, where Var_{j-1} is the EWMA estimate of variance at the previous step and Var_j is the updated variance at this step. If the signal just becomes more reliable $(Var_j > Var_{j-1})$, then $\sqrt{\frac{Var_{j-1}}{Var_j}} < 1$, so the window shrinks, focusing on fewer recent points, because the aim is for the range to reflect the new conditions more responsively. Conversely, if the signal just becomes quieter $(Var_j < Var_{j-1})$, then $\sqrt{\frac{Var_{j-1}}{Var_j}} > 1$, so the window expands to average over more points and avoids overreacting to minor fluctuations.

This sub-step is vital during developing a fault detection algorithm as it enables the window size to widen or shrink in step with the observed variance, so it automatically tunes local range computation. Also, a smaller window prevents peak-to-peak range from being dominated by a few outliers in the noisy state, while a larger window smooths out the occasional blip as in the steady state of signals. In short, it balances noise and normal fluctuations in signals. Moreover, since the dual thresholds (U_j, L_j) depends on α_j, and P_j, making P_j itself adaptive helps to keep the fault alarm neither too twitchy (false positive) nor too sluggish (missed faults). Resulting in keeping the proposed fault detection approach agile.

Step 4: Dual thresholds determination.

Step 4.1: Peak-to-Peak computation.

The determination of dual thresholds employs peak-to-peak measurements within recent signal segments, quantifying the amplitude range within the adaptive windows.

Let n_j, detonate the resampled working segment. And let W_j the adaptive window size at index j. To avoid boundary issues at the beginning of the segment, we define the window start as $s_j = \max\{1,\ j - W_j + 1\}$ and the index set $\varphi_j = \{k : s_j \le k \le j\}$, with effective window size $w_j = j - s_j + 1 \in [1,\ W_j]$. Then local peak-to- peak is computed by:

$$P_j = \max_{k\in\varphi_j} n_k - \min_{k\in\varphi_j} n_k \tag{18}$$

Step 4.2: Proposed dual thresholds calculations

$$U_j = \mu_j + \alpha_j \cdot P_j \tag{19}$$

$$L_j = \mu_j - \alpha_j \cdot P_j \tag{20}$$

where U_j and L_j are the upper and lower thresholds, respectively at the time step j. μ_j is the EWMA estimate of fault-free state mean at time step j. P_j refers to local peak-to-peak range over the most recent W_j points. α_j is the half-range scaling factor (adaptive threshold factor) at time j, which selected and clipped on the ration of initial to current variance, computed by:

$$\alpha_j = clip\left(\alpha_0 \cdot \sqrt{\frac{Var_0}{Var_j}},\ [\alpha_{min},\ \alpha_{max}]\right) \tag{21}$$

where, α_0 is the nominal threshold scaling factor. Var_0 refers to the initial EWMA estimate of the signal variance, typically the variance computed over a seed window to the start of the segment. Var_j is the EWMA estimate of the variance at time j. α_{min}, α_{max} represnt the lower and upper threshold factors respectively.

Step 4.3: Station-specific dual thresholds.

In this sub-step, we consider the EWMA-drive thresholds to establish dual thresholds on each station's particular variability. Concretely, for the station s at each normalized index j

$$U_{base,\,j} = \frac{1}{K}\sum_{K=1}^{K} U_j^k \tag{22}$$

$$L_{base,\,j} = \frac{1}{K}\sum_{K=1}^{K} L_j^k \tag{23}$$

$$RMS_U(j) = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(U_j^k\right)^2} \tag{24}$$

$$RMS_L(j) = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(L_j^k\right)^2} \tag{25}$$

where, U_j^k, L_j^k are the dual (upper and lower) EWMA thresholds computed on the k$^{\text{th}}$ fault-free segment at index j. $U_{base,\,j}$, $L_{base,\,j}$ are their element-wise means. $RMS_U(j)$, $RMS_L(j)$ measure the typical magnitude of those thresholds (a Root Mean Square across cycles).

Then we apply a station-specific scale β_s^U, β_s^L to bump the band (dual thresholds) out or by a fraction of the RMS amplitude:

$$U_{ref,\,j}^{(s)} = U_{base,\,j} + \beta_s^U RMS_U(j) \tag{26}$$

$$L_{ref,\,j}^{(s)} = L_{base,\,j} - \beta_s^L RMS_L(j) \tag{27}$$

By doing this, it enables (1) capturing across-cycle variability; the plain mean $U_{base,\,j}$ ignores how spread out the individual is U_j^k were. By adding a fraction of their RMS, we inflate the band to account for that observed scatter. (2) specific station adjustment to sensitivity; several stations have different mechanical dynamics and sensor noise characteristics. Selecting β_s^U, β_s^L per station allows tuning each band's width detection rate and missed detection rates. (3) offering a robust baseline; as these are all established on EWMA thresholds and then adjusted by their RMS, the final $U_{ref,\,j}^{(s)}$, $L_{ref,\,j}^{(s)}$ blend local adaptivity and global robustness, providing a stable yet responsive reference band for fault detection.

Step 5:Test data analysis and fault detection.

This step aims to analyze and detect faults in new cycles by comparing against the reference model. To achieve this, several steps should be followed as extract and resample working segments from faulty cycles (as in step 1), calculating the mean μ^{test}, (as step 3), defining raw exceedance mask, identify contiguous exceedances intervals (j_s, j_e), time-gate intervals, and final fault detection decision.

Step 5.1: Raw exceedance mask.

In signal processing, a mask is simply a vector of 0 s and 1 s or (False and True) that selects certain elements of another sequence. Mathematically, it is an indicator function:

$$x_A(j) = \begin{cases} 1, \, if condition A holds at index j, \\ 0, \, otherwise \end{cases} \tag{28}$$

Applied to our problem, the mask offers exactly which sample points are exceeding their thresholds, then we define raw exceedance mask. At each resampled index j, we check whether the test EWMA mean μ^{test} lies outside the computed dual thresholds $[L_{ref}[j], U_{ref}[j]]$, by using the following formula:

$$R[j] = \left(\mu^{test}[j] > U_{ref}[j]\right) \vee \left(\mu^{test}[j] < L_{ref}[j]\right) \tag{29}$$

where μ^{test} refers to the EWMA mean of test data (possibly faulty) segment at point j. $U_{ref}[j]$, $L_{ref}[j]$ are the dual thresholds at point j, which built as explained in step 3.

This raw exceedance mask R (Boolean array) flags every sample at which the faulty signal breaches its dual threshold. It is the initial indicator of potential faults, without yet working about the duration.

Identify contiguous exceedances intervals (j_s, j_e)

This sub-step scans the binary vector R for run of consecutive 'true' values. Each run begins at an index j_s where $R[j_s - 1] =$ false but $R[j_s] =$ true, and ends at j_e where $R[j_e] = true$ but $R[j_e + 1] = false$. This sub-step is vital due to the faults are rarely single sample events. Groping adjacent exceedances enables treating them as coherent intervals.

Step 5.2 Time-Gate Intervals (TGI).

It is considered as an intelligent filtering technique to make the fault detection mechanism more reliable and operator-friendly by minimizing noise and signifying relevant faults. Once we get $R[j]$, and identified contiguous runs of intervals (j_s, j_e), where the signal remains outside the dual threshold. Denote these runs by their start and end indices:

$$\left(j_s^{(k)}, j_e^{(k)}\right), k = 1, \ldots, M$$

where k th run:

$j_s^{(k)}$ is the first index in that run (where $R[j_s] = 1$ and $R[j_s - 1] = 0$), and $j_e^{(k)}$ is the last index (where $R[j_e] = 1$ and $R[j_e + 1] = 0$).

Mapping indices to time

After that, we map indices to time, at each resampled index j corresponding to a real time via

$$t_{res}[j] = t_0 + (t_N - t_0)\frac{j}{N-1} \tag{30}$$

where $t_{res}[j]$ indicates linear mapping from index to actual time that preserving the true duration of each interval,t_0 is the actual timestamp of the first working segment point, t_N refers to the timestamp of the last working segment point and N is the fixed resampling length. Thus, the duration of the k th exceedance interval is.

$$\Delta t^{(k)} = t_{res}\left[j_s^{(k)}\right] - t_{res}\left[j_e^{(k)}\right] \tag{31}$$

We introduce a time-gate parameter (T_{min}) which is the minimum acceptable fault duration 0.01 (s). this reflects the domain-specific notion that any excursion shorter than T_{min} is just noise or a harmless transient.

Step 5.3: Fault detection decision.

In this last step, we rebuild a clean Boolean mask $P[j]$ that is true exactly on those indices belonging to one of the time-gated intervals.

$$P[j] = \begin{cases} True, \, \exists (j_s, \, j_e) with j_s \le j \le j_e, \\ False, \, otherwise \end{cases} \tag{32}$$

It reflects only the sustained, time-qualified faults and thus the proposed fault detection algorithm's actual decision about the faulty and healthy states at each resampled point.

In the implementation of the proposed approach, the raw exceedance mask $R[j]$ acts the role of an internal reference for potential faults. Due to it marks every time sample where the EWMA-filtered current exceeds the adaptive dual thresholds. Then the time-gated decision mask $P[j]$ is obtained by retaining only those contiguous runs of $R[j] = 1$ whose physical duration exceeds Δt_{min}. Therefore, the decision set is a subset of the exceedance set.

$$\{j : P[j] = 1\} \subseteq \{j : R[j] = 1\} \tag{33}$$

Hnece, the TGI (Step 5.2) cannot make new alarms at points where the signal is inside the dual thresholds. However, it can only eliminate short alarms. When calculating sample-wise confusion matric using $R[j]$ as ground truth and $P[j]$ as the detector, the number of false positives is identically zero and all losses in sensitivity appear as missed detections of very short alarms. This design choice reflects our intent; the time gate plays as a filter that suppresses short and noisy alarms yet does not introduce new exceedances that are not already presented at the raw threshold level.

In the context of engineering justification of selecting $\Delta t_{min} = 0.01s$, due to all workstations are interconnected and sharing the same conveyor cycle where the sampling frequency $f_s = 100Hz$. For this reason, the Δt_{min} is defined at the system level instead of per station. This time-gate is applied to the contiguous threshold-exceedance intervals derived from the raw mask $R[j]$. Only intervals whose duration exceeds the Δt_{min} are promoted to fault events, and the corresponding indices marked as true in the final decision mask $P[j]$. The rational is that electrical noise and switching spikes may cause very short threshold crossings but do not persist over multiple samples, whereas mechanically meaningful faults produce sustained current deviations lasting several hundreds of milliseconds to seconds. Therefore, choosing $\Delta t_{min} = 0.01s$ is conservative, which is strictly above zero, preventing instantaneous artifacts from being interpreted as events, while remaining much shorter than the shortest fault intervals observed in the measured current profiles.

To evaluate the performance of the proposed fault detection approach, several evaluation metrics are quantified, such as accuracy, Detection Rate (DR), Missed Detection Rate (MDR), False Positive Rate (FPR), and Time to Detect (TTD). For detection issues, the number of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) is utilized to calculate the performance measures. The performance evaluations are calculated through (34) to (38)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{34}$$

$$DR = \frac{TP}{TP + FN} \tag{35}$$

$$MDR = \frac{FN}{FN + TP} \tag{36}$$

$$FPR = \frac{FP}{FP + TN} \tag{37}$$

$$TTD = t_1 - t_{f1} \tag{38}$$

where, t_1 indicates the time which the fault was detected, t_{f1} is the actual occurrence time of the fault. This end-to-end workflow integrates adaptive thresholding, rigorous fault validation, and intuitive visualization, offering a scalable solution for real-time fault detection in industrial systems.

Case study description

System overview

The case study investigates a fault detection framework deployed in ICPS, particularly a conveyor-based sorting system that operates on a closed-loop conveyor architecture as shown in Fig. 2. This system uses motor-driven belts, along with stop gates and lifting/transverse mechanisms, to move belts with carriers between workstations. The system is equipped with position sensors and RFID, which are installed on carriers. RFID readers provide feedback when a specific carrier arrives at each workstation. For the communication of the conveyor with the ERP system, Ethernet cables (orange color) are used. To connect the PC and the Wi-Fi router, other cables are included (gray color).

The process initiates at a selection station (1) where black and white workpieces are individually loaded onto RFID-tagged carriers based on production orders issued by the Manufacturing Execution System (MES). These carriers are then routed via a bypass conveyor to a camera-based inspection station (2), which determines part color. Upon identification, black workpieces are immediately redirected through the bypass conveyor toward the final handling station (5), while white workpieces continue to a milling station (4) for further processing before re-entering the bypass line and proceeding to the handling station. This handling station serves as the system's terminal point, where parts are either prepared for dispatch or subjected to final operations. The entire process is characterized by real-time part identification, automated routing, and inter-station communication, which collectively facilitate autonomous part sorting and adaptive workflow control within a cyber-physical manufacturing environment.

As displayed in Fig. 3, ten runs were performed under nominal conditions, the current amplitudes exhibit clear non-stationary behavior, with low-frequency oscillations drifting. Furthermore, the overall cycle duration varies approximately 37 to 48 (s).

This high dynamic range, characterized by rapid transients, amplitude modulation, and variable timing, renders fixed-threshold detection schemes inadequate and motivates the development of adaptive multi-scale threshold capable of tracking the system's evolving baseline behavior.

Dataset and labeling protocol

The experimental dataset is acquired on the lab-scale conveyor-based sorting system as described in system overview sub-section.

The whole dataset includes 13 cycles. We use 70% for training the composite reference model, 15% for validation

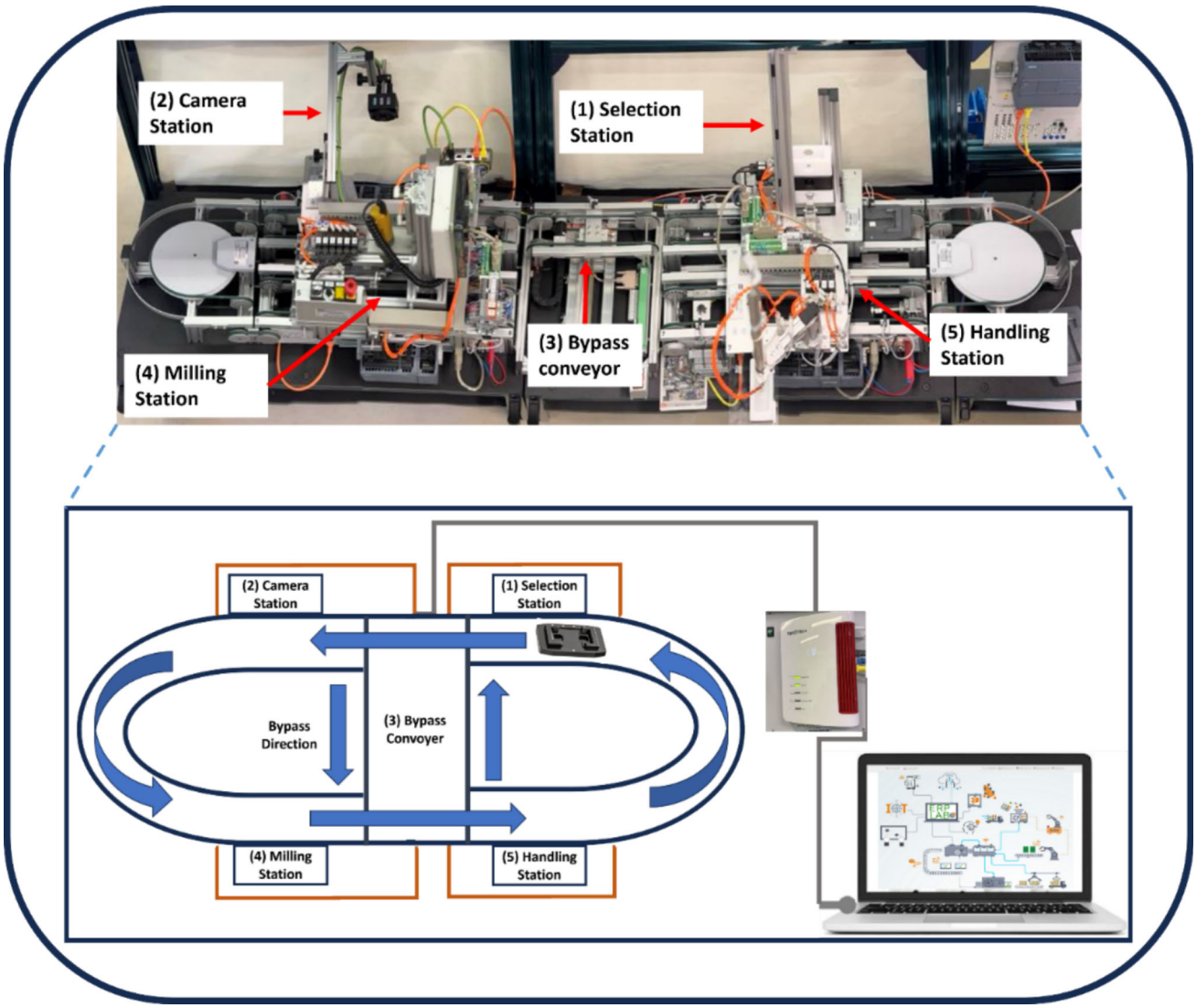

Fig. 2 Layout of the conveyor-based sorting system

and to ensure the proposed fault detection approach performs correctly without false alarms or missed detection. Then, we use 15% for testing and evaluation of the proposed fault detection. The test data the evaluation is performed on the faulty datasets as a test dataset are never used in the construction the reference model or parameter tuning. All performance metrics reported in this article are calculated on this test set.

Under nominal conditions, ten complete fault-free cycles are recorded. Each cycle corresponds to one closed-loop transversal of RFID-tagged carriers through all stations without visible faults or alarms. For each cycle and each station, the working segment is extracted based on the position sensor (Steps 1.1–1.2), filter out low-current idle periods using the station-specific active threshold and resample the resulting segment to a fixed length of $N = 100$ samples (step 1.3). This yields ten fault-free working segments per station.

Additionally, three spontaneous fault scenarios are recorded as explained in the next section entitled 'Fault scenarios analysis and operational symptoms". For each scenario, the working segments are extracted and resampled and labelled as faulty at station level. This gives three faulty working segments per station, one per scenario.

The current signals are acquired at a sampling rate of $f_s = 100Hz$. The 'time' column in the raw data corresponds to this acquisition clock. All durations that are reported in this article in seconds.

The adaptive healthy references (EWMA mean and dual thresholds) are estimated only from healthy data. Particularly we use ten healthy cycles to construct the composite reference model and three healthy cycles to validate it and to ensure the proposed fault detection approach performs correctly without false alarms or missed detection. Then, the evaluation is performed on the faulty datasets as a test data

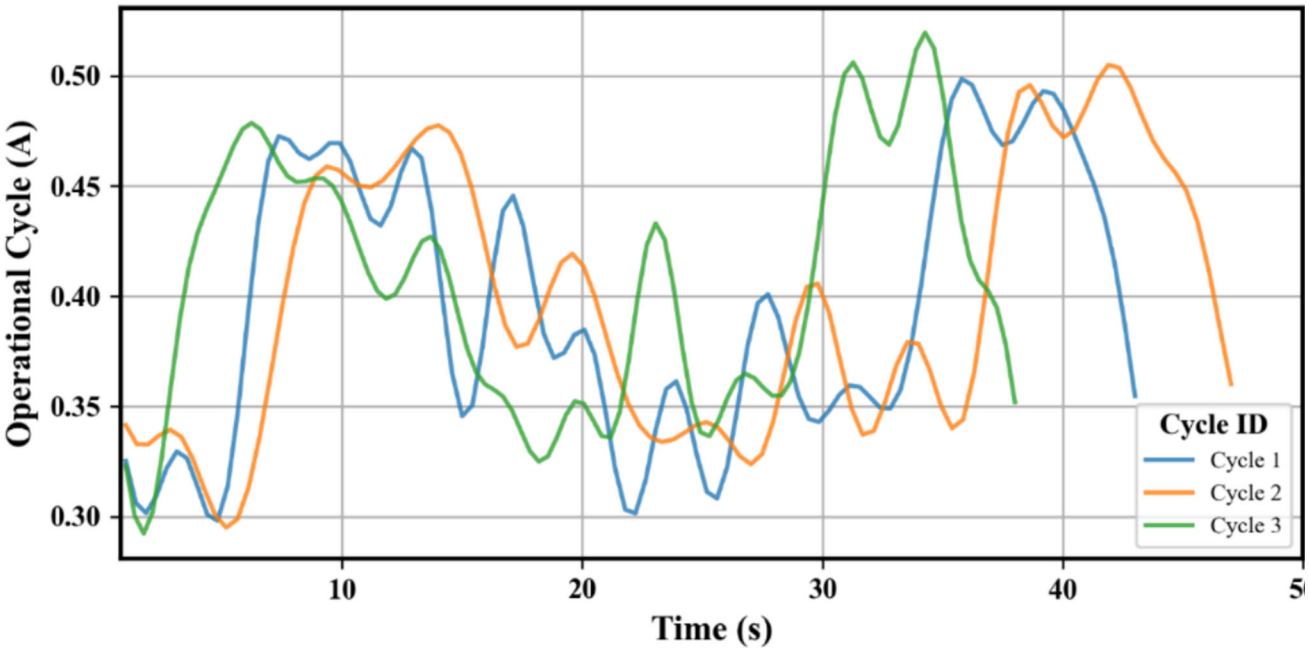

Fig. 3 Multiple operational cycles in nominal conditions

set which is never used during training or validation. All performance metrics reported in this article are calculated on this test set.

Fault scenarios analysis and operational symptoms

In this study, spontaneous faults were observed and recorded at handling stations, camera inspection stations, and milling stations. Three representative fault scenarios were observed during system operation as follows.

Scenario 1: Operational blockage without handling station processing.

Under normal conditions, the stopper beneath the handling station is designed to temporarily halt the carrier with the workpiece, initiating the handling station's operational cycle. After successful processing, the stopper releases the carrier, allowing it to return autonomously to the initial selection station for subsequent operations. However, the observed fault scenario has critical operational symptoms; the stopper beneath the handling station blocks the carrier with the workpiece for a longer and unintended duration as displayed in Fig. 4. Crucially, the handling station does not initiate its intended processing cycle despite the carrier being correctly positioned. Following this anomalous blockage, the stopper spontaneously releases the carrier, allowing it to continue along the conveyor belt. Subsequently, the carrier encounters another unintended blockage at the position of the sensor located on the conveyor segment within the handling station. This additional blockage temporarily halts carrier movement again, causing another brief operational disruption. Eventually, the blockage spontaneously clears, resulting in the carrier moving back toward the initial selection station without undergoing the intended handling process. This fault scenario results in incomplete operations and inefficient system behavior.

The fault is categorized as intermittent mechanical faults, characterized by temporary and spontaneous disruptions in normal operations. Such faults are particularly challenging due to their transient nature, leading to complicating FDD.

Scenario 2: Belt aging and progressive wear at the camera inspection station.

Practically, prolonged operational cycles at the camera inspection station gradually induce mechanical degradation, primarily due to aging conveyor belts and progressive wear in mechanical components such as rollers and bearings. Over time, these conditions lead to increased frictional resistance and intermittent belt misalignments, causing subtle yet measurable fluctuations in conveyor performance. Operational symptoms include minor, intermittent vibrations, occasional squeaking noises, and a gradual, albeit initially imperceptible, decline in conveyor smoothness and efficiency. This progressive degradation scenario can be classified as an incipient intermittent fault, as it involves slow, sporadic deterioration, highlighting the necessity of sensitive fault detection methods to facilitate timely intervention and mitigate the risk of more significant system disruptions.

Scenario 3: Transient mechanical disturbance in the milling station.

Practically, in the current laboratory-scale setup, the milling station operates in a simulated mode, meaning the

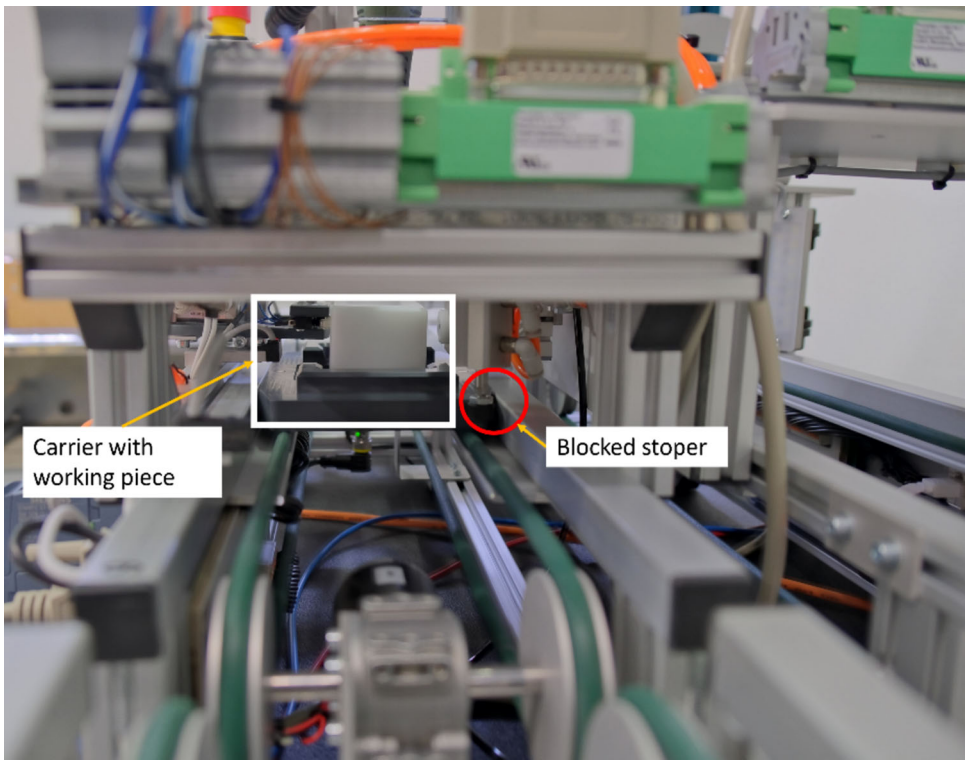

Fig. 4 Mechanical blockage of the stopper at the handling station

spindle and motion systems are active, but no actual material removal occurs on the workpiece. Despite the absence of real cutting, the mechanical components, including the spindle motor, drive system, and conveyor mechanisms, still undergo dynamic motion cycles. During operation, transient mechanical disturbances may arise due to momentary fluctuations in belt tension, slight imbalances in spindle rotation, or control signal inconsistencies, particularly during acceleration or deceleration phases. Operational symptoms include sudden, short duration drops in motor current, as evidenced by abrupt deviations below the lower threshold in the fault detection profile. These disturbances are typically not accompanied by visible mechanical failure but are detectable through sensitive current monitoring. Accordingly, this fault is categorized as an abrupt intermittent fault, reflecting its sudden, irregular occurrence without sustained degradation. Detecting such anomalies is essential in simulated environments to validate system robustness and ensure reliable integration of mechatronic components within cyber-physical manufacturing architectures.

Results and discussion

This section presents and interprets the results of fault detection experiments conducted on a laboratory-scale conveyor-based sorting manufacturing system, focusing on three critical stations: the handling station, the camera inspection station, and the milling station. The evaluation metrics include DR, MDR, FPR, detection delay, and overall accuracy. These evaluations were conducted using a computer characterized by a processor of 12th generation Intel vPro Enterprise with Intel Core i9-12900H with speed of 2.5 GHz; GPU Intel Iris Xe Graphics and Microsoft Windows 10 Enterprise.

In Scenario 1, the fault detection algorithm achieved exceptional performance, with a detection rate of 0.986 and zero false positives, while maintaining a missed detection rate of only 0.014 and no detection delay. This scenario involved a mechanical anomaly where the stopper released the carrier without triggering the expected handling operation. The algorithm accurately distinguished this anomaly due to the sustained deviation in current below the lower threshold, highlighting its effectiveness in capturing significant behavioral shifts related to missed operations as display in Fig. 5. The perfect timing (+ 0.00 delay) indicates a rapid response capability crucial for real-time systems. Notably, the variation in the proposed dual threshold tolerance indicates an adaptive mechanism which increases wariness in a highly dynamic region where faults are critical, by narrowing the tolerance width. Whereas it widens in stable regions to minimize false alarms which are caused by small fluctuations.

In Scenario 2, the system encountered a subtler, incipient intermittent fault, reflecting gradual performance degradation due to mechanical wear. Despite the less abrupt nature

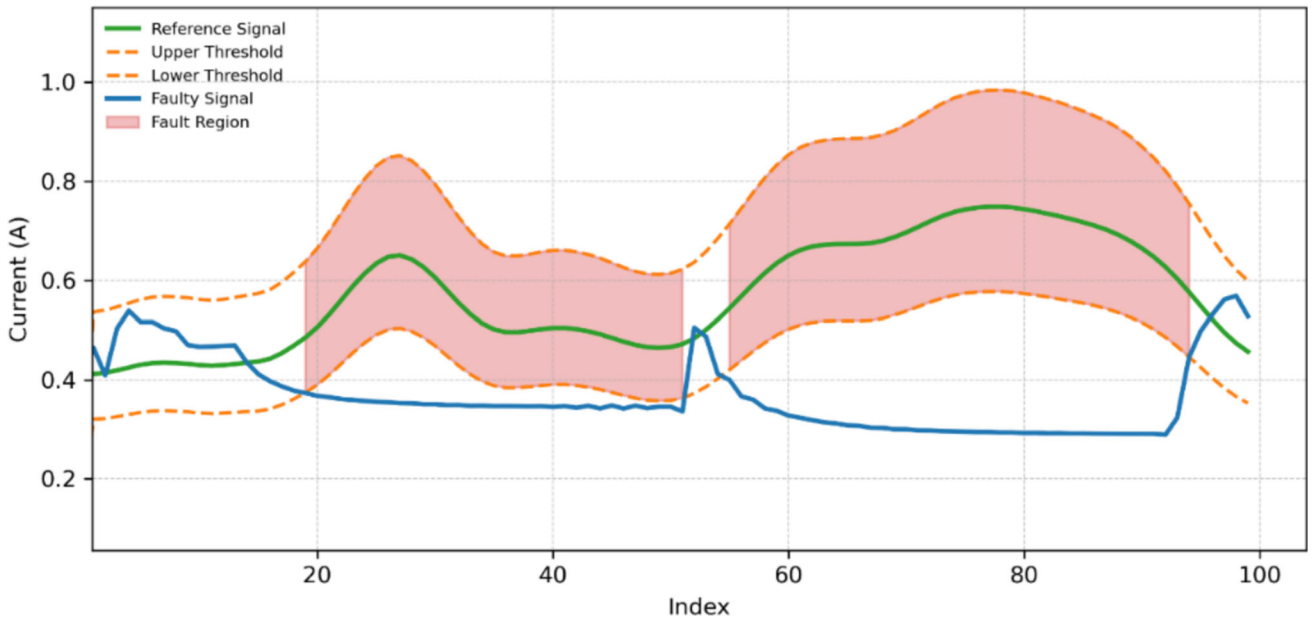

Fig. 5 Fault detection for handling station current signal

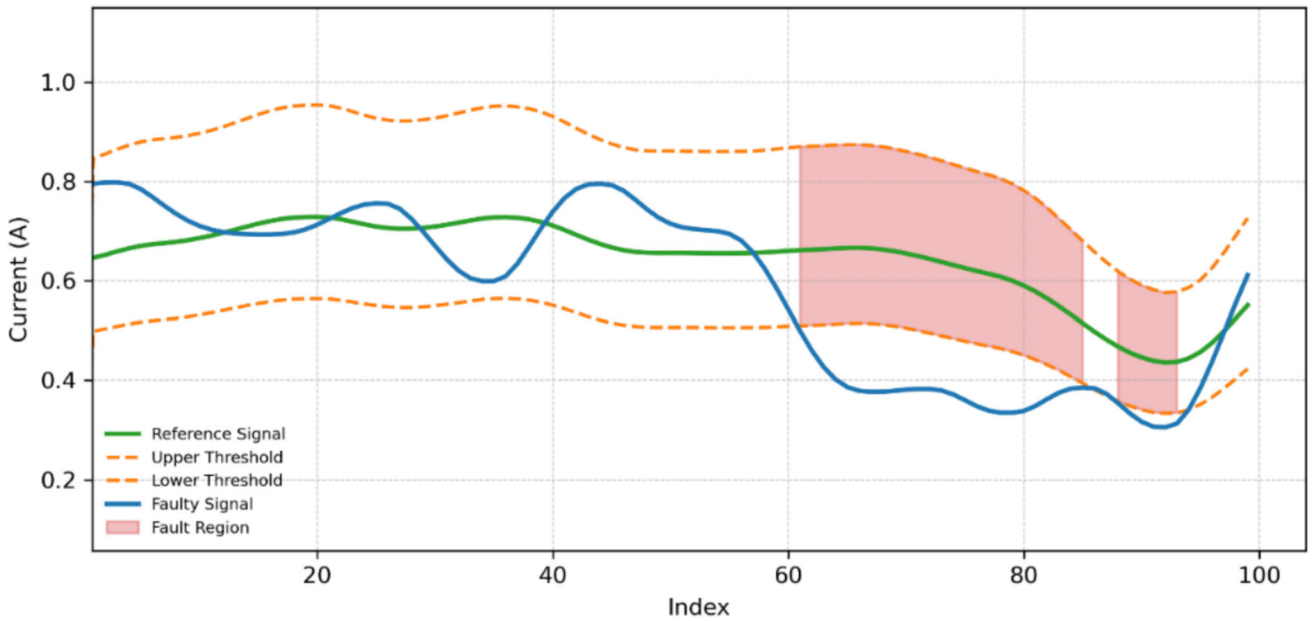

Fig. 6 Fault detection for camera inspection station

of this fault, the algorithm maintained a high detection rate of 0.969 with no false positives, and a missed detection rate of 0.031, again with no delay. This result validates the algorithm's sensitivity to low-amplitude, slowly evolving faults as shown in Fig. 6. A critical requirement in fault detection and early intervention strategies. The high accuracy of 0.99 in this context affirms that the method can reliably distinguish true faults from normal process variability, even in ambiguous signal regimes.

In Scenario 3, the faulty events were more abrupt and short-lived as display in Fig. 7, corresponding to brief disruptions during the simulated milling operation. The algorithm successfully detected these faults with a detection rate of 0.833, and a missed detection rate of 0.167, while preserving a zero false positive rate and + 0.00 detection delay. Though slightly lower in sensitivity compared to the previous scenarios, this performance remains strong given the transient nature and brief duration of the anomaly. The accuracy

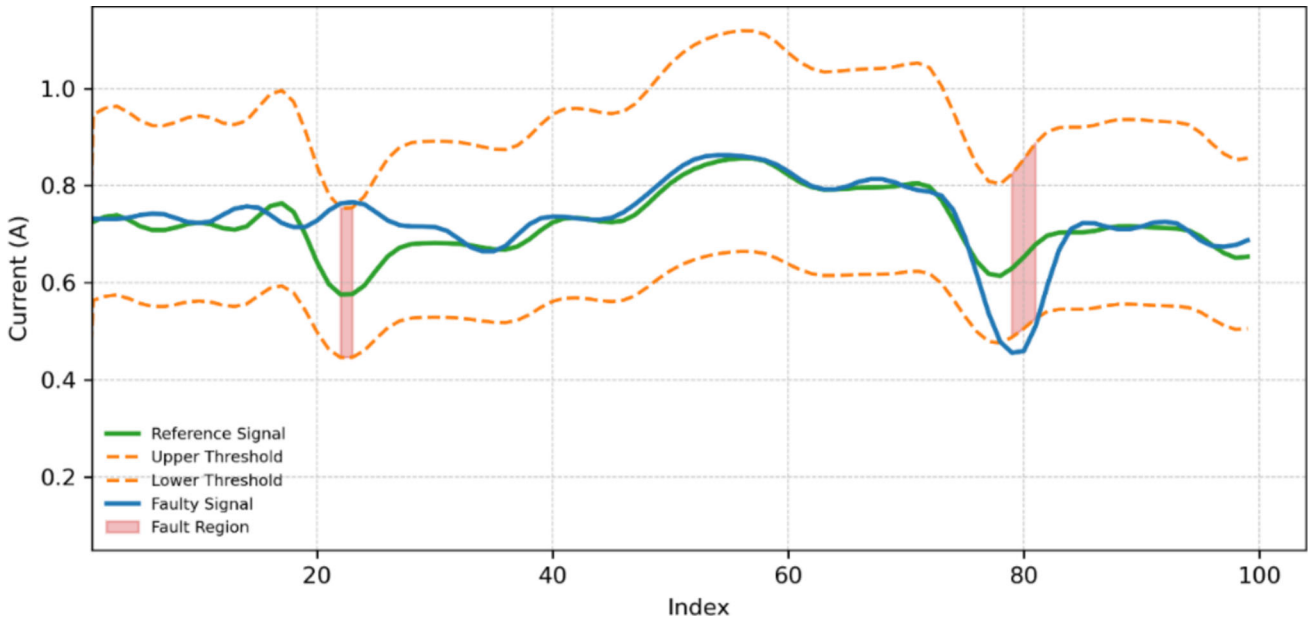

Fig. 7 Fault detection for milling station

remained consistent at 0.99, suggesting robust generalization across fault types. These results indicate that while threshold-based methods may exhibit limitations with high-frequency, short-duration faults, they still provide actionable early warnings in cyber-physical environments when tuned appropriately.

The summarized performance metrics for all three fault scenarios are presented in Table 2. This table highlights the detection consistency of the proposed approach across different fault types, ranging from severe, process-level disruptions to subtle incipient degradations. High detection rates and low missed detection rates were achieved across all scenarios, with perfect detection timing and no false alarms, reinforcing the robustness and practical applicability of the approach in real-time fault monitoring within cyber-physical production environments.

Overall, the results demonstrate that the proposed fault detection approach, based on adaptive thresholding of current signals, is highly effective across diverse fault types, including abrupt, intermittent, and incipient anomalies. The consistently zero false positive rates and accuracies of 0.99 across all scenarios confirm the method's practical reliability in real-time industrial settings. Furthermore, the sensitivity to subtle degradation, as observed in the camera station, and to brief disruptions, as seen in the milling station, underscores the adaptability of the detection framework in dynamic cyber-physical manufacturing systems. These findings support the integration of lightweight, signal-based diagnostic tools for early fault identification and resilient process control in Industry 4.0 environments.

Comparative study with global method (EWMA), local method (Peak-to-peak), and the proposed approach

To further validate the resilience and generalization capabilities of the proposed approach, a comparative study is conducted against two widely adopted methods, EWMA and P2P methods. Firstly, the evaluation performance was conducted under nominal (fault-free) operation conditions. The results are summarized in Table 3. Notably under nominal operation conditions, DR, MDR and TTD cannot be defined because no true faults occur. Hence, for separate evaluations under nominal conditions, we focus on robust metrics including FPR and accuracy using the healthy labels as ground truth.

Secondly, the performance evaluation was conducted under the three fault scenarios in the stations of handling, camera inspection, and milling stations. The results displayed in Fig. 8 for EWMA, and Fig. 9 for P2P. Figure 10 summarizes the performance comparison of EWMA, P2P and the proposed fault detection. For all evaluation metrics, we report the mean and the associated 95% confidence interval. Additionally, a paired Wilcoxon signed-rank tests were performed between the proposed approach and each baseline for every scenario, and we report improvements as statistically significant ($p < 0.05$).

These evaluations were conducted using a computer characterized by a processor of 12^{th} generation Intel vPro Enterprise with Intel Core i9-12900H with speed of 2.5 GHz; GPU Intel Iris Xe Graphics and Microsoft Windows 10 Enterprise. For real-time deployment, the runtime overhead per

Table 2 Summary of fault detection performance across scenarios

Scenario ID	Station	Fault description	Fault type	DR	MDR	FPR	TTD (s)	Accuracy
Scenario 1	Handling station	Operational blockage without handling processing	Intermittent	0.986	0.014	0.00	+ 0.00	0.99
Scenario 2	Camera station	Belt aging and progressive wear	Combined (incipient with intermittent)	0.969	0.031	0.00	+ 0.00	0.99
Scenario 3	Milling station	Transient mechanical disturbance	Combined (abrupt with intermittent)	0.833	0.167	0.00	+ 0.00	0.99

Table 3 Performance results under nominal operating conditions

Approach	Camera station					Milling station					Handling station				
	DR	MDR	FPR	TTD	Acc	DR	MDR	FPR	TTD	Acc	DR	MDR	FPR	TTD	Acc
Proposed	N/A	N/A	0.00	N/A	1	N/A	N/A	0.00	N/A	1	N/A	N/A	0.00	N/A	1
EWMA	N/A	N/A	0.01	N/A	0.99	N/A	N/A	0.01	N/A	0.99	N/A	N/A	0.01	N/A	0.99
P2P	N/A	N/A	0.63	N/A	0.37	N/A	N/A	0.12	N/A	0.88	0.00	N/A	0.75	N/A	0.25

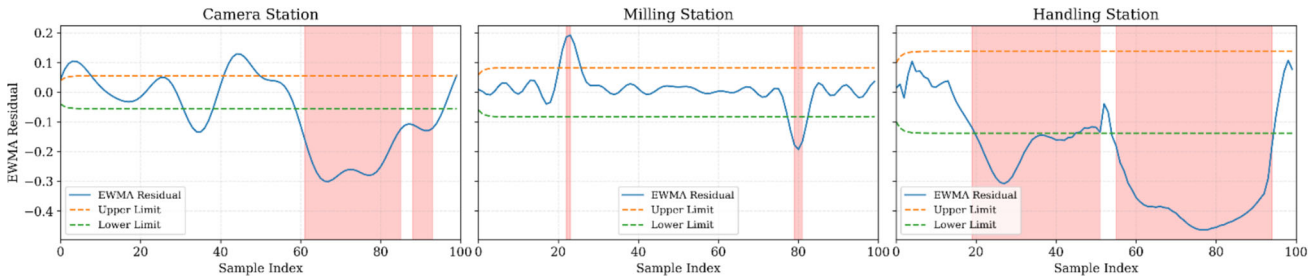

Fig. 8 Fault detection using residual based EWMA

cycle of the proposed approach compared to baseline EWMA and P2P methods is illustrated in Table 4

Scenario 1: Operational blockage without handling processing.

In this scenario, the proposed approach achieved a DR of 98.6% and an accuracy of 99%, outperforming EWMA (DR = 89%, Accuracy = 91%) and P2P (DR = 16%, Accuracy = 22%). Notably, the proposed method completely reduced false positives (FPR = 0%), in contrast to the 3.7% FPR of EWMA and the significantly higher 56% FPR of P2P. This reflects a 100% reduction in false alarms relative to P2P and a 97.3% reduction compared to EWMA. The MDR of the proposed method was only 1.4%, representing an 87.3%

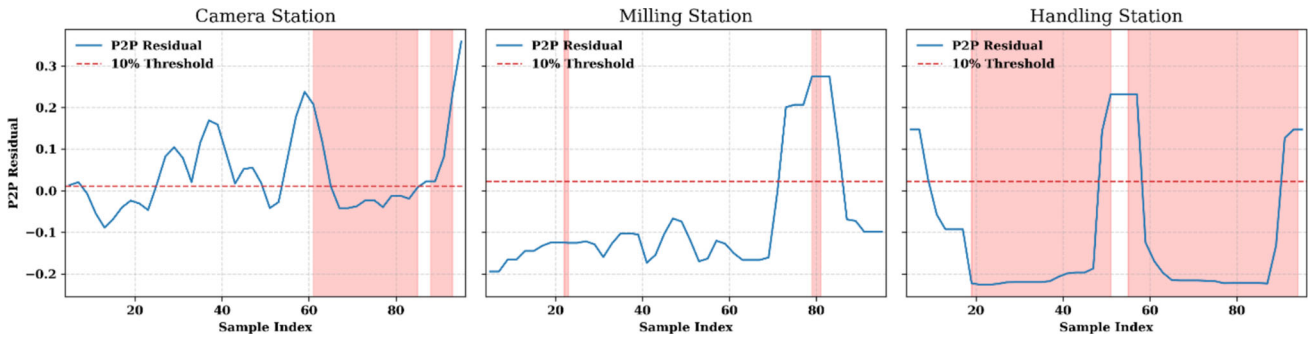

Fig. 9 Fault detection using residual based P2P

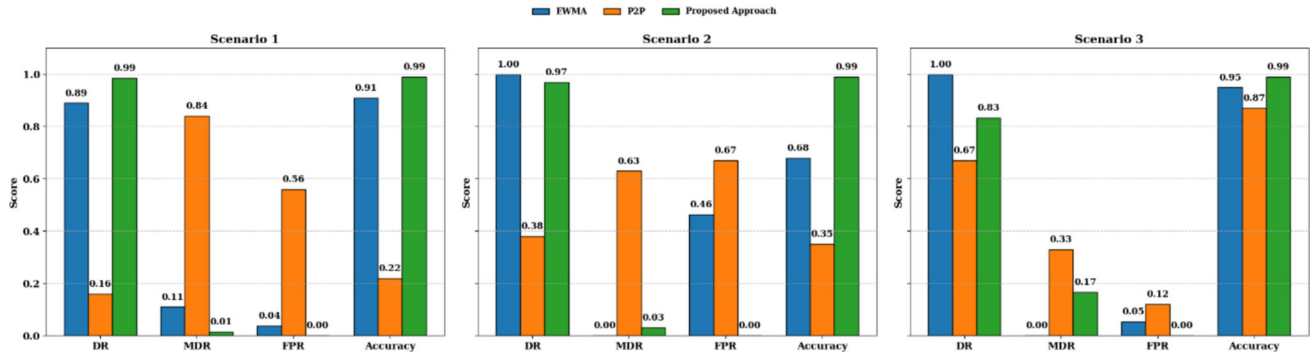

Fig. 10 Performance comparison of EWMA, P2P and proposed method across scenarios

Table 4 Runtime overhead per cycle

Approach	Runtime overhead per cycle (s)	Average runtime per station (s)
EWMA	1.916	0.639
P2P	1.649	0.549
Proposed approach	2.846	0.948

reduction compared to EWMA (MDR = 11%) and a remarkable 98.3% reduction compared to P2P (MDR = 84%). These results underscore the robustness of the proposed system in accurately identifying faults while avoiding false alerts.

Scenario 2: Belt aging and progressive wear at camera station.

This scenario presents a more challenging environment due to higher signal variability. The proposed approach still maintained near-optimal performance with an accuracy of 99% and zero false positives. While EWMA achieved a perfect DR of 100%, it did so at the cost of a high FPR (46.4%), leading to frequent false alarms. Conversely, P2P displayed poor sensitivity with a DR of 83% and the highest FPR of 67%. The proposed method not only balanced these trade-offs but also outperformed both global and local methods in terms of overall reliability, achieving an MDR of just 3.1%. This illustrates the method's resilience in distinguishing between true and spurious anomalies under fluctuating operational conditions.

Scenario 3: Transient mechanical disturbance in milling station.

In this scenario, which is characterized by periodic disturbances and dynamic load changes, the proposed method achieved an accuracy of 99%, a DR of 83.3%, and remain maintained a zero FPR. While EWMA achieved a slightly higher DR of 100%, it introduced a non-negligible FPR of 5.3%, which can be problematic in industrial settings with strict false alarm tolerance. P2P, on the other hand, reached only 67% DR and a 12% FPR. The proposed method achieved a 49.4% reduction in MDR compared to P2P, while simultaneously eliminating all false alarms. Resulting a clear indication of its robustness and precision balance.

Overall, the proposed method demonstrates exceptional resilience and generalization capability across all scenarios. It consistently maintains high detection sensitivity while minimizing both false alarms and missed detections, a balance that neither EWMA nor P2P could achieve. Unlike EWMA, which often struggles for sensitivity, or P2P, which fails under dynamic, varying conditions. The proposed approach presents a comprehensive and scalable solution for real-time fault detection in SM environments. Its consistent ability to achieve near-zero FPR and MDR positions it as a high-confidence, deployable approach for I 4.0 applications, where early, accurate, and reliable detection is critical for maintaining system availability, minimizing downtime, and ensuring process safety.

Ablation studies of the proposed fault detection approach

Ablation studies were conducted over seven configurations, where each algorithmic component was quantified through these confiscations as described in Table 5. The proposed approach (Config 1) combines four elements (adaptive λ, adaptive window, station specific β and time-gated). Config 2 removes only the temporal coherence filter (time-gated). Configs 3 and 6 replace the adaptive window with fixed length (W = 20, and W = 50, respectively). Config 4 disables the station-specific β. Config 5 uses fixed λ, and Config 7 plays as a minimal baseline with fixed λ, fixed W = 50, no station specific β and no time-gating.

The quantitative results (Table 6) show that the proposed approach (Config 1) achieves high detection performance across all stations (DR = 0.969, 0.833 AND 0.986) for camera inspection, milling and handling station, respectively. With low MDR $\leq$ 0.167, zero FPR and overall accuracy of 0.99. The time to detect is essentially zero in all stations, suggesting that once an interval is accepted as a fault, the

Table 5 Ablation configurations of the proposed approach

Configuration	Description	Adaptive λ	Adaptive window	Station-specific β	Time-gated	Fixed window
Config 1	Proposed approach	✓	✓	✓	✓	N/A
Config 2	Without time-gating	✓	✓	✓	x	N/A
Config 3	Without adaptive window (fixed W = 20)	✓	x	✓	✓	20
Config 4	Without station-specific β	✓	✓	x	✓	N/A
Config 5	Without adaptive λ	x	✓	✓	✓	N/A
Config 6	Without adaptive window (fixed W = 50)	✓	x	✓	✓	50
Config 7	Minimal baseline (fixed λ, W = 50, no β, no gating)	x	x	x	x	50

Table 6 Ablation studies evaluation results

Configuration	Camera station					Milling station					Handling station				
	DR	MDR	FPR	TTD	Acc	DR	MDR	FPR	TTD	Acc	DR	MDR	FPR	TTD	Acc
Config 1	0.969	0.031	0.00	+ 0.00	0.99	0.833	0.167	0.00	+ 0.00	0.99	0.986	0.014	0.00	+ 0.00	0.99
Config 2	1.00	0.00	0.00	+ 0.00	1.00	1.00	0.00	0.00	+ 0.00	1.00	1.00	0.00	0.00	+ 0.00	1.00
Config 3	0.968	0.091	0.04	+ 0.00	0.96	0.823	0.18	0.02	+ 0.82	0.97	0.962	0.027	0.03	+ 0.92	0.97
Config 4	0.99	0.00	0.94	+ 0.00	0.36	0.99	0.00	0.777	+ 0.00	0.27	0.99	0.00	0.962	+ 0.00	0.75
Config 5	0.937	0.073	0.04	+ 1.09	0.97	0.5	0.5	0.00	+ 21.55	0.97	0.96	0.024	0.05	+ 0.00	0.96
Config 6	0.968	0.051	0.06	+ 0.00	0.99	0.793	0.217	0.07	+ 6.00	0.95	0.95	0.04	0.03	+ 0.92	0.97
Config 7	1.00	0.00	0.94	+ 0.00	0.36	1.00	0.00	0.72	+ 0.00	0.33	1.00	0.00	1.00	+ 0.00	0.74

alarm is raised from its first out-of-the band sample. These values confirm that our proposed fault detection is sensitive to persistent faults while remaining conservative with respect to spurious fluctuations.

Role of temporal coherence (Config 1 compared to Config 2)

In the ablation studies, the ground truth is defined from instantaneous threshold exceedances of the proposed approach. A sample is labelled faulty whenever the EWMA mean exceeds the adaptive dual thresholds. Config 2 utilizes the same thresholds and EWMA parameters as Config 1 (the proposed approach) yet omits the temporal coherence filter (Time-gated) Δt_{min}. Therefore, its prediction mask coincides typically with this instantaneous ground truth. This leads to DR $= 1$, MDR $= 0$, FPR $= 0$ and Accuracy $= 1$ by construction for all stations.

In comparison to Config 1 which presents Δt_{min}. To enforce temporal persistence (out-of-band intervals shorter than 0.01 s are discarded as non-faulty transients), while longer intervals are retained. The slightly lower DR and non-zero MDR as reported in Config1 therefore reflect intentionally ignored micro-events rather than undetected meaningful faults.

From an engineering perspective, Config 1 provides a more realistic and robust behavior for online deployment; all persistent faults are still detected with essentially zero TTD, and the FPR remains zero in all stations. In the other hand, Config 2 would rise alarms for every short spike and is hence overly sensitive for use in an industrial environment.

Effect of adaptive window, adaptive λ and station-specific β

Disabling the adaptive window (Config 3 and 6) minimizing the robustness. Suggesting that a fixed window unable simultaneous handle dynamic operating phases. Eliminating the station-specific (Config 4) or considering the minimal baseline (Config 7) results in a high detection rate, yet at the cost of extremely high FPR and poor accuracy due to the large portions of the healthy working segments are misclassified as faulty. Finally, fixing adaptive λ (Config 5) degrades the performance of the milling station, demonstrating the importance of adaptation for tracking slow operating point changes without losing responsiveness to sudden faults.

Overall, the ablation study demonstrates that the combination of adaptive λ, adaptive window, station-specific thresholds and temporal coherence (the proposed approach) offer the best trade-off between sensitivity, robustness, and practical deployment.

Ablation study of Δt_{min} across stations

A systematic ablation study of $\Delta t_{\min}$ is conducted over $\Delta t_{\min} \in [0.0 - 10.00]$ s. across camera inspection, milling and handling stations and re-calculating the evaluation metrics involving accuracy, DR, MDR, PFR and TTD for the studied stations as displayed in Fig. 11. The handling station exhibits long fault intervals (see Fig. 5); therefore, its evaluation metrics are steady with respect to $\Delta t_{\min}$. In comparison to camera inspection and milling stations show shorter and intermittent deviations consistent with an incipient fault (see Figs. 6 and 7). For these stations, increasing $\Delta t_{\min}$ to values comparable with the duration of these intervals leads to a gradual reduction of the DR and increasing of the MDR. While the TTD remains close to zero due to the alarm is always raised at the first out-of-the band sample of each acceptable interval. Also FPR is observed in all scenarios is zero with the design of the time-gated dual thresholds, which are tuned so that nominal cycles rarely generate sustained variations above $\Delta t_{\min}$ or below noise-induced crossings are filtered out by construction and do not generate alarms.

From engineering perspective, all stations are interconnected and sharing the same acquisition chain and are synchronized by the same conveyor cycle. Therefore, we selected a single $\Delta t_{\min} = 0.01$ s at the system level according to: (a) the current signals are acquired at a sensor sampling frequency $f_s = 100Hz$ to ensure the time-gate is longer than sensor noise or switching spikes. (b) It remains an order of magnitude shorter than the fastest mechanically meaningful transients in the system, so all fault regimes of interest are preserved.

These results support the selection of $\Delta t_{\min} = 0.01s$ as a proper system-level gate lies safely above electrical noise and still remains orders of magnitude smaller than the mechanical fault durations across all interconnected stations that share the same conveyor cycle.

This selection is consistent with standard alarm management practices, where minimum alarm durations are dimensioned to be several sampling periods yet significantly shorter than the relevant mechanical time constants, to suppress nuisance alarms without making true process faults.

Quantitative comparison study with current state-of-the-art methods

A quantitative comparison study is conducted with the threshold-based fault detection approach form the literature emphasizes the advantages of the proposed adaptive multi-scale fault detection approach as displayed in Table 7. Amin et al., (2019), applied static and dynamic threshold to a continuous stirred-tank heater achieved DR of 36% and 78% respectively. While FPR of the dynamic threshold results 1.75% compared with 0.25% for static threshold. Attar et al.,

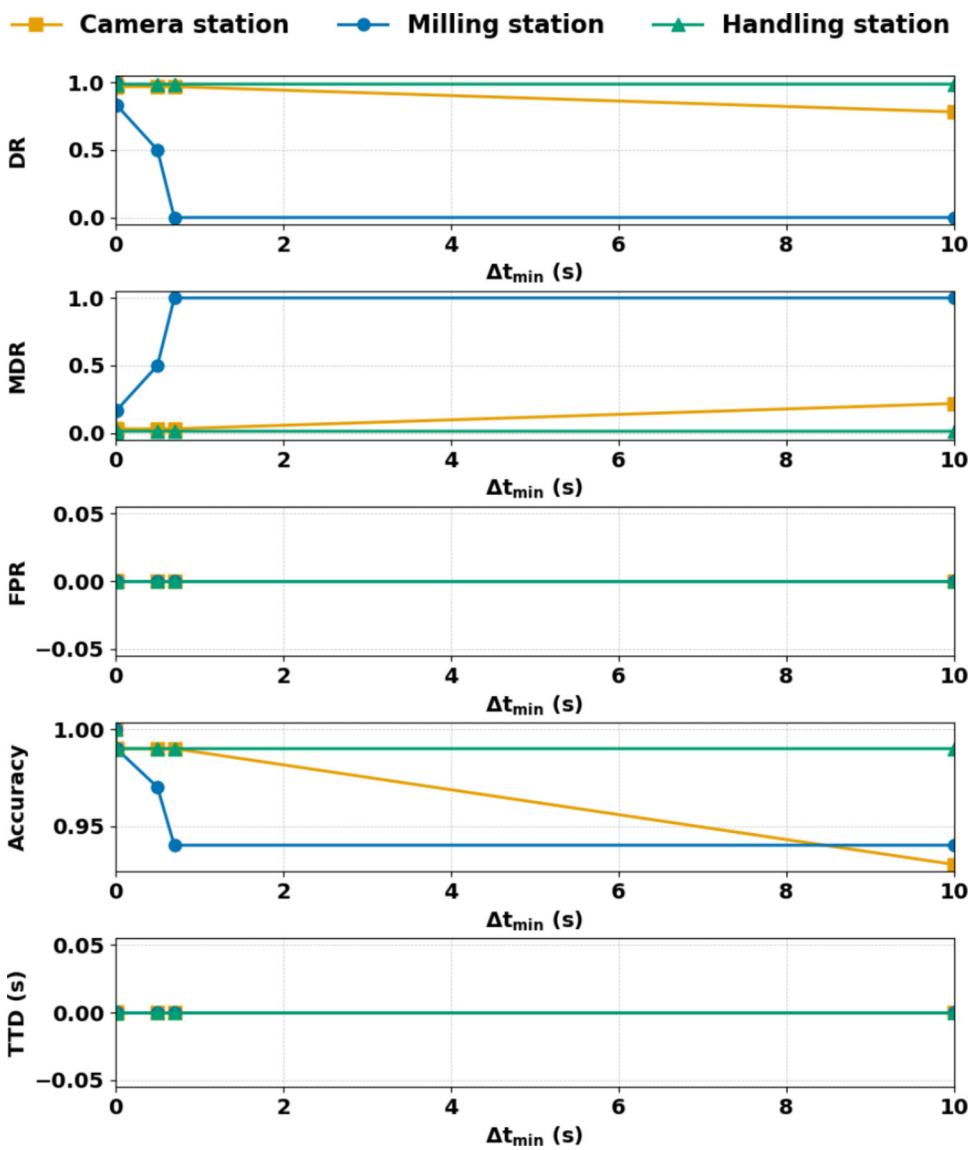

Fig. 11 Evaluation metrics of ablation study of Δt_{min}

Table 7 Summary of quantitative comparison with current state-of-the art

Threshold type	Dataset	Accuracy (%)	DR (%)	FPR (%)	MDR (%)	TTD (s)	Research paper
Static	Continuous stirred tank heater	–	36	0.25	–	–	(Amin et al., 2019)
Dynamic		–	78	1.75	–	–	
Dynamic	Injected sensor faults	–	79.1	–	–	–	(Attar et al., 2024)
Adaptive		–	84.7	–	–	–	
Adaptive	Real marine diesel engine	92	82	0.05	18	–	(Wu et al., 2024)
Adaptive	Sensor bias simulation	–	99.5	–	–	+ 0.42	(Zhao et al., 2022)
Adaptive	PV simulation	90.16	–	–	–	+ 0.21	(Alrifai et al., 2026)
Adaptive multi-scale	Real ICPS	99	98.6	0	1.4	+ 0.00	Our paper

(2024) compared a dynamic and adaptive thresholds for injected sensor faults in an exoskeleton model. The results obtained are improved DR for adaptive threshold (84.7%). However, other quantitative metrics were not reported such as MDR, accuracy, FPR, and TTD limiting their reliability assessment. Other adaptive methods evaluated on simulated sensor bias and PV systems in (Zhao et al., 2022), (Alrifai et al., 2026) achieved high DR 99.5% and 90.16% yet these results are accompanied with TTD values of 0.42 s and 0.21 s respectively and are obtained on simulated datasets. More recent an adaptive threshold was deployed on a marine diesel engine (Wu et al., 2024). The results showed accuracy of 92%, with DR of 82%, FPR of 0.05% and relatively high MDR of 18%, emphasizing a significant number of undetected faults.

In comparison, the proposed adaptive multi-scale threshold which evaluated on a real ICPS, achieved 99% accuracy with DR OF 98.6%, zero of FPR and MDR of 1.4%, while efficiently eliminating detection delay (TTD = 0.00).

These results demonstrated that the proposed approach not only matches the best reported quantitative metrics evaluation, but also in real ICPS data with zero false alarms for spontaneous faults detection. Hence, achieving a significant improvement balance between robustness and responsiveness compared with the current state of the art methods. This confirms that our method achieves a unique balance between responsiveness, robustness, and real-time deployment that is not simultaneously attained by existing state-of-the-art techniques.

The scalability of the proposed approach to larger systems with more stations could be achieved through I/O-aware design. Processing at the edge near to PLC or SCADA publishes only health states and gated alarms upstream through Open Platform Communications Unified Architecture (OPC UA) Publish-Subscribe (Pub-Sub) model or MQTT Sparkplug, so network and historian load increases sub-linearly with the number of stations due to alarm suppression and compact KPIs.

As stations are added, throughput scales near-linearly with available edge gateways, while the uplink historian growth remains muted due to the time-gated alarms and compact KPIs cap event load. This property aligns with the alarm-management goal of reducing nuisance annunciations and ensures that even larger systems with several stations can be monitored without saturating plant networks or historians, consistent with the high accuracy and zero FPR behavior evidenced in our results.

Conclusion

This research introduces a novel, multi-scale fault detection framework specified for dynamic and heterogeneous cyber-physical production systems. By synergistically integrating global EWMA-based trend detection with adaptive peak-to-peak local thresholding, the proposed approach effectively overcomes critical limitations of conventional fault detection methods in non-stationary behaviors. A key innovation lies in the incorporation of adaptive temporal scaling, which dynamically adjusts the analysis window in response to evolving signal variance. This capability allows the system to maintain high sensitivity to both abrupt and incipient faults, ensuring temporal accuracy and detection reliability. In addition, the incorporation of station-specific dual thresholds and a time-gated evaluation mechanism significantly suppress false alarms, improve interpretability and detection specificity, and fault detection with true operational faults.

The proposed approach was validated on a real-world, conveyor-based sorting system encompassing multiple processing stations. The results demonstrated consistent high performance across multiple fault scenarios and process phases comparing with global and local methods. Achieving fault detection accuracy of 99%, zero false positive rates, near-zero missed detection rate, and accurate temporal localization of fault regions. Moreover, the comparative results show a reduction of false alarms of 97% and 100% compared with EWMA and P2P respectively. Also achieve 87% and 98% reduction of MDR compared with EWMA and P2P. These findings affirm the proposed approach's applicability to complex, multi-station SM settings.

In the context of industrial scalability, the pipeline is designed for throughput and footprint characteristics of shop-floor deployments. Each cycle is resampled to a fixed length N and processed in a single pass; the adaptive EWMA and variance updateds are $O(N)$, and the local P2P range adds at most $O(NW_{max})$ work with $W_{max} \leq N$. Memory is $O(N)$ per stream. These bounds make methods computationally light for multi-station, and multi-production lines and flexible to parallelization across stations or production lines. The reliance of the proposed approach on simple streaming recursions (means, ranges, variances) which facilitates edge execution on industrial computers. The time-gated alarm minimizes downstream event load to Manufacturing Execution Systems (MES) layers by suppressing alarm collapses.

The potential practical deployment challenges could be as follows:

- Data quality and timing: Healthy reference drift because of wear variations can be minimized by scheduled re-baseline utilizing recent fault-free (nominal) operating cycles and through accurate time synchronization.

- Alarm management and monitoring: The time-gated and station-specific thresholds should be adjusted to process vulnerabilities and the minimal time-to-consequence and reconfigured within an alarm lifecycle to avoid alarm collapses and chattering.

In the context of integration of the proposed fault detection approach with existing industrial monitoring frameworks, the proposed approach maps clearly onto Open System Architecture for Condition-Based Maintenance (OSA-CBM) and ISO 13374 Condition monitoring and diagnostics of machines by employing data preprocessing (resampling, normalization) and state detection through adaptive EWMA, dual thresholds and time-gated, and it outputs fault-free state suitable for the presentation layer such as dashboards, alarm frames. At the plant connectivity, the health metrics and alarms can be served from OPC UA (Pub-Sub) model or MQTT Sparkplug which allows interoperable connections to MES and cloud analytics that aligning with International Standard for Integrating (ISA-95) enterprise and control systems in manufacturing.

This approach provides a solid foundation for broader integration within Industry 4.0 ecosystems. Future research will explore the extension of the proposed approach to incorporate self-learning threshold adaptation, cross-station fault correlation analysis, and deployment in multi-line and decentralized ICPS architectures under combined fault scenarios under varying loads and operating conditions. By allowing intelligent, context-aware, and adaptive fault detection, the proposed approach contributes significantly to the advances of resilient smart manufacturing systems.

Acknowledgements The authors gratefully acknowledge the financial support of the Luxembourg National Research Fund (FNR) under the grant CORE 2022/17381684/SMOD-SHA, which made this research possible.

Declarations

Conflict of interest The authors declare that this manuscript is original, has not been published before, and is currently not being considered for publication elsewhere.

References

Ahmadini, A. A. H., Khan, I., Alshqaq, S. S. A., AlQadi, H., Ghodhbani, R., & Ahmad, B. (2025). Improved adaptive CUMSUM control chart for industrial process monitoring under measurement error. *Scientific Reports, 15*(1), Article 1. https://doi.org/10.1038/s41598-025-01734-4

Aldrini, J., & Chihi, I. (2025). Towards Responsible AI: Evaluating Intelligent Models for Sensor Fault Detection Through the Lens of Sustainability and Performance Optimization. In *2025 International Conference on Sustainability, Innovation & Technology (ICSIT)*, 1–8. https://doi.org/10.1109/ICSIT65336.2025.11293863

Aldrini, J., Chihi, I., & Sidhom, L. (2023). Fault diagnosis and self-healing for smart manufacturing: A review. *Journal of Intelligent Manufacturing, 35*(6), Article 6. https://doi.org/10.1007/s10845-023-02165-6

Alghushairy, O., Alsini, R., Soule, T., & Ma, X. (2020). A review of local outlier factor algorithms for outlier detection in big data streams. *Big Data and Cognitive Computing, 5*(1), Article 1. https://doi.org/10.3390/bdcc5010001

Alrifai, Y., Aguilera-González, A., Vechiu, I., & Becerra-Becerra, G. (2026). Hybrid fault detection and identification strategy for PV systems combining statistical data-driven and Kalman filter algorithms. *Electric Power Systems Research, 252*, Article 112381. https://doi.org/10.1016/j.epsr.2025.112381

Amin, M. T., Khan, F., & Imtiaz, S. (2019). Fault detection and pathway analysis using a dynamic Bayesian network. *Chemical Engineering Science, 195*, 777–790. https://doi.org/10.1016/j.ces.2018.10.024

Aslansefat, K., Bahar Gogani, M., Kabir, S., Shoorehdeli, M. A., & Yari, M. (2020). Performance evaluation and design for variable threshold alarm systems through semi-Markov process. *ISA Transactions, 97*, 282–295. https://doi.org/10.1016/j.isatra.2019.08.015

Attar, A. A., Bao, K., Hagenmeyer, V., Fabarisov, T., & Morozov, A. (2024). Improving Anomaly Detection with Adaptive Dynamic Threshold: A Review and Enhanced Method. In *2024 8th International Conference on System Reliability and Safety (ICSRS)*, 662–666. https://doi.org/10.1109/ICSRS63046.2024.10927575

Dong, J., Li, D., Cong, Z., & Peng, K. (2025). A new fault detection method based on an updatable hybrid model for hard-to-detect faults in nonstationary processes. *Reliability Engineering & System Safety, 259*, Article 110920. https://doi.org/10.1016/j.ress.2025.110920

Dowdeswell, B., Sinha, R., & MacDonell, S. G. (2020). Finding faults: A scoping study of fault diagnostics for Industrial Cyber–Physical Systems. *Journal of Systems and Software, 168*, Article 110638. https://doi.org/10.1016/j.jss.2020.110638

Eslami, Y., Franciosi, C., Ashouri, S., & Lezoche, M. (2023). A review and analysis of the characteristics of cyber-physical systems in Industry 4.0. *SN Computer Science, 4*(6), Article 6. https://doi.org/10.1007/s42979-023-02268-0

Fan, Z., Liu, H., He, J., Zhang, M., & Du, X. (2021). MPDNet: A 3D Missing Part Detection Network Based on Point Cloud Segmentation. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1810–1814. https://doi.org/10.1109/ICASSP39728.2021.9414867

Haddar, M., Jorani, R. M., Parey, A., Chaari, F., & Haddar, M. (2024). Experimental evaluation for detecting bevel gear failure using univariate statistical control charts. *Journal of the Brazilian Society of Mechanical Sciences and Engineering, 46*(4), Article 4. https://doi.org/10.1007/s40430-024-04816-y

Harrou, F., Nounou, M., & Nounou, H. (2013). A statistical fault detection strategy using PCA based EWMA control schemes. In *2013*

9th Asian Control Conference (ASCC), 1–4. https://doi.org/10.1109/ASCC.2013.6606311

Harrou, F., Nounou, M. N., Nounou, H. N., & Madakyaru, M. (2015). PLS-based EWMA fault detection strategy for process monitoring. *Journal of Loss Prevention in the Process Industries, 36*, 108–119. https://doi.org/10.1016/j.jlp.2015.05.017

Hegazy, M. M., Badawi, A. A., El-Nabarawi, M. A., Eldegwy, M. A., & Louis, D. (2024). One factor at a time and factorial experimental design for formulation of l-carnitine microcapsules to improve its manufacturability. *Heliyon*. https://doi.org/10.1016/j.heliyon.2023.e23637

Iqbal, J., Noor-ul-Amin, M., Khan, I., AlQahtani, S. A., Yasmeen, U., & Ahmad, B. (2023). A novel Bayesian Max-EWMA control chart for jointly monitoring the process mean and variance: An application to hard bake process. *Scientific Reports, 13*(1), Article 21224. https://doi.org/10.1038/s41598-023-48532-4

Kim, M., Jung, S., Kim, B., Kim, J., Kim, E., Kim, J., & Kim, S. (2022). Fault detection method via k-nearest neighbor normalization and weight local outlier factor for circulating fluidized bed boiler with multimode process. *Energies, 15*(17), Article 17. https://doi.org/10.3390/en15176146

Lee, W.-C., Lee, K., & Choi, H.-L. (2024). Multi-modal neural adaptive observer for sensor and actuator fault detection and identification. *International Journal of Aeronautical and Space Sciences, 26*(3), Article 3. https://doi.org/10.1007/s42405-024-00823-4

Leite, D., Andrade, E., Rativa, D., & Maciel, A. M. A. (2024). Fault detection and diagnosis in Industry 4.0: A review on challenges and opportunities. *Sensors, 25*(1), Article 1. https://doi.org/10.3390/s25010060

Li, Y., & Dong, J. (2025). Fault detection unknown input observer for local nonlinear fuzzy autonomous ground vehicles system based on a joint peak-to-peak analysis and zonotopic analysis threshold. *IEEE Transactions on Vehicular Technology, 74*(5), 7226–7236. https://doi.org/10.1109/TVT.2025.3526169

Li, Y., Qin, J., & Wu, C. (2023). A robust adaptive exponentially weighted moving average control chart with a distribution-free design strategy. *Computers & Industrial Engineering, 177*, Article 109083. https://doi.org/10.1016/j.cie.2023.109083

Liu, Y., Kang, J., Wen, L., Bai, Y., Guo, C., & Yu, W. (2023). Fault diagnosis algorithm of gearboxes based on GWO-SCE adaptive multi-threshold segmentation and subdomain adaptation. *Processes, 11*(2), Article 2. https://doi.org/10.3390/pr11020556

Ma, Y., Wang, Z., Meslem, N., Raïssi, T., & Shen, Y. (2023). Fault diagnosis by interval-based adaptive thresholds and peak-to-peak observers. *International Journal of Adaptive Control and Signal Processing, 37*(2), 519–537. https://doi.org/10.1002/acs.3535

Marais, H. L., Zaccaria, V., & Odlare, M. (2022). Comparing statistical process control charts for fault detection in wastewater treatment. *Water Science and Technology, 85*(4), 1250–1262. https://doi.org/10.2166/wst.2022.037

Maraş, S., Arslan, H., & Birgören, B. (2021). Detection of gear wear and faults in spur gear systems using statistical parameters and univariate statistical process control charts. *Arabian Journal for Science and Engineering, 46*(12), Article 12. https://doi.org/10.1007/s13369-021-05930-y

Montgomery, D. C. (2009). *Introduction to Statistical Quality Control*. Wiley.

Rudawska, A., Madleňák, R., Madleňáková, L., & Droździel, P. (2020). Investigation of the effect of operational factors on conveyor belt mechanical properties. *Applied Sciences, 10*(12), Article 12. https://doi.org/10.3390/app10124201

Sorostinean, R., Neghina, C., & Gellert, A. (2025). Boosting anomaly detection with unsupervised K-Means and SOM for energy-efficient factory machines. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-025-02754-7

Stanly Jayaprakash, J., Priyadarsini, M. J. P., Parameshachari, B. D., Karimi, H. R., & Gurumoorthy, S. (2022). Deep Q-network with reinforcement learning for fault detection in cyber-physical systems. *Journal of Circuits, Systems and Computers, 31*(09), Article 2250158. https://doi.org/10.1142/S0218126622501584

Sun, Y., Qin, W., Zhuang, Z., & Xu, H. (2021). An adaptive fault detection and root-cause analysis scheme for complex industrial processes using moving window KPCA and information geometric causal inference. *Journal of Intelligent Manufacturing, 32*(7), Article 7. https://doi.org/10.1007/s10845-021-01752-9

Tang, Y.-C., & Li, K.-H. (2023). A machine-learning approach to setting optimal thresholds and its application in rolling bearing fault diagnosis. *Machine Learning: Science and Technology, 4*(4), Article 045030. https://doi.org/10.1088/2632-2153/ad0ab3

Tran, K. P. (2022). *Control Charts and Machine Learning for Anomaly Detection in Manufacturing*. https://link.springer.com/book/https://doi.org/10.1007/978-3-030-83819-5

Veerasamy, G., Kannan, R., Siddharthan, R., Muralidharan, G., Sivanandam, V., & Amirtharajan, R. (2022). Integration of genetic algorithm tuned adaptive fading memory Kalman filter with model predictive controller for active fault-tolerant control of cement kiln under sensor faults with inaccurate noise covariance. *Mathematics and Computers in Simulation, 191*, 256–277. https://doi.org/10.1016/j.matcom.2021.07.023

Wang, L., You, P., Zhang, X., Jiang, L., & Li, Y. (2025a). Adaptive adjustment graph representation learning method for rotating machinery fault diagnosis under noisy signals. *Frontiers of Mechanical Engineering, 20*(1), Article 1. https://doi.org/10.1007/s11465-024-0818-y

Wang, T., Zheng, R., Li, M., Cai, C., Zhu, S., & Lou, Y. (2025b). Deep learning based self-adaptive modeling of multimode continuous manufacturing processes and its application to rotary drying process. *Journal of Intelligent Manufacturing, 36*(6), 3887–3922. https://doi.org/10.1007/s10845-024-02438-8

Wu, T., Song, H., Gao, H., Wu, Z., & Han, F. (2024). Adaptive dynamic thresholding method for fault detection in diesel engine lubrication systems. *Machines, 12*(12), Article 895. https://doi.org/10.3390/machines12120895

Xu, W., Xu, K.-J., Wu, J.-P., Yu, X.-L., & Yan, X.-X. (2019). Peak-to-peak standard deviation based bubble detection method in sodium flow with electromagnetic vortex flowmeter. *Review of Scientific Instruments, 90*(6), Article 065105. https://doi.org/10.1063/1.5089690

Yan, J., Liu, Y., & Ren, X. (2023). An early fault detection method for wind turbine main bearings based on self-attention GRU network and binary segmentation changepoint detection algorithm. *Energies, 16*(10), Article 10. https://doi.org/10.3390/en16104123

Zhao, W., Guo, Y., & Sun, H. (2022). Research on an adaptive threshold setting method for aero-engine fault detection based on KDE-EWMA. *Journal of Aerospace Engineering, 35*(6), Article 04022087. https://doi.org/10.1061/(ASCE)AS.1943-5525.0001483