SciEn
Journal of Engineering Advancements Vol. 07(04) 2026, pp 133-147 https://doi.org/10.38032/jea.2026.04.001

A Deep Autoencoder-based Adaptive Control Chart Framework under Weak Supervision for Industrial Anomaly Detection

Mehedi Hassan Maruf, Mostafizur Rahman and Md. Mamunur Roshid*

Department of Mechanical Engineering, Chittagong University of Engineering & Technology (CUET), Chattogram-4349, Bangladesh

Received: February 28, 2026, Revised: September 24, 2026, Accepted: October 05, 2026, Available Online: November 08, 2026

ABSTRACT

Nowadays, industrial anomaly detection (IAD) is challenged by multi-dimensional data, nonlinear process behavior, and the lack of reliable fault labels. To address these challenges, this paper proposes a Synergistic Machine Learning–Adaptive Control Chart (SML-ACC) framework under a weakly supervised setting for IAD. The framework combines data-driven feature learning with probabilistic supervision and adaptive statistical decision rules. Furthermore, this allows a reliable and interpretable process monitoring without requiring extensively labeled fault data. It is noted that normal process behavior was learned using only fault-free samples. In addition, deviations from normal operation were measured through fused anomaly scores, which are monitored using ACC with dynamically updated control limits. The proposed framework was evaluated using the three public industrial benchmark datasets namely Tennessee Eastman Process, Steel Plates, and SECOM dataset. Thereafter, the performance was analyzed through comprehensive benchmarking against classical statistical process control methods and unsupervised baselines. Notably, systematic ablation studies were conducted to evaluate the contribution of individual framework components. The results indicate that SML-ACC outperforms the tested classical SPC and unsupervised baseline methods in terms of F1 score and AUC value on the TEP and Steel Plates datasets. Its performance with the SECOM dataset is similar to the evaluated baselines. SML-ACC has an F1 score of 0.604 and an AUC of 0.762 on the Tennessee Eastman Process dataset. The F1 score for cross domain evaluation without retraining is around 69.2% of the value of the source domain. This performance drop is a result of the domain shift, and a motivation for future work on domain adaptation. Collectively, the findings suggest that a hybrid of deep representation learning and adaptive SPC can be an effective and scalable solution for anomaly detection in the context of Industry 4.0.

Keywords: Deep Autoencoders, Adaptive Control Charts, Weak Supervision, Process Monitoring, Industrial Anomaly Detection



Copyright @ All authors

This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.

1 Introduction

Anomaly detection (AD) has become a significant component in product quality, operational safety and economic performance in modern manufacturing systems [1]. Being able to detect abnormal process behavior allows for timely corrective action in order to reduce downtime and prevent equipment damage. Because of the proliferation of sensing, automation and Industrial Internet of Things (IIoT) platforms, the number of industrial data streams is becoming high dimensional, nonlinear, noisy and interrelated, causing difficulties in reliable process monitoring and AD [2],[3]. As a conventional approach for industrial process monitoring (IPM), statistical process control (SPC) has been a long-standing practice due to their simplicity, transparency and interpretability. Classical control charts, such as Shewhart, EWMA and CUSUM, continue to be widely used [4]-[7] in IPM and IAD as well. They work well with independent and stationary process variables. However, real manufacturing systems are seldom as ideal as these assumptions. Additionally, nonlinear, coupled and non-stationary dynamics are prevalent in real-world processes. Consequently, classical control charts can be less effective at detecting abnormal behavior, and can lead to too many false alarms [8]. Therefore, recent studies have investigated the combination of machine learning (ML) and SPC, where Tran et al. [8]. demonstrated that integrating ML into SPC can enhance the monitoring sensitivity in complex process. In the following research, the deep learning (DL) models have been merged with control charts. Tayalati et al. [9] combined the long-short term memory (LSTM) autoencoders with SPC charts to learn the time dependent anomaly patterns. Furthermore, Sabri et al. [9] integrated EWMA chart with LSTM model to enhance the detection of slow and gradual process drift. Auto-pattern recognition using support vector machines and neural networks has also been linked to control charts. These methods can be of great value when the classical statistical assumptions are not satisfied, as is often the case in nonparametric situations, [2],[7],[11],[12].

Although significant progress has been made, the use of fixed control limits in many hybrid methods has not been eliminated. Most of the time, they emphasize performance enhancements for individual control charts instead of the creation of completely adaptive monitoring systems. This means that their performance can deteriorate if the behavior of the underlying process is changing over time. At the same time, the shortcomings of traditional SPC have promoted the adoption of unsupervised feature extraction methods. Deep autoencoders (DAEs) are especially appealing since they may be trained with only “good” (or fault free) information. Givnan et al. [13] proved that autoencoders are able to learn short representations of normal operating conditions and then be used to detect anomalies in the industrial environment. Other research has followed with the addition of attention mechanisms, adversarial learning, and temporal modelling for

*Corresponding Author Email Address: raselmostafiz17@gmail.com

more complex spatial and temporal modelling [14]-[16]. Variants that are resource efficient have also been created for use in resource constrained environments and applications like motor monitoring, screw tightening and additive manufacturing [11],[17].

However, most of the autoencoder based techniques are still based on fixed thresholds defined based on reconstruction errors. In the same way, numerous hybrid ML and SPC methods calculate control limits for the initial reconstruction loss distribution which are maintained at a constant value throughout the monitoring [18]. This practice reduces the adaptive ability that is at the heart of SPC. It also establishes a gap between advanced representation learning and responsive statistical monitoring. A model may be trained to learn changing process patterns and the final statistical decision boundary will stay constant.

A second difficulty is the scarcity of industrial data tagged with labels. In real manufacturing situations, fault and anomaly data are not very well annotated. Faults occur rarely and are not cheap to be labelled. Other limitations of data collection can include safety issues, confidentiality, and limited availability of domain experts [19]. The available labels have therefore been incomplete, ambiguous, and even very imbalanced. These restrictions limit the usefulness of fully supervised AD methods in actual industrial applications. However, if enough labeled data exists, it is possible for supervised classifiers to discover good decision boundaries between normal and abnormal process conditions [20].

This restriction has given rise to the interest in weakly supervised AD. In this context, the information provided for supervision is incomplete or indirect [21],[22]. This may be retrieved from a statistical indicator or a simple heuristic, or pseudo labels produced by an unsupervised model [23],[24]. Pseudo labelling has proven to be promising. To improve the anomaly discrimination, Jiang et al.[21] proposed to use pseudo labels to train classifiers, which is a supervised method. There are also two classifiers that are successful in weakly labeled industrial environments, one of which is a class and another is two class classifiers. Moreover, features derived from autoencoders have been applied to downstream classification, for example, in natural gas pipeline monitoring [25]. By applying these methods, better separation between normal and abnormal states can be achieved. They are typically not coupled with an adaptive layer of statistical decisions, which are still interpretable when operating in real-time.

Another important challenge is cross domain generalization. There may be significant variations in industrial processes between machines, between individual plants, and between operating conditions. These differences can lead to distribution shifts and to lower detection performance. Some possible solutions are domain adaptation and transfer learning. Zhan et al. [25] introduced a graph neural network-based cross domain representation learning approach for industrial control system. Adversarial deep learning was employed by Michau et al. [20] to transfer knowledge between industrial units. But these filters are not often used within comprehensive pipelines that also feature adaptive statistical decision layers. There are three important gaps in the literature in general. First, a lot of the ML based monitoring systems use predetermined thresholds. Secondly, there is a lack of a tight coupling between deep feature learning and adaptive SPC. Third, robustness in various industrial areas is still low. These restrictions are significant because engineers are still using control charts. They offer statistically-based decision limits and clear visual signals [16],[26]. Other learning-based AD techniques, however, generate less interpretable anomaly scores. This discrepancy can decrease the confidence of industries and make the practical adoption difficult [20]. In response to these challenges, this paper introduces a Synergistic Machine Learning and Adaptive Control Chart (SML-ACC) framework for industrial anomaly detection (IAD) problem under weak supervision. The framework fuses unsupervised representation learning with pseudo label-based classification. This design minimizes the need to manually annotate faults. It then combines the anomaly information learned with adaptive statistical process monitoring (SPM). The aim is to enhance detection robustness without compromising the interpretability of statistical monitoring [26].

SML-ACC is not a classical SPC, but it is an extension of SPC based on data driven adaptation. The anomaly information is learned and directly applied to an adaptive control chart (ACC). Three heterogeneous benchmark datasets are used to evaluate the framework: the Tennessee Eastman Process (TEP) [27], the benchmark Steel Plates Faults [28] and the benchmark SECOM for semiconductor manufacturing [29]. These data sets are of different dimensionality, have different noise, and have different process dynamics. This diversity offers a larger foundation for the analysis of the proposed framework. The evaluation is done by comparing SML-ACC to classical SPC methods and unsupervised baseline models. In addition, cross domain evaluation is conducted without re-training. It gives direct assessment of the performance in the distribution shifts and demonstrates how the detection capability varies in individual industrial domains. The adaptive control chart dynamically adjusts detection limits based on the process variability. Systematic ablations are also performed to estimate the individual contribution from different framework components. The analyses combined give a thorough evaluation of the proposed approach for weakly supervised industrial anomaly detection.

2 Methodology

This methodological design presents a SML-ACC system in the IAD. The framework combines the deep unsupervised representation learning, supervised classification, and adaptive SPC into a single pipeline. The general structure of mechanism of the proposed framework is depicted in Fig. 1. Firstly, data preprocessing and feature engineering are applied to raw multivariate process data for providing normalization, consistency, and dimensional alignment between datasets. Next, the DAE is trained on fault free data to learn a low dimensional latent representation of normal operating conditions. The reconstruction error produced by the autoencoder is an unmonitored measure of process abnormality. Thereafter, the latent representations of the autoencoder are combined with chosen engineered process variables and inputted to a trained fault classifier to approximate the likelihood of faulty operation. This is followed by the reconstruction error and classifier probability, which is added together to get a synergistic anomaly score. Subsequently, an ACC using a temporal smoothing and dynamic control limits is used to monitor the anomaly score. When the smoothed anomaly score is above the adaptive upper control limit, the smoothed anomaly score is declared a fault. After being trained on a source domain, the same structure is directly transferred to target domains without retraining, to allow analysis of cross-domain fault detection. The framework operates under weak supervision using reconstruction-error–based pseudo-labels due to the absence of explicit fault annotations.

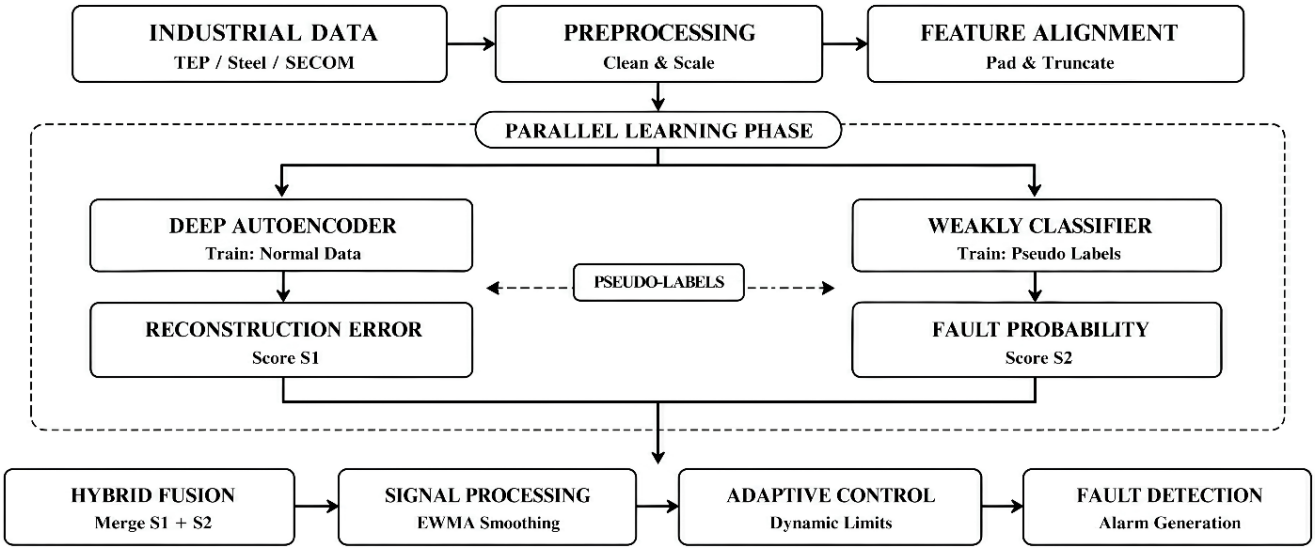


Fig. 1 Working flow diagram of SML-ACC system

2.1 Description of Dataset

Three industrial benchmark datasets, i.e., TEP, Steel Plates Faults, and SECOM are utilized to test the proposed framework. The datasets vary in terms of dimensionality, noise level, and faults. It becomes challenging to perform a single evaluation due to this diversity. The TEP dataset exhibit a nonlinear complex chemical process. Signatures of faults tend to be overlapped making it ticklish to detect them. Additionally, there is a slight defect pattern and imbalance in classes in the Steel Plates dataset. On the contrary, the SECOM dataset is high-dimensional, noisy, and redundant in features. It has become indispensable that datasets are subjected to a common processing pipeline in order to guarantee equal assessment by heterogeneity in data. The processed data is then projected into the fixed latent representation used by the framework using the trained autoencoder. It is important to note that domain adaptation is not performed and the evaluation is done across domains to reveal the impact of domain shift and not to define semantic equivalence between features across datasets.

2.2 Data Preprocessing and Feature Engineering

In order to be robust and ensure computational efficiency, a single pipeline is used to perform preprocessing of heterogeneous datasets of industry. Overall, the procedure is same, and only a few dataset-specific adjustments are introduced in order to adjust with variations in the data characteristics. First, missing values are addressed. In the case of Steel Plates data, linear interpolation is used to maintain local continuity, and thereafter, gaps are filled by mean imputation. Conversely, the mean imputation is directly used since there are no strong time-related assumptions in TEP and SECOM datasets. Moreover, undefined and infinite numbers are substituted to achieve numerical stability. Thereafter, feature engineering is performed to improve fault-relevant information through statistical, time-based, and interaction-based transformations. However, low-variance features are eliminated, and features with high correlation are pruned in order to avoid dimensionality explosion.

In addition, in large scale datasets, the correlation analysis is conducted on representative subsets in order to minimize the computational cost while maintaining the dominance of dependencies. Adaptive feature normalization is performed, i.e., quantile transformation on large datasets, and robust or standard normalization on small datasets. There is selective use of outlier handling in order to reduce excessive data loss. In particular, the normalization of the TEP data is followed by an IQR-based filtering step used on the fault-free data. Subsampling is only applied in feature selection and training the supervised classifier. Finally, the arrangement of the sensors is not uniform, leading to different feature dimensions in different datasets. Hence, only zero-padding/truncation is used to ensure the fixed dimensionality of the input passed to the autoencoder. This dimensional change is considered an architectural change and is not an indication of semantic linkage or physical equivalence of features between datasets. It is recognized that this across domain performance can be influenced by the meaning, distribution, noise characteristics and underlying process dynamics of the features, which is a limitation of the present evaluation.

2.3 Deep Autoencoder-based Representation Learning

Following preprocessing, the SML-ACC framework uses DAE to acquire normal process behavior by reconstructing it. The DAE is trained on the fault-free data of the TEP dataset, which is further used as a source domain due to its high sample size and steady-state normal operating conditions. Single-domain training does not confuse the heterogeneous process dynamics and can make cross-domain evaluations to demonstrate the effects of domain shift explicitly without domain adaptation. Conversely, the learned normal manifold is not the faulty conditions. Such deviation generates increased reconstruction errors. This behavior makes characterization of anomalies effective under minimal supervision.

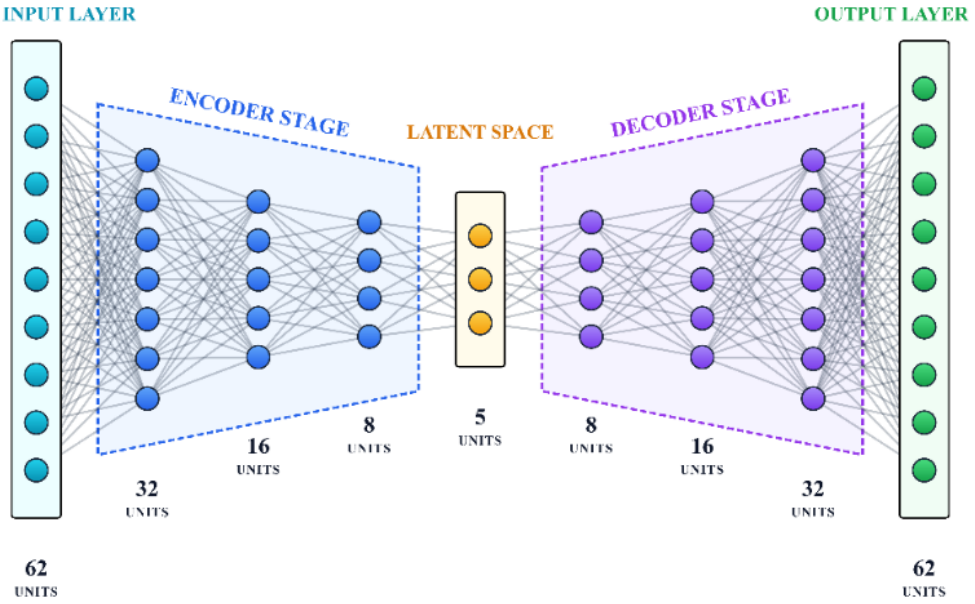


Fig. 2 Architecture of deep autoencoder

As illustrated in Fig. 2, each process sample is represented by a 62-dimensional normalized feature vector. A symmetric encoder–decoder architecture is adopted with gradual compression and expansion (62 → 32 → 16 → 8 → 5 → 8 → 16 → 32 → 62). This step by step diminishing addresses nonlinear association between process variables. Meanwhile, it does not cause the sudden loss of information. The latent space is 5-dimensional which serves as a bottleneck. It suppresses repetition and noises. In the meantime, critical process dynamics are maintained. Mirrored decoder implements consistency of reconstruction. This design stabilizes training. Formally, for an input vector $x \in \mathbb{R}^{62}$, the encoder generates a latent representation

$$z = f_{enc}(x), z \in \mathbb{R}^5 \quad (1)$$

The decoder then reconstructs the input as

$$\hat{x} = f_{dec}(z) \quad (2)$$

Training minimizes the mean squared reconstruction error (MSE) is defined as [30]:

$$L_{rec} = \frac{1}{N}\sum_{i=1}^{N} \| x_i - \hat{x}_i \|_2^2 \quad (3)$$

Computed solely on fault-free samples. After training, a sample-wise reconstruction error is calculated as

$$e_i = \frac{1}{62}\sum_{j=1}^{62} (x_{ij} - \hat{x}_{ij})^2 \quad (4)$$

Large reconstruction errors are usually obtained when the faulty observations are not matched to normal behavior. Along with anomaly characterization, reconstruction error is subsequently used to obtain pseudo-labels in weakly supervised classification, which ensures agreement between unsupervised representation learning and downstream fault discrimination. The basic input into the supervised classifier is the latent representation z. This enhances stability and noise strength. Second, the reconstruction error is kept as a deviation indicator aide-memoir. The ACC then utilizes it. It is noted that reconstruction error is not used as an independent detector. It, instead, augments supervised learning. The training is completely unmonitored, and Adam optimizer is used. Early termination and learning-rate decadence are implemented using validation loss. The dropout regularization enhances robustness in the clean industrial settings. In cross-domain assessment, the feature alignment is used prior to the encoding to meet architectural requirements without removing the effects of domain shifts. The complete architecture and hyperparameter configuration corresponding to Fig. 2 are summarized in Table 1, ensuring reproducibility.

Table 1 Hyperparameter of deep autoencoder

Parameter	Value
Architecture	62 → 32 → 16 → 8 → 5 → 8 → 16 → 32 → 62
Input dimension	62
Latent dimension	5
Hidden activation	ReLU
Output activation	Linear
Loss function	Mean Squared Error (MSE)
Optimizer	Adam
Batch size	128
Maximum epochs	100
Validation split	0.2
Dropout rate	0.2

2.4 Data Splitting and Model Selection

For data splitting to clearly separate model training, model selection, and final evaluation, the data-splitting procedure was specified separately for the autoencoder and the classifier stages. The data that were used in the training of the deep autoencoder were split into 80/20 subsets for training and validation, respectively, with a fixed random seed of 42. The validation subset was used to calculate the validation loss and to stop training the autoencoder early. Stratified sampling with a fixed random seed of 42 was used to split the labeled samples available for the downstream classifier into 75% training and

25% held-out test subsets. Five-fold cross validation was performed on the training subset and the comparison and selection of models among the three models, namely Random Forest, Extra Trees, and Gradient Boosting, were done using the cross validation F1 score. The set aside 25% test set was not used for the selection of the classifiers, but was reserved for final evaluation of the classifiers chosen. The same set of parameters for any pre-processing of the features for the classifiers were optimized on the training set and then used for the held-out test set. This procedure ensures uniformity of separating model development from the final performance evaluation process.

2.5 Supervised Classification

SML-ACC framework is used to supervise fault classification in the weakly supervised environment after learning the unsupervised representation. Thus, supervision is based on pseudo-labels constructed using reconstruction-errors of the trained autoencoder. Samples having reconstruction errors exceeding the threshold at or above the 60th percentile are considered pseudo-anomalous, while others are considered pseudo-normal. The threshold is defined based on the distribution of reconstruction error and without the need of ground-truth fault labels. At the main pipeline, classifiers work on a composite feature representation of extracted latent features of the autoencoder and selected raw process variables and feature engineered. Latent representations offer a noise free and succinct description of global process behavior. Conversely, the raw and engineered features are chosen for enhancing the sensitivity to the fault variation. Reconstruction error is not added to the inputs of the classifier, instead it is applied during pseudo-labeling and anomaly scoring. Three tree-based ensemble classifiers, including Random forest, Extra Trees, and Gradient Boosting, are compared because they are not vulnerable to nonlinearity, feature interactions, and imbalance of the classes. The classifier that gives the best consistent results in F1-score and AUC is chosen to downstream the score on anomalies and merge with the ACC. These pseudo-labels are not actual fault labels; they serve only as an indirect label in the training of the downstream classifier.

2.6 Anomaly Scoring and Adaptive Control Chart

As discussed in the previous section, the autoencoder is trained with only fault-free samples, and gives sample-level reconstruction errors that measure deviations from the learned normal operating pattern. The supervised classification is followed by a calculation of the fault probability of each sample from the classifier's point of view, based on the fused feature representation. Reconstruction error and the probability of the classifiers offer complementary information: Reconstruction error reflects the deviation from the learned normal manifold, while the probability of the classifiers contains more information about discriminating normal and abnormal process states. The values of both quantities are normalized to [0, 1] and fused with the fusion weights that are specific to the dataset and listed in Table 2. As a result, the fused anomaly score is regarded as a one-dimensional (1-D) temporal sequence.

Fused anomaly score is smoothed using an Exponentially Weighted Moving Average (EWMA) [31] to reduce short-term variations but keep track of long-term variations in the anomaly score sequence.

$$z_i = \lambda s_i + (1 - \lambda) z_{i-1} \quad (5)$$

where s_t is the fused anomaly score at time t, z_t is the EWMA-smoothed anomaly score, and λ is the smoothing factor. The values of λ for the data sets are given Table 2.

A layer of adaptive control-chart (ACC) is then added on the EWMA-smoothed anomaly-score sequence. The initial process baseline is set with a dataset specific reference window. Initial estimates of the center and dispersion are the mean and the standard deviation computed from the reference window. The number of the reference and window ranges are presented in Table 2 as 80, 40 and 25 observations for TEP, Steel Plates and SECOM database respectively. This initialization is used to establish a local reference level for monitoring, instead of using a single global reference level for the diverse datasets.

Once the control-chart statistics have been initialized, the anomaly-score sequence is added to it to account for a temporal variation in the process under observation. The adaptive center and dispersion estimates provide the basis for the control limits to be calculated for each monitoring step. The upper control limit (UCL) is:

$$UCL_i = \mu_i + L\sigma_i \quad (6)$$

The adaptive center μ_t and the dispersion estimates σ_t are at time t. $\mu_t \pm 2.0\sigma_t$ defines the control limits, and $\mu_t \pm 1.2\sigma_t$ defines the warning limits which are provided in Table 2.

The adaptive control limits are then compared with the EWMA-smoothed anomaly score to yield the anomaly decision. The samples with smoothed anomaly scores above the corresponding UCL are marked as anomalies. The adaptive formulation can tune the decision boundary for time varying patterns of the anomaly-score distribution, which makes the monitoring layer appropriate for non-stationary industrial processes as opposed to a global threshold. It is important to note that the adaptive monitoring procedure is applied sequentially through the anomaly score stream and that observations that are flagged as anomalous are not subsequently deleted from the monitoring stream. Thus, the following adaptive estimation process still includes the anomalous observations. This is to maintain the online character of the ACC, but if this is the case, it can have an impact on the future estimates of the center and dispersion values, and thus on future control limits. The potential influence is an intrinsic limitation of the present adaptive updating strategy and calls for further research on anomaly resistant updating schemes. In general, the fused machine-learning anomaly score is transformed into a statistical monitoring decision in a meaningful way and the control limits are flexible to accommodate temporal variations in the distribution of the anomaly scores observed.

2.7 Implementation and Parameter Settings

To ensure reproducibility of the proposed SML-ACC framework, the key implementation parameters and dataset-specific settings are summarized in Table 2. The implementation includes the reconstruction-error-based pseudo-label criterion, classifier settings, anomaly-score normalization and fusion, EWMA smoothing, reference-window initialization, adaptive updating, and control-limit parameters.

Table 2 Implementation and parameter settings of the SML-ACC framework.

Component	Parameter	TEP	Steel Plates	SECOM
Pseudo-labeling	Reconstruction-error threshold	60th percentile*	60th percentile*	60th percentile*
Classifier	Candidate models	RF, ET, GB	RF, ET, GB	RF, ET, GB
	Random seed	42	42	42
Score processing	Normalization range	[0, 1]	[0, 1]	[0, 1]
Score fusion	Fusion formulation	Weighted geometric	Weighted arithmetic	Weighted arithmetic
	Reconstruction-error weight (α)	0.70	0.60	0.30
	Classifier-probability weight (β)	0.30	0.40	0.70
EWMA	Smoothing parameter (λ)	0.12	0.18	0.25
Adaptive control chart	Reference/initial window	80	40	25
	Control limits	$\mu \pm 2.0\sigma$	$\mu \pm 2.0\sigma$	$\mu \pm 2.0\sigma$
	Warning limits	$\mu \pm 1.2\sigma$	$\mu \pm 1.2\sigma$	$\mu \pm 1.2\sigma$
Post-processing	Secondary threshold	85th percentile	80th percentile	90th percentile
	Minimum anomaly rate	20%	15%	8%

Note: The 60th-percentile threshold represents the reconstruction-error-based binary-label generation criterion and does not correspond to a ground-truth fault threshold.

These pseudo-labels derived from reconstruction errors are only used for training and selecting the classifiers in the downstream stages, and are not considered as ground-truth fault annotations. Original ground truth labels are not provided during the model training or anomaly score generation phases and are only available for the final performance evaluation based on F1 score, AUC, FAR, and FDR.

Algorithm 1. SML-ACC framework for weakly supervised industrial anomaly detection

No.	Algorithm
1.	Inputs: Fault-free data X_{normal}, process data X
2.	Parameters: τ ← 60th-percentile threshold; α, β ← fusion weights; λ ← EWMA parameter; W ← reference window; k ← 2.0
3.	Output: Anomaly labels ŷ and performance metrics
4.	Preprocess and normalize X_{normal} and X; perform feature engineering and dimensional alignment
5.	Split X_{normal} into training and validation sets (80:20)
6.	Train the deep autoencoder using X_{normal} ,train
7.	E_{normal} ← Calculate reconstruction errors for fault-free samples
8.	τ ← Percentile (E_{normal} , 60)
9.	$\tilde{y}_i$ ← 1 if E_i > τ, otherwise $\tilde{y}_i$ ← 0
10.	Construct classifier features from latent representations and selected process features
11.	Split classifier data into training and held-out test sets (75:25) using stratified sampling
12.	for c ∈ {RF, ET, GB} do
13.	$F1_c$ ← Calculate mean five-fold stratified cross-validation F1-score on the training set
14.	end for
15.	c* ← Select the classifier with the highest cross-validation F1-score
16.	p_t ← Obtain classifier fault probability using c*
17.	S_t ← Fuse normalized reconstruction error and p_t using the predefined α, β
18.	$Z_t \leftarrow \lambda S_t + (1-\lambda) Z_{t-1}$
19.	μ_t, σ_t ← Calculate rolling mean and standard deviation over the reference window W
20.	$UCL_t \leftarrow \mu_t + k\sigma_t$
21.	$\hat{y}_t \leftarrow 1$ if $Z_t > UCL_t$, otherwise $\hat{y}_t \leftarrow 0$
22.	Apply predefined post-processing criteria and evaluate ŷ against ground-truth labels
23.	return ŷ, F1-score, Precision, Recall, AUC, FAR, and FDR

Note: The fusion operator may be arithmetic or geometric according to the predefined dataset-specific settings reported in Table 2.

2.8 Evaluation Metrics

Since the problem is weakly supervised, threshold-independent like AUC metric and imbalance-conscious like F1-score are focused accuracy. The AUC and F1-score are typical metrics used in IAD and classification, which are prioritized. These metrics assess detection accuracy, discrimination capability, and monitoring reliability under weak supervision. Let TP, FP, TN, and FN stand for true positives, false positives, true negatives, and false negatives, respectively. Precision evaluates the reliability of detected faults and is defined as for final evaluation, the predicted anomaly labels are compared with the original ground-truth fault labels provided by each dataset. These ground-truth labels are used exclusively for calculating the final performance metrics and are not used during pseudo-label generation or classifier training.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{7}$$

Recall measures fault detection sensitivity,

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{8}$$

The F1-score provides a balanced measure between Precision and Recall,

$$\mathrm{F1} = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{9}$$

To evaluate discrimination performance independent of decision thresholds, the Area Under the ROC Curve (AUC) is reported. AUC is particularly suitable for imbalanced data and weakly supervised settings. Monitoring reliability is further assessed using the False Alarm Rate (FAR),

$$\mathrm{FAR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{10}$$

and the False Discovery Rate (FDR),

$$\mathrm{FDR} = \frac{\mathrm{FP}}{\mathrm{TP} + \mathrm{FP}} \tag{11}$$

All of these metrics give a complete analysis of the detection accuracy, strength, and the practical use of the proposed framework in the case of industrial monitoring.

3 Results

3.1 Preprocessing Results

This section interprets the preprocessing results using the industrial datasets which are usually skewed, noisy, and uneven. TEP faulty dataset, Steel Plates, and SECOM dataset have 5,24,354 samples, 55 original features, 1,941 samples, 33 features and 1,567 samples, 590 features, respectively. Firstly, datasets are preprocessed and converted to compact and consistent representations of features which are useful in downstream analysis. Final feature dimensions are decreased to 62 for TEP, 58 for Steel Plates and 85 for SECOM, which highlights the heavy elimination of redundant features without losing significant data. Furthermore, 5,930 misrepresented samples are eliminated out of the error-free episode of TEP database, leading to a better data stability.

Fig. 3(a) reveals that the skewness and value range of raw TEP variables are high. Following the preprocessing, the variables in Fig. 3(b) have centered distributions with similar variance. Moreover, the two-dimensional density map in Fig. 3(c) shows that salient inter-feature relations are preserved, and extreme outliers are virtually suppressed. These findings validate that preprocessing enhances the quality and stability of data without compromising the nature of processes. The autoencoder is trained using fault free samples of the TEP data, which is the source domain as described in section 2.3. TEP gives a normal large, stable and well characterized operating regime, thus making the underlying normal process manifold to be learned reliably.

Training on one clean domain avoids mixing heterogeneous process dynamics during representation learning, while the subsequent cross-domain evaluation provides a direct assessment of model transferability under domain shift.

Fig. 4 shows the result of training convergence. The training loss as well as validation loss (MSE) decrease gradually and stabilize after about 30-40 epochs shown in Fig. 4(a). The curves are not divergent, indicating the optimization is stable and not overfitted. Fig. 4(b) shows the development of MAE, including training and validation curves close together during the training process, indicating a consistent reconstruction accuracy and a negligible generalization gap. Table 3 summarizes the final training statistics with low reconstruction errors and convergent values. The autoencoder uses 62 input features to encode them into a latent space with 5 dimensions, with a compression ratio of 12.4:1 and encodes all the necessary process information. Fig. 5 further illustrates the effectiveness of the learned representation. Fig. 5(a) indicates that normal samples are clumped at the low levels of reconstruction error, but faulty samples tend to be systematically deviated. Statistically, this behavior is confirmed in Fig. 5(b), faulty samples generated much higher reconstruction errors. These findings confirm reconstruction error as a sound unsupervised measure of the abnormal behavior of processes.

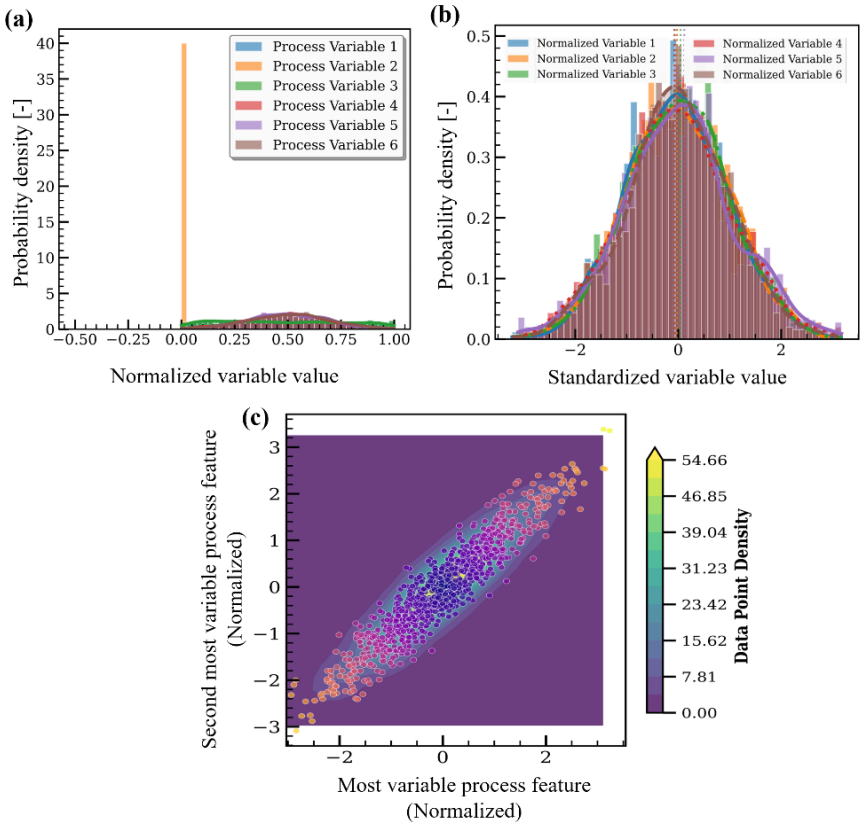


Fig. 3 (a) Raw process variable distributions before preprocessing; (b) Standardized feature distributions after normalization; (c) Two-dimensional density map of the most variable normalized features.

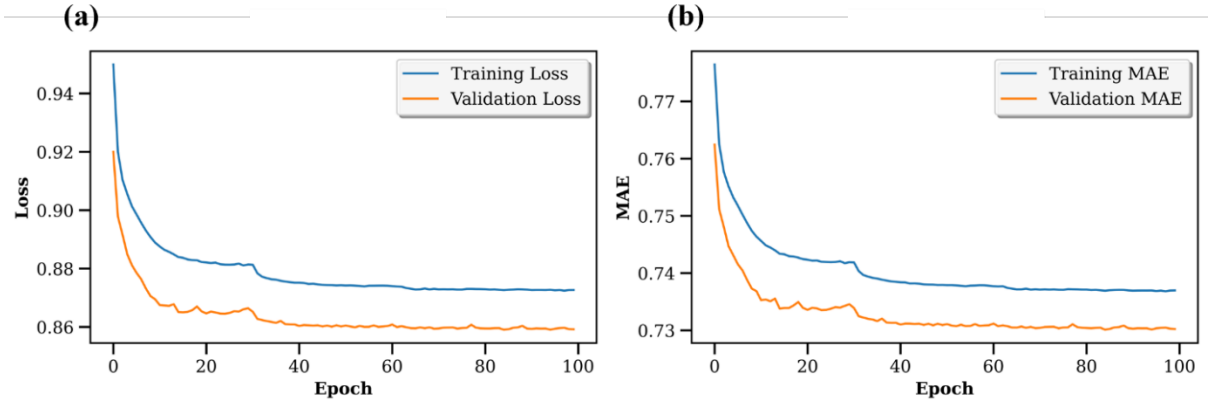


Fig. 4 Training convergence learning curves: (a) Model loss; (b) Model MAE

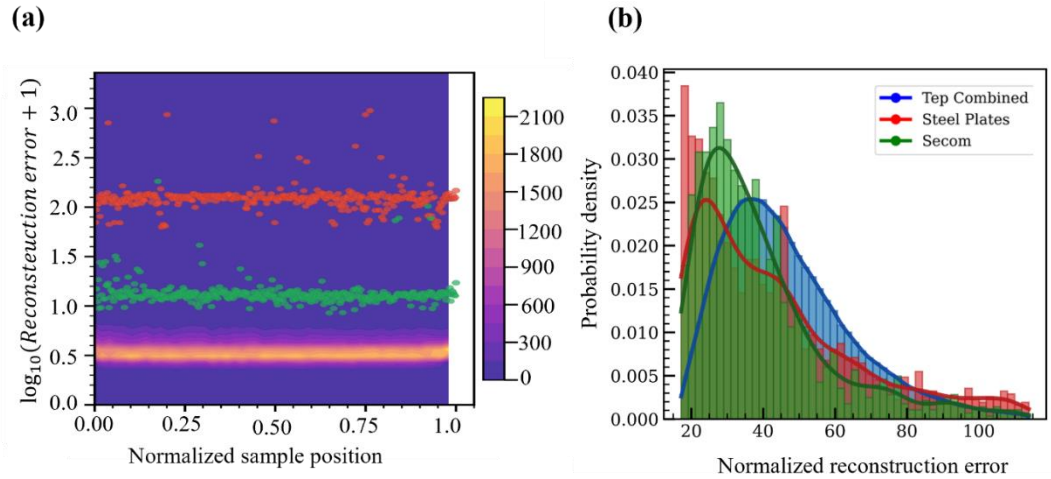


Fig. 5 Autoencoder-based feature extraction and reconstruction error analysis: **(a)** Statistical analysis of reconstruction errors; **(b)** 2D histogram of reconstruction errors.

Table 3 Autoencoder Training and Reconstruction Statistics

Metric	Value
Final training loss (MSE)	0.873
Final validation loss (MSE)	0.859
Final training MAE	0.737
Final validation MAE	0.730
Avg. reconstruction error (TEP faulty)	14.81

3.2 Latent Space Analysis

The latent representations learned by the autoencoder provide a compact description of normal process behavior. However, latent samples corresponding to normal and faulty conditions often overlap, making direct fault separation difficult. To overcome this, the latent features are combined with picked raw variables and engineered features, which both represent the global and fault-sensitive local process structure. The selection based on mutual information is used to obtain the most informative raw variables and a fused feature space of 29 dimensions is obtained. Fig. 6 illustrates how supervised learning affects fault discrimination. The smoothing and aligning distributions of the kernel density estimates the predicted fault probabilities as shown in Fig. 6(a), indicating a stable and well-calibrated output of the classifier. In addition, Fig. 6(b) illustrates the latent feature space before the supervised learning, where both normal and faulty samples are partially overlapped, with near proximity between the centers of the classes. Furthermore, supervised classification yields a better separation between normal and fault centers as shown in Fig. 6(c) with much more compact and structured density contours. This overlap explains the moderate standalone performance observed for reconstruction-based detection in the benchmarking results. The proposed fusion strategy and supervised classification stage significantly improve class separability and yield reliable probabilistic fault scores. These probabilities are combined with reconstruction errors in order to build ACCs in the subsequent section.

3.3 Classification Performance

Fig. 7(a) shows the ROC curves of the evaluated classifiers using the TEP dataset, whereas Fig. 7(b) shows the corresponding precision–recall curves, which better reflect performance under class imbalance. Table 4 summarizes the results of the tested models, among which Gradient Boosting achieved the highest F1-score (0.604) and AUC (0.762) on the evaluated TEP test set. Despite the moderate accuracy (0.616), such mechanism is not counterintuitive given the high-class imbalance and a weak separability of pseudo-labels produced after reconstruction errors, as opposed to actual fault labels produced. In this context, accuracy is less informative than recall and F1-score. These metrics better reflect detection performance, as missed faults are typically more critical than false alarms in industrial monitoring. Therefore, Gradient Boosting is selected because it achieves higher recall (0.675) while maintaining stable precision.

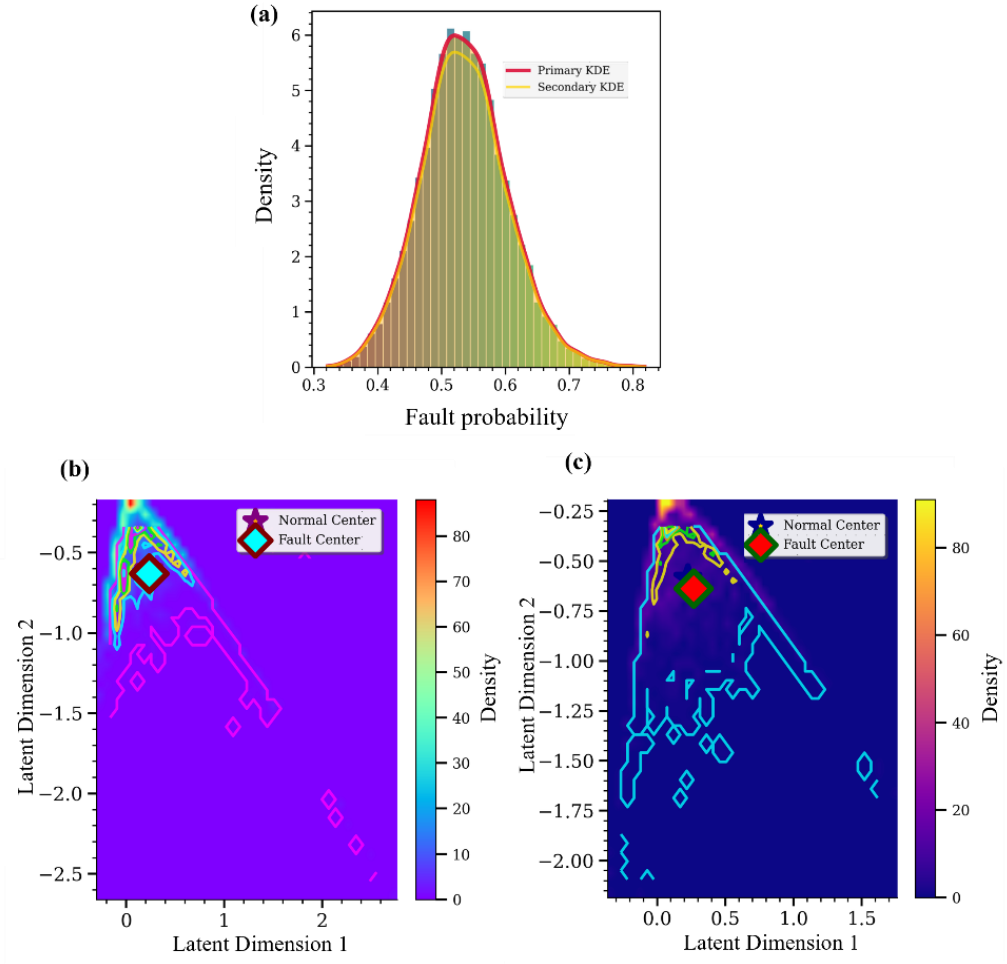


Fig. 6 Classifier probability and latent feature distributions: **(a)** Distribution of predicted fault probabilities; **(b)** SECOM latent feature density in 2D space; **(c)** Steel Plates latent feature density in 2D space.

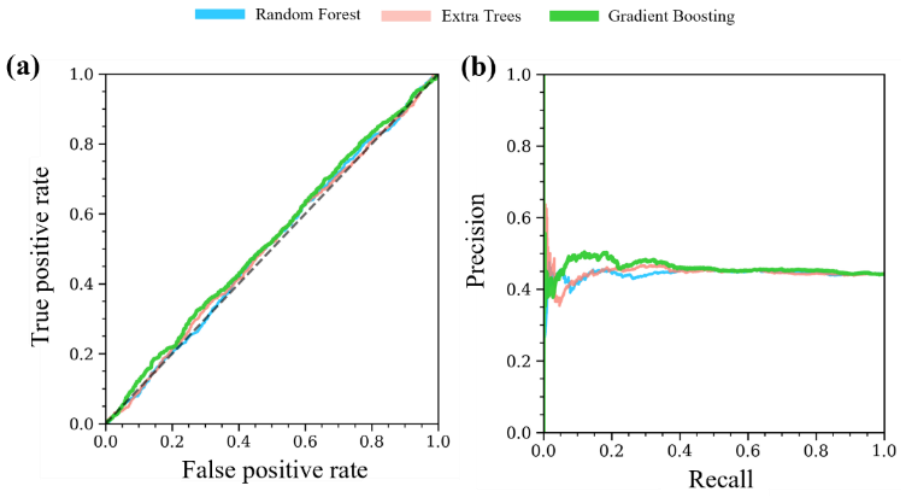


Fig. 7 Classifier performance analysis: (a) ROC curves for all classifiers; (b) Precision-Recall curves.

This makes it suitable for weakly supervised IAD, where missed faults are critical than false alarms. However, these include the indirect construction of supervisory labels, limited class separability in the learned latent space, and the need to subsample large-scale datasets for computational feasibility. The current study is a promising future research as it aims to address these issues by domain-adaptive labeling approaches, as well as more powerful supervisory signals. It should be noted that the reported classification performance reflects discrimination under weak supervision rather than recovery of true fault classes.

Table 4 Classification performance comparison on the TEP dataset

Classifier	Accuracy	Precision	Recall	F1-score	AUC
Random Forest	0.593	0.512	0.561	0.535	0.691
Extra Trees	0.540	0.521	0.559	0.539	0.684
Gradient Boosting	0.616	0.547	0.675	0.604	0.762

3.4 Adaptive Control Chart Analysis

This section illustrates the ultimate results of IAD by the proposed SML-ACC framework. Fig. 8(a) shows the ACC of TEP dataset. The anomaly scores are near the center line when the system is running normally. In the cases where faults are formed, there is sustained exceedance of the adaptive upper control limit (UCL) in scores. These overruns come in adjacent blocks. Such behavior shows that there are long-lasting abnormal conditions. The detected anomalies are of 1,222 among 10,000 samples, and the detection rate is 12.2%. Fig. 8(b) shows the ACC of Steel Plates dataset. In this case, the anomaly score exhibits greater variation, consistent with the characteristics of the manufacturing data. The adaptive UCL changes with the updated anomaly-score distribution, providing a decision boundary that responds to temporal variation. The resulting detection rate is 15.3%. The SECOM dataset shown in Fig. 8(c), shows that high-dimensional sensor noise results in the tightening of the anomaly separation and a decreased detection rate of 8.4%. Table 5 summarizes the sample size, mean anomaly score, score dispersion, adaptive UCL, and anomaly detected in all datasets. The table confirms that the control limits vary in a consistent manner with various characteristics of data. In general, the ACC presents a comprehensible and intensive decision layer. Detected anomalies represent deviations of the anomaly score beyond the adaptive control limits rather than precise fault timestamps, which is consistent with the weakly supervised setting. Overall, the ACC provides an interpretable adaptive decision layer for industrial anomaly monitoring by converting the fused anomaly score into statistically defined monitoring decisions.

Table 5 Adaptive control chart detection statistics

Dataset	Samples	Mean	Std. Dev.	UCL	Anomalies	Rate (%)
TEP	10,000	0.431	0.165	0.662	1,222	12.2
Steel Plates	1,941	0.297	0.156	0.508	297	15.3
SECOM	1,567	0.362	0.196	0.644	132	8.4

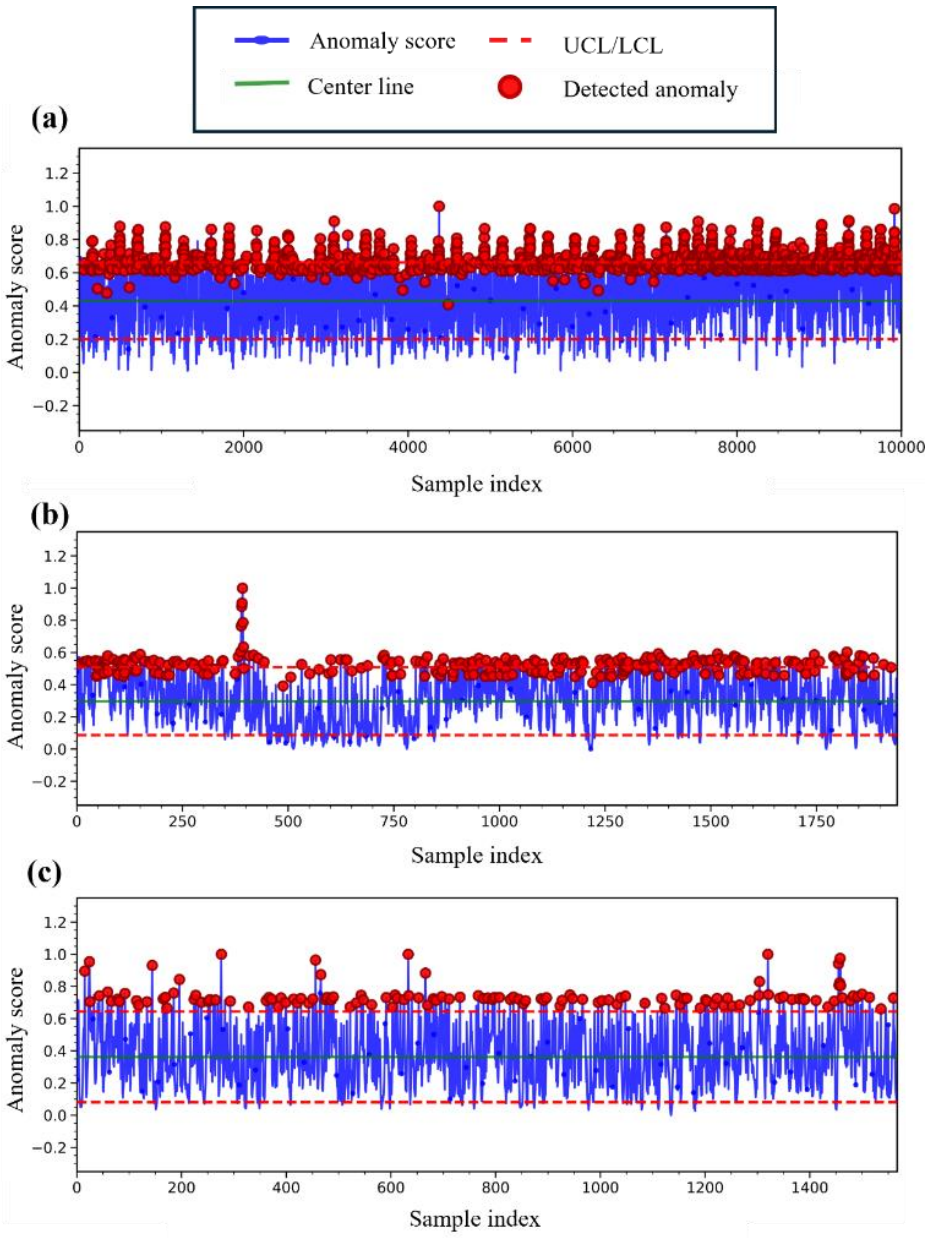


Fig. 8 Adaptive control chart: **(a)** TEP dataset; **(b)** Steel Plates dataset; **(c)** SECOM dataset

3.5 Benchmark Comparison

The SML-ACC framework is compared with classical SPC, unsupervised baselines, Shewhart, EWMA, CUSUM, PCA-SPE, and Isolation Forest. The quantitative results of all the datasets are presented in Table 6, and Fig. 9 represents a more specific comparison on the TEP dataset. Baseline methods are applied with well-known settings as mentioned earlier in the industrial fault detection literature in order to provide fair and repeatable comparison.

Fig. 9(a) demonstrates the ROC curves of the TEP dataset. TEP is limited to visualization since it is the source domain on which training is performed and has large sample sizes, allowing threshold-independent measures to be estimated in a stable manner. As noted, SML-ACC has the best AUC (0.762), indicating that it enables to discriminate the best among all baseline techniques. The corresponding F1-score is shown in Fig. 9(b). Furthermore, SML-ACC has the highest F1-score (0.604) which is more balanced between false alarms and missed detection compared to traditional SPC processes. Table 6 comparatively presents the findings across the three datasets to examine performance across different industrial domains. In addition, SML-ACC achieved higher F1-score than the evaluated baseline methods on the Steel Plates dataset and showed comparable performance on the SECOM dataset. The framework is judged to be suitable on the both numerical significance and physical interpretability. On the numerical significance, the model shows better F1 and AUC, and on the physical interpretability, the ACCs reflect sustained process deviations and not random noise.

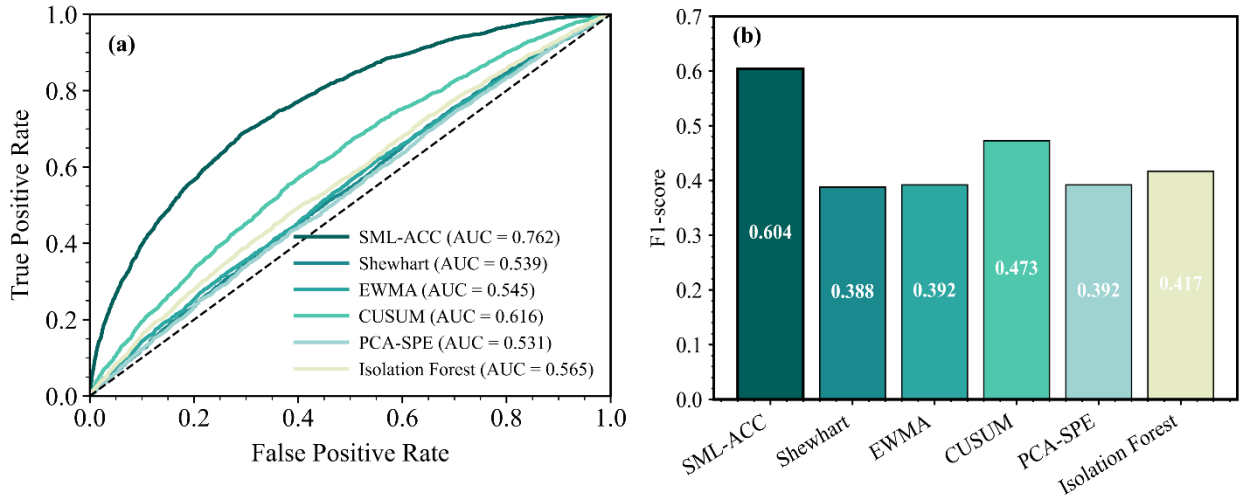


Fig. 9 Performance comparison on the TEP dataset: **(a)** ROC curves of SML-ACC and baseline methods; **(b)** corresponding F1-score.

Table 6 Performance comparison of SML-ACC with baseline methods

Dataset	Method	F1 Score	Precision	Recall	AUC	FAR	FDR
TEP	**SML-ACC**	**0.604**	**0.547**	**0.675**	**0.762**	**0.268**	**0.453**
	Shewhart	0.388	0.351	0.434	0.539	0.384	0.649
	EWMA	0.392	0.355	0.438	0.545	0.382	0.645
	CUSUM	0.473	0.428	0.529	0.616	0.338	0.572
	PCA-SPE	0.392	0.355	0.438	0.531	0.382	0.645
	Isolation Forest	0.417	0.378	0.467	0.565	0.368	0.622
Steel Plates	**SML-ACC**	**0.737**	**0.871**	**0.639**	**0.841**	**0.052**	**0.129**
	Shewhart	0.430	0.406	0.457	0.555	0.368	0.594
	EWMA	0.440	0.415	0.467	0.563	0.363	0.585
	CUSUM	0.438	0.414	0.466	0.575	0.363	0.586
	PCA-SPE	0.356	0.336	0.379	0.444	0.411	0.664
	Isolation Forest	0.365	0.344	0.388	0.526	0.407	0.656
SECOM	**SML-ACC**	**0.522**	**0.513**	**0.532**	**0.708**	**0.136**	**0.487**
	Shewhart	0.427	0.439	0.414	0.651	0.143	0.561
	EWMA	0.389	0.360	0.423	0.637	0.203	0.640
	CUSUM	0.342	0.262	0.492	0.574	0.375	0.738
	PCA-SPE	0.510	0.472	0.556	0.735	0.168	0.528
	Isolation Forest	0.504	0.519	0.489	0.735	0.122	0.481

3.6 Cross-domain Evaluation

This section examines the cross-domain transferability of the SML-ACC framework under direct application without retraining. The model is trained exclusively on the TEP dataset and directly applied to the Steel Plates and SECOM datasets without retraining or domain-specific adaptation. This setup is used to examine performance degradation under domain shift rather than to establish semantic equivalence between the feature spaces of the datasets. Fig. 10 illustrates a heatmap representation of the performance of cross-domain anomaly detection between F1-score, precision, recall, and detection rate, which intuitively compares the performance across the source and target domains. The respective quantitative data are mentioned in Table 7**.**

Table 7 Cross-Domain Anomaly Detection Performance of SML-ACC

Dataset	F1-score	Precision	Recall	Detection Rate
TEP (Source)	0.650	0.877	0.700	0.135
Steel Plates	0.450	0.666	0.500	0.161
SECOM	0.450	0.441	0.500	0.087

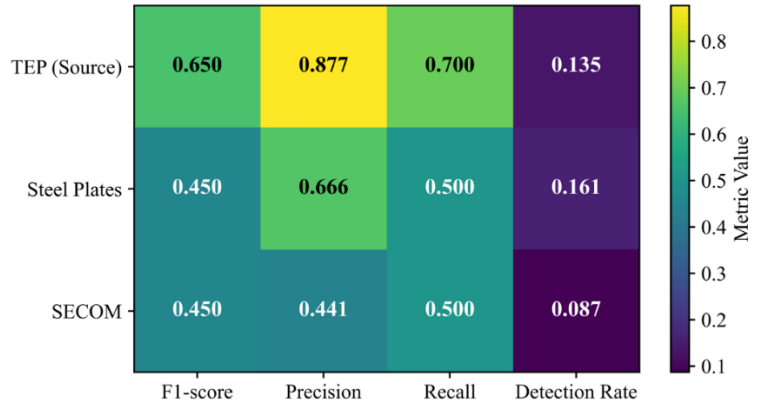


Fig. 10 Heat map of cross domain anomaly detection performance

SML-ACC achieves the F1 score of 0.650 on the input TEP data, which has high precision and recall scores of 0.877 and 0.700, respectively. Moreover, the F1-score on the Steel Plates and SECOM datasets drops to 0.450. This gives an approximation of an F1-score ratio of 69.2%, compared to the F1-score from the source domain of 0.650. The target datasets have similar recall values (0.500) as shown in Fig. 10 despite the domain shift. On the other hand, accuracy drops, especially in the case of SECOM, reflecting a larger percentage of false positive detections. The degradation may be caused by the distribution of features, the noise characteristics, the dimensions of the features and the dynamics of the processes in the different datasets. The results illustrate that no retraining is needed to apply the framework to the target datasets, but the detection performance suffers from the observed domain shifts. The results should thus be viewed as a measure of the direct transferability across domains instead of as evidence for semantic feature alignment or domain-invariant generalization. Fig. 10 shows the tradeoffs between sensitivity and accuracy of detection in various domains. This observation highlights the domain adaptation strategies should be considered in future work in order to enhance robustness and decrease the false positives.

3.7 Ablation Study

The role of every element in SML-ACC framework is explored in a systematic ablation study. One or more individual modules are eliminated leaving the rest of the pipeline intact. Fig. 11(a) shows the comparison of F1-score. The full SML-ACC has the highest F1-score (0.613), whereas the absence of an autoencoder, classifier, adaptive limits, fusion strategy or post-processing reduce performance. This provides evidence that the individual components contribute to the overall performance of the complete SML-ACC framework. The FAR and FDR, as shown in Fig. 11(b) denote operational robustness. The proposed model has the lowest FAR and FDR of 0.222 and 0.401, respectively, highlighting that ablated variant have significantly higher false alarms and false discoveries. Table 8 summarizes all metrics (F1, Precision, Recall, AUC, FAR, and FDR). The full SML-ACC framework provides the best trade-off between detection accuracy and reliability. Therefore, the SML-ACC configuration is selected as the final model, as it simultaneously achieves the highest F1-score and lowest FAR and FDR, ensuring both accurate detection and reliable false-alarm control.

Table 8 Results of ablation study (average across datasets)

Configuration	F1	Precision	Recall	AUC	FAR	FDR
Full SML-ACC	0.6125	0.5993	0.6659	0.7678	0.2220	0.4007
No EWMA	0.5471	0.4943	0.6628	0.7295	0.3183	0.5057
No Post-Processing	0.5233	0.4449	0.6514	0.6975	0.3505	0.5551
No Adaptive Limits	0.4996	0.4160	0.6495	0.6688	0.3978	0.5840
No Autoencoder	0.4890	0.4052	0.6337	0.6559	0.4030	0.5948
Minimal	0.4829	0.4518	0.5396	0.6186	0.3448	0.5482
No Classifier	0.4561	0.3818	0.5670	0.6042	0.3970	0.6182
No Fusion	0.4561	0.3818	0.5670	0.6042	0.3970	0.6182

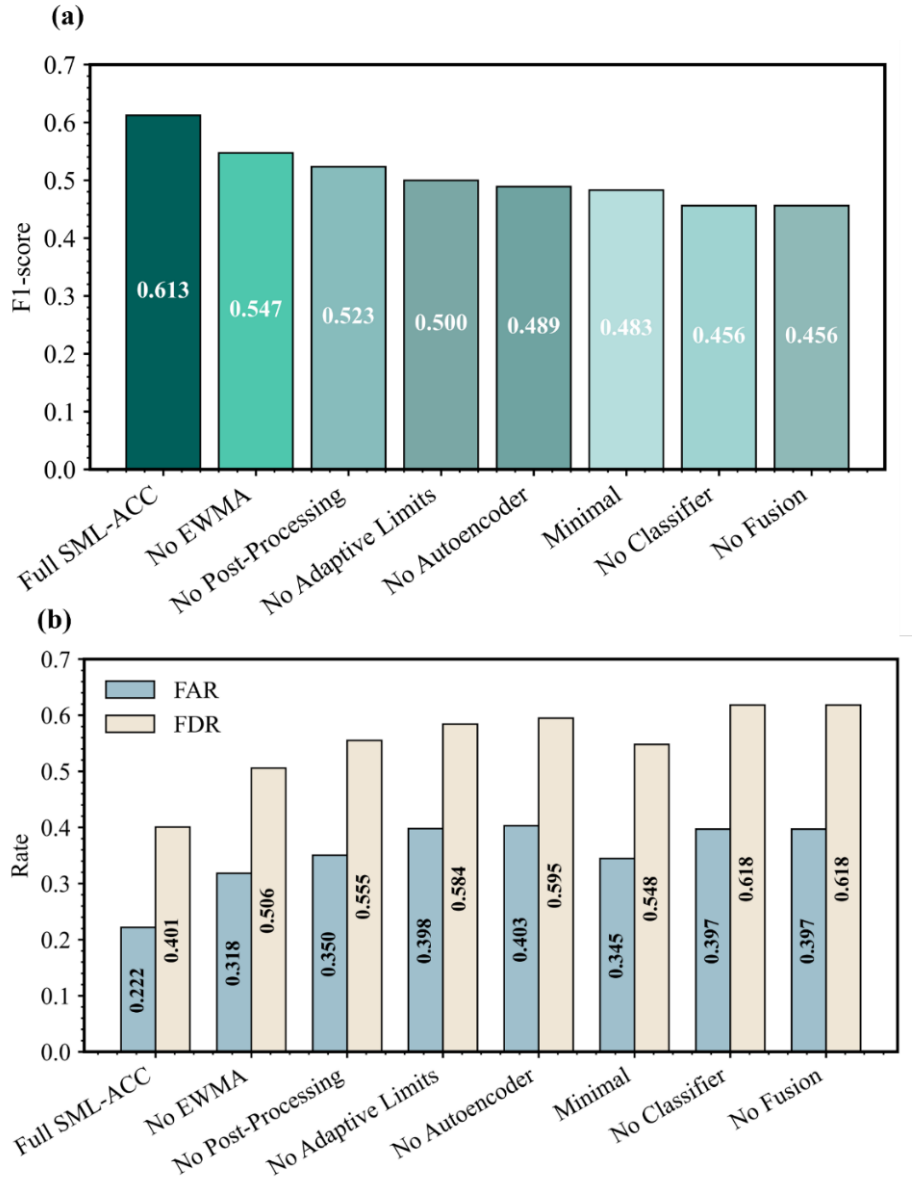


Fig. 11 Results of ablation study (average across all datasets): (a) average F1-score for different SML-ACC configurations; (b) average FAR and FDR.

3.8 Statistical Validation

A five-fold cross-validation approach was used to obtain the fold-level F1-scores and evaluate their statistical significance. Each baseline method was evaluated with the same shuffled five-fold data partitions, using a fixed random seeding of 42 to ensure paired and directly comparable evaluations with SML-ACC. For each pair-wise comparison between SML-ACC and one of the baseline methods at each fold-level in each data set, five observations of the F1 score were collected. Two-sided paired t-tests were used to determine if the differences in F1-scores between SML-ACC and the baseline methods were statistically significant. The null hypothesis (H_0) was that the means of the paired differences between the F1-score of SML-ACC and the baseline method were equal to zero, while the alternative hypothesis (H_1) was that the means of the paired differences were not equal to zero. For each dataset, five pairwise comparisons were made between the following methods: SML-ACC and Shewhart, EWMA, CUSUM, PCA-SPE, and Isolation Forest. The p-values obtained were corrected for multiple comparisons using the Holm-Bonferroni method within each dataset to obtain a family-wise error rate. All p-values were calculated after Holm adjustment of the alpha level of 0.05 were used for statistical significance. Prior to multiple comparisons correction, the paired t-test revealed statistically significant differences for all five baseline comparisons for the TEP dataset: Shewhart ($p = 0.0001$), EWMA ($p < 0.0001$), CUSUM ($p = 0.0006$), PCA-SPE ($p < 0.0001$), and PCA-IF ($p = 0.0003$). For the Steel Plates dataset, all five comparisons showed statistically significant differences (p values of 0.0002, 0.0005, 0.0004, 0.0004, and 0.0002 for Shewhart, EWMA, CUSUM, PCA-SPE, and Isolation Forest, respectively). Statistically significant differences were found between EWMA, CUSUM and the original datasets of SECOM ($p = 0.0376$ and $p = 0.0049$ respectively), while before correction, there were no statistically significant differences between Shewhart, PCA-SPE, and Isolation Forest. Holm adjusted p-values were used in the final significance tests for the five pairwise comparisons made within each set of data. This multiple-comparison adjustment offers a more conservative evaluation of the statistical evidence, and helps avoid false positive results. The reported statistical analysis therefore separates the unadjusted t-test results from the final t-test results after multiple comparisons corrections.

4 Conclusion

This paper introduces SML-ACC framework with weak supervision to detect fault in the industry. The framework combines deep autoencoders for representation learning, weakly supervised classification, and ACCs into a single monitoring pipeline. Fault-free data is used to learn normal process behavior and fault scores (both reconstruction error and probabilistic fault) are used together to identify abnormal conditions. The empirical results obtained from the TEP, Steel Plates, and SECOM datasets provide evidence that SML-ACC achieves competitive performance relative to the evaluated classical SPC algorithms and unsupervised baselines in terms of F1-score, AUC, and false-alarm characteristics. The statistical tests provide evidence that several of the observed performance differences are statistically significant. Additionally, ablation analysis indicates that representation learning, weak supervision, adaptive fusion, and control-limit estimation all positively contribute to overall performance. The cross-domain assessment displays moderate robustness, as it is characterized by about 69.2% performance retention without retraining. Even though the domain shift causes deterioration of performances, it does not affect the recall, which is consistent in target datasets, indicating the reliable anomaly detection in unfamiliar settings. Overall, SML-ACC offers a useful and understandable real-world industrial process monitoring solution when annotations are not much more.

5 Limitations and Future Work

There are some limitations to the proposed SML-ACC framework that outline future research directions. To begin with, the framework is based on reconstruction-error-based pseudo-labels because there are no explicit fault labels. Although this poorly monitored approach is realistic of industrial contexts, pseudo-label noise can limit the discriminative ability of the monitored classifier. Secondly, the autoencoder is trained on one source domain and used on different datasets without retraining. The cross-domain performance is, therefore, influenced by domain shift based on sensor configurations, noise levels and dynamics of the processes. To prevent this effect, the present study fails to use domain adaptation or transfer learning methods. Thirdly, zero-padding or truncation is used to align the characteristics of all datasets to meet the requirements of architectures. This method, though used in the same way, does not explicitly maintain a semantic correspondence between heterogeneous feature sets. Lastly, the framework is tested in an offline environment. Online learning, concept drift, and adaptive model updating issues are not discussed. Future work will focus on domain-invariant representation learning, online adaptation of control limits, and more reliable supervisory signals to improve robustness and long-term applicability in real industrial environments.

Acknowledgements

Authors gratefully acknowledge Department of Mechanical Engineering, Chittagong University of Engineering & Technology, Chattogram-4349, Bangladesh for motivating and supporting to do this research work.

Author Contributions

M. H. Maruf: Conceptualization, Resources, Data Curation, Formal Analysis, Writing-Original & Final draft, Visualization; **M. Rahman:** Methodology, Writing-Review and Editing, Supervision; **M. M. Roshid:** Writing-Review and Editing.

Conflict of Interest Statement

Authors declare that they have no known potential conflict of interest.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Generative AI Statement

No generative AI/LLM was used in any stage of writing this manuscript.

Data Availability

The data will be available on request.

References

[1] Deuse, J., Wiegand, M. and Weisner, K., 2019. Continuous process monitoring through ensemble-based anomaly detection. In Applications in Statistical Computing: From Music Data Analysis to Industrial Quality Improvement (pp. 289-301). Cham: Springer International Publishing.

[2] Velasquez, D., Perez, E., Oregui, X., Artetxe, A., Manteca, J., Mansilla, J.E., Toro, M., Maiza, M. and Sierra, B., 2022. A hybrid machine-learning ensemble for anomaly detection in real-time industry 4.0 systems. IEEE Access, 10, pp.72024-72036.

[3] Yang, W., Chu, W., Wu, X., Zhou, L., Wang, J., Yang, H. and Li, Z., 2025. Industrial multivariate time-series data anomaly detection incorporating attention mechanisms and adversarial training. International Journal of Computer Integrated Manufacturing, 38(12), pp.1697-1715.

[4] Sabri, H.M., Alsaedi, H.J.N. and Alwan, S.J., 2025. Detecting Gradual Trends: Integrating EWMA Control Charts with Artificial Intelligence Algorithms (LSTM).

[5] Wu, C., Liu, F. and Zhu, B., 2015. Control chart pattern recognition using an integrated model based on binary-tree support vector machine. International Journal of Production Research, 53(7), pp.2026-2040. Sabahno, H. and Niaki, S.T.A., 2023. New machine-learning control charts for simultaneous monitoring of multivariate normal process parameters with detection and identification. Mathematics, 11(16), p.3566.

[6] Chongfuangprinya, P., Kim, S.B., Park, S.K. and Sukchotrat, T., 2011. Integration of support vector machines and control charts for multivariate process monitoring. Journal of Statistical Computation and Simulation, 81(9), pp.1157-1173.

[7] Tayalati, F., Boukrouh, I., Bouhsaien, L., Azmani, A. and Azmani, M., 2024. Hybrid approach integrating deep learning-autoencoder with statistical process control chart for anomaly detection: Case study in injection molding process. IEEE Access, 12, pp.95576-95598.

[8] Sabri, H.M., Alsaedi, H.J.N. and Alwan, S.J., 2025. Detecting Gradual Trends: Integrating EWMA Control Charts with Artificial Intelligence Algorithms (LSTM).

[9] Azzaka, M.R., Ihsan, A.F. and Astuti, W., 2025, July. Implementation of Ensemble Learning Methods in Autoencoder Architecture for Anomaly Detection in Natural Gas Pipeline Operational Data. In 2025

International Conference on Data Science and Its Applications (ICoDSA) (pp. 281-286). IEEE.

[10] Yeganeh, A., Abbasi, S.A., Pourpanah, F., Shadman, A., Johannssen, A. and Chukhrova, N., 2022. An ensemble neural network framework for improving the detection ability of a base control chart in non-parametric profile monitoring. Expert Systems with Applications, 204, p.117572.

[11] Givnan, S., Chalmers, C., Fergus, P., Ortega-Martorell, S. and Whalley, T., 2022. Anomaly detection using autoencoder reconstruction upon industrial motors. Sensors, 22(9), p.3166.

[12] Kumar, P., Pandi, S.S., Kumar, L.B. and Karthick, R., 2025, April. Anomaly detection in industrial control systems using machine learning. In 2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS) (pp. 1-6). IEEE.

[13] Kumar, P., Pandi, S.S., Kumar, L.B. and Karthick, R., 2025, April. Anomaly detection in industrial control systems using machine learning. In 2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS) (pp. 1-6). IEEE.

[14] Annaswamy, A.M., Guha, A., Cui, Y., Tang, S., Fisher, P.A. and Gaudio, J.E., 2023. Integration of adaptive control and reinforcement learning for real-time control and learning. IEEE Transactions on Automatic Control, 68(12), pp.7740-7755.

[15] Tan, Y., Jin, B., Nettekoven, A., Chen, Y., Yue, Y., Topcu, U. and Sangiovanni-Vincentelli, A., 2019, December. An encoder-decoder based approach for anomaly detection with application in additive manufacturing. In 2019 18th IEEE international conference on machine learning and applications (ICMLA) (pp. 1008-1015). IEEE.

[16] Sun, Q. and Ge, Z., 2020. Deep learning for industrial KPI prediction: When ensemble learning meets semi-supervised data. IEEE Transactions on Industrial Informatics, 17(1), pp.260-269.

[17] Ma, F., Ji, C., Wang, J., Sun, W., Tang, X. and Jiang, Z., 2024. MOLA: enhancing industrial process monitoring using a multi-block orthogonal long short-term memory autoencoder. Processes, 12(12), p.2824.

[18] Michau, G. and Fink, O., 2021. Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer. Knowledge-Based Systems, 216, p.106816.

[19] Jiang, M., Hou, C., Zheng, A., Hu, X., Han, S., Huang, H., He, X., Yu, P.S. and Zhao, Y., 2023. Weakly supervised anomaly detection: A survey. arXiv preprint arXiv:2302.04549.

[20] Elder IV, J.J.F., Fogelman-Soulié, F., Flach, P. and Zaki, M., 2009. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. Association for Computing Machinery (ACM).

[21] Rizve, M.N., Duarte, K., Rawat, Y.S. and Shah, M., 2021. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. arXiv preprint arXiv:2101.06329.

[22] Arazo, E., Ortego, D., Albert, P., O'Connor, N.E. and McGuinness, K., 2020, July. Pseudo-labeling and confirmation bias in deep semi-supervised learning. In 2020 International joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.

[23] Zhan, D., Zhang, W., Ye, L., Yu, X., Zhang, H. and He, Z., 2024. Anomaly detection in industrial control systems based on cross-domain representation learning. IEEE Transactions on Dependable and Secure Computing, 22(3), pp.2505-2518.

[24] Noor-ul-Amin, M., Kazmi, M.W., Alkhalaf, S., Abdel-Khalek, S. and Nabi, M., 2024. Machine learning based parameter-free adaptive EWMA control chart to monitor process dispersion. Scientific Reports, 14(1), p.31271.

[25] Rieth, C.A., Amsel, B.D., Tran, R. and Cook, M.B., 2017. Additional tennessee eastman process simulation data for anomaly detection evaluation. Harvard Dataverse, 1, p.2017.

[26] Buscema, M. and TS, T., W.: Steel Plates Faults. UCI Machine Learning Repository (2010) [online]

[27] Asuncion, A. and Newman, D., 2007. UCI machine learning repository [online]

[28] Hinton, G.E. and Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. science, 313(5786), pp.504-507.

[29] Roberts, S.W., 2000. Control chart tests based on geometric moving averages. Technometrics, 42(1), pp.97-101.

[30] Hinton, G.E. and Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. science, 313(5786), pp.504-507.

[31] Roberts, S.W., 2000. Control chart tests based on geometric moving averages. Technometrics, 42(1), pp.97-101.