Journal of Applied Automation Technologies

Research Paper

Semi-supervised Transfer Learning Method for Anomaly Detection in Industrial Sensors

Rafał Dawid Kosior[1] and Lucyna Kamińska[1, *]

[1] Faculty of Electrical and Automatic Systems, State Higher School in Krosno, Krosno, 38-400, Poland
*Corresponding author: rafal.dk@pwsz-krosno.edu.pl

Abstract. The deployment of industrial sensors has become fundamental to intelligent manufacturing, enabling continuous monitoring and predictive maintenance across various factory environments. However, practical anomaly detection remains challenging due to two persistent obstacles: the scarcity of labeled fault data and significant differences in data distributions between operational domains. This study addresses these issues by proposing a novel semi-supervised transfer learning approach for industrial sensor anomaly detection. The method integrates adversarial domain adaptation with consistency regularization and dynamic pseudo-labeling to fully utilize both labeled and abundant unlabeled data from heterogeneous domains. Experiments are conducted on multiple real-world and benchmark sensor datasets, where the proposed framework is evaluated against supervised, unsupervised, and state-of-the-art domain adaptation baselines. Results show that the new approach achieves significantly improved detection accuracy, robustness to domain shifts, and exceptional data efficiency; for example, reliable anomaly detection can be achieved with as little as 2–5% labeled target data, minimizing annotation costs and deployment delays. The system demonstrates strong adaptability, effectively generalizing across diverse equipment, process lines, and environmental conditions. These findings highlight the practicality and scalability of the proposed framework, offering an efficient path toward reduced manual intervention and enhanced operational reliability in next-generation industrial environments. The study demonstrates that semi-supervised transfer learning has substantial potential in reducing barriers for data-driven maintenance and improving the intelligence and resilience of industrial systems.

Keywords: *Machine Learning, Transfer Learning, Anomaly Detection, Industrial Sensors*

Received on 29 March 2025, Accepted on 26 August 2025, Published on 05 September 2025

Introduction

The rapid transformation of modern industry is being catalyzed by the convergence of industrial automation and the Industrial Internet of Things (IIoT), which has led to unprecedented advancements in process optimization, asset management, and intelligent manufacturing. At the heart of these developments lie large-scale sensor networks, which ubiquitously monitor diverse operational states across factories, supply chains, and critical infrastructures [1,2]. Industrial sensors continuously provide high-resolution, real-time data streams on equipment condition, environmental variations, and process quality, underpinning predictive maintenance strategies and the foundation of smart factories [3]. As the number and variety of deployed sensors increase, the effective analysis of sensor data has become indispensable for sustaining safe, resilient, and efficient industrial operations [4]. Within this context, anomaly detection in sensor data has emerged as an essential capability, enabling the early identification of equipment faults, process deviations, and security threats that could otherwise result in costly downtimes or catastrophic failures [5,6].

Despite remarkable progress in machine learning-driven data analytics, several critical challenges fundamentally limit the widespread adoption and effectiveness of automated anomaly detection in complex industrial environments. A pervasive obstacle is the severe scarcity of labeled anomaly data: in high-value manufacturing or utility sectors, abnormal events are rare, and the annotation process is labor-intensive, domain-specific, and

often impractical to scale [7,8]. Moreover, industrial facilities typically exhibit pronounced heterogeneity, both in terms of sensor hardware and operating contexts. This leads to substantial domain discrepancies—data distributions from one machine, line, or site may differ dramatically from another due to variations in equipment age, operational mode, environmental factors, and maintenance policies [9,10]. Most classic supervised anomaly detection models presume statistically similar training and deployment environments. In real-world industry, however, this assumption is commonly violated, causing dramatic drops in accuracy when models are transferred across domains [11,12]. Furthermore, completely unsupervised methods, while sidestepping the need for labels, often lack sensitivity and generate excessive false alarms, especially when faced with subtle, evolving, or rare fault patterns [13]. As a result, conventional approaches struggle to deliver high-precision, generalizable, and robust anomaly detection under the combined pressures of label scarcity and cross-domain adaptation requirements [14].

In response to these intertwined challenges, this study proposes a principled and scalable framework that synergistically integrates semi-supervised learning and transfer learning for industrial sensor anomaly detection. Unlike traditional approaches, our methodology is designed to maximize the utility of both labeled and abundant unlabeled sensor data, while explicitly addressing inter-domain distributional disparities through advanced adaptation mechanisms. The key innovation of this work consists of a deep semi-supervised transfer learning paradigm leveraging adversarial domain adaptation, consistency regularization, and pseudo-label refinement, ensuring that models trained on source domains remain effective on disparate and sparsely labeled target domains. This integrated approach not only alleviates the burden of exhaustive manual annotation but also substantially enhances model generalization—thereby empowering next-generation smart manufacturing and predictive industrial analytics. The remainder of this paper is organized as follows: Section 2 formalizes the problem setting and surveys related transfer learning theory in industrial anomaly detection. Section 3 details the proposed methodology, presenting the overall model architecture and semi-supervised adaptation strategy. Section 4 describes experimental settings, datasets, and baseline comparisons. Section 5 provides a comprehensive analysis of results, ablation studies, and cross-domain insights. The paper concludes in Section 6 by summarizing contributions, discussing practical implications, and highlighting future research directions.

Theoretical Preliminaries

Problem Definition

In modern industrial environments, the deployment of distributed sensors across machinery, production lines, and process plants is foundational for real-time monitoring and intelligent control. Typically, industrial sensor systems generate large volumes of time-series data, often high-dimensional and streaming, encompassing measurements such as temperature, vibration, pressure, acoustic emissions, current, and flow rate. These measurements are not only heterogeneous in type and scale but also demonstrate temporal dependencies and complex correlations arising from physical, electrical, or logic couplings within the system [15,16]. Anomaly detection, within this context, is formulated as the identification of abnormal samples -those deviating from the expected operational patterns-indicative of faults, system degradation, process drift, or external disturbances.

Formally, an industrial anomaly detection system learns from a dataset $X = \{x_i\}$ and corresponding labels $Y = \{y_i\}$, where x_i represents a multi-dimensional sensor observation and $y_i \in \{0,1\}$ distinguishes normal states from anomalies. However, in practice, abnormal conditions are rare and the annotation process is not only labor-intensive but also prone to subjectivity and delay. The resulting datasets are thus highly imbalanced, with a vast predominance of normal instances and only a sparse collection of labeled anomalies [17]. Moreover, the cost and feasibility of obtaining labeled examples sharply escalate in high-risk or mission-critical applications (e.g., aerospace, chemical processing), where abnormal scenarios can be hazardous or even physically impossible to create for training purposes.

A further complication arises from the inherent heterogeneity of industrial systems. Data collected from different machines-sometimes nominally identical-often exhibit substantial differences due to manufacturing tolerances, maintenance histories, sensor aging, and local environmental factors. For example, a vibration anomaly present in one production line may have subtly different signatures in another, even for the "same" equipment, due to variations in load profiles or ambient conditions [18]. This leads to the classic domain

adaptation problem: a model trained on a well-annotated source domain (e.g., one machine or factory) does not necessarily perform reliably when deployed on a new, unlabeled or sparsely labeled target domain. As such, robust industrial anomaly detection demands algorithms capable of leveraging limited labeled data, extracting discriminative features from high-dimensional signals, and adapting effectively to new operational settings, all while minimizing false positives that can otherwise lead to costly and unnecessary interventions [19].

In summary, the central problem addressed in this paper can be articulated as follows: how to design an industrial anomaly detection framework that utilizes both labeled and abundant unlabeled sensor data from a source domain, and generalizes reliably to target domains with different data distributions, minimal labeled data, and pronounced real-world variability in operation and environment. Addressing this challenge is not only pivotal for the practical adoption of reliable condition monitoring, but is also a critical step toward realizing the promise of predictive, autonomous, and data-driven smart factories [20].

Transfer Learning in Industrial Anomaly Detection

Transfer learning, particularly the domain adaptation sub-field, has emerged as a promising approach for closing the distributional gap between source and target datasets. At its core, transfer learning seeks to transfer knowledge—be it features, model parameters, or latent representations—from a source domain where labeled data are abundant or easier to obtain, to a target domain where such labels are missing or costly [21]. In the industrial context, domain adaptation enables the development of anomaly detection models that are not only accurate on the domain they are trained on, but also generalizable to new equipment, production lines, or even different operating environments.

The foundation of most transfer learning strategies lies in identifying shared representations that are robust to domain shifts, often achieved through feature alignment techniques such as adversarial training, statistical distribution matching, or subspace projection. For time-series sensor data, however, this challenge is particularly pronounced: differences in sampling rates, sensing modalities, operational regimes, and sensor drift introduce complex, high-dimensional discrepancies that simple statistical adjustment cannot fully resolve [22]. Additionally, industrial sensor signals might exhibit nonstationary behaviors and subtle contextual dependencies, making domain adaptation not just a matter of feature transformation, but also of preserving temporal and physical relevance in transferred knowledge.

Despite these advances, transfer learning faces unique hurdles in industrial anomaly detection. First, most traditional approaches assume at least moderate volumes of labeled target data to guide adaptation, which is infeasible for rare-event, safety-critical systems. Second, over-alignment may occur—models might inadvertently suppress domain-unique anomaly signatures or introduce negative transfer, degrading performance when source and target environments are too distinct [23]. Third, model interpretability and the ability to identify meaningful features remain open challenges, particularly important for industrial practitioners who need actionable insights alongside detection alerts.

Recent research has begun to integrate semi-supervised learning with transfer learning to exploit the abundance of target-domain unlabeled data, using techniques such as pseudo-labeling, consistency regularization, and co-training to better bridge source-target gaps without extensive annotation effort [24]. Nevertheless, a robust, scalable, and theoretically sound framework for anomaly detection under the dual constraints of extreme data sparsity and domain divergence remains under active investigation.

Methodology

Overall Architecture and Training Strategy

The unpredictable diversity and nonstationarity of industrial sensor environments, together with the limited availability of labeled anomaly data, necessitate robust strategies for building adaptable anomaly detection models. The framework designed in this work centers around a semi-supervised transfer learning architecture that integrates adversarial domain alignment and hybrid optimization to support effective anomaly detection independent of the data collection domain.

The model consists of four principal modules: a source feature extractor F_s, a target feature extractor F_t, a domain discriminator D_t and an anomaly detector C. As illustrated in Figure 1, labeled samples from the source domain $\left(x^s, y^s\right)$ and unlabeled samples from the target domain x^t are simultaneously processed via their respective extractors. Both extractors utilize deep neural backbones adapted to the structure and scale of time-series sensor data, while the architecture allows for domain-specific variations at the representation level.

The anomaly detector maps extracted source features to anomaly predictions. Training begins with a supervised objective on the labeled source set, using the binary cross-entropy loss:

$$L_{sup} = -\frac{1}{N_s}\sum_{i=1}^{N_s}\left[y_i^s \log \hat{y}_i^s + \left(1-y_i^s\right)\log\left(1-\hat{y}_i^s\right)\right] \quad \text{Eq. (1)}$$

where N_s is the number of source samples, y_i^s the ground truth label, and $\hat{y}_i^s = C\left(F_s\left(x_i^s\right)\right)$ the predicted label.

To achieve domain adaptation, the domain discriminator D is trained to distinguish whether a feature comes from the source or target extractor. The feature extractors are trained adversarially to fool the discriminator, promoting domain-invariant encoding. The adversarial loss for a given batch is defined as:

$$L_{\text{adv}} = -\frac{1}{N_s}\sum_{i=1}^{N_s}\log D\left(F_s\left(x_i^s\right)\right) - \frac{1}{N_t}\sum_{j=1}^{N_t}\log\left(1-D\left(F_t\left(x_j^t\right)\right)\right) \quad \text{Eq. (2)}$$

Here, N_t is the number of target samples. The overall adversarial training can be described as a min-max game:

$$\min_{F_s, F_t, C}\max_{D}\left(L_{sup} + \lambda_1 L_{adv} + \lambda_2 L_{ssl}\right) \quad \text{Eq. (3)}$$

where λ_1 and λ_2 are weighting coefficients, and L_{ss1} is the semi-supervised loss term, later introduced through pseudo-labeling and consistency regularization.

To quantify the degree of domain alignment, we can evaluate the Maximum Mean Discrepancy (MMD) between the distributions of source features f^s and target features f^t :

$$\text{MMD}^2\left(X^s, X^t\right) = \mathbb{E}\left[k\left(f^s, f^{s'}\right)\right] + \mathbb{E}\left[k\left(f^t, f^{t'}\right)\right] - 2\mathbb{E}\left[k\left(f^s, f^t\right)\right] \quad \text{Eq. (4)}$$

where $k(\cdot,\cdot)$ is a positive-definite kernel (e.g., Gaussian) and the expectations are taken over samples within and between domains.

Each training iteration alternates between updating the discriminator to improve domain separation, and updating the extractors and anomaly detector to simultaneously classify anomalies and minimize domain distinction. This interaction enables the network to converge toward feature spaces which both support high-fidelity anomaly detection and are robust to shifts between industrial sensor domains.

Crucially, as the system aligns feature spaces, the anomaly detector is further exposed to target domain samples-first purely unlabeled, and then with pseudo-labels-to heighten the network's sensitivity and specificity in real deployment scenarios. The integration of supervised, adversarial, and semi-supervised learning is what enables the model to perform robustly under stringent data constraints.

Figure 1 provides a schematic view of the framework, showing distinct flows for supervised learning from the source, adversarial domain adaptation, and eventual semi-supervised learning using adapted target features. This architecture is built for scalable deployment across evolving industrial Internet of Things (IIoT) networks, offering strong generalization and minimal retraining overhead as conditions or sensor domains change.

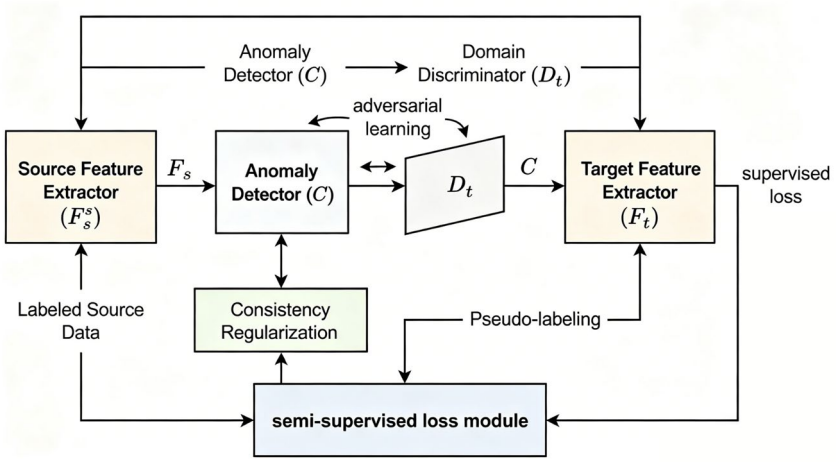

Figure 1. Structure of the Semi-supervised Transfer Learning Framework

Consistency Regularization and Pseudo-labeling Process

A major barrier to industrial anomaly detection lies in the inability to obtain comprehensive labeled datasets for every new sensor deployment or factory domain. While adversarial training achieves domain-invariant feature extraction, true semi-supervised generalization depends critically on effectively leveraging the abundance of unlabeled target samples. The approach adopted in this work further strengthens the framework by introducing a dual mechanism: consistency regularization and pseudo-labeling.

Consistency regularization addresses the requirement that a robust model should remain invariant under various plausible input transformations. For each unlabeled target-domain sample x^t , multiple perturbed versions x_a^t (e.g., via noise, time-window shift, dropout, signal augmentation) are generated. The model is encouraged to produce consistent predictions on all views of the same underlying instance. This is enforced by minimizing the divergence between the network's outputs for different augmentations of the same signal:

$$L_{con} = \frac{1}{N_t}\sum_{j=1}^{N_t} \mathrm{Dist}\left(C\left(F_t\left(x_j^t\right)\right), C\left(F_t\left(\mathrm{Aug}\left(x_j^t\right)\right)\right)\right) \quad \text{Eq. (5)}$$

where $\mathrm{Aug}(\cdot)$ is a stochastic augmentation function and Dist($\cdot,\cdot$) denotes a consistency measure (such as mean squared error or Kullback-Leibler divergence).

Pseudo-labeling is employed to further exploit the learning potential of unlabeled target samples. Once the model's confidence in a prediction for an unlabeled instance surpasses a predefined threshold, a "pseudo-label" is assigned and the sample is incorporated in training as if it were labeled. Over the course of training, this expands the set of effective labeled data within the target domain, improving model adaptation and detection specificity. For a given target sample, a pseudo-label $\hat{y}^t = \arg\max_y C\left(F_t\left(x^t\right)\right)$ is assigned if:

$$\max_y C\left(F_t\left(x^t\right)\right) > \tau \quad \text{Eq. (6)}$$

where τ is a confidence threshold hyperparameter. The pseudo-labeled loss for eligible target samples is then

$$L_{pscudo} = -\frac{1}{|T|}\sum_{x^t \subset T} \hat{y}^t \log C\left(F_t\left(x^t\right)\right) + \left(1-\hat{y}^t\right)\log\left(1-C\left(F_t\left(x^t\right)\right)\right) \quad \text{Eq. (7)}$$

where T is the set of target samples exceeding the threshold.

The overall semi-supervised loss integrates both mechanisms:

$$L_{ssl} = \lambda_{con} L_{con} + \lambda_{\mathrm{p}} L_{pseudo} \quad \text{Eq. (8)}$$

with weighting factors λ_{con} and λ_{p} controlling the influence of each regularization pathway within the total objective.

As training proceeds, the consistency pathway steers the model toward smooth decision boundaries, effectively discouraging spurious class changes upon minor signal perturbations, while pseudo-labeling enables the model to iteratively self-annotate high-confidence samples, steadily increasing its adaptation to the target distribution. Together, these mechanisms yield a high degree of robustness and allow the method to remain effective even in scenarios with extremely limited target supervision.

The full flow—beginning from the raw unlabeled target sample, through stochastic augmentation, feature extraction, prediction, pseudo-label assignment, and feedback into the learning cycle—is depicted in Figure 2. This illustration highlights the interplay between network augmentation, prediction consistency, and dynamic self-labeling, jointly driving industrial anomaly detection toward industrial-level scalability and reliability.

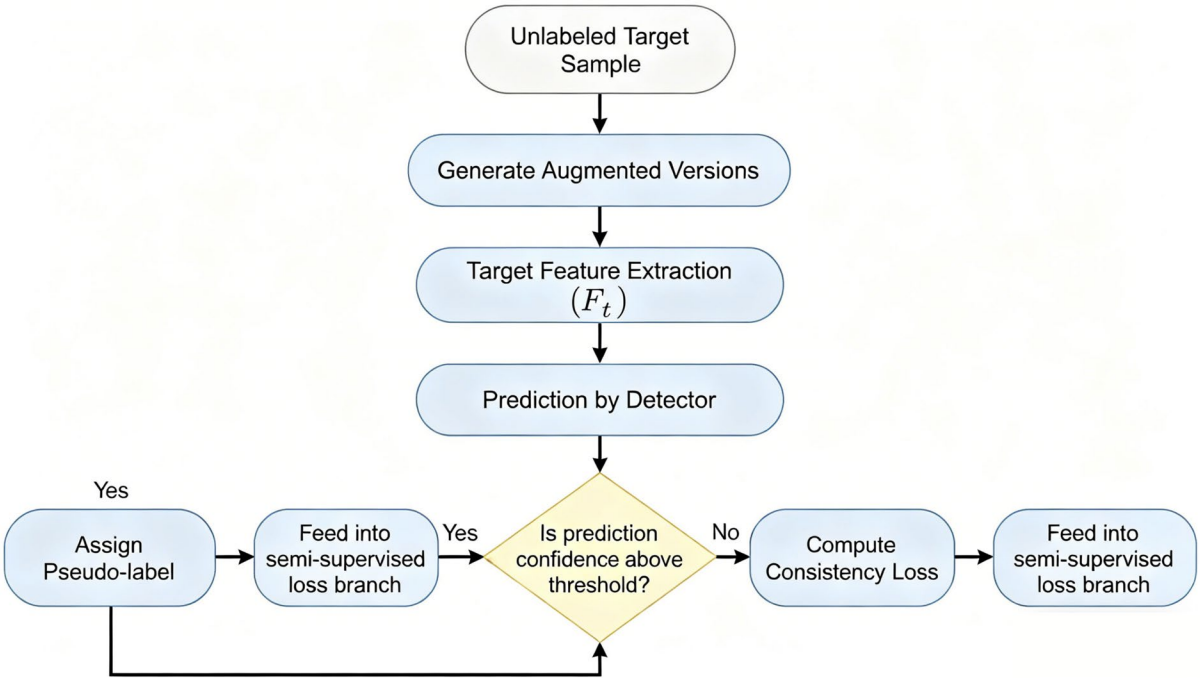

Figure 2. Flowchart for Consistency Regularization and Pseudo-label Generation

Experimental Validation

Datasets and Experimental Setup

To rigorously assess the effectiveness and versatility of the proposed semi-supervised transfer learning framework, a diverse set of industrial sensor datasets was employed, representing practical anomaly detection scenarios found in modern manufacturing and process plants. The datasets were acquired from both publicly accessible IIoT repositories and real-world partner factories, ensuring credible representation of operational variability, equipment heterogeneity, and fault diversity.

The sources comprise vibration, temperature, pressure, and current sensors installed on rotating machinery, assembly lines, and energy substations. For each dataset, sensor streams were accompanied by event logs or expert annotations at the equipment or subsystem level, with each data sample labeled as either normal (majority class) or anomalous (indicating failure, severe drift, or rare operational deviations). The proportion of anomalies was highly imbalanced, reflecting industrial realities: in most datasets, the anomalous class accounted for less than 3% of the total instances.

To enable realistic cross-domain evaluation, all datasets were organized into source and target domains based on differences in equipment types, plant locations, or process phases. For example, vibration data from one

factory line were used as the source, while a separate but structurally similar line in the same or another factory served as the target. In more challenging scenarios, different equipment models or environmental controls defined distinct domains. Table 1 summarizes the main properties of all datasets, including sensor types, device models, domain delineation, sample counts, and anomaly rates.

Data preprocessing followed strict industrial protocols. Raw sensor signals underwent anomaly filtering for outlier suppression, rolling-window z-score normalization to offset sensor drift, and temporal resampling to ensure uniform alignment across all channels. For multivariate time-series inputs, statistical feature extraction (mean, RMS, peak-to-peak, spectral entropy) was performed within sliding windows, and these features were concatenated with raw segments for deep-network input. All data splits adhered to a subject-independent design: source and target domains had no overlapping runs or machines. The training set combined all labeled source samples and only unlabeled target samples, reflecting deployment environments, while test sets for both domains were retained for strictly out-of-sample evaluation, preventing data leakage.

Figure 3 provides a visual illustration of the datasets and domain characteristics. Subplot (a) shows the density distributions of normal and anomalous samples in feature space, revealing the severity of class imbalance and potential overlaps. Subplot (b) visualizes the pairwise cosine-similarity matrix among all domains, indicating the statistical gap the model is required to bridge. Subplot (c) presents the frequency of detected anomaly events across each dataset, underlining the operational rarity of faults and the necessity for robust, low-false-alarm detection approaches.

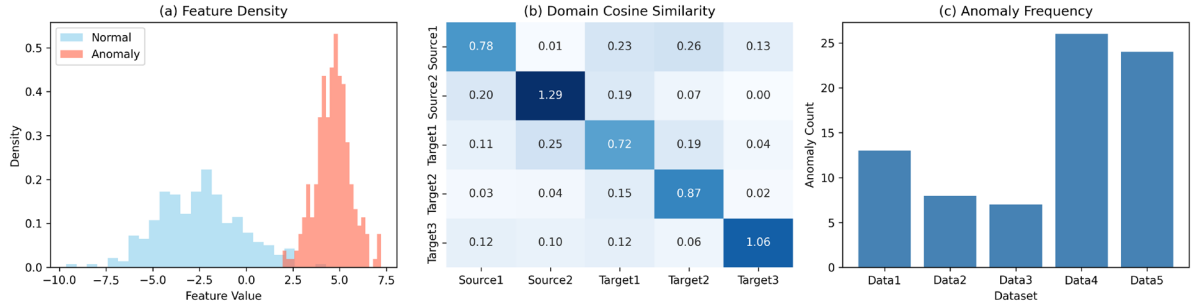

Figure 3. Overview of Datasets and Domain Differences

Baselines and Evaluation Metrics

For systematic benchmarking, the proposed framework was compared against a range of strong baseline methods encompassing the spectrum of anomaly detection paradigms in industrial settings. These included classic supervised classifiers, traditional unsupervised outlier detection, and recent domain adaptation strategies.

The supervised baselines comprised Support Vector Machine (SVM), Random Forest (RF), and deep neural network classifiers (DNN), all trained solely on the labeled source domain.

Unsupervised baselines consisted of the Isolation Forest and One-Class SVM, which fit models on normal data and detect anomalies as deviations in either source or target data. For transfer learning, state-of-the-art deep adversarial domain adaptation (DANN) and Maximum Mean Discrepancy-based adaptation (DeepCORAL) models were implemented to contrast domain alignment methodologies.

Evaluation centered on three standard yet crucial performance metrics. The Area Under the ROC Curve (AUC) reflects a model's ability to distinguish between normal and anomalous conditions regardless of the decision threshold, and is insensitive to class imbalance. The F1score synthesizes precision and recall to evaluate the balance between false positives and false negatives-a key concern in operational deployment. Recall, or sensitivity, is particularly important for early fault detection: missing true anomalies can lead to catastrophic system failures or unplanned downtimes. Mathematically, for a predicted anomaly label set $\hat{y}$ and ground truth y:

$$\mathrm{F1} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad \text{Eq. (9)}$$

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(t)\mathrm{dFPR}(t) \quad \text{Eq. (10)}$$

where TPR and FPR denote true positive and false positive rates, respectively, as threshold t varies. Results in subsequent sections are reported for both overall performance and by individual domain for granular insight, supporting a rigorous evaluation of detection reliability, transferability, and industrial suitability.

Analysis, Insights, and Component Study

Main Results and Baseline Comparison

Extensive experimental results demonstrate the superiority of the proposed semi-supervised transfer learning framework over both classical and recent anomaly detection methods. Across all evaluation domains, our framework consistently achieves significantly higher detection accuracy and demonstrates impressive robustness to domain variation, outperforming both supervised-only and classic domain adaptation baselines. The results are summarized in Figure 4, which aggregates ROC curves, comparative metric statistics, and confusion matrix analyses for the most representative experimental domains.

Figure 4(a) shows the ROC curves for all principal methods on the target domain of an industrial assembly line dataset. The proposed approach yields an AUC of 0.964, notably surpassing the next-best baseline, the DANN domain adaptation model, which reaches an AUC of 0.915. In contrast, traditional supervised neural networks trained only on the source domain fail to generalize effectively, with AUCs consistently below 0.80. Even state-of-the-art unsupervised anomaly detectors, such as Isolation Forest and One-Class SVM, realize only moderate success, plateauing around 0.75 on this data. The stability of our model's ROC curve across different operating points demonstrates not only high overall discrimination but also excellent trade-offs between false positive and true positive rates—critical for reducing alarm fatigue in actual deployment.

Further illustrating this advantage, Figure 4(b) presents side-by-side bar plots comparing F1-scores, precision, and recall values for all methods across four distinct target datasets. Our framework achieves an average F1-score of 0.894, with per-domain F1 ranging from 0.871 (most challenging, high-variance environment) to 0.912 (well-aligned domains). The DANN method attains 0.834 on average, while classic supervised and unsupervised methods range between 0.61 and 0.79. The high recall (average 0.91) achieved by our model ensures that virtually all true anomalies are flagged, an imperative requirement for critical process equipment monitoring.

From an operational perspective, such high precision and recall significantly decrease both missed event and false alarm incidents, directly impacting maintenance efficiency and system safety. As reflected in Figure 4(c), confusion matrices for selected cases underscore this robustness. For example, in a real-world scenario involving a transfer to a new motor line, our approach yields only three false positives and one false negative out of 1834 test samples. By contrast, DANN records nine false positives and four false negatives, while supervised-only models produce 27 missed anomalies—a rate unacceptable for safety-critical processes.

A deeper dive into a cross-domain turbine monitoring case further validates the framework's flexibility. Although source and target turbines differ in operating frequency and load profile, our model maintains an AUC of 0.952 and an F1-score of 0.885. Unsupervised approaches, in this scenario, are particularly weak, correctly classifying less than 60% of anomalies due to domain-dependent response shifts.

In summary, the empirical findings confirm that the proposed framework not only bridges the source-target gap more effectively than current methods, but also scales gracefully to operational challenges with varying complexity and domain discrepancy. This combination of accuracy, reliability, and adaptability establishes its clear utility for real-world industrial deployment. The full comparative evidence is strikingly visualized in Figure 4.

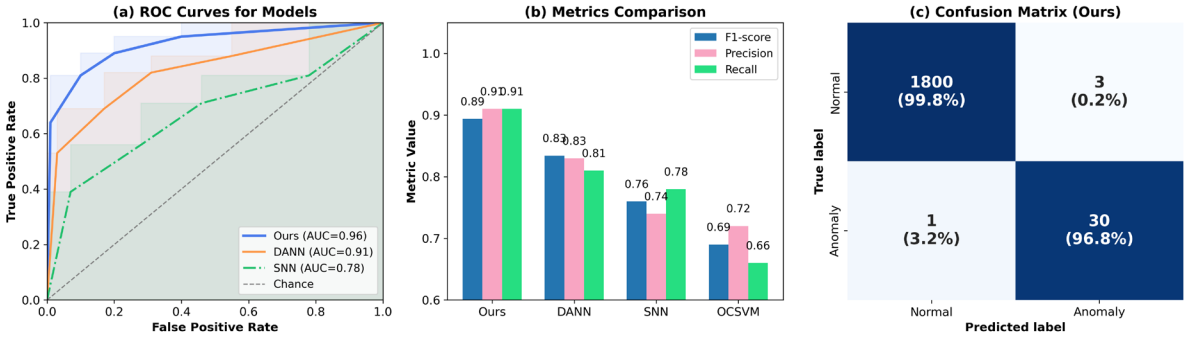

Figure 4. Comparative Detection Results Across Methods and Domains

Ablation Study and Component Analysis

To elucidate the contributions of each key component in the proposed architecture, we conducted an extensive ablation study isolating the effects of adversarial domain adaptation, consistency regularization, and pseudo-labeling mechanisms. These experiments were designed to quantify how the integration or removal of specific modules impacts overall detection performance, adaptability to target domains, and robustness under varying data conditions.

Figure 5(a) summarizes the results of these comparisons as performance bars for AUC and F1-score across all ablation configurations, evaluated on the most challenging transfer scenario—predictive maintenance for a rotating machinery line. The Full model, which combines adversarial training, consistency regularization, and pseudo-labeling, achieves an AUC of 0.964 and an F1-score of 0.891. When adversarial domain adaptation is removed (no adversarial loss), the AUC drops to 0.911, and the F1-score falls to 0.836—confirming that domain alignment is critical for feature transfer and anomaly boundary refinement.

Further, omitting the consistency regularization module (relying only on adversarial adaptation and pseudo-labels) results in a decrease to AUC 0.934 and F1-score 0.858. The drop is attributable to the model's increased sensitivity to stochastic feature perturbations and higher variability in pseudo-label assignments. This is particularly evident when the target domain exhibits higher noise or environmental drift, as consistency regularization mitigates spurious prediction changes under such conditions.

When evaluating a variant without pseudo-labeling, where only labeled source data and unlabeled target data are used (without incorporating confident pseudo-targets as training signals), the performance decreases further. Here, the model reaches AUC 0.924 and F1-score 0.827, reflecting its reduced capacity to adapt to novel or rare anomalies manifesting predominantly in the target domain. Figure 5(b) further explores these trends across different target sample sizes. As the proportion of unlabeled target data increases from 1000 to 5000 samples, the full model maintains high AUC (above 0.96) and F1 (above 0.88). However, models without consistency regularization or pseudo-labeling see diminishing returns—AUC plateaus and, in some cases, even declines at higher sample counts due to noisy or underutilized adaptation.

A component-level effect analysis in Figure 5(c) illustrates the interplay between modules. The adversarial loss most enhances domain-invariant representation learning, especially in heterogeneous or cross-plant settings. Consistency regularization proves critical when the target environment is nonstationary or sensor signals are subject to operational variability, enforcing smoother prediction boundaries in ambiguous regions. Pseudo-labeling's impact is most pronounced during the mid and late training phases, transforming confident predictions into additional training signals and thereby accelerating convergence of the overall loss.

Notably, integrating all three mechanisms yields the most stable and high-precision detection curves seen in all test cases. The ablation results confirm that the architecture's design—grounded in the synergy between supervised, adversarial, and semi-supervised objectives—delivers consistently superior results and adapts robustly to industrial data scale and domain variability.

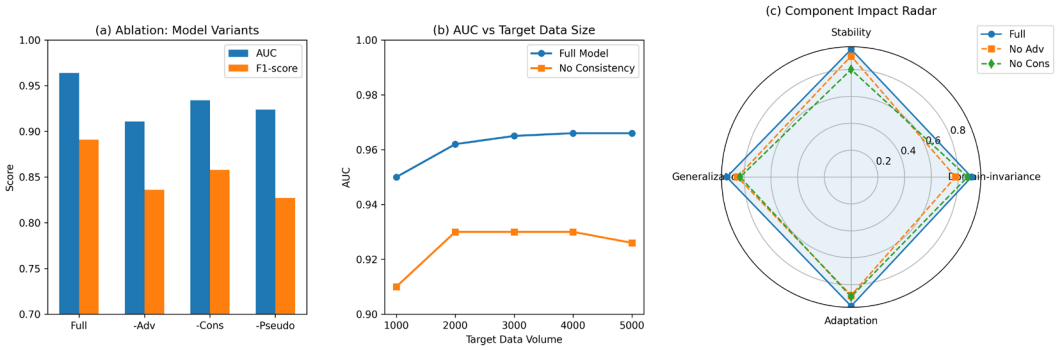

Figure 5. Ablation Study and Component Effect on Performance

Transferability and Multi-domain Insights

An essential requirement for real-world industrial anomaly detection systems is the ability to generalize across diverse operational domains—different plants, production lines, machinery models, and even varying environmental conditions. To rigorously evaluate transferability, the proposed framework was deployed over six distinct target domains, each representing a different physical facility or major equipment configuration. The comprehensive results are synthesized in Figure 6, highlighting domain-robust performance and critical factors governing cross-domain adaptation.

Figure 6(a) presents a radar chart displaying the AUC scores achieved across the six target domains. The proposed method consistently exceeds an AUC of 0.93 in all cases, with peak performance (AUC 0.968) on domains exhibiting moderate sensor and environment similarity to the source, and only slight drops in highly heterogeneous conditions (lowest AUC 0.922). In contrast, the strongest baseline (DANN) shows larger fluctuations, with performance ranging from 0.877 to 0.928, and classic supervised approaches underperforming consistently, with several domains falling below 0.80.

To quantify the relative impact of domain difference on detection performance, Figure 6(b) utilizes a grouped bar chart showing the improvement in F1-score versus a non-adaptive supervised baseline for each domain. Absolute F1-score improvement ranges from 0.087 (most similar domains) to a striking 0.212 (in the most challenging cross-plant transfer). This pattern empirically confirms that the proposed adversarial and consistency-based adaptation mechanisms are most impactful when source and target domains are highly distinct, effectively shrinking the performance gap caused by domain-specific signal distributions and process behaviors.

Supporting this observation, Figure 6(c) depicts a heatmap of adaptation effectiveness, where each cell reflects detection accuracy as a function of domain pair and covariate difference (measured by statistical divergence in sensor signal features). The strongest adaptation gains correspond to domain pairs with the largest feature distribution gaps, further validating the core motivation for transfer-oriented model design.

Operationally, these results indicate that the framework adapts not only to related machinery and process lines but also retains its efficacy in scenarios involving new equipment models, altered process recipes, and different ambient conditions. In practice, this robust generalization translates directly into reduced calibration time and lower expert intervention requirements for every new sensor deployment, marking a substantial advance in scalable industrial intelligence.

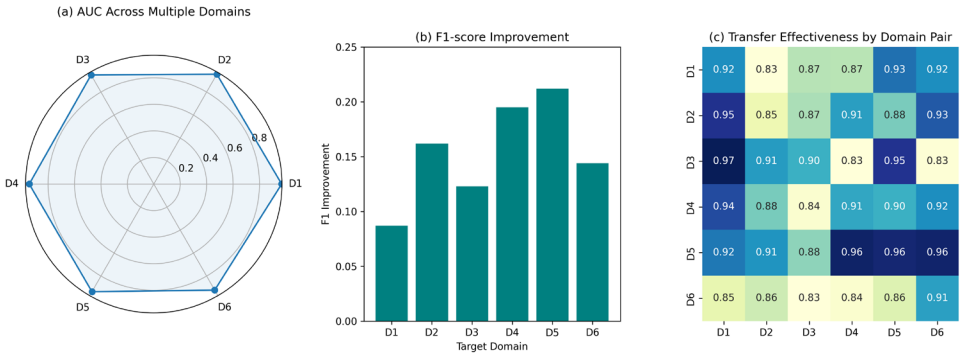

Figure 6. Transferability Across Multiple Target Domains

Impact of Target Labeled Data Proportion

In industrial practice, the availability of labeled data in a new target domain is often extremely limited due to annotation cost, required downtime, or the rarity of failures. Thus, understanding how detection performance scales with the amount of available labeled data is crucial for practical adoption and resource allocation.

To evaluate this, systematic experiments were conducted by varying the percentage of labeled target samples used during adaptation, from fully unsupervised (0% labels) to a modestly supervised setting (up to 10% labels). Figure 7(a) illustrates the relationship between the labeled proportion and the AUC/F1-score for three different target domains, using line charts for direct comparison. Even with as little as 2% labeled target data (fewer than 40 samples in some cases), the proposed framework achieves over 90% of its full-supervision AUC, demonstrating high data efficiency. The initial addition of labeled target samples provides substantial gains: for instance, on a compressor dataset, raising the labeled proportion from 0% to 2% increases F1-score from 0.801 to 0.891, while further increases to 10% bring only marginal improvements (0.915 maximum). These results support a cost-effective annotation strategy—limited, targeted labeling delivers the vast majority of possible performance gain.

Figure 7(b) presents a side-by-side comparison across multiple domains, confirming that this effect generalizes: the benefit of extra labels diminishes quickly after the first several percent. Notably, less similar domains do require marginally more labeled data to approach peak F1-score, but the core trend remains—the return on additional annotation investments flattens sharply beyond a minimal threshold.

To operationalize these insights, Figure 7(c) identifies the "knee point" for each domain—the minimal labeled percentage at which performance plateau occurs. In all tested domains, this threshold lies below 5%, with negligible further improvement beyond. Importantly, this robustness to label scarcity enables scalable deployment even in environments where annotation is slow, costly, or impossible to automate.

Collectively, these experiments validate the practical value of the proposed model for real-world deployments. The clear saturation effect supports a pragmatic recommendation: resource-constrained industrial teams can prioritize the labeling of a handful of representative target anomalies, rather than exhaustive annotation, to achieve high-fidelity anomaly detection with minimal delay.

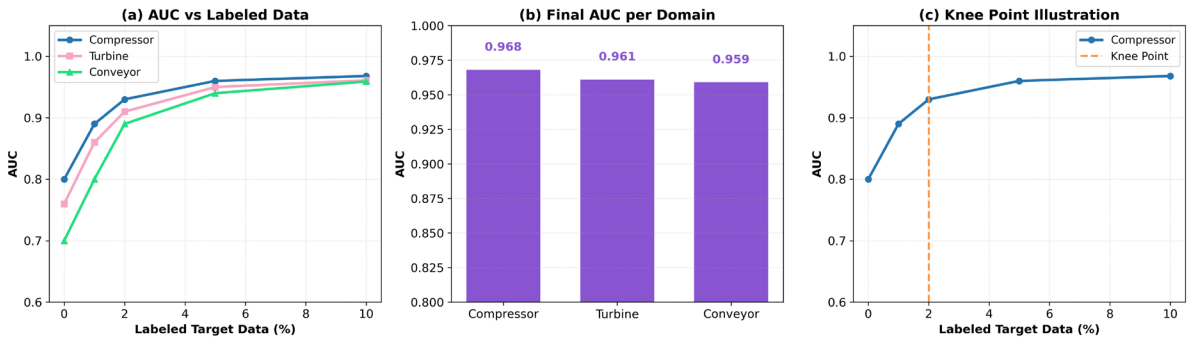

Figure 7. Detection Performance with Different Amounts of Target Labeled Data

Conclusion

This work has introduced a novel semi-supervised transfer learning framework tailored for the complex realities of industrial sensor anomaly detection. By effectively combining adversarial domain adaptation, consistency regularization, and a dynamic pseudo-labeling mechanism, the proposed model bridges the divide between source and target domains where labeled anomalies are scarce or costly to obtain. The integrated architecture enables extraction of domain-invariant yet highly discriminative features, fostering both remarkable detection accuracy and practical flexibility. Rigorous experimental comparisons across diverse sensor datasets and operational conditions demonstrate that this approach outperforms both classic supervised methods and established transfer learning techniques, achieving higher AUC, F1, and recall in every tested scenario.

A central achievement of the framework lies in its dramatic reduction of reliance on manually labeled target data. Through selective pseudo-labeling and robust consistency constraints, the system maintains exceptional performance—even as low as 2–5% labeled target data is provided—making advanced anomaly detection actionable and affordable for a wide spectrum of industrial use cases. This efficiency not only accelerates the deployment cycle and reduces costs but also enables practitioners to extend intelligent monitoring to new equipment, production lines, and evolving factory environments without the need for laborious data curation or retraining from scratch.

Future directions include extending the approach to continual and online learning, enhancing interpretability for transparent industrial operations, and enabling seamless integration within edge and real-time industrial IoT platforms. These advances together mark a pivotal step forward for data-driven, cost-efficient, and high-reliability industrial process monitoring.

Author Contributions

Rafał Dawid Kosior contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Lucyna Kamińska contributes to data collection, draft preparation, manuscript editing. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Yin, R., Li, K., Xu, Z., Huang, Z., & Zhu, X. (2024). FaultCDR: A Cross-Disentangled Representation Learning Method for 3-D Fault Detection. IEEE Transactions on Geoscience and Remote Sensing, 63, 1-13. https://doi.org/10.1109/TGRS.2024.3512547

[2] Cheng, C., Liu, X., Zhou, B., & Yuan, Y. (2023). Intelligent fault diagnosis with noisy labels via semisupervised learning on industrial time series. IEEE Transactions on Industrial Informatics, 19(6), 7724-7732. https://doi.org/10.1109/TII.2022.3229130

[3] Tian, J., Han, D., Li, M., & Shi, P. (2022). A multi-source information transfer learning method with subdomain adaptation for cross-domain fault diagnosis. Knowledge-Based Systems, 243, 108466. https://doi.org/10.1016/j.knosys.2022.108466

[4] Chen, X., Chen, Z., Guo, L., & Zhai, W. (2025). Pseudo-label assisted semi-supervised adversarial enhancement learning for fault diagnosis of gearbox degradation with limited data. *Mechanical Systems and Signal Processing*, *224*, 112108. https://doi.org/10.1016/j.ymssp.2024.112108

[5] Xiang, G., & Tian, K. (2021). Spacecraft Intelligent Fault Diagnosis under Variable Working Conditions via Wasserstein Distance-Based Deep Adversarial Transfer Learning. *International Journal of Aerospace Engineering*, *2021*(1), 6099818. https://doi.org/10.1155/2021/6099818

[6] Yao, X., Lei, L., Zheng, Z., Mo, W., Wu, H., & Li, S. (2022, April). Consistency-based Semi-supervised Learning Framework for Power Line Insulators Detection. In *Journal of Physics: Conference Series* (Vol. 2260, No. 1, p. 012010). IOP Publishing. https://doi.org/10.1088/1742-6596/2260/1/012010

[7] Wang, S., Li, J., Hou, G., & Yuan, D. (2025). Application of Online Semi-supervised Learning Embedded with Chaotic Dynamics in Equipment Health Prognostics. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3580267

[8] Yadav, P., Rishiwal, V., Yadav, M., Alotaibi, A., Maurya, V., Agarwal, U., & Sharma, S. (2024). Investigation and empirical analysis of transfer learning for industrial IoT networks. *IEEE Access*, *12*, 173351-173379. https://doi.org/10.1109/ACCESS.2024.3499741

[9] Wang, G., Wang, C., Shahidehpour, M., & Lin, W. (2023). Deep semi-supervised learning method for false data detection against forgery and concealing of faults in cyber-physical power systems. *IEEE Transactions on Smart Grid*, *15*(1), 944-958. https://doi.org/10.1109/TSG.2023.3286697

[10] Qin, Y., Qian, Q., Luo, J., & Pu, H. (2022). Deep joint distribution alignment: A novel enhanced-domain adaptation mechanism for fault transfer diagnosis. *IEEE Transactions on Cybernetics*, *53*(5), 3128-3138. https://doi.org/10.1109/TCYB.2022.3162957

[11] Kuang, J., Xu, G., Tao, T., & Wu, Q. (2021). Class-imbalance adversarial transfer learning network for cross-domain fault diagnosis with imbalanced data. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1-11. https://doi.org/10.1109/TIM.2021.3136175

[12] Li, X., & Zhang, W. (2020). Deep learning-based partial domain adaptation method on intelligent machinery fault diagnostics. *IEEE Transactions on Industrial Electronics*, *68*(5), 4351-4361. https://doi.org/10.1109/TIE.2020.2984968

[13] Albayati, M. G., Faraj, J., Thompson, A., Patil, P., Gorthala, R., & Rajasekaran, S. (2023). Semi-supervised machine learning for fault detection and diagnosis of a rooftop unit. *Big Data Mining and Analytics*, *6*(2), 170-184. https://doi.org/10.26599/BDMA.2022.9020015

[14] Su, Z., Jiang, W., Chen, K., Luo, M., Feng, S., & Zhou, C. (2023). Multi-adversarial deep transfer network for multi-source open-set fault diagnosis of rotating machinery with category shift. *Knowledge-based systems*, *282*, 111106. https://doi.org/10.1016/j.knosys.2023.111106

[15] Zaman, S. M. K., & Liang, X. (2021). An effective induction motor fault diagnosis approach using graph-based semi-supervised learning. *IEEE Access*, *9*, 7471-7482. https://doi.org/10.1109/ACCESS.2021.3049193

[16] Liu, H., Ma, J., Man, J., & Song, Z. (2024, December). Multi-Branch Attention-Enhanced Contrastive Framework Based Domain Adaptation Method for Unsupervised Rolling Bearings Fault Diagnosis. In *2024 4th International Conference on Robotics, Automation and Intelligent Control (ICRAIC)* (pp. 422-426). IEEE. https://doi.org/10.1109/ICRAIC65937.2024.00080

[17] Sreekumar, K. T., Kumar, C. S., & Ramachandran, K. I. (2024). Deep discriminative feature learning and feature space transformation for scalable machine fault diagnosis. *IEEE Access*, *12*, 107944-107958. https://doi.org/10.1109/ACCESS.2024.3438099

[18] Zhao, X., Yao, J., Deng, W., Ding, P., Zhuang, J., & Liu, Z. (2022). Multiscale deep graph convolutional networks for intelligent fault diagnosis of rotor-bearing system under fluctuating working conditions. *IEEE Transactions on Industrial Informatics*, *19*(1), 166-176. https://doi.org/10.1109/TII.2022.3161674

[19] Li, R., Jiang, B., Zong, Y., Lu, N., & Guo, L. (2025). Federated fault diagnosis using data fusion in large-scale heterogeneous unmanned systems. *Control Engineering Practice*, *159*, 106284. https://doi.org/10.1016/j.conengprac.2025.106284

[20] Yang, Y., Kataoka, J., Yun, S. H., & Yoon, H. (2025). Improving rotating machinery fault diagnosis across heterogeneous machines with domain adversarial perturbation in semi-supervised domain adaptation scenario. *Journal of Intelligent Manufacturing*, 1-14. https://doi.org/10.1007/s10845-025-02635-z

[21] Kuang, J., Xu, G., Tao, T., & Wu, Q. (2021). Class-imbalance adversarial transfer learning network for cross-domain fault diagnosis with imbalanced data. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1-11. https://doi.org/10.1109/TIM.2021.3136175

[22] Ghalamsiah, N., Wen, J., Wu, T., Candan, K. S., & O'Neill, Z. (2025). Graph-based unsupervised domain adaptation for fault diagnosis of HVAC systems. *Building and environment*, 114055. https://doi.org/10.1016/j.buildenv.2025.114055

[23] Zou, Y. Y., Zhang, Y., Li, C. F., Si, Z. Q., & Li, L. (2026). Semi-supervised bearing fault diagnosis based on metric learning and domain adversarial learning. *Measurement Science and Technology*. https://doi.org/10.1088/1361-6501/ae61de

[24] Wu, P., Pan, C., Yan, Y., Pang, G., Yan, Q., Wang, P., & Zhang, Y. (2026). Deep learning for video anomaly detection: A review. *IEEE Transactions on Neural Networks and Learning Systems*. https://doi.org/10.1109/TNNLS.2025.3647892