KADF: Knowledge-Guided Aligned Multi-Source Data Fusion for ICS Anomaly Detection

Yanqun Wu[1,2,a], Yongtao Zhang[2,b], Mingxing Liu[1,2,c]
Quan Ma[2,d], Fan Wen[2,e], Hengxu Shen[2,f],
[1]College of Computer Science, Sichuan University, Chengdu, China
[2]National Key Laboratory of Nuclear Reactor Technology, Nuclear Power Institute of China, Chengdu, China
[a]m15720123548@163.com, [b]475590320@qq.com, [c]lmxstar@163.com,
[d]maq_np@163.com, [e]starsoul_w@qq.com, [f]1804879169@qq.com

***Abstract*—Anomaly detection in industrial control systems (ICS)—a critical application safeguarding mission-critical infrastructure—requires the integration of multi-source signals such as network traffic, security logs, and physical process data. However, existing methods typically rely on single modality models or simplistic fusion approaches, making models vulnerable to noise or distributional biases from any single data source. This work re-examines the "alignment" step prior to fusion and proposes a knowledge-guided aligned multi-source data fusion method for ICS anomaly detection (KADF). KADF decouples the conventional end-to-end black-box paradigm into an "align first, then fuse" process: first, we convert formalizable asset information into learnable embedding vectors; subsequently, during the fusion stage, attention bias informed by security priors is injected to compel the model to focus on high-risk cross-domain correlations. We devise alignment loss and security prior knowledge loss to facilitate embed security priors. Experiments across public industrial control datasets show that KADF outperforms existing single-modality models and recent benchmarks, achieving mean AUC and F1-score values of 96.7% and 96.6%. This confirms the effectiveness of the knowledge-guided alignment paradigm for network attack detection models.**

***Index Terms*—Knowledge-guided alignment, multi source data fusion, attention mechanism bias, industrial control system**

I. Introduction

The rapid advancement of the industrial internet has shifted Industrial Control Systems (ICS), a foundational class of critical infrastructure, from isolated to interconnected environments, exposing them to increasingly severe and complex cyber threats [1]. In this evolving landscape, anomaly detection relying solely on a single data source, such as network traffic, is no longer sufficient. To achieve holistic security awareness, integrating multi-sources information—including network traffic, security logs, and physical process data—thus has therefore emerged as an essential direction [2] [3].

Network traffic data reflects anomalies in communication behaviors, security log data documents sequences of system security events, and physical process data indicates the operational status of controlled objects. Theoretically, the complementarity of such multi-source information can effectively reveal threats that are subtle in individual dimensions but conspicuous in cross-domain correlations [4]. However, the ideal fusion is confronted with a core practical challenge: the heterogeneity of multi-source data in industrial control scenarios. This heterogeneity is manifested in three aspects: asynchrony (varying sampling rates), modal heterogeneity (diverse data modalities), and distribution divergence (distinct data distributions) [5], which results in a significant semantic gap between multi-source data.

Existing methods exhibit two critical limitations when addressing such heterogeneous data: 1) Neglect of feature alignment leads to semantic mismatch in direct fusion. Most existing studies adopt either "feature concatenation" [6] or "late-stage decision-making" [7], [8] fusion strategies, performing brute-force fitting on multi-source data without proper handling of data heterogeneity. This further causes the fusion model to be susceptible to domination by noise or distributional characteristics from a single data source, making it difficult to learn consistent cross-domain representations of anomalies. 2) Lack of domain knowledge guidance results in an undirected fusion process. ICS are endowed with abundant formalizable domain knowledge (e.g., asset topology and physical constraints). However, existing methods generally overlook the guiding role of such knowledge and fail to incorporate prior information into the fusion process. Consequently, the model struggles to focus on cross-domain correlations with genuine security implications, and its performance degrades significantly especially when confronted with attack variants.

To address the issues, inspired by the applications of Domain Adaptation and Transfer Learning in other fields [9], [10], we proposes the Knowledge-Guided Aligned Multi-Source Data Fusion for ICS Anomaly Detection Framework(KADF). The core idea of KADF is to decompose the traditional "end-to-end black-box fusion" paradigm into a clear workflow of "knowledge injection—feature alignment—guided fusion", which fundamentally resolves the problems of semantic mismatch and directional ambiguity in heterogeneous data fusion. Specifically, by introducing prior knowledge such as asset topology, KADF guides multi-source data to achieve semantic space alignment prior to fusion; subsequently, through a knowledge-guided fusion mechanism, it enhances the model's capability to perceive critical security correlations and interpretability.

The main contributions are summarized as follows:

1. We propose the KADF framework, which innovatively establishes knowledge-guided feature alignment as a core preprocessing stage for multi-source data fusion detection

in ICS, thereby providing a novel paradigm for addressing heterogeneous data.

2. We design an asset-aware projection alignment module, which creatively converts asset identifiers into learnable embedding vectors and devises a alignment loss to achieve latent-space semantic normalization for asynchronous and heterogeneous data.

3. We propose a Gated Query Fusion Module to enable fine-grained multi-view fusion. Built upon this, our two-stage architecture first aggregates intra-window multi-view features, then fuses them with security prior knowledge via attention-based alignment.

4. Experiments on multiple public datasets demonstrate that KADF significantly outperforms mainstream baseline methods in terms of detection rate and generalization performance, while also achieving superior interpretability.

II. Related Work

A. Single-source Anomaly Detection

Early research on ICS security has primarily focused on single data sources. Network traffic-based detection identifies anomalous commands or communication patterns by parsing industrial control protocols (e.g., Modbus/TCP, DNP3) [11]. Host or device log-based detection leverages log parsing and sequence analysis techniques to detect anomalous event flows [12]. Physical process-based detection, by contrast, focuses on the sequence analysis of sensor and control signals, utilizing state machines or machine learning methods to identify anomalies that violate physical laws [13]. Although these methods are effective within specific dimensions, they form isolated security silos that fail to detect and correlate coordinated attacks across cyber-physical boundaries.

B. Multi-source Data Fusion Detection

To break through the limitations of single-source detection, multi-source data fusion detection has become a research hotspot in recent years. Based on the fusion level, existing methods can be categorized into two types:

Feature-level fusion: This type of method performs early concatenation on feature vectors from different sources and then feeds them into a classifier (e.g., Support Vector Machines, deep neural networks). For instance, [6] directly concatenates statistical features of network traffic and count features of system logs for model training. Despite its simplicity and ease of implementation, this method ignores the distribution differences and semantic misalignment between features. The high-dimensional feature space generated by concatenation is filled with noise, leading to limited improvements in model performance.

Model-level fusion: This type of method adopts more sophisticated deep learning architectures, such as multi-modal neural networks, which design dedicated subnetworks for feature extraction from different data sources before conducting late-stage fusion [7], [8]. Although these methods exhibit strong representational capabilities, their fusion process is blind. They rely on data-driven approaches to implicitly learn cross-domain correlations, yet lack explicit mechanisms to ensure that features from different sources are mapped to a relatively unified latent space. As a result, the models are prone to overfitting to source-specific noise. Meanwhile, the black-box fusion process suffers from poor interpretability.

C. Alignment and Adaptation Methods for Data Heterogeneity

The issue of data heterogeneity has been extensively investigated in other domains of machine learning. Domain adaptation and transfer learning aim to reduce the distribution divergence between the source domain and the target domain, typically implemented via Maximum Mean Discrepancy (MMD) or adversarial training [14]. These ideas provide inspiration for addressing multi-source data distribution shifts in ICS analytics. More enlightening work stems from bioinformatics. To eliminate the batch effect in single-cell sequencing, researchers have developed methods such as Harmony [9], Seurat v3 [15], and the Regularized Centroid Mapping proposed by Chen et al. [10]. Their core idea is to calibrate data from different batches into a common latent space via linear or nonlinear mapping, enabling the clustering of cells with the same type while eliminating batch-induced variations.

III. Methodology

A. Overview Framework

Task Formulation. Given a multi-source asynchronous raw sequence $\mathcal{S} = \{x^{(v)}\}_{v\in\mathcal{V}}$, where $x^{(v)} = \{(t_k^{(v)}, x_k^{(v)})\}_{k=1}^{K_v}$ denotes the subsequence from view $v \in \mathcal{V}$ with $\mathcal{V} = \{\text{net}, \text{dev}, \text{log}, \dots\}$. Each subsequence possesses independent timestamps $t_k^{(v)}$, variable sampling rates, and distinct feature dimensions d_v. The entire sequence is associated with a global anomaly label $y \in \{0, 1\}$. The objective is to learn a model that outputs anomalous discrimination for $\mathcal{S}$.

Figure 1 illustrates the proposed KADF framework.

B. Security Prior Knowledge Construction

Asset Knowledge: To unify cross-view asset representations, we construct an Extensible asset Knowledge Base.

$$\mathcal{A} = \{\text{ID}_i\}_{i=1}^{|\mathcal{A}|}, \qquad \text{ID}_i \in \mathbb{Z}^+. \tag{1}$$

The asset ID_i is generated by joint hashing of discrete fields (IP, port, etc.). This ensures consistent identifiers across views for the same asset.

Security Prior Knowledge: The Security Prior Knowledge Base comprises textual descriptions of attack manifestations across multiple views, encoded via a domain-fine-tuned BERT, denoted as $h_p^{(v)} \in \mathbb{R}^{d_k}$. Example prior knowledge entries are shown in Table I.

C. Asset-aware projection alignment module

For any view v, raw features $x^{(v)} \in \mathbb{R}^{d_v}$ are mapped to asset representations via direct embedding lookup.

$$a^{(v)} = \mathbf{E}\big[\text{ID}(x^{(v)})\big] \in \mathbb{R}^{d_a}, \tag{2}$$

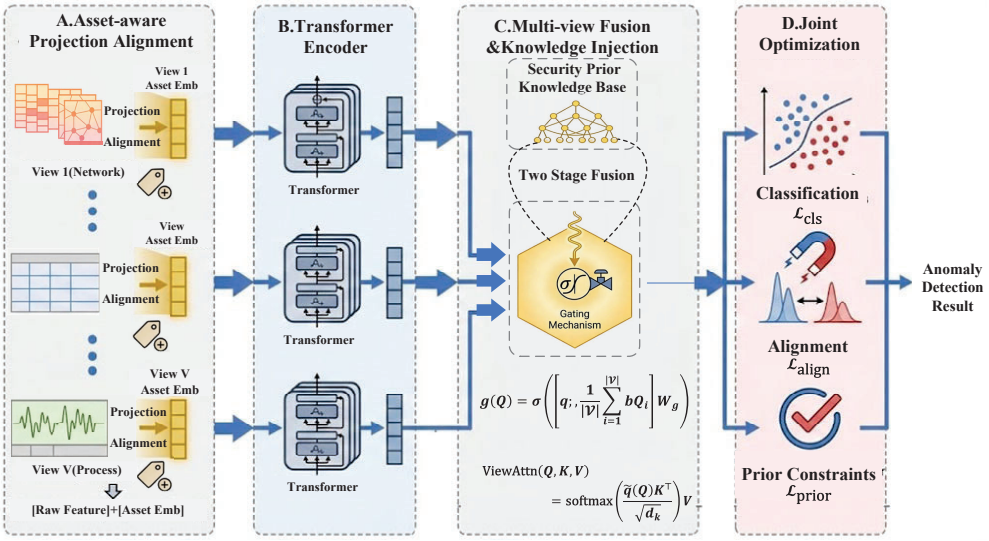


Fig. 1. Framework of KADF. **A. Asset-aware Projection Alignment:** Asset identifiers retrieved from the knowledge base are projected into embeddings via learnable modules and concatenated to raw view features. An asset-aware alignment loss optimizes these projections to enforce consistent cross-view mappings. **B. View-Specific Transformer Encoder:** Dedicated Transformers independently extract embedded representations from each individual view. **C. Multi-view Fusion and Security Prior Knowledge Injection.:** Cross-view discrepancies are minimized while security prior knowledge is injected to enhance representation alignment. **D. Joint Optimization:** The entire architecture is trained end-to-end via a composite loss function $\mathcal{L}$.

TABLE I
PRIOR KNOWLEDGE TABLE

Attack type	Net view	Log view	Physics view
Port scanning	Network traffic has rapid sequential single-source connection attempts to multiple ports.	Device logs and firewall logs remain normal.	Parameters are within normal ranges, no anomalous control signals observed.
Brute force login	A sharp surge in network traffic.	Authentication logs show numerous repeated login failures.	Device operating parameters remain within the normal range.
Man-in-the-middle attack	ARP spoofing occurs in network traffic.	IDS triggers ARP alerts, Modbus protocol violation alerts and logs.	Device operating parameters are abnormal.
.....	.....	.....	.....

$\mathbf{E} \in \mathbb{R}^{|\mathcal{A}| \times d_a}$ denotes a trainable embedding matrix, d_a is the asset embedding dimension. It is concatenated with raw features. This forms the final input for each view.

$$\tilde{x}^{(v)} = \left[x^{(v)}\,;\, a^{(v)}\right] \in \mathbb{R}^{d_v + d_a}, \tag{3}$$

$[;]$ denotes vector concatenation.

D. Transformer Encoder

Each view employs a Transformer encoder with identical architecture, and outputs $h_{\text{tf}}^{(v)}$

$$h_{\text{tf}}^{(v)} = \text{Transformer}_v(\tilde{x}^{(v)}) \in \mathbb{R}^{d_k}, \tag{4}$$

d_k denotes the embedding dimension.

E. Multi-view information fusion and security prior knowledge injection.

To effectively fuse multi-view semantic information, we propose a Gated Query Fusion Module. Based on this module, we design a two-stage fusion architecture. Stage one fuses features from all views within the current time window. Stage two matches and fuses embedding with security prior knowledge.

Gated Query Fusion Module. We design a gating mechanism for dynamic fusion of learnable query $\mathbf{q}$ and view queries $\mathbf{Q}$. Compute the mean representation of view queries and concatenate it with $\mathbf{q}$.

$$\mathbf{g}(\mathbf{Q}) = \sigma\left(\left[\mathbf{q};\, \frac{1}{|\mathcal{V}|}\sum_{i=1}^{|\mathcal{V}|}\mathbf{Q}_i\right]\mathbf{W}_g\right) \in \mathbb{R}^{d_k}, \tag{5}$$

σ denotes the Sigmoid function. $\mathbf{W}_g \in \mathbb{R}^{2d_k \times d_k}$ is a learnable weight matrix. $|\mathcal{V}|$ represents the number of views. The gated fused query $\tilde{\mathbf{q}}(Q)$ is computed as:

$$\tilde{\mathbf{q}}(Q) = \mathbf{g}(\mathbf{Q}) \odot \mathbf{q} + (1 - \mathbf{g}(\mathbf{Q})) \odot \frac{1}{|\mathcal{V}|}\sum_{i=1}^{|\mathcal{V}|}\mathbf{Q}_i \tag{6}$$

Multi-view features are fused via cross attention:

$$\text{ViewAttn}(Q, K, V) = \text{softmax}\left(\frac{\tilde{\mathbf{q}}(Q)\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V} \tag{7}$$

Stage One: Intra-view Fusion. Given the Transformer encoder outputs $h_{\text{tf}}^{(v)}$ for $|\mathcal{V}|$ views in the current window. Then stack them vertically into $\mathbf{h}_{\text{view}} \in \mathbb{R}^{|\mathcal{V}| \times d_k}$. Alignment and fusion is performed via Equation 7.

$$\mathbf{h}_{\text{v}}^{'} = \text{ViewAttn}\left(h_{view}W_q, h_{view}W_k, h_{view}W_v\right) \in \mathbb{R}^{d_k}, \tag{8}$$

$\mathbf{W}_{q,k,v} \in \mathbb{R}^{d_k \times d_k}$ are projection matrices.

Stage Two: Security Prior Knowledge Prior Matching Fusion. Given records from the Security Prior Knowledge Base,

each with multi-view structure. Also stack their view features. Apply the view fusion module with parameters shared with Equation (8):

$$\mathbf{p}_j = \text{ViewAttn}\left(h_p W_q, h_p W_k, h_p W_v\right) \in \mathbb{R}^{d_k}. \tag{9}$$

We query with current window features. Retrieve relevant security prior knowledge. Perform weighted fusion:

$$\mathbf{h}_{\text{p}}^{'} = \text{softmax}\left(\frac{\mathbf{h}_{\text{v}}\mathbf{P}^{\top}}{\sqrt{d_k}}\right)\mathbf{P} \in \mathbb{R}^{d_{\text{k}}}, \tag{10}$$

We concatenate intra-view features with security prior knowledge features. The final representation $\mathbf{h}_{\text{final}}$ is:

$$\mathbf{h}_{\text{final}} = \left[\mathbf{h}_{\text{v}}^{'};\, \mathbf{h}_{\text{p}}^{'}\right]\mathbf{W}_{\text{out}}, \tag{11}$$

$\mathbf{W}_{\text{out}} \in \mathbb{R}^{2d_{\text{k}} \times d_{\text{out}}}$ denotes the output projection matrix.

F. Loss Function

The total loss comprises three terms.

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda_1 \mathcal{L}_{\text{align}} + \lambda_2 \mathcal{L}_{\text{prior}}. \tag{12}$$

$\mathcal{L}_{\text{cls}}$ denotes the classification loss. $\mathcal{L}_{\text{align}}$ denotes the alignment loss. $\mathcal{L}_{\text{prior}}$ denotes the security prior knowledge loss. $\lambda_1, \lambda_2 \in \mathbb{R}^{+}$ are trade-off hyperparameters.

1) Classification loss:

$$p = \sigma(\mathbf{h}_{\text{final}} W_{cls}), \quad \mathcal{L}_{\text{cls}} = \text{BCE}(p, y), \tag{13}$$

σ denotes the activation function. BCE denotes binary cross-entropy. $\mathbf{W}_{\text{cls}} \in \mathbb{R}^{d_k \times 2}$ is a learnable weight matrix.

2) Alignment loss: In ICS anomaly detection, cross-view alignment focuses on enforcing instance-level semantic consistency for the same asset rather than matching global feature distributions. Distribution-based methods such as MMD are sensitive to sample imbalance and non-stationarity, which are common in ICS data. An L2-based alignment loss provides a simple and stable constraint that reduces cross-modal semantic drift while preserving modality-specific discriminative features.

$$\mathcal{L}_{\text{align}} = \frac{2}{|\mathcal{V}|(|\mathcal{V}|-1)} \sum_{1 \le i < j \le |\mathcal{V}|} \|h_{\text{tf}}^{(i)} - h_{\text{tf}}^{(j)}\|^2, \tag{14}$$

$h_{\text{tf}}^{(i)}$ denotes the embedding vector of the i-th view.

3) Security Prior Knowledge loss: $\mathcal{L}_{\text{prior}}$ aggregates three practical constraints:

$$\mathcal{L}_{\text{prior}} = \mathcal{L}_{phys} + \mathcal{L}_{asset} + \mathcal{L}_{ord} \tag{15}$$

Physical consistency constraint $\mathcal{L}_{\text{phys}}$: Leverages real-world relationships between physical variables θ and q to compute the loss.

$$\hat{q} = f_\theta(z), \quad \mathcal{L}_{\text{phys}} = \frac{1}{N}\sum_{i=1}^{N}\left\|\hat{q}^{(i)} - q^{(i)}\right\|^2 \tag{16}$$

Asset access constraint $\mathcal{L}_{\text{asset}}$: During preprocessing, a violation label $v_a \in \{0, 1\}$ is computed for each sample. p_{viol} is an auxiliary classification head.

$$p_{\text{asset}} = \sigma\big(\mathbf{h}_{\text{final}}^{'} W_{asset}\big), \quad \mathcal{L}_{\text{asset}} = \text{BCE}(p_{\text{asset}}, v_a). \tag{17}$$

Event order constraint $\mathcal{L}_{\text{order}}$: During preprocessing, an order label $v_o \in \{0, 1\}$ is computed for critical event sequences. p_{ord} is an auxiliary binary classification head.

$$p_{\text{ord}} = \sigma\big(\mathbf{h}_{\text{final}} W_{ord}\big), \quad \mathcal{L}_{\text{ord}} = \text{BCE}(p_{\text{ord}}, v_o). \tag{18}$$

IV. Experiments

A. Datasets

ICS-Flow [19] simultaneously captures network flows, process variables, and attack logs. Network traffic has 45k bidirectional flows, each characterized by 50 features. PLC tank levels and valve states exhibit the physical impact of adversarial activities, while automatic IT/NST labels designate reconnaissance, distributed denial-of-service (DDoS), man-in-the-middle (MITM), and replay events. We constructed the dataset by sliding a 500ms time window, yielding a total of 16,016 samples.

ICS-ADD [20] condenses a comprehensive attack scenario into a seven-minute window: a span-port PCAP trace, OpenPLC register logs, and IDS/firewall syslogs are collected in parallel, while DNS tunneling, port scanning, Modbus scanning, ARP spoofing, false-data injection, and slow-body DoS are executed in a stepwise manner. We constructed the dataset by sliding a 100ms time window, yielding a total of 2,208 samples.

TON_IoT [21] focuses on edge-fog-cloud IoT/IIoT security: Security-Onion captures complete packet data, and Zeek extracts 44 flow features; Windows/Linux audit trails and MQTT telemetry data are logged concurrently, yielding 22 million records across nine attack families. We constructed the dataset by sliding a 500ms time window, yielding a total of 12,018 samples.

The statistical information of the dataset is summarized in Table III. For all the datasets, we consistently split them into training, validation, and test sets in a 7:2:1 ratio.

B. Baselines

Feature-level Fusion. These methods, e.g., the traditional MLP [16], LSTM [17] and Transformer [18], concatenate multi-source features and feed them into a classifier without accounting for modal asynchrony or semantic mismatch.

Model-level Fusion. These methods, e.g., MMCD [8] , employ separate sub-networks to encode each modality and perform late fusion, yet remain agnostic to ICS-specific knowledge.

Cross-domain Alignment (DAN). These methods, e.g., ADDA [14], align source and target domains by minimizing distribution divergence, but without incorporating industrial-control priors, thus failing to highlight security-critical cross-modal correlations.

C. Experiments Settings

Implementation Details. All experiments are conducted on a server with NVIDIA A10 GPUs. We implement our method KADF in Python 3.7 with PyTorch 1.11.0 and CUDA 11.3. For KADF, we adopt the following configurations: Transformer encoder with 4 attention heads and 4 layers; embedding dimension $d_k = 128$; trade-off hyperparameters $\lambda_1 \in \{0.5, 1, 2\}$

TABLE II
ANOMALY DETECTION RESULTS FOR DIFFERENT BASELINE METHODS (%)

	ICS-flow				ICS-ADD				TON_IoT			
Methods	F1	ACC	Recall	AUC	F1	ACC	Recall	AUC	F1	ACC	Recall	AUC
MLP [16]	92.28	96.87	85.91	92.92	65.05	83.71	48.20	74.10	74.10	76.58	86.58	84.45
LSTM [17]	93.09	97.19	87.07	93.53	91.87	94.80	93.53	94.45	93.09	94.19	94.07	93.53
Transformer [18]	93.44	97.32	87.93	93.93	94.40	96.38	97.12	**96.58**	92.44	96.32	95.23	93.93
MMCD [8]	94.15	97.68	88.35	94.65	94.58	97.05	95.42	94.31	94.22	95.61	95.75	94.72
ADDA [14]	94.62	97.95	88.92	95.18	94.75	98.18	94.96	94.87	93.73	97.89	95.98	95.24
ours	**97.18**	**98.52**	**90.85**	**97.47**	**96.35**	**98.63**	**97.68**	96.12	**96.45**	**98.44**	**96.15**	**96.58**

TABLE III
DATASET STATISTICS

	Window Size(ms)	Total Samples	Anomaly Ratio
ICS-Flow	500	16,016	22%
ICS-ADD	100	2,208	32%
TON_IOT	500	12,018	26%

and $\lambda_2 \in \{0.05, 0.1, 0.2\}$; window sizes of 500ms for ICS-Flow, 100ms for another dataset, and 500ms for a third dataset; learning rate $2e^{-3}$. For fair comparison, all baseline methods use 128-dimensional hidden layers.

D. Main Results

Table II presents the anomaly detection results of five baseline methods. Among them, three typical "feature-level fusion" baselines—MLP, LSTM, and Transformer—directly concatenate network traffic, logs, and physical signals as input to a unified model, without incorporating any industrial control domain knowledge or addressing semantic mismatches between heterogeneous modalities. It can be seen that on the ICS-Flow, ICS-ADD and TON_IoT datasets, their F1 and AUC scores are generally lower than those of the comparison schemes based on late fusion (MMCD) and cross-domain fusion without industrial priors (ADDA), verifying the upper-limit limitation of the simple concatenation strategy. Our KADF framework (the last row) comprehensively incorporates industrial asset topology, cross-domain temporal alignment, and security priors through the three steps of asset embedding, contrastive alignment and prior guidance. Thus, all its metrics significantly outperform the model-level and cross-domain fusion baselines in the table: its F1 scores on the three datasets are improved by approximately 2.5%, 1.6% and 2.0%, respectively, compared with the best baseline (ADDA).

E. Ablation Study

TABLE IV
ABLATION STUDY ON ICS-FLOW AND ICS-ADD DATASETS (%)

	ICS-Flow		ICS-ADD	
	F1	AUC	F1	AUC
Full KADF	97.18	97.47	96.35	96.12
w/o $\mathcal{L}_{align}$	96.59	96.09	95.81	95.62
w/o $\mathcal{L}_{prior}$	96.44	95.62	95.69	95.35
w/o $asset$	95.71	95.18	95.17	95.01
w/o $info$	95.38	95.05	94.98	94.65

For quantitative analysis of component contributions, we progressively removed the contrastive alignment loss ($\mathcal{L}_{align}$), prior information loss ($\mathcal{L}_{prior}$), asset embedding ($asset$), and security prior knowledge infomation($info$) on the ICS-Flow and ICS-ADD datasets while keeping the remaining structure unchanged. Results are shown in Table IV. Removing $\mathcal{L}_{align}$ decreased F1 by 0.59% and 0.54%, proving latent space alignment effectively bridges semantic gaps across heterogeneous sources. Excluding $\mathcal{L}_{prior}$ led to F1 drops of 0.74% and 0.66%, confirming security priors are essential for highlighting high-risk cross-domain correlations. Removing $asset$ caused the most notable performance decline, with F1 falling by 1.47% and 1.18%, emphasizing asset identifiers' critical role in alleviating modality asynchrony. Finally, replacing the two-stage fusion with simple early concatenation further reduced F1 by 1.80% and 1.37%, demonstrating that historical prior matching aids in capturing the temporal evolution of attacks. In summary, the full KADF achieves the highest F1 and AUC scores on both datasets, and the absence of any module leads to consistent performance degradation. This monotonic decline fully validates the effectiveness of all proposed components.

F. Parameter Sensitivity Analysis

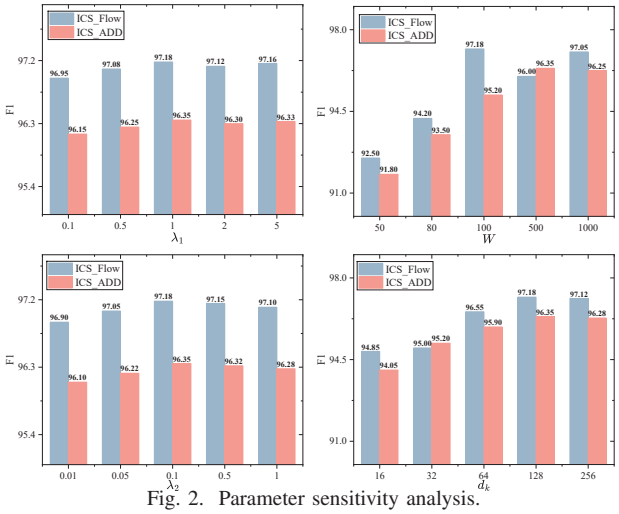
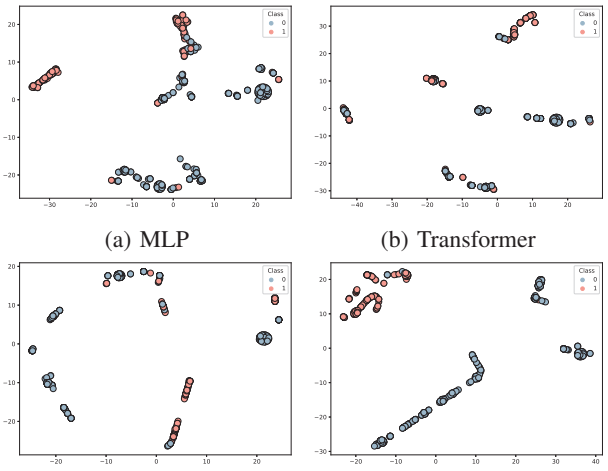


Fig. 2. Parameter sensitivity analysis.

We conduct hyperparameter sensitivity analysis of KADF on the ICS-Flow and ICS-ADD datasets. As illustrated in Figure 2, the proposed method demonstrates remarkable robustness to variations in λ_1 and λ_2. In contrast, the smaller windows provide insufficient temporal information, thereby impeding the model's ability to learn meaningful attack patterns. Similarly, overly small embedding dimensions (d_k) limit the model's capacity to capture discriminative features.

G. Visualization



(a) MLP (b) Transformer

(c) ADDA (d) KADF

Fig. 3. t-SNE visualization of ICS-ADD under four methods.

Figure 3 presents the t-SNE visualization of representation vectors extracted by four methods on an ICS-ADD test set. Blue and pink points denote positive and negative samples. MLP exhibits considerable overlap in the boundary regions, indicating limited discriminative capacity. Transformer achieves more compact intra-class clustering compared to MLP, yet partial interleaving of positive and negative samples persists with an indistinct decision boundary. ADDA leverages industrial data with domain adaptation techniques, yielding improved separation over Transformer through adversarial alignment of source and target distributions; however, the overlap remains in certain regions due to the absence of explicit security prior guidance. KADF demonstrates superior separability with maximal inter-class margins and highly concentrated intra-class distributions. These qualitative findings corroborate that KADF achieves optimal representation discriminability for the ICS-ADD binary classification task.

V. Conclusions

This paper addresses the challenge of multi-source heterogeneous data fusion in industrial control scenarios. We propose KADF, Knowledge-Guided Aligned Multi-Source Data Fusion for ICS Anomaly Detection framework. KADF injects domain knowledge into the representation space via extensible asset embedding, explicitly reduces cross-domain distribution gaps with a contrastive alignment loss. A security-prior-based attention bias mechanism is designed to actively focus the model on high-risk cross-domain correlations. Experiments on multiple public datasets demonstrate that KADF improves the F1 and AUC scores by approximately 2.6% and 2.3% respectively compared with the best baseline. Ablation studies further validate the significant contributions of key modules. Future work will explore the more cyber kill chain knowledge to construct graph-structured priors with enhanced interpretability, and conduct online validation in real PLC/DCS closed-loop environments to evaluate real-time inference latency and false alarm rate for developing deployable security monitoring tools for industrial sites.

References

[1] R. Kumar, R. Kela, S. Singh, and R. Trujillo-Rasua, "Apt attacks on industrial control systems: A tale of three incidents," International Journal of Critical Infrastructure Protection, vol. 37, JUL 2022.

[2] N. Bakalos, A. Voulodimos, N. Doulamis, A. Doulamis, A. Ostfeld, E. Salomons, J. Caubet, V. Jimenez, and P. Li, "Protecting water infrastructure from cyber and physical threats: Using multimodal data fusion and adaptive deep learning to monitor critical systems," IEEE Signal Processing Magazine, vol. 36, no. 2, pp. 36–48, 2019.

[3] C. Feng, V. R. Palleti, A. Mathur, and D. Chana, "A systematic framework to generate invariants for anomaly detection in industrial control systems," Proceedings 2019 Network and Distributed System Security Symposium, 2019.

[4] Y. Zhang, "Intelligent control system for industrial robots based on multi-source data fusion," Journal of Intelligent Systems, vol. 32, no. 1, SEP 7 2023.

[5] K. Liu and P. Xu, "Anomaly detection based on multi-source heterogeneous data fusion," Proceedings of SPIE, 2022.

[6] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, and R. Candell, "A survey of physics-based attack detection in cyber-physical systems," ACM Computing Surveys, vol. 51, no. 4, pp. 1–36, 2018.

[7] Y. Du, Y. Huang, G. Wan, and P. He, "Deep learning-based cyber–physical feature fusion for anomaly detection in industrial control systems," Mathematics, vol. 10, no. 22, 2022. [Online]. Available: https://www.mdpi.com/2227-7390/10/22/4373

[8] S. Bahadoripour, E. Macdonald, and H. Karimipour, "A deep multi-modal cyber-attack detection in industrial control systems," in 2023 IEEE International Conference on Industrial Technology (ICIT), 2023 2023, Conference Paper, pp. 1–6, 2023 IEEE International Conference on Industrial Technology (ICIT), 4-6 April 2023, Orlando, FL, USA.

[9] I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, and S. Raychaudhuri, "Fast, sensitive and accurate integration of single-cell data with harmony," Nature Methods, vol. 16, no. 4, pp. 1–8, 2019.

[10] S. Zhu, H. Hua, and S. Chen, "Rigorous integration of single-cell atac-seq data using regularized barycentric mapping," Nature Machine Intelligence, vol. 7, no. 9, pp. 1461–1477, 2025.

[11] A. Dehlaghi-Ghadim, M. H. Moghadam, A. Balador, and H. Hansson, "Anomaly detection dataset for industrial control systems," IEEE Access, vol. 11, pp. 107 982–107 996, 2023.

[12] M. Hussain, E. Foo, and S. Suriadi, "An improved industrial control system device logs processing method for process-based anomaly detection," in 2019 International Conference on Frontiers of Information Technology (FIT), 2019, pp. 150–1505.

[13] X. T. WANG Jingcai, LI Yan, "Modeling of false data injection attack and rapid screening of vulnerable lines under attack," Electric Power Construction, vol. 43, no. 1, p. 9, 2022.

[14] S. Bahadoripour, H. Karimipour, A. N. Jahromi, and A. Islam, "An explainable multi-modal model for advanced cyber-attack detection in industrial control systems," Internet Of Things, vol. 25, APR 2024.

[15] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, and R. Satija, "Comprehensive integration of single-cell data," Cell, vol. 177, no. 7, pp. 1888–1902.e21, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0092867419305598

[16] H. Taud and J.-F. Mas, "Multilayer perceptron (mlp)," in Geomatic approaches for modeling land change scenarios. Springer, 2017, pp. 451–455.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," Advances in neural information processing systems, vol. 30, 2017.

[19] A. Dehlaghi-Ghadim, M. H. Moghadam, A. Balador, and H. Hansson, "Anomaly detection dataset for industrial control systems," IEEE Access, vol. 11, pp. 107 982–107 996, 2023.

[20] G. B. Gaggero, A. Armellin, G. Portomauro, and M. Marchese, "Industrial control system-anomaly detection dataset (ics-add) for cyber-physical security monitoring in smart industry environments," IEEE Access, vol. 12, pp. 64 140–64 149, 2024.

[21] N. Moustafa, "A new distributed architecture for evaluating ai-based security systems at the edge: Network ton_iot datasets," Sustainable Cities and Society, vol. 72, p. 102994, 2021.