
A DOMAIN-INVARIANT SPATIO-TEMPORAL GRAPH ATTENTION NETWORK FOR CROSS-DOMAIN ANOMALY DETECTION IN INDUSTRIAL CONTROL SYSTEMS

**Emejuo Nichodemus Chibueze¹, Idris Shehu Musa², Basira Yahaya^{2,*}, Sani Saleh
Saminu², Mohammed Bashir Abdulrazaq²**

¹National Research Institute for Chemical Technology (NARICT), Zaria

²Ahmadu Bello University, Zaria

Email: emejue.chibueze@narict.gov.ng, Idris.ms@icloud.com, basirae@gmail.com*,
Sssaleh@engineering.abu.edu.ng, mbabdulrazaq@abu.edu.ng

***Corresponding author:** Basira Yahaya

DOI: <https://doi.org/10.67203/sp.ICESID.2026>

ABSTRACT

Industrial Control Systems (ICSs) are cyber-physical systems responsible for monitoring and controlling critical industrial processes. Their increasing integration with Information Technology (IT) networks, the Industrial Internet of Things (IIoT), and cloud-based services has expanded their exposure to cyber threats, creating a need for intelligent anomaly detection models capable of transferring learned knowledge across heterogeneous industrial environments. However, many existing deep learning-based approaches are developed for individual ICS domains and therefore have limited capability to exploit transferable spatio-temporal patterns across systems with different operational characteristics and feature structures. This study proposes a Domain-Invariant Spatio-Temporal Graph Attention Network (DI-STGAT) for cross-domain anomaly detection in ICSs. The proposed model integrates a Spatio-Temporal Graph Attention Network (ST-GAT) with Domain-Adversarial Neural Network (DANN) learning through a Gradient Reversal Layer (GRL). The ST-GAT captures temporal behaviour and spatial dependencies among industrial process variables, while the DANN/GRL mechanism facilitates domain adaptation by reducing domain-specific information in the learned representations. A fixed-dimensional graph-level representation enables the same model architecture to operate across heterogeneous ICS datasets with different numbers of process variables and graph structures. DI-STGAT achieved cross-domain anomaly detection across heterogeneous ICS environments by combining spatio-temporal graph learning with domain-adversarial adaptation to learn transferable representations from labeled source data and unlabeled target data, thereby improving anomaly discrimination across industrial domains without requiring target-domain anomaly labels for training. The framework was evaluated using the BATADAL, SWaT, and WADI benchmark datasets across six source-to-target transfer scenarios. Experimental results show that DI-STGAT improved mean cross-domain balanced accuracy from 50.51% to 51.90% and mean F1-score from 4.13% to 4.37% compared with the ST-GAT baseline at a fixed decision threshold of 0.50. The strongest improvement was achieved for SWaT→WADI, where balanced accuracy increased from 53.04% to 61.37% and F1-score from 7.48% to 8.94%. The proposed framework achieved mean cross-domain ROC-AUC and PR-AUC values of 72.21% and 36.36%, respectively. Domain-representation analysis further demonstrated measurable adversarial alignment between source and target representations, providing evidence of transferable representation learning across heterogeneous industrial domains. Overall, the findings demonstrate that DI-STGAT provides a viable domain-adaptive spatio-temporal learning framework for cross-domain anomaly detection in heterogeneous ICS environments, supporting more adaptable cybersecurity monitoring and improved protection of critical industrial infrastructure.

Keywords: Industrial Control Systems (ICS); Cross-Domain Anomaly Detection; Domain Adaptation; Spatio-Temporal Graph Attention Network (ST-GAT); Domain-Adversarial Neural Network (DANN); Gradient Reversal Layer (GRL)

1. INTRODUCTION

Industrial Control Systems (ICSs) are critical components of modern industrial infrastructure, providing monitoring and control functions for essential production processes. Traditionally, ICS environments were designed as isolated operational technology (OT) systems, with primary emphasis on safety, reliability, and continuous availability. However, the adoption of Industry 4.0, Industrial

Internet of Things (IIoT), remote monitoring, cloud services, Artificial Intelligence and information technology–operational technology (IT-OT) convergence has transformed these environments into highly interconnected cyber physical systems. While this transformation has improved automation, monitoring, and operational efficiency, it has also increased system exposure to cyber threats and introduced vulnerabilities associated with network connectivity, remote access, heterogeneous devices, and interconnected control components (Ahmadi-Assalemi *et al.*, 2025; Rehman & Bahşi, 2024; Rotibi *et al.*, 2024; Wu *et al.*, 2026). Unlike conventional IT systems, cyber-attacks against ICSs can propagate from the cyber layer to physical processes, potentially causing equipment damage, production interruption, environmental consequences, and safety incidents. Consequently, reliable detection of abnormal system behaviour has become an important component of ICS cybersecurity and resilience (Rehman & Bahşi, 2024; Wu *et al.*, 2026).

The increasing complexity of ICS environments has exposed limitations in conventional intrusion detection approaches. Thus, signature and rule-based methods are effective for previously known attack patterns but are less capable of identifying novel or evolving attacks that do not match predefined signatures. Anomaly-based detection provides an alternative by learning normal system behaviour and identifying deviations associated with attacks, faults, or abnormal operating conditions (Choi & Kim, 2024; Ahmadi-Assalemi *et al.*, 2025). The availability of high-dimensional operational data from sensors, actuators, controllers, and communication systems has consequently encouraged the application of machine learning (ML) and deep learning (DL) techniques for ICS anomaly detection. Recent studies have explored recurrent networks, convolutional architectures, autoencoders, attention mechanisms, and adaptive learning approaches to capture complex temporal patterns in industrial data (Aslam *et al.*, 2025; Liu *et al.*, 2024; Xu *et al.*, 2025). For instance, Liu *et al.* (2024) introduced a time-frequency fusion feature-attention approach to address long-sequence and complex temporal dependencies in ICS operational data, while Xu *et al.* (2025) proposed a dual-attention network that uses attention-based causal relationships among industrial components to improve anomaly detection and interpretation. These studies demonstrate the potential of deep learning for modelling complex ICS behaviour, but they primarily focus on improving detection within individual operational environments.

A major challenge in ICS anomaly detection is that industrial processes are inherently spatio-temporal. Sensor and actuator variables do not operate independently; they interact through physical processes, controllers, communication networks, and process stages. Consequently, treating industrial measurements as independent or purely sequential variables may overlook important relationships between interconnected components. Graph neural networks (GNNs) provide a natural representation for such systems because nodes can represent sensors, actuators, controllers, or process variables, while edges represent their physical, functional, or statistical relationships. Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) have therefore attracted increasing attention for multivariate anomaly detection because they can learn dependencies among interconnected variables (Deng & Hooi, 2021; Veličković *et al.*, 2018). In ICS applications, Yang *et al.* (2024) demonstrated that centrality-aware graph convolution can exploit relationships among industrial components, while Wang *et al.* (2023) showed that spatio-temporal graph neural networks can capture spatial and temporal characteristics in high-dimensional and noisy ICS data. Similarly, Fan *et al.* (2024) developed a spatio-temporal gated graph attention network for IIoT anomaly detection, and Zhang *et al.* (2025) proposed PdGAT-ID to combine periodic feature extraction with spatio-temporal graph attention for ICS intrusion detection. These studies indicate that graph-based spatio-temporal modelling can provide richer representations than conventional independent-variable approaches.

Recent research has further demonstrated the benefits of explicitly modelling hierarchical and dynamic relationships within industrial systems. Lin *et al.* (2025) proposed a multi-level graph attention network that represents relationships among physical processes and controllers, while Li *et al.* (2025) integrated static and dynamic graph learning with attention-based temporal modelling to capture spatial and temporal dependencies in ICS data. Canonico *et al.* (2025) similarly demonstrated that reducing unnecessary graph connections can improve both detection efficiency and computational requirements in reconfigurable cyber–physical systems. These findings reinforce the importance of jointly modelling spatial dependencies and temporal dynamics rather than treating industrial observations as isolated time-series measurements.

Despite these advancements, an important limitation remains: most existing ICS anomaly detection models are developed and evaluated within a single operational domain. Models trained on one industrial environment generally assume that the statistical distribution, sensor relationships, process dynamics, and attack characteristics of the training environment remain applicable to the test environment. This assumption is problematic because real industrial systems are heterogeneous. Water treatment, water distribution, energy, manufacturing, chemical, and oil-and-gas systems differ substantially in process topology, sensor configurations, control logic, communication protocols, operating ranges, and anomaly characteristics. Recent reviews have identified limited cross-domain datasets, distribution shifts, concept drift, and inadequate real-world validation as continuing challenges for AI-based ICS anomaly detection (Moriano *et al.*, 2025; Aslam *et al.*, 2025; Wu *et al.*, 2026). The problem is therefore not simply to achieve high detection accuracy within one dataset, but to develop representations that remain informative when the model encounters a different industrial domain.

Domain adaptation provides a promising approach to this problem by encouraging a model to learn features that are discriminative for anomaly detection while being less dependent on the identity of the training domain. Domain-Adversarial Neural Networks (DANN) achieves this by introducing a domain classifier and a Gradient Reversal Layer (GRL), which encourages the feature extractor to learn representations that are useful for the primary task but less distinguishable across domains. Although adversarial learning has been extensively investigated in general domain adaptation, its integration with spatio-temporal graph representations for cross-domain ICS anomaly detection remains comparatively limited. Existing graph-based ICS studies have primarily concentrated on improving spatial representation, temporal modelling, interpretability, or detection accuracy within selected benchmark datasets rather than explicitly enforcing domain-invariant representations. This limitation is particularly important because graph structures themselves may differ between industrial environments, making direct transfer of learned representations difficult (Nahalka *et al.*, 2024; Fan *et al.*, 2024; Lin *et al.*, 2025; Ganin *et al.*, 2016).

The challenge is further highlighted by the nature of commonly used ICS benchmarks. Datasets such as SWaT, WADI, and BATADAL provide valuable experimental environments for evaluating anomaly detection methods, but they represent different industrial processes and therefore introduce distributional differences. SWaT represents a water-treatment environment, WADI represents a water-distribution system, and BATADAL focuses on cyber-physical attacks against water-distribution networks (Goh *et al.*, 2016; Ahmed *et al.*, 2017; Mathur & Taormina *et al.*, 2018). These datasets have enabled substantial progress in ICS anomaly detection, but their differences also provide an appropriate basis for evaluating whether a model can generalize beyond its source domain. Studies using graph-based approaches have demonstrated strong performance on individual datasets, yet the problem of systematically evaluating transferability between heterogeneous domains remains insufficiently explored.

Against this background, this study identifies three interrelated research problems that motivate the development and evaluation of the proposed framework.

1. Inadequate joint representation of spatial and temporal dependencies across heterogeneous ICS environments.
2. Poor cross-domain generalization caused by distribution shifts between heterogeneous ICS environments.
3. Limited systematic evidence for cross-domain transfer and domain-adversarial representation learning across heterogeneous ICS benchmarks.

To answer these questions, the study adopts a controlled experimental framework involving the three benchmark datasets. Each dataset is independently preprocessed and represented using its domain-specific graph structure, thereby preserving the characteristics of the corresponding industrial environment rather than forcing heterogeneous variables into an artificial common node space. Temporal sequences are subsequently processed through the ST-GAT architecture to generate a fixed-dimensional graph representation. The conventional ST-GAT is first trained and evaluated as the baseline. DI-STGAT then extends the same architecture with a GRL and domain classifier for unsupervised source-to-target adaptation. Six transfer directions: BATADAL to SWaT, BATADAL to WADI, SWaT to BATADAL, SWaT to WADI, WADI to BATADAL, and WADI to SWaT are

evaluated under a controlled protocol. Target-domain anomaly labels are excluded from adaptation training and model selection and are reserved for final evaluation.

Accordingly, the main contributions of this study are:

1. **Domain-invariant ST-GAT:** A spatio-temporal graph attention framework integrated with DANN and GRL to learn transferable anomaly representations across heterogeneous ICS domains.
2. **Cross-domain evaluation:** A systematic source-to-target evaluation using the three datasets to assess the generalization capability of the proposed model beyond its training environment.
3. **Controlled baseline and ablation analysis:** A conventional ST-GAT baseline is established and compared with the DANN/GRL-enhanced model to isolate the contribution of domain-adversarial learning.

Further, the proposed framework has potential relevance to addressing the increasing need to strengthen the resilience and cybersecurity of critical infrastructure in Nigeria and other African countries, where increasing digitalization and interconnection of industrial and utility systems can introduce additional cyber security risks. Nigeria's National Cyber security Policy and Strategy identifies the protection and resilience of Critical National Information Infrastructure as a national cybersecurity priority (Federal Republic of Nigeria, 2021). Water treatment and distribution systems provide a particularly relevant application context because reliable water infrastructure remains important to Nation's sustainable development, while challenges relating to infrastructure, service delivery, and access to clean water persist (Ogunbode, 2026). The use of water-system benchmark datasets in this study therefore provides an experimental basis for investigating transferable anomaly-detection techniques applicable to heterogeneous water infrastructure. By reducing dependence on target-domain anomaly labels, the proposed cross-domain approach could support the development of adaptable cybersecurity monitoring for infrastructure environments where labeled attack data are limited. Such capabilities are relevant to Nigeria's sustainable infrastructure development and align with SDG 6 (Clean Water and Sanitation) and SDG 9 (Industry, Innovation and Infrastructure), which emphasize sustainable water management and resilient infrastructure, respectively (United Nations, 2026).

2. METHODOLOGY

This study proposes a Domain-Invariant Spatio-Temporal Graph Attention Network (DI-STGAT) for cross-domain anomaly detection in heterogeneous Industrial Control Systems (ICSs). The proposed framework combines temporal attention, graph attention, and domain-adversarial learning to capture temporal behaviour and structural dependencies among industrial process variables while encouraging transferable representations across different ICS domains. Unlike approaches that require identical feature sets across source and target systems, DI-STGAT retains the domain-specific variables and graph structures of each ICS and transforms them into a common fixed-dimensional graph-level representation suitable for cross-domain adaptation.

2.1 Model Overview

The proposed DI-STGAT consists of three principal components: (i) an ST-GAT feature extractor, (ii) an anomaly-classification branch, and (iii) a domain-adaptation branch based on a Domain-Adversarial Neural Network (DANN) and Gradient Reversal Layer (GRL). The ST-GAT feature extractor first applies temporal attention to identify informative observations within each process variable's temporal sequence. The resulting node representations are subsequently processed by a multi-head Graph Attention Network (GAT) to capture dependencies among interconnected industrial variables. Mean graph pooling then converts the variable-size node representations into a fixed 32-dimensional graph-level embedding. This processing sequence corresponds to the implemented architecture described in the original model specification.

The graph-level embedding is supplied to two branches. The first is a binary anomaly classifier that distinguishes normal from anomalous process behaviour. The second is a domain classifier connected through the GRL. During domain adaptation, labeled source-domain samples provide supervision for anomaly classification, whereas source and unlabeled target samples jointly

contribute to domain classification. Target-domain anomaly labels are not used during adaptation training or model selection. Accordingly, the overall architecture can be summarized as:

$$\begin{aligned} & \text{Input} \rightarrow \text{Temporal Attention} \rightarrow \text{Multi-Head GAT} \rightarrow \text{Graph Embedding} \rightarrow \\ & \left\{ \begin{array}{l} \text{Anomaly Classifier} \\ \text{GRL} \rightarrow \text{Domain Classifier} \end{array} \right. \quad (1) \end{aligned}$$

As show in Equation (1), the anomaly and domain-classification branches operate on the same graph-level representation. This shared representation provides the connection between spatio-temporal anomaly learning and domain-adversarial adaptation as shown in figure 2.1.

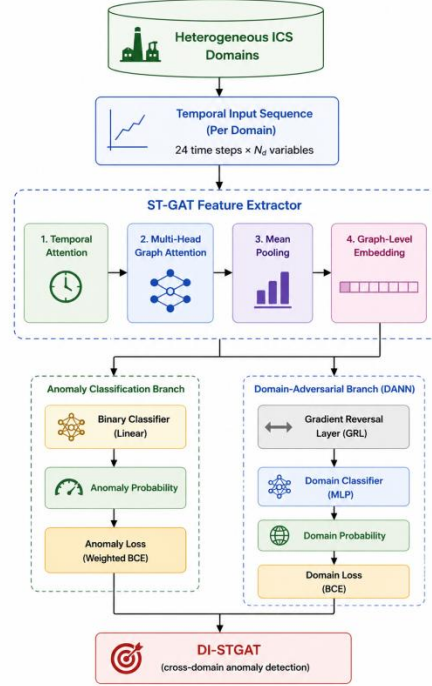


Figure 2.1: Architecture of the Proposed DI-STGAT Model

2.2 Domain-Specific Graph and Temporal Input Representation

The three ICS domains considered in this study contain different numbers and configurations of process variables. Rather than forcing these heterogeneous datasets into an artificial common feature space, each domain retains its own set of process variables and corresponding graph structure. For an ICS domain d , a temporal input sequence is represented as in equation 2:

$$\mathbf{X}_t^{(d)} = [\mathbf{x}_{t-T+1}^{(d)}, \mathbf{x}_{t-T+2}^{(d)}, \dots, \mathbf{x}_t^{(d)}] \in \mathbb{R}^{T \times N_d}, \quad (2)$$

where T is the temporal window length and N_d is the number of process variables in domain d . In this study, $T = 24$; therefore, each input contains the previous 24 observations of the corresponding industrial variables. As expressed in Equation (2), N_d is domain-dependent, allowing the model to accommodate heterogeneous ICS environments rather than assuming an identical input dimension. The implemented datasets contain 38 variables for BATADAL, 45 for SWaT, and 123 for WADI. Each ICS domain is represented by a graph

$$G^{(d)} = (V^{(d)}, E^{(d)}), \quad (3)$$

where $V^{(d)}$ represents the process variables and $E^{(d)}$ represents their inferred relationships. Equation (3) therefore permits each domain to maintain a graph topology corresponding to its own process structure. The relationships between variables are determined using the absolute Pearson correlation computed from the corresponding training partition:

$$r_{ij}^{(d)} = \left| \frac{\sum_t (x_{t,i}^{(d)} - \bar{x}_i^{(d)})(x_{t,j}^{(d)} - \bar{x}_j^{(d)})}{\sqrt{\sum_t (x_{t,i}^{(d)} - \bar{x}_i^{(d)})^2} \sqrt{\sum_t (x_{t,j}^{(d)} - \bar{x}_j^{(d)})^2}} \right|. \quad (4)$$

In Equation (4), $r_{ij}^{(d)}$ measures the strength of the statistical relationship between process variables i and j . For each node, the three variables with the highest absolute correlations are selected as neighbours. The resulting graph is symmetrized and self-loops are added. This produces domain-specific graphs containing 38 nodes and 188 directed edges for BATADAL, 45 nodes and 253 directed edges for SWaT, and 123 nodes and 707 directed edges for WADI. Importantly, the graph topology is derived exclusively from training data to prevent validation or test information from influencing graph construction.

2.3 Temporal Attention

Industrial anomalies frequently develop over a sequence of observations rather than appearing as isolated events. The temporal attention module is therefore used to determine the relative importance of the 24 observations associated with each process-variable node. For node i , the attention score at time t is calculated as

$$e_{i,t} = \mathbf{w}_2^\top \tanh(\mathbf{W}_1 x_{i,t} + \mathbf{b}_1) + b_2, \quad (5)$$

where $\mathbf{W}_1, \mathbf{w}_2, \mathbf{b}_1$, and b_2 are trainable parameters. The implemented temporal-attention network uses a $1 \rightarrow 8 \rightarrow 1$ transformation. Equation (5) assigns a learned importance score to every observation within the temporal window. The attention scores are normalized across the temporal dimension using the softmax function:

$$\alpha_{i,t} = \frac{\exp(e_{i,t})}{\sum_{\tau=1}^T \exp(e_{i,\tau})}. \quad (6)$$

As defined by Equation (6), the normalized temporal attention coefficients indicate the relative contribution of each time step to the representation of node i . The temporal representation is subsequently obtained through the weighted aggregation

$$h_i^{(T)} = \sum_{t=1}^T \alpha_{i,t} x_{i,t}. \quad (7)$$

Thus, Equation (7) transforms the 24-step history of each process variable into a learned scalar representation in which more informative time steps receive greater attention. This allows the model to capture temporal behaviour without introducing recurrent or convolutional components.

2.4 Multi-Head Graph Attention

Following temporal attention, the resulting node representations are processed by a multi-head graph-attention mechanism to capture spatial dependencies among related industrial process variables. Unlike conventional aggregation methods that treat all neighbouring nodes equally, graph attention learns the relative contribution of each neighbouring variable. For an edge connecting nodes i and j , the unnormalized attention score for attention head k is calculated as

$$e_{ij}^{(k)} = \text{LeakyReLU} \left[\mathbf{a}^{(k)\top} (\mathbf{W}^{(k)} h_i^{(T)} \parallel \mathbf{W}^{(k)} h_j^{(T)}) \right], \quad (8)$$

where $\mathbf{W}^{(k)}$ denotes the trainable transformation matrix, $\mathbf{a}^{(k)}$ represents the attention vector for head k , and $\parallel$ denotes concatenation. Equation (8) determines the learned importance of node j to node i within the domain-specific graph. The attention scores are normalized over the neighbourhood of node i as

$$\beta_{ij}^{(k)} = \frac{\exp(e_{ij}^{(k)})}{\sum_{m \in N_i} \exp(e_{im}^{(k)})}, \quad (9)$$

where N_i contains the graph neighbours of node i , including its self-loop. Equation (9) ensures that the attention coefficients represent the relative importance of connected variables. The output of attention head k is then obtained as

$$\mathbf{h}_i^{(k)} = \sum_{j \in N_i} \beta_{ij}^{(k)} \mathbf{w}^{(k)} h_j^{(T)}. \quad (10)$$

According to Equation (10), each node representation is updated using an attention-weighted combination of information from its graph neighbourhood. The implemented model uses four graph-attention heads with eight output dimensions per head. Their outputs are concatenated to produce

$$\mathbf{h}_i^{(G)} = \parallel_{k=1}^4 \mathbf{h}_i^{(k)} \in \mathbb{R}^{32}. \quad (11)$$

Consequently, Equation (11) produces a 32-dimensional spatial representation for each process variable node. The use of multiple attention heads allows the model to capture different patterns of relationships among industrial variables while maintaining the same parameterization across domains with different graph sizes.

2.5 Graph-Level Representation and Anomaly Classification

Although the same ST-GAT parameters can operate on different graph sizes, the resulting node level representations cannot be directly combined across domains because the datasets contain different numbers of nodes. A fixed-dimensional graph representation is therefore generated using mean pooling:

$$\mathbf{z}^{(d)} = \frac{1}{N_d} \sum_{i=1}^{N_d} \mathbf{h}_i^{(G)}, \mathbf{z}^{(d)} \in \mathbb{R}^{32}. \quad (12)$$

As shown in Equation (12), the node representations are aggregated into a 32-dimensional graph level embedding, irrespective of the number of process variables in the input domain. This mechanism is important to cross-domain adaptation because each of the datasets can retain their respective domain-specific graphs while still producing representations with an identical dimensionality. The graph-level representation is subsequently passed to a binary anomaly classifier:

$$\ell_y = \mathbf{w}_y^T \mathbf{z} + b_y, p_y = \sigma(\ell_y), \quad (13)$$

where ℓ_y represents the anomaly logit, $\mathbf{w}_y$ and b_y are trainable classifier parameters, and p_y denotes the predicted anomaly probability. Equation (13) therefore maps the 32-dimensional graph representation to the probability that the corresponding temporal graph observation is anomalous. Because anomalous samples are substantially less frequent than normal samples, the anomaly-classification branch is optimized using weighted binary cross-entropy:

$$L_A = -\frac{1}{n_s} \sum_{i=1}^{n_s} [w_+ y_i^s \log p_i^s + (1 - y_i^s) \log(1 - p_i^s)], \quad (14)$$

where n_s denotes the number of source samples and $w_+ = N_{\text{normal}} / N_{\text{anomaly}}$ is calculated from source-domain training data only. Equation (14) increases the contribution of the minority anomaly class during optimization without resampling the training data.

2.6 Domain-Adversarial Learning

The principal extension from the conventional ST-GAT baseline to DI-STGAT is the incorporation of domain-adversarial learning. The objective is to retain representations that are useful for anomaly classification while reducing information that makes the source and target domains easily distinguishable. Let F_θ denote the shared ST-GAT feature extractor. For source and target observations, the corresponding graph-level representations are

$$\mathbf{z}_s = F_\theta(\mathbf{X}_s, G_s), \mathbf{z}_t = F_\theta(\mathbf{X}_t, G_t). \quad (15)$$

As expressed in Equation (15), the source and target inputs are processed using their respective domain-specific graphs, G_s and G_t , while the shared ST-GAT parameters map both inputs into the same 32-dimensional representation space. This enables domain adaptation without requiring the source and target systems to contain identical process variables. The resulting representations are

passed through the GRL to a domain classifier implemented as a 32 to 16 to 1 network with a ReLU hidden activation:

$$p_d = \sigma[\mathbf{W}_{d2}\text{ReLU}(\mathbf{W}_{d1}\mathbf{z}^{GRL} + \mathbf{b}_{d1}) + b_{d2}]. \quad (16)$$

In Equation (16), p_d represents the predicted domain probability and $\mathbf{z}^{GRL}$ is the graph embedding after passing through the GRL. Source samples are assigned domain label 0, while target samples are assigned domain label 1. The domain-classification loss is calculated as:

$$L_D = \frac{1}{2} [\text{BCE}(p_d^s, 0) + \text{BCE}(p_d^t, 1)]. \quad (17)$$

As shown in Equation (17), both source and target samples contribute to the domain-adversarial objective. However, target anomaly labels are not involved. Consequently, DI-STGAT performs unsupervised target-domain adaptation, using labeled source data for anomaly learning and unlabeled target data for domain alignment.

2.7 Gradient Reversal Layer

The GRL provides the adversarial mechanism through which the ST-GAT representation is encouraged to become less domain-specific. During forward propagation, the GRL behaves as an identity function:

$$R_\lambda(\mathbf{z}) = \mathbf{z}. \quad (18)$$

Thus, Equation (18) shows that the numerical values of the graph-level representation are passed unchanged to the domain classifier. During back propagation, however, the gradient is reversed according to

$$\frac{\partial R_\lambda}{\partial \mathbf{z}} = -\lambda \mathbf{I}, \quad (19)$$

where λ controls the strength of the adversarial signal. According to Equation (19), the domain classifier is optimized to correctly distinguish source and target representations, while the reversed gradient encourages the ST-GAT feature extractor to generate representations that make this distinction more difficult. This adversarial relationship encourages the extraction of more transferable features while retaining the anomaly-discriminative objective of the source task. Rather than applying the maximum adversarial pressure from the beginning of training, λ is progressively increased according to

$$\lambda(p) = \frac{2}{1 + \exp(-10p)} - 1, 0 \leq p \leq 1, \quad (20)$$

where p represents normalized training progress. Equation (20) gradually increases the influence of domain-adversarial learning as training progresses, allowing the feature extractor initially to learn anomaly-relevant representations before stronger domain-confusion pressure is introduced.

2.8 Joint DI-STGAT Training Objective

DI-STGAT jointly optimizes anomaly discrimination and domain adaptation. The overall objective is defined as

$$L_{\text{total}} = L_A + L_D, \quad (21)$$

where L_A is the source-domain anomaly-classification loss defined in Equation (14) and L_D is the source-target domain-classification loss defined in Equation (17). Therefore, Equation (21) jointly trains the ST-GAT representation for anomaly discrimination and domain-adversarial adaptation. Importantly, the adversarial coefficient λ is not multiplied by L_D again in Equation (21). Instead, λ acts exclusively through the reversed gradient defined in Equation (19) and follows the progressive schedule specified in Equation (20). This reflects the actual implementation and prevents the domain-adversarial contribution from being weighted twice.

During optimization, labeled source-domain samples contribute to the anomaly loss L_A , while both source-domain and unlabeled target-domain samples contribute to the domain loss L_D . The anomaly-classification objective encourages the 32-dimensional ST-GAT embedding to retain information required to distinguish normal from anomalous behaviour, whereas the GRL-based

domain objective encourages the feature extractor to reduce information that permits straightforward discrimination between source and target domains.

Consequently, the proposed DI-STGAT combines temporal attention for modelling process evolution, multi-head graph attention for capturing relationships among industrial variables, fixed-dimensional graph pooling for accommodating heterogeneous graph sizes, and DANN/GRL for cross-domain representation adaptation. This integrated architecture enables knowledge learned from a labeled source ICS to be adapted to heterogeneous target ICS using unlabeled target-domain observations, without requiring identical process-variable sets or target-domain anomaly labels during training.

3. RESULTS AND DISCUSSION

This section presents and discusses the experimental results of the conventional ST-GAT baseline and the proposed Domain-Invariant Spatio-Temporal Graph Attention Network (DI-STGAT). The analysis considers within-domain ST-GAT performance, baseline cross-domain transfer, cross-domain adaptation using DANN/GRL, comparative performance between ST-GAT and DI-STGAT, decision-threshold behaviour, and domain-representation alignment. The quantitative results are complemented with graphical analyses of discrimination performance, cross-domain transfer, domain-adaptation effects, and training behaviour.

3.1 Within-Domain Performance of the ST-GAT Baseline

The ST-GAT baseline was first evaluated independently on BATADAL, SWaT, and WADI to determine whether the underlying architecture could learn anomaly-relevant representations before investigating cross-domain transfer. Table 3.1 presents the results obtained using the fixed classification threshold of 0.50.

Table 3.1: Within-domain ST-GAT performance

Dataset	Accuracy	Balanced Accuracy	Precision	Recall	F1-score	ROC-AUC	PR-AUC
BATADAL	0.0192	0.5058	0.0078	1.0000	0.0154	0.9197	0.0465
SWaT	0.0372	0.5000	0.0372	1.0000	0.0718	0.9655	0.3312
WADI	0.3677	0.6718	0.0547	1.0000	0.1037	0.7256	0.0620

The results reveal a clear distinction between the ranking capability of the ST-GAT and its classification behaviour at the fixed threshold. BATADAL and SWaT obtained ROC-AUC values of 0.9197 and 0.9655, respectively, indicating strong discrimination between normal and anomalous observations across possible decision thresholds. WADI achieved a lower but still informative ROC-AUC of 0.7256.

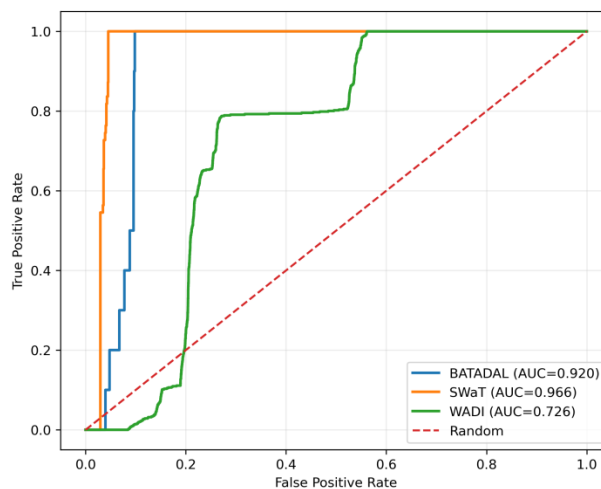


Figure 3.1: Within-domain ROC curves of the ST-GAT baseline

Figure 3.1 shows the ROC curves obtained by the ST-GAT baseline for BATADAL, SWaT, and WADI. The curves confirm the strong ranking capability observed numerically in Table 3.1,

particularly for SWaT and BATADAL. SWaT achieved the highest ROC-AUC of 0.9655, followed by BATADAL at 0.9197, whereas WADI achieved 0.7256. The figure therefore demonstrates that the ST-GAT extracted useful anomaly-related representations within each domain, although the fixed 0.50 threshold did not provide an equally effective operating point.

This difference is important because the models obtained a recall of 1.0000 for all three datasets but produced large numbers of false positives. For BATADAL, all 10 anomalous test samples were detected, but 1,277 of 1,292 normal samples were classified as anomalous. Similarly, all 55 SWaT attacks were detected, but all 1,422 normal samples were classified as attacks. WADI produced a comparatively better operating point, correctly identifying all 631 attacks and 5,714 normal observations, although 10,913 normal observations were still incorrectly classified as anomalous. The findings indicate that the ST-GAT baseline learned useful anomaly-ranking information but experienced substantial fixed-threshold calibration difficulty. This distinction becomes particularly important when examining cross-domain transfer.

3.2 Cross-Domain Generalization of the ST-GAT Baseline

The baseline models were subsequently evaluated under cross-domain transfer. In each experiment, a model trained on one source domain was evaluated on another target domain without target anomaly-label training or target-domain fine-tuning.

Table 3.2: Cross-domain ST-GAT baseline performance

Source → Target	Balanced Accuracy	F1-score	ROC-AUC	PR-AUC
BATADAL → SWaT	0.5000	0.0718	1.0000	1.0000
BATADAL → WADI	0.5001	0.0706	0.4734	0.0325
SWaT → BATADAL	0.5000	0.0152	1.0000	1.0000
SWaT → WADI	0.5304	0.0748	0.4909	0.0335
WADI → BATADAL	0.5000	0.0152	0.9305	0.0852
WADI → SWaT	0.5000	0.0000	0.4914	0.0367
Mean	0.5051	0.0413	0.7310	0.3646

As shown in table 3.2 the mean cross-domain ROC-AUC was 0.7310, whereas balanced accuracy and F1-score averaged 0.5051 and 0.0413, respectively. Considerable variation occurred among the six transfer directions.

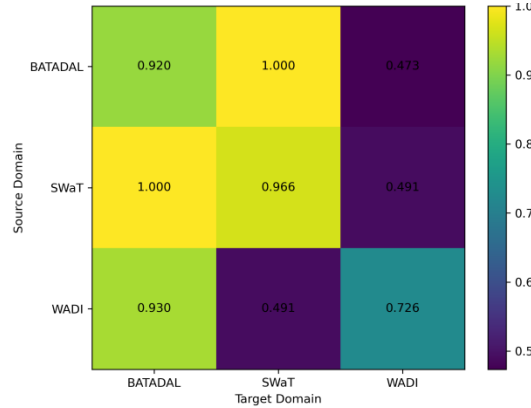


Figure 3.2: ROC-AUC within and Cross-domain performance

The Figure 3.2 provides a visual representation of the direction-dependent transfer behaviour of the baseline model. Strong transfer ranking was obtained for BATADAL to SWaT and SWaT to BATADAL, both of which achieved ROC-AUC values of 1.0000, while WADI to BATADAL achieved 0.9305. In contrast, BATADAL to WADI, SWaT to WADI, and WADI to SWaT produced values close to 0.50. The heatmap therefore shows that cross-domain transfer was not symmetric and that the effectiveness of a source representation depended strongly on the selected target environment.

The results also demonstrate that the ability of the architecture to process heterogeneous graphs does not automatically guarantee distributional generalization. Although all target graphs could be processed through the common ST-GAT architecture, substantial source-target distribution shifts remained, providing the motivation for the DANN/GRL adaptation mechanism.

3.3 Cross-Domain Performance of DI-STGAT

The proposed DI-STGAT was subsequently evaluated across the same six transfer directions. During adaptation, anomaly labels were obtained exclusively from the source domain, while unlabeled target-domain training samples participated only in domain-adversarial learning.

Table 3.3: Cross-domain performance of DI-STGAT

Source → Target	Accuracy	Balanced Accuracy	Precision	Recall	F1-score	ROC-AUC	PR-AUC
BATADAL → SWaT	0.0372	0.5000	0.0372	1.0000	0.0718	1.0000	1.0000
BATADAL → WADI	0.0367	0.5001	0.0366	1.0000	0.0706	0.4742	0.0325
SWaT → BATADAL	0.0077	0.5000	0.0077	1.0000	0.0152	1.0000	1.0000
SWaT → WADI	0.2556	0.6137	0.0468	1.0000	0.0894	0.4408	0.0306
WADI → BATADAL	0.0077	0.5000	0.0077	1.0000	0.0152	0.9296	0.0818
WADI → SWaT	0.9628	0.5000	0.0000	0.0000	0.0000	0.4882	0.0366
Mean	—	0.5190	—	—	0.0437	0.7221	0.3636

The proposed DI-STGAT as shown in table 3.3 achieved a mean balanced accuracy of 51.90% and mean F1-score of 4.37%, compared with 50.51% and 4.13%, respectively, for the conventional ST-GAT baseline. The largest improvement occurred for SWaT to WADI, where balanced accuracy increased from 53.04% to 61.37%, while F1-score increased from 7.48% to 8.94%. This improvement resulted primarily from better recognition of normal WADI observations while maintaining complete attack recall. However, a notable failure case occurred for WADI→SWaT, where DI-STGAT achieved 96.28% accuracy but zero precision, recall, and F1-score. As shown in Table 3.4, all 1,477 target samples were classified as normal (TN = 1,422, FN = 55), resulting in the failure to detect any of the 55 anomalies. Thus, the high accuracy reflects the strong class imbalance rather than effective anomaly detection. The near-chance ROC-AUC of 0.4882 further indicates weak anomaly discrimination, while the domain-classification balanced accuracy of 0.6687 suggests incomplete source–target alignment. This result demonstrates that DI-STGAT does not transfer uniformly across all domain pairs and that high accuracy can mask complete anomaly-detection failure. For practical deployment, transfer-specific validation using balanced accuracy, recall, F1-score, and PR-AUC is therefore essential before deploying the model in an unseen ICS environment.

Table 3.4: Confusion Matrix for the WADI→SWaT Transfer

Actual Class	Predicted Normal	Predicted Anomaly	Total
Normal	1,422 (TN)	0 (FP)	1,422
Anomaly	55 (FN)	0 (TP)	55
Total	1,477	0	1,477

3.4 Comparative Performance of ST-GAT and DI-STGAT

To determine the contribution of domain-adversarial adaptation, the performance of DI-STGAT was directly compared with the corresponding ST-GAT baseline. As summarized in Table 3.5, DI-STGAT produced modest increases in mean balanced accuracy and F1-score, while mean ROC-AUC and PR-AUC showed slight decreases, indicating that the effect of domain-adversarial adaptation differed between fixed-threshold decision performance and overall ranking performance.

Table 3.5: Aggregate comparison of ST-GAT and DI-STGAT

Metric	ST-GAT	DI-STGAT	Change
Balanced Accuracy	0.5051	0.5190	+0.0139
F1-score	0.0413	0.0437	+0.0024
ROC-AUC	0.7310	0.7221	−0.0089
PR-AUC	0.3646	0.3636	−0.0010

Figure 3.3 compares the F1-scores of the conventional ST-GAT and DI-STGAT for each source-target pair. The most visible difference occurs for SWaT to WADI, where DI-STGAT increases F1-score from 0.0748 to 0.0894. The remaining transfer directions show either equivalent or nearly equivalent fixed-threshold performance. The figure therefore indicates that the benefit of adversarial adaptation is transfer-dependent, with the largest decision-level improvement occurring when WADI is adapted from SWaT.

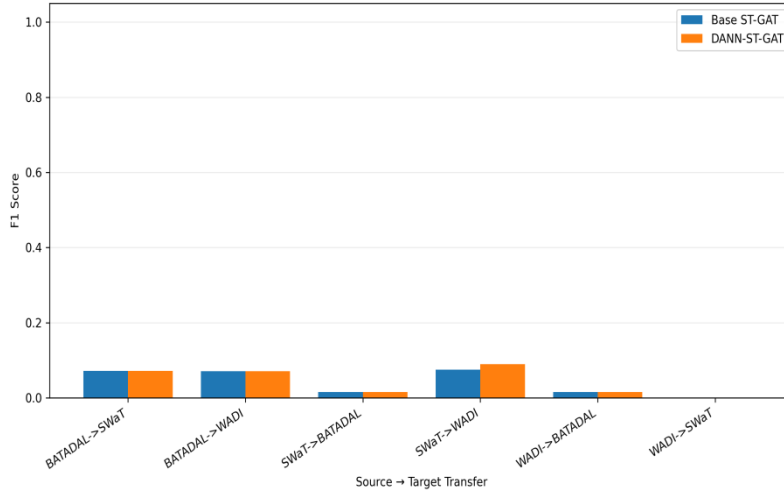


Figure 3.3: Cross-domain F1-score comparison between ST-GAT and DI-STGAT

Across all six transfers, mean balanced accuracy increased by 1.39 percentage points, while mean F1-score increased by only 0.24 percentage points. The largest F1 improvement occurred for SWaT→WADI, from 0.0748 to 0.0894; however, the absolute F1-score remains low. Conversely, mean ROC-AUC decreased by 0.89 percentage points and PR-AUC by 0.10 percentage points.

Therefore, although DANN/GRL produced a measurable numerical improvement in fixed-threshold decision performance for some source–target relationships, the magnitude and absolute performance levels are insufficient to demonstrate deployment-ready anomaly detection. Rather, the results provide evidence of the feasibility of domain-adversarial adaptation within the ST-GAT framework, while highlighting the need for substantial improvement in cross-domain anomaly discrimination and calibration before practical deployment. Accordingly, DI-STGAT should be regarded as an experimental cross-domain framework rather than an immediately deployable ICS anomaly-detection solution.

Further, the reported comparisons are based on the fixed experimental runs and were not evaluated across multiple random seeds. Consequently, the observed differences between ST-GAT and DI-STGAT, including the 1.39 percentage-point increase in mean balanced accuracy, should be interpreted as descriptive experimental differences rather than statistically significant improvements. Multiple-seed evaluation with confidence intervals and statistical significance testing is therefore required in future work to determine the robustness of these differences to random initialization and training variability.

3.5 Training Behaviour of DI-STGAT

The evolution of the anomaly and domain-learning objectives was monitored during DANN training. Because six source-target directions were evaluated, displaying every training curve in the

main manuscript would be repetitive. The SWaT to WADI curve is selected as the representative example because this transfer produced the largest improvement in balanced accuracy and F1-score.

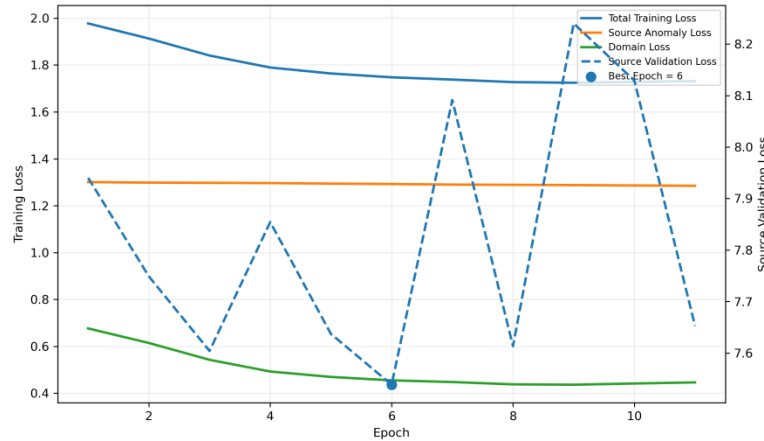


Figure 3.4: DI-STGAT training behaviour for SWaT to WADI

Figure 3.4 illustrates the evolution of the source anomaly-training loss, source validation loss, domain-classification loss, and GRL adaptation strength during SWaT to WADI training. The figure demonstrates the simultaneous optimization of anomaly discrimination and domain adaptation. As training progresses, the GRL coefficient increases according to the adopted schedule, progressively strengthening the adversarial signal applied to the ST-GAT feature extractor. Model selection remains governed by source-validation anomaly loss rather than target-domain labels.

The training behaviour is consistent with the unsupervised adaptation design of DI-STGAT: labeled source observations guide anomaly classification, whereas source and unlabeled target representations jointly contribute to the domain-adversarial objective. The remaining five training curves may be reported as supplementary material if required by the journal.

3.6 Domain-Representation Analysis

An important objective of the DANN/GRL component is to reduce the amount of domain-specific information retained in the learned representation. Domain-classification balanced accuracy was therefore analyzed after adaptation. A value approaching 0.50 indicates stronger source-target confusion, whereas a value approaching 1.00 indicates that the two domains remain readily distinguishable.

Table 3.6: Domain-representation diagnostic results

Source → Target	Domain Balanced Accuracy	Interpretation
BATADAL → SWaT	0.9888	Weak domain confusion
BATADAL → WADI	0.7911	Partial alignment
SWaT → BATADAL	0.8317	Limited alignment
SWaT → WADI	0.7931	Partial alignment
WADI → BATADAL	0.5000	Strong domain confusion
WADI → SWaT	0.6687	Moderate alignment
Mean	0.7622	Domain-dependent alignment

Table 3.6 shows substantial variation in the degree of source-target domain confusion produced by adversarial training. WADI to BATADAL achieved a domain balanced accuracy of 0.5000, indicating that the trained domain classifier could not distinguish the two representations better than chance. WADI to SWaT also showed increased confusion with a value of 0.6687. Conversely, BATADAL→SWaT remained highly distinguishable at 0.9888. The figure therefore demonstrates that DANN/GRL produced domain-dependent representation alignment rather than uniform domain invariance.

Table 3.7: Aggregate Domain-Representation Diagnostics for DI-STGAT

Diagnostic Measure	Mean Value	Interpretation
Balanced accuracy	0.7622	Indicates that source and target representations remained distinguishable overall; 0.50 represents chance-level domain discrimination.
Binary Cross-Entropy (BCE)	0.5442	Measures the domain classifier's prediction error; values approaching the binary chance-loss level indicate greater domain confusion.
Source–target probability	0.2134	Measures the difference between predicted domain probabilities for source and target representations; a smaller gap indicates stronger alignment.
Centroid Euclidean distance (L2)	0.4339	Measures the distance between source and target embedding centroids; a smaller value indicates greater proximity in the representation space.
Centroid cosine similarity	0.9995	Indicates very high directional similarity between source and target embedding centroids; values closer to 1 indicate greater directional similarity.

Table 3.7 shows the mean domain-classification balanced accuracy of 0.7622. Additional diagnostics produced a mean domain-classification (BCE) of 0.5442, mean source-target probability gap of 0.2134, mean centroid Euclidean distance of 0.4339, and mean centroid cosine similarity of 0.9995. These measures provide complementary evidence that adversarial learning influenced source-target representations, although complete domain confusion was not obtained consistently. An important observation is that stronger domain confusion did not automatically result in higher anomaly-detection performance. For example, WADI to BATADAL achieved the strongest domain confusion but did not produce a corresponding F1-score improvement. This suggests that successful domain adaptation requires not only reducing domain-specific information but also preserving anomaly-discriminative structure in the shared representation.

3.7 Decision-Threshold Analysis

A secondary experiment investigated whether selecting the classification threshold from source-validation data could improve target-domain performance. Target validation and test anomaly labels were not used in threshold selection.

Table 3.8: DI-STGAT threshold comparison

Threshold Strategy	Mean Balanced Accuracy	Mean F1-score
Fixed threshold = 0.50	0.5190	0.0437
Source-validation threshold	0.5012	0.0263

The source validation derived threshold reduced mean balanced accuracy from 0.5190 to 0.5012 and F1-score from 0.0437 to 0.0263. Consequently, the fixed 0.50 threshold was retained as the primary operating point as shown in table 3.8.

This result demonstrates that classification calibration itself does not transfer reliably between heterogeneous ICS domains. Even when representations contain transferable ranking information, the probability scale learned from one environment may not correspond directly to the optimal decision boundary in another. This observation helps explain the disparity between relatively strong ROC-AUC values and low fixed-threshold F1-scores observed in several experiments.

3.8 Overall Discussion

The experimental findings demonstrate three main outcomes. (i) The ST-GAT architecture successfully processed heterogeneous ICS environments containing different numbers of process variables and graph structures. The fixed 32-dimensional graph-level representation enabled the three datasets to be processed using the same architecture without requiring identical feature sets, thereby providing a common representation mechanism for heterogeneous ICS domains. (ii) The integration

of DANN and GRL enabled unsupervised target-domain adaptation using labeled source-domain data and unlabeled target-domain data without requiring target-domain anomaly labels during training. DI-STGAT increased mean cross-domain balanced accuracy from 50.51% to 51.90% and mean F1-score from 4.13% to 4.37%, with the strongest improvement occurring for SWaT to WADI, where balanced accuracy increased from 53.04% to 61.37% and F1-score from 7.48% to 8.94%. The ST-GAT versus DI-STGAT comparison in Figure 3.3 demonstrates this improvement. (iii) The domain-representation analysis in Table 3.5 and 3.6 showed that adversarial learning influenced source–target separability, although the degree of alignment varied across transfer directions. These findings demonstrate that DI-STGAT achieved the study's aim of developing a domain-adaptive spatio-temporal framework capable of cross-domain anomaly detection across heterogeneous ICS environments without requiring target-domain anomaly labels for training. However, the transfer-dependent improvements and variations in domain alignment indicate that the framework did not achieve uniform domain invariance or performance improvement across all source–target combinations. Thus, DI-STGAT establishes the feasibility of domain-adversarial cross-domain anomaly detection for heterogeneous ICSs while identifying cross-domain calibration and representation alignment as areas requiring further improvement.

4. CONCLUSION

This study proposed DI-STGAT, a domain-adaptive spatio-temporal graph learning framework for cross-domain anomaly detection in heterogeneous Industrial Control Systems. By integrating ST-GAT with DANN/GRL, the framework provides a mechanism for transferring anomaly-relevant knowledge from a labeled source domain to an unlabeled target domain while accommodating differences in process variables and graph structures. Evaluation across BATADAL, SWaT, and WADI demonstrated the feasibility of this approach, with DI-STGAT providing measurable improvements in cross-domain decision performance over the conventional ST-GAT baseline, particularly for the SWaT→WADI transfer. The findings also showed that the effectiveness of domain adaptation depends on the source–target relationship, indicating that representation alignment alone does not guarantee improved anomaly discrimination. The principal contribution of this work is therefore the establishment of a heterogeneity-aware, domain-adaptive ST-GAT framework that supports cross-domain anomaly detection without requiring target-domain anomaly labels for training. Future work will focus on improving cross-domain probability calibration and graph-level representation learning, strengthening domain alignment while preserving anomaly-discriminative information, and extending the framework to ICS datasets from other industrial sectors to further evaluate its generalizability beyond the water-related environments considered in this study.

REFERENCES

- Ahmadi-Assalemi, G., Al-Khateeb, H., Benson, V., Adamyk, B., & Ammi, M. (2025). Adaptive learning anomaly detection and classification model for cyber and physical threats in industrial control systems. *IET Cyber-Physical Systems: Theory & Applications*, 10(1), e70004. <https://doi.org/10.1049/cps2.70004>
- Ahmed, C. M., Palleti, V. R., & Mathur, A. P. (2017). WADI: A water distribution testbed for research in the design of secure cyber physical systems. *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks*, 25–28. <https://doi.org/10.1145/3055366.3055375>
- Aslam, M. M., Tufail, A., & Irshad, M. N. (2025). Survey of deep learning approaches for securing industrial control systems: A comparative analysis. *Cyber Security and Applications*, 3, 100096. <https://doi.org/10.1016/j.csa.2025.100096>

- Canonico, R., Lista, F., Navarro, A., Sperli, G., & Vignali, A. (2025). Threat detection in reconfigurable cyber–physical systems through spatio-temporal anomaly detection using graph attention network. *Computers & Security*, 156, 104509. <https://doi.org/10.1016/j.cose.2025.104509> (Iris)
- Choi, W.-H., & Kim, J. (2024). Unsupervised learning approach for anomaly detection in industrial control systems. *Applied System Innovation*, 7(2), 18. <https://doi.org/10.3390/asi7020018>
- Deng, A., & Hooi, B. (2021). Graph neural network-based anomaly detection in multivariate time series. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5), 4027–4035. <https://doi.org/10.1609/aaai.v35i5.16523>
- Fan, Y., Fu, T., Listopad, N. I., Liu, P., *et al.* (2024). Utilizing correlation in space and time: Anomaly detection for Industrial Internet of Things (IIoT) via spatiotemporal gated graph attention network. *Alexandria Engineering Journal*, 106, 560–570. <https://doi.org/10.1016/j.aej.2024.08.048>
- Federal Republic of Nigeria. (2021). National cybersecurity policy and strategy 2021. *Office of the National Security Adviser*. National Cybersecurity Policy and Strategy
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59), 1–35.
- Goh, J., Adepu, S., Junejo, K. N., & Mathur, A. (2016). A dataset to support research in the design of secure water treatment systems. In *Critical Information Infrastructures Security* (pp. 88–99). Springer. https://doi.org/10.1007/978-3-319-71368-7_8
- Li, J., Deng, X., & Yao, B. (2025). Enhanced anomaly detection of industrial control systems via graph-driven spatio-temporal adversarial deep support vector data description. *Expert Systems with Applications*, 270, 126573. <https://doi.org/10.1016/j.eswa.2025.126573>
- Lin, L., Gu, A., Min, F., & Zhou, S. (2025). Multi-level graph attention network-based anomaly detection in industrial control system. *Actuators*, 14(5), 210. <https://doi.org/10.3390/act14050210>
- Liu, J., Sha, Y., Zhang, W., Yan, Y., & Liu, X. (2024). Anomaly detection method for industrial control system operation data based on time–frequency fusion feature attention encoding. *Sensors*, 24(18), 6131. <https://doi.org/10.3390/s24186131>
- Mathur, A. P., & Tippenhauer, N. O. (2016). SWaT: A water treatment testbed for research and training on ICS security. In *2016 International Workshop on Cyber-Physical Systems for Smart Water Networks (CySWater)* (pp. 31–36).
- Moriano, P., Hespeler, S. C., Li, M., & Mahbub, M. (2025). Adaptive anomaly detection for identifying attacks in cyber-physical systems: A systematic literature review. *Artificial Intelligence Review*, 58, 283. <https://doi.org/10.1007/s10462-025-11292-w>
- Nahalka, M., Cook, M. M., & Pezaros, D. (2024). The good, the bad and the ugly: Investigating the effectiveness of graph deep neural networks for anomaly detection in industrial control systems. In *IFIP International Internet of Things Conference* (pp. 21–36). Springer. https://doi.org/10.1007/978-3-031-81900-1_2
- Ogunbode, T. O. (2026). The state of SDG 6 in Nigeria (2016–2024): A critical review of progress and future strategies. *Environmental Health Insights*.
- Rehman, M. U., & Bahşi, H. (2024). Process-aware security monitoring in industrial control systems: A systematic review and future directions. *International Journal of Critical Infrastructure Protection*, 47, 100719. <https://doi.org/10.1016/j.ijcip.2024.100719>
- Rotibi, A., Saxena, N., & Burnap, P. (2024). Winning the battle with cyber risk identification tools in industrial control systems: A review. *IET Cyber-Physical Systems: Theory & Applications*, 9(4), 350–365. <https://doi.org/10.1049/cps2.12105>
- United Nations. (2026). Sustainable Development Goal 6: *Clean water and sanitation*, Nigeria. United Nations in Nigeria. UN Nigeria SDG 6

- United Nations. (2026). Sustainable Development Goal 9: *Industry, innovation and infrastructure*, Nigeria. United Nations in Nigeria. UN Nigeria SDG 9
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. *International Conference on Learning Representations*. ([OpenReview](#))
- Wang, Y., Peng, H., Wang, G., Tang, X., Wang, X., & Liu, C. (2023). Monitoring industrial control systems via spatio-temporal graph neural networks. *Engineering Applications of Artificial Intelligence*, 122, 106144. <https://doi.org/10.1016/j.engappai.2023.106144>
- Wu, T., Zhou, D., Ou, Q., & Luo, F. (2026). Intrusion detection systems in industrial control systems: Landscape, challenges and opportunities. *Computers, Materials & Continua*, 86(3), 4. <https://doi.org/10.32604/cmc.2025.073482> (TechScience)
- Xu, L., Wang, B., Zhao, D., & Wu, X. (2025). DAN: Neural network based on dual attention for anomaly detection in ICS. *Expert Systems with Applications*, 263, 125766. <https://doi.org/10.1016/j.eswa.2024.125766>
- Yang, J., Sheng, Y.-Q., Wang, J.-L., & Ni, H. (2024). CAGCN: Centrality-aware graph convolution network for anomaly detection in industrial control systems. *Journal of Computer Science and Technology*, 39(4), 967–983. <https://doi.org/10.1007/s11390-022-2149-y>
- Zhang, D., Wang, M., Bu, Y., Yu, J., & Yang, L. (2025). PdGAT-ID: An intrusion detection method for industrial control systems based on periodic extraction and spatiotemporal graph attention. *Computers & Security*, 149, 104210. <https://doi.org/10.1016/j.cose.2024.104210>