AIMS Energy

https://www.aimspress.com/journal/energy

AIMS Energy, 13(5): 1219–1240.
DOI: 10.3934/energy.2025045
Received: 20 June 2025
Revised: 21 August 2025
Accepted: 09 September 2025
Published: 24 September 2025

Research article

A graph attention network-based framework for dynamic topology change-aware FDIA detection in smart grids

Xing Liu[1], **Weicheng Shen**[1,*]**and Fengyong Li**[1,2]

[1] College of Computer Science and Technology, Shanghai University of Electric Power, Shanghai, 201306, China

[2] Engineering Research Center of Offshore Wind Technology Ministry of Education (Shanghai University of Electric Power), Yangpu District, Shanghai 200090, China

* **Correspondence:** Email: shenwits721@gmail.com.

Abstract: False data injection attacks (FDIAs) pose persistent threats to the security of modern power systems by compromising data integrity through the manipulation of measurements. While data-driven detection models can effectively identify these attacks under stable grid topologies, ensuring accurate results, their performance significantly deteriorates when the grid topology changes due to events like fault restoration, routine maintenance, or power flow redistribution. This lack of adaptability in traditional data-driven methods leads to a substantial decline in detection accuracy in such dynamic environments. Addressing this challenge, this study proposes a dynamic topology change-aware FDIA detection method based on graph attention networks, named the AST-TGT model. The model incorporates two key sub-modules: an attention mechanism module and a temporal convolutional module, to address the spatial interference effects arising from dynamic topological changes. First, a spatio-temporal attention mechanism was introduced to enhance the representational capacity of network nodes by dynamically assigning attention weights, thereby improving the model's understanding of both the topological structure and measurement data. Subsequently, a spatio-temporal convolutional block, composed of temporal convolutional layers and graph attention layers, was proposed to capture the joint temporal and spatial dependencies of the power grid topology and dynamic operational data. Extensive experiments conducted on the IEEE 14-bus and IEEE 118-bus standard test systems demonstrate that, compared to other data-driven detection models, the proposed model effectively improves the accuracy and stability of FDIA detection in the face of grid topology changes and exhibits high robustness in complex interference environments.

Keywords: false data injection attack; smart grid; topology changes; attention mechanism; graph attention networks

1. Introduction

The continuous advancement of information technology has led to the close integration of traditional power systems with sophisticated information control equipment and communication sensor networks. This convergence has given rise to cyber-physical systems (CPSs) in the power grid [1]. These advanced systems can react in real time to various operational conditions and external events, optimizing resource allocation and significantly enhancing system adaptability and resilience to interference. However, this technological progress, while bringing numerous benefits, also introduces new security challenges, particularly in the face of increasingly complex cyberattacks. Among these, false data injection attacks (FDIAs) stand out as a primary security threat to smart grids [2]. Attackers can jeopardize grid security by injecting carefully crafted false data into the power system through information-sensing devices. This allows them to bypass bad data detection (BDD) mechanisms, manipulate measurement results, and ultimately mislead the power system into making incorrect decisions [3].

In response to false data injection attacks (FDIAs), researchers both domestically and internationally have proposed numerous detection methods in recent years. Existing countermeasures can be broadly categorized into two main classes: traditional model-based analytical methods and data-driven machine learning approaches [4]. Model-based analytical methods, which generally do not require historical datasets to train independent systems, rely on establishing system models based on the relationship between measurements and system states. Liu et al. [5] proposed a detection mechanism that separates the nominal grid state from anomalous data. Rashed et al. [6] proposed a model-driven estimation method using the Kalman filter (KF) for FDIA detection in power grids. Furthermore, numerous studies estimate system parameters for FDIA detection using techniques such as kernel density estimation [7], the unscented Kalman filter (UKF) [8], Kullback-Leibler (KL) divergence [9], and maximum likelihood estimation [10]. Chakhchoukh et al. [11] employed a statistical outlier approach to remove FDIA based on a continuous batch-mode regression representation of the extended Kalman filter (EKF). Abbaspour et al. [12] presented an anomaly detection scheme incorporating an observer and the EKF, enhancing the capability of FDIA detection. Kurt et al. [13] utilized DKF in conjunction with blockchain technology to protect network databases and network communication channels from FDIAs. Wang et al. [14] proposed a method combining interval state estimation (ISE) with deep learning to improve the detection accuracy of FDIAs. Yang et al. [15] employed the minimum linear variance gain as the Kalman gain, while the optimal gain of the unknown input observer (UIO) was obtained through pole placement, detecting attack signals by treating them as unknown inputs. However, these model-based detection algorithms are highly susceptible to variations in system parameters, and uncertain factors such as noise can easily affect the detection results.

In contrast to model-based detection algorithms, data-driven artificial intelligence detection algorithms for FDIAs do not rely on complex system models or parameters [16]. Moreover, due to the complexity and stealth of FDIAs, traditional model-based detection methods often struggle to capture their traces, especially in the context of dynamic power system changes and massive data flow. With the sharp increase in data volume and the significant improvement in hardware computing capabilities within cyber-physical power systems, the application of artificial intelligence technologies such as machine learning and deep learning in FDIA detection has a more robust foundation. For example, Ding et al. [17] employed a conditional deep belief network (CDBN) to analyze the temporal dependencies between data in supervisory control and data acquisition (SCADA) systems' time-series measurements. This method not only leverages the powerful feature extraction capabilities of deep learning but also effectively achieves FDIA detection by automatically searching for thresholds based on the prediction differences with the test set data. Furthermore, Wang et al. [18] utilized the advantages of recurrent neural networks (RNNs) in time-series prediction to identify compromised measurements. Lin et al. [19] developed a federated learning framework leveraging edge computing. This framework aims to utilize distributed data from various owners for attack detection. In a similar vein, Shabbir et al. [20] proposed a fusion-enhanced federated learning framework to bolster security against adversarial attacks.

However, these methods treat measurement data as Euclidean data and do not consider the inherent spatial characteristics of the power grid. Therefore, to enhance the accuracy of FDIA detection, some studies have begun to consider the power system's inherent graph topology and the spatial correlations of measurement data, employing graph neural network models to capture the spatio-temporal relationship features between data. Zhang et al. [21] developed an FDIA detection model based on a graph attention network, which introduces an attention mechanism to dynamically adjust the connection weights between network nodes, thereby more accurately capturing and utilizing the spatial features in the power grid. The principles of graph attention have also proven effective in other domains, such as in knowledge graph-based recommendation systems [22]. Recent research has explored using graph neural networks (GNNs) for detecting stealthy FDIAs in smart grids. For instance, Boyaci et al. [23] leveraged GNNs to model FDIAs and detect them using spatial correlations in measurement data. Similarly, Li et al. [24] designed an FDIA detection method based on gated GNNs, extracting spatial features from power grid topology and measurement data to boost accuracy. Additionally, Su et al. [25] introduced an interpretable deep learning-based FDIA detection approach that enhances performance by combining node feature attention and spatial topology attention. Huang et al. [26] proposed a GraphSAGE- and BiLSTM-SE-based model for smart grid FDIA detection. To address insufficient spatial topology capture and poor interpretability, their model integrates shapley additive explanations (SHAP) for spatio-temporal explainability and outperforms state-of-the-art models on IEEE test systems. Li et al. [27] proposed a novel false data injection attack detection method for power grids, utilizing a spatial-temporal transformer network with self-attention and graph convolutional layers to effectively capture the complex spatio-temporal dependencies of power grid topology and data, demonstrating superior accuracy and robustness. Qu et al. [28] proposed a new dummy data injection attack (DDIA) localization method, which leverages temporal and spatial attention matrices with gated stacked causal convolution and graph wavelet sparse convolution to extract spatio-temporal features from power grid data and topology, enabling accurate, robust, and generalizable detection and localization of attacks.

Nevertheless, the aforementioned methods rely on the assumption that the power grid topology and system parameters are invariant. This poses a significant limitation, as power systems are frequently adjusted in actual operation due to maintenance, upgrades, or emergency responses, and these adjustments can lead to changes in the grid's topological structure. When the power grid topology changes, using traditional graph network models that solely consider operational data can result in normal data being misclassified as anomalous, thus reducing detection accuracy. Therefore, we propose a spatio-temporal convolutional block, composed of temporal convolutional layers and graph attention layers, to capture the joint temporal and spatial dependencies of the power grid topology and dynamic operational data. Additionally, a spatio-temporal attention mechanism is introduced to optimize the representational capacity of network nodes by assigning attention weights to the topological structure and measurement data, effectively enhancing the model's understanding of the input data and further improving its robustness and detection accuracy. This approach addresses the issue of reduced detection accuracy when the power grid topology changes. Simulation experiments on the IEEE 14-bus and 118-bus systems demonstrate that this method effectively improves FDIA detection accuracy during power grid topological changes, reduces the impact of anomalous data changes in perturbed nodes, and enhances the robustness of the detection model in complex environments.

Motivated by the aforementioned challenges, we have designed a dynamic topology change-aware FDIA detection method based on graph attention convolutional networks, which primarily makes the following novel contributions:

- We propose a spatio-temporal convolutional block that combines temporal convolutional layers and graph attention layers to capture the joint temporal and spatial dependencies of the power grid topology and dynamic operational data.
- We introduce a spatio-temporal attention mechanism to dynamically assign attention weights, thereby optimizing the representational capacity of network nodes and enhancing the model's understanding of both the topological structure and measurement data. This approach demonstrates superior detection performance compared to other data-driven methods when the power grid topology changes.
- We conduct a comprehensive evaluation of our proposed model against several state-of-the-art FDIA detection models on the IEEE 14-bus and IEEE 118-bus test systems. The results show that our detection model achieves higher accuracy and greater robustness than other detection models in complex environments such as topological changes.

The rest of this paper is organized as follows: Section 2 introduces background information on smart grid state estimation, bad data detection, and false data injection attacks. Section 3 provides a detailed description of the proposed FDIA detection framework. Comprehensive experiments were conducted to evaluate the performance of the proposed scheme. Section 4 presents the experimental results and corresponding discussions. Finally, Section 5 concludes this paper.

2. Related work

2.1. State estimation and bad data detection

Power system state estimation is a crucial technique that utilizes measurement information to estimate the operating state of the power system, which is typically represented by the voltage magnitudes and phase angles of all nodes. State estimation provides the fundamental data for the secure, stable, and economic operation of the power system. However, due to factors like measurement errors and communication failures, the measurement data may contain bad data. These bad data points can severely impact the accuracy of state estimation and even lead to incorrect operational decisions. Therefore, bad data detection and identification are essential components of state estimation.

The measurement model describes the relationship between the measured values and the state variables, and is typically represented as:

$$z = Hx + e \tag{2.1}$$

where z is the measurement vector, x is the state vector, H is the Jacobian matrix representing the network topology, and e is the measurement error vector.

The most commonly used state estimation algorithm is the weighted least squares (WLS) method. Its objective is to find a state vector $\hat{x}$ that minimizes the weighted sum of the squared differences between the measured values and their estimates:

$$\min F(x) = (z - Hx)^T W(z - Hx) \tag{2.2}$$

where W is the weight matrix, typically chosen as the inverse of the measurement error covariance matrix R, i.e., $W = R^{-1}$. In this way, measurements with higher accuracy have a greater weight in the objective function.

By solving the aforementioned minimization problem, the estimated value of the state vector $\hat{x}$ is obtained as:

$$\hat{x} = \left(H^T R^{-1} H\right)^{-1} H^T R^{-1} z \tag{2.3}$$

After obtaining the state estimate $\hat{x}$, bad data detection needs to be performed. A common method for bad data detection is residual testing. The residual reflects the difference between the measured values z and the estimated measured values $\hat{z} = H\hat{x}$, and its Euclidean norm is defined as:

$$r = ||z - \hat{z}||_2 = ||z - H\hat{x}||_2 \tag{2.4}$$

The presence of bad data is determined by comparing the residual r with a predefined threshold τ. If $r > \tau$, it is considered that bad data exists. If $r < \tau$, the data is considered normal.

2.2. *False data injection attack*

False data injection attacks (FDIAs) in power grids are a cybersecurity threat that targets the state estimation process. Attackers manipulate a subset of the measurement data, causing the control center to perform state estimation based on erroneous information. This can lead to incorrect decisions and operations, jeopardizing the secure and stable operation of the grid. The core principle of FDIAs is to inject specifically crafted attack vectors into the measurement data in a way that allows them to evade detection by traditional bad data detection methods, thereby arbitrarily altering the state estimation results.

An FDIA involves superimposing an attack vector a onto the true measurement data z, resulting in the compromised measurement data z':

$$z' = z + a \tag{2.5}$$

The attacker's objective is to design an attack vector a such that the state estimate $\hat{x}'$ computed from the compromised data z' deviates arbitrarily from the true state, while ensuring that the resulting residual r' passes traditional bad data detection tests. If the attack vector a lies within the column space of the Jacobian matrix H, meaning there exists a non-zero vector c such that:

$$a = Hc \tag{2.6}$$

then, the attacked state estimate is obtained as:

$$\hat{x}\prime = \hat{x} + c \tag{2.7}$$

In this scenario, the resulting residual after the attack is:

$$r' = r \tag{2.8}$$

The residual produced by the attack is identical to the normal residual. Therefore, traditional bad data detection methods based on residual magnitude will fail to detect this type of attack.

3. Proposed method

3.1. Spatio-temporal attention network module

The model employs a temporal feature attention mechanism for the input power measurement data and a spatial topology attention mechanism for the graph topology. The temporal feature attention layer captures the significance of each feature at different time points, while the spatial topology attention layer captures the contribution of different sensors within the spatial topology. This dual attention mechanism effectively mitigates the impact of dynamic topological changes.

The temporal feature attention layer is designed to compute feature-wise attention scores, allowing the model to dynamically re-calibrate feature responses by explicitly modeling the importance of each feature channel. We assume the input data consists of node features $X \in \mathbb{R}^{B\times N\times F}$, where B is the batch size, N is the number of nodes, and F is the feature dimension. To create a robust attention mechanism, the input X is passed through three parallel linear transformations to generate three unique score matrices:

$$S_1 = \sigma(XW_1) \quad S_2 = \sigma(XW_2) \quad S_3 = \sigma(XW_3) \tag{3.1}$$

where $W_1, W_2, W_3 \in \mathbb{R}^{F\times F}$ are learnable weight matrices for the three parallel layers, and σ is the sigmoid activation function. The score matrices capture different aspects of feature importance and are then fused by averaging and normalized to produce the final feature attention weights:

$$S_f = \text{softmax}\left(\frac{S_1 + S_2 + S_3}{3}\right) \tag{3.2}$$

where the `softmax` operation is applied along the feature dimension (last dimension) to generate a set of weights that sum to one for each node. Finally, the learned attention weights are applied to the original feature matrix through element-wise multiplication to obtain the re-weighted feature representation:

$$X_{out} = X \odot S_f \tag{3.3}$$

where $\odot$ denotes the Hadamard (element-wise) product. This operation selectively amplifies informative features and suppresses less relevant ones.

The spatial topology attention layer is designed to learn the relative importance of connections between nodes directly from the graph structure. It operates on the adjacency matrix $A \in \mathbb{R}^{N\times N}$ (for a single graph, applied batch-wise) to generate a re-weighted adjacency matrix that highlights more significant spatial dependencies. This is achieved using a multi-head attention mechanism. For each attention head h ($h = 1, \ldots, H$), a unique set of attention coefficients is computed:

$$E^{(h)} = \text{softmax}(AW^{(h)}) \tag{3.4}$$

where $W^{(h)} \in \mathbb{R}^{N \times N}$ is a learnable weight matrix specific to head h. The `softmax` function is applied to each row, normalizing the learned edge weights originating from each node. The resulting attention coefficients are then used to re-weight the original adjacency matrix for each head:

$$A'^{(h)} = A \odot E^{(h)} \tag{3.5}$$

This produces H different weighted adjacency matrices, each representing a distinct learned spatial relationship. The outputs from all attention heads are then concatenated to aggregate the learned information:

$$A_{\text{multi-head}} = \text{concat}(A'^{(1)}, A'^{(2)}, \ldots, A'^{(H)}) \tag{3.6}$$

where the concatenation occurs along the last dimension, resulting in a tensor of shape $\mathbb{R}^{N \times (N \cdot H)}$. Finally, the aggregated representation is passed through a linear transformation to produce the final, refined adjacency matrix:

$$A_{out} = A_{\text{multi-head}} W_{out} \tag{3.7}$$

where $W_{out} \in \mathbb{R}^{(N \cdot H) \times N}$ is a learnable weight matrix that combines the multi-head outputs into a single final matrix of shape $\mathbb{R}^{N \times N}$.

3.2. Temporal convolutional network module

The temporal convolutional module integrates gated temporal convolutional layers and graph attention layers. The gated temporal convolutional layers adjust the convolutional outputs through element-wise multiplication, automatically learning the importance of features at different time steps to dynamically regulate the flow of topological information. This further enhances the model's adaptability to node information disturbances caused by topological changes. A temporal-graph-temporal sandwich structure is employed to achieve cross-modal interaction: the first temporal convolution extracts temporal patterns, the graph attention captures spatial dependencies, and the second temporal convolution fuses spatio-temporal features. This cyclical iterative mechanism enables the model to continuously perceive the effect of topological changes on node features in the topological structure, improving its ability to model dynamic topologies. The input, after adjustment by the attention network layers, is then fed into the temporal convolutional network module.

The temporal gated convolution module is designed to capture temporal dependencies from sequence data. It uses one-dimensional convolutions combined with a gating mechanism, known as a gated linear unit (GLU), to control the flow of information. An additional parallel convolutional path is added to the output of the GLU, acting as a residual-like connection. Assuming the input data is a tensor $X \in \mathbb{R}^{B \times N \times C_{in}}$, where B is the batch size, N is the number of nodes (or time steps), and C_{in} is the input feature dimension.

First, the input tensor is transposed to align with the standard 1D convolution input format of (batch, channels, length), resulting in $X' \in \mathbb{R}^{B \times C_{in} \times N}$. The core of the module is a gated linear unit (GLU), which applies two parallel 1D convolutions to the input. The output of one path is passed through a sigmoid function to act as a gate for the other:

$$X_{GLU} = (X' \circledast W_1) \odot \sigma(X' \circledast W_2) \tag{3.8}$$

where $\circledast$ denotes the 1D convolution operation, W_1 and W_2 are the kernels for the two convolutional layers, σ is the sigmoid activation function, and $\odot$ is the Hadamard (element-wise) product.

In parallel, a third 1D convolution is computed, which serves as a skip connection:

$$X_{skip} = X' \circledast W_3 \tag{3.9}$$

where W_3 is the kernel of the third convolutional layer. The final output within the temporal block is the sum of the GLU output and the skip connection output:

$$X_f = X_{GLU} + X_{skip} \tag{3.10}$$

The resulting tensor X_f has a shape of $\mathbb{R}^{B \times C_{out} \times N}$, where C_{out} is the output feature dimension. Finally, the tensor is transposed back to the original dimension order to produce the module's final output:

$$X_{out} = X_f^T \tag{3.11}$$

The final output X_{out} has a shape of $\mathbb{R}^{B \times N \times C_{out}}$.

3.3. *Detection algorithm overall process*

The measurement data and power grid topology are processed by temporal feature attention and spatial topology attention, respectively. They are then further processed by a sequence of gated temporal convolution, graph attention, and another gated temporal convolution. Finally, a fully connected layer is added to further extract effective features from a wider range of dynamic topologies and to ensure that the final output has the same dimensions as the predicted target. The overall framework of the proposed AST-TGT model is shown in Figure 1, and the specific steps for FDIA detection are outlined in Algorithm 1.

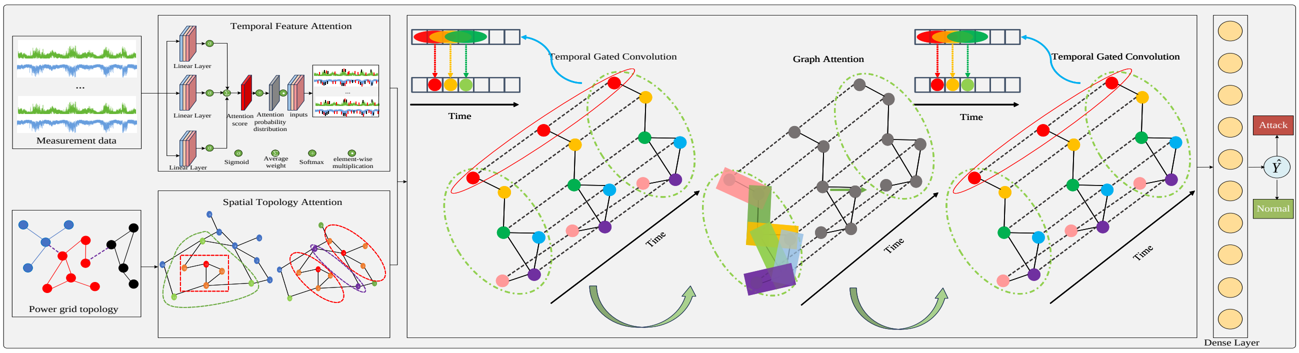


Figure 1. Proposed FDIA detection framework.

Algorithm 1 Proposed FDIA Detection Algorithm

```
Input: $X \in \mathcal{R}^{B\times N\times F}$: input feature matrix, $A \in \mathcal{R}^{N\times N}$: adjacency matrix, $L_r$: training labels, $L_t$: test labels, $E$: number of training epochs, $B$: batch size, $LOF$: loss function
Output: $\hat{Y}$: model detection result
for $i = 1$ to $E$ do
    for $j = 1$ to $B$ do
        $X'_j \leftarrow$ Temporal feature attention($X_j$)
        $A' \leftarrow$ Spatial topology attention($A$)
        $X^t_j \leftarrow$ Temporal gated convolution($X'_j, A'$)
        $X^g_j \leftarrow$ GAT($X^t_j, A'$)
        $X^f_j \leftarrow$ Temporal gated convolution($X^g_j, A'$)
    end for
    $\hat{Y} \leftarrow \sigma$(FC(Flatten($X^f_j$)))
    $loss \leftarrow LOF(\hat{Y}, L_r)$
    Update parameters
end for
for $k = 1$ to $N_t$ do
    $X'_k \leftarrow$ Temporal feature attention($X_k$)
    $A' \leftarrow$ Spatial topology attention($A$)
    $X^t_k \leftarrow$ Temporal gated convolution($X'_k, A'$)
    $X^g_k \leftarrow$ GAT($X^t_k, A'$)
    $X^f_k \leftarrow$ Temporal gated convolution($X^g_k, A'$)
    $\hat{Y} \leftarrow \sigma$(FC(Flatten($X^f_k$)))
    if $\hat{Y} = L_t$ then
        Node Attacked
    else
        Node Normal
    end if
end for
```

4. Experimental results and discussions

4.1. Experimental setup

In this experiment, two classic power system datasets, the IEEE 14-bus system data and the IEEE 118-bus system data, were used to evaluate the model's performance. These datasets utilize load data from different zones of the New York Independent System Operator (NYISO), which has been adjusted and mapped onto the loads of the IEEE 14-bus and IEEE 118-bus systems. Power flow calculations were performed to simulate power grid measurement data, and the Jacobian matrix and grid topology were obtained using MATPOWER, based on which the FDIA dataset was constructed. For each bus system, we generated 15,000 data samples, comprising 7500 normal samples and 7500 attack samples. For ease of model detection, normal data were labeled as 0, and attack data were labeled as 1. Correspondingly, all data samples were divided into training and testing sets, with 75% allocated for training and 25% for testing. Furthermore, to ensure the fairness of the experiments, all experiments utilized a batch size of 32 and were trained for 100 epochs. All computations were conducted on a machine equipped with an AMD Ryzen R9-7945HX CPU, 32GB RAM, and an NVIDIA GeForce RTX 4060 GPU.

The proposed model was designed for offline training followed by online deployment. Initially, the model was trained on a large historical dataset of power system data to learn to detect FDIAs. Once trained, it was deployed to continuously monitor live data streams from the grid. In this online phase, the model operated in a detection-only mode. For sub-synchronous or quasi-real-time systems, the model processed new measurement data as it became available in batches, rather than needing to process every single data point instantaneously. This allowed for slightly delayed but highly accurate predictions about the presence of an attack. To maintain accuracy and adapt to evolving grid conditions over time, the model could be periodically fine-tuned using a strategy of incremental learning. This approach updated the model with new data in small batches, preventing performance degradation from domain drift without requiring a full, time-consuming retraining cycle. This ensured the model remained effective even as the grid's operating patterns changed.

In addition, in order to quantify the impact of topological variations on power system performance, we first modeled the grid's topology as an undirected graph:

$$G = (V, E) \tag{4.1}$$

where V represents the set of nodes (buses) and E represents the set of edges (lines). The graph Laplacian matrix L is defined as:

$$L = D - A \tag{4.2}$$

where A is the adjacency matrix and D is the diagonal degree matrix. The eigenvalues of the Laplacian matrix provide critical insights into the graph's structure and connectivity. To measure the severity of a topological change, such as line disconnections, we propose the topology perturbation score (TPS). This metric quantifies the Euclidean distance between the eigenvalue spectra of the original and modified Laplacian matrices:

$$TPS = \sum_{i=1}^{N} (\lambda_i - \lambda_i')^2 \tag{4.3}$$

where λ_i and λ_i' are the i-th eigenvalues of the original and perturbed Laplacian matrices, respectively, and N is the total number of nodes. Based on this quantitative score, we can classify topology changes into three distinct categories. A weak topology change is characterized by a low TPS ($TPS < T_1$), typically corresponding to a single non-critical line tripping or the failure of a low-load bus. A moderate topology change falls within a moderate TPS range ($T_1 \leq TPS < T_2$), often associated with the loss of a major transmission line or the simultaneous failure of a few buses. Finally, a strong topology change is defined by a high TPS ($TPS \geq T_2$), which usually indicated a wide-area outage or the loss of multiple critical buses, posing a serious threat to system stability and reliability.

4.2. Evaluation metrics

To comprehensively evaluate our model's performance, we utilized four metrics: accuracy, recall, precision, and the F_1-score. True positive (TP) was the number of correctly detected attack samples;

true negative (TN) was the number of normal samples correctly identified; false positive (FP) was the number of normal samples incorrectly identified as attacks; and false negative (FN) was the number of attack samples incorrectly identified as normal. These metrics were defined as follows:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{4.4}$$

$$Pre = \frac{TP}{TP + FP} \tag{4.5}$$

$$Rec = \frac{TP}{TP + FN} \tag{4.6}$$

$$F_1 = \frac{2 \times Pre \times Rec}{Pre + Rec} \tag{4.7}$$

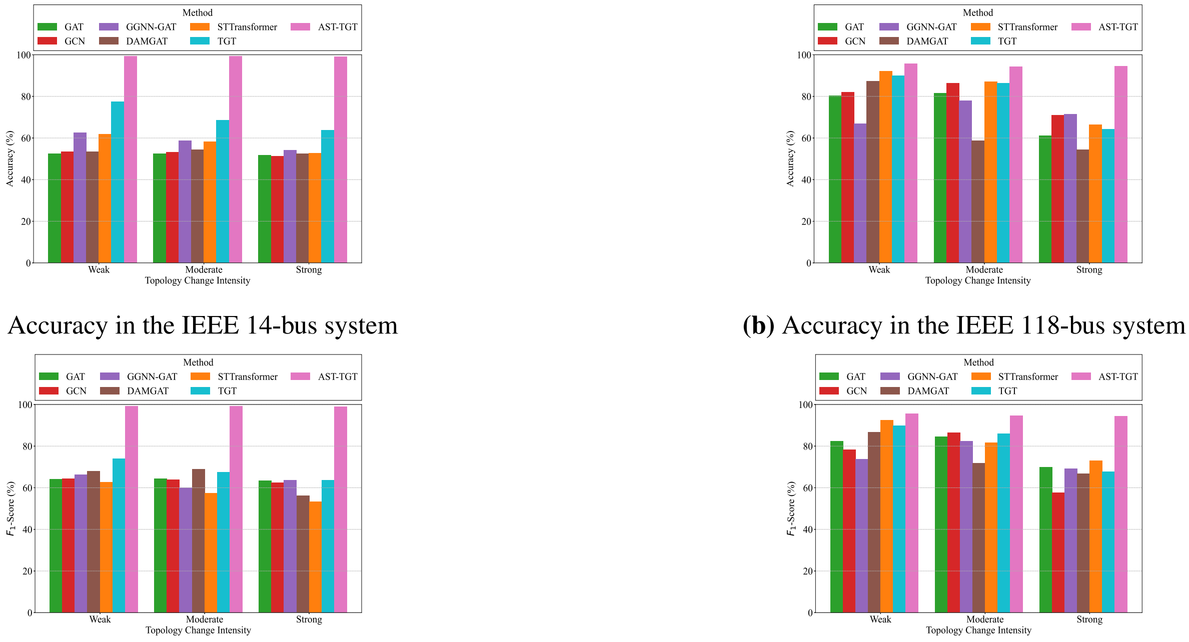
4.3. Performance comparison for static topology change

To confirm our model's detection effectiveness and reliability in static topology environments, we constructed simulation scenarios for both the IEEE 14-bus and IEEE 118-bus systems. Subsequently, we conducted a comparative analysis between our proposed AST-TGT model and six conventional FDIA detection models: GRU, GAT, GCN, GGNN-GAT, DAMGAT, and STTransformer. Additionally, to isolate the impact of the attention mechanism, we performed an ablation study contrasting the full AST-TGT model with a TGT-only variant that omitted this component. As shown in Table 1, the proposed AST-TGT method demonstrated the best detection performance on both datasets compared to other models. Specifically, in the IEEE 14-bus system, our model improved the accuracy by 7.18% and 4.05%, the precision by 7.82% and 4.47%, the recall by 6.48% and 3.64%, and the F_1-score by 7.15% and 4.06% compared to the DAMGAT model and STTransformer, respectively. Similarly, in the more complex and larger IEEE 118-bus system, our model exhibited higher improvements across all four metrics compared to the other models. In addition, the proposed model demonstrated significant improvements in all four indicators compared to its TGT-only variant that lacked the attention mechanism, which demonstrated the effectiveness of the attention mechanism. These experimental results indicated that the proposed AST-TGT model demonstrated superior detection performance on both small-scale and large-scale datasets. We attributed the superior performance of AST-TGT to the synergistic combination of the spatio-temporal attention module and the temporal convolutional module. Within the spatio-temporal attention module, the spatial topology attention layer was designed to capture the contributions arising from the associations between different sensors within the spatial topology by concentrating on processing the spatial topological structure. Simultaneously, the temporal attention layer meticulously analyzed the data in the temporal dimension, enabling the model to automatically focus on significant and effective nodes while disregarding those that had become less important due to topological changes. The input, refined by the spatio-temporal attention module, was then fed into the temporal convolutional module. This module employed a structure where a graph attention layer acted as a bridge between two temporal gated convolutional layers. The initial temporal convolution extracted temporal patterns, followed by the graph attention capturing spatial dependencies, and finally, a secondary temporal convolution fused these spatio-temporal features. This cyclical iterative mechanism allowed the model to continuously perceive the perturbation of node features caused by topological structure changes, thereby enhancing its ability to model dynamic topologies. Furthermore, the temporal convolutional module leveraged the advantages of temporal gated convolutional layers and introduced a dual dynamic regulation mechanism. By applying element-wise multiplication to weight the convolutional outputs, this structure enabled the model to adaptively learn temporal feature weights, automatically strengthening the information propagation of critical time points and weakening noise interference.

Table 1. A comprehensive comparison of eight detection methods—GRU, GAT, GCN, GGNN-GAT, DAMGAT, STTransformer, TGT, and AST-TGT—was conducted in this study. Their performance under a static topology was assessed using four critical metrics: accuracy, precision, recall, and the F_1-score, with all values reported as percentages. The experimental evaluations were performed on two standard power systems: the IEEE 14-bus system and the IEEE 118-bus system.

Methods	IEEE 14-Bus System				IEEE 118-Bus System			
	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
GRU [29]	96.03	95.76	96.33	96.04	92.87	93.81	91.81	92.79
GAT [30]	91.97	92.33	91.53	91.97	89.63	89.56	89.74	89.65
GCN [23]	92.60	92.43	92.80	92.61	87.83	88.80	86.61	87.69
GGNN-GAT [24]	92.73	92.56	92.93	92.74	91.87	92.38	91.27	91.83
DAMGAT [25]	92.47	92.07	92.93	92.50	92.13	91.37	93.07	92.21
STTransformer [27]	95.60	95.42	95.77	95.59	93.84	94.40	93.27	93.83
TGT [32]	96.69	97.24	96.09	96.66	94.93	94.52	95.40	94.96
AST-TGT	**99.65**	**99.89**	**99.41**	**99.65**	**96.50**	**96.16**	**96.87**	**96.52**

4.4. Performance comparison for dynamic topology change



(a) Accuracy in the IEEE 14-bus system

(b) Accuracy in the IEEE 118-bus system

(c) F_1-score in the IEEE 14-bus system

(d) F_1-score in the IEEE 118-bus system

Figure 2. A comparison of the accuracy and F_1-score of seven detection methods—GAT, GCN, GGNN-GAT, DAMGAT, STTransformer, TGT, and AST-TGT—was conducted in this study, considering different topology change intensities.

Table 2. A comprehensive comparison of seven detection methods—GAT, GCN, GGNN-GAT, DAMGAT, STTransformer, TGT, and AST-TGT—was conducted in this study. Their performance under different topology change intensities was assessed using four critical metrics: accuracy, precision, recall, and the F_1-score, with all values reported as percentages. The experimental evaluations were performed on two standard power systems: the IEEE 14-bus system and the IEEE 118-bus system.

Intensity	Method	IEEE 14-Bus System				IEEE 118-Bus System			
		Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
Weak	GAT [30]	52.44	51.35	85.18	64.07	80.36	71.15	97.92	82.42
	GCN [23]	53.51	52.04	84.29	64.35	82.04	90.77	68.81	78.28
	GGNN-GAT [24]	62.67	60.17	73.93	66.35	67.02	58.92	98.68	73.78
	DAMGAT [25]	53.42	51.67	99.64	68.05	87.38	85.25	88.47	86.83
	STTransformer [27]	61.87	70.88	56.11	62.63	92.17	90.06	94.80	92.37
	TGT [32]	77.36	68.78	79.95	73.95	90.03	91.22	88.60	89.90
	AST-TGT	**99.29**	**99.11**	**99.46**	**99.29**	**95.82**	**93.50**	**96.92**	**95.66**
Moderate	GAT [30]	52.43	51.78	85.45	64.48	81.49	75.26	96.62	84.62
	GCN [23]	53.14	52.32	82.16	63.93	86.36	91.02	82.21	86.39
	GGNN-GAT [24]	58.84	58.96	61.03	59.98	77.94	71.22	97.52	82.32
	DAMGAT [25]	54.45	52.60	99.77	68.88	58.84	56.18	99.32	71.77
	STTransformer [27]	58.27	61.28	54.19	57.52	87.17	84.08	79.35	81.65
	TGT [32]	68.56	72.36	63.24	67.61	86.40	88.30	83.94	86.07
	AST-TGT	**99.29**	**99.76**	**98.83**	**99.29**	**94.19**	**93.22**	**95.95**	**94.56**
Strong	GAT [30]	51.82	50.10	85.95	63.30	61.14	54.88	96.63	70.00
	GCN [23]	51.34	49.80	83.33	62.35	71.06	92.54	41.75	57.54
	GGNN-GAT [24]	54.19	51.63	82.68	63.57	71.56	70.53	67.68	69.07
	DAMGAT [25]	52.45	50.66	63.07	56.19	54.34	50.69	98.32	66.90
	STTransformer [27]	52.85	52.64	53.93	53.28	66.33	78.60	68.14	73.00
	TGT [32]	63.73	70.21	58.30	63.70	64.17	59.38	78.81	67.73
	AST-TGT	**99.05**	**99.02**	**99.02**	**99.02**	**94.63**	**93.11**	**95.62**	**94.35**

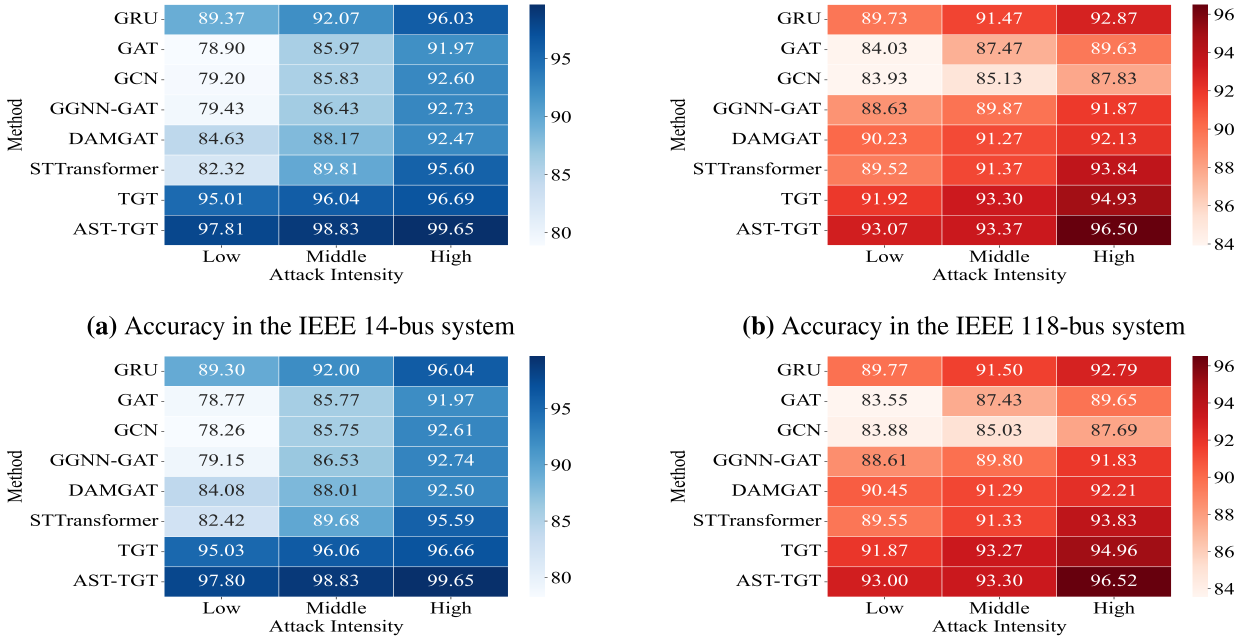
In the actual operation of power systems, the grid topology changes dynamically as a result of unforeseen events such as equipment failures and natural disasters. These dynamic changes pose a major challenge to data-driven detection methods. For a rigorous analysis of the effects of power grid topological variations on detection performance, we categorized these changes—stemming from node-failure-induced line disconnections—into three levels based on the topology perturbation score (TPS), where weak topology changes are defined as those with a TPS less than 1.0 ($TPS < 1.0$), moderate changes are those with a TPS between 1.0 and 3.0 ($1.0 \leq TPS < 3.0$), and strong changes are those with a TPS of 3.0 or greater ($TPS \geq 3.0$). We then compared our proposed AST-TGT model with five mainstream FDIA detection models: GAT, GCN, GGNN-GAT, DAMGAT, and STTransformer. Furthermore, an ablation study was performed to ascertain the efficacy of the attention modules by contrasting the full AST-TGT model with its attention-free TGT-only counterpart. Table 2 and Figure 2 clearly demonstrate the superior performance of the proposed AST-TGT model in FDIA detection under dynamic topology. In the weak topology change of the IEEE 14-bus system, the accuracy of the AST-TGT model reached 99.29%, which was significantly higher than other comparison models, far exceeding the 62.67% of GGNN-GAT and 61.87% of STTransformer. Similarly, in the more complex strong topology change of the IEEE 118-bus system, the accuracy of AST-TGT reached 94.63%, which was also better than other models. Moreover, the proposed model significantly outperformed its TGT-only variant across all four indicators, highlighting the effectiveness of the attention mechanism. These results show that AST-TGT had stronger detection capabilities when dealing with complex attack scenarios. This was because traditional data-driven detection methods often had difficulty effectively perceiving mutations in topology structures, causing the model to mistake normal data for abnormal data, thereby reducing detection performance. The proposed AST-TGT model effectively mitigated the impact of topology changes by assigning node weight values. In addition, the spatio-temporal attention layer performed a detailed analysis of the measurement data and the power grid topology, allowing the model to automatically focus on important and valid nodes while ignoring those that had become irrelevant due to topology changes. A unique temporal-graph-temporal design formed a hierarchical representation for a controlled training process, effectively processing and strengthening dynamic topology change information. This combination of spatio-temporal attention and the temporal-graph-temporal structure enabled the model to continuously perceive the interference of node features caused by topology changes, enhancing its ability to model dynamic topology.

4.5. Robustness test for attack intensity

Table 3. A comprehensive comparison of eight detection methods—GRU, GAT, GCN, GGNN-GAT, DAMGAT, STTransformer, TGT, and AST-TGT—was conducted in this study. Their performance under different attack intensities was assessed using four critical metrics: accuracy, precision, recall, and the F_1-score, with all values reported as percentages. The experimental evaluations were performed on two standard power systems: the IEEE 14-bus system and the IEEE 118-bus system.

Intensity	Method	IEEE 14-Bus System				IEEE 118-Bus System			
		Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
Low	GRU [29]	89.37	89.81	88.79	89.30	89.73	89.53	90.01	89.77
	GAT [30]	78.90	79.22	78.32	78.77	84.03	86.24	81.01	83.55
	GCN [23]	79.20	81.91	74.92	78.26	83.93	84.22	83.54	83.88
	GGNN-GAT [24]	79.43	80.21	78.12	79.15	88.63	88.82	88.41	88.61
	DAMGAT [25]	84.63	87.18	81.19	84.08	90.23	88.57	92.41	90.45
	STTransformer [27]	82.32	81.70	83.15	82.42	89.52	89.14	89.95	89.55
	TGT [32]	95.01	94.50	95.56	95.03	91.92	92.20	91.56	91.87
	AST-TGT	**97.81**	**98.01**	**97.59**	**97.80**	**93.07**	**93.75**	**92.25**	**93.00**
Middle	GRU [29]	92.07	92.69	91.33	92.00	91.47	91.25	91.74	91.50
	GAT [30]	85.97	86.92	84.66	85.77	87.47	87.73	87.14	87.43
	GCN [23]	85.83	86.19	85.32	85.75	85.13	85.67	84.41	85.03
	GGNN-GAT [24]	86.43	85.87	87.19	86.53	89.87	90.47	89.14	89.80
	DAMGAT [25]	88.17	89.12	86.92	88.01	91.27	91.11	91.47	91.29
	STTransformer [27]	89.81	90.37	89.00	89.68	91.37	91.73	90.94	91.33
	TGT [32]	96.04	95.45	96.69	96.06	93.30	93.80	92.74	93.27
	AST-TGT	**98.83**	**98.46**	**99.20**	**98.83**	**93.37**	**94.35**	**92.27**	**93.30**
High	GRU [29]	96.03	95.76	96.33	96.04	92.87	93.81	91.81	92.79
	GAT [30]	91.97	92.33	91.53	91.97	89.63	89.56	89.74	89.65
	GCN [23]	92.60	92.43	92.80	92.61	87.83	88.80	86.61	87.69
	GGNN-GAT [24]	92.73	92.56	92.93	92.74	91.87	92.38	91.27	91.83
	DAMGAT [25]	92.47	92.07	92.93	92.50	92.13	91.37	93.07	92.21
	STTransformer [27]	95.60	95.42	95.77	95.59	93.84	94.40	93.27	93.83
	TGT [32]	96.69	97.24	96.09	96.66	94.93	94.52	95.40	94.96
	AST-TGT	**99.65**	**99.89**	**99.41**	**99.65**	**96.50**	**96.16**	**96.87**	**96.52**

To validate the proposed model's superior detection performance under varying attack intensities, we categorized the generated FDIA samples into three levels: low, medium, and high intensity. Specifically, low attack intensity referred to scenarios with an average injected power deviation of less than 10% of the actual measured values, medium attack intensity corresponded to deviations between 10% and 30%, and high attack intensity involved deviations exceeding 30% [31]. Then, we compared our proposed AST-TGT model with six mainstream FDIA detection models: GRU, GAT, GCN, GGNN-GAT, DAMGAT, and STTransformer. Additionally, we conducted an ablation study comparing the full AST-TGT model with its TGT-only variant to verify the contribution of the attention mechanism. As clearly depicted in Table 3 and Figure 3, the proposed AST-TGT model consistently outperformed other models across all attack intensities in both bus systems. Specifically, under low-intensity attacks on the IEEE 14-bus system, AST-TGT outperformed DAMGAT and STTransformer, respectively, with improvements of 13.18% and 15.49% in accuracy, 10.83% and 16.31% in precision, 16.4% and 14.44% in recall, and 13.72% and 15.38% in F_1-score. Similarly, comparable detection results were observed in the IEEE 118-bus system. Additionally, the superior performance of the proposed model across all four indicators, when compared to the TGT-only variant, was attributed to the inclusion of the attention mechanism. Indeed, this phenomenon was readily explained. The spatio-temporal attention network module in our proposed model captured the importance of each node at different time points and the contribution of different nodes within the spatial topology. Furthermore, the temporal convolutional network module extracted features from the data adjusted by the spatio-temporal attention module, enabling the identification of subtle anomalies that traditional detection methods often fail to recognize.



(a) Accuracy in the IEEE 14-bus system

(b) Accuracy in the IEEE 118-bus system

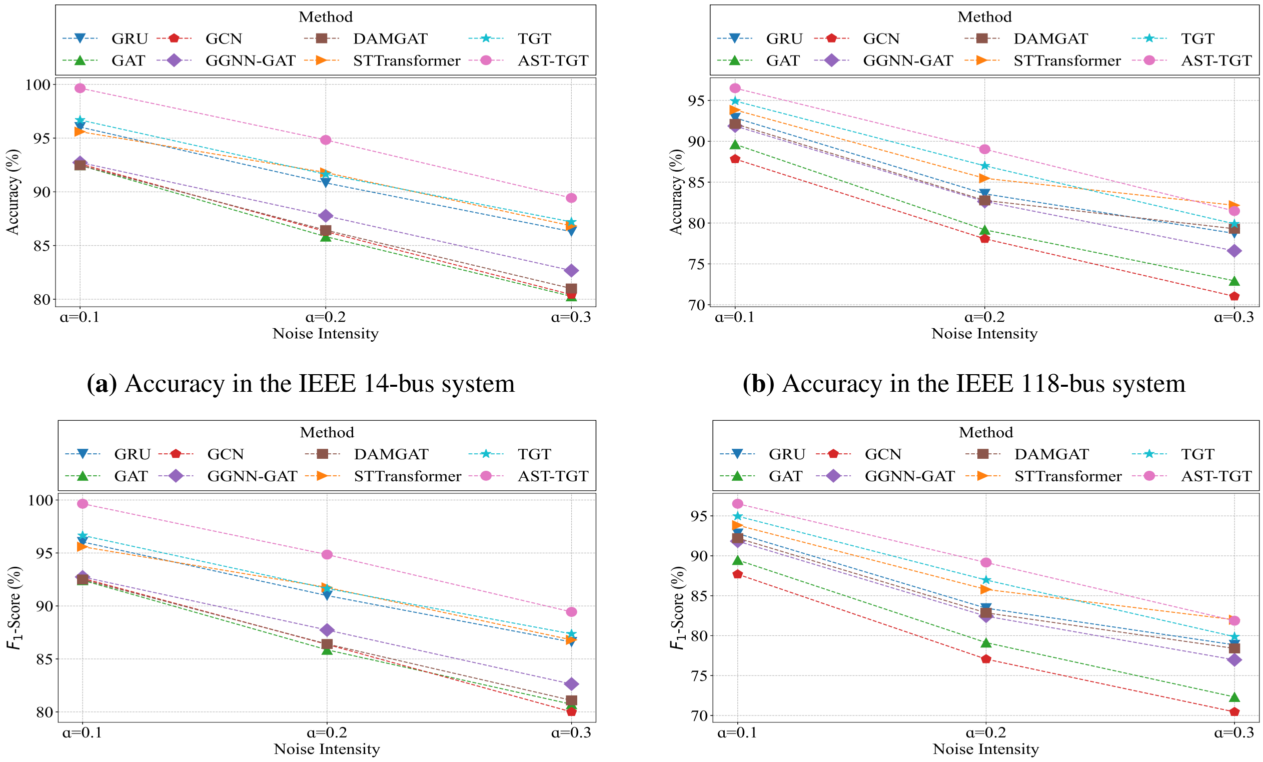
(c) F_1-score in the IEEE 14-bus system

(d) F_1-score in the IEEE 118-bus system

Figure 3. A comparison of the accuracy and F_1-score of eight detection methods—GRU, GAT, GCN, GGNN-GAT, DAMGAT, STTransformer, TGT, and AST-TGT—was conducted in this study, considering different attack intensities.

4.6. *Robustness test for noise intensity*

To further validate the proposed model's anti-interference capability and robustness in complex noisy environments, we employed the additive white Gaussian noise (AWGN) model to add noise to the data samples. Based on the noise intensity, the generated FDIA samples were categorized into three levels with noise variances of $\alpha = 0.1$, $\alpha = 0.2$, and $\alpha = 0.3$. The proposed AST-TGT model was then compared against six prominent FDIA detection models: GRU, GAT, GCN, GGNN-GAT, DAMGAT, and STTransformer. Additionally, we conducted an ablation study comparing the full AST-TGT model with its TGT-only variant to verify the contribution of the attention mechanism. As clearly observed in Table 4 and Figure 4, the proposed AST-TGT model significantly outperformed other methods across all noise levels and maintained a leading position in all four detection metrics. Specifically, in the IEEE 14-bus system, when the noise variance was $\alpha = 0.3$, the proposed AST-TGT model achieved an accuracy of 89.43%, precision of 89.35%, recall of 89.53%, and F_1-score of 89.44%, while the other models were generally below 85%. Similar results were observed in the IEEE 118-bus system. Moreover, the proposed model significantly outperformed its TGT-only variant across all four indicators and highlighted the effectiveness of the attention mechanism. Furthermore, as the noise level increased, the overall detection performance of all schemes declined. This decline was attributed to the fact that higher noise levels can obscure the features of attack data, which makes it challenging for detection models to distinguish between attack and normal data. However, under the same noise conditions, the proposed scheme exhibited the smallest performance degradation. This robustness was attributed to the model's attention mechanism, which could dynamically adjust the focus on different parts of the input data, and the temporal convolutional module, which combined the advantages of temporal gated convolutional layers and introduced a dual dynamic regulation mechanism. By weighting the convolutional outputs through element-wise multiplication, this enabled the model to adaptively learn temporal feature weights and automatically strengthen the information propagation of critical time points, which effectively reduced the interference of noise on the decision-making process.



(a) Accuracy in the IEEE 14-bus system

(b) Accuracy in the IEEE 118-bus system

(c) F_1-score in the IEEE 14-bus system

(d) F_1-score in the IEEE 118-bus system

Figure 4. A comparison of the accuracy and F_1-score of eight detection methods—GRU, GAT, GCN, GGNN-GAT, DAMGAT, STTransformer, TGT, and AST-TGT—was conducted in this study, considering different noise intensities.

Table 4. A comprehensive comparison of eight detection methods—GRU, GAT, GCN, GGNN-GAT, DAMGAT, STTransformer, TGT, and AST-TGT—was conducted in this study. Their performance under different noise intensities was assessed using four critical metrics: accuracy, precision, recall, and the F_1-score, with all values reported as percentages. The experimental evaluations were performed on two standard power systems: the IEEE 14-bus system and the IEEE 118-bus system.

Intensity	Method	IEEE 14-Bus System				IEEE 118-Bus System			
		Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
$\alpha = 0.1$	GRU [29]	96.03	95.76	96.33	96.04	92.87	93.81	91.81	92.79
	GAT [30]	92.47	92.86	91.99	92.43	89.63	90.64	88.34	89.47
	GCN [23]	92.60	92.43	92.80	92.61	87.83	88.80	86.61	87.69
	GGNN-GAT [24]	92.73	92.56	92.93	92.74	91.87	92.38	91.27	91.83
	DAMGAT [25]	92.47	92.07	92.93	92.50	92.13	91.37	93.07	92.21
	STTransformer [27]	95.60	95.42	95.77	95.59	93.84	94.40	93.27	93.83
	TGT [32]	96.69	97.24	96.09	96.66	94.93	94.52	95.40	94.96
	AST-TGT	**99.65**	**99.89**	**99.41**	**99.65**	**96.50**	**96.16**	**96.87**	**96.52**
$\alpha = 0.2$	GRU [29]	90.83	89.48	92.53	90.98	83.57	84.15	82.74	83.44
	GAT [30]	85.83	85.70	85.99	85.86	79.17	79.32	78.95	79.13
	GCN [23]	86.30	85.88	86.86	86.37	78.07	80.83	73.62	77.06
	GGNN-GAT [24]	87.77	87.94	87.53	87.73	82.63	83.42	81.48	82.44
	DAMGAT [25]	86.43	86.50	86.32	86.41	82.77	82.45	83.28	82.86
	STTransformer [27]	91.80	92.25	91.26	91.75	85.47	83.98	87.67	85.79
	TGT [32]	91.63	91.43	91.86	91.65	87.00	87.21	86.74	86.97
	AST-TGT	**94.83**	**94.50**	**95.20**	**94.85**	**89.03**	**88.20**	**90.14**	**89.16**
$\alpha = 0.3$	GRU [29]	86.30	84.69	88.59	86.60	78.73	78.48	79.21	78.85
	GAT [30]	80.27	78.87	82.66	80.72	72.93	74.04	70.69	72.32
	GCN [23]	80.43	81.75	78.32	80.00	71.03	71.94	69.02	70.45
	GGNN-GAT [24]	82.67	82.74	82.52	82.63	76.60	75.82	78.15	76.97
	DAMGAT [25]	81.00	80.66	81.52	81.09	79.30	82.02	75.08	78.40
	STTransformer [27]	86.83	86.95	86.66	86.80	82.17	82.86	81.15	81.99
	TGT [32]	87.20	86.18	88.59	87.37	79.90	80.05	79.68	79.87
	AST-TGT	**89.43**	**89.35**	**89.53**	**89.44**	**81.47**	**80.19**	**83.61**	**81.87**

4.7. Comparison for a different number of attack nodes

Table 5. A comprehensive comparison of eight detection methods—GRU, GAT, GCN, GGNN-GAT, DAMGAT, STTransformer, TGT, and AST-TGT—was conducted in this study. Their performance under single-node and multi-node attacks was assessed using four critical metrics: accuracy, precision, recall, and the F_1-score, with all values reported as percentages. The experimental evaluations were performed on two standard power systems: the IEEE 14-bus system and the IEEE 118-bus system.

Attack Nodes	Method	IEEE 14-Bus System				IEEE 118-Bus System			
		Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
Single-Node	GRU [29]	93.73	93.38	94.13	93.75	92.33	92.45	92.21	92.33
	GAT [30]	87.10	87.07	87.12	87.10	87.67	87.57	87.81	87.69
	GCN [23]	85.70	85.06	86.59	85.82	87.03	87.62	86.28	86.94
	GGNN-GAT [24]	87.37	86.22	88.93	87.55	91.57	91.66	91.47	91.56
	DAMGAT [25]	87.67	86.49	89.26	87.85	90.40	89.98	90.94	90.46
	STTransformer [27]	89.12	87.79	90.80	89.27	93.20	93.26	93.14	93.20
	TGT [32]	96.69	97.24	96.09	96.66	93.37	94.12	93.54	93.82
	AST-TGT	**99.57**	**99.73**	**99.41**	**99.57**	**94.40**	**95.06**	**93.67**	**94.36**
Multi-Node	GRU [29]	96.03	95.76	96.33	96.04	92.87	93.81	91.81	92.79
	GAT [30]	92.47	92.86	91.99	92.43	89.63	89.56	89.74	89.65
	GCN [23]	92.60	92.43	92.80	92.61	87.83	88.80	86.61	87.69
	GGNN-GAT [24]	92.73	92.56	92.93	92.74	91.87	92.38	91.27	91.83
	DAMGAT [25]	92.47	92.07	92.93	92.50	92.13	91.37	93.07	92.21
	STTransformer [27]	95.60	95.42	95.77	95.59	93.84	94.40	93.27	93.83
	TGT [32]	96.00	96.64	95.29	95.96	94.93	94.52	95.40	94.96
	AST-TGT	**99.65**	**99.89**	**99.41**	**99.65**	**96.50**	**96.16**	**96.87**	**96.52**

In real-world power grid systems, attackers often have limited resources, meaning they might only possess local grid topology information and might consider attacking the fewest possible nodes to achieve their objectives. Conversely, if attackers possessed comprehensive grid topology information, exhibiting high stealth and destructive capabilities, they might target multiple nodes. To validate the detection performance of the proposed model under varying numbers of attack nodes, we conducted experiments involving both single-node and multi-node attacks. As shown in Table 5, the detection efficiency of all models on the multi-node attack dataset was superior to that of the single-node attack scenario in both the IEEE 14 and IEEE 118 bus systems. This was because multi-node attacks led to more pronounced data perturbations in the attacked nodes, which made them easier for detection models to identify. In both cases, the proposed AST-TGT model outperformed other models as well as the TGT model without attention.

4.8. Complexity analysis

To demonstrate the superior performance of our proposed solution, we conducted a comprehensive evaluation of its computational complexity. Although our AST-TGT model incurred increased computational overhead due to its complex multi-head attention and multi-layer convolutional structure—as shown in Table 6, its average training time on the IEEE 14-bus and 118-bus systems was indeed longer than methods like GRU, GAT, GCN, STTransformer, GGNN-GAT, and the attention-free TGT—this design achieved detection performance that far surpassed its rivals. We argue that this modest trade-off in time complexity is entirely justified by the significant leap in performance. To ensure fairness and stability, all complexity comparison experiments were conducted under identical environmental conditions, with results being averaged over six runs.

Table 6. A comparison of the computational complexity of eight detection methods—GRU, GAT, GCN, GGNN-GAT, DAMGAT, STTransformer, TGT, and AST-TGT—was conducted in this study. Their performance was assessed based on three key metrics: training time, testing time, and the number of parameters. These experimental evaluations were performed on two standard power systems: the IEEE 14-bus system and the IEEE 118-bus system.

Method	IEEE 14-Bus System			IEEE 118-Bus System		
	Train(s)	Test(s)	Parameter(K)	Train(s)	Test(s)	Parameter(K)
GRU [29]	22.03	0.03	0.54	25.53	0.04	2.41
GAT [30]	28.80	0.15	0.29	29.67	0.15	2.17
GCN [23]	23.96	0.11	0.28	25.47	0.11	2.15
GGNN-GAT [24]	31.89	0.16	0.81	32.03	0.16	2.47
DAMGAT [25]	56.38	0.17	0.93	82.54	0.19	43.25
STTransformer [27]	67.18	0.13	3.95	66.79	0.15	5.62
TGT [32]	54.31	0.26	1.17	59.60	0.27	3.45
AST-TGT	71.02	0.24	1.60	72.41	0.25	31.56

5. Conclusions

This paper proposed a false data injection attack (FDIA) detection method based on graph attention convolutional networks for dynamic topological changes. It introduced a spatio-temporal attention mechanism to optimize the representational capacity of network nodes by dynamically assigning attention weights, thereby enhancing the model's understanding of both the topological structure and measurement data. Furthermore, it established a temporal-graph-temporal (TGT) module, which combines temporal gated convolutional layers and graph attention layers to achieve cross-modal information interaction through a temporal-graph-temporal sandwich structure. This cyclical iterative mechanism enabled the model to continuously perceive the perturbation of node features due to changes in the topological structure, which improved its ability to model dynamic topologies. Simulation verifications conducted on the IEEE 14-bus and IEEE 118-bus systems demonstrated that the AST-TGT model exhibited superior accuracy, precision, recall, and F_1-score compared to current mainstream detection models, and possessed better FDIA detection capabilities in dynamic topological environments. Additionally, ablation experiments showed that our attention mechanism was effective and the proposed method demonstrated good anti-interference capability and robustness in complex noisy environments. While our approach demonstrated robust FDIA detection performance in the face of dynamic power grid topologies, our model currently lacked inherent interpretability. This led to opacity in the model's decision-making process, potentially reducing trust among users and operators. Within the context of robust FDIA detection, this remained an open challenge, as repeatedly evidenced in existing research. The future work aimed to explore the interpretability of the proposed model.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research is supported by grants from the Engineering Research Center of Offshore Wind Technology Ministry of Education (Shanghai University of Electric Power).

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

Conceptualization, Xing Liu; methodology, Xing Liu and Weicheng Shen; validation, Xing Liu; formal analysis, Weicheng Shen and Fengyong Li; writing—original draft preparation, Xing Liu; writing—review and editing, Weicheng Shen and Fengyong Li; supervision, Fengyong Li. All authors have read and agreed to the published version of the manuscript.

References

1. Li F, Shen W, Bi Z, et al. (2024) Sparse Adversarial Learning for FDIA Attack Sample Generation in Distributed Smart Grids. *Comput Model Eng Sci (CMES)* 139: 2. https://doi.org/10.32604/cmes.2023.044431

2. Li K, Li F, Wang B, et al. (2024) False data injection attack sample generation using an adversarial attention-diffusion model in smart grids. *AIMS Energy* 12: 1271–1293. https://doi.org/10.3934/energy.2024058

3. Han Y, Feng H, Li K, et al. (2023) False data injection attacks detection with modified temporal multi-graph convolutional network in smart grids. *Comput Secur* 124: 103016. https://doi.org/10.1016/j.cose.2022.103016

4. Liang G, Zhao J, Luo F, et al. (2016) A review of false data injection attacks against modern power systems. *IEEE Trans Smart Grid* 8: 1630–1638. https://doi.org/10.1109/TSG.2015.2495133

5. Liu L, Esmalifalak M, Ding Q, et al. (2014) Detecting false data injection attacks on power grid by sparse optimization. *IEEE Trans Smart Grid* 5: 612–621.

6. Rashed M, Kamruzzaman J, Gondal I, et al. (2022) False data detection in a clustered smart grid using unscented Kalman filter. *IEEE Access* 10: 78548–78556. https://doi.org/10.1109/ACCESS.2022.3193781

7. Chen Y, Huang S, Liu F, et al. (2018) Evaluation of reinforcement learning-based false data injection attack to automatic voltage control. *IEEE Trans Smart Grid* 10: 2158–2169. https://doi.org/10.1109/TSG.2018.2790704

8. Živković N, Sarić AT (2018) Detection of false data injection attacks using unscented Kalman filter. *J Mod Power Syst Clean Energy* 6: 847–859. https://doi.org/10.1007/s40565-018-0413-5

9. Singh SK, Khanna K, Bose R, et al. (2017) Joint-transformation-based detection of false data injection attacks in smart grid. *IEEE Trans Ind Inf* 14: 89–97. https://doi.org/10.1109/TII.2017.2720726

10. Moslemi R, Mesbahi A, Velni JM (2017) A fast, decentralized covariance selection-based approach to detect cyber attacks in smart grids. *IEEE Trans Smart Grid* 9: 4930–4941. https://doi.org/10.1109/TSG.2017.2675960

11. Chakhchoukh Y, Lei H, Johnson BK (2019) Diagnosis of outliers and cyber attacks in dynamic PMU-based power state estimation. *IEEE Trans Power Syst* 35: 1188–1197. https://doi.org/10.1109/TPWRS.2019.2939192

12. Abbaspour A, Sargolzaei A, Forouzannezhad P, et al. (2019) Resilient control design for load frequency control system under false data injection attacks. *IEEE Trans Ind Electron* 67: 7951–7962. https://doi.org/10.1109/TIE.2019.2944091

13. Kurt MN, Yılmaz Y, Wang X (2019) Secure distributed dynamic state estimation in wide-area smart grids. *IEEE Trans Inf Forensics Secur* 15: 800–815. https://doi.org/10.1109/TIFS.2019.2928207

14. Wang H, Ruan J, Ma Z, et al. (2019) Deep learning aided interval state prediction for improving cyber security in energy internet. *Energy* 174: 1292–1304. https://doi.org/10.1016/j.energy.2019.03.009

15. Yang T, Murguia C, Kuijper M, et al. (2020) An unknown input multiobserver approach for estimation and control under adversarial attacks. *IEEE Trans Control Network Syst* 8: 475–486. https://doi.org/10.1109/TCNS.2020.3029160

16. Ding Y, Ma K, Pu T, et al. (2021) A deep learning-based classification scheme for cyber-attack detection in power system. *IET Energy Syst Integr* 3: 274–284. https://doi.org/10.1049/esi2.12034

17. Ding Y, Ma K, Pu T, et al. (2021) A deep learning-based classification scheme for false data injection attack detection in power system. *Electronics* 10: 1459. https://doi.org/10.3390/electronics10121459

18. Wang Y, Zhang Z, Ma J, et al. (2021) KFRNN: An effective false data injection attack detection in smart grid based on Kalman filter and recurrent neural network. *IEEE Int Things J* 9: 6893–6904. https://doi.org/10.1109/JIOT.2021.3113900

19. Lin WT, Chen G, Huang Y (2022) Incentive edge-based federated learning for false data injection attack detection on power grid state estimation: A novel mechanism design approach. *Appl Energy* 314: 118828. https://doi.org/10.1016/j.apenergy.2022.118828

20. Shabbir A, Manzoor HU, Zoha A, et al. (2025) Smart grid security through fusion-enhanced federated learning against adversarial attacks. *Eng Appl Artif Intell* 157: 111169. https://doi.org/10.1016/j.engappai.2025.111169

21. Zhang W, Deng C, Su X, et al. (2022) Spatial-temporal attention based interpretable deep framework for FDIA detection in smart grid. *2022 IEEE Sustainable Power and Energy Conference (iSPEC)* https://doi.org/10.1109/iSPEC54162.2022.10032978

22. Elahi E, Anwar S, Al-kfairy M, et al. (2025) Graph attention-based neural collaborative filtering for item-specific recommendation system using knowledge graph. *Expert Syst Appl* 266: 126133. https://doi.org/10.1016/j.eswa.2024.126133

23. Boyaci O, Narimani MR, Davis KR, et al. (2021) Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks. *IEEE Trans Smart Grid* 13: 807–819. https://doi.org/10.1109/TSG.2021.3117977

24. Li X, Wang Y, Lu Z (2023) Graph-based detection for false data injection attacks in power grid. *Energy* 263: 125865. https://doi.org/10.1016/j.energy.2022.125865

25. Su X, Deng C, Yang J, et al. (2024) DAMGAT-Based interpretable detection of false data injection attacks in smart grids. *IEEE Trans Smart Grid* 15: 4182–4195. https://doi.org/10.1109/TSG.2024.3364665

26. Huang S, Li F, Li K, et al. (2025) Dependency-Aware GraphSAGE based interpretable FDIA detection using BiLSTM with SE-Attention in smart grids. *IEEE Int Things J* https://doi.org/10.1109/JIOT.2025.3578369

27. Li X, Hu L, Lu Z (2024) Detection of false data injection attack in power grid based on spatial-temporal transformer network. *Expert Syst Appl* 238: 121706. https://doi.org/10.1016/j.eswa.2023.121706

28. Qu Z, Dong Y, Li Y, et al. (2024) Localization of dummy data injection attacks in power systems considering incomplete topological information: A spatio-temporal graph wavelet convolutional neural network approach. *Appl Energy* 360: 122736. https://doi.org/10.1016/j.apenergy.2024.122736

29. Wang J, Si Y, Zhu Y, et al. (2024) Cyberattack detection for electricity theft in smart grids via stacking ensemble GRU optimization algorithm using federated learning framework. *Int J Electr Power Energy Syst* 157: 109848. https://doi.org/10.1016/j.ijepes.2024.109848

30. Liao W, Zhu R, Yang Z, et al. (2023) Electricity theft detection using dynamic graph construction and graph attention network. *IEEE Trans Ind Inf* 20: 5074–5086. https://doi.org/10.1109/TII.2023.3331131

31. Wu Y, Zu T, Guo N, et al. (2023) Laplace-domain hybrid distribution model based FDIA attack sample generation in smart grids. *Symmetry* 15: 1669. https://doi.org/10.3390/sym15091669

32. Zheng Q, Zheng J, Mei F, et al. (2023) TCN-GAT multivariate load forecasting model based on SHAP value selection strategy in integrated energy system. *Front Energy Res* 11: 1208502. https://doi.org/10.3389/fenrg.2023.1208502