Computer Modeling in
Engineering & Sciences

Tech Science Press

DOI: 10.32604/cmes.2024.055442

ARTICLE

Self-Attention Spatio-Temporal Deep Collaborative Network for Robust FDIA Detection in Smart Grids

Tong Zu and Fengyong Li*

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

*Corresponding Author: Fengyong Li. Email: fyli@shiep.edu.cn

Received: 27 June 2024 Accepted: 22 August 2024 Published: 27 September 2024

ABSTRACT

False data injection attack (FDIA) can affect the state estimation of the power grid by tampering with the measured value of the power grid data, and then destroying the stable operation of the smart grid. Existing work usually trains a detection model by fusing the data-driven features from diverse power data streams. Data-driven features, however, cannot effectively capture the differences between noisy data and attack samples. As a result, slight noise disturbances in the power grid may cause a large number of false detections for FDIA attacks. To address this problem, this paper designs a deep collaborative self-attention network to achieve robust FDIA detection, in which the spatio-temporal features of cascaded FDIA attacks are fully integrated. Firstly, a high-order Chebyshev polynomials-based graph convolution module is designed to effectively aggregate the spatio information between grid nodes, and the spatial self-attention mechanism is involved to dynamically assign attention weights to each node, which guides the network to pay more attention to the node information that is conducive to FDIA detection. Furthermore, the bi-directional Long Short-Term Memory (LSTM) network is introduced to conduct time series modeling and long-term dependence analysis for power grid data and utilizes the temporal self-attention mechanism to describe the time correlation of data and assign different weights to different time steps. Our designed deep collaborative network can effectively mine subtle perturbations from spatiotemporal feature information, efficiently distinguish power grid noise from FDIA attacks, and adapt to diverse attack intensities. Extensive experiments demonstrate that our method can obtain an efficient detection performance over actual load data from New York Independent System Operator (NYISO) in IEEE 14, IEEE 39, and IEEE 118 bus systems, and outperforms state-of-the-art FDIA detection schemes in terms of detection accuracy and robustness.

KEYWORDS

False data injection attacks; smart grid; deep learning; self-attention mechanism; spatio-temporal fusion

1 Introduction

With the advancement of smart sensor and wireless communication technologies, traditional power systems are gradually transitioning to intelligent grid cyber-physical systems to enhance energy utilization efficiency and system stability [1,2]. The advantages of smart grids are becoming increasingly apparent, as the cyber-physical systems in smart grids offer end-to-end bi-directional power flow, which allows the users to feedback energy into the grid, thus improving system scalability, efficiency, and stability [3,4]. However, the openness and diversity of cyber-physical systems make the smart grid more susceptible to diverse malicious network attacks [5], especially at the boundary of interaction between the public Internet and the power system private network. For example, denial of service attack [6], man-in-the-middle attack [7], network topology attack [8], channel measurement attack [9] and false data injection attack [10].

False Data Injection Attacks (FDIAs) have garnered increasing attention due to their destructive and covert nature [11]. Generally, in the cyber-physical systems of smart grids, crucial operations such as emergency analysis, Volt-VAR optimization, real-time pricing, etc., are always conducted using power system state estimators. Sensor measurements are utilized as inputs to these estimators, which then generate corresponding outputs including a range of voltage magnitudes and phase angles. Control signals are dispatched by Supervisory Control and Data Acquisition (SCADA) systems based on the output of state estimation, which is crucial for the efficient operation of smart grids [12]. Although the cyber-physical system can mitigate the information interference of the public Internet by deploying Bad Data Detection (BDD) algorithms in the state estimator, FDIAs can circumvent these checks by injecting carefully crafted attack vectors, thereby tampering with measurement data from SCADA systems and impacting the accuracy of state estimation. This could lead to severe consequences for grid computation, scheduling operations, and the overall stability of the system [13]. Fig. 1 presents a complete schematic diagram of an FDIA attack on a cyber-physical system.

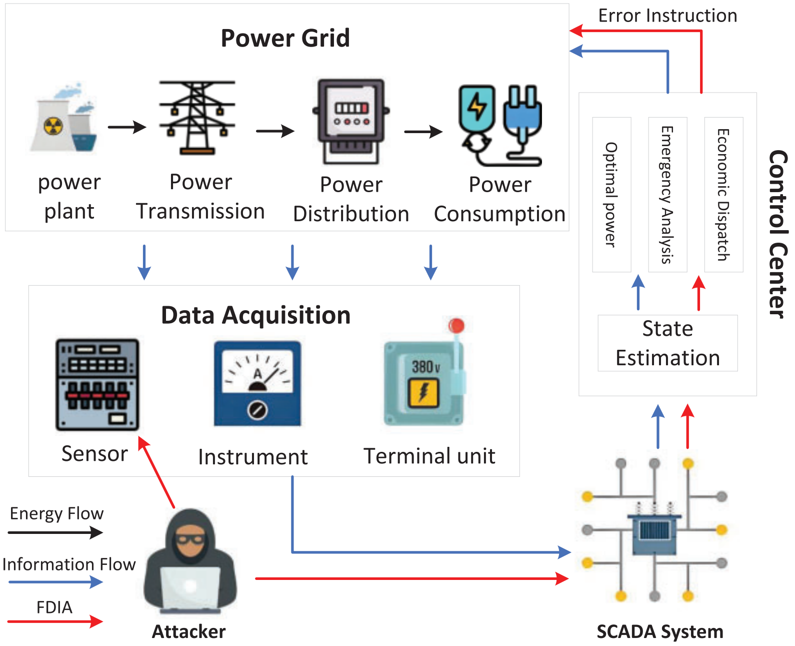


Figure 1: Typical scenarios of FDIA in Smart Grid

In general, FDIA attack detection techniques mainly contain traditional machine learning model-based methods and data-driven methods [14]. Model-based detection methods rely on system modeling to compare with expected behavior to detect attacks. For example, Wei et al. [15] proposed a method using Forecast-Aided State Estimation (FASE) and Square Root Unscented Kalman Filter (SR-UKF) for FDIA detection, which involved Generalized Likelihood Ratio Test (GLRT). Qu et al. [16] suggested an FDIA detection approach that utilizes the Hellinger distance to track measurement value changes and determine attack presence. Shen et al. [17] introduced a detection method based on random matrix theory, utilizing random variables from load short-term forecasts to construct random matrices for FDIA detection.

Data-driven detection methods usually leverage large amounts of historical data and employ statistical analysis or machine learning techniques to find unusual patterns to detect potential attack behavior. These methods require no prior knowledge and can adapt to various complex attack scenarios. For instance, James et al. [18] utilized discrete wavelet transform (DWT) and deep neural networks (DNN) to analyze system states continuously over time, effectively capturing FDIAs. Habibi et al. [19] proposed a neural network based on time series analysis and the Nonlinear Autoregressive Exogenous model (NARX) for detecting network attacks using estimated errors. Lu et al. [20] introduced a convolutional neural network called representation learning CNN (RL-CNN) to capture local data features, exhibiting superior performance in locating network attacks as a multi-label classifier. With the gradual increase of time series data in the smart grid, the detection effect of FDIAs can be improved more effectively by using the measurement of time series dependence in time series data. For example, Wang et al. [21] introduced a two-level learner-based scheme for detecting FDIAs, integrating linear and nonlinear time series data from the power grid and employing a combination of Kalman filter and Recurrent Neural Network (KFRNN). To effectively capture long-term dependencies in the data. Ayad et al. [22] proposed a deep learning (DL) method based on Long Short-Term Memory (LSTM) framework to detect FDIAs. However, the above methods focus on the time characteristics of the measured data, but ignore the topology of the grid, resulting in the loss of many key information in the learning process. Furthermore, the researchers try to explore spatial correlations in grid topology to further address the FDIA detection problem. Boyaci et al. [23] proposed an FDIA identification method based on Graph Neural Network (GNN) using the graph topology of the power system and the spatial correlation of the measurement data. In order to better extract the spatial characteristics of power grid topology information and operation data, Li et al. [24] designed a detection method of FDIAs based on Gated Graph Neural Network (GGNN) to improve the detection accuracy under the change of power grid topology. Su et al. [25] proposed a Dual-Attention Multi-head Graph Attention Network (DAMGAT) for FDIAs detection to improve the interpretability of graph neural network model and the representation ability of power topology nodes. Considering the spatio-temporal dependence of power grid data, Zhang et al. [26] analyzed temporal correlation and spatial correlation by volumetric Kalman filter and Gaussian process regression to capture the dynamic characteristics of the state vector to evaluate and localize FDIAs. Han et al. [27] designed a multi-graph mechanism and a time correlation layer to mine the correlation features of power data, and constructed a graph topology for the detection of FDIAs.

Overall, existing FDIAs detection methods can already utilize data and power grid topology to achieve efficient attack detection. However, they rarely consider the spatio-temporal dependencies of power grid data, which limits the detection performance of FDIAs. *First*, in terms of spatial correlation, most methods only consider the influence between nodes connected by the power grid topology, ignoring the potential impacts from other nodes, significantly weakening the understanding of the overall correlation of the power grid topology. *Second*, in terms of temporal correlation, current works often overlook the long-term correlation of sequences and the correlation of data features with different time steps. *Third*, most methods hardly consider the subtle perturbations in spatiotemporal feature information, making it difficult for existing methods to distinguish between power grid noise and FDIA attacks, resulting in lower robustness of detection models.

Facing the aforementioned problems, we are thus motivated to design an efficient deep collaborative self-attention network in the context of robust FDIAs detection, which makes the following novel contributions:

- We design a deep collaborative self-attention network to achieve effective robust FDIA detection. Our proposed collaborative network model can effectively capture subtle perturbations from spatiotemporal feature information, efficiently distinguish power grid noise from FDIA attacks, and adapt to diverse attack intensities.
- We design a graph convolution module based on Chebyshev polynomials, which utilizes the characteristics of the graph convolution network to aggregate the node information in the power grid and introduces the spatial self-attention mechanism to adjust the degree of attention to different nodes. The proposed module can better adapt to different grid structures and characteristics, thus improving the robustness of the model to potential changes and anomalies in grid data.
- Bi-directional LSTM network with a self-attention mechanism is introduced to conduct time series modeling and long-term dependence analysis on power grid data and utilizes the temporal self-attention module to describe the time correlation of data and assign different weights to different time steps.
- Comprehensive experiments are performed over three standard datasets and demonstrate that our method can outperform state-of-the-art FDIA detection schemes in terms of detection accuracy and robustness.

The rest of this paper is organized as follows. Section 2 reviews the preliminary state estimation of smart gird and false data injection attacks. In Section 3, we propose a robust FDIA detection scheme by designing an efficient deep collaborative self-attention network. Extensive experiments are performed to evaluate the overall performance of the proposed scheme, and the corresponding results and discussions are presented in Section 4. Finally, Section 5 concludes the paper.

2 Preliminary

2.1 State Estimation and Bad Data Detection

In general, the control center estimates the state through the measurement information in the monitoring and data acquisition system to ensure the safety and stability of the power grid [15]. State estimation primarily uses the redundancy of measurement data to estimate the operating state of the grid, including bus voltage, transmission line power flow, and bus power [24]. In energy management systems, state estimation can enable the functions of power flow calculation and load forecasting and the DC power model is usually used to ensure the convergence of the state estimation. The unit voltage of each node in the system is assumed to be 1, and the effect of line resistance and ground branch is ignored, the active power between bus a and bus b in the test system can be accordingly represented by the following model:

$$P_{ab} = \frac{\beta_a - \beta_b}{X_{ab}} + e \tag{1}$$

where β_a and β_b represent voltage amplitude and voltage phase of bus a and bus b, respectively. X_{ab} is the branch impedance and e is the measurement error. The active power injection on busbar a can be expressed as the sum of the active power flow of each adjacent branch, which can be calculated as follows:

$$P_a = \sum_{a\in N} P_{ab} + e \tag{2}$$

where P_a represents active injection of bus a and N is the set of branches adjacent to bus a. Accordingly, the generalization can be expressed as:

$$z = \mathbf{H}x + e \tag{3}$$

where z is the measurement data obtained in the SCADA system. $\mathbf{H}$ is the Jacobian matrix. x is the system state vector. e is the measurement error vector. The covariance matrix is diagonal when the measurement errors are assumed to be uncorrelated. Correspondingly, according to the principle of residual and least squares, the minimization objective function $\mathscr{F}(x)$ can be calculated as follows:

$$\min \mathscr{F}(x) = (z - \mathbf{H}x)^T R^{-1}(z - \mathbf{H}x) \tag{4}$$

where $\mathbf{H}$ is the weight matrix of direction-finding quantity, and the state estimator variable x can be obtained from the objective function as:

$$\tilde{x} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{H}^{-1}z \tag{5}$$

In order to grasp the real-time operation status of the smart power grid, the SCADA system uses intelligent terminals and other devices to collect measurement data [25]. Since these measurements are easily affected by traditional power system faults, such as equipment aging, communication failure, and noise interference, the SCADA system introduces the bad data detection (BDD) module to identify and eliminate such independent and accidental natural faults, where the constructed residual vector r in BDD module is the difference between the real direction-finding vector $\tilde{z}$ and the estimated theoretical vector, which can be expressed as:

$$r = \|z - \tilde{z}\|_2 = \left\|z - \mathbf{H}\tilde{x}\right\|_2 \tag{6}$$

In the process of FDIA bad data detection, the Euclidean norm of the residual r is first compared with the threshold τ. If $r > \tau$, the presence of bad data is determined and further processed by identification and correction operations. The aforementioned bad data detection process is repeated until all residual vectors meet the pre-determined criteria.

2.2 *False Data Injection Attack*

During FDIAs, false data vector $a = [a_1, a_2, \cdots, a_m]^T$ is usually injected into the vector measurement data $z = [z_1, z_2, \cdots z_m]^T$, which inevitably cause the input vector $z_a = z + a$ of the state estimate to be biased from the real case [28]. Accordingly, the state variable x may be also offset, resulting in $\tilde{x}_a = \tilde{x} + c$, where $c = [c_1, c_2, \cdots c_n]^T$ is the deviation of the state variable before and after the attack. If the attacker can invade the configuration information of the power system and obtain the Jacobian matrix $\mathbf{H}$ from the system, the attacker could construct a false data attack vector $a = \mathbf{H}c$ so that bad data detection module cannot recognize it. Correspondingly, the attacked measurement vector z_a and residual r_a can be represented as follows:

$$z_a = \mathbf{H}\tilde{x} + e + \mathbf{H}c = \mathbf{H}(\tilde{x} + c) + e = \mathbf{H}\tilde{x}_a + e \tag{7}$$

$$\begin{aligned} r_a &= \left\|z_a - \mathbf{H}\tilde{x}_a\right\|_2 = \left\|z + a - \mathbf{H}(\tilde{x} + c)\right\|_2 \\ &= \left\|z - \mathbf{H}\tilde{x} + a - \mathbf{H}c\right\|_2 = r \end{aligned} \tag{8}$$

Meanwhile, since the bad data detection module may fail to detect FDIAs, the control center is likely to make wrong decision instructions according to the estimated state x [29]. It is worth noting that the attack effect of the FDIA method described above is related to the attacker's mastery of the grid information. When the attacker can use more power grid topology information and electrical parameters, it is entirely possible for them to construct a Jacobian matrix $\mathbf{H}$ that is closer to the topology information of the power grid, and the constructed false data vectors are very similar to the weak fluctuation noise of the power grid. Correspondingly, the bad data detection module is highly likely to identify these FDIA attacks as general noise from the power grid, allowing them to easily bypass the detection model for effective attacks.

3 Proposed Method

3.1 Overview of Proposed Detection Scheme

Our designed deep collaborative network mainly consists of two network branches. One branch is a graph convolution network based on Chebyshev polynomials, which aggregates the node information in the power grid, and adaptively adjusts the degree of attention to different nodes by introducing the spatio self-attention mechanism. Another branch is the bi-directional LSTM network, which can conduct time series modeling and long-term dependence analysis for power grid data describe the time correlation of data, and assign different weights to different time steps by introducing the temporal self-attention mechanism. The overall detection process can be described as follows. The operational data of sensors is first collected through the SCADA system, including the bus injection active power P_i and reactive power Q_i, branch power flow P_{ij} and Q_{ij}, to form the grid measurement data $Z = \left[P_i, Q_i, P_{ij}, Q_{ij}\right]$. Subsequently, after obtaining the power grid information, the attacker performs an FDIA attack to evade the BDD module. Furthermore, the operational data that is not detected by the BDD module is then fed into the deep collaborative network. Our deep collaborative network extracts temporal and spatial self-attention features through graph convolutional networks and bidirectional LSTM networks, respectively. Finally, spatio-temporal self-attention features are fully integrated by the deep collaborative network to make a final judgment decision. The overview of the proposed network architecture is shown in Fig. 2.

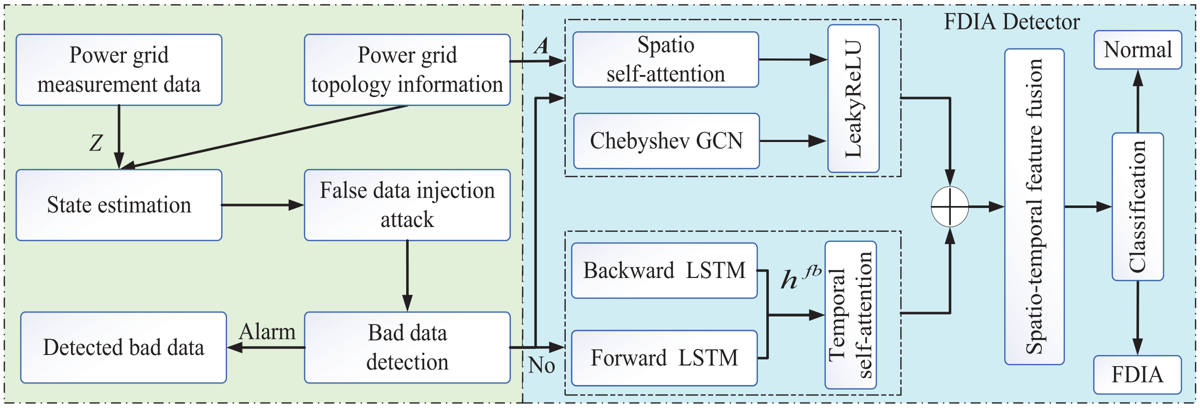


Figure 2: The overall architecture of proposed robust FDIA detection scheme

3.2 High-Order Chebyshev Graph Convolution Network

Graph convolution network (GCN) is usually used in node classification, graph classification, link prediction, and other tasks [30,31]. In general, a traditional graph convolution network only calculates the result of Chebyshev graph convolution at $k = 1$ and $\lambda = 2$. Nevertheless, since the k-order convolution operator of Chebyshev graph convolution can cover the $k - 1$-order adjacent nodes, the general GCN can only extract the spatial correlation of the first-order adjacent nodes while ignoring the effective spatial information of high-order adjacent nodes. In order to better represent the topology characteristics of the power grid, we re-construct graph convolution network by introducing k-order Chebyshev polynomial ($k \geq 2$), which can effectively capture the spatial relationship between higher-order adjacent nodes, and realize the information dissemination and feature aggregation between nodes through the adjacency matrix. Correspondingly, the connection mode and interaction between nodes in the power grid data can be better understood.

To be specific, we model the power grid topology as an undirected graph $\mathscr{G} = (\mathscr{V}, \mathscr{E}, \mathbf{A})$, where $\mathscr{V}$ represents the set of N nodes in the power grid, and $\mathscr{E}$ is the set of branches connected between nodes. Through the connection relationship between nodes in the power grid, the adjacency matrix $\mathbf{A} \in \mathscr{R}^{n\times n}$ of the graph can be obtained. If two bus nodes are directly connected, the corresponding element of two nodes in matrix $\mathbf{A}$ is 1; that is, $\mathbf{A}_{ij} = 1$, otherwise $\mathbf{A}_{ij} = 0$. The complete procedure for using Chebyshev graph convolution to update node information can be defined as follows:

$$\hat{X} = \sigma\left(\sum_{k=0}^{K} \beta_k T_k(\tilde{\Lambda})x\right) \tag{9}$$

where x is the input data. k is the order of Chebyshev polynomial. $\beta_k = (\beta_1, \beta_2, \cdots, \beta_k)$ is the parameter matrix of k-order Chebyshev polynomial. $T_k(\tilde{\Lambda})$ is a Chebyshev polynomial function, which can be defined as follows:

$$\begin{aligned} T_0(\tilde{\Lambda}) &= I \\ T_1(\tilde{\Lambda}) &= \tilde{\Lambda} \\ T_k(\tilde{\Lambda}) &= 2\tilde{\Lambda}T_{k-1}(\tilde{\Lambda}) - T_{k-2}(\tilde{\Lambda}), k \geq 2 \end{aligned} \tag{10}$$

where I is the identity matrix and $\tilde{\Lambda}$ stands for the symmetric normalized Laplace matrix, which can maintain the structural information, sparsity and positive definiteness of graph, and can be expressed as:

$$\tilde{\Lambda} = \frac{2\tilde{L}}{\lambda_{\max}} - I \tag{11}$$

where $\lambda_{\max}$ is the largest characteristic of Laplacian matrix and $\tilde{L} \in R^{n\times n}$ is the normalized Laplace matrix, which can be calculated as:

$$\tilde{L} = I - D^{-\frac{1}{2}}\mathbf{A}D^{-\frac{1}{2}} \tag{12}$$

where $D = diag\,([d_1, d_2, \cdots, d_n])$ is the degree matrix recording the degree of each node. Since Laplacian matrix is a positive semi-definite matrix and eigenvalues are greater than or equal to 0, each eigenvalue of $\frac{\tilde{L}}{\lambda_{\max}}$ is accordingly within [0, 1], $\frac{2\tilde{L}}{\lambda_{\max}}$ eigenvalues is within [0, 2], and the eigenvalue of $\frac{2\tilde{L}}{\lambda_{\max}} - I$ is within $[-1, 1]$. This result can ensure that the range of independent variables of Chebyshev polynomials is limited to $[-1, 1]$ interval, and guarantee the stability and convergence of numerical calculation in graph convolution network, thereby adapting to the properties of Chebyshev polynomials.

Although high-order Chebyshev GCN can effectively propagate information and control the propagation range, all nodes in each order polynomial only share one parameter, which makes it impossible to adaptively allocate neighbor node weights according to node differences. Therefore, we introduce the spatio self-attention mechanism to further improve its feature representation capability. The spatio self-attention mechanism can dynamically adjust the attention weight between nodes. Accordingly, the network model can accurately learn the importance of each node relationship and then pay more attention to the nodes that are crucial to FDIA detection task. The attention scoring for node i and its adjacent node j is performed through dot multiplication, and the formula is as follows:

$$e_{ij} = \frac{Q \cdot K^T}{\sqrt{d_k}} \tag{13}$$

where $Q = X \cdot W^q$, $K = X \cdot W^k$, and $V = X \cdot W^v$ are the queries, keys, and values obtained from the linear transformation of node feature X, respectively. W^q, W^k and W^v are learnable parameter matrices. d_k stands for the feature dimension used to prevent large dot multiplication results. $\hat{A}$ is the auxiliary transformation matrix to facilitate the calculation of dimension matching.

Furthermore, the attention scores of all adjacent nodes of node i are normalized by using the softmax function to obtain the attention coefficient α_{ij}. Finally, the value matrix V can be weighted by the calculated node attention coefficient α_{ij} to obtain the updated feature representation of each node.

$$\alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{i=1}^{N} \exp\left(e_{ij}\right)} \tag{14}$$

$$\tilde{S}_{at} = \hat{A}\sigma\left(\alpha_{ij} V\right) \tag{15}$$

When the network aggregates the node information according to k-order Chebyshev polynomials, we can multiply the elements of the attention weighting matrix and the polynomial $T_k(\tilde{\Lambda})$ to adjust the weight distribution. And then focus on those vulnerable nodes:

$$\tilde{T}_k(\tilde{\Lambda}) = T_k(\tilde{\Lambda}) \odot \tilde{S}_{at} \tag{16}$$

$$\tilde{X} = LeakyReLU\left(\sum_{k=0}^{K} T_k(\tilde{\Lambda})\beta_k x\right) \tag{17}$$

Finally, after all node information is aggregated, each node can dynamically adjust the feature representation according to its influence in the power grid to better reflect its role in the whole detection model.

3.3 BiLSTM Network with Temporal Self-Attention Mechanism

Bidirectional Long Short-Term Memory (BiLSTM) is a variant of recurrent neural networks (RNNs). Compared to traditional unidirectional LSTMs, BiLSTM can simultaneously consider both past and future information in a sequence, thus better capturing the long-term dependencies and contextual information of sequences [32]. A BiLSTM consists of two opposing LSTM layers, combining forward and backward information flow to enhance the capture of sequence data features [33]. In forward propagation, the BiLSTM can process the entire time sequence to capture past-to-present temporal information.

$$\vec{h}_t = \overrightarrow{\text{LSTM}}\left(\vec{h}_{t-1}, x_t, \vec{c}_t\right), t \in [1, T] \tag{18}$$

while in backward propagation, it can process from the time sequence end and mainly focus on future-to-present temporal information. BiLSTM allows each time point to access contextual information, comprehensively understanding temporal dependencies.

$$\overleftarrow{h}_t = \overleftarrow{\text{LSTM}}\left(\overleftarrow{h}_{t+1}, x_t, \overleftarrow{c}_t\right), t \in [1, T] \tag{19}$$

where x_t represents the input of the current sample. $\overleftarrow{c}_t$ and $\overrightarrow{c}_t$ represent the backward and forward cell states of the current sample. T is the length of the input time series. After processing the entire sequence, BiLSTM can ensure the output contains important features from both the beginning and end of the sequence by merging the hidden states from forward and backward propagation, which can significantly enhance the understanding and predictive accuracy of time series data.

$$h_t^{fb} = \delta_1 \overrightarrow{h}_t + \delta_2 \overleftarrow{h}_t \tag{20}$$

where δ_1 and δ_2 are the weight ratios of forward and backward hidden states. In general, a complete LSTM unit is composed of input gate, forgetting gate, output gate, and cell state [32]. The input gate that controlled by the sigmoid function can regulate the impact of the current input on the cell state. The forget gate determines the extent to which the cell state from the previous time step is forgotten, while the output gate regulates the influence of the cell state on the hidden state. The cell state is responsible for transmitting and storing long-term dependency information, and gate mechanisms control the flow and storage of information, which enable LSTM to capture long-term dependencies in sequences while maintaining gradient stability.

In view of the dynamic and complexity of the time series data involved in the FDIAs detection task, we further introduce the temporal self-attention mechanism to enhance the performance and robustness of the BiLSTM model. Our goal is to improve the model's perception of the dynamic characteristics of time series data, so as to more accurately find the possible abnormal data, and further improve the model's robustness. Specifically:

$$\tilde{h}_t = \alpha_j \odot h_t^{fb} \tag{21}$$

$$Te_{at} = B\sigma\left(\tilde{h}_t V_t\right) \tag{22}$$

where $\alpha_j = \left[\alpha_1, \alpha_2, \cdots, \alpha_j\right]$ is temporal self-attention weight matrix. h_t^{fb} is the forward and backward hidden state extracted by the current BiLSTM, and B is an auxiliary transformation matrix. Then, we combine the attention weight matrix of BiLSTM for weighted output to update the representation of temporal features. Through the temporal self-attention mechanism, the weight of the power grid time series data at different time points can be efficiently allocated, which helps in processing the long series data more effectively, taking into account the context dependency of the time series and the importance of different time points. Notably, a dropout layer is added at the end of the network to prevent its overfitting and improve the generalization capability of the model.

3.4 Spatio-Temporal Feature Fusion

Considering the spatial correlation and temporal dependence, the features can be further fused to improve the representation capability of the features by combining self-attention and temporal self-attention features. To be specific, we use the pre-processed measurement data as the input $X = [x_1, x_2, \cdots, x_n]$ of the network model. In the processing of spatial correlation, the topology of the power grid can be firstly obtained to construct the corresponding adjacency matrix $\mathbf{A}$, and the corresponding

Laplace matrix $\tilde{\Lambda}$ is sequentially calculated. Then, we perform Chebyshev graph convolution by setting appropriate hyperparameters to effectively aggregate the information of adjacent nodes.

$$\tilde{X} = \text{LeakyReLU}(\text{CGCN}\,(X)\;W_1 + b_1) \tag{23}$$

where LeakyReLU is nonlinear activation function, W_1 is parameter matrix, and b_1 is the deviation. After obtaining the spatial characteristics, we further consider the temporal characteristics of the grid measurement data.

$$Te = \text{LeakyReLU}(\text{BiLSTM}\,(X)\;W_2 + b_2) \tag{24}$$

Finally, the spatio correlation information and temporal dependence information can be effectively fused to improve the representation capability of the features.

$$F_u = W_c\tilde{X} \odot W_b Te \tag{25}$$

where W_c and W_b are two weighted matrices. After performing spatio-temporal feature fusion, the fused features can be fed to the full connection layer, and then output the results through the activation function to determine the normal and abnormal probability of the grid measurement data.

$$\hat{Y} = \sigma\left(W_y F_u + b_y\right) \tag{26}$$

where W_y are weighted matrices, b_y is the deviation, and σ is sigmoid activation function. By combining the high-order Chebyshev graph convolution network and BiLSTM network with temporal self-attention mechanism, an efficient deep collaborative network can be integrated, which is called as CGCN-BiLSTM(Chebyshev Graph Convolution-Bidirectional Long Short-Term Memory). The complete network architecture is shown in Fig. 3.

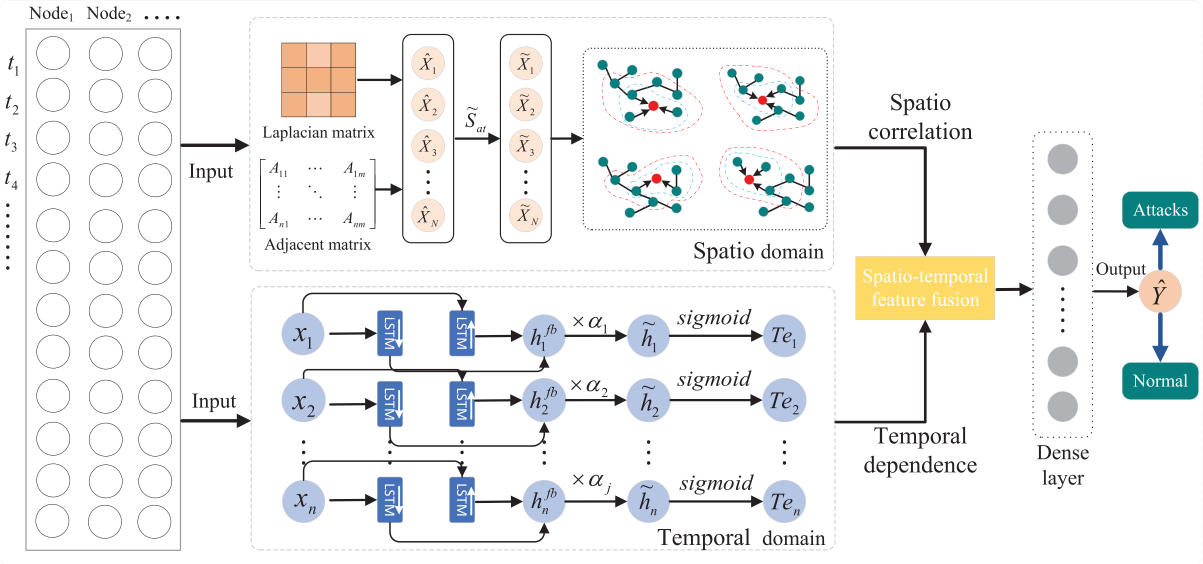


Figure 3: Structure diagram of a deep collaborative network with spatiotemporal self-attention mechanism

In addition, it is necessary to construct the calculation error and back-propagation error of the loss function to guide the updating of network parameters during the training process of the neural network. In our network model, the cross entropy loss function is used to calculate the loss of the model, and the cosine annealing learning rate scheduler is also involved to dynamically adjust the learning rate and optimize the training process of the deep learning model.

$$\mathscr{L}_{loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{Y}_i) + (1 - y_i)\log(1 - \hat{Y}_i)] \tag{27}$$

where N is the input sequence length, y_i stands for the real label (–1 or 1) of the i-th sample, and $\hat{Y}_i$ represents the prediction probability that the model belongs to the positive class for the i-th sample.

3.5 Robust FDIA Detection Procedure

Based on the above-mentioned the spatiotemporal feature fusion model, we can build the detailed FDIA detection procedure, which can be described detailedly as follows, e.g., **Algorithm 1**, which can be described detailedly as follows:

- **Step 1:** We conduct power flow calculations on the data collected from the power grid to obtain measurement data, which is then pre-processed to get input features $X \in \mathscr{R}^{N\times T}$.
- **Step 2:** We further compute the adjacency matrix $\mathbf{A} \in R^{N\times N}$ and Laplacian matrix $\tilde{\Lambda} \in \mathscr{R}^{N\times N}$ corresponding to the bus system, and then normalize the standardized X, $\mathbf{A}$, $\tilde{\Lambda}$ as the inputs to the CGCN-BiLSTM model.
- **Step 3:** The information of various nodes in the power grid is sequentially aggregated by using CGCN to get the feature representation $\hat{X}$. Subsequently, spatial information can be further integrated to obtain feature representation $\tilde{X}$ by combining spatial self-attention $\tilde{S}_{at}$ with Chebyshev polynomials, $\hat{X}$ and the Laplacian matrix.
- **Step 4:** Input feature X into bidirectional LSTM units to capture its long-term dependencies. Then, the forward and backward hidden states h_t^{fb} can be generated. Subsequently, the state feature $\tilde{h}_t$ is further updated by using temporal attention and obtaining the temporal feature representation Te_{at} through activation functions.
- **Step 5:** Finally, the spatio-temporal features are fully fused to output the model decision $\hat{Y}$ through fully connected layers. Notably, the parameters can be gradually optimized through the loss function.

Algorithm 1: Spatio-Temporal Feature Fusion Procedure

Input: $X \in \mathscr{R}^{N\times T}$ → input feature value; $\mathbf{A} \in \mathscr{R}^{N\times N}$ → adjacent matrix; $\tilde{\Lambda} \in \mathscr{R}^{N\times N}$ → Laplacian matrix; k → chebyshev order; E → epochs; B → batchsize;

Output: $\hat{Y}$ → model detection result;

```
for $i = 1 \rightarrow i = E$ do
    for $j = 1 \rightarrow j = B$ do
        $\tilde{S}_{at} = \hat{A}\sigma\,(\alpha_s V)$;
        $\tilde{T}_k\left(\tilde{\Lambda}\right) = T_k\left(\tilde{\Lambda}\right) \odot \tilde{S}_{at}$;
        $\tilde{X} = LeakyReLU\left(\sum_{k=0}^{K} T_k\left(\tilde{\Lambda}\right)\beta_k X_{ij}\right)$;
    end for
    for $t = 1 \rightarrow t = T$ do
        $h_t^{fb} = \delta_1 \overrightarrow{h_t} + \delta_2 \overleftarrow{h_t}$;
```

(Continued)

Algorithm 1 (continued)

```
        $\tilde{h}_t = \alpha_t \odot h_t^{fb}$;
        $Te_{at} = B\sigma\left(\tilde{h}_t V_t\right)$;
    end for
    $F_u = W_c\tilde{X} \odot W_b Te$;
    $\hat{Y} = \sigma\left(W_y F_u + b_y\right)$;
    $loss = \mathscr{L}_{loss}\left(y_i, \hat{Y}_i\right)$;
    if $\hat{Y} = y_i$ then
        attacked;
    else
        normal;
    end if
end for
```

4 Experimental Results and Discussions

In this section, we first introduced the experimental setup and evaluation metrics in Sections 4.1 and 4.2, respectively. Then, a series of comparisons and discussions aiming at overall detection performance were sequentially performed in Section 4.3. Furthermore, we discussed the robustness between our scheme and several state-of-the-art schemes in Section 4.4. Last but not least, the computation complexity of different detection schemes was tested in Section 4.5.

4.1 Experimental Setup

In our experiment, the load information of different regions of NYISO was utilized as the basic power gird data to generate normal measurement data. We simulated the measured data $\mathscr{X}$ of the power system by power flow calculation based on this load information. Matpower toolkit was used to obtain relevant information, such as Jacobian matrix **H** and power grid topology, and FDIA sample sets were constructed to maximize the avoidance of BDD detection.

In order to ensure the authenticity of the experiment, we added the standard Gaussian distribution noise with a variance of 0.01 and mean value of 0, which was applied to different datasets, IEEE 14 bus system, IEEE 39 bus system, and IEEE 118 bus system, to generate experimental data with varying degrees of noise interference. Our goal is to verify that our detection model can still efficiently and accurately identify FDIA attacks under different levels of noise interference. In addition, for each bus system, we generated 15,000 groups of measurement data, including 7500 groups of FDIA data and 7500 groups of normal data. We labeled FDIA data as 1 and normal data as -1 to facilitate model detection. Correspondingly, all data samples were divided the data set into 75% training set and 25% test set for each experiment.

Moreover, in the process of model training, our network model can gradually update the weight parameters by calculating the loss value and gradient. In order to speed up the convergence speed of model training and prevent gradient explosion, we normalized the input data by the *minmax* function, and then adopted the exit strategy to temporarily stop the work of some neurons with a certain probability p to avoid the over-fitting of the model. All experiments were performed over Pytorch 1.7 framework with the Intel Core i9-12900hx CPU, 16GB RAM and NVIDIA GeForce RTX 4060 GPU.

The power flow calculation and state estimation of power grid data were simulated using Matpower toolkit on MATLAB.

4.2 Evaluation Metrics

To provide sufficient performance comparison in terms of detection performance, we introduced four evaluation metrics, i.e., accuracy, precision, recall and F_1 score, to give the experimental results. In general, the definition of statistical variables involved in the evaluation indicators were true negative (TN), false positive (FP), true positive (TP), and false negative (TN), respectively, which are as follows: 1) TN represents the number of correctly identified normal measurement data of the power grid as normal data, denoted as β_{TN}. 2) FP refers to the number of errors in identifying normal measurement data of the power grid as FDIA, denoted as β_{FP}. 3) TP stands for the number of FDIA data correctly identified as FDIA in the power grid, denoted as β_{TP}. 4) TN means the number of times FDIA data in the power grid is incorrectly identified as normal data, denoted as β_{FN}. Correspondingly, the calculation expressions for the four evaluation indicators are given as follows:

$$Acc = \frac{\beta_{TP} + \beta_{TN}}{\beta_{TP} + \beta_{FP} + \beta_{TN} + \beta_{FN}} \tag{28}$$

Acc is the ratio of all correctly judged FDIA samples. The higher the value of *Acc*, the better the overall performance of the detection model.

$$Pre = \frac{\beta_{TP}}{\beta_{TP} + \beta_{FP}} \tag{29}$$

Pre is the ratio of real FDIA samples in the predicted FDIA samples. The larger the value of *Pre*, the lower the false alarm rate of the detection model and the better the detection effect.

$$Rec = \frac{\beta_{TP}}{\beta_{TP} + \beta_{FN}} \tag{30}$$

Rec is the ratio of correctly predicted FDIA samples in real FDIA samples. The higher the recall rate, the lower the missed detection rate of the detection model.

$$F_1 = \frac{2 \times Pre \times Rec}{Pre + Rec} \tag{31}$$

F_1 is the harmonic average of precision and recall. The higher the F_1 score, the better the overall performance of the detection model.

4.3 Comparison with the State of the Arts

In this section, we conducted a series of experiments to compare the proposed scheme with existing state-of-the-art FDIA detection schemes, CNN [20], LSTM [22], GCN [24], CNN-LSTM [34] and DAMGAT [25]. In order to verify the effectiveness and reliability of the proposed FDIAs detection method, four different evaluation metrics, Accuracy, Precision, Recall, F_1 score, were sequentially tested. All experiments were carried out over three standard data sets, IEEE 14 bus system, IEEE 39 bus system, and IEEE 118 bus system, to provide the experimental results. In each experiment, the input batch of all models was set to 64, and the number of training iterations was 80. The initial learning rate was set to 0.01 and adjusted using the Adam optimizer. The parameter probability of the dropout layer was initially set to 0.4. The order of Chebyshev polynomial of the proposed model is set to 3, the number of layers of LSTM is 2, and the number of layers of self-attention is 2.

The corresponding experimental results are shown in Table 1. As can be observed from this table, our scheme can obtain the best overall detection performance compared to other state-of-the-art schemes, no matter which dataset is used. To be specific, on IEEE 14 bus system, compared with the current best scheme DAMGAT [25], our scheme can obtain a significant performance gain with 3.25% for accuracy, 4.37% for precision, 1.81% for recall, and 3.02% for F_1 score. For IEEE 39 bus system, our solution can achieve approximate improvements with 3.60% for accuracy, 3.67% for precision, 3.25% for recall, and 3.47% for F_1 score. Similarly, for the larger scale data from the IEEE 118 bus system, our scheme can still achieve an approximate performance advantage, that is, the performance gain with 3.14% for accuracy, 2.92% for precision, 3.24% for recall, and 3.08% for F_1 score. In addition, we conducted the experimental comparison of FPR (False Positive Rate) on IEEE 14, 39 and 118 bus systems, respectively, and the results are shown in Table 2. It can be seen from the data in the table that the FPR value of our model is significantly lower than that of other advanced models, indicating that the proposed method performs better in distinguishing between true and false samples.

Table 1: Overall performance comparison for six detection methods, CNN, LSTM, GCN, CNN-LSTM, DAMGAT, CGCN-BiLSTM. Four evaluation metrics, accuracy, precision, recall, F_1 score, were used to provide the results. All experiments were performed on IEEE 14, IEEE 39 and IEEE 118 bus systems

Methods	IEEE 14 bus system				IEEE 39 bus system				IEEE 118 bus system			
	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
CNN [20]	0.8151	0.8522	0.7651	0.7963	0.8109	0.8170	0.7360	0.7634	0.8030	0.8069	0.7260	0.7526
LSTM [22]	0.8662	0.8331	0.8864	0.8721	0.8553	0.8354	0.8845	0.8593	0.8457	0.9015	0.7757	0.8339
GCN [24]	0.9157	0.9299	0.9039	0.9167	0.9067	0.8754	0.9539	0.9130	0.8883	0.8855	0.8987	0.8920
CNN-LSTM [34]	0.9223	0.9062	0.9419	0.9237	0.9110	0.9168	0.9091	0.9129	0.8983	0.8792	0.9232	0.9007
DAMGAT [25]	0.9493	0.9460	0.9558	0.9509	0.9360	0.9282	0.9487	0.9383	0.9263	0.9347	0.9208	0.9277
CGCN-BiLSTM	**0.9818**	**0.9897**	**0.9739**	**0.9811**	**0.9720**	**0.9649**	**0.9812**	**0.9730**	**0.9577**	**0.9639**	**0.9532**	**0.9585**

Note: The significance of bold value shows the maximum in current column.

Table 2: Overall performance comparison for six detection methods, CNN, LSTM, GCN, CNN-LSTM, DAMGAT, CGCN-BiLSTM. Evaluation metrics false positive rate (FPR) was used to provide the results. All experiments were performed on IEEE 14, IEEE 39 and IEEE 118 bus systems

Methods	Different datasets		
	IEEE 14 bus system	IEEE 39 bus system	IEEE 118 bus system
CNN [20]	0.1986	0.2028	0.2364
LSTM [22]	0.1412	0.1571	0.1849
GCN [24]	0.0970	0.1069	0.1269
CNN-LSTM [34]	0.0850	0.0976	0.1158
DAMGAT [25]	0.0591	0.0731	0.0812
CGCN-BiLSTM	**0.0193**	**0.0314**	**0.0489**

Note: The significance of bold value shows the minimum in current column.

The above experimental results demonstrated that our scheme can achieve significant performance improvements on both small-scale and large-scale datasets. In fact, this phenomenon can be easily explained by the following two reasons. Firstly, our proposed deep collaborative network combined the characteristics of the graph convolution network and Bi-LSTM to optimize the network structure, which can better adapt to different power grid structures and effectively capture the difference between the slight changes and actual FDIA attacks in the power grid, thus improving the robustness of the model to potential FDIA attacks. Secondly, the self-attention mechanism was introduced in spatiotemporal feature construction, which can further enhance the representation capability of spatiotemporal features, thereby efficiently guiding detection features to pay more attention to the differences between FDIA attack samples and normal data, resulting in a significant improvement of detection accuracy.

Furthermore, we can observe an interesting phenomenon from Table 1, that is, the overall detection performance on the IEEE 118 bus system is slightly lower than that on the IEEE 39 and IEEE 14 bus systems. To be specific, for accuracy measurement, the average reductions for IEEE 118 bus system were approximately 2.41% for IEEE 14 bus system, 1.43% for IEEE 39 bus system. For F_1 score, the average reductions for IEEE 118 bus system were approximately 2.26% for IEEE 14 bus system, 1.45% for IEEE 39 bus system. This is mainly because if a bus system has a larger scale and more complex topology, it means that more system nodes and more complex power grid structures need to be processed in the FDIA attack detection process. This is inevitably more susceptible to external interference on the topology of the power grid, which can introduce more complex noise interference and lead to attack misjudgment by the detection model. Correspondingly, our proposed network model can deeply analyze and understand the behavior and interaction of each power node, which can efficiently process the spatial-temporal complexity of the power grid and make the detection model more sensitive to identify and respond to abnormal data in various power grid operations, thereby enhancing the adaptability and detection accuracy in the changing complex power grid environment.

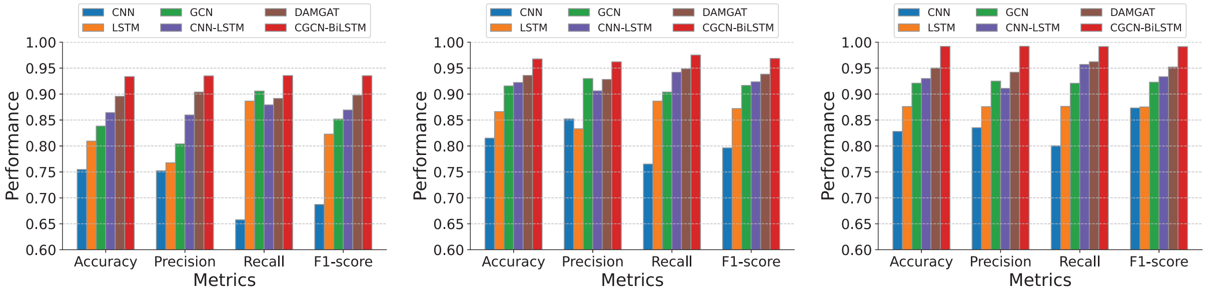
4.4 Robustness Analysis

In order to gain more insight, we further verified the applicability and robustness of the proposed detection scheme in the real power grid environment. In our experiments, a series of comparative experiments, including different attack intensities, different noise environments and different node attacks, were conducted over standard datasets from three bus systems, IEEE 14, IEEE 39 and IEEE 118 bus systems.

4.4.1 Attack Intensity

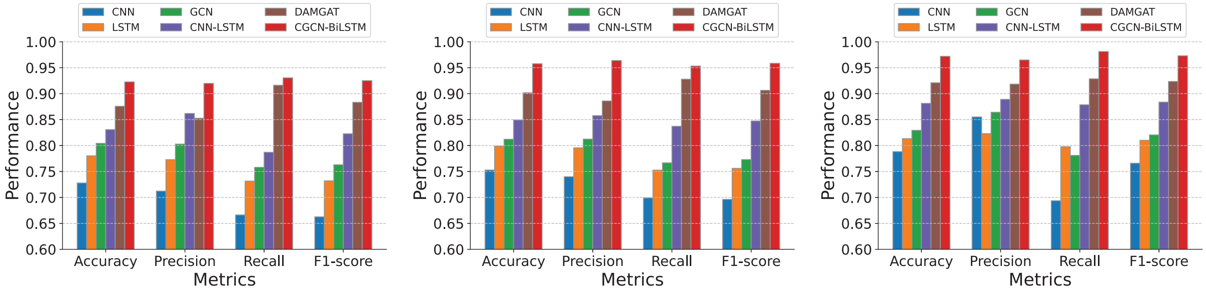
Attack intensity refers to the degree of deviation between the false data injected by the attacker and the actual measured data. A higher attack intensity means that there is a large deviation between the injected false data and the actual data, while a lower attack strength means that there is only a slight difference. Intuitively, FDIA attackers always hope that the attack data can bypass the detection model as much as possible while completing an effective attack, which forces the detection model to adapt to attack levels of different densities as much as possible. To give a more valid observation, we divided the generated FDIAs samples into the strong attack, medium attack, and weak attack [35] and observed the robustness of our proposed detection model under different levels of attack intensity, where strong attack is that the ratio of the average injection power deviation to the actual measurement is greater than 30%, medium attack is that the ratio ranges between 10% and 30%, while weak attack is that the ratio is less than 10%.

Figs. 4–6 present the experimental comparison of different attack intensities on IEEE 14, IEEE 39 and IEEE 118 bus systems, respectively. The abscissa in these figures shows the accuracy rate, accuracy rate, recall rate and F_1 score index, while the ordinate is the overall detection performance of each metric. From the results, we can easily observe that the proposed scheme can almost always obtain superior detection performance values compared with most of the existing schemes, no matter which data set from IEEE bus system is used. To be specific, for a high-intensity attack on IEEE 118 bus system, the proposed scheme can obtain F_1 score about 3.23% higher than DAMGAT scheme [25] and F_1 score about 10.33% higher than CNN-LSTM scheme [34]. Even if for low-intensity attack on IEEE 118 bus system, our scheme still showed an obvious advantage compared with other schemes, e.g., the F_1 score gains of our scheme were 2.95% for DAMGAT scheme, 12.09% for CNN-LSTM scheme, 16.05% for GCN scheme, respectively. We believe that this mainly benefits from our proposed deep collaborative network and the introduction of a spatio-temporal self-attention mechanism. On the one hand, the deep collaborative network can better adapt to different power grid structures and effectively capture the difference between the slight changes and actual FDIA attacks in the power grid. On the other hand, the introduction of self-attention mechanism further enhances the representation capability of spatio-temporal features, resulting in the improvement of detection performance.



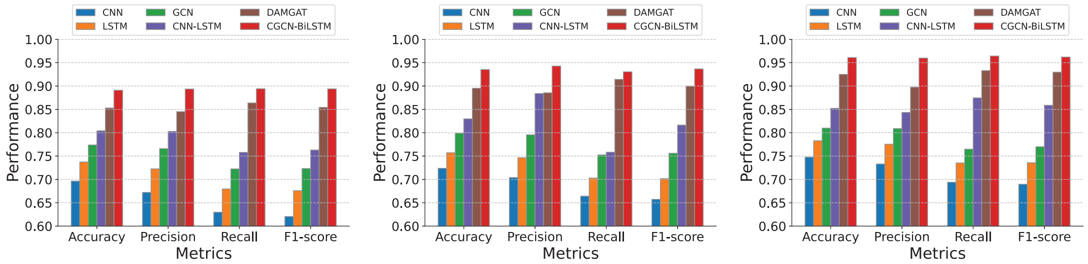
(a) Low intensity attack (9%) (b) Medium intensity attack (20%) (c) High intensity attack (35%)

Figure 4: Performance comparison with four evaluation metrics for different attack intensities on IEEE 14 bus system. Three types of intensity attacks, Low (9%), Medium (20%), High (35%), were tested in each experiment



(a) Low intensity attack (9%) (b) Medium intensity attack (20%) (c) High intensity attack (35%)

Figure 5: Performance comparison with four evaluation metrics for different attack intensities on IEEE 39 bus system. Three types of intensity attacks, Low (9%), Medium (20%), High (35%), were tested in each experiment



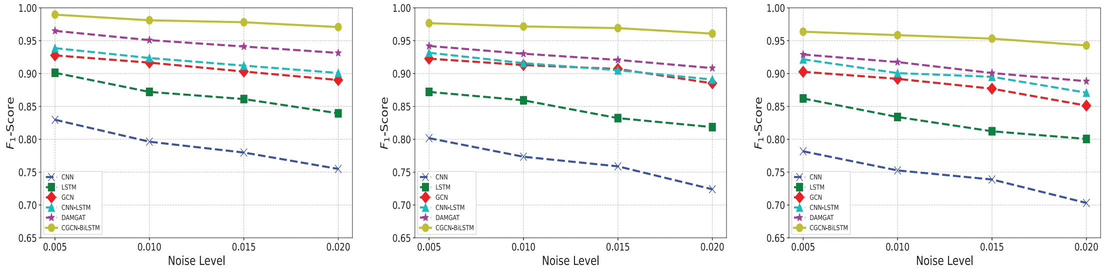
(a) Low intensity attack (9%) (b) Medium intensity attack (20%) (c) High intensity attack (35%)

Figure 6: Performance comparison with four evaluation metrics for different attack intensities on IEEE 118 bus system. Three types of intensity attacks, Low (9%), Medium (20%), High (35%), were tested in each experiment

4.4.2 Noise Interference

To give more insight, we implemented a series of experiments to further verify the robustness of the proposed method against measurement noise interference. Standard Gaussian noise with different variances were introduced into power grid measurement to simulate the real power grid environment, and performed specific experiments over IEEE 14, IEEE 39 and IEEE 118 bus systems.

The experimental results were shown in Fig. 7, where the abscissa represents the noise variance, while the ordinate is F_1 score, which is used to measure the detection performance. Moreover, we compared the accuracy rate with other advanced detection models at different noise levels, and the experimental results are shown in Table 3, where the noise error is between 0.5% and 2%. As can be seen from the figure and table, our scheme always outperformed other state-of-the-art detection methods in various noise levels, demonstrating the superior detection capability of our scheme. In addition, we can also observe that with the noise level increasing, the overall detection performance of all schemes decreases. This is because high noise levels may obscure the characteristics of attack data, making the detection model difficult to distinguish between attack and normal data. However, our scheme exhibited the smallest performance decreasing under identical noise interference, suggesting that the model's self-attention mechanism dynamically adjusts its focus on different parts of the input data, effectively reducing the interference of noise on the decision-making process.



(a) IEEE 14 bus system (b) IEEE 39 bus system (c) IEEE 118 bus system

Figure 7: Performance comparison under different noise environments. All experiments were performed over IEEE 14, IEEE 39 and IEEE 118 bus systems and standard Gaussian noises were used to simulate the real power grid environment

Table 3: Comparison of accuracy of six detection methods: CNN, LSTM, GCN, CNN-LSTM, DAMGAT and CGCN-BiLSTM under different ambient noise. All experiments were performed on IEEE 14, IEEE 39, and IEEE 118 bus systems

Methods	IEEE 14 bus system				IEEE 39 bus system				IEEE 118 bus system			
	0.5%	1.0%	1.5%	2.0%	0.5%	1.0%	1.5%	2.0%	0.5%	1.0%	1.5%	2.0%
CNN [20]	0.8198	0.7963	0.7789	0.7549	0.8067	0.7738	0.7598	0.7249	0.7876	0.7526	0.7398	0.7032
LSTM [22]	0.9022	0.8711	0.8672	0.8385	0.8731	0.8583	0.8322	0.8175	0.8621	0.8349	0.8142	0.8005
GCN [24]	0.9277	0.9167	0.9051	0.8902	0.9177	0.9130	0.9070	0.8852	0.9018	0.8931	0.8779	0.8516
CNN-LSTM [34]	0.9329	0.9196	0.9041	0.9012	0.9244	0.9128	0.9071	0.8879	0.9055	0.8962	0.8813	0.8702
DAMGAT [25]	0.9651	0.9509	0.9419	0.9324	0.9427	0.9308	0.9211	0.9094	0.9298	0.9197	0.9019	0.8902
CGCN-BiLSTM	**0.9897**	**0.9812**	**0.9782**	**0.9709**	**0.9769**	**0.9719**	**0.9694**	**0.9563**	**0.9639**	**0.9587**	**0.9502**	**0.9411**

Note: The significance of bold value shows the maximum in current column.

4.4.3 Node Number Discussions

To accurately simulate real-world conditions in power grids, we further study the comparative experiments on single-node and multi-node attacks. In general, for FDIA attackers, single-node attacks can significantly reduce attack costs when the resources are limited. However, in some specific scenarios, the attackers may simultaneously implement FDIA attacks on multiple nodes to achieve maximum attack effectiveness.

We tested the detailed comparative experiments with different numbers of attack nodes on the IEEE 14, IEEE 39, and IEEE 118 test systems. The corresponding experimental results were shown in Tables 4–6, respectively. From these tables, the proposed method is significantly superior to the existing several detection methods on three bus systems, whether the single node or multiple nodes scenario was tested. Furthermore, the experimental results indicated that in single-node attack scenarios, the attack data are often limited to a very small number of data points, greatly increasing the difficulty of identifying FDIAs within a vast dataset of normal operations. In contrast, for multi-node attacks, due to strong inter-node relationships and the wide distribution of data, the model can obtain a richer set of data points for analysis, thereby generally achieving higher detection performance than single-node attacks. This observation further validates the effectiveness and high adaptability of the proposed scheme in complex attack scenarios.

Table 4: Performance comparison with four metrics for single-node FDIAs detection and multi-node FDIAs detection on IEEE 14 bus system

Methods	Single-node detection				Multi-node detection			
	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
CNN [20]	0.7820	0.8074	0.7728	0.7755	0.8175	0.8412	0.7863	0.7925
LSTM [22]	0.8539	0.8564	0.9067	0.8699	0.8710	0.8330	0.9364	0.8817
GCN [24]	0.9097	0.8953	0.9331	0.9138	0.9173	0.9017	0.9416	0.9212
CNN-LSTM [34]	0.9093	0.8783	0.9558	0.9154	0.9210	0.8941	0.9597	0.9258
DAMGAT [25]	0.9417	0.9756	0.9091	0.9412	0.9510	0.9462	0.9591	0.9526
CGCN-BiLSTM	**0.9733**	**0.9662**	**0.9825**	**0.9742**	**0.9871**	**0.9945**	**0.9796**	**0.9865**

Note: The significance of bold value shows the maximum in current column.

Table 5: Performance comparison with four metrics for single-node FDIAs detection and multi-node FDIAs detection on IEEE 39 bus system

Methods	Single-node detection				Multi-node detection			
	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
CNN [20]	0.7743	0.8058	0.7313	0.7405	0.8127	0.8363	0.7813	0.7869
LSTM [22]	0.8381	0.8054	0.8889	0.8397	0.8575	0.8638	0.9010	0.8715
GCN [24]	0.8909	0.8969	0.9052	0.8957	0.9093	0.8783	0.9558	0.9154
CNN-LSTM [34]	0.8959	0.8995	0.9094	0.8998	0.9173	0.9017	0.9416	0.9212
DAMGAT [25]	0.9143	0.9160	0.9105	0.9116	0.9280	0.9066	0.9584	0.9318
CGCN-BiLSTM	**0.9649**	**0.9627**	**0.9681**	**0.9643**	**0.9727**	**0.9754**	**0.9708**	**0.9725**

Note: The significance of bold value shows the maximum in current column.

Table 6: Performance comparison with four metrics for single-node FDIAs detection and multi-node FDIAs detection on IEEE 118 bus system

Methods	Single-node detection				Multi-node detection			
	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
CNN [20]	0.7634	0.7916	0.7535	0.7561	0.8016	0.8309	0.7858	0.7952
LSTM [22]	0.8274	0.8501	0.7982	0.8045	0.8456	0.8190	0.8828	0.8447
GCN [24]	0.8664	0.8730	0.9018	0.8779	0.8903	0.8373	0.9760	0.9013
CNN-LSTM [34]	0.8864	0.8912	0.9059	0.8926	0.9057	0.9392	0.8727	0.9047
DAMGAT [25]	0.8895	0.8956	0.9049	0.8948	0.9137	0.9378	0.8909	0.9138
CGCN-BiLSTM	**0.9367**	**0.9470**	**0.9235**	**0.9339**	**0.9587**	**0.9527**	**0.9675**	**0.9601**

Note: The significance of bold value shows the maximum in current column.

4.5 Computational Complexity Analysis

To further demonstrate the superiority of our proposed scheme, we performed a detailed analysis of the complexity and compared the running time of different FDIA detection schemes. To estimate the complexity of each model, we compared the training times of different detection schemes. The running time of five existing state-of-the-art FDIA detection schemes, CNN [20], LSTM [22], GCN [24], CNN-LSTM [34] and DAMGAT [25], and our proposed scheme. The experiments were performed under identical environmental conditions to ensure a fair comparison. The training time per epoch was the average of ten epochs, while the total training time was the average of three complete model training. In addition, on IEEE 14 bus system, the total maximum execution time of the model is 158.12 s; On the larger IEEE 39 bus system, this time is extended to 491.49 s; On the most complex IEEE 118 bus system, the total maximum execution time reached 1472.16 s. In the self-attention mechanism we introduce, we use linear multiplication, the time complexity is mainly with the input node, its time complexity is maintained in $O(n)$, and the space complexity is maintained in the range of constant change, that is $O(1)$. In addition, the computational overhead of the bi-directional LSTM network is mainly related to the input sequence length T. Its consumption mainly increases linearly with T, so the time complexity of the bi-directional LSTM network can be simplified to $O(T)$ and the space complexity to $O(T)$. In order to verify the feasibility of the method, we carry out experiments to detect the delay and throughput. The average detection delay of IEEE 14 bus system is 0.0034 s, and the throughput is 8245.95 samples per second. On the IEEE 39 bus system, the average detection delay is 0.0106 s and the throughput is 2621.37 samples per second. The IEEE 118 bus system has an average detection delay of 0.0312 s and a throughput of 897.83 samples per second (taking the average of the results of 20 iterations each). Different datasets were tested over the computer platform with Intel Core i9-12900hx CPU, 16 GB RAM and NVIDIA GeForce RTX 4060 GPU.

The corresponding experimental results were presented in Table 7. As can be seen from this table the average single training and total training time of our scheme were significantly lower than that of DAMGAT scheme, whichever IEEE bus system was used. Specifically, our scheme can reduce by 0.74 and 58.35 s for IEEE 14 bus system, 1.83 and 156.04 s for IEEE 39 bus system, and 8.49 and 623.72 s for IEEE 118 bus system. This is because the two-layer multi-head attention mechanism adopted by DAMGAT increases the computational complexity and leads to a long training time. Our proposed spatio-temporal self-attention mechanism simplifies the attention calculation and is more suitable for detection tasks, which significantly shortens the training time. In addition, compared to some traditional detection schemes, e.g., CNN, LSTM, our scheme got an obviously longer average single training and total training times. This is because the internal structure of traditional detection models is simpler, and their corresponding training process is thus faster. However, their feature extraction and representation capability is also correspondingly limited, which inevitably lowers their detection performance.

Table 7: Average training time for six different FDIA detection schemes. The experiments were performed over IEEE 14, IEEE 39 and IEEE 118 bus systems, and average single training time (s) (S-training in short) and average total training time (s) (T-training in short) were discussed in each test

Methods	IEEE 14 bus system		IEEE 39 bus system		IEEE 118 bus system	
	S-training (s)	T-training (s)	S-training (s)	T-training (s)	S-training (s)	T-training
CNN [20]	0.96	78.15	2.72	255.85	9.17	767.43
LSTM [22]	1.08	87.44	3.14	259.69	10.38	842.04
GCN [24]	1.31	107.81	4.03	328.92	12.78	1051.14
CNN-LSTM [34]	1.77	144.65	4.97	420.93	16.54	1362.60
DAMGAT [25]	2.56	206.89	7.71	620.67	25.90	2062.69
CGCN-BiLSTM	1.82	148.54	5.88	464.63	17.41	1438.97

5 Conclusions

This paper proposed a deep collaborative self-attention network to achieve effective and robust FDIA detection. The proposed network designed a high-order Chebyshev polynomials-based graph convolution module to aggregate the node information in the power grid and introduced spatial self-attention mechanism to adjust the degree of attention given to different nodes. Furthermore, a bidirectional LSTM network with a self-attention mechanism was introduced to conduct time series modeling and long-term dependence analysis and assign different weights to different time steps. The proposed network model can effectively capture subtle perturbations from spatio-temporal feature information, efficiently achieving robust FDIA detection, and adapting to diverse attack intensities.

Extensive experiments demonstrated that the proposed method outperformed existing state-of-the-art FDIA detection schemes in terms of detection accuracy and robustness.

While our scheme can improve the efficiency and robustness of FDIA detection, it should be noted that the training of the proposed network model may be more complex and time-consuming, especially for power grid topology with dynamic changes. In terms of future work, we aim to refine our new scheme in two ways. First, we intend to investigate FDIA detection method applied to dynamic topology changes in the power grid, which may be more practical for new power systems. Second, we intend to explore the lightweight of the deep collaborative network model by optimizing the self-attention structure. These two issues are left for our future work.

Acknowledgement: The authors would like to thank anonymous reviewers for their valuable suggestions which helped to improve this article.

Funding Statement: This work was supported in part by the Research Fund of Guangxi Key Lab of Multi-Source Information Mining & Security (MIMS21-M-02).

Author Contributions: The authors confirm contribution to the paper as follows: data collection: Tong Zu; analysis and interpretation of results: Tong Zu; draft manuscript preparation: Tong Zu; study conception and design: Fengyong Li; supervision and revision: Fengyong Li. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: All data used or analyzed during this study are included in this article and its references.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Ghiasi M, Niknam T, Wang Z, Mehrandezh M, Dehghani M, Ghadimi N. A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: past, present and future. Elect Power Syst Res. 2023;215:108975. doi:10.1016/j.epsr.2022.108975.
2. Islam MS, Rahman MA, Bin Ameedeen MA, Ajra H, Ismail ZB, Zain JM. Blockchain-enabled cybersecurity provision for scalable heterogeneous network: a comprehensive survey. Comput Modeling Eng Sci. 2024;138(1):43–123. doi:10.32604/cmes.2023.028687.
3. Reda HT, Anwar A, Mahmood AN, Tari Z. A taxonomy of cyber defence strategies against false data attacks in smart grids. ACM Comput Surv. 2023;55(14s):1–37. doi:10.1145/3592797.
4. Ghiasi M, Wang Z, Niknam T, Dehghani M, Ansari HR. Cyber-physical security in smart power systems from a resilience perspective: concepts and possible solutions. In: Power systems cybersecurity: methods, concepts, and best practices. Cham, Switzerland: Springer International Publishing, 2023. p. 67–89.
5. An D, Zhang F, Yang Q, Zhang C. Data integrity attack in dynamic state estimation of smart grid: attack model and countermeasures. IEEE Trans Autom Sci Eng. 2022;19(3):1631–44. doi:10.1109/TASE.2022.3149764.
6. Elsisi M, Altius M, Su S-F, Su C-L. Robust kalman filter for position estimation of automated guided vehicles under cyberattacks. IEEE Trans Instrum Meas. 2023;72:1–12. doi:10.1109/TIM.2023.3250285.

7. Zhang J, Zhou J, Gu Z, Zhang Z, Wang L, Yu ZL, et al. CS-LeCT: chained secure and low-energy consumption data transmission based on compressive sensing. IEEE Trans Instrum Meas. 2023;72:1–10. doi:10.1109/TIM.2023.3280495.

8. Wang Z, He H, Wan Z, Sun Y. Coordinated topology attacks in smart grid using deep reinforcement learning. IEEE Trans Ind Inform. 2020;17(2):1407–15. doi:10.1109/TII.2020.2994977.

9. Selvam R, Tyagi A. Residue number system (RNS) and power distribution network topology-based mitigation of power side-channel attacks. Cryptography. 2023;8(1):1. doi:10.3390/cryptography8010001.

10. Ning C, Xi Z. Improved stealthy false data injection attacks in networked control systems. IEEE Syst J. 2024;18(1):505–15. doi:10.1109/JSYST.2024.3350179.

11. Guo H, Sun J, Pang Z-H, Liu G-P. Event-based optimal stealthy false data-injection attacks against remote state estimation systems. IEEE Trans Cybern. 2023;53(10):6714–24. doi:10.1109/TCYB.2023.3255583.

12. Bhattar PL, Pindoriya NM. False data injection attack with max-min optimization in smart grid. Comput Secur. 2024;140:103761. doi:10.1016/j.cose.2024.103761.

13. Habib AA, Hasan MK, Alkhayyat A, Islam S, Sharma R, Alkwai LM. False data injection attack in smart grid cyber physical system: issues, challenges, and future direction. Comput Electr Eng. 2023;107:108638.

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

15. Wei S, Xu J, Wu Z, Hu Q, Yu X. A false data injection attack detection strategy for unbalanced distribution networks state estimation. IEEE Trans Smart Grid. 2023;14(5):3992–4006.

16. Qu Z, Yang J, Wang Y, Georgievitch PM. Detection of false data injection attack in power system based on hellinger distance. IEEE Trans Ind Inform. 2023;20(2):2119–28. doi:10.1109/TII.2023.3286895.

17. Shen Y, Qin Z. Detection, differentiation and localization of replay attack and false data injection attack based on random matrix. Sci Rep. 2024;14(1):2758. doi:10.1038/s41598-024-52954-z.

18. James J, Hou Y, Li VO. Online false data injection attack detection with wavelet transform and deep neural networks. IEEE Trans Ind Inform. 2018;14(7):3271–80. doi:10.1109/TII.2018.2825243.

19. Habibi MR, Baghaee HR, Dragičević T, Blaabjerg F. Detection of false data injection cyber-attacks in DC microgrids based on recurrent neural networks. IEEE J Emerg Sel Top Power Electron. 2020;9(5): 5294–310. doi:10.1109/JESTPE.2020.2968243.

20. Lu K-D, Zhou L, Wu Z-G. Representation-learning-based CNN for intelligent attack localization and recovery of cyber-physical power systems. IEEE Trans Neural Netw Learn Syst. 2023;35(5):6145–55. doi:10.1109/TNNLS.2023.3257225.

21. Wang Y, Zhang Z, Ma J, Jin Q. KFRNN: an effective false data injection attack detection in smart grid based on kalman filter and recurrent neural network. IEEE Internet Things J. 2021;9(9):6893–904. doi:10.1109/JIOT.2021.3113900.

22. Ayad A, Khalaf M, Salama M, El-Saadany EF. Mitigation of false data injection attacks on automatic generation control considering nonlinearities. Elect Power Syst Res. 2022;209(6):107958. doi:10.1016/j.epsr.2022.107958.

23. Boyaci O, Narimani MR, Davis KR, Ismail M, Overbye TJ, Serpedin E. Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks. IEEE Trans Smart Grid. 2021;13(1):807–19. doi:10.1109/TSG.2021.3117977.

24. Li X, Wang Y, Lu Z. Graph-based detection for false data injection attacks in power grid. Energy. 2023;263(5):125865. doi:10.1016/j.energy.2022.125865.

25. Su X, Deng C, Yang J, Li F, Li C, Fu Y, et al. Damgat based interpretable detection of false data injection attacks in smart grids. IEEE Trans Smart Grid. 2024;15(4):4182–95. doi:10.1109/TSG.2024.3364665.

26. Zhang G, Li J, Bamisile O, Cai D, Hu W, Huang Q. Spatio-temporal correlation-based false data injection attack detection using deep convolutional neural network. IEEE Trans Smart Grid. 2021;13(1):750–61. doi:10.1109/TSG.2021.3109628.

27. Han Y, Feng H, Li K, Zhao Q. False data injection attacks detection with modified temporal multi-graph convolutional network in smart grids. Comput Secur. 2023;124:103016. doi:10.1016/j.cose.2022.103016.

28. Feng H, Han Y, Si F, Zhao Q. Detection of false data injection attacks in cyber-physical power systems: an adaptive adversarial dual autoencoder with graph representation learning approach. IEEE Trans Instrum Meas. 2024;73:1–11. doi:10.1109/TIM.2023.3331398.

29. Li F, Shen W, Bi Z, Su X. Sparse adversarial learning for FDIA attack sample generation in distributed smart srids. Comput Model Eng Sci. 2024;139(2):2095–115. doi:10.32604/cmes.2023.044431.

30. Zhao S, Peng R, Hu P, Tan L. Heterogeneous network embedding: a survey. Comput Model Eng Sci. 2023;137(1):83–130. doi:10.32604/cmes.2023.024781.

31. Liao L, Hu Z, Zheng Y, Bi S, Zou F, Qiu H, et al. An improved dynamic Chebyshev graph convolution network for traffic flow prediction with spatial-temporal attention. Appl Intell. 2022;52(14):16104–116. doi:10.1007/s10489-021-03022-w.

32. Liu X, Zhou H, Guo K, Li C, Zu S, Wu L. Quantitative characterization of shale gas reservoir properties based on BiLSTM with attention mechanism. Geosci Front. 2023;14(4):101567. doi:10.1016/j.gsf.2023.101567.

33. Rahim MA, Farid FA, SalehMusaMiah A, Puza AK, Alam MN, Hossain MN, et al. An enhanced hybrid model based on CNN and BiLSTM for identifying individuals via handwriting analysis. Comput Model Eng Sci. 2024;140(2):1689–710. doi:10.32604/cmes.2024.048714.

34. Dao F, Zeng Y, Qian J. Fault diagnosis of hydro-turbine via the incorporation of bayesian algorithm optimized CNN-LSTM neural network. Energy. 2024;290:130326. doi:10.1016/j.energy.2024.130326.

35. Wu Y, Zu T, Guo N, Zhu Z, Li F. Laplace-domain hybrid distribution model based FDIA attack sample generation in smart grids. Symmetry. 2023;15(9):1669. doi:10.3390/sym15091669.