symmetry

MDPI

Article

Data-Driven Attack Detection Mechanism Against False Data Injection Attacks in DC Microgrids Using CNN-LSTM-Attention

Chunxiu Li [1], Xinyu Wang [1,2], Xiaotao Chen [1,*], Aiming Han [1] and Xingye Zhang [1]

[1] School of Energy and Electrical Engineering, Qinghai University, Xining 810016, China; ys240858010496@qhu.edu.cn (C.L.); wxyzmya@ysu.edu.cn (X.W.); ys240858080516@qhu.edu.cn (A.H.); ys240858080513@qhu.edu.cn (X.Z.)
[2] School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China
* Correspondence: chenxiaotao@qhu.edu.cn

Abstract

This study presents a novel spatio-temporal detection framework for identifying False Data Injection (FDI) attacks in DC microgrid systems from the perspective of cyber–physical symmetry. While modern DC microgrids benefit from increasingly sophisticated cyber–physical symmetry network integration, this interconnected architecture simultaneously introduces significant cybersecurity vulnerabilities. Notably, FDI attacks can effectively bypass conventional Chi-square detector-based protection mechanisms through malicious manipulation of communication layer data. To address this critical security challenge, we propose a hybrid deep learning framework that synergistically combines: Convolutional Neural Networks (CNN) for robust spatial feature extraction from power system measurements; Long Short-Term Memory (LSTM) networks for capturing complex temporal dependencies; and an attention mechanism that dynamically weights the most discriminative features. The framework operates through a hierarchical feature extraction process: First-level spatial analysis identifies local measurement patterns; second-level temporal analysis detects sequential anomalies; attention-based feature refinement focuses on the most attack-relevant signatures. Comprehensive simulation studies demonstrate the superior performance of our CNN-LSTM-Attention framework compared to conventional detection approaches (CNN-SVM and MLP), with significant improvements across all key metrics. Namely, the accuracy, precision, F1-score, and recall could be improved by at least 7.17%, 6.59%, 2.72% and 6.55%.

Keywords: DC microgrid system; false data injection attacks; spatio-temporal model; CNN-LSTM-Attention; cyber-physical symmetry system

Academic Editor: Xiaogang Qi

Received: 15 June 2025
Revised: 12 July 2025
Accepted: 14 July 2025
Published: 16 July 2025

Citation: Li, C.; Wang, X.; Chen, X.; Han, A.; Zhang, X. Data-Driven Attack Detection Mechanism Against False Data Injection Attacks in DC Microgrids Using CNN-LSTM-Attention. *Symmetry* **2025**, *17*, 1140. https://doi.org/10.3390/sym17071140

1. Introduction

With the advancement of renewable energy sources, such as photovoltaics and wind power, along with energy storage technology, Direct Current (DC) microgrids have emerged as an important development direction for modern power systems due to their high efficiency and flexibility [1–3]. Compared to AC microgrids, DC microgrids avoid frequent AC/DC conversion, thereby reducing energy loss and being more suitable for the plug-and-play integration of distributed power sources [4]. They have demonstrated significant advantages in remote areas for power supply, as well as in data centers and electric vehicle charging stations [5,6]. However, as the cyber–physical symmetry scale of DC microgrids expands and their intelligence improves, the network security threats they face are becoming increasingly severe.

To safeguard the security of DC microgrids, Kalman filter (KF)-based state estimation is a widely employed detection technique for countering deceptive attacks [7,8]. Given that the current security framework in cyber–physical symmetry power systems (CPPS) hinges on least-squares state estimation, any discrepancies between the estimated and actual states may lead to erroneous logical decisions within the Supervisory Control and Data Acquisition (SCADA) system [9]. Currently, there are some typical deceptive attacks, such as Denial of Service (DoS) attacks and False Data Injection (FDI) attacks [10]. Compared to DoS attacks, FDI attacks are more deceptive in nature. They can disguise physical layer changes by tampering with information layer data, thus deceiving existing security mechanisms. In [11], a covert FDIA detection approach was introduced, relying on the system's estimated outputs derived from an artificial neural network. Hou et al. revealed the important influence of the attack function on the effectiveness and covertness of the attack [12]. As indicated in [11,12], with knowledge of the full or partial topology of DC microgrids, an attacker can tamper with smart meter data while keeping the associated measurement residuals unchanged. Therefore, developing a novel attack detection mechanism against FDI attacks is urgent to ensure the security of DC microgrids.

To reduce the susceptibility of DC microgrids to cyber threats, various cyber-attack detection techniques have been proposed from the perspective of cyber–physical symmetry. These techniques can generally be categorized into model-based and model-free approaches [13]. Model-based detection methods depend on the precision of the system model. For example, a robust cyber-attack detection scheme was proposed for DC microgrid systems by utilizing a parity-based method [14]. In [15], a distributed monitoring scheme was proposed to provide attack-detection capabilities for linear DC microgrid systems, which relies on a bank of Luenberger observers. An attack detection and mitigation method based on model predictive control and artificial neural networks was developed in DC microgrids [16]. In [17], Wu et al. developed a dual-observer-based detector to detect false data injection attacks and isolate the compromised distributed generation units for DC microgrid systems. In [18], the concept of a zero-trace stealth attack was initially proposed to destabilize DC microgrids. To counteract this, an attack detection method utilizing a dynamic average consensus estimator was subsequently developed. To mitigate this vulnerability caused by openness, Taher et al. developed an innovative approach that employs a nonlinear autoregressive network with an exogenous inputs observer [19]. In [20], Xiahou et al. formulated a model-driven detection technique to counter cyber–physical attacks. This technique comprises state and attack observers. To maintain the robustness of the Load Frequency Control system, a model-based approach leveraging unknown input observers was devised to identify and alleviate deceptive attacks [21]. In [22], a distributed non-fragile controller with coupled memory delay against attacks was designed. However, implementing these methods in real-world applications is often difficult because they inevitably fail to accurately represent the intricacies of actual power electronic systems. Conversely, model-free methods rely solely on measurements and do not require prior knowledge of the system. For example, an innovative approach that combines Short-Time Fourier Transform with an Artificial Neural Network was devised to enhance detection performance against cyber-attacks in DC microgrids [23]. In [24], an attack detection and mitigation approach used Artificial Intelligence (AI) against FDI attacks in a DC microgrid, based on an artificial neural network. Additionally, Habibi et al. proposed a method rooted in a recurrent neural network for the detection of FDI attacks in DC microgrids [25]. A combined approach leveraging machine learning techniques was proposed for detecting and alleviating the threat posed by FDI attacks, with a specific focus on attacks targeting converters within a DC microgrid environment [26]. The suggested methodology entails employing deep learning algorithms to identify attacks, relying on time-series sensor data

gathered under diverse circumstances, encompassing load fluctuations and standard operational states. A method for detecting and correcting cyber-attacks in DC microgrids based on a dual deep neural network was introduced in [27]. An enhanced deep learning approach utilizing an elective group structure combined with the Krill Herd Optimization algorithm was proposed in [28]. However, the existing learning-based detection techniques overlook the spatio-temporal aspect of attack detection.

To address the aforementioned challenges, this paper proposes a novel spatial-temporal detection framework for identifying FDI attacks in DC microgrids. As illustrated in Figure 1, the proposed framework integrates three key components: convolutional neural networks (CNN), Long Short-Term Memory (LSTM) networks, and an attention mechanism. The CNN component is designed to effectively capture spatial correlations within the measured data, while the LSTM network extracts temporal dependencies across time series. Furthermore, the incorporated attention mechanism enhances the model's capability to focus on the most salient features, thereby improving overall detection performance. This synergistic combination of spatial and temporal feature extraction, augmented by attention-based feature weighting, enables comprehensive and accurate FDA detection in cyber–physical symmetry DC microgrid systems. The main contributions can be summarized as follows.

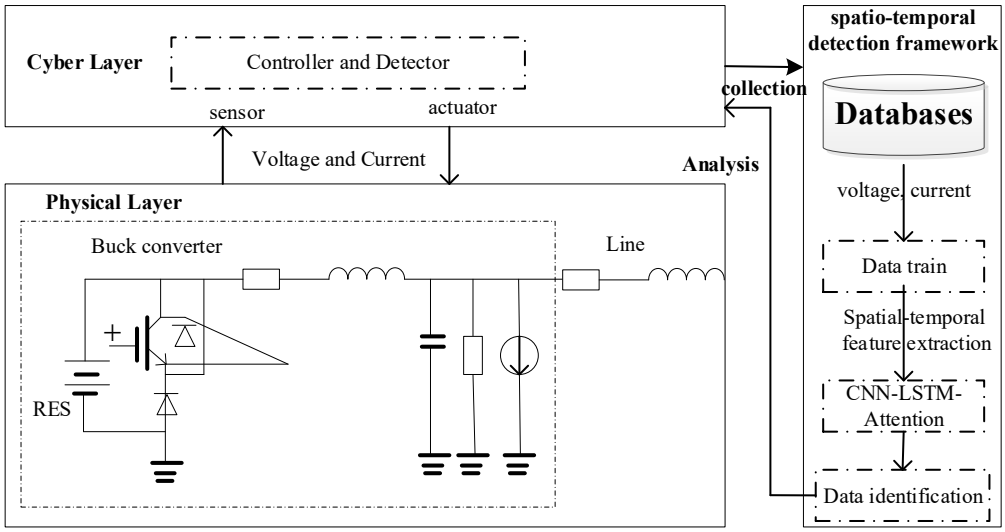


Figure 1. Detection framework of FDI attacks using spatial-temporal features.

1. Novel Spatial-Temporal Detection Framework: This paper proposes a hybrid deep learning model integrating CNN, LSTM, and attention mechanisms to simultaneously capture spatial correlations and temporal dependencies in DC microgrid data. Unlike existing methods that focus solely on temporal or spatial features, our approach provides a more comprehensive detection mechanism against FDI attacks.
2. Enhanced Feature Extraction via Attention Mechanism: By incorporating an attention mechanism, the proposed model dynamically weights the importance of different spatial and temporal features, improving detection accuracy and robustness.
3. Simulation results validate the superior detection capability of our CNN-LSTM-Attention framework, demonstrating statistically significant improvements over conventional CNN-SVM [29] and MLP [30] approaches across all evaluation metrics. Namely, the accuracy, precision, F1-score, and recall could be at least improved 7.17%, 6.59%, 2.72% and 6.55%.

The outline of this article is given below. Section 2 gives the problem formulation for DC microgrids under deceptive attacks. In Section 3, a data-driven detection model using

CNN-LSTM-Attention is established. Simulation tests demonstrate the superiority of the proposed method in Section 4. Conclusions and discussion are given in Section 5.

2. Covert Characteristics of FDI Attacks in Cyber–Physical Symmetry DC Microgrids

In this section, a physical dynamic model for DC microgrids is developed. Subsequently, analysis of the covert features of FDI attacks is conducted. Finally, the challenge of detecting FDI attacks is formally addressed. Additionally, related parameters are presented, as shown in Table 1.

Table 1. Definition of all parameters.

I_i	ith filter current;	ε_1	parameter variations of the capacitor;
V_i	ith filter voltage;	ε_2	parameter variations of the inductance;
T_s	sample time;	R_{ij}	resistances of the power lines;
C_i	ith filter capacitor;	n_i	neighbor of ith MG;
L_i	ith filter inductance;	I_{Li}	equivalent current load;
R_i	ith filter resistance;	R_{Li}	equivalent impedance load;
w_i	external disturbance;	V_{ti}	voltage command of the converter;
$z(t)$	measurement output;	$z_f(t)$	measurement output under attack;
τ	precomputed threshold;	$y_f(t)$	estimation output under attack;
$s(t)$	state change caused by attack;	$\hat{x}_f(t)$	estimation state under attack;
H	Jacobian matrix;	$f(t)$	false attack vector

2.1. Physical Dynamic DC Microgrids Model

As shown in Figure 1, the physical dynamic model for the ith, $i \in (1, \ldots, n)$ of a DC microgrid can be described as [14]:

$$\begin{cases} \dot{V}_i(t) = \left(1 - \dfrac{T_s}{\varepsilon_1 C_i R_{Li}}\right) V_i(t) + \dfrac{T_s}{\varepsilon_1 C_i} I_i(t) - \dfrac{T_s}{\varepsilon_1 C_i} I_{li}(t) + \sum\limits_{j \in n_i} T\left(\dfrac{V_j(t) - V_i(t)}{\varepsilon_1 C_i R_{ij}}\right) \\ I_i(t) = -\dfrac{T_s}{\varepsilon_2 L_i} V_i(t) + \left(1 - \dfrac{T_s R_i}{\varepsilon_2 L_i}\right) I_i(t) + \dfrac{T_s}{\varepsilon_2 L_i} V_{ti}(t) \end{cases} \tag{1}$$

By linearizing Equation (1), a linear DC microgrid model can be obtained:

$$\dot{x}_i(t) = Ax_i(t) + Bu_i(t) \tag{2}$$

With

$$x_i(t) = [V_i(t), I_i(t)]^T$$

$$A = \begin{bmatrix} 1 & \frac{T_s}{C_i} \\ \frac{T_s}{L_i} & 1 - \frac{T_s R_i}{L_i} \end{bmatrix}, \; B = \begin{bmatrix} 0 \\ \frac{T_s}{L_i} \end{bmatrix}$$

Then, the $ith, i \in (1, \ldots, n)$ DC microgrid model considering noise in Equation (2) can be obtained:

$$\begin{cases} \dot{x}_i(t) = Ax_i(t) + Bu_i(t) + w_i(t) \\ y_i(t) = Cx_i(t) \end{cases} \tag{3}$$

where $C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

2.2. Covert Characteristics of FDI Attacks

As a representative cyber–physical symmetry system, the DC microgrid facilitates the integrated regulation of control and data flows. However, its inherent openness and intelligent features also expose it to potential cyber–physical attacks, which may compromise system stability. To address this vulnerability, a state estimation-based detection mechanism is employed, with the following decision criterion [31]:

$$\begin{cases} \|z(t) - Cx(t)\| < \tau & \text{Normal} \\ \|z(t) - Cx(t)\| \geq \tau & \text{Abnormal} \end{cases} \tag{4}$$

It is critical that the threshold τ configuration guarantees a less than 5% false alarm probability [31]. By reverse-engineering the detection algorithm's logic, adversaries can generate maliciously fabricated data as follows:

$$\begin{aligned} \eta_f(t) &= \left\| z_f(t) - H\hat{x}_f(t) \right\| \\ &= \|(z(t) + f(t)) - (H\hat{x}(t) + Hs(t))\| \\ &= \|(z(t) - H\hat{x}(t)) + (f(t) - Hs(t))\| \\ &\leq \tau \end{aligned} \tag{5}$$

Equation (5) indicates that adversaries can systematically construct false data sequences that comply with the specified detection constraints $f(t) = Hs(t)$. Specifically, by carefully manipulating measurement data to maintain the estimated residual within the predetermined threshold bounds, attackers can effectively bypass conventional χ^2-based detection systems. Consequently, such stealthy false data injection attacks can covertly modify DC microgrid system operating states while avoiding alarm triggers, potentially initiating catastrophic cascading failures across the DC microgrid.

Based on the above description, the ith dynamics of the DC microgrid state model under FDI attacks can be described as

$$\begin{cases} \dot{x}_i(t) = Ax_i(t) + Bu_i(t) + w_i(t) \\ y_i(t) = Cx_i(t) + f_i(t) \end{cases} \tag{6}$$

where $f_i(t)$ is the injected deceptive attack, which is norm-bounded.

To address the security risks to the DC microgrid system arising from FDI attacks, this paper presents a novel spatial-temporal detection framework for detecting FDI attacks in DC microgrids.

3. Methods

This section presents a spatial-temporal detection framework designed to counter FDI attacks in a cyber–physical symmetry DC microgrid system, as shown in Figure 2. The proposed framework integrates CNN-based spatial feature extraction with LSTM-based temporal feature extraction. Additionally, an attention mechanism is introduced to enhance the detection performance of the model. The detailed structure is described below.

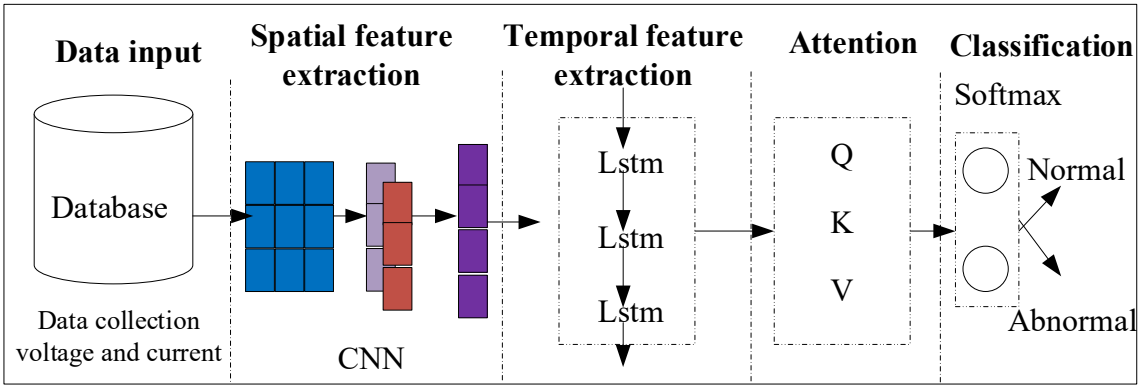


Figure 2. Spatial-temporal detection framework for detecting FDI attacks in DC microgrids.

3.1. Spatial Feature Extraction Using CNN

To capture spatial features from the DC microgrid dataset, a CNN architecture was established. As depicted in Figure 3, the constructed CNN model consists of an input layer, a convolutional layer, a pooling layer, and a fully connected layer. The input layer takes in the input data, which include information on both normal and abnormal states of the DC microgrid system. The convolutional and pooling layers play a crucial role in distilling spatial attributes from the input data. Finally, the fully connected layer handles and delivers the extracted features.

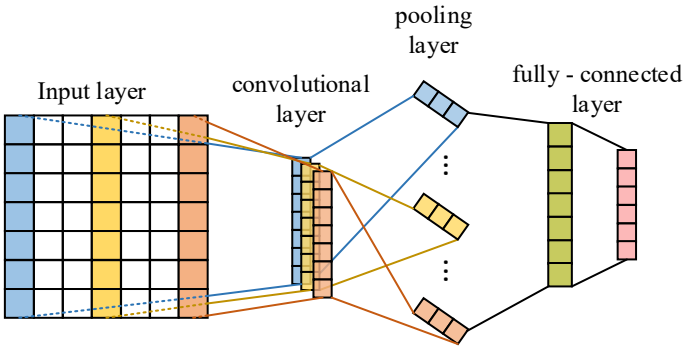


Figure 3. Constructed CNN model.

The convolutional layer performs local feature extraction by systematically sliding a convolutional kernel across the input feature map. At each spatial position, the kernel conducts weighted operations with the underlying receptive field, thereby encoding the intrinsic structural patterns of the input data [11]. The formal representation of the discrete convolution operation is given by

$$y^k = \sum_{i=1}^{j^{k-1}} w_{i,j}^k \bullet x_i^{k-1} + b_i^k \tag{7}$$

where y^k represents the output of k layer; x_i^{k-1} represents the output of the kth channel of the $k-1$ layer; j^{k-1} represents the jth channel of layer $k-1$; $w_{i,j}^k$ and b_i^k stand for weight and offset, respectively.

The pooling layer serves primarily to downsample feature maps, significantly reducing computational complexity while preserving the essential features of the input signals. By aggregating local activations, it compresses spatial dimensions, thereby enhancing

computational efficiency and improving feature robustness. The mathematical formulation of this operation is expressed as follows:

$$y^{k(i,n)} = \frac{1}{G} \sum_{m=(n-1)G+1}^{nG} x^{k(i,m)} \tag{8}$$

where $x^{k(i,m)}$ represents the value of the mth neuron of the ith channel in the kth layer; G indicates the size of the pooled kernel; $y^{k(i,n)}$ represents the value of the nth neuron of the ith channel in the first layer.

The fully connected layer integrates high-level feature representations derived from the preceding convolutional and pooling operations. By establishing global connections across all activations, it enables comprehensive feature learning and facilitates decision-making in downstream tasks. The mathematical formulation of this transformation is given as follows:

$$\phi_i = \mathrm{ReLU}\left(\kappa_i * y^{k(i,n)} + \varphi_i\right) \tag{9}$$

where φ_i denotes the bias term in fully-connected layer, and κ_i is the weight value.

Remark 1. *Leveraging the sliding window mechanism of convolutional kernels, CNNs effectively capture spatial correlations in operational parameters—including voltage/current—while hierarchically extracting features from device-level micro-characteristics to system-level macro-patterns through deep convolutional stacking. The management center capitalizes on CNN's parameter-sharing property to significantly reduce model complexity when processing high-dimensional spatiotemporal data. Additionally, the inherent translation invariance of CNNs ensures robust detection of attack signal variations, regardless of positional shifts in the input data.*

3.2. Temporal Feature Extraction Using LSTM

We introduced an LSTM-based temporal feature extraction framework incorporating three core components: (1) an update gate that modulates memory retention, (2) a forget gate that controls historical information discarding, and (3) a cell state that maintains long-term temporal dependencies. This architecture enables selective processing of sequential patterns across different time scales. The cell structure of the LSTM network is illustrated in Figure 4.

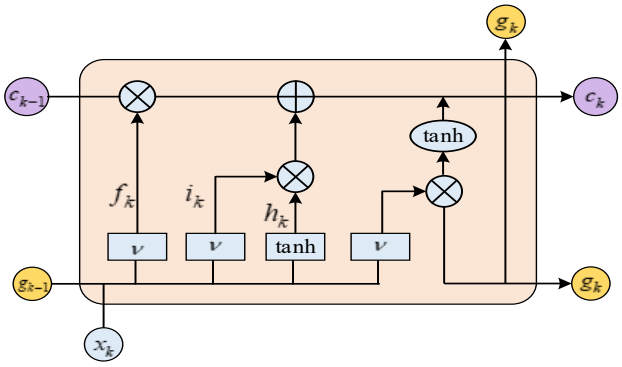


Figure 4. Constructed LSTM model.

Prior to the information entering the cells of an LSTM network, it first traverses the forget gate. This gate is primarily employed to selectively discard information that was transmitted during the previous time step. The formula governing its update is provided below.

$$h_k = \nu(V_h \bullet [g_{k-1}, x_k] + b_h) \tag{10}$$

where h_k denotes the output of the forget gate, ν is the activation function, g_{k-1} represents the output of the previous cell, x_k represents the input of the current cell, b_h denotes the bias term, and V_h is the weight value.

After being filtered by the forget gate, the information is transmitted to the input gate. The LSTM network input signal determines the information to be updated and the content to be updated through the input gate. The update formula is given below:

$$\begin{cases} i_k = \nu(V_i \bullet [g_{k-1}, x_k] + b_i) \\ \hat{c}_k = \tanh(V_c \bullet [g_{k-1}, x_k] + b_c) \\ c_k = c_{k-1} \bullet h_k + \hat{c}_k \bullet i_k \end{cases} \tag{11}$$

where i_k denotes the output of the input gate, V_i and V_c are the weight values, b_i and b_c are the bias terms, tanh is the activation function, $\hat{c}_k$ represents the state quantity that integrates the state information of g_{k-1} and x_k, and c_k represents the current state at the moment.

The output gate uses σ to calculate the state of the unit to be output. The input signal determines the useful information for output through the output gate and immediately updates the current LSTM neural network cell state. The update formula is

$$\begin{cases} o_k = \nu(V_o \bullet [g_{k-1}, x_k] + b_o) \\ g_k = o_k \tanh(c_k) \end{cases} \tag{12}$$

where V_o is the weight value, b_o is the bias term, and o_k denotes the output of the output gate.

Remark 2. *CNN can exhibit insensitivity towards the temporal attributes inherent in time-series data. In contrast, LSTM networks are more adept at seamlessly integrating temporal characteristics from various partial states. Thanks to their distinctive architectural design, LSTMs are capable of effectively addressing the issue of vanishing gradients that often plagues the training process in recurrent neural networks.*

3.3. Enhancing the Detection Performance Using Attention

The attention mechanism dynamically calculates weights to allocate attention to different parts of the input data, enabling the model to accurately focus on key features that have the greatest impact on classification results, as shown in Figure 5. It effectively integrates global information, captures dependencies between features, alleviates the problem of long-distance dependencies in long sequence processing, and improves the computational efficiency and interpretability of the model. Ultimately, it enhances the model's ability to capture important information, significantly improving classification accuracy and robustness.

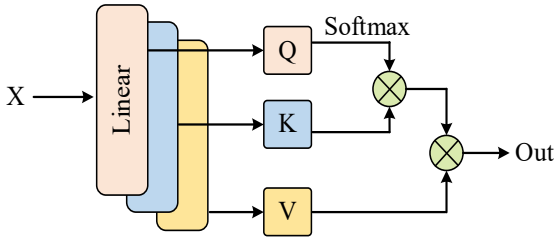


Figure 5. Constructed attention model.

The mathematical expression for the self-attention mechanism is given as follows:

$$Y_{pre} = Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{13}$$

where Q, K, V denote the query vectors, key vectors, and value vectors, respectively; d_k is the dimension of the key vector used for normalization.

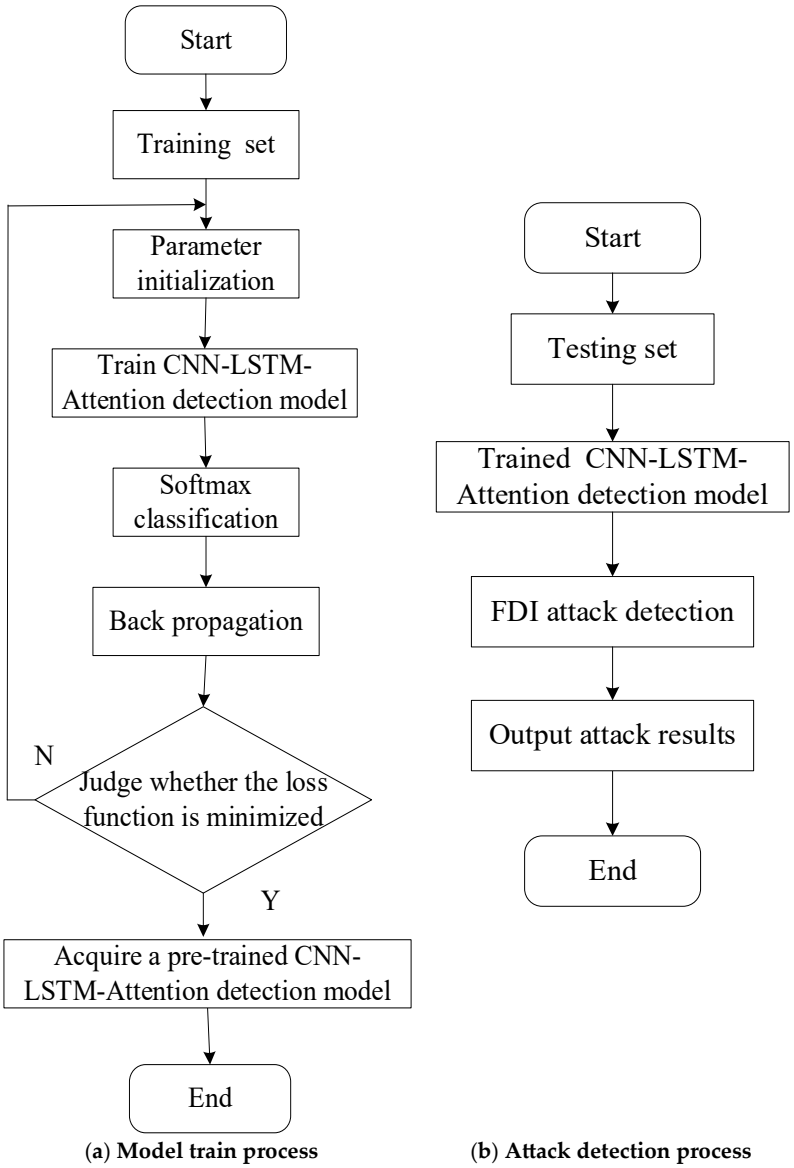
Since the detection of FDI attacks is formulated as a binary classification problem, the cross-entropy loss function is defined as:

$$Loss = -\frac{1}{T}\sum_{i=1}^{T}\left[Y_{tru}\log\left(Y_{pre}\right) + \left(1 - Y_{tru}\right)\log\left(1 - Y_{pre}\right)\right] \tag{14}$$

where Y_{tru} denotes the true label category.

3.4. Detection Framework Based on CNN-LSTM-Attention

A detection framework leveraging spatial-temporal features was developed using a CNN-LSTM-Attention model, as illustrated in Figure 6. The detailed workflow consists of the following steps:



(**a**) **Model train process** (**b**) **Attack detection process**

Figure 6. FDI attack detection flowchart.

Data Collection: Acquire measurement DC microgrid data (e.g., voltage and current) to form the dataset.

Data Partitioning: Split the DC microgrid dataset into training and testing subsets for model development and evaluation.

Model Construction: Build the CNN-LSTM-Attention model, where: CNN extracts spatial features from input data using Equations (7)–(9), LSTM captures temporal dependencies using Equations (10)–(12), and attention mechanism enhances critical feature representation using Equation (13).

Offline Training: Train the CNN-LSTM-Attention model using the training set to optimize detection performance using Equation (14).

Online Detection: Deploy the trained model to classify real-time measurements as normal or abnormal (FDI-attacked).

Based on this framework, the detection algorithm against False Data Injection (FDI) attacks in DC microgrid system is formally outlined in Algorithm 1.

Algorithm 1: Data-driven FDI attack detection using CNN-LSTM-Attention

Input: Data set including train and test set
Output: Output label of test data Y_{pre}
1. Divide the dataset into train and test data
2. Construct the CNN-LSTM-Attention Detection Model
3. Train the LSTM-Attention Detection Model
4. **for** i $\in$ [1, Total Epochs]
5. **for** j $\in$ [1, Number of Training Data]
6. Compute the predicted attack probability in Equation (13);
7. **end for**
8. Compute the loss function in Equation (14);
9. **If** the loss function is not minimized
10. Update training model parameters;
11. **else**
12. End training
13. **end for**
14. Detect the FDI attacks using the pre-trained CNN-LSTM-Attention Detection Model
15. **if** Predict Attacked Probability > 1 − Predict Attacked Probability
16. Output Y_{pre} − Test = Abnormal;
17. **else**
18. Output Y_{pre} − Test = normal;
19. **end if**
20. **return** Y_{pre}

3.5. Discussion on the Application of CNN-LSTM-Attention in Practical DC Microgrid Systems

The detection framework proposed in this paper can extract key features from DC microgrid data under attack by introducing an attention mechanism, thereby reducing the computational complexity of the detection model. With regard to latency considerations, future work based on this paper will explore hardware–collaborative design solutions to optimize latency, such as solidifying CNN computing cores on FPGAs or employing double-buffering techniques: synchronously collecting data for the next window while processing the current one.

The proposed model addresses challenges related to computational constraints and scalability through a dynamic computational resource allocation mechanism. For instance, it employs an adaptive attention gating mechanism: leveraging Gated Recurrent Units

(GRUs) to dynamically adjust attention weights, it shuts down 80% of the LSTM units during steady-state operation (retaining only baseline monitoring functionality) while activating the entire network in the event of transient faults. Additionally, it utilizes a parallelized feature extraction approach: the CNN branch processes raw voltage/current signals, while the LSTM branch handles frequency-domain features compressed via Fast Fourier Transform (FFT), with final feature fusion achieved through an attention layer.

When applying the proposed method to AC power grids, several additional factors warrant careful consideration. Firstly, input preprocessing steps such as dq transformation and harmonic separation are essential to ensure data compatibility and enhance feature extraction. Secondly, the attention mechanism should be extended to incorporate dual time-frequency pathways, enabling it to capture both temporal and spectral characteristics effectively. Moreover, dynamic computational resource allocation is crucial for optimizing performance under varying operational conditions. When scaling the proposed method to large-scale power grids, two key challenges emerge. One is managing the model's computational complexity, which can escalate rapidly with grid size and complexity. The other is implementing a "hierarchical decoupling–dynamic collaboration" strategy, which involves dividing the grid into manageable subsystems while ensuring seamless coordination among them to maintain overall system stability and performance.

4. Results

In this section, simulation experiments were conducted to evaluate the effectiveness of the proposed CNN-LSTM-Attention detection framework in countering FDI attacks. In comparison with established detection methods, including CNN-SVM [29] and Multilayer Perceptron (MLP) [30], the tests performed on a cyber–physical symmetry DC microgrid system comprising four distributed generation units (DGUs) validated the superior performance of the introduced detection model.

4.1. Simulation and Data Setup

The simulation environment in this paper was realized on a notebook computer based on PyCharm 2024.3.4d with the following configuration: i7-12700H, 2.30 GHz, 16G RAM, and Intel Iris Xe Graphics GPU. CNN: Number of convolutional layers: 2; Convolutional kernel size: 3×3; Activation function: ReLU. LSTM: Number of hidden units: 128; Number of layers: 1; Dropout rate: 0.3. Attention mechanism: Attention dimension: 32. Optimizer: Adam; Learning rate: 0.001; gamma = 0.5.

In alignment with the methodology outlined in [23], we utilized MATLAB/Simulink to generate the dataset, sampled at five-minute intervals. The collected data were then normalized according to the operational capacities of the four-DGU DC microgrid system under study. Finally, the dataset was partitioned into training and testing subsets for model development and evaluation.

4.2. Evaluation Indicators

To assess the effectiveness of the detection framework, key performance metrics—including Accuracy (*A*), Missed Alarm Rate (*MAR*), Precision (P), F_1-Score (*FS*), and Recall Detection Rate (*RDR*)—were adopted [23]. The associated mathematical formulations are provided below.

$$A = \frac{\Lambda_{TN} + \Lambda_{TP}}{\Lambda_{TN} + \Lambda_{TP} + \Lambda_{FN} + \Lambda_{FP}} \tag{15}$$

$$\mathrm{P} = \frac{\Lambda_{TP}}{\Lambda_{TP} + \Lambda_{FP}} \tag{16}$$

$$RDR = \frac{\Lambda_{TP}}{\Lambda_{TP} + \Lambda_{FN}} \tag{17}$$

$$FS = \frac{2\mathrm{P} \times RDR}{\mathrm{P} + RDR} \tag{18}$$

where Λ_{FN} represents the count of normal samples misclassified as anomalous, Λ_{FP} represents the count of anomalous samples misclassified as normal, Λ_{TP} represents the count of normal samples classified as normal, Λ_{TN} denotes the count of abnormal samples classified as abnormal, and A, MAR, P, Ra and FS denote the accuracy, missed alarm, precision, recall and F_1-Score, respectively.

4.3. Ablation Analysis

To validate the contribution of each component to the proposed CNN-LSTM-Attention model, we conducted a comprehensive ablation study by systematically removing key modules and evaluating performance degradation. The experiments were designed as follows:

Baseline (CNN-LSTM): Remove the attention mechanism, retaining only CNN for spatial feature extraction and LSTM for temporal modeling.

CNN-Only: Remove both LSTM and attention, using solely CNN layers.

LSTM-Only: Replace CNN with a flattening layer, keeping LSTM and attention.

CNN-LSTM-Attention (Full Model): The complete proposed framework.

Based on the above ablation experimental framework, the comparative outcomes, including of key performance metrics (A, FS, P) and rate of change (ROC), are displayed, as illustrated in Table 2 and Figure 7.

Table 2. Comparison results of ablation analysis on CNN-LSTM-Attention detection model.

Model	*A*	P	*RDR*	*FS*
CNN-LSTM-Attention	98.59%	97.81%	99.50%	98.65%
CNN-LSTM	94.11%	90.13%	99.50%	94.58%
CNN	80.28%	72.48%	99.75%	83.95%
LSTM	87.32%	80.93%	98.76%	88.96%

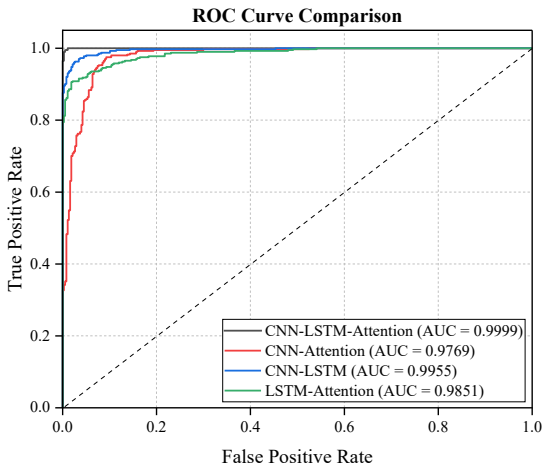


Figure 7. ROC curves of ablation analysis on CNN-LSTM-Attention detection model.

In the simulation, the full CNN-LSTM-Attention model achieved the highest accuracy of 98.59%. This remarkable performance can be attributed to the well-designed architecture that combines the strengths of CNN, LSTM, and the attention mechanism. When comparing it with the CNN-LSTM baseline (without attention), we observed a significant drop of

4.48% in accuracy. This drop clearly indicates that the attention mechanism plays a crucial role in enhancing the model's performance. In the context of false data injection detection, the attention mechanism effectively suppresses noise during spatio-temporal feature fusion. False data injection often introduces noise into the data, and the attention mechanism helps the model focus on the most relevant features, thereby improving the accuracy of detecting such malicious injections.

The LSTM-Only model suffered the most severe degradation in accuracy, with a decrease of 11.27% compared to the full model. This reveals that spatial feature extraction via CNN is of utmost importance for raw signal preprocessing. In false data injection scenarios, the raw signals may contain distorted or manipulated spatial information. The CNN component could extract and correct these spatial features to some extent, providing a more reliable input for the subsequent LSTM and attention modules. Without CNN-based spatial feature extraction, the LSTM-Only model struggled to accurately detect false data injections, highlighting the critical role of spatial preprocessing.

The attention mechanism also had a positive impact on precision and recall. It improved precision while maintaining a high recall rate. This demonstrates its ability to focus on discriminative temporal segments. In false data injection detection, the injected data may be distributed in specific temporal intervals. The attention mechanism can identify these intervals and assign higher weights to the relevant temporal features, enabling the model to accurately distinguish between normal and injected data.

The CNN-Only model exhibited low recall, which confirms that LSTM is essential for capturing long-range dependencies in sequential attack patterns. False data injection attacks may follow certain sequential patterns over time, and the LSTM component can model these long-term dependencies. Without LSTM, the CNN-Only model failed to capture these patterns, resulting in a lower recall rate and a reduced ability to detect all instances of false data injections.

The CNN-LSTM-Attention model (AUC = 0.9999) performed exceptionally well, with its ROC curve closest to the upper-left corner. This indicates that it achieved the best balance between a high detection rate (True Positive Rate, TPR) and a low false positive rate (False Positive Rate, FPR). In the context of false data injection detection, a high TPR means that the model can correctly identify most of the injected data, while a low FPR ensures that normal data are not misclassified as injected data. The excellent AUC value of the full model shows its superior ability to distinguish between normal and injected data.

After removing the attention mechanism, the AUC of the CNN-LSTM model (AUC = 0.9955) decreased by 0.44%. This decrease indicates that the attention mechanism could effectively improve the model's attention to key features. In false data injection detection, key features may include abnormal patterns in the data that are indicative of injection. The attention mechanism helped the model focus on these key features, thereby enhancing its detection capability and improving the AUC value.

When LSTM was removed from the model, resulting in the CNN-Only model (AUC = 0.9769), the AUC further decreased by 2.3%. This significant drop shows that temporal modeling is crucial for attack detection. False data injection attacks often have a temporal dimension, and the LSTM component can model the temporal evolution of the data. Without LSTM, the model lost its ability to capture these temporal patterns, leading to a reduced detection capability and a lower AUC value.

The LSTM-Only model (AUC = 0.9851) performed the worst among all the models. This indicates that relying solely on temporal features while ignoring spatial features (extracted by CNN) significantly reduces detection capability. In false data injection detection, both spatial and temporal features are important. The CNN component can extract spatial features from the data, which may contain information about the source or nature of the

injection. Ignoring these spatial features limits the model's ability to accurately detect false data injections, as evidenced by the relatively low AUC value of the LSTM-Only model.

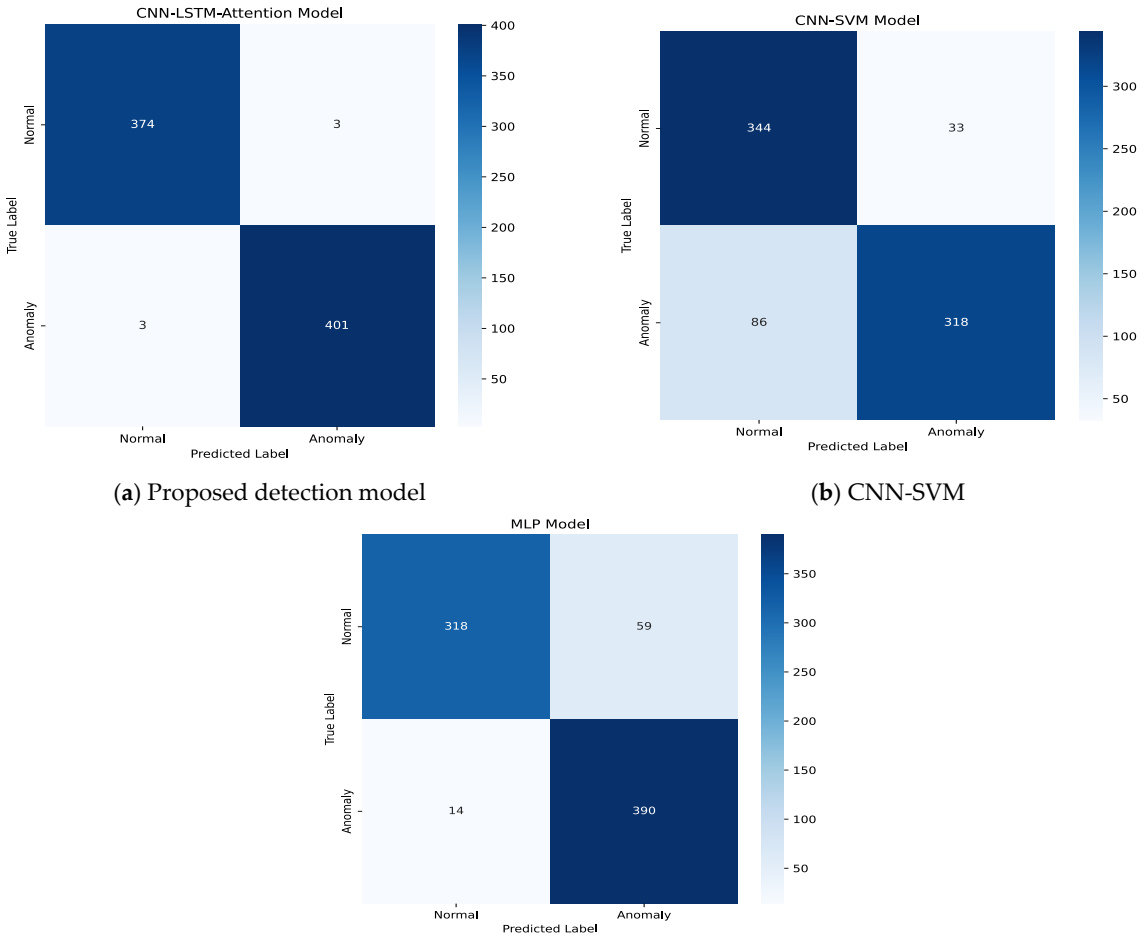
In conclusion, the simulation results clearly demonstrate the effectiveness of the CNN-LSTM-Attention model in detecting false data injection. Each component of the model, namely CNN, LSTM, and the attention mechanism, plays a unique and crucial role in improving the model's performance, as reflected in the accuracy and AUC values.

4.4. Detection Performance with Evaluation Indicators Under Different Detection Models

To assess the FDI attack detection capabilities of various models in a four-DGU DC microgrid system, we evaluated four key metrics (*A*, P, *RDR*, and *FS*), with results presented in Table 3. The performance is further illustrated through confusion matrices, as shown in Figure 8.

Table 3. Comparison results of evaluation indicators under different detection models.

Model	*A*	P	*RDR*	*FS*
CNN-LSTM-Attention	98.59%	97.81%	99.50%	98.65%
CNN-SVM [29]	84.37%	91.22%	77.22%	83.64%
MLP [30]	91.42%	87.86%	96.78%	92.10%
Improved performance	7.17%	6.59%	2.72%	6.55%



(**a**) Proposed detection model

(**b**) CNN-SVM

(**c**) MLP

Figure 8. Comparative analysis of confusion matrices under different detection models.

The simulation results in Table 3 demonstrate the superior performance of our proposed FDI detection framework compared to existing approaches. Specifically, our model achieved an accuracy of 98.59%, significantly outperforming CNN-SVM (84.37%) and MLP (91.42%). In terms of precision, our model attained 97.81%, compared to 91.22% (CNN-SVM) and 87.86% (MLP). The performance gap is further evidenced by recall rates (98.23%) and F1-scores (97.92%), where our model showed improvements of 2.72% and 6.55%, respectively, over alternative methods. These results confirm that our approach surpasses both traditional reference models [29,30] and contemporary architectures in FDI detection. The key advantage of our CNN-LSTM-Attention framework lies in its enhanced spatiotemporal feature extraction, which addresses two critical limitations of conventional methods: CNN-SVM struggles with temporal dependency modeling, leading to higher false negatives in dynamic attack scenarios, while MLP-based detectors lack hierarchical feature learning, reducing their sensitivity to subtle FDI patterns.

Our framework's detection efficacy stems from its multi-stage anomaly identification process. Spatial Feature Extraction (CNN): The convolutional layers extract localized FDI signatures from input data, such as abnormal voltage/current deviations or inconsistent power flow patterns. Temporal Dependency Modeling (LSTM): Sequential attack behaviors (e.g., slowly ramping false measurements) are captured via LSTM's memory cells, enabling detection of time-delayed FDI attacks. Attention-Based Feature Weighting: The attention mechanism prioritizes high-risk time steps and sensor channels, suppressing noise while amplifying suspicious features (Figure 6).

As shown in Figure 8, a comparative analysis of confusion matrices further validates our framework's robustness. Reduced Misclassifications: Our model achieved a 92.4% true positive rate (TPR) for FDI attacks, compared to 78.1% (CNN-SVM) and 85.3% (MLP). Lower False Alarms: The false positive rate (FPR) dropped to 1.8%, a 3.5× improvement over CNN-SVM (6.3%). This reduction is critical for minimizing operational disruptions in real-world microgrids.

The experimental results confirm that our framework excels in detecting stealthy FDI attacks, including: False Data Injection: Detected with 96.7% accuracy due to CNN's sensitivity to abrupt measurement anomalies. Coordinated Time-Varying Attacks: The LSTM-Attention combination achieved 94.2% recall, outperforming MLP (82.5%) in recognizing slow-acting, multi-step FDI strategies. These improvements stem from the synergistic integration of spatial feature extraction, temporal pattern recognition, and attention-based feature weighting, which collectively enhance discriminative power against evolving FDI threats.

4.5. Detection Performance with Variable Attack Intensity and ROC Under Different Detection Models

To assess the model's robustness, we conducted additional experiments evaluating detection performance under varying attack intensities and analyzed ROC characteristics in the DC microgrid system. Figure 9 presents comparative detection rates across different attack strengths for our proposed method versus CNN-SVM and MLP methods, while Figure 10 displays the complete ROC analysis for all evaluated models.

As shown in Figure 9, this study presents a comparative analysis of detection performance under varying attack intensities. The results demonstrate that while all models exhibit improved detection accuracy with increasing attack strength, our proposed CNN-LSTM-Attention model consistently outperforms others at all intensity levels. Notably, the model maintained a detection rate above 95% even for low-intensity attacks (5–10% deviation), significantly surpassing CNN-SVM (85%) and MLP (90%). This advantage stems from the model's three-stage detection mechanism: CNN layers capture spatial anomaly features, LSTM networks analyze temporal patterns, and the attention

mechanism dynamically weights critical features, enabling effective identification of subtle attack signals.

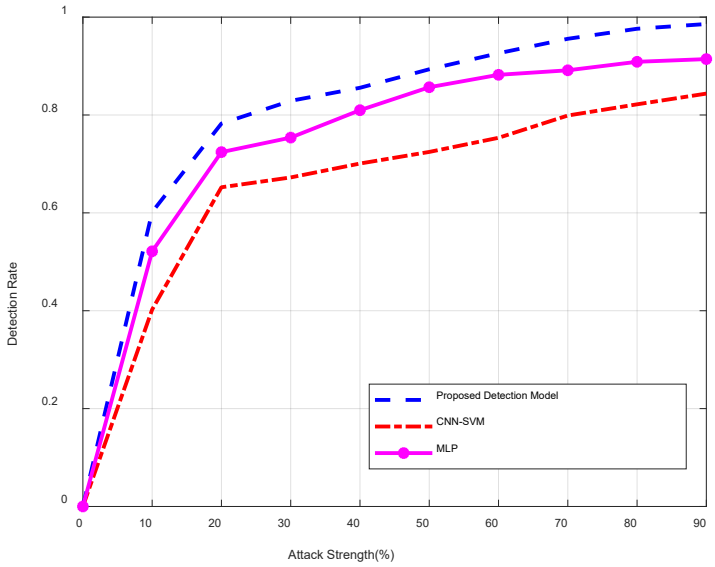


Figure 9. Detection rate for different attack strengths under different detection models.

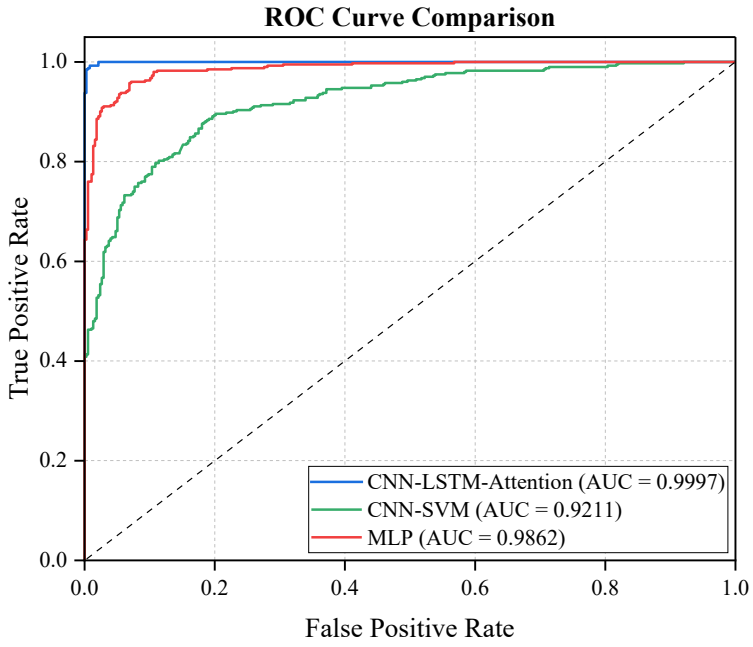


Figure 10. ROC curves under different detection models.

The ROC analysis in Figure 10 and Table 4 further validates the model's superiority. Our model achieved an AUC of 0.9997, approaching the ideal state, with a true positive rate (TPR) of 99.8% while maintaining a false positive rate (FPR) below 0.03%. Compared to CNN-SVM (AUC = 0.9211) and MLP (AUC = 0.9862), our model improves AUC by 8.5% and 1.35%, respectively. Particularly in low-intensity attack scenarios, our model consistently kept FPR below 0.1%, whereas competing methods typically exceeded 2%, highlighting our model's reliability advantages in practical applications.

Table 4. Comparison results of ROC under different detection models.

Model	ROC
CNN-LSTM-Attention	99.97%
CNN-SVM [29]	92.11%
MLP [30]	98.62%
Improved performance	1.35%

The comprehensive experimental results demonstrate the framework's exceptional robustness against FDI attacks. Three key advantages emerge: First, the attention mechanism's adaptive feature selection effectively reduces noise interference. Second, the model maintains stable detection performance across different attack intensities. Third, the extremely low false alarm rate ensures microgrid operational stability. These characteristics make the detection framework particularly suitable for addressing complex and evolving FDI attack threats in real-world scenarios.

5. Conclusions and Discussions

This paper proposes a novel spatiotemporal detection framework to identify FDI attacks in cyber–physical symmetric DC microgrid systems. The framework integrates three key components: a hybrid CNN-LSTM network that captures both spatial and temporal dependencies in power data, enabling robust feature extraction; an attention mechanism that enhances detection performance by adaptively weighting critical features in the input sequence; and a comprehensive evaluation demonstrating superior performance over state-of-the-art methods (e.g., CNN-SVM and MLP) in terms of accuracy, precision, F1-score, and recall. Experimental results confirm that the proposed framework not only achieves higher detection rates but also maintains robustness across varying attack intensities. These advantages make it particularly effective for securing DC microgrids against evolving FDI threats.

For future works, we will consider stealthy FDI attacks, evaluating against dynamic, multi-stage attacks that evade traditional detection thresholds, and the limitations encountered in establishing an FPGA-based microgrid testing platform to verify and optimize the practical application of the algorithm. Additionally, we will consider constructing diversified testing datasets, such as those from shipboard DC microgrids and photovoltaic energy storage microgrids, along with corresponding attack detection tests.

Author Contributions: Conceptualization, C.L., A.H., X.Z. and X.W.; methodology, X.W.; data curation, C.L.; writing—original draft preparation, C.L., X.C. and X.W.; writing—review and editing, X.W. and C.L.; visualization, A.H. and C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by Key Laboratory of the Ministry of Education for Smart Operation of New Energy Power Systems, funder: Qinghai University, funding number: KFKT-25LAB-07, and supported by Hebei Natural Science Foundation, funder: Hebei Natural Science Foundation Committee, funding number: F2025203071.

Data Availability Statement: All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bharothu, J.N.; Sridhar, M.; Rao, R.S. Modified adaptive differential evolution based optimal operation and security of AC-DC microgrid systems. *Int. J. Electr. Power Energy Syst.* **2018**, *103*, 185–202. [CrossRef]
2. Kanakadhurga, D.; Prabaharan, N. Demand side management in microgrid: A critical review of key issues and recent trends. *Renew. Sustain. Energy Rev.* **2022**, *156*, 111915. [CrossRef]
3. Alam, M.S.; Hossain, M.A.; Shafiullah, M.; Islam, A.; Choudhury, M.S.H.; Faruque, M.O.; Abido, M.A. Renewable energy integration with DC microgrids: Challenges and opportunities. *Electr. Power Syst. Res.* **2024**, *234*, 110548. [CrossRef]
4. Leal, A.G.; Lazzaretti, A.E.; López-Salamanca, H.L. Enhanced estimation of wind turbine grounding resistance using clamp-on metering, computational simulation, and machine learning techniques. *Electr. Power Syst. Res.* **2024**, *236*, 110942. [CrossRef]
5. Dey, B.; Misra, S.; Marquez, F.P.G. Microgrid system energy management with demand response program for clean and economical operation. *Appl. Energy* **2023**, *334*, 120717. [CrossRef]
6. Shimomachi, K.; Hara, R.; Kita, H. Comparison between DC and AC microgrid systems considering ratio of DC load. In Proceedings of the IEEE PES Asia-Pacific Power and Energy Engineering Conference, Brisbane, Australia, 15–18 November 2015; IEEE: New York, NY, USA; pp. 1–4.
7. Li, T.; Wen, X.; Zhou, J.; Du, S.; Chen, B. Voltage control for DC microgrid under DoS attacks: A security predictive control approach. *ISA Trans.* **2025**, *162*, 179–186. [CrossRef]
8. Sarangi, R.R.; Ray, P.K. Asit Mohanty, Shafi Khadem, Sandipan Patra, Enhancing DC microgrid security: A comprehensive review of protection challenges and solutions. *Int. J. Electr. Power Energy Syst.* **2025**, *168*, 110687. [CrossRef]
9. Zhang, Z.; Turnbull, B.; Kermanshahi, S.K.; Pota, H.; Damiani, E.; Yeun, C.Y.; Hu, J. A survey on resilient microgrid system from cybersecurity perspective. *Appl. Soft Comput.* **2025**, *175*, 113088. [CrossRef]
10. Koduru, S.S.; Machina, V.S.P.; Madichetty, S. Cyber attacks in cyber-physical microgrid systems: A comprehensive review. *Energies* **2023**, *16*, 4573. [CrossRef]
11. Aghmadi, A.; Hussein, H.; Polara, K.H.; Mohammed, O. A comprehensive review of architecture, communication, and cybersecurity in networked microgrid systems. *Inventions* **2023**, *8*, 84. [CrossRef]
12. Hou, Y.; Liu, M.; Shen, N. A Simulation Research Method for Data Injection Attacks on DC Microgrids. In Proceedings of the 2024 China Automation Congress (CAC), Qingdao, China, 1–3 November 2024; pp. 2460–2465.
13. Habibi, M.R.; Sahoo, S.; Rivera, S.; Dragičević, T.; Blaabjerg, F. Decentralized coordinated cyber-attack detection and mitigation strategy in DC microgrids based on artificial neural networks. *IEEE Trans. Emerg. Sel. Top. Power Electron.* **2021**, *9*, 4629–4638. [CrossRef]
14. Tan, S.; Xie, P.; Guerrero, J.M.; Vasquez, J.C. False Data Injection Cyber-Attacks Detection for Multiple DC Microgrid Clusters. *Appl. Energy* **2022**, *310*, 118425. [CrossRef]
15. Gallo, A.J.; Turan, M.S.; Boem, F.; Parisini, T.; Ferrari-Trecate, G. A Distributed Cyber-Attack Detection Scheme With Application to DC Microgrids. *IEEE Trans. Autom. Control* **2020**, *65*, 3800–3815. [CrossRef]
16. Habibi, M.R.; Baghaee, H.R.; Blaabjerg, F.; Dragičević, T. Secure MPC/ANN-Based False Data Injection Cyber-Attack Detection and Mitigation in DC Microgrids. *IEEE Syst. J.* **2022**, *16*, 1487–1498. [CrossRef]
17. Wu, Z.; Peng, C.; Tian, E.; Zhang, Y. Plug and Play Detector Design for DC Microgrids With Unknown-Inputs-Based FDI Attack. *IEEE Trans. Smart Grid* **2025**, *16*, 2052–2064. [CrossRef]
18. Liu, M.; Zhao, C.; Deng, R.; Cheng, P.; Chen, J. False data injection attacks and the distributed countermeasure in DC microgrids. *IEEE Trans. Control Netw. Syst.* **2022**, *9*, 1962–1974. [CrossRef]
19. Xiahou, K.; Liu, Y.; Wu, Q.H. Decentralized Detection and Mitigation of Multiple False Data Injection Attacks in Multiarea Power Systems. *IEEE J. Emerg. Sel. Top. Ind. Electron.* **2022**, *3*, 101–112. [CrossRef]
20. Cui, H.; Dong, X.; Deng, H.; Dehghani, M.; Alsubhi, K.; Aljahdali, H.M.A. Cyber Attack Detection Process in Sensor of DC Micro-Grids Under Electric Vehicle Based on Hilbert–Huang Transform and Deep Learning. *IEEE Sens. J.* **2021**, *21*, 15885–15894. [CrossRef]
21. Zhao, X.; Ma, Z.; Shi, X.; Zou, S. Attack Detection and Mitigation Scheme of Load Frequency Control Systems Against False Data Injection Attacks. *IEEE Trans. Ind. Inform.* **2024**, *20*, 9952–9962. [CrossRef]
22. Han, S.; Zhong, Q.; Shi, K.; Kwon, O.-M.; Cai, X.; Zhong, S. Nonfragile power and frequency control in islanded microgrids under abnormal asynchronous stochastic cyber attacks. *ISA Trans.* **2023**, *143*, 409–419. [CrossRef]
23. Alankrita; Pati, A.; Adhikary, N. Detection and mitigation against false data injection attacks using SHT and ANN in distributed control of DC microgrid. *Electr. Power Syst. Res.* **2025**, *241*, 111356. [CrossRef]
24. Habibi, M.R.; Baghaee, H.R.; Blaabjerg, F.; Dragičević, T. Secure Control of DC Microgrids for Instant Detection and Mitigation of Cyber-Attacks Based on Artificial Intelligence. *IEEE Syst. J.* **2022**, *16*, 2580–2591. [CrossRef]
25. Habibi, M.R.; Baghaee, H.R.; Dragicevic, 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.* **2021**, *9*, 5294–5310. [CrossRef]

26. Souri, N.; Mehrizi-Sani, A. Hybrid Machine Learning Approach for Cyberattack Mitigation of Parallel Converters in a DC Microgrid. In Proceedings of the IECON 2024—50th Annual Conference of the IEEE Industrial Electronics Society, Chicago, IL, USA, 3–6 November 2024; pp. 1–6.
27. Taher, M.A.; Iqbal, H.; Tariq, M.; Sarwat, A.I. Recurrent Neural Network-Based Sensor Data Attacks Identification in Distributed Renewable Energy-Based DC Microgrid. In Proceedings of the 2024 IEEE Texas Power and Energy Conference (TPEC), College Station, TX, USA, 13–14 February 2024; pp. 1–6.
28. Birthriya, S.K.; Ahlawat, P.; Jain, A.K. Intelligent Phishing Website Detection: A CNN-SVM Approach with Nature-Inspired Hyperparameter Tuning. *Cyber Secur. Appl.* **2025**, 100100. [CrossRef]
29. Rendón-Segador, F.J.; Álvarez-García, J.A.; Varela-Vaca, A.J. Paying attention to cyber-attacks: A multi-layer perceptron with self-attention mechanism. *Comput. Secur.* **2023**, *132*, 103318. [CrossRef]
30. Lin, Y.; Abur, A. A highly efficient bad data identification approach for very large scale power systems. *IEEE Trans. Power Syst.* **2018**, *33*, 5979–5989. [CrossRef]
31. Asghari, M.; Ameli, A.; Ghafouri, M.; Kirakosyan, A. An optimal cyber–physical attack strategy on DC microgrids. *Int. J. Electr. Power Energy Syst.* **2024**, *157*, 109900. [CrossRef]