Information Fusion 122 (2025) 103183

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Piece-wise loss guided multi-resolution data fusion learning for false data injection attack detection in smart grids

Haosen Yang [b], Bifei Tan [a,*], Zipeng Liang [b], Xin Shi [c], Hanjiang Dong [b], Chongyu Wang [d], Shibo Chen [e]

[a] Wuyi University, Guangdong, China
[b] Hong Kong Polytechnic University, Kowloon, Hong Kong
[c] North China Electric Power University, Beijing, China
[d] Illinois Institute of Technology, Chicago, IL, United States of America
[e] The Hong Kong University of Science and Technology, Hong Kong

ARTICLE INFO

Keywords:
False data injection attack detection
Multi-resolution data fusion
Semi-supervised learning
Piece-wise loss function

ABSTRACT

Cyber-attacks perform a significant threat to the stability and security of electrical power grids. This paper initially analyzes various kinds of detective data regions related to false data injection attacks (FDIAs) detection methods in literature. And then, to detect FDIAs in a high-accuracy and high-robustness way, this paper presents a novel multi-resolution data fusion approach. The proposed approach integrates high-resolution data from phasor measurement units (PMUs) and low-resolution data from supervisory control and data acquisition (SCADA) systems. This combination leverages the distinct yet complementary characteristics of PMU and SCADA data, both of which are crucial for modern power grid management. Besides, given the scarcity of real attack data, we propose a piece-wise loss guided semi-supervised learning methodology. It effectively merges limited FDIA data with abundant normal operating data. Therefore, the proposed approach operates in the scenario totally consistent to the practice, which utilizes PMU and SCADA data as much as possible, at the same time incorporates limited real FDIA data and other data of normal operating states. In detail, we introduce a resolution enlargement network to extract features at multiple resolutions. These features, along with original PMU and SCADA data, are then fed into the semi-supervised learning framework consisting of an unsupervised branch for normal operational data and a supervised branch guided by piece-wise loss for sparse FDIA data. Extensive validations across a wide range of case studies illustrate that the proposed method demonstrates significant improvements in detection accuracy.

1. Introduction

Cyber-attacks in power grids refer to malicious acts of an adversary target on computer systems, communication networks, and control systems that manage and operate electrical power grids. These attacks aim to disrupt or manipulate the functioning of power grid infrastructures, potentially leading to power outages, equipment damage, economic losses, and even compromising the social security [1–4]. False data injection attacks (FDIAs), as one of the most prevalent ways of cyber-attacks, attracted great attention in recent years. Commonly-studied FDIA methods include load redistribution attacks [5–7], topological attacks [8,9], profit-oriented attacks in markets [10], parameters-aided attacks [11], sequential outages attacks [12,13] and blind attacks [14–17].

To against these numerous kinds of FDIAs, many countermeasures were proposed. The initial FDIA detection methods are traditional statistical methods, such as χ^2 method [18], largest normalized residual test (LNRT) [18] and largest absolute value (LAV) [19]. These detection approaches can only identify the bad data that are not aligned with the physical principles of power grids, and they cannot identify the new coming unobservable attacks. To enhance the FDIA detection performance, in recent years, numerous machine learning approaches were presented, which can better utilize the statistical properties behind the tampered data. These machine learning approaches include deep belief network (DBN) [20–22], principal component analysis (PCA) [23], Gaussian mixture model (GMM) [24], extreme learning machine (ELM) [25], convolutional neural network (CNN) [26–28],

* Corresponding author.
E-mail addresses: 31910019@sjtu.edu.cn (H. Yang), tanbifei0487@hotmail.com (B. Tan), zipliang.liang@connect.polyu.hk (Z. Liang), xinshi@ncepu.edu.cn (X. Shi), hanjiang.dong@connect.polyu.hk (H. Dong), chongyu.wang@outlook.com (C. Wang), shibochen.ustc@gmail.com (S. Chen).

https://doi.org/10.1016/j.inffus.2025.103183
Received 4 October 2024; Received in revised form 30 January 2025; Accepted 1 April 2025
Available online 15 April 2025

long-short-term memory (LSTM) neural network [29], gated recurrent unit (GRU) [30,31], graphical neural network (GNN) [32,33], Transformer [34], deep autoencoder (DAE) [35,36], stacked autoencoder (SAE) [37], adversarial autoencoder (AAE) [38], deep autoencoding Gaussian mixture model [39], temporal convolutional network (TCN) [40] and generative adversarial network (GAN) [41].

We classified these machine learning FDIA detection approaches in two ways: (1) according to the inputting data; (2) according to the training methodology.

The input data for FDIA detection primarily consist of phasor measurement unit (PMU) data and supervisory control and data acquisition (SCADA) data. SCADA data include active and reactive power injections at nodes, as well as nodal voltage magnitudes. In contrast, PMU data provide additional information such as the phase of nodal voltages and currents. PMUs have significantly high sampling frequency of nearly 30–120 Hz, compared to the much lower sampling rate of approximately 0.1–0.5 Hz of SCADA systems. Furthermore, PMUs offer greater measurement accuracy than SCADA, with an error of almost 0.3–1% compared to the error of 1%–3% of SCADA. Despite these advantages, the high costs of PMUs limit their installation to partial critical nodes within the power grid, resulting in relatively sparse PMU data. Consequently, SCADA data offer a more vital role in power grid operations due to the general deployment [42,43]. However, the detailed temporal characteristics provided by PMUs can greatly enhance FDIA detection capabilities. This leads to the **problem 1**: How to collaboratively fuse insufficient high-resolution and high-accuracy PMU data as well as enough low-frequency yet low-precision SCADA data to benefit the FDIA detection.

As for the training property, machine learning based FDIA approaches can be generally divided into three categories: supervised, unsupervised and semi-supervised learning. Supervised learning [25–34] aims to establish a relationship between measurement data and a target outcome, such as determining if there exists an attack. However, these approaches face significant challenges in the context of FDIA detection. First, the rapid increase in the volume of measurement data, combined with the scarcity of labeled FDIA instances, makes it difficult to collect timely and accurate data. Secondly, there may be considerable differences between future FDIAs and historical samples, as attack strategies evolve quickly and are very difficult to predict. About unsupervised learning [20–24,35–37,39], they rely solely on measurement data without labels. As a result, they fail to take advantages of the available labeled FDIA data from historical records. To effectively leverage the limited labeled FDIA data alongside abundant normal operational data, semi-supervised learning emerges as a suitable approach for FDIA detection in power grids. In [38,41], adversarial semi-supervised learning was applied which identifies FDIA by whether the data can be properly reconstructed by neural networks. Ref. [40] used a TCN with a traffic attention module to capture long-term temporal correlations. **Problem 2**: However, one problem for current semi-supervised learning is that due to the rarity of labeled FDIA data, it is difficult to define the threshold which is used to finally split the results as false data or not. Considering that whether an FDIA is successful depends not only on the attack strategy but also on the magnitude of the overall FDIA vector or partial elements, for an FDIA data $\mathbf{a}$, its linear enlargement $\boldsymbol{\kappa}\mathbf{a}$ ($\kappa \geq 1$) is also an FDIA data in all likelihood. But conversely, the linear reduction $\boldsymbol{\kappa}\mathbf{a}$ ($\kappa < 1$) may be not. Therefore, in the scenario of extremely small amount of effective FDIA data, for the data $\boldsymbol{\kappa}\mathbf{a}$ ($\kappa < 1$), it is difficult to figure out the determination boundary of FDIA detection.

To address the problem 1 and 2 as mentioned above, this article proposes a multi-resolution learning approach for FDIA detection, where different PMU and SCADA data are excellently fused. The proposed method at first utilizes a resolution enlargement neural network to learn multi-resolution features, and then these features along with SCADA and PMU data are fused for FDIA detection. Secondly, to properly utilize the extreme rare FDIA data and large amounts of normal operating data, a piece-wise loss function in a semi-supervised environment is used to guide the training process. The piece-wise loss sets many branches corresponding to FDIA data with different magnitudes. If an FDIA data exists, the piece-wise loss is expected to judge the almost linear enlargement of this data as FDIA samples. Combining these, the proposed approach is capable of achieving higher detection accuracy. Numerous tests are conducted to demonstrate the effectiveness and advantages of the proposed method.

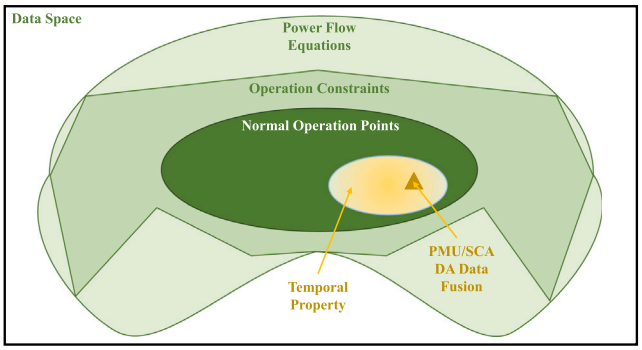


Fig. 1. The exhibition of various regions in the data space of power grids. A smaller data region means a stricter false data detection rule, and is more difficult to be attacked. The regions in order are formed by power flow equations, operation constraints, normal operating points, temporal property, and PMU/SCADA data fusion.

The main contributions are listed as follows:

(1) The detective data regions for multiple FDIA detection methods in literature were analyzed. We show the finding that the data region formed by PMU/SCADA data fusion is more difficult to be tampered by malicious attackers than that using power grid physical equations, operating constraints, common operating states, and temporal characteristics. Hence, the collaboration between abundant low-resolution, low-accuracy SCADA data and limited high-resolution, high-accuracy PMU data is very beneficial for FDIA detection.

(2) According to the above finding, we propose a multi-resolution fusion learning approach that integrates data from PMUs and SCADA. The proposed approach utilizes a resolution enlargement network to create a connection between SCADA and PMU data, facilitating the extraction of embedded multi-resolution features. These features, in conjunction with PMU and SCADA data, are subsequently integrated into a semi-supervised learning framework for the detection of FDIAs. Comparative analysis with other approaches highlights the advantages of the proposed method in improving detection accuracy.

(3) To tackle the practical problem posed by the extreme rarity of successful FDIA data, we still introduce a piece-wise loss function within the semi-supervised learning framework. This piece-wise loss function amplifies the influence of the limited labeled FDIA data, effectively addressing the data scarcity issue. Furthermore, the proposed semi-supervised learning framework leverages both the sparse FDIA data and abundant normal operating data.

The remainder of this article is organized as follows. Section 2 introduces the FDIA detection problem in smart grids, and Section 3 elaborates the proposed methodology. Section 4 is case studies, as well as Section 5 summarizes this work.

2. Problem statement

2.1. Basics of FDIA

Power system state estimation calculates the nodal voltages of all nodes from numerous measurement data, which is generally based on the measurement data of SCADA. It can be written as:

$$\mathbf{z} = h(\mathbf{x}) + \mathbf{e} \tag{1}$$

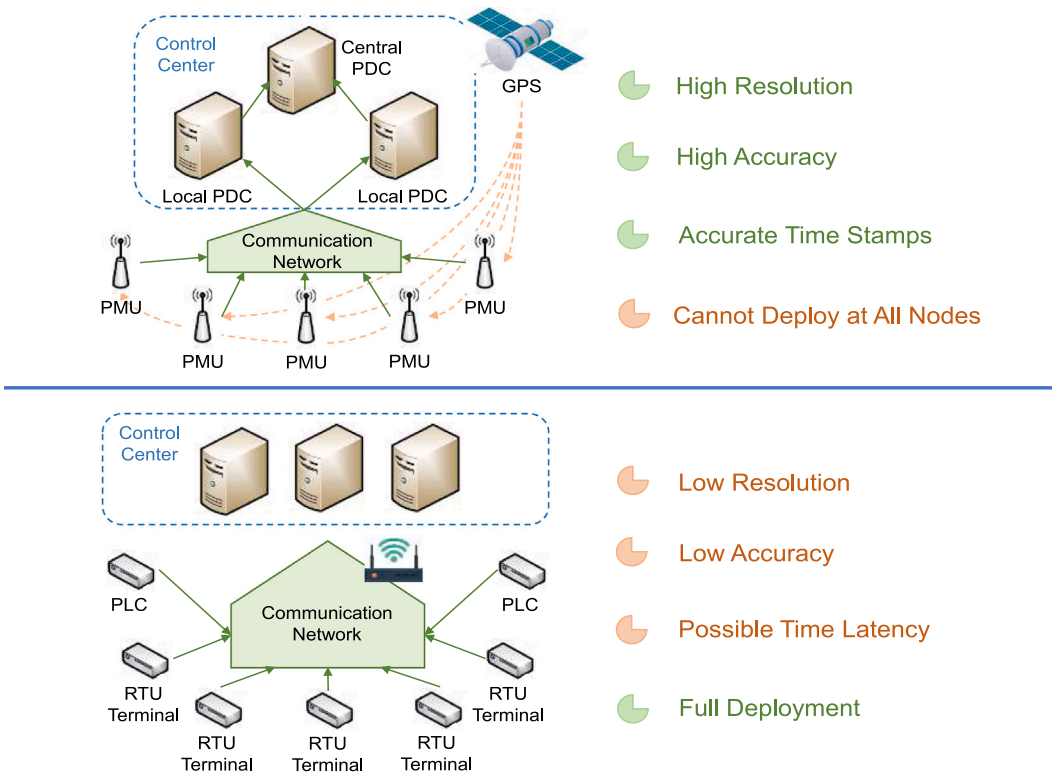


Fig. 2. The comparison between PMU and SCADA data. The upper part is the introduction of PMU, while the bottom one is about SCADA. PMUs have the advantages of high resolution, high accuracy, accurate time stamps, but it is not possible now to deploy PMUs at all nodes due to their high expense. In contrast, SCADA can be deployed at all nodes but it suffers from low resolution, low accuracy and possible inaccurate time stamps.

where $\mathbf{z}$ denotes the vector of measurement data from SCADA, including the magnitudes of nodal voltage, nodal injection power, power flows (branch transmission power). $\mathbf{x}$ is the state variables (nodal voltage magnitudes and angles). $\mathbf{e}$ represents the measurement noise, and $h()$ represents the nonlinear state estimation equations of power systems. State estimation uses the measurement data vector $\mathbf{z}$ to estimate the state variables $\mathbf{x}$.

The general FDIA model against state estimation is:

$$\mathbf{z}+\mathbf{a}=h(\mathbf{x}+\mathbf{c})+\mathbf{e} \tag{2}$$

where $\mathbf{a}$ and $\mathbf{c}$ denote the vector of injected false data, deviation of state variables, respectively. The attacker injects the false data $\mathbf{a}$ to lead the deviation of state variables from $\mathbf{x}$ to $\mathbf{x}+\mathbf{c}$, so that the changed state variables $\mathbf{x}+\mathbf{c}$ will lead the central energy management system (EMS) take inaccurate dispatch and control actions.

2.2. FDIA detection issue

The injected false data may be a single point or a continuous series of multiple measuring variables. The purpose of FDIA detection is to identify these malicious injected data. Traditional model-based false data detection, such as χ^2, LNRT, LAV detection, essentially uses the residual of state estimation $\|(\mathbf{z}+\mathbf{a})-h(\mathbf{x}+\mathbf{c})\|^2$, $\max\{(\mathbf{z}+\mathbf{a})-h(\mathbf{x}+\mathbf{c})\}$, $|(\mathbf{z}+\mathbf{a})-h(\mathbf{x}+\mathbf{c})|$ to detect false data, respectively. However, the state estimation residual essentially represents if the injected data $\mathbf{z}+\mathbf{a}$ is corresponding to the system power flow equations (physical equations of power grids), i.e., whether there exists a solution $\mathbf{x}+\mathbf{c}$ satisfying $\mathbf{z}+\mathbf{a}=h(\mathbf{x}+\mathbf{c})$. To clearly show this, Fig. 1 is included, where the largest green region refers to the area surrounded by power flow equations, thereby FDIA data falling at this region can bypass traditional residuals based detection methods, called unobservable FDIA [2–4,14–16,44]. Moreover, various operational constraints, such as the upper and lower limits on power transmission and nodal voltage magnitudes, define a narrower operational area. Additionally, a smaller feasible region is formed by the abundance of normal operational points, namely it is constructed by the range of common operating states, as shown by the dark green region in Fig. 1. When considering the time series properties of power grid data, a narrower and safer region can be established, as illustrated by the yellow area in Fig. 1. The time series characteristics provide more comprehensive features that enhance the effectiveness of FDIA detection.

More recent, the emergence of PMU data largely enhances the technology of FDIA detection [45]. Due to the high resolution and high accuracy of PMU data, bringing PMU data into FDIA detection can significantly improve the FIDA detection accuracy, as shown in the yellow triangle in Fig. 1. The advantages of PMU data can be listed as follows: (1) PMU data are more high-resolution (the resolution of PMU is around 30–120 Hz, while SCADA is almost 0.1–0.5 Hz); (2) PMU data are more accurate than SCADA, with an error of almost 0.3–1% compared to the error of 1%–3% of SCADA; (3) the time stamps of PMU data are determined by the global position system (GPS), thus the time stamps of PMU data are very precise. However, due to the expense of PMUs, the PMU deployments are limited to partial critical nodes rather than all buses. Therefore, how to utilize the rare high-resolution high-accuracy PMU data to aid the general low-resolution low-accuracy SCADA data becomes a crucial problem in FDIA detection. The advantages and disadvantages of PMU versus SCADA are shown in Fig. 2.

On the other hand, FDIA detection in practice is naturally a semi-supervised learning problem. This is caused by that the practical up-to-date successful FDIA is very rare, i.e., most of FDIA attempts are ineffective. So it is important for FDIA detection to take advantages of both seldom successful FDIA data and sufficient normal operating data of a power grid, and semi-supervised learning is extremely suitable to this scenario.

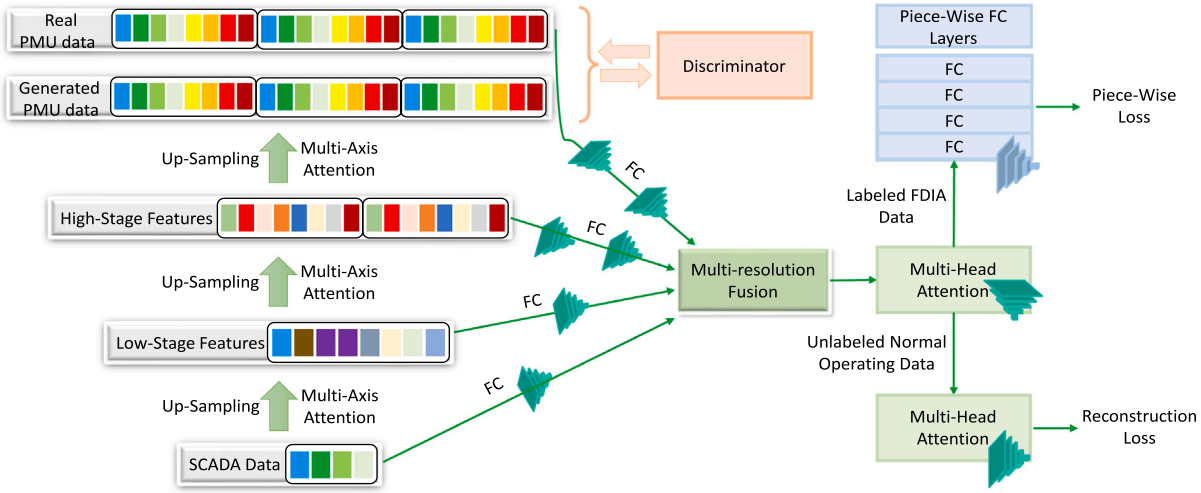


Fig. 3. The framework of the proposed data fusion method for FDIA detection. The left part shows the resolution enlargement framework where from the bottom to top are SCADA data, low-stage features, high-stage features and PMU data. The right part is the FDIA detection method, which at first integrates the features and data with multiple resolutions, then processes it through multi-head attention blocks. Finally, the unlabeled normal data and labeled FDIA data are sent into the unsupervised branch with piece-wise loss and the supervised branch with reconstruction loss.

3. Proposed methodology

The proposed method mainly has two stages: multi-resolution features extraction and piece-wise loss guided semi-supervised learning. The first part aims to generate features with multiple resolutions which contain the spatial–temporal knowledge of power grid operation data. Utilizing these features, the second part identifies the false data caused by FDIA in a semi-supervised way.

3.1. Resolution enlargement framework

To extract multi-resolution features, we propose a resolution enlargement framework based on GAN. It learns the relationship between SCADA data and PMU data, i.e., enlarges the low-frequency SCADA data to generate high-frequency PMU data. As shown in the left part of Fig. 3, the detailed framework of this resolution enlargement network has numerous layers which hierarchically enlarge the resolution of SCADA data and finally approximate PMU data.

Considering a matrix of SCADA data $\mathbf{Z_s}$ for a time period with length T, it collects measurement data of N variables to form a $N \times T$ matrix $\mathbf{Z_s}$. Then we propose to use numerous up-sampling and multi-axis attention layers (USMAAs) to enlarge $\mathbf{Z_s}$ to estimate PMU data $\mathbf{Z_p}$. Each USMAA employs a up-sampling process and a multi-axis attention part. The up-sampling, as shown in Fig. 4, essentially copies the original data along with the temporal direction. The multi-axis attention refers to the simultaneous attention calculation for temporal axis and spatial axis. As shown in Fig. 5, the proposed method has a spatial attention and a temporal attention part, and they are both multi-head attention networks (see Section 3.B (3)) but applied on different data slices. As shown in Fig. 5, the temporal attention processes each row in $\mathbf{Z_s}$ (time series of a measurement variable), as well as the spatial attention processes every column in $\mathbf{Z_s}$ (all measurement variables at a certain time). Then these two attention results are added as the final result of a USMAA. So one layer of up-sampling and multi-axis attention can be written as:

$$\mathbf{O_l} = MAA(US(\mathbf{Z_s})) \tag{3}$$

where $US()$ and $MAA()$ denote the calculation process of up-sampling and multi-axis attention. $\mathbf{O_l}$ is the output of the first USMAA layer, namely the low-stage features. Generally, we use three USMAA layers,

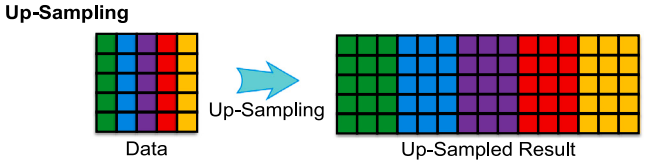


Fig. 4. The exhibition of up-sampling process, which copies the original data along with the temporal direction.

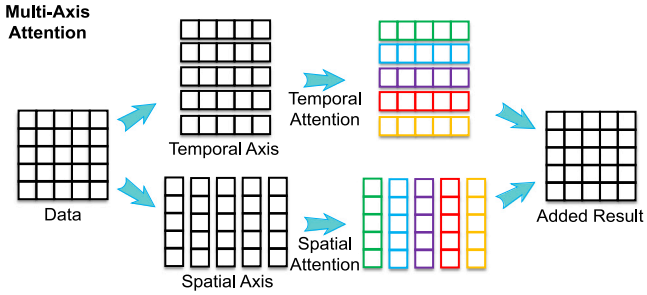


Fig. 5. The exhibition of multi-axis attention mechanism. The upper part is the temporal attention part where the multi-head attention is used for temporal axis, and the bottom part means that the multi-head attention is operated for data with different nodes, i.e., in spatial axis.

thus we have:

$$\begin{aligned} \mathbf{O_l} &= MAA(US(\mathbf{Z_s})) \\ \mathbf{O_h} &= MAA(US(\mathbf{O_l})) \\ \mathbf{O_f} &= MAA(US(\mathbf{O_h})) \end{aligned} \tag{4}$$

where $\mathbf{O_l}$, $\mathbf{O_h}$, $\mathbf{O_f}$ denote the low-stage features (output of the first USMAA layer), high-stage features (output of the second USMAA layer) and the final output (approximated PMU data).

Then to make the final output $\mathbf{O_f}$ accurately approximate the real PMU data $\mathbf{Z_p}$, we employ the structure of GAN [46]. Specifically, a discriminator, composed of numerous fully-connected (FC) layers, is used to distinguish the output $\mathbf{O_f}$ and the real PMU data $\mathbf{Z_p}$:

$$Dis(\mathbf{O_f}, \mathbf{Z_p}) = MFC([\mathbf{O_f}, \mathbf{Z_p}]) \tag{5}$$

where $Dis()$ denotes the discriminator, and $MFC()$ represents multiple FC layers. Similar to GAN [46], by an iterative yet adversarial training

process, the USMAA layers and the discriminator will eventually reach the Nash equilibrium. And forced by the supervise of the discriminator, the USMAA layers are able to precisely approximate the PMU data from SCADA data inputs, thus the two features in the middle, i.e., the low-stage and high-stage features, are generated. The purpose of using features with different resolutions are twofold: (1) The features with various resolutions can represent the spatial–temporal characteristics with different time scales. Specifically, as the resolution of PMU is much larger than that of SCADA, the supplement of features with middle resolutions (lower than PMU but higher than SCADA) is useful to improve the FDIA detection performance. (2) PMU data include voltage phases (angles) which cannot be measured by SCADA. Thus the relationship mapping from SCADA data to PMU data is very possible to contain other information about power grid physical principles.

3.2. FDIA detection methodology

After the training process of the resolution enlargement framework as mentioned above, we obtain low-stage features $\mathbf{O}_l$ and high-stage features $\mathbf{O}_h$. These features along with PMU and SCADA data are then input into the proposed piece-wise loss guided semi-supervised framework. As shown in the middle part of Fig. 3, they are fused by FC layers with different sizes, i.e., we have:

$$\mathbf{O_{FU}} = [FC(FC(\widetilde{\mathbf{Z}}_p)), FC(FC(\mathbf{O}_h)), FC(\mathbf{O_l}), FC(\mathbf{Z}_s)] \tag{6}$$

where $FC()$ notates an FC layer. $\widetilde{\mathbf{Z}}_p$ represents a mixed PMU data for all measuring variables due to that PMUs cannot be deployed at all buses. For a node without PMU, the approximated PMU data from resolution enlargement network $\mathbf{O_f}$ is used. Thereby it can be written as:

$$\widetilde{\mathbf{Z}_{p,j}} = \begin{cases} \mathbf{Z_{p,j}}, & if \quad \xi_j = 1 \\ \mathbf{O_{f,j}}, & if \quad \xi_j = 0 \end{cases} \tag{7}$$

where $\mathbf{Z_{p,j}}$, $\mathbf{O_{f,j}}$ represent the jth row of $\mathbf{Z_p}$, $\mathbf{O_f}$, namely the measurement series of the jth single measuring variable. ξ_i is an indicator representing if the ith measurement variable is accessible by PMUs.

3.2.1. Unsupervised branch

As mentioned above, FDIA detection generally operates in a semi-supervised way, where there are only seldom realistic FDIA data $\{\mathbf{O^F_{FU}}, y\}$ and large amounts of normal operating data $\mathbf{O^N_{FU}}$. The rare FDIA data are input into the supervised branch while normal operating data are handled by the unsupervised branch. As shown in the right bottom part of Fig. 3, when inputting normal operating data, a multi-head attention [47] structure is used. And a reconstruction loss measuring the difference between the outputs and PMU data is used:

$$L_{rec} = \|MHA(MHA(\mathbf{O^N_{FU}})) - \mathbf{Z_p}\|^2 \tag{8}$$

where $MHA()$ denotes the multi-head attention part.

3.2.2. Piece-wise loss guided supervised branch

The supervised branch uses a piece-wise loss function to handle seldom FDIA data. Due to the rarity of effective practical FDIA data and the fact that whether an FDIA is unobservable depends not only on the attack strategy but also the magnitude of attack vectors, a very large-magnitude FDIA vector is easy to be identified, while a small-magnitude FDIA vector is more stealthy but has less damage. To make the FDIA realize sufficient damages, attackers generally tend to launch a large-magnitude attack within the stealth requirement. So in the perspective of a defender, in the scenario of very seldom FDIA data, it is not easy to acknowledge the threshold that determines what magnitude of FDIA data should be identified. For instance, if there exists an FDIA data $\mathbf{a}$ with the magnitude $|\mathbf{a}|$, the linear enlargement $\boldsymbol{\kappa}\mathbf{a}$ ($\kappa \geq 1$) is also supposed to be an FDIA data, but the linear reduction $\boldsymbol{\kappa}\mathbf{a}$ ($\kappa < 1$) may be not. This is due to that the small-magnitude data changes may also result from fluctuations of operating states, but the linear enlargement of an FDIA data $\boldsymbol{\kappa}\mathbf{a}$ ($\kappa \geq 1$) is still an FDIA data by the similar FDIA strategy in all likelihood. To address this problem, we present a piece-wise loss function to guide the supervised branch.

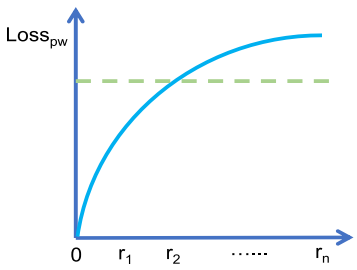


Fig. 6. The expected profile of the piece-wise loss function. It generates large values for the piece with large intervals. For an FDIA data sequence falling at the interval r_i to r_{i+1}, the FC layers corresponding to larger intervals (larger than r_{i+1}) are expected to indicate it as false data.

For an FDIA data sequence, we set some intervals according to the data magnitudes. These intervals are denoted by $0 < r_1 < r_2 < r_3 < \cdots < r_n$, where n denotes the number of intervals. Then the piece-wise loss function for FDIA data is represented by:

$$L_f = exp(-\int_{r_i}^{r_n} EFC_i(MHA(\mathbf{O^F_{FU}}))dr_i) \tag{9}$$

where $MHA(\mathbf{O^F_{FU}})$ represents the output of a multi-head attention part when inputting $\mathbf{O^F_{FU}}$. $EFC_i()$ is the FC layer with an Exponential function corresponding to the interval from r_{i-1} to r_i, so we have:

$$EFC_i(MHA(\mathbf{O^F_{FU}})) = exp(\mathbf{w_i^T}MHA(\mathbf{O^F_{FU}}) + \mathbf{b}_i) \tag{10}$$

where $\mathbf{w_i}$ and $\mathbf{b_i}$ are the weight and bias vector of the ith FC layer (corresponding to the interval r_{i-1} to r_i), as shown in the right upper part in Fig. 3. Specifically, each interval (e.g., $0 \sim r_1$, $r_1 \sim r_2$, $\ldots$, $r_i \sim r_{i+1}$) has an FC layer with parameters $\mathbf{w_i}$ and $\mathbf{b_i}$. When $|MHA(\mathbf{O^F_{FU}})|$ falls into the interval between r_i and r_{i+1}, this piece-wise loss function integrates the results of FC layers $exp(\mathbf{w_i^T}MHA(\mathbf{O^F_{FU}}) + \mathbf{b}_i)$ whose corresponding lower interval bound is larger or equal to r_i. In other words, if a data sequence is false, the FC layers corresponding to larger intervals are expected to unanimously indicate it as a false data. The loss function (9) can be rewritten as:

$$L_f = exp(-\sum_{r_i}^{r_n} EFC_i(MHA(\mathbf{O^F_{FU}}))) \tag{11}$$

Conversely, if the data without FDIA $\mathbf{O^N_{FU}}$ is used for the supervised branch, the loss function (9) becomes:

$$L_n = 1 - exp(-\sum_{0}^{r_i} EFC_i(MHA(\mathbf{O^N_{FU}}))) \tag{12}$$

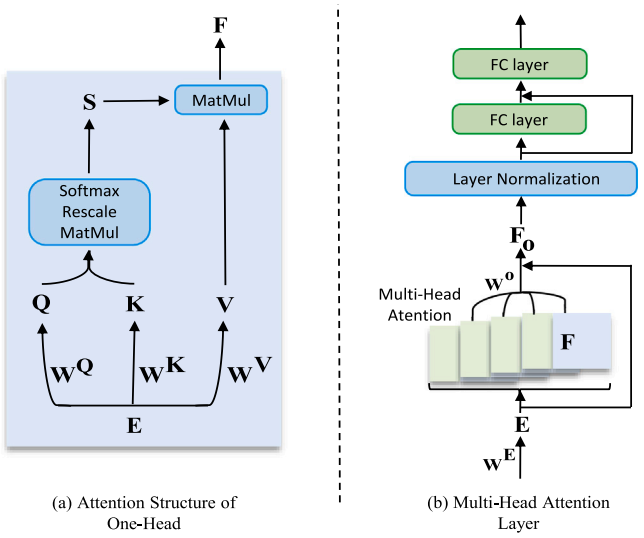
Then the overall piece-wise loss function is defined as:

$$L_{pw} = \sum_{\mathbf{O^F_{FU}} \in S_F} L_f + \sum_{\mathbf{O^N_{FU}} \in S_{N_P}} L_n \tag{13}$$

where S_F is the set of successful FDIA data, and S_{N_P} is a set of partial randomly-selected normal operating data paired with FDIA data S_F ($|S_F| = |S_{N_P}|$). When inputting FDIA data into the proposed approach, these paired normal operating data are also input into the supervised branch. The piece-wise loss is expected to provide large values for FDIA data while small values for normal data. An ideal profile of the piece-wise loss is shown in Fig. 6.

When training the proposed method, large amounts of normal operating data are input into the unsupervised branch, and hence the parameter-updating process is guided by the reconstruction loss L_{rec}. Therefore, the total loss of training can be described by:

$$L = \sum_{\mathbf{O^N_{FU}} \in S_N} L_{rec} + \sum_{\mathbf{O^F_{FU}} \in S_F} L_f + \sum_{\mathbf{O^N_{FU}} \in S_{N_P}} L_n \tag{14}$$



(a) Attention Structure of One-Head

(b) Multi-Head Attention Layer

Fig. 7. The exhibition of multi-head attention structure. The left part shows the flowchart of the attention calculation in one-head, corresponding to the formula (15). The right part exhibits the entire multi-head attention structure, shown as the formula (16) and Ref. [48,49].

3.2.3. Multi-head attention

We then introduce the multi-head attention structure [47] which serves as the bottom block of the proposed method. The self-attention structure of one head here is:

$$\mathbf{E} = \mathbf{Z}_{\mathbf{in}}^{\mathbf{T}}\mathbf{W}^{\mathbf{E}}, \quad \mathbf{Q} = \mathbf{E}\mathbf{W}^{\mathbf{Q}}, \quad \mathbf{K} = \mathbf{E}\mathbf{W}^{\mathbf{K}},$$
$$\mathbf{V} = \mathbf{E}\mathbf{W}^{\mathbf{V}}, \quad \mathbf{S} = Sm(\frac{\mathbf{Q}\mathbf{K}^{\mathbf{T}}}{\sqrt{d}}), \quad \mathbf{F} = \mathbf{S}\mathbf{V} \tag{15}$$

where $\mathbf{Z}_{\mathbf{in}}$ is the input measurement data matrix defined above. $Sm()$ denotes the well-known Softmax function, and $\mathbf{W}^{\mathbf{E}}$, $\mathbf{W}^{\mathbf{Q}}$, $\mathbf{W}^{\mathbf{K}}$, $\mathbf{W}^{\mathbf{V}}$ are weight matrices, as well as d is the hidden size. The flowchart of this calculation is shown in Fig. 7(a).

A multi-head attention block has multiple similar single-head blocks, as shown in the bottom part in Fig. 7(b). The final output of the multi-head attention is the linear integration of the result $\mathbf{F}$ of all heads with a skip-connection part:

$$\mathbf{F}_{\mathbf{o}} = [\mathbf{F}_{\mathbf{1}}, \mathbf{F}_{\mathbf{2}}, \dots, \mathbf{F}_{\mathbf{H}}]\mathbf{W}^{\mathbf{o}} + \mathbf{E} \tag{16}$$

where $\mathbf{F}_{\mathbf{i}}$ denotes the output of the ith head, and H is the number of heads of the multi-head attention. $\mathbf{W}^{\mathbf{o}}$ is the weight matrix, and $\mathbf{F}_{\mathbf{o}}$ is the integrated result.

As shown in Fig. 7(b), after the multi-head attention, the output $\mathbf{F}_{\mathbf{o}}$ is then sent into a layer normalization [49], an FC layer with skip-connection [48] and an FC layer sequentially. The first FC layer with skip-connection aims to enhance the representation ability, while the final FC layer is used to transform the results into any form. All above mentioned parts are concatenated to form one layer of the proposed method, as shown in Fig. 7(b).

3.3. Detection threshold

Considering we simultaneously use the reconstruction loss for normal operating data and the piece-wise loss for rare FDIA data, we define a mixture criterion integrating these unsupervised and supervised branches to represent the likelihood of false data:

$$S = L_f - \alpha L_n + \beta L_{rec} \tag{17}$$

where L_f, L_n and L_{rec} here represent the loss ((11), (12), (8)) when inputting the incoming data. Then the threshold η of this score (17) is based on the contamination rate (rate of FDIA data) assumed in the training dataset. For instance, if the contamination rate is 0.05, η is set at the 5% location which excludes the data with the largest 5% criterion S in the training dataset as FDIA data. After the threshold η is determined, the incoming data point larger than η will be identified as false data.

4. Case studies

In this section, we show the performance and advantages of the presented method against other methods. In the first case, we test the impact of every part in the proposed method. In the second case, numerous supervised, unsupervised, and semi-supervised approaches are compared. And the third case tests the robustness of the proposed approach against different ratios of labeled FDIA data, and data windows with various time scales.

4.1. Settings

We use two datasets simulated in IEEE 14 and 118-bus system to evaluate the performance of the proposed method and compare it with other methods. In terms of data simulation, three attack methods are simulated, including load redistribution attack [10], blind FDIA [17] and the direct replacement of data by copying a period of data from another time. Load redistribution attacks refer to the malicious behaviors that make the grid operator have an inaccurate perception of load demands. Load redistribution attacks are generally achieved by direct invading of EMS computers. Blind FDIA means that the attackers do not have the knowledge of system parameters and attempt to manipulate the measurement data through the wireless communication network. Since this is a semi-supervised learning situation, there are 1000 labeled data with (33% FDIA data and 67% normal operating data), 6500 unlabeled normal operating data, and 2500 testing data (also with 33% FDIA data and 67% normal operating data). As mentioned above, the potential attacks in the future may be dissimilar to historical attacks, so we make the training data different from the testing data by including random attacks (randomly-generated FDIA data) in the testing data. The proposed method is achieved by Pytorch. The evaluation metrics used for FDIA detection include the well-known mean classification accuracy and $f1$ index.

As mentioned above, PMUs are difficult to be deployed at all nodes due to the very high expense of PMUs recently. In case studies, the weighted least square (WLS) based PMU placement strategy [50] is used. The detailed deployments of PMUs are shown in Figs. 12 and 13 in Appendix. SCADA covers all nodes, and measurements of nodal voltage magnitudes, nodal injected active and reactive power are used. For PMU measurements, nodal voltage magnitudes and angles, as well as nodal injected current magnitudes and angles, are selected. Thus, there are at total 3 kinds of measurement variables from SCADA while 4 types of variables from PMUs. The SCADA and PMU sampling rates are assumed as 0.5 Hz and 30 Hz. The time scale of data window is set as 1 min, thereby the sampling data of SCADA and PMUs have 30 and 1800 points. Moreover, we set the length of low-stage and high-stage features as 120 and 480 points. Then, to precess SCADA data, low-stage features, high-stage features and PMU data, numerous FC layers are employed to make them into the identical size 30×1 and integrate them. In terms of the training process, the Adam algorithm [51] with learning rate 1×10^{-4} is used. The detailed parameters settings of the proposed method are listed in Table 1.

4.2. Effectiveness of design of the proposed method

To prove the effectiveness of the structure of the presented method, we test the weakened method without partial structure in the proposed method. In detail, these changes include: (1) only use SCADA data; (2) only use PMU data; (3) use PMU and SCADA data without low-stage and high-stage features extraction; (4) without the discriminator; (5)

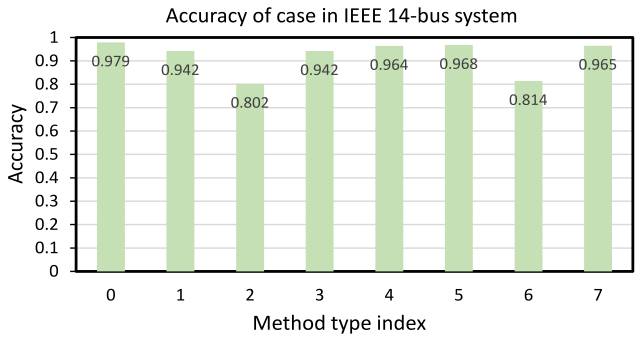


Fig. 8. The mean accuracy of all weakened methods in IEEE 14-bus system. The abscissa is the index of method type, where the method 0 refers to the entire proposed method. The ordinate is the detection accuracy, and the values are placed near to the top of the bar.

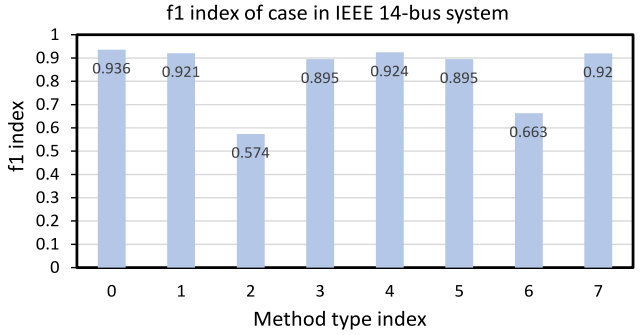


Fig. 9. The $f1$ index of all weakened methods in IEEE 14-bus system. The abscissa is the index of method type, where the method 0 refers to the entire proposed method. The ordinate is the $f1$ index, and the values are placed near to the top of the bar.

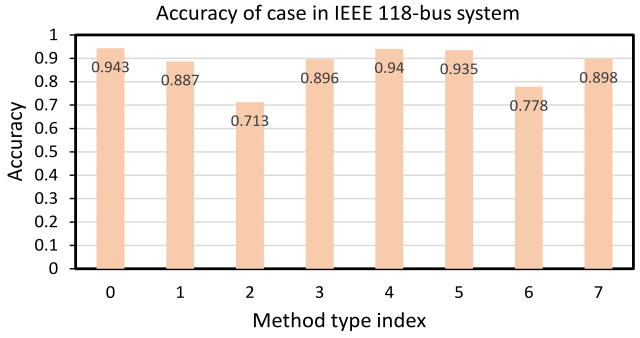


Fig. 10. The mean accuracy of all weakened methods in IEEE 118-bus system. The abscissa is the index of method type, where the method 0 refers to the entire proposed method. The ordinate is the detection accuracy, and the values are placed near to the top of the bar.

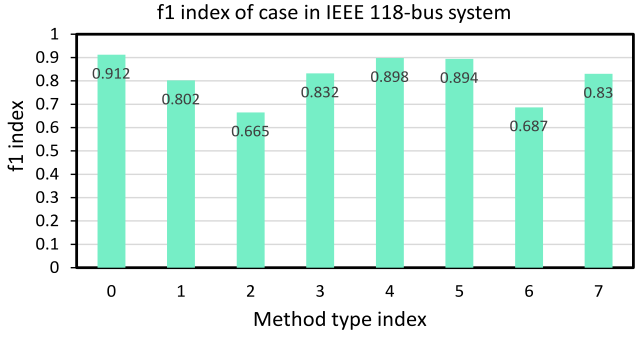


Fig. 11. The $f1$ index of all weakened methods in IEEE 118-bus system. The abscissa is the index of method type, where the method 0 refers to the entire proposed method. The ordinate is the $f1$ index, and the values are placed near to the top of the bar.

Table 1
The parameters settings of the proposed method.

System	IEEE 14-bus	IEEE 118-bus
Nodes	14	118
PMUs	4	32
SCADA	14	118
PMU measurements	4 × 4	4 × 32
SCADA measurements	3 × 14	3 × 118
Time scale	1 min	1 min
Length of SCADA data	30	30
Length of low-stage features	120	120
Length of high-stage features	480	480
Length of PMU data	1800	1800
FC layers of SCADA data	30 × 30	30 × 30
FC layers of low-stage features	120 × 30	120 × 30
FC layers of high-stage features	480 × 120 120 × 30	480 × 120 120 × 30
FC layers of PMU data	1800 × 360 360 × 30	1800 × 360 360 × 30
FC layers in the discriminator	3	3
Multi-head attention block	Layers: 3 Heads: 16	Layers: 3 Heads: 16
Multi-head attention in unsupervised branch	Layers: 3 Heads: 16	Layers: 3 Heads: 16
FC layers in supervised branch	4	6
Intervals for piece-wise FC layers	16	32
Batch size	256	528
Learning rate	0.0001	0.0001

without piece-wise loss; (6) only utilize unsupervised branch; (7) only utilize supervised branch. The first four weakened methods are related to the inputs $\mathbf{O}_{\mathbf{FU}}$ for FDIA detection network, i.e., the outputs of the resolution enlargement framework. The first three scenarios imply that there is no resolution enlargement framework, and we directly input SCADA and PMU data. The fourth scenario is that no discriminator exists, suggesting that the resolution enlargement framework is merely a mapping from SCADA to PMU data, instead of the GAN structure. Furthermore, the fifth weakened method is that the well-known cross-entropy is used rather than the proposed piece-wise loss at the supervised branch. Finally, the sixth and seventh scenarios are that only unsupervised or supervised learning branch is used, namely, they become unsupervised or supervised FDIA detection methods.

The performances of the proposed approach along with various weakened methods are illustrated in Figs. 8 to 11. These figures display the mean accuracy and $f1$ score for IEEE 14-bus and 118-bus systems, respectively. The results indicate that the proposed method consistently achieves the best performance. In contrast, the performance in weakened method 2 is notably poor, primarily because PMUs cannot be installed at all nodes due to their high costs. Consequently, relying solely on PMU data may overlook partial critical information of nodes without PMUs. Scenario 6 also exhibits weak performance, as it relies entirely on the unsupervised learning approach. This method depends solely on unlabeled normal operational data, failing to utilize any labeled FDIA data, which compromises its effectiveness. The results from scenarios 1, 3, and 4 highlight the advantages of the multi-resolution features extracted by our resolution enlargement framework, which significantly enhances the FDIA detection. Furthermore, the superior performance in scenarios 5 and 7 demonstrates that the proposed piece-wise loss guided semi-supervised learning approach effectively improves detection capabilities when faced with limited FDIA data.

4.2.1. Computational time

It is essential to report the operational time of the proposed method, which is calculated using NVIDIA GeForce RTX 3050. The results are presented in Table 2, where the batch size refers to the number of

Table 2
The computation time (Unit: ms) of an epoch with various batch sizes.

Batch size	128	256	528	1024
IEEE 14-bus, training	25.2	64.6	116.6	243.5
IEEE 14-bus, testing	18.2	22.2	31.2	64.5
IEEE 118-bus, training	124.6	278.0	436.7	890.7
IEEE 118-bus, testing	72.2	93.2	116.8	235.6

Table 3
The comparison results.

Type	Methods	IEEE 14		IEEE 118	
		Accuracy	$f1$ index	Accuracy	$f1$ index
Supervised	KNN	0.670	0.279	0.701	0.215
	RF	0.788	0.524	0.734	0.602
	GBDT	0.864	0.736	0.766	0.613
	LSTM	0.871	0.772	0.833	0.755
	biLSTM	0.902	0.865	0.889	0.779
	Transformer	0.941	0.921	0.887	0.802
Unsupervised	SVDD	0.642	0.322	0.655	0.389
	GAN	0.812	0.699	0.789	0.661
	VAE	0.745	0.356	0.712	0.425
Semi- supervised	S3VM	0.853	0.701	0.803	0.522
	SSVAE	0.872	0.722	0.788	0.633
	Proposed	0.979	0.936	0.943	0.912

data sequences processed concurrently. We, respectively, count the computation time of one epoch with batch size 128, 256, 528, 1024, with time unit millisecond, as shown in Table 2. It should be noted that the training time is always larger than the testing time because the training also includes parameters updating process.

4.3. Comparison with other methods

Other deep learning methods are utilized for comparison, including K-nearest neighborhood (KNN), random forest (RF), gradient-boosted decision trees (GBDT), LSTM [52], bidirectional LSTM (biLSTM) [52], transformer [47], deep support vector data description (DSVDD) [53], GAN [46], variational autoencoder (VAE) [54], semi-supervised support vector machines (S3VM) [55], semi-supervised variational autoencoder (SSVAE) [56], categorized as supervised, unsupervised and semi-supervised learning shown in Table 3. It is noteworthy that the inputting data of these methods are totally different. For supervised learning approaches, they can only use labeled FDIA data, while unsupervised learning approaches are only able to utilize normal operating data, as well as semi-supervised learning methods can use both.

The comparison results are presented in Table 3. As shown, the proposed method significantly outperforms other approaches. This high accuracy is primarily attributed to the resolution enlargement framework and the piece-wise loss function. The resolution enlargement framework generates multi-resolution features that effectively leverage both low-frequency SCADA data and high-frequency PMU data. Meanwhile, the piece-wise loss function enhances the impact of the limited effective FDIA data.

Additionally, semi-supervised learning approach capitalizes on the combination of scarce labeled data and abundant unlabeled normal data, which improves the precision of FDIA data identification. Thus, the proposed piece-wise loss guided semi-supervised method proves to be more effective in scenarios with limited FDIA data compared to traditional semi-supervised learning techniques.

4.4. Sensitivity to the amount of FDIA data and time scales

The proposed approach effectively leverages both scarce FDIA data and a substantial volume of normal operational data. To validate its robustness in the face of data scarcity, it is essential to assess the performance of the proposed method using varying amounts of FDIA data. In previous cases, the ratio of labeled FDIA data is set as 13.3%. In this context, we will examine scenarios with sparser FDIA data to further evaluate the effectiveness of the presented method.

Table 4
The performance under various ratios of FDIA data amount.

Index	System	1%	3%	5%	13.3%
Accuracy	IEEE 14	0.976	0.975	0.978	0.979
$f1$ index	IEEE 14	0.896	0.918	0.934	0.936
Accuracy	IEEE 118	0.942	0.943	0.945	0.943
$f1$ index	IEEE 118	0.891	0.902	0.904	0.912

Table 5
The performance under various settings of time window.

Index	System	10 s	30 s	1 min	2 min
Accuracy	IEEE 14	0.904	0.969	0.979	0.981
$f1$ index	IEEE 14	0.835	0.924	0.936	0.935
Accuracy	IEEE 118	0.887	0.940	0.943	0.943
$f1$ index	IEEE 118	0.805	0.905	0.912	0.913

The results are presented in Table 4, where the mean accuracy remains stable even with very small amounts of FDIA data, such as 1% and 3% compared to normal operational data. However, the $f1$ index shows a slight decline in the scenario with a 1% FDIA data ratio. This decrease is attributed to the extreme sparsity of FDIA data, implying that the characteristics of FDIA are not fully revealed. Nevertheless, the proposed method still achieves a high $f1$ score, demonstrating its effectiveness despite the limited FDIA data.

We still test the sensitivity of the proposed approach when data sequences with different time scales are input. As shown in Table 5, a very short time scale, such as 10 s, hinders the performance of the proposed approach, evidently due to that the time series characteristics are difficult to reveal by too short data sequences. The performance when inputting data sequences of 30 s is slightly worse than those of 1 and 2 min. In addition, longer data sequences are difficult to bring about a significant improvement.

5. Conclusion

We conduct the comprehensive analysis on the detective data regions of multiple FDIA detection methods in literature. The results demonstrate that the data region generated through the fusion of PMU and SCADA data is more resilient to malicious data compared to the data region based on power grid physical equations, operating constraints, common operating states, and temporal characteristics. Accordingly, this article proposes a multi-resolution data fusion learning approaches for FDIA detection in smart grids. We still take into account the imbalance between rare labeled FDIA data and large amounts of normal operating data. We design experimental cases to test the effectiveness of every part in the presented approach, and the comparisons with other approaches are conducted. In the future, we aim to consider the scenario that the time stamps of SCADA systems are not precise, such as data latency and data time misalignment, to improve the robustness and generality of the proposed method.

CRediT authorship contribution statement

Haosen Yang: Writing – original draft, Validation, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Bifei Tan:** Project administration, Methodology, Investigation, Funding acquisition. **Zipeng Liang:** Data curation. **Xin Shi:** Methodology. **Hanjiang Dong:** Software. **Chongyu Wang:** Validation. **Shibo Chen:** Validation.

Declaration of competing interest

We declare that there is not any conflict of interests here.

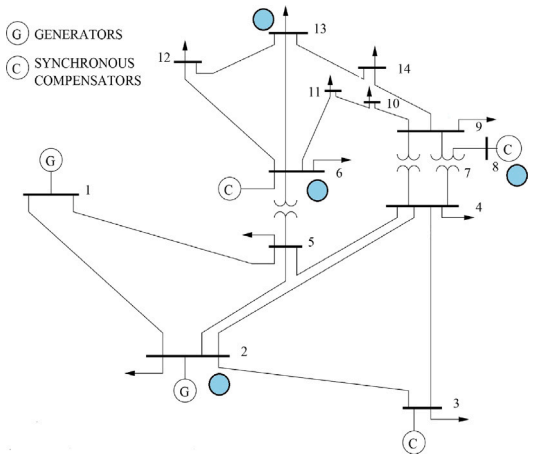


Fig. 12. The PMU deployment in IEEE 14-bus system. The nodes with installed PMUs include 2, 6, 8, 13.

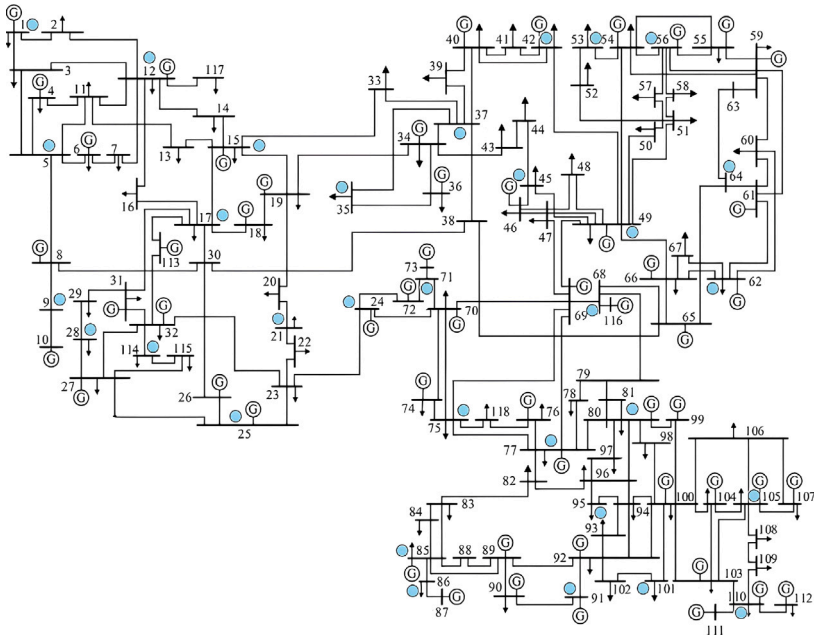


Fig. 13. The PMU deployment in IEEE 118-bus system. The deployed PMUs are at node 1, 5, 9, 12, 15, 17, 21, 24, 25, 28, 35, 37, 42, 45, 49, 53, 56, 62, 64, 69, 71, 75, 77, 81, 85, 86, 91, 95, 101, 105, 110, 114.

Acknowledgments

This paper is supported by Department of Education of Guangdong Province's 2024 Ordinary University Accreditation Research Project (2024KQNCX033) and Jiangmen Basic and Theoretical Science Research Project (2023JC01018).

Appendix. PMU placement

The detailed deployments of PMUs are shown in Figs. 12 and 13.

Data availability

Data will be made available on request.

References

[1] Y. Liu, P. Ning, M.K. Reiter, False data injection attacks against state estimation in electric power grids, ACM Trans. Inf. Syst. Secur. 14 (1) (2011) 1–33.

[2] H. Yang, X. He, Z. Wang, R.C. Qiu, Q. Ai, Blind false data injection attacks against state estimation based on matrix reconstruction, IEEE Trans. Smart Grid 13 (4) (2022) 3174–3187, http://dx.doi.org/10.1109/TSG.2022.3164874.

[3] Z.-H. Yu, W.-L. Chin, Blind false data injection attack using PCA approximation method in smart grid, IEEE Trans. Smart Grid 6 (3) (2015) 1219–1226.

[4] H. Yang, W. Zhang, C.Y. Chung, Z. Wang, W. Qiu, Z. Liang, AC false data injection attack based on robust tensor principle component analysis, IEEE Trans. Ind. Informatics (2024) 1–11.

[5] X. Liu, Z. Li, Z. Shuai, Y. Wen, Cyber attacks against the economic operation of power systems: A fast solution, IEEE Trans. Smart Grid 8 (2) (2017) 1023–1025, http://dx.doi.org/10.1109/TSG.2016.2623983.

[6] X. Liu, Z. Li, Local load redistribution attacks in power systems with incomplete network information, IEEE Trans. Smart Grid 5 (4) (2014) 1665–1676, http://dx.doi.org/10.1109/TSG.2013.2291661.

[7] R.-P. Liu, X. Wang, B. Zeng, R. Zgheib, Modeling load redistribution attacks in integrated electricity-gas systems, IEEE Trans. Smart Grid (2024) 1, http://dx.doi.org/10.1109/TSG.2024.3350992.

[8] G. Liang, S.R. Weller, F. Luo, J. Zhao, Z.Y. Dong, Generalized FDIA-based cyber topology attack with application to the Australian electricity market trading mechanism, IEEE Trans. Smart Grid 9 (4) (2018) 3820–3829, http://dx.doi.org/10.1109/TSG.2017.2677911.

[9] G. Liang, S.R. Weller, J. Zhao, F. Luo, Z.Y. Dong, A framework for cyber-topology attacks: Line-switching and new attack scenarios, IEEE Trans. Smart Grid 10 (2) (2019) 1704–1712, http://dx.doi.org/10.1109/TSG.2017.2776325.

[10] Q. Zhang, F. Li, Q. Shi, K. Tomsovic, J. Sun, L. Ren, Profit-oriented false data injection on electricity market: Reviews, analyses, and insights, IEEE Trans. Ind. Informatics 17 (9) (2021) 5876–5886, http://dx.doi.org/10.1109/TII.2020.3036104.

[11] C. Liu, H. Liang, T. Chen, Network parameter coordinated false data injection attacks against power system AC state estimation, IEEE Trans. Smart Grid 12 (2) (2021) 1626–1639, http://dx.doi.org/10.1109/TSG.2020.3033520.

[12] L. Che, X. Liu, Z. Li, Y. Wen, False data injection attacks induced sequential outages in power systems, IEEE Trans. Power Syst. 34 (2) (2019) 1513–1523, http://dx.doi.org/10.1109/TPWRS.2018.2871345.

[13] L. Che, X. Liu, Z. Shuai, Z. Li, Y. Wen, Cyber cascades screening considering the impacts of false data injection attacks, IEEE Trans. Power Syst. 33 (6) (2018) 6545–6556, http://dx.doi.org/10.1109/TPWRS.2018.2827060.

[14] J. Kim, L. Tong, R.J. Thomas, Subspace methods for data attack on state estimation: A data driven approach, IEEE Trans. Signal Process. 63 (5) (2015) 1102–1114, http://dx.doi.org/10.1109/TSP.2014.2385670.

[15] W.-L. Chin, C.-H. Lee, T. Jiang, Blind false data attacks against AC state estimation based on geometric approach in smart grid communications, IEEE Trans. Smart Grid 9 (6) (2018) 6298–6306.

[16] S. Lakshminarayana, A. Kammoun, M. Debbah, H.V. Poor, Data-driven false data injection attacks against power grids: A random matrix approach, IEEE Trans. Smart Grid 12 (1) (2021) 635–646.

[17] H. Yang, Z. Wang, A false data injection attack approach without knowledge of system parameters considering measurement noise, IEEE Internet Things J. (2023) 1, http://dx.doi.org/10.1109/JIOT.2023.3288983.

[18] A. Abur, A.G. Exposito, Power System State Estimation: Theory and Implementation, CRC Press, Boca Raton, Florida, USA, 2004.

[19] M. Gol, A. Abur, LAV based robust state estimation for systems measured by PMUs, IEEE Trans. Smart Grid 5 (4) (2014) 1808–1814.

[20] K.-D. Lu, G.-Q. Zeng, X. Luo, J. Weng, W. Luo, Y. Wu, Evolutionary deep belief network for cyber-attack detection in industrial automation and control system, IEEE Trans. Ind. Informatics 17 (11) (2021) 7618–7627, http://dx.doi.org/10.1109/TII.2021.3053304.

[21] I. Sohn, Deep belief network based intrusion detection techniques: A survey, Expert Syst. Appl. 167 (2021) 114170, http://dx.doi.org/10.1016/j.eswa.2020.114170, URL https://www.sciencedirect.com/science/article/pii/S0957417420309088.

[22] W. Qiu, Q. Tang, K. Zhu, W. Wang, Y. Liu, W. Yao, Detection of synchrophasor false data injection attack using feature interactive network, IEEE Trans. Smart Grid 12 (1) (2021) 659–670, http://dx.doi.org/10.1109/TSG.2020.3014311.

[23] A. Parizad, C.J. Hatziadoniu, Cyber-attack detection using principal component analysis and noisy clustering algorithms: A collaborative machine learning-based framework, IEEE Trans. Smart Grid 13 (6) (2022) 4848–4861, http://dx.doi.org/10.1109/TSG.2022.3176311.

[24] X. Yang, P. Zhao, X. Zhang, J. Lin, W. Yu, Toward a Gaussian-mixture model-based detection scheme against data integrity attacks in the smart grid, IEEE Internet Things J. 4 (1) (2017) 147–161, http://dx.doi.org/10.1109/JIOT.2016.2631520.

[25] D. Xue, X. Jing, H. Liu, Detection of false data injection attacks in smart grid utilizing ELM-based OCON framework, IEEE Access 7 (2019) 31762–31773, http://dx.doi.org/10.1109/ACCESS.2019.2902910.

[26] G. Zhang, J. Li, O. Bamisile, D. Cai, W. Hu, Q. Huang, Spatio-temporal correlation-based false data injection attack detection using deep convolutional neural network, IEEE Trans. Smart Grid 13 (1) (2022) 750–761, http://dx.doi.org/10.1109/TSG.2021.3109628.

[27] S. Wang, S. Bi, Y.-J.A. Zhang, Locational detection of the false data injection attack in a smart grid: A multilabel classification approach, IEEE Internet Things J. 7 (9) (2020) 8218–8227, http://dx.doi.org/10.1109/JIOT.2020.2983911.

[28] W. Qiu, Q. Tang, K. Zhu, W. Wang, Y. Liu, W. Yao, Detection of synchrophasor false data injection attack using feature interactive network, IEEE Trans. Smart Grid 12 (1) (2021) 659–670, http://dx.doi.org/10.1109/TSG.2020.3014311.

[29] H. Li, C. Dou, D. Yue, G.P. Hancke, Z. Zeng, W. Guo, L. Xu, End-edge-cloud collaboration based false data injection attack detection in distribution networks, IEEE Trans. Ind. Informatics (2023) 1–13, http://dx.doi.org/10.1109/TII.2023.3281664.

[30] W. Lei, Z. Pang, H. Wen, W. Hou, W. Han, FDI attack detection at the edge of smart grids based on classification of predicted residuals, IEEE Trans. Ind. Informatics 18 (12) (2022) 9302–9311, http://dx.doi.org/10.1109/TII.2022.3174159.

[31] J.J.Q. Yu, Y. Hou, V.O.K. Li, Online false data injection attack detection with wavelet transform and deep neural networks, IEEE Trans. Ind. Informatics 14 (7) (2018) 3271–3280, http://dx.doi.org/10.1109/TII.2018.2825243.

[32] O. Boyaci, M.R. Narimani, K.R. Davis, M. Ismail, T.J. Overbye, E. Serpedin, Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks, IEEE Trans. Smart Grid 13 (1) (2022) 807–819, http://dx.doi.org/10.1109/TSG.2021.3117977.

[33] Z. Du, Z. Yan, Y. Xu, A dimensional augmentation-based data-driven method for detecting false data injection in smart meters, IEEE Trans. Smart Grid (2023) 1, http://dx.doi.org/10.1109/TSG.2023.3309530.

[34] Y. Li, X. Wei, Y. Li, Z. Dong, M. Shahidehpour, Detection of false data injection attacks in smart grid: A secure federated deep learning approach, IEEE Trans. Smart Grid 13 (6) (2022) 4862–4872, http://dx.doi.org/10.1109/TSG.2022.3204796.

[35] M.M.N. Aboelwafa, K.G. Seddik, M.H. Eldefrawy, Y. Gadallah, M. Gidlund, A machine-learning-based technique for false data injection attacks detection in industrial IoT, IEEE Internet Things J. 7 (9) (2020) 8462–8471, http://dx.doi.org/10.1109/JIOT.2020.2991693.

[36] J. Wang, D. Shi, Y. Li, J. Chen, H. Ding, X. Duan, Distributed framework for detecting PMU data manipulation attacks with deep autoencoders, IEEE Trans. Smart Grid 10 (4) (2019) 4401–4410, http://dx.doi.org/10.1109/TSG.2018.2859339.

[37] A. Bhattacharjee, A.K. Mondal, A. Verma, S. Mishra, T.K. Saha, Deep latent space clustering for detection of stealthy false data injection attacks against AC state estimation in power systems, IEEE Trans. Smart Grid 14 (3) (2023) 2338–2351, http://dx.doi.org/10.1109/TSG.2022.3216625.

[38] Y. Zhang, J. Wang, B. Chen, Detecting false data injection attacks in smart grids: A semi-supervised deep learning approach, IEEE Trans. Smart Grid 12 (1) (2021) 623–634, http://dx.doi.org/10.1109/TSG.2020.3010510.

[39] C. Chen, Y. Wang, M. Cui, J. Zhao, W. Bi, Y. Chen, X. Zhang, Data-driven detection of stealthy false data injection attack against power system state estimation, IEEE Trans. Ind. Informatics 18 (12) (2022) 8467–8476, http://dx.doi.org/10.1109/TII.2022.3149106.

[40] M. Abdel-Basset, H. Hawash, R.K. Chakrabortty, M.J. Ryan, Semi-supervised spatiotemporal deep learning for intrusions detection in IoT networks, IEEE Internet Things J. 8 (15) (2021) 12251–12265, http://dx.doi.org/10.1109/JIOT.2021.3060878.

[41] M. Farajzadeh-Zanjani, E. Hallaji, R. Razavi-Far, M. Saif, M. Parvania, Adversarial semi-supervised learning for diagnosing faults and attacks in power grids, IEEE Trans. Smart Grid 12 (4) (2021) 3468–3478, http://dx.doi.org/10.1109/TSG.2021.3061395.

[42] W. Qiu, Q. Tang, J. Liu, W. Yao, An automatic identification framework for complex power quality disturbances based on multifusion convolutional neural network, IEEE Trans. Ind. Informatics 16 (5) (2020) 3233–3241, http://dx.doi.org/10.1109/TII.2019.2920689.

[43] W. Qiu, Q. Tang, Y. Wang, L. Zhan, Y. Liu, W. Yao, Multi-view convolutional neural network for data spoofing cyber-attack detection in distribution synchrophasors, IEEE Trans. Smart Grid 11 (4) (2020) 3457–3468, http://dx.doi.org/10.1109/TSG.2020.2971148.

[44] N. Costilla-Enriquez, Y. Weng, Attack power system state estimation by implicitly learning the underlying models, IEEE Trans. Smart Grid 14 (1) (2023) 649–662, http://dx.doi.org/10.1109/TSG.2022.3197770.

[45] H. Yang, X. Shi, R.C. Qiu, X. He, Q. Ai, Z. Wang, Monitoring data factorization of high renewable energy penetrated grids for probabilistic static voltage stability assessment, IEEE Trans. Smart Grid 13 (2) (2022) 1273–1286, http://dx.doi.org/10.1109/TSG.2021.3128503.

[46] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, Vol. 27, 2014.

[47] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, Vol. 30, 2017.

[48] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 770–778, http://dx.doi.org/10.1109/CVPR.2016.90.

[49] J.L. Ba, J.R. Kiros, G.E. Hinton, Layer normalization, 2016, arXiv preprint arXiv:1607.06450.

[50] N.M. Manousakis, G.N. Korres, A weighted least squares algorithm for optimal PMU placement, IEEE Trans. Power Syst. 28 (3) (2013) 3499–3500.

[51] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2017, arXiv preprint arXiv:1412.6980.

[52] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, http://dx.doi.org/10.1162/neco.1997.9.8.1735.

[53] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S.A. Siddiqui, A. Binder, E. Müller, M. Kloft, Deep one-class classification, in: International Conference on Machine Learning, 2018, pp. 4393–4402.

[54] D.P. Kingma, M. Welling, Auto-encoding variational Bayes, 2013, arXiv preprint arXiv:1803.01271.

[55] K. Bennett, A. Demiriz, Semi-supervised support vector machines, in: M. Kearns, S. Solla, D. Cohn (Eds.), Advances in Neural Information Processing Systems, Vol. 11, MIT Press, 1998.

[56] D.P. Kingma, S. Mohamed, D. Jimenez Rezende, M. Welling, Semi-supervised learning with deep generative models, Adv. Neural Inf. Process. Syst. 27 (2014).