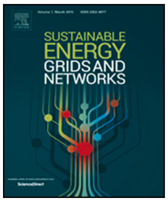
Sustainable Energy, Grids and Networks 36 (2023) 101219

Contents lists available at ScienceDirect

Sustainable Energy, Grids and Networks

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



Deep semi-supervised electricity theft detection in AMI for sustainable and secure smart grids

Ruobin Qi [a], Qingqing Li [b], Zhirui Luo [a], Jun Zheng [a,*], Sihua Shao [c]

[a] Department of Computer Science and Engineering, New Mexico Institute of Mining and Technology, 801 Leroy Place, Socorro, NM, 87801, USA
[b] Department of Computer and Information Sciences, Towson University, 8000 York Road, Towson, MD, 21252, USA
[c] Department of Electrical Engineering, New Mexico Institute of Mining and Technology, 801 Leroy Place, Socorro, NM, 87801, USA

ARTICLE INFO

Keywords:
Smart grid
Advanced metering infrastructure (AMI)
Electricity theft detection
Machine learning
Semi-supervised anomaly detection
Observer meter

ABSTRACT

Electricity theft is a major issue that affects the sustainability and security of smart grids. This paper proposes a deep semi-supervised method for electricity theft detection in the Advanced Metering Infrastructure (AMI) of smart grids. By only using normal samples to train the detection model, the proposed method has the capability of detecting unknown attacks in a short time frame. The method utilizes the ratio profile generated from the readings of the observer meter and the user's smart meter as the input, to reduce false positives, which is then transformed into a 2D image with the continuous wavelet transform (CWT) to capture the time–frequency information. Discriminative features are extracted from the CWT image with a deep convolutional autoencoder (CAE) and principal component analysis (PCA), which are fed into a semi-supervised autoencoder for classification. The performance of the proposed method was evaluated and compared with a set of baselines and four supervised machine learning and deep learning methods under 11 different false data injection (FDI) attacks using smart meter data from both business and residential users. The results show that the proposed method significantly outperforms the baselines and is more capable of detecting unknown attacks than supervised methods.

1. Introduction

Smart grids feature bi-directional power and information flow to provide more efficient and resilient power transmission, distribution, and management than traditional power networks [1]. The Advanced Metering Infrastructure (AMI) is the essential part of smart grids, consisting of smart meters, communication modules, and Meter Data Management Systems (MDMS) [2,3]. They are responsible for collecting large amounts of high-frequency electricity consumption data from customers, sending the collected information to the analysis computer and receiving operation commands from the operation center, and long-term data storage and event management, respectively [3].

While smart grids bring significant benefits, they are also exposed to potential risks and threats. As the essential component of smart grids, the AMI faces a variety of cyber-attacks, including the electricity theft attacks [4,5], in which fraudsters reduce their electricity bills by injecting false data into smart meters to trick the utility companies. According to the statistics [6], global economic losses due to electricity theft amount to $89.3 billion annually. A majority of the losses come from emerging countries, about $58.7 billion. In the United States, the revenue loss due to non-technical loss (NTL) including electricity theft also estimated to be $6.5 billion by the Electric Power Research Institute (EPRI) [7]. In addition to financial losses, electricity theft could lead to power quality degradation due to underestimated demands and even raise safety concerns [8]. Therefore, methods for efficient and timely detection of electricity theft attacks are greatly needed for the sustainability and security of smart grids.

Traditional electricity theft detection methods include sending employees to customers' homes to check facilities [8], installing additional smart meters for verification [9,10], or going door-to-door to verify the meter data [11]. All these methods have significant drawbacks, such as being labor-consuming, inefficient, or having high costs. With the growing amount of smart meter data acquired by AMI, data-driven electricity theft detection using machine learning has become popular in recent years. Many of the proposed methods are based on supervised machine learning and deep learning algorithms. Shallow machine learning algorithms like decision trees [12], support vector machines (SVM) [12–14], and extreme gradient boosting (XGBoost) [15,16], have been applied for detecting electricity theft using smart meter data. Due

[*] Corresponding author.
E-mail addresses: ruobin.qi@student.nmt.edu (R. Qi), qingqingli@towson.edu (Q. Li), zhirui.luo@student.nmt.edu (Z. Luo), jun.zheng@nmt.edu (J. Zheng), sihua.shao@nmt.edu (S. Shao).

https://doi.org/10.1016/j.segan.2023.101219
Received 17 March 2023; Received in revised form 4 November 2023; Accepted 9 November 2023
Available online 17 November 2023

to its promising performance in many areas, deep learning has been used for building electricity theft detection models in recent years. In [17], a wide & deep convolutional neural network (CNN) model was developed for electricity theft detection, which consists of a wide component and a deep component to learn the global knowledge and capture the periodicity of smart meter data, respectively. Inspired by the work of [17], Xia et al. [18] proposed an improved detection model that enhances the wide & deep CNN with dilated convolution and a channel-dimensional adaptive attention module. In [19], a hybrid deep model was built for electricity theft detection, which integrates a CNN and long short-term memory (LSTM). Electricity theft detection methods based on supervised learning require data and labels from both normal users and fraudulent users to train the models. However, the scarcity of real electricity theft data leads to the data imbalance problem, which causes the overfitting of trained models. In addition, the performance of supervised learning-based detection methods generally degrades significantly when dealing with unknown attacks [13,20]. Unlike supervised learning, unsupervised learning-based methods [1,21–24] only use unlabeled data to identify potential fraudulent users, which build detection models by using clustering and/or correlation analysis. Unsupervised learning-based detection methods have some limitations, such as a relatively long detection time and limited detection performance on certain attack types.

To overcome the limitations of methods based on supervised learning and unsupervised learning, we propose a new electricity theft detection method based on semi-supervised outlier detection in this paper. Fig. 1 illustrates the distinction between semi-supervised outlier detection and supervised learning. It can be seen that supervised learning utilizes both normal and fraudulent samples to train the detection model, whereas semi-supervised outlier detection relies solely on normal samples for training the detection model. Consequently, the semi-supervised model possesses the capability to detect unknown attack samples with different characteristics than the existing fraudulent samples, whereas the supervised model often falters in such cases. To the best of our knowledge, there are only a few works in the literature that use semi-supervised outlier detection for electricity theft detection. For example, one-class SVM (OCSVM) was tried in [13] but the detection performance was poor. The main contributions of our work are summarized as follows: (1) We propose a new deep semi-supervised electricity theft detection method which utilizes the ratio profile generated from the readings of the observer meter and the user's smart meter as input, which can significantly reduce false positives of detection due to low electricity usage; (2) the proposed method extracts discriminative features from the ratio profile through a combination of continuous wavelet transform (CWT)-based time–frequency representation, deep feature extraction with a Convolutional Autoencoder (CAE), and dimension reduction using principal component analysis (PCA); (3) the proposed method utilizes a semi-supervised autoencoder classifier for the classification of the extracted features; (4) we adopted two publicly available smart meter datasets, one for business users and one for residential users, to evaluate the performance of the proposed method and compare it with a set of semi-supervised baseline methods; (5) we also compared the proposed method with popular supervised machine learning and deep learning methods for detecting unknown attacks.

The rest of this paper is organized as follows. The background information about the AMI architecture and false data injection (FDI) attacks are introduced in Section 2. Section 3 presents the details of the proposed deep semi-supervised method for electricity theft detection. The performance evaluation experiments and results are described in Section 4. We finally conclude the paper in Section 5.

2. Background

2.1. AMI architecture

Fig. 2 is an illustration of the system architecture of AMI. In the home area networks (HAN), smart meters installed in residential houses or business buildings collect a large amount of high-resolution electricity consumption data, which is then sent to the concentrators through the neighborhood area networks (NAN). Each concentrator is responsible for gathering the smart meter data collected by a group of smart meters in a neighborhood area. The concentrators are connected to the control center through the wide area network (WAN). The control center manages the collected smart meter data and performs electricity distribution automation and other intelligent applications. In the AMI architecture, one or more observer meters can be installed in a NAN, and each of them is responsible for recording the total power consumption of a group of consumers. Observer meters can be integrated with concentrators, which are much harder to hack and tamper with by attackers than smart meters [1,22,24]. Therefore, it is reasonable to assume that the data recorded by the observer meter is valid and reliable.

Table 1
Eleven types of FDIAs.

Attack type	Modification
1	$\tilde{x}_t = \alpha x_t,\ 0.2 < \alpha < 0.8$
2	$\tilde{x}_t = f(t)x_t,\ f(t) = \begin{cases} 0 & \text{if } t_1 < t < t_2 \\ 1 & \text{otherwise} \end{cases}$
3	$\tilde{x}_t = \alpha_t x_t,\ 0.2 < \alpha_t < 0.8$
4	$\tilde{x}_t = f(t)x_t,\ f(t) = \begin{cases} \alpha,\ 0.2 < \alpha < 0.8 & \text{if } t_1 < t < t_2 \\ 1 & \text{otherwise} \end{cases}$
5	$\tilde{x}_t = \alpha_t mean(\mathbf{x}),\ 0.2 < \alpha_t < 0.8$
6	$\tilde{x}_t = \begin{cases} x_t & \text{if } x_t \leq \delta \\ \delta & \text{if } x_t > \delta \end{cases} \quad \delta < max(\mathbf{x})$
7	$\tilde{x}_t = max(x_t - \gamma, 0),\ \gamma < max(\mathbf{x})$
8	$\tilde{x}_t = (1 - f(t))x_t,\ f(t) = \begin{cases} \alpha_{max} & t \geq t_{max} \\ \beta(t - t_s) & t_s < t < t_{max} \\ 0 & t < t_s \end{cases}$
9	$\tilde{x}_t = mean(\mathbf{x})$
10	$\tilde{x}_t = x_{N-t-1}$
11	$\tilde{x}_t = \begin{cases} x_t - \alpha x_t & 0.2 < \alpha < 0.8,\ t_1 < t < t_2 \\ x_t + \frac{\Delta}{N-n} & \text{otherwise} \end{cases}$

2.2. FDI attacks

Since it is hard if not impossible to collect real-world smart meter data of electricity thieves, research studies usually generate synthetic data for performance evaluation by modeling their malicious behaviors as various FDI attacks (FDIAs). In this study, we consider eleven types of FDIAs defined in Table 1, which have been widely used in other research work [13,15,19,24,25]. The eleven FDI attacks can be divided into two categories based on their attack characteristics: reduced consumption attacks and load profile shifting attacks [8]. Table 2 summarizes the symbols used throughout this paper and their definitions.

Reduced Consumption Attacks: In Table 1, types 1 to 8 attacks belong to the category of reduced consumption attacks which reduce smart meter readings to lower the attackers' electricity bills. The type 1 attack reduces all smart meter readings by multiplying them with an attack intensity factor α randomly generated in the range of $(0.2, 0.8)$. The type 2 attack randomly chooses a time interval of the day longer than 4 h and sets the readings in the interval to 0. All other readings remain the same. The type 3 attack is similar to the type 1 attack but randomly generates an attack intensity factor α_t for each smart meter reading instead of generating a factor for all readings. The type 4 attack reduces the smart meter readings in a selected time interval by a randomly generated attack intensity factor α in the range of $(0.2, 0.8)$. The type 5 attack replaces a smart meter reading x_t with the average daily consumption $mean(x)$ multiplied with a randomly

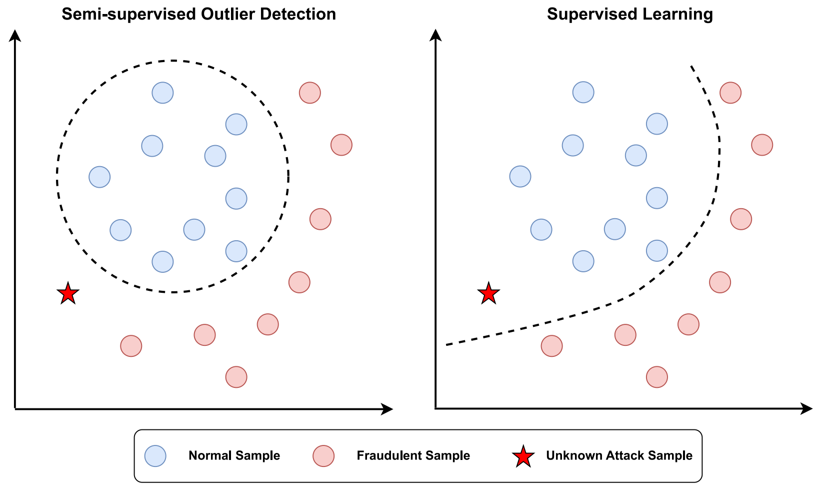


Fig. 1. Semi-supervised outlier detection vs. supervised learning.

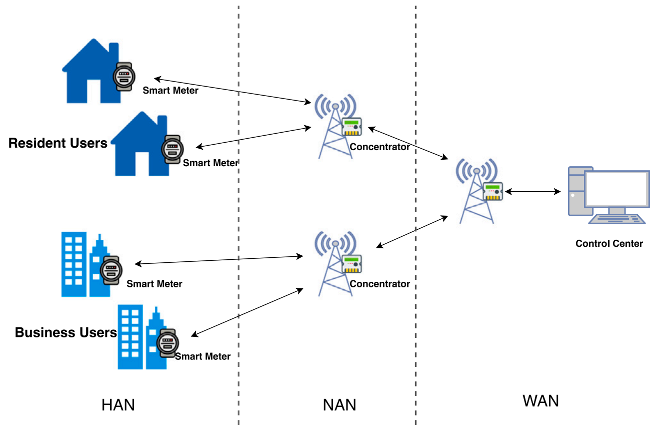


Fig. 2. The system architecture of AMI.

generated attack intensity factor α_t in the range of $(0.2, 0.8)$. The type 6 attack selects a cutoff threshold δ between 0 and the maximum reading of the day $max(x)$ and replaces any reading higher than δ with the value of δ. Other readings lower than or equal to δ remain the same. The type 7 attack selects a cutoff value γ between 0 and $max(x)$ which is subtracted from all readings. If the operation generates a negative value, the result will be set to 0. The type 8 attack randomly generates a maximum attack intensity α_{max} and gradually increases the attack intensity after the start time t_s with a change rate β. The attack intensity reaches α_{max} at t_{max} and remains at α_{max} after that.

Load Profile Shifting Attacks: Types 9 to 11 attacks defined in Table 1 belong to load profile shifting attacks which keep the total amount of daily power consumption the same but change the locations of the peaks and valleys in daily load profiles in different ways to subvert high electricity prices at specific time intervals set by the utility companies. The type 9 attack replaces each smart meter reading with the average consumption of the load profile $mean(\mathbf{x})$ while the type 10 attack flips the load profile. The type 11 attack selects the time interval of peak hours (t_1, t_2) which usually have high electricity prices, and reduces the n readings of the interval with a randomly generated attack intensity factor α in the range of $(0.2, 0.8)$. The total amount of reduced consumption in the selected time interval Δ is then evenly added back to the readings of other times to keep the total consumption unchanged.

A sample electricity consumption profile and its corresponding tampered profiles under reduced consumption attacks and load profile shifting attacks are shown in Figs. 3 and 4, respectively. Note that an electricity thief can tamper with the smart meter data by using a mixture of different types of FDIAs, which is also modeled in our experiments and noted as the MIX attack. Under the MIX attack, the fraudulent user randomly selects one of the eleven FDI attacks with equal chances on an attack day.

Table 2
Symbols used in this paper and their definitions.

Notations	Description
$\mathbf{x}$	Smart meter readings of a load profile
$\mathbf{x}'$	Readings of a load profile not including missing values
$mean(\mathbf{x})$	Average power consumption of a load profile
$max(\mathbf{x})$	Maximum power consumption of a load profile
$\sigma(\mathbf{x})$	Standard deviation of a load profile
N	Number of smart meter readings of a load profile
$t \in [0, N-1]$	Time to perform a smart meter reading
x_t	Real smart meter reading at t
$\tilde{x}_t$	Tampered smart meter reading at t
(t_1, t_2)	Selected time interval to launch an attack
α	Randomly generated attack intensity factor
α_t	Randomly generated attack intensity factor at t
δ	Cutoff threshold for FDIA 6
γ	Cutoff value for FDIA 7
α_{max}	Maximum attack intensity for FDIA 8
β	Attack intensity change rate for FDIA 8
Δ	Total reduced consumption in the attack interval for FDIA 11
t_s	Time to start an FDIA 8
t_{max}	Time to reach α_{max} for FDIA 8
n	Number of readings in a selected attack interval
O_t	Observer meter reading at t
$x_{i,t}$	ith user's smart meter reading at t
$R_{i,t}$	Ratio profile value for the ith user at t
$\mathbf{R}$	A 1-D daily ratio profile
$\mathbf{X}$	2-D CWT image generated from $\mathbf{R}$
k	Number of features obtained from the flatten layer of the CAE
$\mathbf{f}$	Feature vector extracted by the CAE
l	Vector length of $\mathbf{f}$
m	Number of instances in the training dataset
$\mathbf{f}_r$	Feature vector generated by PCA
r	Number of PCA features

3. Methodology

The workflow of the proposed deep semi-supervised method for electricity theft detection is illustrated in Fig. 5. In the following, we describe the details of the proposed method.

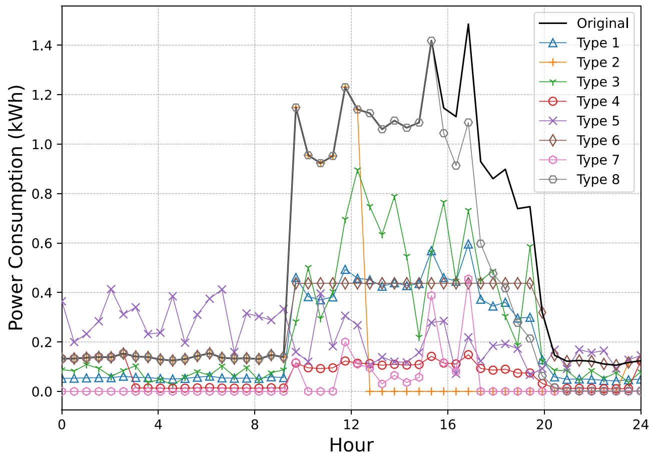


Fig. 3. A sample electricity consumption profile and its corresponding tampered profiles by applying eight types of reduced consumption attacks.

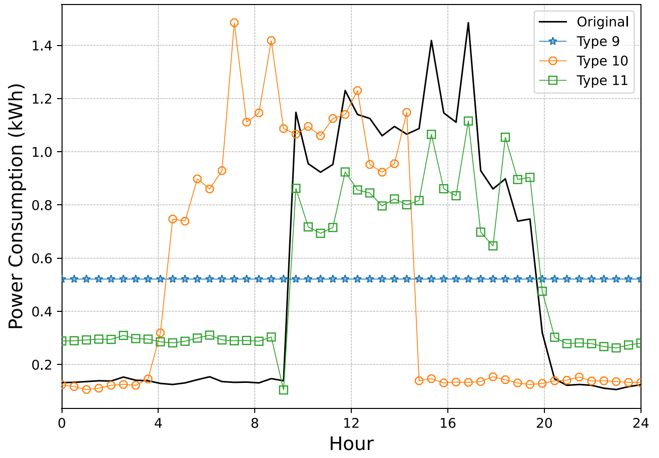


Fig. 4. A sample electricity consumption profile and its corresponding tampered profiles by applying three types of load profile shifting attacks.

3.1. Data pre-processing

Due to reasons like smart meter failures, transmission errors, poor connections, and system maintenance [17], there are often missing or erroneous values in smart meter readings. In our study, these data problems are dealt with in the pre-processing stage.

For missing values, we use the mean value method [15] to recover the data as shown in Eq. (1).

$$x_t = mean(\mathbf{x}') \quad \text{if} \quad x_t = NaN \tag{1}$$

For erroneous values, we use the “three-sigma rule of thumb” [15] to recover the data as shown in Eq. (2):

$$x_t = \begin{cases} \frac{x_{t-1}+x_{t+1}}{2} & \text{if } x_t > 3\sigma(\mathbf{x}) \quad \text{and} \quad x_{t-1},\ x_{t+1} \neq NaN \\ x_t & \text{otherwise} \end{cases} \tag{2}$$

3.2. Ratio profile

In general, users may have significantly different usage amounts on different days. For example, the electricity consumption of business users on weekends will be significantly lower than that of weekdays. Directly using load profiles to build detection models could result in a significant number of false positives. Thus, the ratio profile was proposed in [1] which is calculated as the ratio between the load profile of the observer meter and a user's load profile. Since the ratio of a user's consumption to the total consumption in an area is considered to be relatively stable and the readings of the observer meter are hard to be tampered with, malicious changes caused by FDI attacks will result in significant changes in ratio profiles which makes the detection easier. On the other hand, the ratio profile lessens the impact of low usage of normal users on certain days such that the false positives of detection can be significantly reduced.

The calculation of the ratio profile is shown in Eq. (3), where $R_{i,t}$ is the value of the ratio profile for the ith user in the area at time t, O_t and $x_{i,t}$ are the readings of the observer meter and the user's smart meter at time t, respectively:

$$R_{i,t} = \frac{O_t}{x_{i,t}} \tag{3}$$

Fig. 6 illustrates how ratio profile can help the detection of electricity theft. The two plots on the top of the figure show the untampered load profile of a business user over 10 days and the corresponding load profile of the observer meter covering the area, respectively. A usage pattern can be easily observed from the user's load profile that the user has high consumption on weekdays and very low consumption on weekends. The plot in the middle of the figures shows the ratio profile calculated based on Eq. (3) using the untampered user load profile. It can be seen that the ratio profile significantly reduces the effect of low consumption on weekends. The two plots at the bottom of the figure show the tampered load profile of the user and the corresponding ratio profile, respectively. For the tampered load profile, two reduced consumption attacks were applied on days 3 and 9, and two load profile shifting attacks were applied on days 2 and 4. The data of those tampered days in the tampered load profile and the corresponding ratio profile are colored in red. Apparently, the values of the ratio profile on those tampered days are significantly higher than those on untampered days which greatly facilitates the detection of malicious changes caused by FDIAs.

3.3. CWT

CWT provides an overcomplete representation of a signal by decomposing a continuous time function into several wavelets. Mathematically a time series function $x(t)$ could be transformed by Eq. (4), where $\bar{\psi}(t)$ is the mother wavelet which is continuous on both the time and frequency domains, a and b represent the scale and translational values, respectively.

$$X_{\psi}(a,b) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} x(t)\bar{\psi}(\frac{t-b}{a})dt \tag{4}$$

In this study, we applied CWT on a 1-D daily ratio profile $\mathbf{R}$ to transform it into a 2-D image $\mathbf{X}$, which is a time–frequency representation of the ratio profile. Fig. 7 shows an example of transforming a sample ratio profile into a CWT image. We adopted the Mexican hat wavelet as the mother wavelet for CWT. For a ratio profile with length N, the size of the CWT transformed image is $N \times N$.

3.4. Feature extraction

3.4.1. Deep feature extraction

Autoencoder is a popular neural network architecture that consists of an encoder and a decoder. The encoder transforms the input data into a low-dimensional representation. The decoder then tries to reconstruct the input data from the low-dimensional representation. The autoencoder is trained by minimizing the error between the input data and the reconstructed data.

In our study, a CAE is developed to extract features from the CWT image generated from the ratio profile as shown in Fig. 5. The parameters of the CAE are listed in Table 3, where the size of the input CWT image is $N \times N$, and k is the number of features obtained from the flatten layer. In the training stage, the encoder compresses an

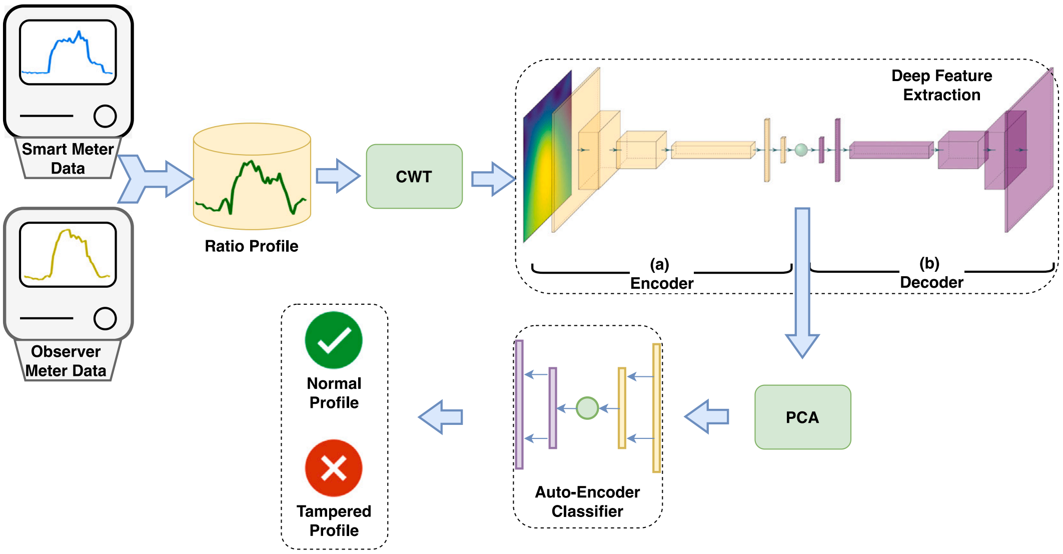


Fig. 5. The workflow of the proposed deep semi-supervised method for electricity theft detection.

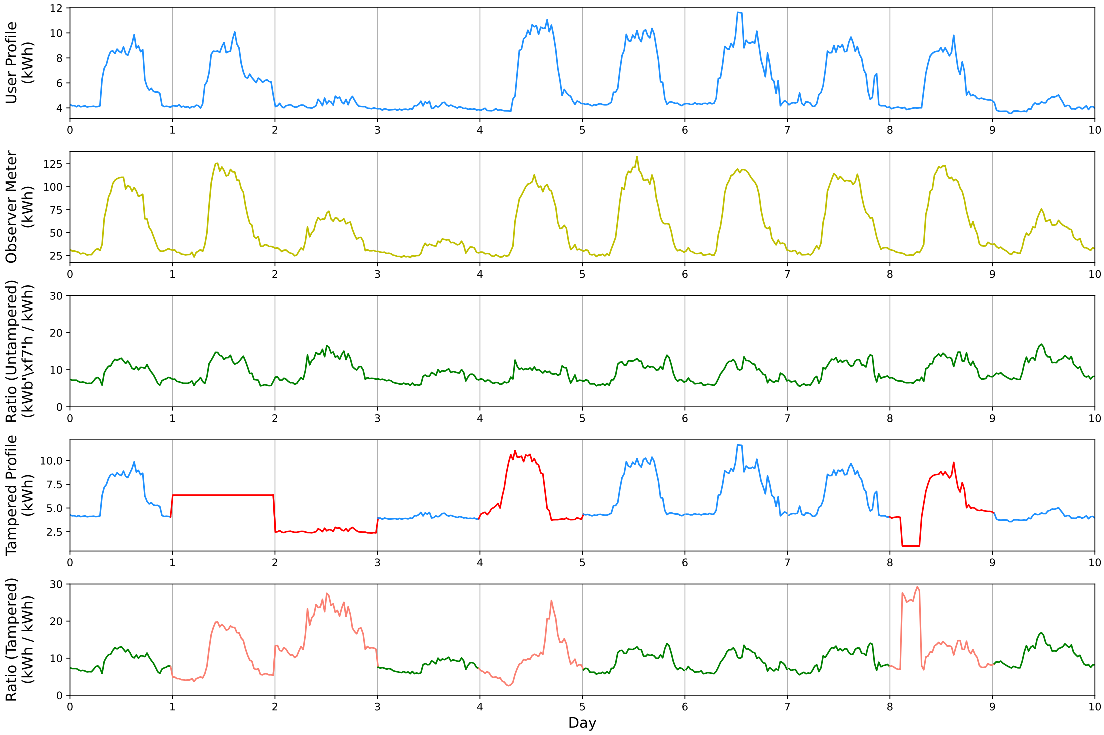


Fig. 6. An illustration of using ratio profile to help detect electricity theft. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

input image into a latent space and reconstructs it through the decoder. The mean-squared error (MSE) loss function is used to compute the error between the input image and the reconstructed one, and the Adam optimizer updates the CAE's weights to minimize this error. After training, the decoder is removed from the CAE, and the encoder serves as a deep feature extractor. Its output is the feature vector **f** extracted from the input CWT image, with a length of $k/4$, representing most of the critical information of the input data in a lower dimension.

3.4.2. *PCA for dimension reduction*

As the feature vector generated by the CAE still has a high dimension, we applied PCA to further reduce its dimensionality while

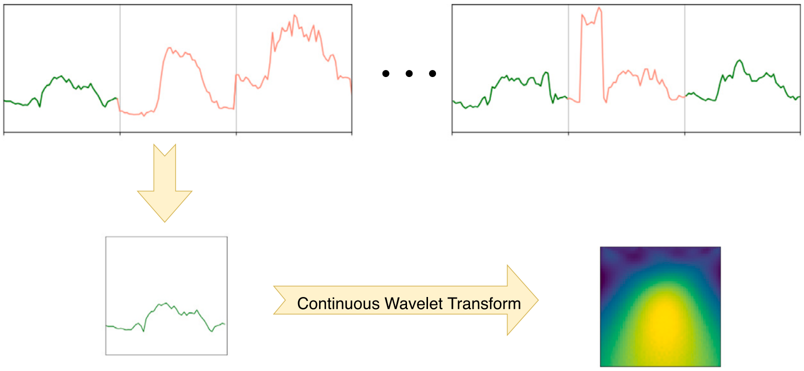


Fig. 7. A sample ratio profile **R** and the corresponding CWT transformed image **X**.

Table 3
Parameters for the CAE.

	Layer	Parameters
	Input	(1, N, N)
Encoder	Conv2d-1	8 filters, 3 × 3 kernel, 2 strides
	Conv2d-2	16 filters, 3 × 3 kernel, 2 strides
	Conv2d-3	32 filters, 3 × 3 kernel, 2 strides
	Flatten	–
	Linear-1	k input features, $k/2$ output features
	Linear-2	$k/2$ input features, $k/4$ output features
Decoder	Linear-3	$k/4$ input features, $k/2$ output features
	Linear-4	$k/2$ input features, k output features
	Unflatten	–
	Deconv-1	16 filters, 3 × 3 kernel, 2 strides
	Deconv-2	8 filters, 3 × 3 kernel, 2 strides
	Deconv-3	1 filter, 3 × 3 kernel, 2 strides

Table 4
Parameters for the semi-supervised autoencoder classifier.

	Layer	Input features	Output features
Encoder	L1	r	$r/2$
	L2	$r/2$	$r/4$
Decoder	L3	$r/4$	$r/2$
	L4	$r/2$	r

maintaining the most important discriminative information of input data. PCA is a technique commonly used for dimension reduction in deep learning. PCA projects high-dimensional data into a lower-dimensional space by identifying the most important directions of variation in the data, which are known as principal components. Typically, the first few principal components are retained since they are enough to capture most of the variation in the data. Given the feature matrix $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n] \in \mathbb{R}^{m \times l}$ extracted from the training data, where m denotes the number of training instances and $l = k/4$ is the number of features extracted by the CAE, its covariance matrix $\mathbf{S}$ can be factorized as:

$$\mathbf{S} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathbf{T}} \tag{5}$$

where $\mathbf{U} \in \mathbb{R}^{l \times l}$ serves as the utility matrix, and $\boldsymbol{\Lambda} \in \mathbb{R}^{l \times l}$ is a diagonal matrix filled with the eigenvalues. The columns of $\mathbf{U}$ are the eigenvectors of the matrix corresponding to the principal components of the data. For the purpose of dimension reduction in the proposed method, a feature vector $\mathbf{f}$ undergoes the following transformation:

$$\mathbf{f}_r = \mathbf{f}\mathbf{U}_r \tag{6}$$

where $\mathbf{f}_r$ is the reduced feature vector with r features, and $\mathbf{U}_r$ consists of the first r eigenvectors from $\mathbf{U}$. In our study, we set r to 20 for the data of business users and 40 for the data of residential users.

3.5. Semi-supervised autoencoder classifier

As shown in Fig. 5, the output of PCA will be finally fed into a semi-supervised autoencoder for classification which has a structure shown in Table 4. In Table 4, r is the number of features extracted by PCA. The semi-supervised autoencoder classifier is trained with the PCA features extracted from normal ratio profiles in the training dataset with a training process similar to the CAE. The MSE loss function and Adam optimizer are used for training. In the testing stage, the anomaly score of an input PCA feature vector $\mathbf{f}_r$, $AS_{\mathbf{f}_r}$, is evaluated as the MSE between $\mathbf{f}_r$ and the reconstructed output of the autoencoder, $\mathbf{f}'_r$, as shown in Eq. (7). A threshold determined in the training stage is used to classify the input as normal or abnormal based on the anomaly score. Because the classifier is trained only with normal profiles, the reconstructed vector $\mathbf{f}'_r$ of a tampered profile is expected to differ significantly from the input vector $\mathbf{f}_r$, resulting in a high anomaly score.

$$AS_{\mathbf{f}_r} = MSE(\mathbf{f}_r, \mathbf{f}'_r) \tag{7}$$

4. Performance evaluation and results

4.1. Smart meter datasets

To evaluate the performance of the proposed method, we use smart meter data of both business users and residential users since they have different electricity consumption characteristics.

The smart meter dataset of business users is from the Irish CER smart meter project [26], which has been widely used for evaluating electricity theft detection methods. The dataset contains more than 500 days of smart meter data collected from more than 5000 residential users and small and medium-sized business users during 2009 and 2010. For this dataset, the data were collected by smart meters every half an hour. In our study, we used the data of business users in 180 days from July 14, 2009, to Jan. 10, 2010. Of these, the data from 125 days were randomly selected as the training set, and the remaining data were used for testing. We randomly selected half of the testing days and tampered with the data of those days with a selected attack type. In an experiment, we randomly picked 30 business users to form an area that is covered by an observer meter. Each experiment was repeated ten times.

The smart meter data of residential users for evaluation were collected by Los Alamos Department of Public Utilities (LADPU) from 1757 households in Los Alamos, New Mexico, USA, from 2013 to 2019 [27]. The data was collected by smart meters every 15 min. In our study, we used the data in 180 days from Jan. 1, 2017, to June 30, 2017. The preparation of training and testing sets and experimental settings were the same as the Irish CER dataset.

4.2. Performance metrics

In our study, the area under the curve (AUC) was used as the primary metric for performance evaluation, which has been widely adopted for evaluating the performance of electricity theft detection methods [17,22,24,28]. AUC is calculated as the area under the Receiver Operating Characteristics (ROC) curve which plots the true positive rate (TPR) versus the false positive rate (FPR) of the detection model by using different detection thresholds. A higher AUC value indicates a better detection capability of the model.

When comparing the proposed method with supervised detection methods, we added F_1 score as an additional performance metric which is popular for evaluating supervised detection methods [19,29,30]. Since the F_1 score is obtained by determining a detection threshold first, we adopted the method of [31] by setting the detection threshold based on the outlier ratio.

4.3. Evaluation results of the proposed method

In this section, the performance evaluation results of the proposed method are presented. Due to the limited work in semi-supervised electricity theft detection, we compared the proposed method (PRM) with a set of baselines including one existing method and the variations of the existing method and proposed method, as shown in the following:

- $OCSVM_U$: This is an existing semi-supervised electricity theft detection method proposed in [13] which uses the OCSVM classifier with the user load profile as the input.
- $OCSVM_R$: This is a variation of $OCSVM_U$ which uses the ratio profile instead of the user load profile as the input of the OCSVM classifier.
- PRM_U: Instead of using the ratio profile as the input, this variation of the proposed method uses the user load profile as the input.
- $PRM_{w.o.PCA}$: This variation of the proposed method does not use PCA for further discriminative information extraction and dimension reduction. The deep features extracted by the CAE are directly used as the input of the semi-supervised auto-encoder classifier.

Tables 5 and 6 show the evaluation results in terms of AUC for business users and residential users, respectively. In each cell, the average AUC value of the results obtained from 300 users is reported, with the standard deviation shown in parenthesis. It can be observed from the results that methods using ratio profiles significantly outperform the corresponding methods directly using user profiles, which demonstrates the superiority of ratio profiles in capturing the malicious behaviors of electricity thieves. For both business users and residential users, PRM and $OCSVM_R$ achieve comparable performance for attack types 2, 3, 5, and 7, while PRM has significantly better performance than $OCSVM_R$ for the other seven attack types. Thus, under the MIX-type attack, the average AUC values of PRM are 5.0% and 3.7% higher than those of $OCSVM_R$ for business users and residential users, respectively. Compared with the variation $PRM_{w.o.PCA}$, PRM achieves significantly better performance for the majority of the attack types for business users, which results in a 2.7% higher average AUC value under the MIX attack. For residential users, PRM and $PRM_{w.o.PCA}$ have comparable performance for the majority of the attack types, while PRM significantly outperforms $PRM_{w.o.PCA}$ for attack types 4, 10, and 11. Thus, under the MIX attack, PRM still achieves better performance than $PRM_{w.o.PCA}$ for residential users. The results indicate that PCA not only reduces the computational cost but also improves the generalization performance of the proposed method. Overall, it has been proven by the results that the proposed method is a viable solution for electricity theft detection under a variety of FDI attacks.

Table 5
Performance comparison results for business users in terms of AUC (%).

$Type$	PRM	PRM_U	$OCSVM_U$	$OCSVM_R$	$PRM_{w.o.PCA}$
1	**88.9 (1.5)**	78.3 (3.0)	75.1 (2.7)	82.8 (3.2)	87.0 (1.8)
2	**99.8 (0.3)**	93.6 (2.0)	95.5 (1.0)	99.4 (0.7)	99.8 (0.5)
3	**99.2 (0.3)**	82.3 (2.5)	85.0 (2.5)	98.9 (1.3)	98.5 (0.8)
4	**83.8 (1.4)**	76.8 (2.9)	70.2 (2.2)	77.9 (2.0)	80.2 (1.5)
5	97.5 (0.9)	83.0 (3.1)	84.8 (4.1)	**98.2 (1.4)**	93.8 (1.4)
6	**69.9 (1.1)**	68.6 (3.1)	57.4 (1.8)	57.8 (2.4)	65.3 (1.7)
7	95.7 (0.9)	87.1 (2.2)	91.2 (1.9)	94.3 (1.6)	**95.9 (0.9)**
8	**83.8 (0.9)**	74.4 (1.6)	68.9 (1.4)	79.0 (1.7)	81.6 (1.3)
9	**78.6 (3.4)**	76.6 (3.6)	63.6 (4.6)	70.3 (2.3)	69.1 (3.1)
10	**82.9 (2.0)**	81.5 (2.9)	80.4 (2.6)	71.8 (2.3)	75.2 (1.9)
11	**87.9 (1.5)**	81.2 (2.2)	66.2 (4.2)	82.2 (1.7)	85.2 (1.7)
MIX	**87.6 (1.1)**	80.9 (1.7)	76.4 (2.6)	82.6 (1.4)	84.9 (1.6)

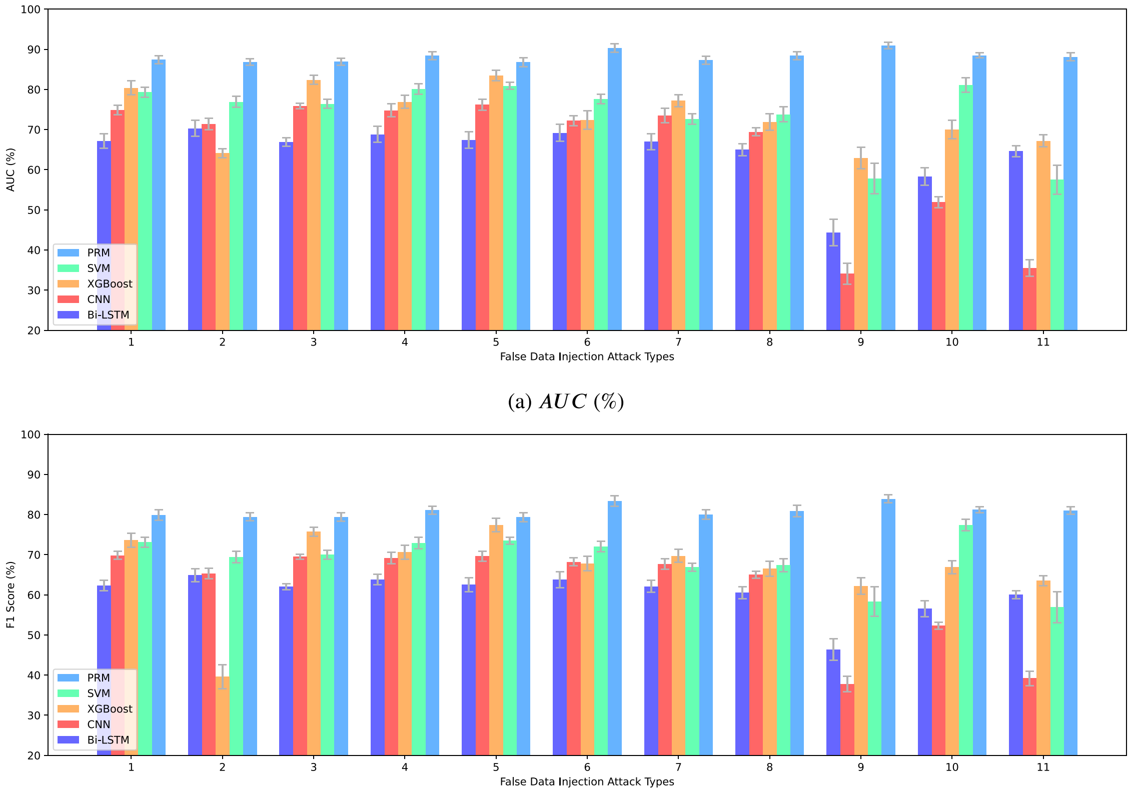
Table 6
Performance comparison results for residential users in terms of AUC (%).

$Type$	PRM	PRM_U	$OCSVM_U$	$OCSVM_R$	$PRM_{w.o.PCA}$
1	91.6 (0.4)	84.0 (1.6)	84.7 (1.2)	89.5 (0.5)	**91.8 (0.6)**
2	99.6 (0.2)	97.6 (0.3)	96.2 (0.3)	99.3 (0.3)	**99.7 (0.2)**
3	**100.0 (0.0)**	87.9 (0.9)	94.8 (0.5)	**100.0 (0.0)**	**100.0 (0.0)**
4	**84.0 (0.7)**	78.7 (1.1)	74.8 (0.8)	81.4 (0.7)	82.8 (0.7)
5	99.9 (0.1)	85.1 (0.9)	88.9 (0.7)	**100.0 (0.0)**	99.9 (0.1)
6	62.1 (1.3)	58.4 (1.3)	53.6 (1.6)	53.6 (2.0)	**63.5 (1.7)**
7	97.1 (0.4)	93.3 (1.0)	95.1 (0.7)	**97.3 (0.4)**	97.2 (0.5)
8	**87.6 (0.8)**	82.9 (0.7)	81.1 (0.8)	85.7 (0.8)	87.4 (0.7)
9	70.2 (0.7)	54.8 (1.1)	32.6 (0.1)	46.8 (0.2)	**70.8 (0.8)**
10	**86.8 (0.6)**	82.3 (0.6)	79.5 (0.8)	82.2 (0.8)	82.0 (1.1)
11	**81.1 (1.3)**	76.9 (1.1)	65.3 (1.4)	77.9 (1.2)	79.3 (1.0)
MIX	**87.7 (0.6)**	81.1 (0.8)	78.1 (0.7)	84.0 (0.6)	87.1 (0.7)

4.4. Performance comparison with supervised methods

To demonstrate PRM's capability of detecting unknown attacks, we performed a performance comparison with two supervised shallow machine learning methods, SVM [13] and XGBoost [16], and two recently proposed supervised deep learning methods, CNN [32] and bidirectional long short-term memory (Bi-LSTM) [33]. Unlike supervised methods which need the data of both normal users and fraudulent users to train the detection models, semi-supervised methods only need the data of normal users for training. Thus, we adopt the approach of [13] to prepare the training and testing sets for the proposed method and the two supervised methods. The 180-day smart meter data of a user was first split into two sets, the data of randomly selected 125 days for training and the data of the remaining 55 days for testing. For the proposed method, the data from the 125 training days were directly used as the training set. For the four supervised methods, the 125-day data were used as the normal samples of the training set. We then tampered with the 125-day data using a selected FDI attack type where the tampered data were used as the malicious samples of the training set. The testing sets are the same for the proposed method and the four supervised methods. The data from the 55 testing days were used as normal samples of the testing set. The malicious samples of the testing set were obtained by tampering with the data of the testing days using a mixture of attack types except the one used for generating malicious samples of the training set.

Figs. 8(a) and 8(b) show the results for business users in terms of AUC and F1 Score, respectively. It can be seen from the results that PRM achieves significantly better performance than the four supervised methods in terms of both the AUC and F1 score in all cases. The results show that PRM has a stable performance as it only uses normal data for training. On the other hand, the detection performance of the four supervised methods is heavily influenced by the malicious samples used in the training set. For instance, when training with only malicious samples generated by attack types 9 or 11, SVM and CNN exhibit significantly poorer performance compared to when trained with malicious samples from other attack types. Similarly, XGBoost's performance is



(a) AUC (%)

(b) F_1 score (%)

Fig. 8. Performance comparison results for business users.

affected significantly when trained exclusively with malicious samples generated by attack type 2, and Bi-LSTM's performance suffers when trained solely with malicious samples from attack type 9. Notably, the deep learning methods generally yield inferior results compared to the shallow machine learning methods due to the limited training data available. The results for residential users are shown in Fig. 9, which have similar trends to the results for business users. In summary, the results prove that PRM is more capable of dealing with unknown attacks than the four supervised methods.

5. Conclusion

Electricity theft is a serious issue affecting the sustainability and security of smart grids that can overload power lines, damage home appliances, cause raised rates for legitimate users, and undermine overall grid stability. Traditional machine learning-based electricity theft detection methods generally use supervised or unsupervised learning algorithms, both with their limitations. In this paper, we propose a deep semi-supervised method for electricity theft detection with the aim of detecting unknown attacks in a short time frame. The method utilizes the ratio profile obtained from the observer meter and smart data as input and performs deep feature extraction using CWT, CAE, and PCA. A semi-supervised autoencoder classifier then classifies the extracted features as normal or fraudulent. Our evaluation results using smart meter data from both business and residential users under 11 different FDI attacks show that the proposed method achieves significantly better performance in terms of AUC than a set of baselines including existing methods and variations of the proposed method. The results also demonstrate that the proposed method greatly outperforms supervised learning-based methods in terms of both AUC and F1 score when detecting unknown attacks. It should be noted that supervised detection methods have superior performance compared to semi-supervised methods when there are enough malicious samples for training. Therefore, the purpose of the proposed method is to complement existing supervised methods for detecting unknown attacks. It can be observed from the results in Tables 5 and 6 that the proposed method has significantly lower detection performance for attack types 6 and 9 than other attack types. Thus, our future work will focus on improving the detection performance of these two attack types. We consider combining manually crafted features targeting these two attack types with deep features to achieve the goal.

CRediT authorship contribution statement

Ruobin Qi: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization. **Qingqing Li:** Software, Data curation. **Zhirui Luo:** Software, Data curation. **Jun Zheng:** Conceptualization, Methodology, Writing – original draft, Supervision, Project administration, Funding acquisition. **Sihua Shao:** Writing – review & editing, Funding acquisition.

Declaration of competing interest

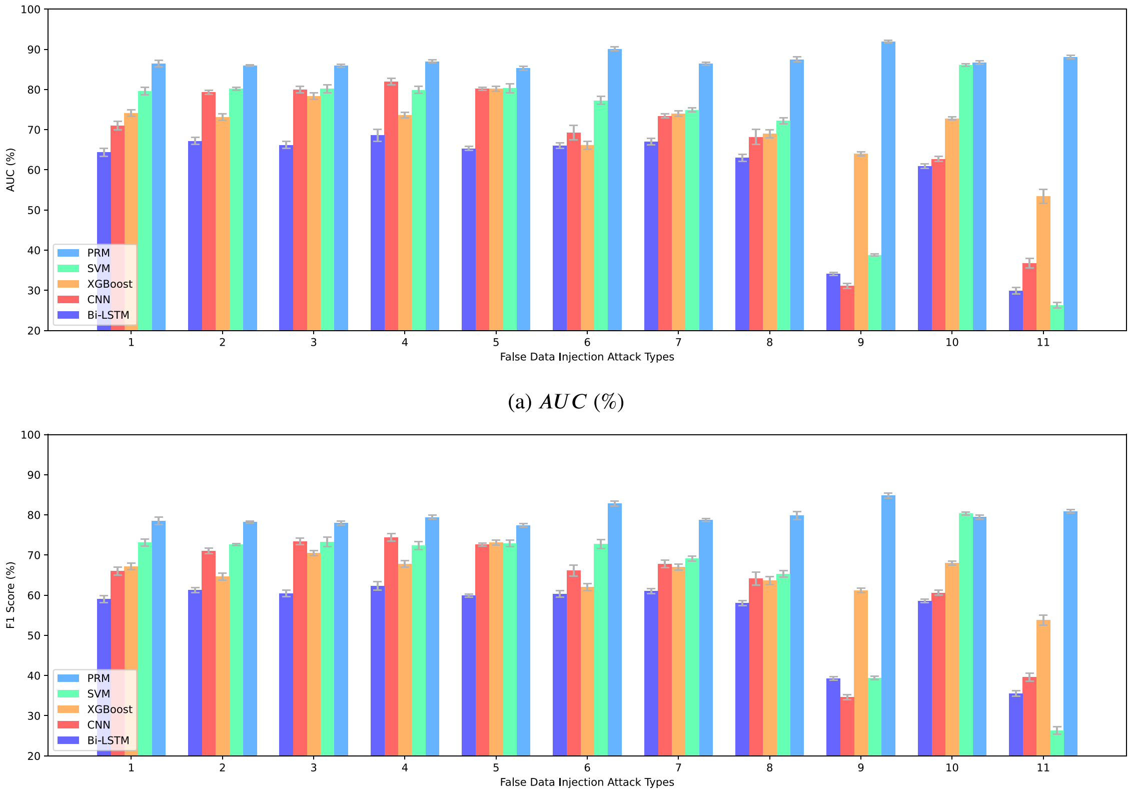
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been use is publicly available.

Acknowledgments

This material is based upon work supported by the National Science Foundation EPSCoR Cooperative Agreement OIA-1757207.



(a) AUC (%)

(b) F_1 score (%)

Fig. 9. Performance comparison results for residential users.

References

[1] R. Qi, J. Zheng, Z. Luo, Q. Li, A novel unsupervised data-driven method for electricity theft detection in AMI using observer meters, IEEE Trans. Instrum. Meas. 71 (2022) 1–10.

[2] R.R. Mohassel, A. Fung, F. Mohammadi, K. Raahemifar, Application of advanced metering infrastructure in smart grids, in: 22nd Mediterranean Conference on Control and Automation, IEEE, 2014, pp. 822–828.

[3] NETL Modern Grid Strategy, Advanced metering infrastructure, 2008, URL: https://www.smartgrid.gov/files/documents/NIST_SG_Interop_Report_Postcommentperiod_version_200808.pdf.

[4] F. Mohammadi, Emerging challenges in smart grid cybersecurity enhancement: A review, Energies 14 (5) (2021) 1380.

[5] S. Shapsough, F. Qatan, R. Aburukba, F. Aloul, A. Al Ali, Smart grid cyber security: Challenges and solutions, in: 2015 International Conference on Smart Grid and Clean Energy Technologies (ICSGCE), IEEE, 2015, pp. 170–175.

[6] K. Katyora, Electricity theft detection & prevention using artificial intelligence for african utilities, 2021, URL: https://energycentral.com/c/pip/electricity-theft-detection-prevention-using-artificial-intelligence-african.

[7] EPRI, Advanced metering infrastructure technology - limiting non-technical distribution losses in the future, 2008, URL: https://www.epri.com/research/products/1016049.

[8] X. Xia, Y. Xiao, W. Liang, J. Cui, Detection methods in smart meters for electricity thefts: A survey, Proc. IEEE 110 (2) (2022) 273–319.

[9] Z. Xiao, Y. Xiao, D.H.-C. Du, Non-repudiation in neighborhood area networks for smart grid, IEEE Commun. Mag. 51 (1) (2013) 18–26.

[10] Z. Xiao, Y. Xiao, D.H.-C. Du, Building accountable smart grids in neighborhood area networks, in: 2011 IEEE Global Telecommunications Conference-GLOBECOM 2011, IEEE, 2011, pp. 1–5.

[11] Z. Xiao, Y. Xiao, D.H.-C. Du, Exploring malicious meter inspection in neighborhood area smart grids, IEEE Trans. Smart Grid 4 (1) (2012) 214–226.

[12] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, S. Mishra, Decision tree and SVM-based data analytics for theft detection in smart grid, IEEE Trans. Ind. Inform. 12 (3) (2016) 1005–1016.

[13] P. Jokar, N. Arianpoo, V.C. Leung, Electricity theft detection in AMI using customers' consumption patterns, IEEE Trans. Smart Grid 7 (1) (2015) 216–226.

[14] F. Shehzad, N. Javaid, S. Aslam, M.U. Javaid, Electricity theft detection using big data and genetic algorithm in electric power systems, Electr. Power Syst. Res. 209 (2022) 107975.

[15] R. Punmiya, S. Choe, Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing, IEEE Trans. Smart Grid 10 (2) (2019) 2326–2329.

[16] Z. Yan, H. Wen, Electricity theft detection base on extreme gradient boosting in AMI, IEEE Trans. Instrum. Meas. 70 (2021) 1–9.

[17] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, Y. Zhou, Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids, IEEE Trans. Ind. Inform. 14 (4) (2017) 1606–1615.

[18] R. Xia, Y. Gao, Y. Zhu, D. Gu, J. Wang, An attention-based wide and deep CNN with dilated convolutions for detecting electricity theft considering imbalanced data, Electr. Power Syst. Res. 214 (2023) 108886.

[19] M.N. Hasan, R.N. Toma, A.-A. Nahid, M.M. Islam, J.-M. Kim, Electricity theft detection in smart grid systems: A CNN-LSTM based approach, Energies 12 (17) (2019) 3310.

[20] R. Qi, C. Rasband, J. Zheng, R. Longoria, Detecting cyber attacks in smart grids using semi-supervised anomaly detection and deep representation learning, Information 12 (8) (2021) 328.

[21] P.P. Biswas, H. Cai, B. Zhou, B. Chen, D. Mashima, V.W. Zheng, Electricity theft pinpointing through correlation analysis of master and individual meter readings, IEEE Trans. Smart Grid 11 (4) (2019) 3031–3042.

[22] Y. Peng, Y. Yang, Y. Xu, Y. Xue, R. Song, J. Kang, H. Zhao, Electricity theft detection in AMI based on clustering and local outlier factor, IEEE Access 9 (2021) 107250–107259.

[23] J. Tao, G. Michailidis, A statistical framework for detecting electricity theft activities in smart grid distribution networks, IEEE J. Sel. Areas Commun. 38 (1) (2019) 205–216.

[24] K. Zheng, Q. Chen, Y. Wang, C. Kang, Q. Xia, A novel combined data-driven approach for electricity theft detection, IEEE Trans. Ind. Inform. 15 (3) (2018) 1809–1819.

[25] M.G. Chuwa, F. Wang, A review of non-technical loss attack models and detection methods in the smart grid, Electr. Power Syst. Res. 199 (2021) 107415.

[26] Commission for Energy Regulation, Cer smart metering project— electricity customer behaviour trial, 2009–2010, 2012, Available: https://www.ucd.ie/issda/data/commissionforenergyregulationcer/.

[27] V. Souza, T. Estrada, A. Bashir, A. Mueen, LADPU smart meter data, 2020, Dryad, Dataset, Available: http://dx.doi.org/10.5061/dryad.m0cfxpp2c.

[28] D. Gu, Y. Gao, K. Chen, J. Shi, Y. Li, Y. Cao, Electricity theft detection in AMI with low false positive rate based on deep learning and evolutionary algorithm, IEEE Trans. Power Syst. 37 (6) (2022) 4568–4578.

[29] I.U. Khan, N. Javaid, C.J. Taylor, X. Ma, Data driven analysis for electricity theft attack-resilient power grid, IEEE Trans. Power Syst. (2022).

[30] R. Razavi, A. Gharipour, M. Fleury, I.J. Akpan, A practical feature-engineering framework for electricity theft detection in smart grids, Appl. Energy 238 (2019) 481–494.

[31] S. Wang, J. Liu, G. Yu, X. Liu, S. Zhou, E. Zhu, Y. Yang, J. Yin, Multi-view deep one-class classification: A systematic exploration, 2021, arXiv preprint arXiv:2104.13000.

[32] E.U. Haq, C. Pei, R. Zhang, H. Jianjun, F. Ahmad, Electricity-theft detection for smart grid security using smart meter data: A deep-CNN based approach, Energy Rep. 9 (2023) 634–643.

[33] R. Kaur, G. Saini, Electricity theft detection system for smart metering application using bi-LSTM, in: Proceedings of Second International Conference on Computational Electronics for Wireless Communications: ICCWC 2022, Springer, 2023, pp. 581–592.