PE-DOCC: A Novel Periodicity-enhanced Deep One-class Classification Framework for Electricity Theft Detection

Zhijie Wu * and Yufeng Wang *

Posted Date: 26 January 2025

doi: 10.20944/preprints202501.1914.v1

Keywords: electricity theft detection; one-class classification; transformer encoder; unsupervised representation learning

 doi:10.20944/preprints202501.1914.v1

Article

PE-DOCC: A Novel Periodicity-enhanced Deep One-class Classification Framework for Electricity Theft Detection

Zhijie Wu * and Yufeng Wang *

Nanjing University of Posts and Telecommunications
* Correspondence: 1022010113@njupt.edu.cn (Z.W.); wfwang@njupt.edu.cn (Y.W.)

Abstract: Electricity theft, emerging as one of severe cyberattacks in smart grids, causes significant economic losses. Due to the powerful expressive ability of deep neural network(DNN), supervised and unsupervised DNN-based Electricity theft detection (ETD) schemes have witnessed great deployment. However, the existing works have the following weakpoints. Supervised DNN-based schemes require abundant labelled anomalous samples for training, and even worse, can't detect unseen theft patterns. To avoid the extensively labor-consuming behaviour of labelling anomalous samples, unsupervised DNNs-based schemes aim to learn the normality of time-series, and infer anomaly score for each data instance, but they fail to capture periodic features effectively. To address these challenges, this paper proposes a novel periodicity-enhanced deep one-class classification framework (PE-DOCC) based on periodicty-enhanced transformer encoder, named Periodicformer encoder. Specifically, within the encoder, a novel criss-cross periodic attention is proposed to capture both horizontal and vertical periodic features. The Periodicformer encoder is pre-trained by reconstructing partially masked input sequences, and the learned latent representations are then fed into a one-class classification for anomaly detection. Extensive experiments on real-world datasets demonstrate that our proposed PE-DOCC framework outperforms state-of-the-art unsupervised ETD methods.

Keywords: electricity theft detection; one-class classification; transformer encoder; unsupervised representation learning

1. Introduction

Electricity theft, a longstanding challenge for utility companies, has evolved with the advent of smart grids. In addition to traditional physical attacks (such as meter bypassing), modern cyberattacks (like spoofing and man-in-the-middle attacks) now target electronic meters, often leveraging low-cost, easily accessible tools. This shift towards cyberattacks is particularly prominent in advanced power systems. The economic impact is severe, with global losses reaching $96 billion in 2017 [1], especially in developing countries such as India. Electricity theft not only results in financial losses but also leads to inaccurate power demand forecasts, contributing to poor power quality and infrastructure damage. Furthermore, it hampers investments in advanced technologies and safety measures. Despite strict regulations in various countries, the strong economic incentives continue to fuel electricity theft.

Recently, based on the massive amount of data provided by the advanced metering infrastructure (AMI), deep neural network (DNN)-based electricity theft detection (ETD) schemes have been widely developed [2,3], since DNNs can automatically extract features and have powerful functional approximation ability. Generally, from the perspective of learning paradigm, the existing schemes can be classified into two categories: supervised learning-based and unsupervised learning-based. (1) Supervised DNNs are trained using both normal and abnormal samples, typically requiring a large number of labeled samples while ensuring that the quantities of normal and abnormal samples are balanced. However, in smart grids, it is impractical to obtain a large number of electricity theft samples, leading to a data imbalance between abnormal and normal samples. Under such circumstances, supervised

 doi:10.20944/preprints202501.1914.v1

models struggle to learn the general features of abnormal samples, resulting in poor detection performance when faced with new electricity theft methods. Many supervised learning-based studies have leveraged the periodicity in electricity consumption data to distinguish between electricity theft and normal usage. For example, several studies have transformed one-dimensional electricity consumption data into two-dimensional matrices and employed various DNNs to capture the underlying periodicity. ETD-ConvLSTM (Electricity Theft Detector based upon Convolutional Long Short Term Memory neural networks) [4] uses a ConvLSTM network to capture the periodic patterns in users' consumption. HORLN (Hybrid-order Representation Learning Network) [5] uses the autocorrelation function to determine whether there is periodicity in different electricity consumption data. Additionally, some studies (e.g., Reference [6]) convert the electricity data into graphs and leverage Graph Convolutional Networks (GCNs) to capture the periodicity within the graph structure. However, the aforementioned DNNs do not inherently capture temporal dependencies and fail to fully exploit the periodic features. Although HORLN uses autocorrelation to highlight the capture of periodic characteristics, it only incorporates it as a second-order feature, merging it with first-order feature information. (2) Unsupervised DNNs focus on capturing the characteristics of normal electricity consumption data, offering an effective solution to the data imbalance problem commonly faced by supervised DNNs. Reference [7] proposes a deep learning-based electricity theft cyberattack detection method for AMIs, combining stacked autoencoders with an LSTM (Long Short Term Memory neural networks)-based sequence-to-sequence (seq2seq) structure to capture sophisticated patterns and temporal correlations in electricity consumption data. Reference [8] proposes a novel unsupervised two-stage approach for electricity theft detection, combining a Gaussian mixture model to cluster consumption patterns and an attention-based bidirectional LSTM encoder-decoder to enhance robustness against non-malicious changes. A novel anomaly score is proposed, which comprehensively considers the similarity of consumption patterns and reconstruction errors. Reference [9] proposes a stacked sparse denoising autoencoder for electricity theft detection, leveraging the reconstruction error of anomalous behaviors to identify theft users. However, existing unsupervised DNNs also have not fully exploited the periodic features in electricity consumption data.

Based on the above considerations, we propose a periodicity-enhanced deep one-class classification framework (PE-DOCC) based on periodicty-enhanced transformer encoder, named Periodicformer encoder. The main contributions of this paper can be summarized as follows.

1. We propose a novel PE-DOCC framework for detecting electricity theft, which incorporates unsupervised representation learning (i.e., pre-train Periodicformer encoder by recovering partially masked input sequences) and one-class classification (anomaly detection based on local outlier factor);
2. To extract richer periodic features, within encoder a novel Criss-Cross Periodic Attention (CCPA) is proposed, which comprehensively considers both the horizontal and vertical periodic features;
3. Extensive experiments of ETD using the Irish data set are conducted to compare the performance with various state-of-the-art methods. Appropriate metrics are selected for evaluation and the effectiveness of the proposed ETD scheme is verified.

The rest of the paper is organized as follows. In Section II, typical one-class (OC) classification (including traditional machine learning-based and recent deep learning-based) ETD schemes are summarized. Section III presents the proposed ETD framework PE-DOCC and describes its main module in detail. Section IV provides the comprehensive performance comparison between the proposed framework and state-of-the-art ETD methods. Finally, we briefly conclude this paper.

2. Related Work

Unsupervised anomaly detection for electricity theft (and similar issues) is a highly challenging task, and over the past few years, various methods have been proposed to address it.

One traditional type is classifcation-based methods. Reference [10] proposes a feature selection algorithm for mixed load data with statistical features based on random forest (FS-MLDRF) to reduce

 doi:10.20944/preprints202501.1914.v1

the dimension of user data, and then use the hybrid random forest and weighted support vector data description (HFR-WSVDD) to address the data imbalance issue. Reference [11] proposes a privacy-preserving anomaly detection approach for LoRa-based IoT networks using Federated Learning. SVOC [12], a data-driven approach to detect gas-theft suspects among boiler room users, combines scenario-based data quality detection, deformation-based normality detection, and an OCSVM (One-class Support Vector Machine)-based anomaly detection algorithm. Another common method is ensemble-based methods. TiWS-iForest [13] incorporates weak supervision to reduce complexity and enhance anomaly detection performance. The approach addresses the limitations of standard iForest (Isolation Forest) in terms of memory, latency, and efficiency, particularly in low-resource environments like TinyML implementations on ultra-constrained microprocessors. Additionally, density-based methods e.g. ELOF [14] (Ensemble Local Outlier Factor) integrates multiple LOF models with optimized hyperparameters and distance metrics.

Besides traditional methods, deep learning-based unsupervised anomaly detection algorithms have gained a lot attention recently. Tab-iForest [15] combines the TabNet deep learning model with the Isolation Forest algorithm, where TabNet selects relevant features through pre-training and inputs them into Isolation Forest for anomaly detection. Reference [16] combines autoencoding and one-class classification, designed to benefit from strong abstraction by neural networks (using an autoencoder) and the removal of the complex threshold selection (using an OC classifier). A new model selection method that discovers well-optimized models from a variety of combinations is also proposed. Reference [17] proposes a two-step electricity theft detection strategy to enhance economic returns for power companies. It uses a convolutional autoencoder (CAE) to identify electricity theft and the Tr-XGBoost regression algorithm which combines XGBoost (Extreme Gradient Boosting) with TrAdaBoost (Transfer Adaptive Boosting) strategy, to predict potentially stolen electricity (PSE).

To sum up, deep learning-based unsupervised anomaly detection methods typically begin by using deep neural networks (e.g., autoencoders) to extract temporal dependencies and periodic features present in electricity data. The features output by the DNN are often low-dimensional, addressing the curse of dimensionality that traditional machine learning struggles with. These extracted features are then analyzed through one-class classification, effectively tackling the data imbalance issue between normal and anomalous data.

3. Proposed Framework: PE-DOCC

Figure 1 presents the proposed ETD framework, which explicitly consists of two main phases: model training and model testing. The training phase consists of two steps: Step 1: Pretrain the Periodicformer encoder by recovering partially masked input sequences (i.e., unsupervised representation learning). Step 2: Feed the latent representations output by the encoder into local outlier factor (LOF) to learn the anomaly detection boundary. In the testing phase, the input data is fed through the trained Periodicformer encoder and classifier to determine whether testing latent representation is within the trained boundary (1) or not (-1).

 doi:10.20944/preprints202501.1914.v1

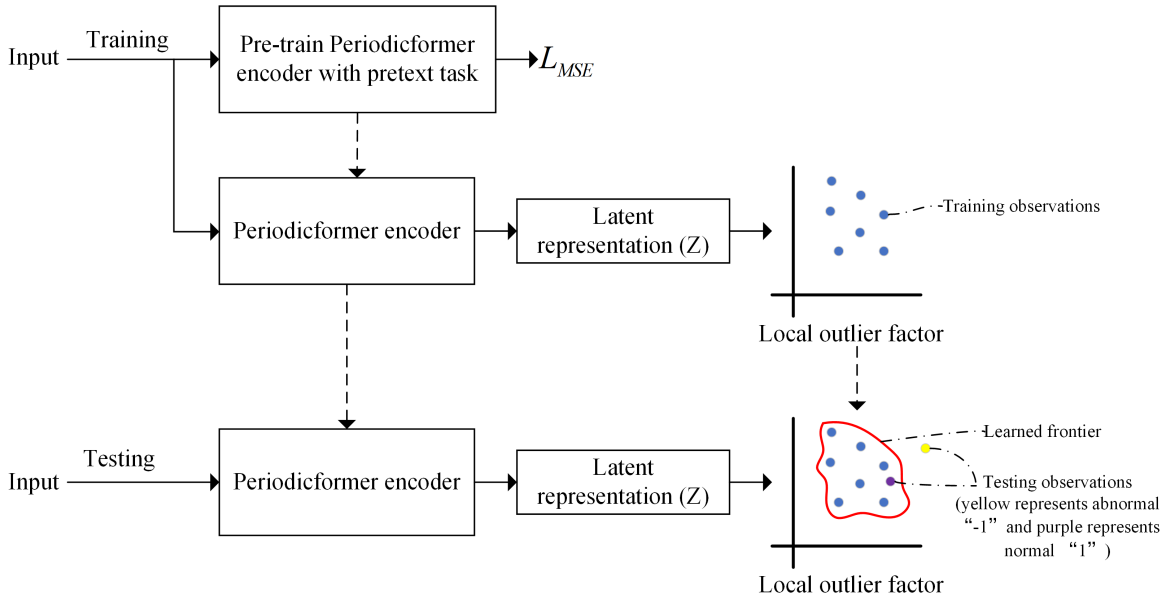

Figure 1. Proposed framework integrating unsupervised representation learning with one-class classification.

3.1. Periodicformer Encoder for Unsupervised Representation Learning

Figure 2 illustrates the overall architecture of the proposed Periodicformer encoder, which is composed of three modules: patching module, the proposed Criss-cross periodic attention module and the multilayer perceptron (MLP) module (consisting of two linear layers with a gelu activation).

In the layer i, we denote the patch size as p_i, the length of input feature as L_{i-1}, and the embedding dimension as d.

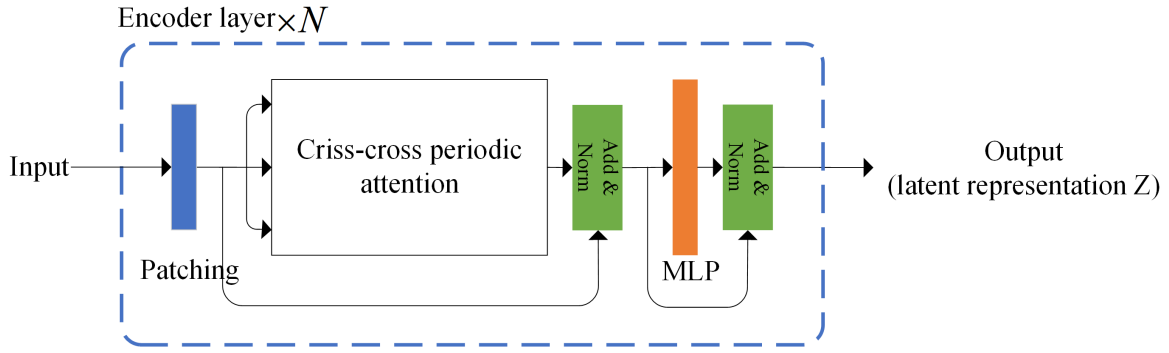

Figure 2. The overall architecture of the proposed Periodicformer encoder.

3.1.1. Patching Operation

The input feature $F_{i-1} \in \mathbb{R}^{L_{i-1} \times d}$ for layer i is first divided into multiple patches $P = [patch^1, ..., patch^{\frac{L_{i-1}}{p_i}}] \in \mathbb{R}^{\frac{L_{i-1}}{p_i} \times (p_i \times d)}$ ($\frac{L_{i-1}}{p_i}$ is the number of the patches, each patch $\in \mathbb{R}^{p_i \times d}$), then flatten these patches and embeds them into a new feature map $F'_{i-1} = [flatten(patch^1), ..., flatten(patch^{\frac{Li-1}{p_i}})]W_P + b_P$.

It is worth noting that we do not use positional encoding here, as our experimental results show that adding positional encoding actually reduces detection performance, which is consistent with the findings in Reference [18].

3.1.2. Criss-Cross Periodic Attention

As shown in Figure 3, we propose the criss-cross periodic attention mechanism to mine richer periodic scale characteristics. Row autocorrelation and column autocorrelation explore periodic dependencies by calculating autocorrelation.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2025 doi:10.20944/preprints202501.1914.v1

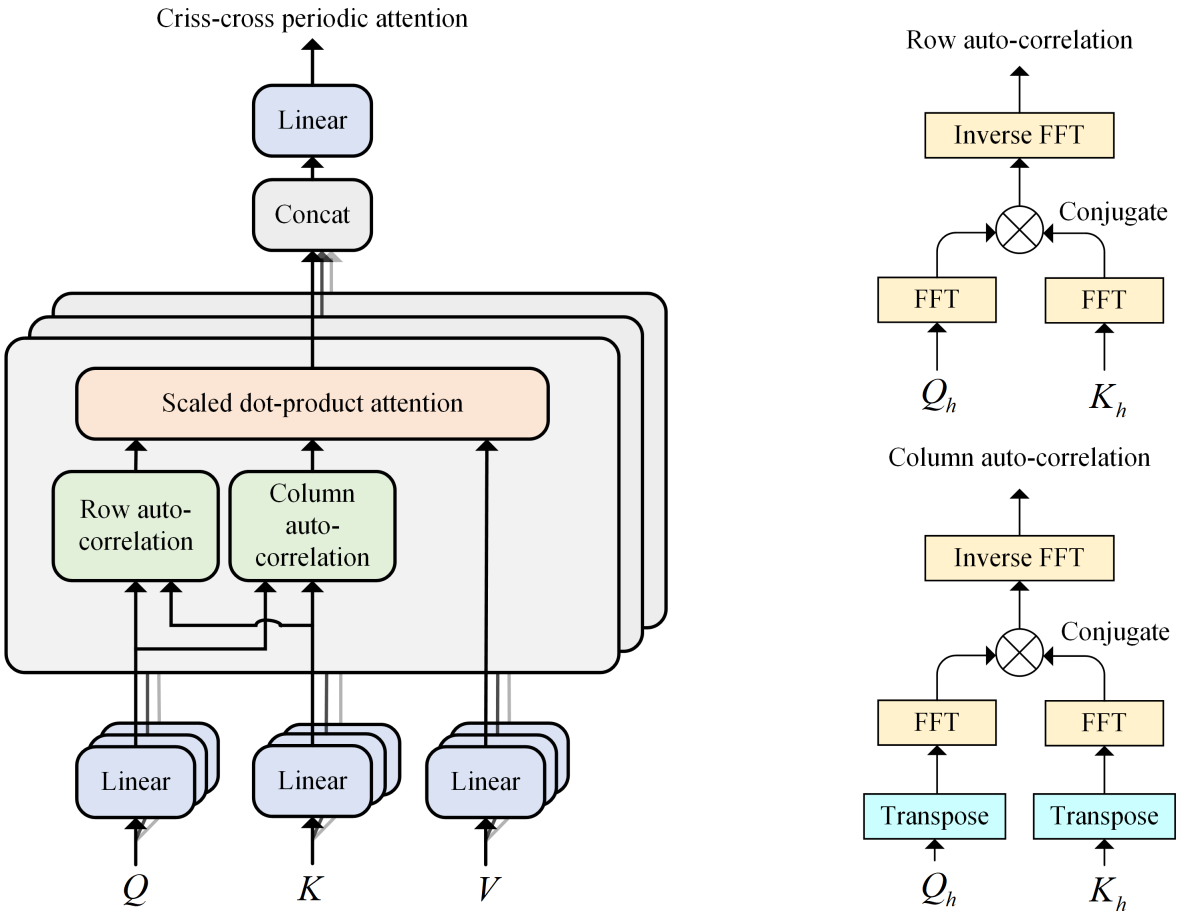

Figure 3. Criss-cross periodic attention (left) and the corresponding row and column auto-correlation for a single head h (right).

- Period-based dependencies: Inspired by Auto-Correlation mechanism [19], for an input sequence X_t with length L, we can obtain the autocorrelation $R_{XX}(\tau)$ by the following Equation 1.

$$R_{XX}(\tau) = \lim_{L\to\infty} \frac{1}{L} \sum_{t=1}^{L} X_t X_{t-\tau}, \tag{1}$$

$R_{XX}(\tau)$ reflects the time-delay similarity between $\{X_t\}$ and its τ lag series $\{X_{t-\tau}\}$. Taking the Row auto-correlation in Figure 3 as an example, for single head h and input sequence $F'_{i-1} \in \mathbb{R}^{L_i \times d}$ where $L_i = \frac{L_{i-1}}{p_i}$, after the linear projector, we get query Q_h, key K_h and value V_h. At this point, we consider Q_h as sequences along the row direction, i.e.,$Q_h = [row_1, ..., row_{d_k}]$ where d_k is the embedding dimension of K_h, with K_h treated in the same manner. Then we can obtain the Row auto-correlation $\hat{R}_{Row}(Q_h, K_h)$ by the following Equation 2.

$$\hat{R}_{Row}(Q_h, K_h) = [R_{Q_h,K_h}(1), ..., R_{Q_h,K_h}(d_k)], \tag{2}$$

where R_{Q_h,K_h} is the autocorrelation between Q_h and K_h. The calculation of column auto-correlation is similar to that of Row auto-correlation, with the only difference being that Q_h and K_h are treated as sequences along the column direction. The formula for calculating Column auto-correlation is provided in Equation 3.

$$\hat{R}_{Column}(Q_h, K_h) = [R_{Q_h^T,K_h^T}(1), ..., R_{Q_h^T,K_h^T}(L_i)], \tag{3}$$

where $R_{Q_h^T,K_h^T}$ is the autocorrelation between the transpose of Q_h and the transpose of K_h;

 doi:10.20944/preprints202501.1914.v1

- Criss-cross periodic attention: To comprehensively capture the different periodic features extracted by row and column auto-correlation, we use the scaled dot product attention, with the product of row and column auto-correlation serving as the new attention weights. For single head h, the Criss-cross periodic attention (CCPA) mechanism is as the following Equation 4.

$$\text{CCPA}(Q_h, K_h, V_h) = softmax(\frac{\hat{R}_{Row}(Q_h, K_h)\hat{R}^T_{Column}(Q_h, K_h)}{\sqrt{d_k}})V_h, \tag{4}$$

For the multi-head version, with H heads, the process is as Equation 5;

$$\begin{aligned} \text{Multi-head}(Q, K, V) &= concat(head_1, ..., head_H)W_o, \\ \text{where } head_i &= \text{Criss-cross periodic attention}(Q_i, K_i, V_i), \end{aligned} \tag{5}$$

- Efficient computation: For the autocorrelation computation (Equation 1), given input sequence X_t, $R_{XX}(\tau)$ can be calculated by the Fast Fourier Transforms (FFT) as Equation 6:

$$\begin{aligned} S_{XX}(f) &= \mathcal{F}(X_t)\mathcal{F}^*(X_t) = \int_{-\infty}^{\infty} X_t e^{-i2\pi tf} dt \overline{\int_{-\infty}^{\infty} X_t e^{-i2\pi tf} dt}, \\ R_{XX}(\tau) &= \mathcal{F}^{-1}(S_{XX}(f)) = \int_{-\infty}^{\infty} S_{XX}(f) e^{i2\pi f\tau} df, \end{aligned} \tag{6}$$

where $\tau \in \{1, ..., L\}$, $\mathcal{F}$ denotes FFT and $\mathcal{F}^{-1}$ is its inverse transformation. $*$ denotes the conjugate operation and $S_{XX}(f)$ is in the frequency domain.

3.1.3. Unsupervised Representation Learning

Electricity consumption data is treated as a univariate time series of length T, expressed as $X = \{x_1, ..., x_T\}$. To determine whether the data point x_t at time t is an anomaly, we analyze a contextual window of length L, defined as $w_t = \{x_{t-L+1}, ..., x_t\}$, instead of focusing solely on a single point. For simplicity, windows w_t are omitted when $t < L$ [20].

Unsupervised representation learning involves two steps: creating a Boolean mask M and calculating the mean squared error (MSE) loss. The mask M is generated with alternating segments of 0s and 1s, where the lengths of the 0s and 1s follow geometric distributions with means l_m and l_u, respectively. The proportion r determines the ratio of 0s in the mask, and the length l_u is computed as $l_u = \frac{1-r}{r} l_m$. The masked detection window $\tilde{w}_t$ is then passed through the encoder, producing feature representation Z. The output Z is flattened and passed through a linear layer to forecast the masked values, yielding $\tilde{w}_t$. The MSE loss is computed by comparing the predicted values $\tilde{w}_t(i)$ to the actual masked values $w_t(i)$ at the indices in M_0, which are the positions where $M(i) = 0$. The overall process of unsupervised representation learning is described in detail in Algorithm 1.

Note that, the objective of the unsupervised representation learning is to learn latent representations Z for subsequent one-class classification-based anomaly detection. In the subsequent anomaly detection process, the linear layer originally used for outputting the masked time-series is discarded.

 doi:10.20944/preprints202501.1914.v1

Algorithm 1: Unsupervised representation learning

Input:
L: Length of mask and sequence to be masked,
l_m: Average length of masking subsequences (streaks of 0s),
r: Proportion of L to be masked,
W: Detection windows,
ϕ: Mapping function of Periodicformer encoder,
M: A mask of length L with all values initialized to one,
E: Maximum number of epochs.
Output: Z: Latent representation of W.

Initialize:
$p_m \leftarrow \frac{1}{l_m}$; `// Probability of each masking sequence stopping. Parameter of geometric distribution`
$p_u \leftarrow p_m \cdot \frac{r}{1-r}$; `// Probability of each unmasked sequence stopping. Parameter of geometric distribution.`
$p \leftarrow [p_m, p_u]$;
$state \leftarrow \mathbb{1}(\text{rand}() > r)$; `// State 0 means masking, 1 means not masking`

for *epoch* $\leftarrow 1$ ***to*** E **do**
  foreach ***batch of windows*** w_t ***in*** W **do**
    // Randomly create a boolean mask M
    for $i \leftarrow 1$ ***to*** L **do**
      $M[i] \leftarrow state$
      if $rand() < p[state]$ **then**
        $state \leftarrow 1 - state$
      end
    end
    // Calculate MSE loss
    $\hat{w}_t \leftarrow M \odot w_t$
    $Z \leftarrow \phi(\hat{w}_t)$
    $\tilde{w}_t \leftarrow \text{flatten}(Z)W + b$
    $\mathcal{L}_{MSE} \leftarrow \sum_{i \in M_0} [\tilde{w}_t(i) - w_t(i)]^2$; `// Compute MSE loss over masked elements`
    Make a gradient step to minimize $\mathcal{L}_{MSE}$.
  end
end

return $Z \leftarrow \phi(W)$

3.2. One-Class Classifier for Anomaly Detection

Once the unsupervised training is completed, the training data is rejected into the trained Periodicformer encoder and the latent representation ($Z = \{z_1, ...z_{|X_{tr}|}\}$) is fed to the local outlier factor (LOF) as input where $|X_{tr}|$ is the number of training data.

LOF is a density-based outlier detection algorithm. The determination of the outlier is judged based on the density between each data point and its neighbor points. The lower the density of the point, the more likely it is to be identified as the outlier. The key definitions for the LOF are as follows:

Defination 1: k-distance of data point z_p. For a given dataset D and any positive integer k, the k-distance of z_p denoted as k-distance(z_p), is defined as the distance $d(z_p, z_o)$ between data point z_p and data point z_o in the following conditions:

1. for at least k data points $z_{o'} \in D \setminus \{z_p\}$ it holds that $d(z_p, z_{o'}) \leq d(z_p, z_o)$;
2. for at most $k-1$ objects $z_{o'} \in D \setminus \{z_p\}$ it holds that $d(z_p, z_{o'}) < d(z_p, z_o)$.

 doi:10.20944/preprints202501.1914.v1

Defination 2: k-distance neighborhood of data point z_p. Given the k-distance of z_p, the k-distance neighborhood of z_p contains every data point z_q whose distance from z_p is not greater than the k-distance, as described in Equation 7.

$$N_{\text{k-distance}(z_p)}(z_p) = \{z_q \in D \setminus \{z_p\} | d(z_p, z_q) \leq \text{k-distance}(z_p)\} \tag{7}$$

Definition 3: reachability distance of data point z_p with respect to data point z_o. Let k be a natural number. The reachability distance of z_p with respect to z_o is defined as Equation 8.

$$\text{reach-dist}_k(z_p, z_o) = \max(\text{k-distance}(z_o), d(z_p, z_o)) \tag{8}$$

Definition 4: local reachability density of data point z_p. The local reachability density of z_p is defined as Equation 9.

$$lrd_k(z_p) = 1 / \frac{\sum_{z_o \in N_k(z_p)} \text{reach-dist}_k(z_p, z_o)}{|N_k(z_p)|} \tag{9}$$

where $N_k(z_p)$ is the k-distance neighborhood of data point z_p.

Definition 5: local outlier factor of data point z_p. The local outlier factor of z_p is defined as Equation 10.

$$LOF_k(z_p) = \frac{\sum_{z_o \in N_k(z_p)} \frac{lrd_k(z_o)}{lrd_k(z_p)}}{|N_k(z_p)|} \tag{10}$$

The contamination rate hyper-parameter determines the threshold based on the LOF of the training samples. During testing, data points with LOF values greater than the threshold are considered outliers.

4. Results

The entire setup was implemented using Python 3.8.17. The Periodicformer Encoder and the unsupervised representation learning process were developed with PyTorch 2.0.1 and PyTorch-Ignite 0.3.0, utilizing the Adam optimizer with a learning rate of 0.001 during back-propagation. The one-class classifier and its training process were implemented using scikit-learn. Experiments were conducted three times on a system equipped with an Intel i3-12100F CPU and an NVIDIA GeForce RTX 2060 12G GPU, and the average results are reported.

4.1. Dataset

The dataset utilized in this study for experimental evaluation is the publicly available electricity consumption dataset from the Irish CER Smart Metering Project [21]. It includes daily electricity consumption records for approximately 5,000 residential and commercial users. Details of Irish dataset is shown in Table 1.

Table 1. Details of Irish dataset.

Descriptions	Values
Number of smart meters	25730
Time span	14/7/2009-31/12/2010
Recording interval	30 minutes

For the experiments, we use data collected over a 365-day period. We exclude users with missing values (NaN), correct erroneous data, and apply min-max normalization to preprocess the dataset [22]. To enhance detection performance, the detection window is set to cover 7 days ($L = 336$) instead of a single day. The electricity consumption data from 5,000 residential users is split into training, validation, and test sets in a ratio of 8:1:1. The training set is treated as clean, normal data for model

 doi:10.20944/preprints202501.1914.v1

training. In the validation and test sets, 10% of the data is randomly selected and tempered using the attack models described in Section 4.2. The validation set is utilized for hyper-parameter tuning, while the test set is reserved for evaluating the model's anomaly detection performance.

4.2. Attack Model

False data injection is a type of cyber attack in which manipulated or false information is deliberately introduced into a system to mislead its operations or decision-making processes. It is commonly employed to compromise the integrity of data in power systems, sensor networks, and other critical infrastructures.

In this study, to simulate various electricity theft patterns, six different false data injection attacks are implemented, as shown in Table 2. These attacks are broadly categorized into two types: reduced consumption attacks and load profile shifting attacks [1]. Types 1–4 primarily reduce the electricity consumption recorded by the meter, while Types 5 and 6 shift electricity usage from high-price periods to low-price periods.

The physical interpretations of these six attacks are as follows: FDI1: Electricity consumption recorded at each time point is scaled down by a randomly selected proportion. FDI2: A constant value is subtracted from the electricity consumption data. FDI3: Electricity consumption values exceeding a preset threshold are capped at the threshold value. FDI4: Consumption data within a randomly selected time interval during the day is replaced with zeros. FDI5: The total daily electricity consumption is evenly distributed across all recording points and reduced by a fixed proportion. FDI6: The entire daily electricity consumption profile is inverted. Additionally, Figure 4 illustrates examples of both normal electricity consumption data and data affected by each of these six attack types.

Table 2. Six types of FDI attacks.

Attack Type	Modification
1	$\tilde{x}_t = \alpha_t x_t, \alpha_t = random(0.2, 0.8)$
2	$\tilde{x}_t = \max(x_t - \gamma, 0), \gamma < \max(X)$
3	$\tilde{x}_t = \min(x_t, \gamma), \gamma < max(X)$
4	$\tilde{x}_t = \begin{cases} 0, & t_{start} < t < t_{start} + \Delta t \\ x_t, & otherwise \end{cases}$
5	$\tilde{x}_t = \alpha mean(X), \alpha = random(0.2, 0.8)$
6	$\tilde{x}_t = x_{n-t}$

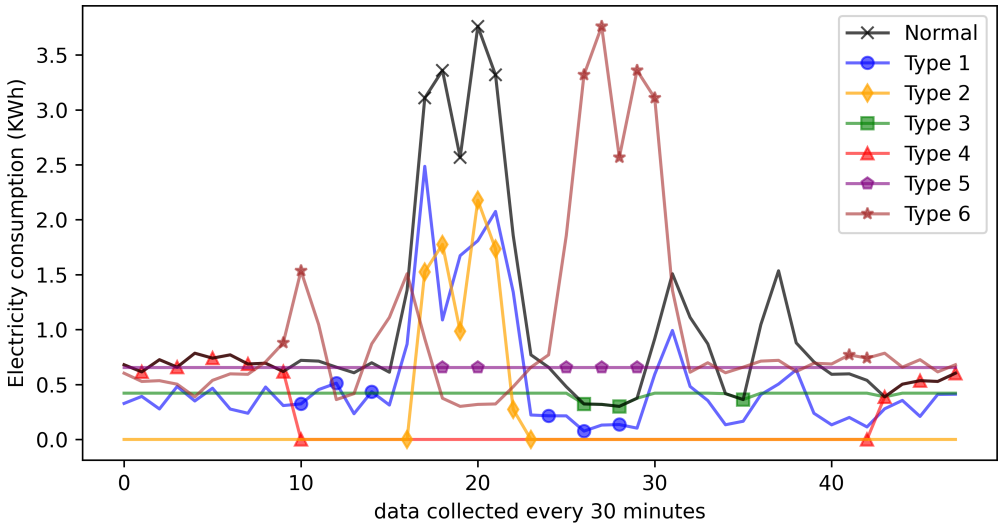

Figure 4. Example of daily electricity consumption and its tampered data corresponding to six types of FDI attacks.

 doi:10.20944/preprints202501.1914.v1

4.3. Evaluation Metrics

The confusion matrix quantifies all classification outcomes of a model, providing an intuitive representation of classification accuracy. In this paper, electricity theft samples are considered as the positive class, while normal electricity usage samples are treated as the negative class. The definitions of true positive (TP), false positive (FP), false negative (FN), and true negative (TN) are summarized in Table 3.

Table 3. Confusion matrix of electricity theft detection results.

Samples	Actually positive	Actually negative
Predicted positive	TP	FP
Predicted negative	FN	TN

Based on the above confusion matrix, we use the F1 score (F1), the area under the receiver operating characteristic (ROC) curve (AUC), Recall, and false positive rate (FPR) to evaluate the performance of the proposed electricity theft detection framework.

Recall, also known as true positive rate (TPR), evaluates a model's effectiveness in identifying all actual positive instances within a dataset, highlighting its sensitivity to true positives. It is calculated as follows.

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

FPR quantifies the rate at which negative samples are misclassified as positive, reflecting the model's tendency for false alarms. It is computed as follows.

$$FPR = \frac{FP}{FP + TN} \tag{12}$$

In anomaly detection, we generally aim for a model that has a low FPR while maintaining a high TPR.

Precision quantifies the accuracy of positive predictions, calculated as follows.

$$precision = \frac{TP}{TP + FP} \tag{13}$$

F1 captures the trade-off between Precision and Recall, offering a single metric to evaluate performance, particularly in scenarios with imbalanced data. It is calculated as follows.

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{14}$$

AUC evaluates a model's capability to differentiate between classes over varying decision thresholds, with larger values reflecting stronger discriminative power.

4.4. Analysis of Model Hyper-Parameters

We applied grid search on the validation set to identify the optimal hyper-parameters. The selection ranges for each parameter and the corresponding best results are presented in Table 4.

For the Periodicformer encoder, the hyper-parameters primarily consist of three components: the number of heads H, the encoder embedding dimension d and the number of layers N. The best result H, d and N in this paper is consistent with that in Reference [23]. As for the patch size for each layer p_i, this paper empirically concludes that the patch size should initially be larger and then gradually decrease. For LOF, we explore three values of number of neighbors k, and test three different values for the contamination rate parameters to control the expected proportion or number of anomaly points. Additionally, we experiment two different distance metrics. The best combination of parameters was selected based on performance.

 doi:10.20944/preprints202501.1914.v1

Table 4. Selection of hyper-parameters.

Hyper-parameter	Range of values	Optimal value
k	100, 200, 300	200
contamination rate	0.01, 0.1, 0.2	0.01
distance metric	'manhattan', 'euclidean'	'euclidean'
d	64,128	128
H	8,16	8
N	2,3,4	3

4.5. Comparison with Other Methods

In this part, we compare the the proposed method with six one-class classification-based ETD methods. We implement all these methods on the preprocessed data for fair comparison.

1. OCSVM [24]: OCSVM is a support vector machine-based algorithm designed for anomaly detection. It distinguishes normal samples from anomalies by maximizing the margin between the data and the origin in the feature space. We set the kernel function, ν parameters as 'linear' and 0.2 respectively;
2. iForest [25]: iForest is an efficient unsupervised anomaly detection method leveraging random forests. It isolates anomalies by exploiting their tendency to be easier to separate from the majority of the data through random partitioning. We set the contamination, maxFeatures and nEstim parameters as 0.1, 0.6 and 300 respectively;
3. LOF [26]: LOF is a density-based anomaly detection algorithm. It detects outliers by comparing the local density of a sample with the densities of its neighbors, identifying anomalies as points with significantly lower densities. We set the contamination rate as 0.1 while keeping other parameters as default;
4. Autoencoder with one-class (OC) classifier [27]: Reference [27] incorporates the concept of autoencoding and OC classification. We implement this method using CNN-AE(1D) as the autoencoder along with three OC classifiers.

Table 5 presents the comparison results between our proposed method and six baseline methods, and following findings can be observed.

- Overall, the combination of the Periodicformer encoder and LOF achieved the best performance, with F1, AUC, Recall, and FPR scores of 0.833, 0.973, 0.877, and 0.025, respectively. Compared to the second-best approach, the combination of Autoencoder and LOF (F1, AUC, Recall, and FPR scores of 0.814, 0.952, 0.856, and 0.027, respectively), it achieved improvements of 2.3%, 1.7%, 2.4%, and 8% in F1, AUC, Recall, and FPR, respectively;
- The three approaches that rely solely on the OC classifier exhibit relatively poor anomaly detection performance, with some even failing (defined as AUC or Recall values below 0.5). This observation aligns with the findings reported in Reference [27]. This suggests that the one-class classifier fails to effectively capture the distribution of normal samples in the training dataset, making it difficult to distinguish between normal and anomalous samples in the test set. This underscores the importance of designing a robust unsupervised representation learning method to extract meaningful features, aiding the one-class classifier in better modeling the distribution of normal samples and improving anomaly detection performance;
- Comparing the proposed combination of the Periodicformer encoder and OC classifier with the combination of Autoencoder and OC classifier, specifically the combination of Periodicformer encoder and LOF versus Autoencoder and LOF, the Periodicformer encoder demonstrates superior suitability for modeling time series data. Additionally, our proposed criss-cross periodic attention further enhances the model's ability to capture richer periodic features, which helps distinguish normal data from anomalous data. In contrast, the Autoencoder relies solely on CNN to capture local features without considering the periodicity of time series data, resulting in poorer performance;

 doi:10.20944/preprints202501.1914.v1

- In the three combinations of the Periodicformer encoder with different OC classifiers, the scheme combined with iForest yields the poorest performance. According to Reference [28], iForest is sensitive only to global outliers and struggles with detecting local outliers. The distribution of power theft behavior data is particularly unfavorable for iForest detection. For instance, when considering an entire year (365 days), power theft typically occurs on specific days or weeks. These anomalies may not be noticeable when viewed in the context of the whole year. However, when compared to adjacent days or weeks, these outliers may become more apparent. This is precisely where iForest's limitations lie. Regarding the combination with OCSVM, Reference [29] states that OCSVM is sensitive to noise and prone to false positives, which aligns with our experimental findings. Although it achieves a high recall, it also results in a relatively high false positive rate.

For anomaly detection tasks, our goal is to develop a model that achieves a low false positive rate (FPR) while maintaining a high true positive rate (TPR). To showcase the performance of all the methods in this regard, we have plotted the ROC curve, as shown in Figure 5. The proposed combination of the Periodicformer encoder and LOF achieves the highest TPR at the lowest FPR, and the combination of Autoencoder and iForest yields the poorest performance.

Table 5. Performance of different methods.

Model for unsupervised representation learning	OC classifier	Metrics			
		F1	AUC	Recall	FPR
-	OCSVM	0.464	0.893	0.816	0.217
	iForest	-	-	-	-
	LOF	-	-	-	-
Periodicformer encoder	OCSVM	0.517	0.940	0.918	0.189
	iForest	0.434	0.860	0.752	0.194
	LOF	0.833	0.973	0.877	0.025
Autoencoder	OCSVM	0.505	0.917	0.890	0.209
	iForest	0.420	0.840	0.743	0.223
	LOF	0.814	0.952	0.856	0.027

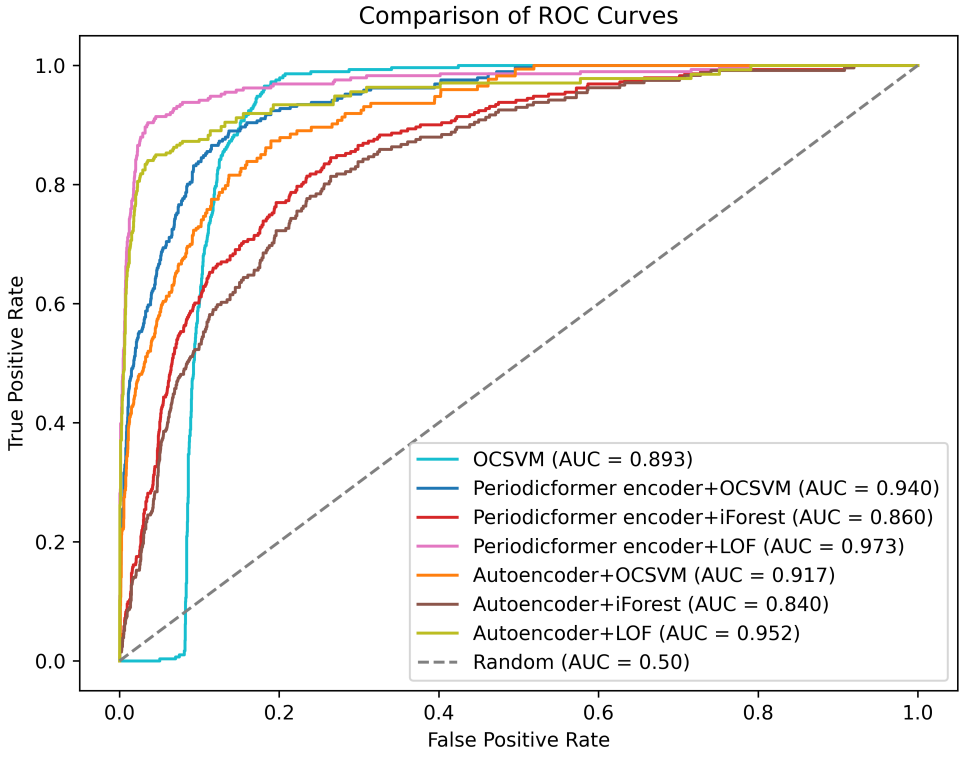

Figure 5. ROC curves of the proposed method and other methods and the value of the area under each curve AUC.

 doi:10.20944/preprints202501.1914.v1

4.6. Ablation Analysis

An ablation study is conducted to demonstrate the impact of the main components in the proposed framework on the ETD performance. Specifically, three variants of our proposed method are implemented.

1. RA-DOCC: This method uses a combination of a Periodicformer encoder with only row auto-correlation and LOF;
2. CA-DOCC: This method uses a combination of a Periodicformer encoder with only column auto-correlation and LOF;
3. Vanilla-DOCC: This method uses a combination of a vanilla Transfomer encoder [30] and LOF.

Figure 6 shows the results of the ablation experiments, from which the following findings can be drawn:

- RA-DOCC and CA-DOCC produce comparable experimental results, but show notable improvements over Vanilla-DOCC. Specifically, F1, AUC, Recall, and FPR improve by approximately 5%, 1%, 4%, and 34%, respectively. This suggests that row auto-correlation and column auto-correlation are all effective in capturing periodic features, which help the model distinguish anomalous samples;
- Our proposed Criss-cross periodic attention effectively integrates multiple periodic features. Compared to the vanilla Transformer encoder, F1, AUC, Recall, and FPR have improved by 15.3%, 3.2%, 12.5%, and 104%, respectively.

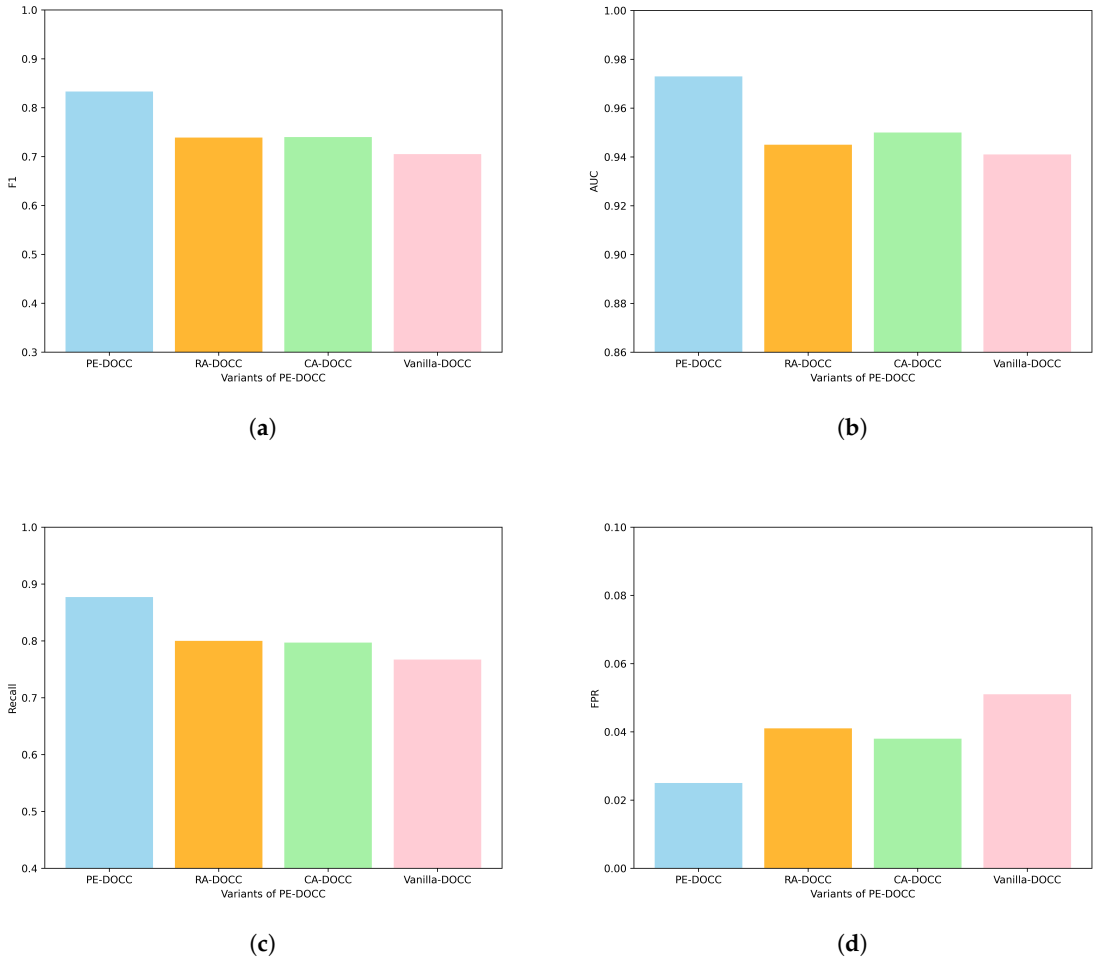

Figure 6. Performance of proposed method and its variants. (**a**) F1 scores of PE-DOCC and its variants. (**b**) AUC of PE-DOCC and its variants. (**c**) Recall of PE-DOCC and its variants. (**d**) FPR of PE-DOCC and its variants.

 doi:10.20944/preprints202501.1914.v1

5. Discussion and Conclusions

This paper proposes a novel periodicity-enhanced deep one-class classification framework for electricity theft detection, named PE-DOCC. Overall, the training samples first undergo unsupervised representation learning (i.e., pre-train the Periodicformer encoder). Then, the training samples are passed through the pretrained encoder to obtain latent representations, which are finally fed into a one-class classifier. During testing, the trained Periodicformer encoder and classifier are used to determine whether the testing latent representation falls within the trained boundary. To enhance the model's ability to capture periodic features and improve its capacity to distinguish anomalous samples from normal ones, the paper introduces a criss-cross periodic attention mechanism that comprehensively considers both horizontal and vertical periodic features. Experiments are conducted on a real-world electricity consumption dataset, and the results show that the proposed ETD framework can effectively detect anomalous electricity behaviors and outperforms state-of-the-art ETD methods. In future work, we will explore how to use Graph Neural Networks (GNN) to uncover the relationships between different variables in multivariate datasets to further improve model performance.

References

1. Xia, X.; Xiao, Y.; Liang, W.; Cui, J. Detection Methods in Smart Meters for Electricity Thefts: A Survey. *Proceedings of the IEEE* **2022**, *110*, 273–319. https://doi.org/10.1109/JPROC.2021.3139754.
2. Yan, Z.; Wen, H. Performance Analysis of Electricity Theft Detection for the Smart Grid: An Overview. *IEEE Transactions on Instrumentation and Measurement* **2022**, *71*, 1–28. https://doi.org/10.1109/TIM.2021.3127649.
3. Stracqualursi, E.; Rosato, A.; Di Lorenzo, G.; Panella, M.; Araneo, R. Systematic review of energy theft practices and autonomous detection through artificial intelligence methods. *Renewable and Sustainable Energy Reviews* **2023**, *184*, 113544. https://doi.org/https://doi.org/10.1016/j.rser.2023.113544.
4. Xia, X.; Lin, J.; Jia, Q.; Wang, X.; Ma, C.; Cui, J.; Liang, W. ETD-ConvLSTM: A Deep Learning Approach for Electricity Theft Detection in Smart Grids. *IEEE Transactions on Information Forensics and Security* **2023**, *18*, 2553–2568. https://doi.org/10.1109/TIFS.2023.3265884.
5. Zhu, Y.; Zhang, Y.; Liu, L.; Liu, Y.; Li, G.; Mao, M.; Lin, L. Hybrid-Order Representation Learning for Electricity Theft Detection. *IEEE Transactions on Industrial Informatics* **2023**, *19*, 1248–1259. https://doi.org/10.1109/TII.2022.3179243.
6. Liao, W.; Yang, Z.; Liu, K.; Zhang, B.; Chen, X.; Song, R. Electricity Theft Detection Using Euclidean and Graph Convolutional Neural Networks. *IEEE Transactions on Power Systems* **2023**, *38*, 3514–3527. https://doi.org/10.1109/TPWRS.2022.3196403.
7. Takiddin, A.; Ismail, M.; Zafar, U.; Serpedin, E. Deep Autoencoder-Based Anomaly Detection of Electricity Theft Cyberattacks in Smart Grids. *IEEE Systems Journal* **2022**, *16*, 4106–4117. https://doi.org/10.1109/JSYST.2021.3136683.
8. Liang, Q.; Zhao, S.; Zhang, J.; Deng, H. Unsupervised BLSTM-Based Electricity Theft Detection with Training Data Contaminated. *ACM Trans. Cyber-Phys. Syst.* **2024**, *8*. https://doi.org/10.1145/3604432.
9. Huang, Y.; Xu, Q. Electricity theft detection based on stacked sparse denoising autoencoder. *International Journal of Electrical Power & Energy Systems* **2021**, *125*, 106448. https://doi.org/https://doi.org/10.1016/j.ijepes.2020.106448.
10. Cai, Q.; Li, P.; Wang, R. Electricity theft detection based on hybrid random forest and weighted support vector data description. *International Journal of Electrical Power & Energy Systems* **2023**, *153*, 109283. https://doi.org/https://doi.org/10.1016/j.ijepes.2023.109283.
11. Senol, N.S.; Baza, M.; Rasheed, A.; Alsabaan, M. Privacy-Preserving Detection of Tampered Radio-Frequency Transmissions Utilizing Federated Learning in LoRa Networks. *Sensors* **2024**, *24*. https://doi.org/10.3390/s24227336.
12. Yi, X.; Yang, X.; Huang, Y.; Ke, S.; Zhang, J.; Li, T.; Zheng, Y. Gas-Theft Suspect Detection Among Boiler Room Users: A Data-Driven Approach. *IEEE Transactions on Knowledge and Data Engineering* **2022**, *34*, 5796–5808. https://doi.org/10.1109/TKDE.2021.3062707.
13. Barbariol, T.; Susto, G.A. TiWS-iForest: Isolation forest in weakly supervised and tiny ML scenarios. *Information Sciences* **2022**, *610*, 126–143. https://doi.org/https://doi.org/10.1016/j.ins.2022.07.129.

14. Tirulo, A.; Chauhan, S.; Issac, B. Ensemble LOF-based detection of false data injection in smart grid demand response system. *Computers and Electrical Engineering* **2024**, *116*, 109188. https://doi.org/https://doi.org/10.1016/j.compeleceng.2024.109188.
15. Liang, D.; Wang, J.; Gao, X.; Wang, J.; Zhao, X.; Wang, L. Self-supervised Pretraining Isolated Forest for Outlier Detection. In Proceedings of the 2022 International Conference on Big Data, Information and Computer Network (BDICN), 2022, pp. 306–310. https://doi.org/10.1109/BDICN55575.2022.00065.
16. Kim, C.; Chang, S.Y.; Kim, J.; Lee, D.; Kim, J. Automated, Reliable Zero-Day Malware Detection Based on Autoencoding Architecture. *IEEE Transactions on Network and Service Management* **2023**, *20*, 3900–3914. https://doi.org/10.1109/TNSM.2023.3251282.
17. Cui, X.; Liu, S.; Lin, Z.; Ma, J.; Wen, F.; Ding, Y.; Yang, L.; Guo, W.; Feng, X. Two-Step Electricity Theft Detection Strategy Considering Economic Return Based on Convolutional Autoencoder and Improved Regression Algorithm. *IEEE Transactions on Power Systems* **2022**, *37*, 2346–2359. https://doi.org/10.1109/TPWRS.2021.3114307.
18. Zeng, A.; Chen, M.; Zhang, L.; Xu, Q. Are transformers effective for time series forecasting? In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2023, Vol. 37, pp. 11121–11128.
19. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In Proceedings of the Advances in Neural Information Processing Systems; Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; Vaughan, J.W., Eds. Curran Associates, Inc., 2021, Vol. 34, pp. 22419–22430.
20. Tuli, S.; Casale, G.; Jennings, N.R. TranAD: deep transformer networks for anomaly detection in multivariate time series data. *Proc. VLDB Endow.* **2022**, *15*, 1201–1214. https://doi.org/10.14778/3514061.3514067.
21. for Energy Regulation (CER), C. CER smart metering project-electricity customer behaviour trial, 2009-2010, 2012.
22. Qi, R.; Zheng, J.; Luo, Z.; Li, Q. A Novel Unsupervised Data-Driven Method for Electricity Theft Detection in AMI Using Observer Meters. *IEEE Transactions on Instrumentation and Measurement* **2022**, *71*, 1–10. https://doi.org/10.1109/TIM.2022.3189748.
23. Zerveas, G.; Jayaraman, S.; Patel, D.; Bhamidipaty, A.; Eickhoff, C. A Transformer-based Framework for Multivariate Time Series Representation Learning. In Proceedings of the Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, New York, NY, USA, 2021; KDD '21, p. 2114–2124. https://doi.org/10.1145/3447548.3467401.
24. Schölkopf, B.; Platt, J.C.; Shawe-Taylor, J.; Smola, A.J.; Williamson, R.C. Estimating the support of a high-dimensional distribution. *Neural computation* **2001**, *13*, 1443–1471.
25. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation Forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, 2008, pp. 413–422. https://doi.org/10.1109/ICDM.2008.17.
26. Breunig, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: identifying density-based local outliers. *SIGMOD Rec.* **2000**, *29*, 93–104. https://doi.org/10.1145/335191.335388.
27. Kim, C.; Chang, S.Y.; Kim, J.; Lee, D.; Kim, J. Automated, Reliable Zero-Day Malware Detection Based on Autoencoding Architecture. *IEEE Transactions on Network and Service Management* **2023**, *20*, 3900–3914. https://doi.org/10.1109/TNSM.2023.3251282.
28. Cheng, Z.; Zou, C.; Dong, J. Outlier detection using isolation forest and local outlier factor. In Proceedings of the Proceedings of the Conference on Research in Adaptive and Convergent Systems, New York, NY, USA, 2019; RACS '19, p. 161–168. https://doi.org/10.1145/3338840.3355641.
29. Lu, T.; Wang, L.; Zhao, X. Review of Anomaly Detection Algorithms for Data Streams. *Applied Sciences* **2023**, *13*. https://doi.org/10.3390/app13106353.
30. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.u.; Polosukhin, I. Attention is All you Need. In Proceedings of the Advances in Neural Information Processing Systems; Guyon, I.; Luxburg, U.V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; Garnett, R., Eds. Curran Associates, Inc., 2017, Vol. 30.