# Research on Electricity Theft Detection Technology Based on Deep Learning

**Ren Lu, Gong Dianxuan*, Zhang Hao**
*School of Science, North China University of Science and Technology, Tangshan, Hebei, China*
**Corresponding Author*

**Abstract: The global issue of electricity theft has become increasingly severe, posing significant threats to the stability of power systems and public safety. This paper introduces a Multi-Dimensional Perception Network for Electricity Anomaly Detection, which integrates the DSAE module and MSSPAM module to enhance the model's capability in capturing both global and local features and improving temporal dependency modeling. Experimental results demonstrate that the proposed model outperforms existing baseline methods across multiple metrics, including accuracy, precision, and recall, providing an efficient and reliable solution for electricity theft detection.**

**Keywords: Electricity Theft Detection; Anomaly Detection; Deep Learning; Feature Fusion; Attention Mechanism**

## 1. Introduction

Electricity theft is a global issue that undermines public safety and power system stability, contributing to grid failures and safety risks while hindering the industry's sustainable growth. Annual losses exceed $25 billion worldwide. In Pakistan, theft rates surpass 80% in Dera Ismail Khan and 95% in Bannu, resulting in power outages over 12 hours in some areas [1] . the World Bank reports that electricity theft reduces India's GDP by 1.5%, causing losses of around $4.5 billion annually [2] [3] . In Guangdong, China, theft and infrastructure damage have led to direct losses exceeding 70 million yuan over the past decade [4] . These issues threaten public safety, system stability, and long-term industry growth.

Electricity theft detection methods can be classified into three categories: rule-based, statistical, and machine learning-based approaches. Rule-based methods use predefined rules, such as consumption pattern analysis and load anomaly detection, but struggle with complex theft behaviors and exhibit low accuracy. Statistical methods, like Z-Score and time series analysis, model normal user behavior to detect anomalies but fail to capture intricate patterns. Machine learning techniques, including classification algorithms, neural networks, and anomaly detection, offer improved accuracy by automatically extracting features from large datasets, with deep learning and reinforcement learning showing particular promise in complex scenarios.

The main contributions of this paper can be summarized as follows:

An innovative attention mechanism is proposed that integrates feature information from multiple input channels, enhancing the model's ability to capture key features. This mechanism enables flexible and precise adjustment of attention distribution.

Explicit position encoding, combined with a multi-scale feature aggregation strategy, enhances the model's ability to capture temporal dependencies in sequential data. This approach improves the model's ability to model sequential features while ensuring computational efficiency.

The structure of this paper is as follows: Section 2 reviews related work, Section 3 outlines the technologies used, Section 4 details the experimental setup, and Section 5 concludes with future directions.

## 2. Related Work

Machine learning and data analytics are crucial for improving the accuracy and efficiency of electricity theft detection in smart grids, with recent advancements in pattern recognition further enhancing detection performance..

RK Ahir et al. [5] proposed a method combining Dynamic Time Warping (DTW) with k-Nearest Neighbors (k-NN), achieving an F1 score of 94% and a True Positive Rate (TPR) of 93%, demonstrating effective anomaly detection. Deep learning approaches have further boosted detection accuracy. LJ Lepolesa et al. [6] developed a DNN-based method integrating time-domain and frequency-domain features, improving AUC by 1% and accuracy by 1.8%. Xiaofang Xia et al. [7] introduced the ETD-ConvLSTM method, converting time series into spatiotemporal data, reducing detection time and enhancing efficiency. K Zheng et al. [8] combined the Maximum Information Coefficient (MIC) and CFSFDP clustering, improving AUC by 10% and MAP@200 by 20% through shape similarity and amplitude correlation. S Munawar et al. [9] proposed a hybrid method with BiGRU and BiLSTM networks, increasing accuracy from 88.7% to 95%. KV Blazakis et al. [10] combined ANN and ANFIS, outperforming SVM and RBF in low electricity theft scenarios. R Akram et al. [11] introduced a method using CNN and RUSBoost optimization, achieving 93.5% accuracy with bio-inspired algorithms for improved training and classification performance.

While existing methods have achieved notable results, they often rely on single data features for electricity anomaly detection, particularly in feature selection and utilization. Electricity consumption data possess complex spatiotemporal characteristics, encompassing both global features (e. g., seasonal trends, weather, and holiday effects) and local features (e. g., daily patterns and sudden energy spikes). As a result, current models often struggle to fully capture these complexities when focusing on only one type of feature. the solution proposed in this paper addresses this limitation by combining both global and local features, leading to improved classification performance.

## 3. Related Methods

### 3.1 Dual-Source Attention Enhancer

The Channel Attention Mechanism [12] optimizes feature representation in deep learning by dynamically assigning weights to channels, allowing the model to prioritize relevant information. In CNNs, feature maps are composed of multiple channels, each representing different patterns, with varying levels of importance. the mechanism learns optimal weights for each channel, improving the model's discriminative power.

The Dual-Source Attention Enhancer (DSAE), shown in **Figure 1,** enhances feature representation by using a dual-input feature map structure based on the SE attention mechanism. Unlike traditional modules that process a single input feature map, this dual-input structure allows the model to integrate information from multiple sources or levels, improving the flexibility and precision of channel attention.

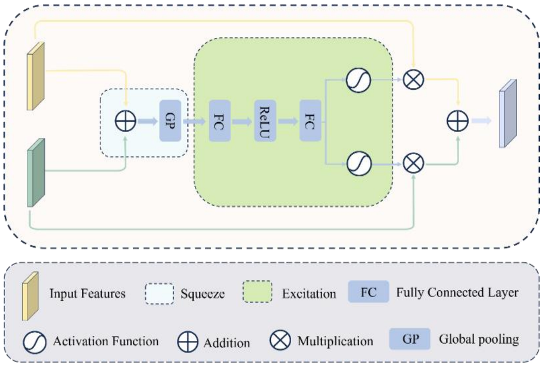


**Figure 1. Dual-Source Attention Enhancer**

The DSAE module takes two input feature maps, $F_1$ and $F_2$, both of which have identical dimensions, $F_1, F_2 \in R^{C\times H\times W}$. These feature maps are then fused on a per-channel basis:

$$F_{merge} = F_1 + F_2$$

Where $F_{merge}$ represents the fused feature map. This design enables the model to simultaneously capture the global semantic information from multiple sources of features.

The fused feature map is processed through global average pooling to generate a global feature vector:

$$g_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{merge,c}(i,j)$$

The global feature vector is passed through two fully connected layers to generate two independent channel weight vectors, $z_1$ and $z_2$, which are used to weight the two input feature maps individually:

$$z_1 = \sigma\left(W_{2,1}\cdot\delta(W_{1,1}\cdot g)\right), z_2 = \sigma\left(W_{2,2}\cdot\delta(W_{1,2}\cdot g)\right)$$

Where $W_{1,1}, W_{1,2}, W_{1,2}$, and $W_{1,2}$ are the weight matrices of the fully connected layers.

The weight vectors are applied to the two input feature maps, generating the re-weighted feature maps:

$$F'_{1,c} = z_{1,c}\cdot F_{1,c}, F'_{2,c} = z_{2,c}\cdot F_{2,c}$$

Finally, the two re-weighted feature maps are fused together:

$$F_{out} = F_1' + F_2'$$

The output feature map is added to the original fused feature map to alleviate the vanishing gradient problem and facilitate the flow of information:

$$F_{final} = F_{out} + F_{merge}$$

The DSAE module enhances feature representation through a dual-input structure, allowing the model to integrate feature information from multiple sources while strengthening both global and local representations. By using two independent fully connected layers, the module optimizes channel dependencies, adjusting weights to highlight essential features and reduce redundancy. the addition of residual connections improves training stability and prevents the vanishing gradient problem in deep networks. This module is highly effective in multi-level feature fusion tasks and excels in deep learning applications like classification, detection, and sequence modeling.

### 3.2 Multi-Scale Sequential Perception Attention Module

The Multi-Scale Linear Attention Mechanism [13] is an advanced method that combines the low complexity of linear attention with the benefits of multi-scale feature representations. This approach enhances feature extraction and improves task performance. Traditional self-attention mechanisms capture dependencies by computing a global similarity matrix with a complexity of $O(N^2)$, where NN is the sequence length. This becomes costly in tasks with long sequences. the linear attention mechanism uses a kernel function to map the query (Q) and key (K) vectors nonlinearly, avoiding the direct computation of the global similarity matrix, and reducing the complexity to $O(N)$. This effectively reduces computational costs while preserving the ability to capture global dependencies.

Multi-scale feature representation enhances the attention mechanism by downsampling input features at various scales, generating multiple sets of sub-features. Each set undergoes independent attention computation, and the resulting feature representations are weighted and fused to obtain the final output. This method enables the model to capture both global information and local details, improving its overall feature representation capabilities.

To overcome the limitations of traditional multi-scale linear attention mechanisms in sequence modeling, the Multi-Scale Sequential Perception Attention Module (MSSPAM) incorporates explicit positional encoding (PE). This addition enables the model to capture the relative relationships between data points more accurately, especially for data with strong temporal or spatial dependencies. Unlike traditional methods, MSSPAM combines multi-scale feature aggregation with positional encoding, significantly enhancing the model's ability to process sequential data.

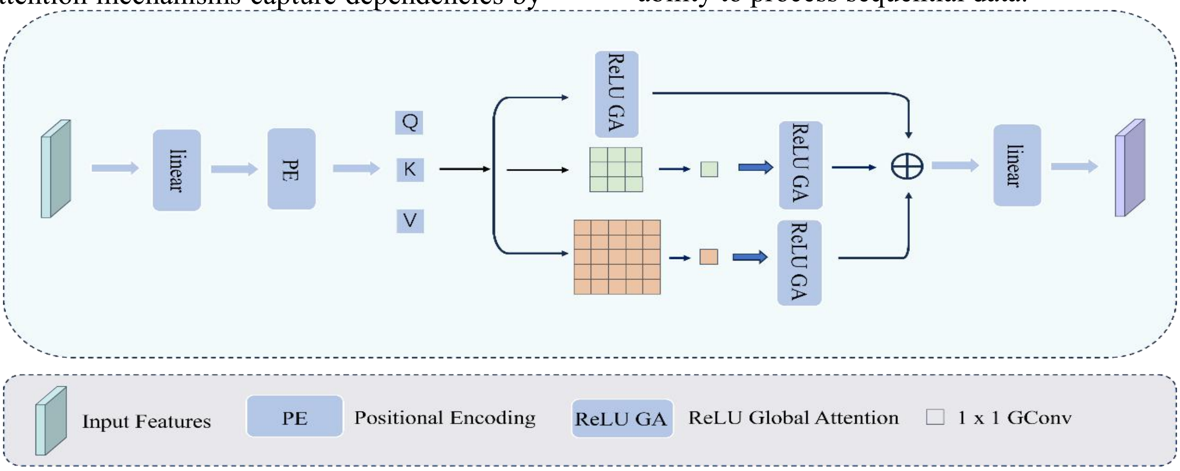


**Figure 2. Multi-Scale Sequential Perception Attention Module**

The MSSPAM module generates multiple scales of feature subsets $F^s$ by downsampling the input feature map $F \in R^{C\times H\times W}$, where $C$, $H$, and $W$ represent the number of channels, height, and width, respectively. At each scale, positional encoding $PE^s$ is explicitly added to the feature subsets, providing positional information for each feature point:

$$F_{pe}^s = F^s + PE^s$$

Here, $PE^s \in R^{C\times H_s\times W_s}$ represents the learnable positional encoding matrix, where $H_s$ and $W_s$ denote the dimensions of the downsampled feature subset. the inclusion of positional information effectively enhances the model's ability to capture both global and local dependencies in sequence tasks.

For each feature subset $F^s_{pe}$ with added positional information, independent linear transformations are applied to generate the query (Q), key (K), and value (V) matrices:

$$Q^s = W^s_q F^s_{pe}, K^s = W^s_k F^s_{pe}, V^s = W^s_v F^s_{pe}$$

Where $W^s_q, W^s_k$, and $W^s_v$ are the weight matrices for the linear transformations. Then, the attention output is computed separately for each scale:

$$A^s = \frac{\phi(Q^s)\cdot\left(\phi(K^s)^T \cdot V^s\right)}{\phi(Q^s)\cdot\phi(K^s)^T}$$

Finally, by performing weighted fusion of the attention outputs from different scales, a consolidated feature representation is generated:

$$A_{\text{multi}} = \sum_s w_s \cdot A^s$$

where $w_s$ is the weight learned through training.

By explicitly incorporating positional information, the MSSPAM module can accurately capture the temporal or spatial relationships between feature points. When handling long sequence tasks or complex spatial data, the model's ability to capture both local details and global dependencies is significantly enhanced.

To reduce model complexity and improve computational efficiency, the attention module employs a residual connection mechanism, where the features with added positional information are directly summed with the attention output:

$$F_{\text{out}} = F_{\text{in}} + A_{\text{multi}}$$

This design preserves the original information of the features while optimizing the model's training efficiency and stability.

Experimental results show that the MSSPAM module outperforms traditional methods across various tasks. the inclusion of positional information enhances the model's ability to capture long-range dependencies and local details, particularly improving performance on data with complex temporal features. By combining multi-scale features with positional information, the model achieves excellent results in classification, anomaly detection, and sequence prediction.

This design efficiently addresses the performance limitations of traditional linear attention mechanisms in long-sequence modeling by explicitly modeling positional information. the combination of multi-scale fusion and positional encoding provides an effective feature extraction solution that balances efficiency and robustness for long-sequence tasks.

### 3.3 Multi-Dimensional Perception for Electricity Anomaly Detection Network

Building on the theoretical foundations and innovations of multi-scale linear attention and channel attention mechanisms, an efficient feature fusion model for electricity anomaly detection, called the Multi-Dimensional Perception for Electricity Anomaly Detection Network (MDPAN), has been developed. As shown in **Figure 3**, MDPAN integrates multi-level feature extraction and information optimization strategies, enabling collaborative modeling of global and local features, and demonstrating exceptional performance in complex data scenarios.

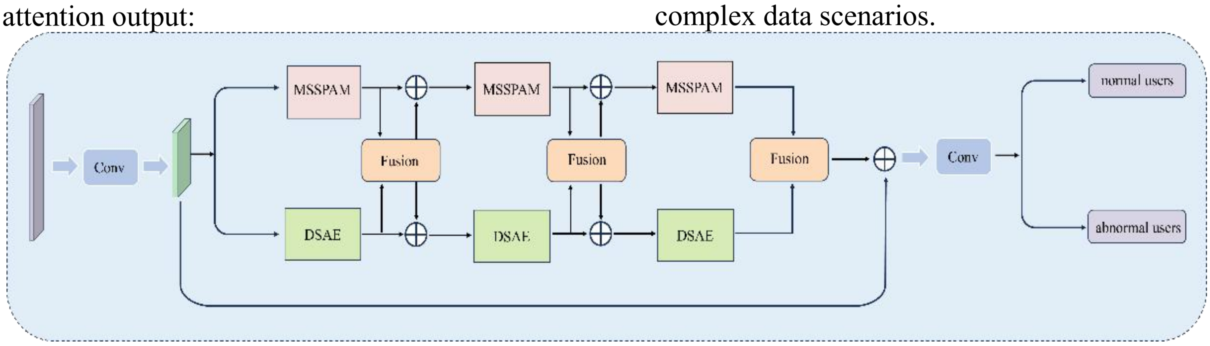


**Figure 3. Multi-Dimensional Perception for Electricity Anomaly Detection Network**

The MDPAN model consists of several key modules designed to enhance feature extraction and anomaly detection. Initially, convolutional layers process the input data, extracting local features while preserving spatial structure. the MSSPAM module aggregates multi-scale features and incorporates positional encoding to capture both global dependencies and local details, improving the modeling of temporal relationships. the DSAE module optimizes multi-channel features through adaptive weighting, enhancing feature complementarity and flexibility in selection. Residual connections improve model stability and feature flow. Multi-level features are consolidated through a fusion module, producing high-dimensional representations that combine global and local information. Finally, these features are mapped to the output layer for anomaly recognition and classification.

The MDPAN model integrates multi-scale features and multi-channel information through the synergistic interaction of the MSSPAM and DSAE modules. the MSSPAM module enhances contextual awareness by capturing key features at various resolutions, while the DSAE module improves the discriminative power of global features through inter-channel selection and weight allocation. the introduction of explicit positional encoding addresses the limitations of traditional linear attention mechanisms, enabling the model to capture complex dependencies in temporal or spatial data. the combination of positional information and multi-scale feature fusion optimizes performance in long-sequence tasks. Additionally, the lightweight architecture and simplified attention calculations, along with residual connections, reduce computational complexity, improving training efficiency and inference speed.

In the electricity anomaly detection task, the MDPAN model effectively identifies anomaly patterns by leveraging deep feature extraction and optimization. the integration of multi-scale features and explicit positional encoding enables the model to capture both long-range dependencies and short-term fluctuations in electricity usage data. Experimental results show that the MDPAN model outperforms traditional methods in classification accuracy across various test scenarios. the lightweight design of the MSSPAM module reduces computational complexity compared to models using traditional self-attention mechanisms, ensuring higher efficiency in long-sequence tasks with fewer computational resources. the dynamic weighting and residual connection mechanisms in the DSAE module enhance feature selection and training stability. Overall, the model demonstrates strong adaptability and robustness across different data distributions and task environments, supporting its practical application.

## 4. Experimental Procedure

### 4.1 Dataset

The dataset used in the experiment is sourced from the State Grid Corporation of China (SGCC) and contains daily electricity usage data from 42, 372 users over a period of 4, 034 days. the statistical information is presented in **Table 1**.

**Table 1. Information of the State Grid Corporation of China (SGCC) Dataset**

| Description | Value |
|---|---|
| Data Collection Period | 2014/1/1~2016/11/1 |
| Data Type | Time Series Data |
| Total Number of Users | 42372 |
| Number of Users with Normal Electricity Usage | 38757 |
| Number of Users with Abnormal Electricity Usage | 3615 |

### 4.2 Data Preprocessing

To enhance classification accuracy, preprocessing steps must address issues such as missing values and outliers. Missing values are handled through interpolation; if both neighboring values are numeric, the missing value is replaced with their average; otherwise, it is set to zero. Outliers are detected using the following formula:

$$x_i > \text{avg}(X) + 2 \times \text{std}(X)$$

where $X$ represents the dataset, and $\text{avg}(X)$ and $\text{std}(X)$ denote its mean and standard deviation, respectively. Normalization is performed using the formula:

$$x^{'} = \frac{x - \min}{\max - \min}$$

where $x^{'}$ is the normalized value, and min and max are the minimum and maximum values of the column. To address class imbalance, the SMOTE method is employed, which generates new minority class samples by interpolating between the k-nearest neighbors of each minority sample. This approach improves the model’s ability to detect minority class instances.

### 4.3 Performance Metrics

In electricity anomaly detection, evaluating model performance requires selecting appropriate metrics. the confusion matrix visualizes the relationship between predictions and actual labels, while accuracy, precision, and recall assess correct classification, positive prediction accuracy, and the ability to detect true positives, respectively. the Matthews Correlation Coefficient (MCC) offers a balanced evaluation, especially for imbalanced datasets. the ROC curve and AUC value show the model's performance across thresholds, indicating its ability to distinguish between positive and negative samples. These metrics together provide a comprehensive performance assessment for anomaly detection.

### 4.4 Experimental Results

The proposed method was evaluated by comparing its performance with benchmark models, including Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Convolutional Neural Network (CNN), and Long Short-Term Memory Network (LSTM), using reliable performance metrics. the comparison was conducted with a 70% training set proportion, and the results are shown in the **Table 3**.

**Table 3. Results of Various Methods**

| Methods | ACC | AUC | Precision | Recall | MCC |
|---|---|---|---|---|---|
| LR | 0.907 | 0.541 | 0.407 | 0.562 | 0.354 |
| SVM | 0.906 | 0.689 | 0.731 | 0.398 | 0.566 |
| RF | 0.901 | 0.732 | 0.795 | 0.223 | 0.603 |
| CNN | 0.916 | 0.751 | **0.785** | 0.608 | 0.639 |
| LSTM | 0.912 | 0.743 | 0.706 | 0.590 | 0.692 |
| Model with only DSAE | 0.909 | 0.760 | 0.783 | 0.496 | 0.625 |
| Model with only MSSPAM | 0.918 | 0.804 | 0.732 | 0.589 | 0.707 |
| **Proposed Model** | **0.922** | **0.820** | 0.737 | **0.613** | **0.728** |

A comprehensive analysis of various machine learning models reveals that Logistic Regression (LR) and Support Vector Machine (SVM) are effective for linear problems but struggle with nonlinear features in complex datasets. Random Forest (RF), while resistant to overfitting and capable of handling large feature sets, may suffer from low recall in certain cases. Convolutional Neural Networks (CNN) and Long Short-Term Memory Networks (LSTM) excel in tasks involving images or sequential data, offering automatic feature extraction and long-term dependency capture, though they require more training time and computational resources. In contrast, the proposed model outperforms others across different training ratios, particularly in terms of AUC and MCC, demonstrating superior generalization, stability, and effectiveness in handling diverse datasets.

## 5. Conclusion

This study presents a multidimensional perception-based electricity anomaly detection network integrating the DSAE and MSSPAM modules to enhance anomaly detection accuracy in electricity usage data. Convolutional layers are used for initial data processing, extracting local features while preserving spatial structure. the MSSPAM module combines multi-scale feature aggregation and position encoding to capture global dependencies and local details. the DSAE module optimizes multi-channel features through adaptive weighting, improving flexibility in feature selection. Experimental results on the State Grid Corporation of China (SGCC) dataset show that the proposed model outperforms benchmark methods by approximately 5% in accuracy and offers improved computational efficiency and numerical stability.

The multidimensional perception-based electricity anomaly detection network offers an effective solution for power theft detection, demonstrating exceptional performance across various metrics. Its promising results highlight its potential for real-world applications. Future work will focus on optimizing the model structure and exploring broader application scenarios to further enhance detection efficiency and accuracy.

## References

[1] Smith T B. Electricity theft: a comparative analysis [J]. Energy policy, 2004, 32(18):2067-2076.
[2] Gaur V, Gupta E. the determinants of electricity theft: An empirical analysis of Indian states [J]. Energy Policy, 2016,

93:127-136.

[3] Golden M A, Min B. Theft and loss of electricity in an Indian state [M]. International Growth Centre, 2012.

[4] Friedman E, Lee C K. Remaking the world of Chinese labour: A 30-year retrospective [J]. British journal of industrial relations, 2010, 48(3):507-533

[5] Ahir R K, Chakraborty B. Pattern-based and context-aware electricity theft detection in smart grid [J]. Sustainable Energy, Grids and Networks, 2022, 32:100833.

[6] Lepolesa L J, Achari S, Cheng L. Electricity theft detection in smart grids based on deep neural network [J]. Ieee Access, 2022, 10:39638-39655.

[7] Xia X, Lin J, Jia Q, et al. ETD-ConvLSTM: A deep learning approach for electricity theft detection in smart grids [J]. IEEE Transactions on Information Forensics and Security, 2023, 18:2553-2568.

[8] Zheng K, Chen Q, Wang Y, et al. A novel combined data-driven approach for electricity theft detection [J]. IEEE Transactions on Industrial Informatics, 2018, 15(3):1809-1819.

[9] Munawar S, Javaid N, Khan Z A, et al. Electricity theft detection in smart grids using a hybrid BiGRU–BiLSTM model with feature engineering-based preprocessing [J]. Sensors, 2022, 22(20):7818.

[10] Blazakis K V, Kapetanakis T N, Stavrakakis G S. Effective electricity theft detection in power distribution grids using an adaptive neuro fuzzy inference system [J]. Energies, 2020, 13(12):3110.

[11] Akram R, Ayub N, Khan I, et al. Towards big data electricity theft detection based on improved rusboost classifiers in smart grid [J]. Energies, 2021, 14(23):8029.

[12] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020:11534-11542.

[13] Cai H, Li J, Hu M, et al. Efficientvit: Multi-scale linear attention for high-resolution dense prediction [J]. arXiv preprint arXiv:2205.14756, 2022.