International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 8383-8396

https://doi.org/10.70517/ijhsa464714

Time Series Analysis Model for Electricity Consumption Behavior Monitoring and Anti-Theft Electricity Research

Shuai Yang[1,*], Qiong Cao[1], Wei Zhang[1] and Hao Guo[1]
[1] State Grid Shanxi Marketing Service Center, Taiyuan, Shanxi, 030000, China
Corresponding authors: (e-mail: shuaiyang8080@163.com).

Abstract Electricity user behavior data is complex and diverse, resulting in significant variability and uncertainty in user behavior data, which increases the difficulty of monitoring electricity user behavior and leads to low monitoring rates. This paper utilizes the singular value equivalent matrix to obtain a non-Hermitian matrix and performs standardization processing on the aforementioned matrix. Considering the ARMA equation system for time series stationarity, the proposed numerical solution is used to calculate the expression, thereby extending RMT from a purely Gaussian environment to a non-Gaussian environment. An ETD-SAC electricity theft detection model framework is constructed to determine whether users are engaging in electricity theft during the detection period. Through user electricity consumption behavior detection, it was found that the electricity load trend of electricity theft users fluctuated between [8.54, 38.54] kWh after July 15, 2023. One of the suspected users detected bypassed the meter for electricity theft, with the meter current ranging from -0.1 to 0.4 A, while the actual incoming current was 0.6 to 1 A, constituting electricity theft behavior. Using the same method for electricity theft behavior analysis, CZ Factory was found to have engaged in electricity theft on October 1, 2023, requiring the recovery of 1,354 units of electricity and 1,126.528 yuan in electricity fees. The anti-electricity theft application model based on ARMA achieved good results.

Index Terms non-Hermitian matrix, ARMA model, ETD-SAC electricity theft detection model, electricity theft behavior analysis

I. Introduction

Electricity, as a critical strategic resource for economic development, plays an indispensable role in societal progress. However, in recent years, energy shortages have become increasingly frequent. As the unit cost of electricity has risen, electricity theft by users has intensified, disrupting the normal order of the electricity market economy and creating a significant economic black hole, resulting in substantial economic losses for the nation and electricity-related enterprises [1]-[3]. Additionally, electricity theft poses significant social hazards. Accidents caused by transformer and line failures due to electricity theft, resulting in power outages of varying scales, occur frequently, disrupting power supply and even leading to civil disputes, public order issues, fires, and criminal cases, thereby affecting social stability [4]-[6]. Currently, electricity theft by users is primarily detected through regular inspections and analysis of electricity consumption reports. This places high demands on the meticulousness and professionalism of power company staff, as well as requiring them to have extensive practical experience [7]-[9]. Additionally, this monitoring method relies heavily on periodic inspections of the power system, resulting in some waste of manpower and low monitoring efficiency [10]. With the development of smart meters and electricity meter data collection technology, the power industry has entered the digital age. Electricity consumption data from power users is now fully accessible and utilized, leading to progress in electricity consumption behavior monitoring and anti-theft efforts, thereby reducing instances of electricity theft [11]-[13].

Against the backdrop of rapidly increasing electricity consumption data, traditional electricity consumption behavior monitoring has exposed its shortcomings, such as rising false alarm rates and difficulty in detecting new forms of electricity theft, thereby increasing the challenges for anti-theft efforts [14], [15]. Time series analysis models are a method of modeling and analyzing time series data, playing a crucial role in analyzing data with temporal characteristics. They aim to reveal trends, seasonality, cycles, and irregular fluctuations in data, enabling prediction and decision-making. These models are widely applied in fields such as economics, finance, markets, meteorology, and medicine [16]-[18]. Electricity data exhibits distinct temporal characteristics such as multiple periodicity, nonlinear fluctuations, and spatial correlation. By applying intelligent algorithms based on time series analysis models to monitor user electricity consumption behavior, analyzing the collected electricity consumption information can identify electricity theft targets and better protect the interests of power supply companies

[19]-[21].

In studies related to electricity consumption and theft, relevant time series analysis models are primarily used for electricity consumption prediction and electricity theft detection. Literature [22] points out that time series econometric models and machine learning can effectively extract statistical features of power consumption, with the autoregressive partial differenced moving average model yielding better results in estimating power consumption. Literature [23] compared the long short-term memory (LSTM) model and the seasonal autoregressive integrated moving average (SARIMA) model in predicting power consumption, with the LSTM model achieving a lower error rate. Literature [24] uses power consumption data sensed by different sensors at the terminal unit and the user terminal unit as the basis, employing an autoregressive model for feature extraction, analysis, and prediction to identify meter-bypassing power theft behavior. However, when the sensor data received at the user terminal unit is 0, the coefficients of the autoregressive model exhibit discrepancies. Literature [25] utilized LSTM and fuzzy inference systems to detect and confirm power consumption data from different types of users, thereby accurately identifying power theft behavior. Literature [26] integrated the Shaoxing swarm algorithm, gated convolutional autoencoder, cost-sensitive learning, and LSTM to construct a power theft detection model, which achieved a power theft detection accuracy rate of 99.45% under real-time power consumption data. In Reference [27], electricity consumption data balancing, feature extraction, and electricity theft classification were performed using electricity theft attacks, LSTM, and gated recurrent units (GRU), respectively, to create an electricity theft detection strategy. Reference [28] proposed an electricity theft detection model based on bidirectional gated recurrent units (GRU) and bidirectional LSTM, using K-means minority class oversampling technology to address the electricity data imbalance issue, combined with random feature engineering for preprocessing, to improve the classification accuracy of whether users engage in electricity theft. Reference [29] established an anti-electricity theft diagnosis method dominated by the Long Short-Term Memory (LSTM) network model. LSTM is primarily used to screen the features of users identified by electricity theft detection devices as engaging in electricity theft or abnormal electricity consumption behavior, and the results are reported to the main station.

This paper selects multiple nodes as samples of user electricity consumption, performs discrimination processing based on the status of each node, standardizes the original matrix, and forms a non-Hermitian matrix. Using the singular value equivalent matrix, the non-Hermitian matrix is solved to obtain the matrix standardization results. RMT is extended from a purely Gaussian environment to a non-Gaussian environment. Considering the global characteristics of the electricity theft detection algorithm, we identified features with high computational complexity and critical importance. We proposed a time series stationary ARMA model and performed data preprocessing, sequence decomposition, sequence embedding, feature extraction, and classifier classification operations to determine whether users engaged in electricity theft during the detection period. We designed simulation experiments to detect user electricity load behavior and conducted anti-theft application tests using the ARMA-based user electricity feature extraction algorithm.

II. Detection of electricity consumption behavior based on the ARMA time series model

II. A. Data processing

II. A. 1) Acquisition of raw data

Select the electricity consumption of n nodes as samples, and select k state variables for each node for discrimination, resulting in a total of $N = n \times k$ variables. For each sampling time t_i, the collected data can be organized into a column vector:

$$x(t_i) = \left[x_1(t_i), x_2(t_i), \cdots, x_N(t_i)\right]^{\mathrm{T}} \tag{1}$$

Due to the continuous expansion of the sampling time, the N column vectors are expressed in matrix form, and the sliding time window method is selected to collect samples. The window width is set to T, that is, the historical data collected is used to collect the power data at time i, which together form a random matrix and serve as the data source for analysis:

$$X_{N\times T} = \left[x(t_{i-T+1}), \cdots, x(t_i)\right] \tag{2}$$

For any $N \times T$ original matrix $\hat{X}$, perform normalization processing:

$$\tilde{x}_{i,j} = (\tilde{x}_{i,j} - \overline{x}_i) \times \frac{\sigma(\tilde{x}_i)}{\sigma(\hat{x}_i)} + \overline{\tilde{x}_i} \tag{3}$$

In the formula: $\hat{x}_i = (\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_T)$, $\sigma(\hat{x}_i)$ is the standard deviation of $\hat{x}_i$, $\overline{x}_i$ is the mean of $\hat{x}_i$, $\sigma(\tilde{x}_i)$ is the standard deviation of $\tilde{x}_i$ and $\sigma(\tilde{x}_i) = 1$, $\overline{\tilde{x}}_i$ is the mean of $\tilde{x}_i$ and $\overline{\tilde{x}}_i = 0, 1 \le i \le N, 1 \le j \le T$.

The matrix obtained through the above processing becomes an $N \times T$ non-Hermitian matrix $\tilde{x} \in C^{N \times T}$.

II. A. 2) Solving the singular value equivalent matrix

For non-Hermitian matrices, the singular value equivalent matrix of $\tilde{X}$ can be obtained from the following formula:

$$X_u = \sqrt{\tilde{X}\tilde{X}^*}U \tag{4}$$

In the formula: U is a unitary matrix that conforms to the Haar distribution.

Then we can see that:

$$X_u X_u^* = \tilde{X}\tilde{X}^* \tag{5}$$

When there are L matrices $\hat{X}_i (i = 1, 2, \cdots, L)$, perform random matrix analysis on them to obtain L processed independent non-Hermitian matrices, where:

$$z = \sum_{i=1}^{L} X_{u,i} \in C^{n \times n} \tag{6}$$

II. A. 3) Matrix Standardization

The above matrix needs to be standardized to obtain the matrix $\overline{z}$:

$$\tilde{z}_i = z_i / (\sqrt{N}\sigma(z_i)) \tag{7}$$

In the formula: $z_i = (z_{i1}, z_{i2}, \cdots, z_{iN})$, $\tilde{z}_i = (\tilde{z}_{i1}, \tilde{z}_{i2}, \cdots, \tilde{z}_{iN})$, and $\sigma(z_i)$ is the standard deviation of the matrix z_i.

II. A. 4) Limit spectral distribution of ARMA models for time series

First, consider the stationary ARMA(p, q) equation for time series [30]:

$$\phi(B)y_t = \theta(B)\varepsilon_t \tag{8}$$

In the equation: $\{y_t : t = 0, \pm 1, ...\}$ is a sequence of real variables, $\{\varepsilon_t : t = 0, \pm 1, ...\}$ is a white noise vector obeying $N(0, \sigma^2)$, and B is a delay operator.

Let X_i be N independent copies of $y = (y_1, y_2, ..., y_T)$:

$$X_i = (X_{1i}, X_{2i}, \cdots, X_{Ti}), i \in [1, N] \tag{9}$$

$$X = (X_1, X_2, \cdots, X_N)^T \tag{10}$$

If $c = N / T \in (0,1]$, then the empirical spectrum corresponding to the covariance matrix $S = 1/N(XX^H)$ tends toward the Stieltjes transform probability distribution F:

$$z = -\frac{1}{s} + \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{cs + \{2\pi f(\lambda)\}} d\lambda \tag{11}$$

In the formula: $s_F(z) = \int \frac{1}{x - z} F(dx), z \in C^+$, $f(\lambda)$ is the spectral density of the ARMA(p, q) model:

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|\frac{\theta(e^{-i\lambda})}{\phi(e^{-i\lambda})}\right|^2, \lambda \in [0, 2\pi) \tag{12}$$

For more complex situations, a numerical solution is proposed to calculate the expression of F as follows:

$$s = \frac{1}{-z + A(s(z))} \tag{13}$$

In the equation: $A(s(z)) = \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{cx + \{2\pi f(\lambda)\}^{-1}} d\lambda$.

Let ε be a sufficiently small positive value, and set $z = x + i\varepsilon$. Choose the initial value $s_0(z) = u + i\varepsilon$. Iterate according to the iterative equation $k \geq 0$:

$$s_{k+1}(z) = \{-z + A(s_k(z))\}^{-1} \tag{14}$$

Until $s_k(z)$ converges, then the density function $f_T(x)$:

$$f_T(x) = \frac{1}{\pi} I s_k(z) \tag{15}$$

Through the above processing, RMT is extended from a purely Gaussian environment to a non-Gaussian environment.

II. B.Electricity theft detection model

II. B. 1) Model Framework

Figure 1 shows the ETD-SAC electricity theft detection model framework. Addressing the issues of high computational complexity in global feature extraction and the tendency for key features to be obscured by complex redundant global features in existing electricity theft detection algorithms, this chapter proposes an electricity theft detection model based on a sequence-level connection autocorrelation mechanism [31]. One of the advantages of the ETD-SAC model is its ability to simultaneously capture seasonal features with similar sub-sequences in power data, trend-based features after sequence decomposition, and local similarity features in power data.

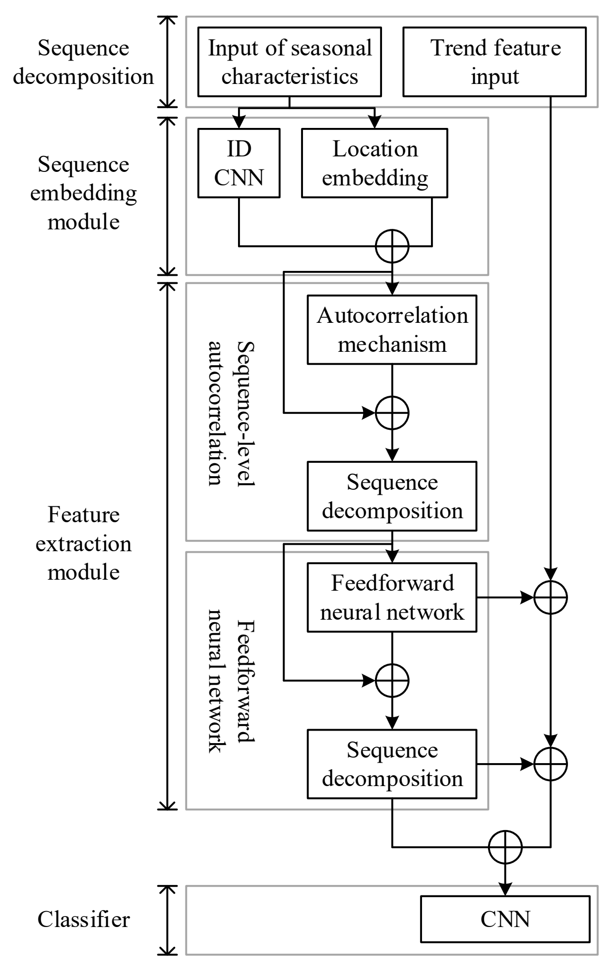


Figure 1: Framework of ETD-SAC electricity Theft Detection Model

II. B. 2) Data preprocessing

(1) Missing value handling

User electricity consumption data often contains missing values. This is mainly caused by various factors, such as smart meter malfunctions, unreliable transmission of measurement data, unscheduled system maintenance, and storage issues. Therefore, in this chapter, we use linear interpolation methods to fill in missing values based on the following equation:

$$q_i = \begin{cases} \frac{q_{i-1}+q_{i+1}}{2} & q_i \in NaN, q_{i-1}, q_{i+1} \notin NaN \\ 0 & q_i \in NaN, q_{i-1} or q_{i+1} \in NaN \\ q_i & q_i \notin NaN \end{cases} \tag{16}$$

where q_i represents the value of the power data at a certain time. If the value is an empty character or a non-numeric character, it is represented as NaN. When the power data at a certain time is an empty character, the average value of the adjacent time is used to update the current time's power consumption value. If the value of the adjacent time is also NaN, the current time's power consumption value is represented as 0.

This chapter restores outliers based on the “three sigma rule” using the following equation:

$$\begin{cases} avg(q)+3\cdot std(q) & if\ q_i > avg(q)+3\cdot std(q) \\ q_i & otherwise \end{cases} \tag{17}$$

In this context, $avg(q)$ represents the average value of the sequence q, while $std(q)$ denotes the standard deviation of the sequence q. This method effectively reduces outliers in the data.

(3) Data normalization

Before training the model, the sample data needs to be normalized. After interpolating the data and removing outliers, the MAX-MN normalization method is selected. For each training sample $X_k = q_{T,T,}$, normalization is performed according to the following equation:

$$f(q_i) = \frac{q_i - \min(X_k)}{\max(X_k) - \min(X_k)} \tag{18}$$

where $\min(X_k)$ denotes the minimum value of the sample sequence X_k, $\max(X_k)$ denotes the maximum value of the sample sequence X_k, and $i \in [T_s : T_c]$.

II. B. 3) Sequence decomposition

This paper uses the average pooling process $AvgPool(\cdot)$ with the same padding strategy to implement sliding average, from which the trend features after sequence decomposition are obtained. Then, the power data is subtracted from the trend features to obtain the periodic features. The specific formula is as follows:

$$X_t = AvgPool(Padding(X)) \tag{19}$$

$$X_s = X - X_t \tag{20}$$

Among these, $X \in R^{L\times d}$ is the input data of historical electricity consumption samples of length L, $X_s \in R^{L\times d}$ is the seasonal feature obtained from sequence decomposition, and $X_t \in R^{L\times d}$ is the trend feature obtained from sequence decomposition. The convolution kernel size in the pooling layer is 25, and the stride is 1.

II. B. 4) Sequence Embedding

For complex time series feature extraction, low-dimensional data needs to be elevated to amplify features. In order to output seasonal features $X_s \in R^{L\times d}$ to the defined d_{model} dimension and convert them into encoded vectors, the model uses convolutional layers to embed features. Specifically, this is achieved using one-dimensional convolutional layers:

$$X_{conv} = W_{conv} * X_s + B_{conv} \tag{21}$$

where $X_{conv} \in R^{L \times d_{m \times dil}}$, W_{conv}, B_{conv} are the output, convolution kernel parameters, convolution layer input, and bias term of the convolution layer, respectively. In the convolution operation, the convolution kernel size is 3×3, and the stride is 1. To prevent gradient vanishing during training, the ReLU linear function is used. Additionally, the "same padding strategy" is applied.

In order to enable the model to fully utilize the sequential information of electricity data, the ETD-SAC model adds time-location encoding to the embedding layer. There are several options for location embedding. Referring to the transformer, sine and cosine functions of different frequencies are used:

$$PE_{(pos,2i)} = \sin(pos / 10000^{2i/d_{modol}}) \tag{22}$$

$$PE_{(pos,2i+1)} = \cos(pos / 10000^{2i/d_{modol}}) \tag{23}$$

In this context, pos represents the position in the time series, i denotes the length of the sine and cosine functions, and d_{model} refers to the dimension of the position encoding. The choice of sine and cosine functions to represent position encoding stems from the fact that the temporal characteristics of power data exhibit periodicity similar to that of sine and cosine functions. This makes it convenient to use sine and cosine functions to represent relative positions within power data. Additionally, the sine function allows for variable lengths of the input sequence. Both sine and cosine functions are used because only when both are used together can PE_{pos+k} be represented as a linear function of PE_{pos}. This allows each point in the position encoding sequence to be represented by a point in the sine function, achieving the position embedding effect.

The sequence embedding layer then adds the output of the one-dimensional convolution to the output of the position embedding:

$$X_{ous} = X_{conv} + PE \tag{24}$$

II. B. 5) Feature extraction

Assume there are Layer sequence-level autocorrelation mechanism layers. The overall equation for the l th sequence-level autocorrelation mechanism layer can be summarized as $X_{autocor}^{l} = SeriesAutoCorrelation(X_{autocor}^{l-1})$. The formula is as follows:

$$S_{autocor}^{l}, \tau_{autocor}^{l} = SeriesDecomp(Auto-Correlation(X_{autocor}^{l-1}) + X_{autocor}^{l-1}) \tag{25}$$

Among these, $X_{autocor}^{l} = S_{autocor}^{l}, l \in \{1, \cdots, L\}$ denotes the output of the l th sequence-level autocorrelation mechanism layer, and $X_{autocor}$ denotes the embedding of X_{ems}, which is used to extract periodic features. $S_{autocor}^{l}$ denotes the seasonal features decomposed by the sequence decomposition module in the l th layer. $\tau_{autocor}^{l}$ denotes the trend features decomposed by the sequence decomposition module in the l th layer. Additionally, a residual network is applied in each autocorrelation mechanism to mitigate network degradation issues.

As shown in the autocorrelation mechanism flow in Figure 2, under the single-head autocorrelation mechanism, for a seasonal embedding sequence X_{ems} with a time length of L that has undergone feature embedding. First, the embedded sequence X_{ems} is mapped to three vectors $query\ Q$, $key\ K$ and $value\ V$. Then, the similarity between the delayed sequences of $query\ Q$ and $key\ K$ is calculated, and the $Topk$ positions with the highest similarity as the time delay starting points. Time delay operations are then performed on $value\ V$ at the $Topk$ starting points to obtain a new time delay sequence. The similarity values are then weighted and summed with the time-delayed $value\ V$ to obtain the output vector of the autocorrelation mechanism. This allows the attention mechanism to be replaced by the following formula. The specific autocorrelation mechanism formula is:

$$\tau_1, \ldots, \tau_k = \underset{\tau \in [1, \cdots, L]}{argTovk} (R_{Q,K}(\tau_k)) \tag{26}$$

$$\hat{R}_{Q,K}(\tau_1), \cdots, \hat{R}_{Q,K}(\tau_k) = SoftMax(R_{Q,K}(\tau_1), \cdots, R_{Q,K}(\tau_k)) \tag{27}$$

$$Auto-Correlation(Q, K, V) = \sum_{i=1}^{k} Roll(V, \tau_i) \hat{R}_{Q,K}(\tau_i) \tag{28}$$

Among them, τ represents the starting point of time delay. $argTopk(\cdot)$ is used to obtain the starting points of similar subsequences in the time series $query\ Q$ and the time series $key\ K$ to obtain $Topk$ time series, and let $k=[c\times\log L]$, where c is a hyperparameter. $R_{Q,K}$ is the autocorrelation between the sequence $query\ Q$ and the sequence $key\ K$. $Roll(X,\tau)$ denotes the sequence rollback operation on $value\ V$ starting from the time delay τ. During sequence rollback, elements beyond the first position are reintroduced into the last position to construct a time-delayed sequence with similar seasonal characteristics to the original time series within the cycle.

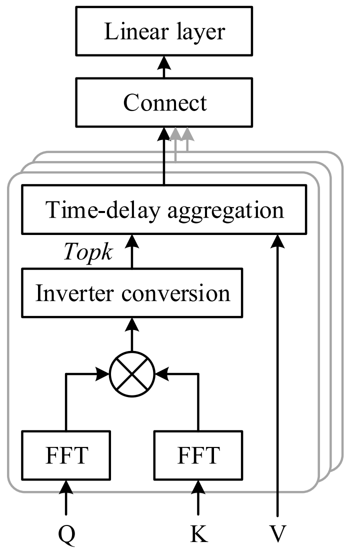


Figure 2: Autocorrelation mechanism

Analysis of user electricity data reveals that electricity data exhibits daily periodicity, with similar sub-sequences occurring between phases of the same period. For a power data set X_t with seasonal similarity, the correlation $R_{xx}(\tau)$ between different delay sequences within the sequence can be calculated using the following equation:

$$R_{xx}(\tau)=\lim_{L\to\infty}\frac{1}{L}\sum_{t=1}^{L}x_t x_{t-\tau} \tag{29}$$

$R_{\alpha}(\tau)$ reflects the time-delay similarity between the sequence X_t and its delayed sequence $X_{t-\tau}$ starting at time τ.

II. B. 6) Classifier

After feature extraction in the previous stage, model validation is performed in this stage. Specifically, the features extracted in the previous subsection are used as input, and a binary classification neural network layer is used to calculate the degree of abnormality in the electricity consumption curve. The binary classification neural network layer consists of two consecutive 1D-CNN layers, with the formula as follows:

$$Y_{out}=\sigma(W_2(W_1Y+b_1)+b_2) \tag{30}$$

W_1,W_2 are the learnable parameters of the one-dimensional convolution layer, b_1,b_2 are the bias terms of the one-dimensional convolution layer, and $\sigma(\cdot)$ is the sigmoid activation function. The final output of the binary classification neural network layer is a single neuron, meaning that the values of the elements in Y_{out} are output to the interval (0, 1) via the sigmoid activation function, containing only a single element value. Then, the output value Y_{out} is compared with the set threshold for the degree of abnormality in the electricity consumption curve to determine whether the user engaged in electricity theft during the detection period.

III. Electricity usage behavior detection and anti-electricity theft application analysis

III. A. Detection and analysis of user electricity consumption behavior

III. A. 1) User electricity load

Trend indicators are used to reflect fluctuations in users' electricity consumption data. For this purpose, this paper randomly selected one honest user and one electricity thief, and plotted their continuous 10-month electricity consumption trends as shown in Figure 3 below.

It can be observed that the electricity load of users forms a continuous time series with significant fluctuations. The electricity load trends of honest and electricity-stealing users are inconsistent. The electricity load trend of honest users fluctuates between 4.38 kW•h and h to 49.28 Kw•h, while the electricity consumption trend of electricity-stealing users fluctuated between [8.54, 38.54] Kw•h. That is, the volatility of electricity consumption patterns may differ between different types of users. To better capture the volatility patterns of load time series, this paper employs a sliding window method to extract trend-related feature indicators from the load time series of different user types. The implementation is carried out using the Python software platform.

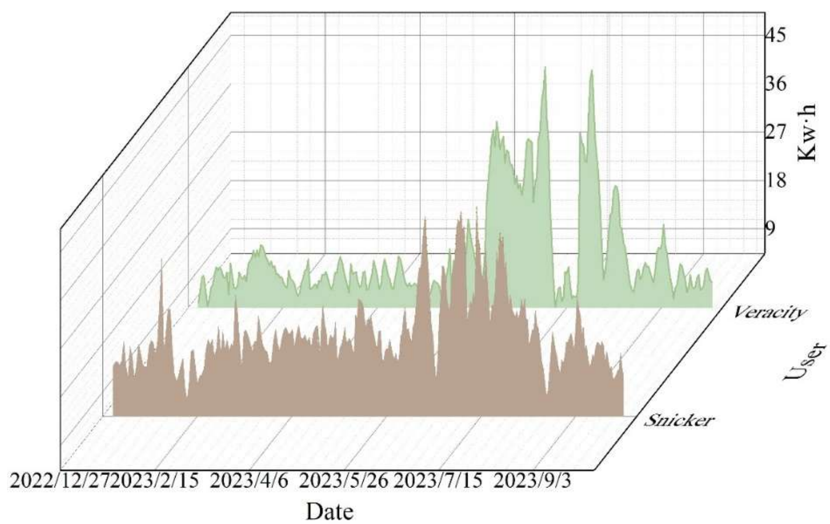


Figure 3: General trend diagram of user's electricity load

III. A. 2) Distribution of electricity load characteristic dimension data

Data standardization (normalization) processing is a foundational task in data analysis. To eliminate the influence of different units of measurement among features, data must first undergo standardization processing. Data standardization involves scaling the data proportionally to fit it into a smaller, specific range, thereby transforming it into dimensionless numerical data. After processing, data from different units and scales can be compared and evaluated comprehensively. Figure 4 shows the distribution of two-dimensional feature data. It can be seen that the actual user electricity consumption feature 1 is distributed in the range [-3.05, 14.12], and feature 2 is distributed in the range [-3.64, 7.47].

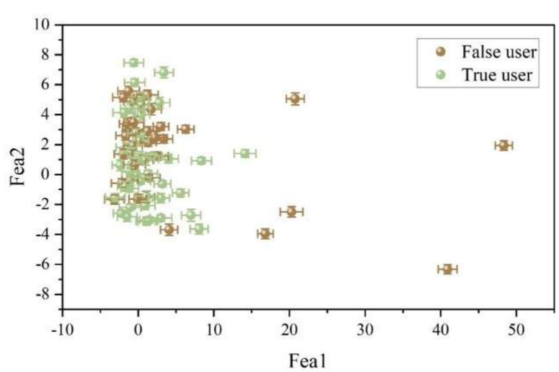


Figure 4: Two dimensional eigendata distribution

III. A. 3) User Behavior Detection Results

Figure 5 shows the accuracy of the user anomaly detection model. The classification accuracy of the user behavior gray list generated by the ARMA model at different detection rates is compared with the classification accuracy of the gray list generated by the unsupervised detection model at different detection rates. The horizontal axis represents the detection rate, i.e., the number of gray list users detected, while the vertical axis represents the accuracy of the detection results.

In the experiments, the trends of the curves were largely similar across the three different datasets. As shown in the figure, the ARMA model consistently outperformed the unsupervised detection model throughout the entire detection rate improvement process. In other words, under the same detection rate, the detection model based on the ARMA model achieved higher accuracy throughout the detection process compared to the detection model based solely on unsupervised learning. The accuracy of the ARMA-based detection model reaches its peak at a detection rate of approximately 30%-40%, achieving an accuracy rate of 84.85%. Such a high accuracy rate holds significant value for on-site detection.

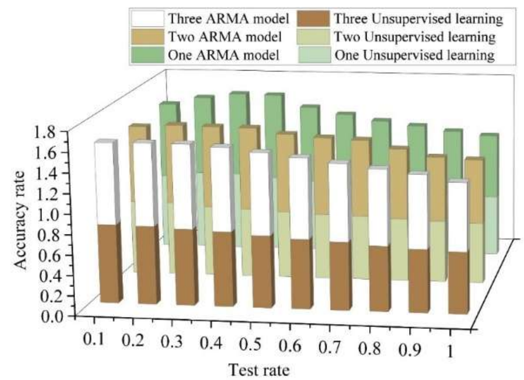


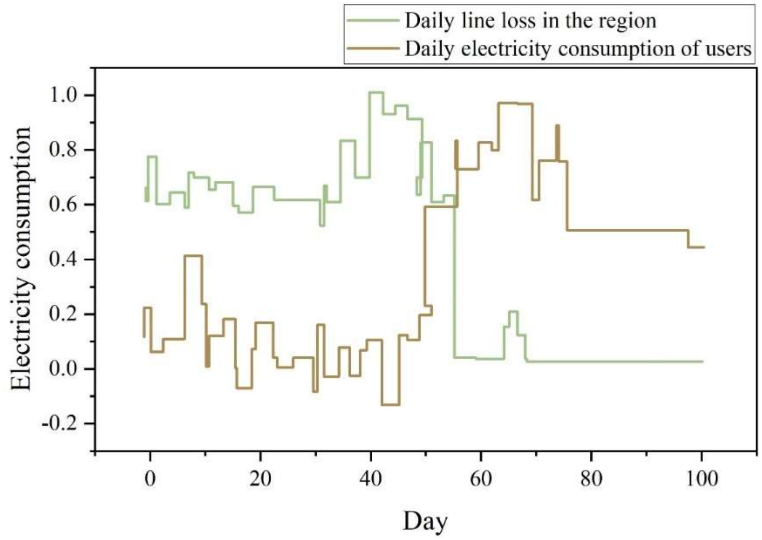
Figure 5: User anomaly detection model accuracy

III. B. Detection of electricity theft

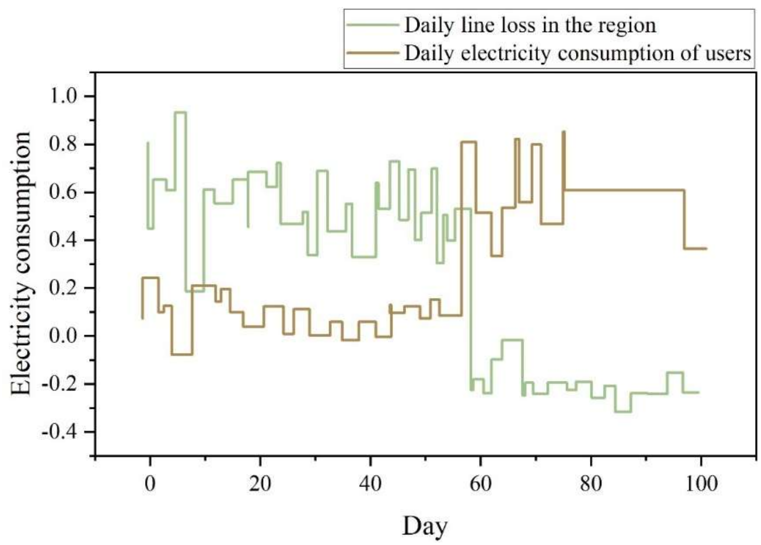
III. B. 1) Model Results

The model was selected for validation testing in a certain county in CZ. Analysis was conducted on certain characteristics of some suspicious users. Based on the output results, a ranking of suspicion levels was established. Users with higher suspicion levels were selected, and their electricity consumption data was plotted in a graph. Figure 6 shows the correlation between user transformer district line loss and daily electricity consumption. Figure (a) represents User 1, and Figure (b) represents User 2.

In May 2023, the suspected users corresponding to Figures (a) and (b) were selected for on-site inspections in a county in CZ. Ultimately, both users were confirmed as electricity thieves. One user engaged in bypassing the meter to steal electricity, with the meter reading a current of -0.1 to 0.4 A, while the actual incoming current was 0.6 to 1 A, constituting electricity theft. The other user tampered with the meter to steal electricity, causing inaccurate meter readings to achieve the purpose of electricity theft.



(a)User 1



(b)User 2

Figure 6: The correlation between line loss and daily electricity in the user platform area

III. B. 2) Electricity theft detection results for different models

Figure 7 compares the results of different models in identifying electricity theft types with actual on-site inspections. It can be seen that the ARMA model proposed in this study provides anti-theft results that are closer to the actual on-site inspection values. For the three types of electricity theft—undervoltage theft, altering the wiring method of high-voltage metering devices, and neutral point creation method theft—the diagnostic results align with the actual inspection results, with the number of diagnostic results being 16, 12, and 12, respectively. In contrast, traditional anti-theft models show greater discrepancies with actual on-site inspection data, with some even exceeding the actual on-site inspection results. This demonstrates the superiority of the ARMA-based anti-theft model proposed in this study for diagnosing electricity theft types.

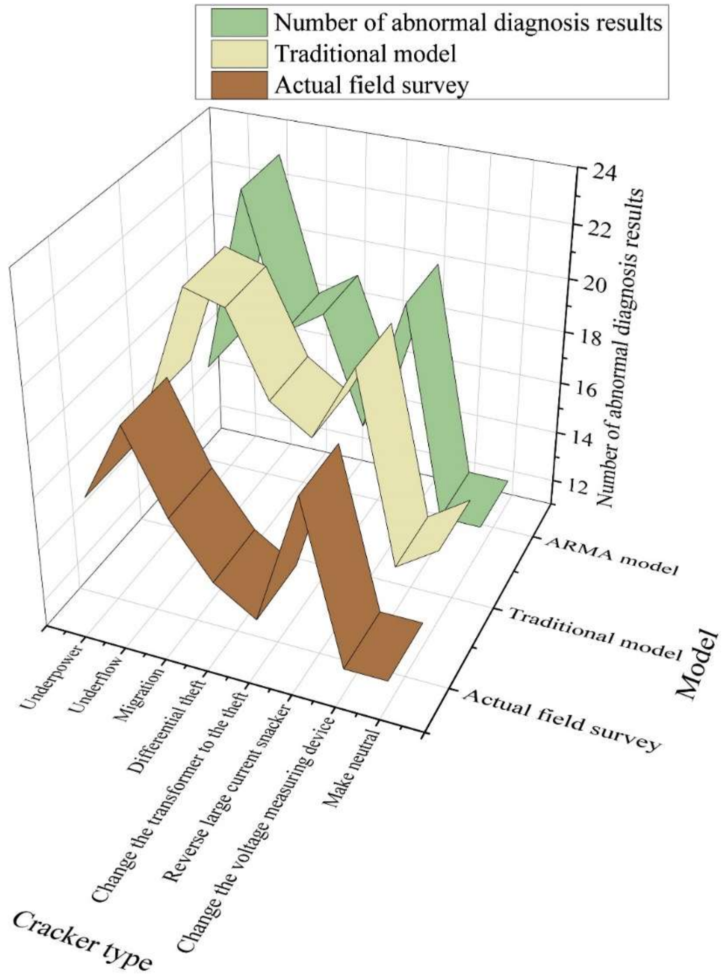


Figure 7: Different models diagnose the type of cracker and the actual field survey

III. C. Application Verification

III. C. 1) Treatment Methods

Abnormal user electricity consumption data primarily includes abnormal electricity consumption, current, and voltage anomalies. By analyzing user data, it is possible to distinguish between normal and abnormal user electricity consumption characteristics and use this to predict future user electricity consumption. Experimental testing analyzed the data using an ARMA-based user electricity consumption feature extraction algorithm to extract user electricity consumption features. The algorithm is then used for electricity theft diagnosis. By analyzing daily data for electricity theft and unauthorized usage, the algorithm identifies users suspected of electricity theft. Further confirmation is required, and the following procedures should be followed:

(1) Anti-theft verification

On-site inspection to identify related electricity theft devices or damaged power equipment. For example, photographing the electricity theft scene and identifying the electricity theft user.

(2) Estimating the time of electricity theft

Carefully understand the actual electricity usage situation, compare and analyze monthly electricity consumption over the past few years to identify suspicious points, and do not rely on the electricity thief's own statements to determine the time of electricity theft.

(3) Determining the amount of stolen electricity

1) The amount of electricity stolen by illegally connecting to the power supply facilities of the power supply company is calculated by multiplying the rated capacity of the illegally connected equipment (kVA is considered kW) by the actual electricity usage time.

2) For electricity theft through other means, the stolen electricity volume is calculated by multiplying the rated current value of the billing electricity meter (for meters with current limiters, the current value set by the limiter) by the capacity (kVA is equivalent to kW) and the actual electricity theft duration.

3) When the duration of electricity theft cannot be determined, the number of days of theft shall be calculated based on a minimum of 180 days. The daily theft duration shall be calculated as 12 hours for power users and 6 hours for lighting users.

(4) Electricity theft must be dealt with quickly and accurately.

Once electricity theft is confirmed, the power supply shall be cut off immediately in accordance with relevant regulations. While protecting the scene, the type of electricity theft shall be determined and the “Work Order for Handling Electricity Breach and Electricity Theft” shall be filled out in a timely manner and signed by the person responsible for the electricity theft.

III. C. 2) Test Cases

Using the same analytical method, we once again identified electricity theft at a woodworking factory in a certain city within our jurisdiction, as shown in Tables 1–3:

An on-site inspection found that the user had privately modified the electricity meter to reduce its readings. According to Article 100 of the “Power Supply Business Rules,” CZ factory was required to make up for 1,354 units of electricity and 1,126.528 yuan in electricity fees for the electricity theft that occurred on October 1, 2023. The algorithm proposed in this paper can effectively detect abnormal electricity usage by users. When electricity usage fluctuations exceed 15%, further analysis of relevant indicator data can be conducted to assess the extent of electricity theft. When the probability of electricity theft reaches the threshold, personnel are dispatched for on-site inspection. Through statistical analysis of actual data, power supply personnel found that the probability of users engaging in electricity theft has increased compared to previous periods, achieving good results.

Table 1: User electricity theft test case

Customer Number	User name	Recharge power	Make up for the electricity bill	Filing mark	Occurrence time	Nature
6900000496	Cz factory	1354	1126.528	NO	2023/10/1	Electricity theft

Table 2: Annual electricity bill for users

Electricity bill year and month	Battery power	Electricity bill	Type of electricity charge	Electricity bill frequency	Planning status"	Current status Date
2023/12/1	1189	989.248	Normal electricity bill	1	Issue	2023/12/16
2023/11/1	1248	1038.336	Normal electricity bill	1	Issue	2023/11/17
2023/10/1	**1354**	**1126.528**	**Compensation for illegal theft charges**	**1**	Issue	**2023/10/1**
2023/10/1	679	564.928	Normal electricity bill	1	Issue	2023/10/1
2023/9/1	1063	884.416	Normal electricity bill	1	Issue	2023/9/1
2023/8/1	2348	1953.536	Normal electricity bill	1	Issue	2023/8/1
2023/7/1	1293	1075.776	Normal electricity bill	1	Issue	2023/7/1
2023/6/1	1452	1208.064	Normal electricity bill	1	Issue	2023/6/1
2023/5/1	449	373.568	Normal electricity bill	1	Issue	2023/5/1
2023/5/1	900	748.8	Normal electricity bill	1	Issue	2023/5/1
2023/4/1	1165	969.28	Normal electricity bill	1	Issue	2023/4/1
2023/3/1	963	801.216	Normal electricity bill	1	Issue	2023/3/1
2023/2/1	1936	1610.752	Normal electricity bill	1	Issue	2023/2/1
2023/1/1	1262	1049.984	Normal electricity bill	1	Issue	2023/1/1

Table 3: The base figure of the user's annual

Reading type	Combined ratio	Last number	This number	Transcribe	Submeter	Check for residual electricity	Last time I checked the current amount of electricity	Sheet time
Work (total)	100	657.49	772.46	6185	0	1765	1132	2023/12/11
Active (peak)	100	253.18	279.86	2486	0	1438	4168	2023/12/11
Work (valley)	100	130.96	145.35	1342	0	1352	2542	2023/12/11
No work (total)	100	494.65	498.18	91	0	0	7408	2023/12/11
The maximum quantity is work (total)	100	0	0	0	0	0	38	2023/12/11
Work (total)	15	731.63	800.46	1642	0	1168	894	2023/11/14
No work (total)	15	945.31	981.36	0	0	0	5348	2023/11/15
Work (total)	15	694.08	730.49	595	0	596	1065	2023/10/18
No work (total)	15	969.44	985.23	0	0	0	8869	2023/10/19

IV. Conclusion

This paper uses an ARMA time series model to obtain user electricity consumption data, construct a Hermitian matrix, and standardize the matrix, thereby extending RMT from a purely Gaussian environment to a non-Gaussian environment. Based on the extracted user electricity consumption behavior features, this paper proposes an ETD-SAC electricity theft detection model based on a time series-level connection autocorrelation mechanism.

Through a combination of simulation and empirical verification, the effectiveness of electricity consumption behavior detection and anti-theft applications is validated. The ARMA model constructed in this paper achieves a consistently higher accuracy rate than unsupervised learning detection models in user electricity consumption behavior detection. The accuracy rate of the ARMA detection model reaches its peak at a detection rate of approximately 30%-40%, achieving an accuracy rate of 84.85%, which holds significant value for anti-theft behavior detection.

The ARMA model's results are closer to actual on-site measurements in anti-theft applications. The diagnostic results for the three types of electricity theft—undervoltage theft, altering high-voltage metering device wiring methods, and neutral point creation methods—remain consistent, with diagnostic result counts of 16, 12, and 12, respectively.

In anti-theft electricity applications, the model detected an electricity theft incident at a certain factory in CZ on October 1, 2023, requiring the recovery of 1,354 units of electricity and 1,126.528 yuan in electricity fees. After using the model for anti-theft electricity detection on users, positive results were achieved.

References

[1] Arango, L. G., Deccache, E., Bonatto, B. D., Arango, H., & Pamplona, E. O. (2017). Study of electricity theft impact on the economy of a regulated electricity company. Journal of Control, Automation and Electrical Systems, 28, 567-575.

[2] Stevenson, E. S., Cook, J., & Bihl, T. (2025, March). Understanding Electricity Theft: Causes, Consequences, and AI-Based Detection. In 2025 ASEE North Central Section (NCS) Annual Conference.

[3] Jamil, F. (2018). Electricity theft among residential consumers in Rawalpindi and Islamabad. Energy Policy, 123, 147-154.

[4] Osadebamwen, K. O. (2023). ELECTRICITY THEFT AND POWER SUPPLY IN EDO STATE, NIGERIA: A CORRELATIONAL ANALYSIS. International Journal of Functional Research in Science and Engineering (IJFRSE), 1(2).

[5] Rohalin, S. (2023). Criminal Characteristics of Criminal Offenses Related to Electricity Theft. Krakowskie Studia Małopolskie, 4(40), 188-210.

[6] Onat, N. (2018). Electricity Theft Problem and Effects of Privatization Policies on Distribution Losses of Turkey. Celal Bayar University Journal of Science, 14(2), 163-176.

[7] Ali, S., Yongzhi, M., & Ali, W. (2023). Prevention and detection of electricity theft of distribution network. Sustainability, 15(6), 4868.

[8] Yan, Z., & Wen, H. (2021). Performance analysis of electricity theft detection for the smart grid: An overview. IEEE Transactions on Instrumentation and Measurement, 71, 1-28.

[9] Zhou, W., Li, B., Xiao, H., Xiao, H., Wang, W., Zheng, Y., & Su, S. (2023). Electricity theft detection of residential users with correlation of water and electricity usage. IEEE Transactions on Industrial Informatics, 20(4), 5339-5351.
[10] Wan, Z., Li, J., & Gao, Y. (2018). Monitoring and diagnosis process of abnormal consumption on smart power grid. Neural Computing and Applications, 30, 21-28.
[11] Nadeem, A., & Arshad, N. (2021). A data-driven approach to reduce electricity theft in developing countries. Utilities Policy, 73, 101304.
[12] Xia, X., Xiao, Y., Liang, W., & Cui, J. (2022). Detection methods in smart meters for electricity thefts: A survey. Proceedings of the IEEE, 110(2), 273-319.
[13] Ndulaka, J. S., Nwele, E. O., Olubiwe, M., & Akwukwaegbu, I. O. (2025). Anti-Electricity Theft Model for Sustainable Economic Development in Nigeria. International Journal of Electrical, Energy and Power System Engineering, 8(2), 128-141.
[14] Susanti, L., Fatrias, D., Ichwana, D., Kamil, H., & Putri, M. V. (2018, October). A configuration system for real-time monitoring and controlling electricity consumption behavior. In 2018 International Conference on Information Technology Systems and Innovation (ICITSI) (pp. 442-447). IEEE.
[15] Ibrahem, M. I., Nabil, M., Fouda, M. M., Mahmoud, M. M., Alasmary, W., & Alsolami, F. (2020). Efficient privacy-preserving electricity theft detection with dynamic billing and load monitoring for AMI networks. IEEE Internet of things journal, 8(2), 1243-1258.
[16] Pirkis, J., John, A., Shin, S., DelPozo-Banos, M., Arya, V., Analuisa-Aguilar, P., ... & Spittal, M. J. (2021). Suicide trends in the early months of the COVID-19 pandemic: an interrupted time-series analysis of preliminary data from 21 countries. The Lancet Psychiatry, 8(7), 579-588.
[17] Jiang, P., Dong, Q., Li, P., & Lian, L. (2017). A novel high-order weighted fuzzy time series model and its application in nonlinear time series prediction. Applied Soft Computing, 55, 44-62.
[18] Alharbi, F. R., & Csala, D. (2022). A seasonal autoregressive integrated moving average with exogenous factors (SARIMAX) forecasting model-based time series approach. Inventions, 7(4), 94.
[19] Liu, X., Guo, W., Feng, Q., & Wang, P. (2022). Spatial correlation, driving factors and dynamic spatial spillover of electricity consumption in China: A perspective on industry heterogeneity. Energy, 257, 124756.
[20] Díaz Redondo, R. P., Fernández Vilas, A., & Estévez Caldas, A. (2020). Monitoring Electricity Consumption based on Time Series Analysis. In Intelligent Environments 2020 (pp. 321-330). IOS Press.
[21] Maki, S., Ashina, S., Fujii, M., Fujita, T., Yabe, N., Uchida, K., ... & Chandran, R. (2018). Employing electricity-consumption monitoring systems and integrative time-series analysis models: A case study in Bogor, Indonesia. Frontiers in energy, 12, 426-439.
[22] Charfeddine, L., Zaidan, E., Alban, A. Q., Bennasr, H., & Abulibdeh, A. (2023). Modeling and forecasting electricity consumption amid the COVID-19 pandemic: Machine learning vs. nonlinear econometric time series models. Sustainable Cities and Society, 98, 104860.
[23] Bilgili, M., & Pinar, E. (2023). Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye. Energy, 284, 128575.
[24] ABDULLATEEF, A. I., SALAMI, M. J. E., & AKOREDE, M. F. (2018). Intelligent technique for electricity theft identification using autoregressive model. LAUTECH Journal of Engineering and Technology, 12(1), 1-12.
[25] Otuoze, A. O., Mustafa, M. W., Sultana, U., Abiodun, E. A., Jimada-Ojuolape, B., Ibrahim, O., ... & Abdullateef, A. I. (2024). Detection and confirmation of electricity thefts in Advanced Metering Infrastructure by Long Short-Term Memory and fuzzy inference system models. Nigerian Journal of Technological Development, 21(1), 112-130.
[26] Pamir, Javaid, N., Javed, M. U., Houran, M. A., Almasoud, A. M., & Imran, M. (2023). Electricity theft detection for energy optimization using deep learning models. Energy Science & Engineering, 11(10), 3575-3596.
[27] Pamir, Javaid, N., Javaid, S., Asif, M., Javed, M. U., Yahaya, A. S., & Aslam, S. (2022). Synthetic theft attacks and long short term memory-based preprocessing for electricity theft detection using gated recurrent unit. Energies, 15(8), 2778.
[28] Munawar, S., Javaid, N., Khan, Z. A., Chaudhary, N. I., Raja, M. A. Z., Milyani, A. H., & Ahmed Azhari, A. (2022). Electricity theft detection in smart grids using a hybrid BiGRU–BiLSTM model with feature engineering-based preprocessing. Sensors, 22(20), 7818.
[29] Shen, Y., Shao, P., Chen, G., Gu, X., Wen, T., Zang, L., & Zhu, J. (2021). An identification method of anti-electricity theft load based on long and short-term memory network. Procedia Computer Science, 183, 440-447.
[30] J. Marulanda Durango,A. Escobar Mejía & E. Giraldo. (2025). Control System of an SVC Based on ARMA Model to Improve the Displacement Power Factor of an Electric Arc Furnace. Journal of Electrical and Computer Engineering,2025(1),9945400-9945400.
[31] Zhen Si,Zhaoqing Liu,Changchun Mu,Meng Wang,Tongxin Gong,Xiaofang Xia... & Yang Xiao. (2025). A new deep learning based electricity theft detection framework for smart grids in cloud computing. Computer Standards & Interfaces,94,104007-104007.