energies

MDPI

*Article*

# A Novel Electricity Theft Detection Strategy Based on Dual-Time Feature Fusion and Deep Learning Methods

**Qinyu Huang [1], Zhenli Tang [2,*], Xiaofeng Weng [2], Min He [2], Fang Liu [2], Mingfa Yang [1] and Tao Jin [1,*]**

[1] Department of Electrical Engineering, Fuzhou University, Fuzhou 350116, China; 220127039@fzu.edu.cn (Q.H.); yangmf@fzu.edu.cn (M.Y.)

[2] Fujian YILI Information Technology Co., Ltd., Fuzhou 350001, China; wengxf@139.com (X.W.); 18059105943@189.cn (M.H.); rickydc@163.com (F.L.)

* Correspondence: 13950360570@163.com (Z.T.); jintly@fzu.edu.cn (T.J.)

**Abstract:** To enhance the accuracy of theft detection for electricity consumers, this paper introduces a novel strategy based on the fusion of the dual-time feature and deep learning methods. Initially, considering electricity-consumption features at dual temporal scales, the paper employs temporal convolutional networks (TCN) with a long short-term memory (LSTM) multi-level feature extraction module (LSTM-TCN) and deep convolutional neural network (DCNN) to parallelly extract features at these scales. Subsequently, the extracted features are coupled and input into a fully connected (FC) layer for classification, enabling the precise detection of theft users. To validate the method's effectiveness, real electricity-consumption data from the State Grid Corporation of China (SGCC) is used for testing. The experimental results demonstrate that the proposed method achieves a remarkable detection accuracy of up to 94.7% during testing, showcasing excellent performance across various evaluation metrics. Specifically, it attained values of 0.932, 0.964, 0.948, and 0.986 for precision, recall, F1 score, and AUC, respectively. Additionally, the paper conducts a comparative analysis with mainstream theft identification approaches. In the comparison of training processes, the proposed method exhibits significant advantages in terms of identification accuracy and fitting degree. Moreover, with adjustments to the training set proportions, the proposed method shows minimal impact, indicating robustness.

**Keywords:** deep learning; electricity theft detection; feature fusion; parallel model

check for updates

**Citation:** Huang, Q.; Tang, Z.; Weng, X.; He, M.; Liu, F.; Yang, M.; Jin, T. A Novel Electricity Theft Detection Strategy Based on Dual-Time Feature Fusion and Deep Learning Methods. *Energies* **2024**, *17*, 275. https://doi.org/10.3390/en17020275

Academic Editor: Valentina E. Balas

Received: 28 November 2023
Revised: 25 December 2023
Accepted: 27 December 2023
Published: 5 January 2024

## 1. Introduction

The theft of electricity by users is a significant factor contributing to non-technical losses (NTL) in the operation of the power grid. This behavior not only causes substantial economic losses to the country [1–4] but also has adverse effects on the safe and stable operation of the electrical power system [5]. Traditional methods of electricity theft include deliberately damaging the electricity meter, bypassing the meter to obtain electricity illegally, and so on. Traditional methods of detecting electricity theft typically require employees from the power company to physically record the electricity consumption of meters and identify meters with anomalies. This approach is both time-consuming and labor-intensive, resulting in low efficiency [6]. In recent years, with the upgrading of smart grids and the widespread adoption of advanced metering infrastructure (AMI), power companies can now record electricity-consumption information in real-time [7]. Although the widespread use of AMI systems can reduce the occurrence of electricity theft events, such illicit activities persist and have become more covert and sophisticated. This is due to the fact that AMI systems need to transmit user electricity-consumption information to cloud storage. As a result, methods of electricity theft have evolved to include interference with information communication in AMI systems or network attacks on cloud data storage centers to manipulate electricity-consumption information [8], achieving the goal of electricity theft.

With the popularization of AMI systems and the wealth of user electricity information they provide, research on data-driven electricity theft detection is gradually becoming a highly regarded and popular field. Currently, many scholars have conducted extensive research in this field, and the main research directions can be summarized as electricity theft detection based on system state estimation, electricity theft detection based on game theory, and electricity theft detection based on artificial intelligence algorithms [9]. The electricity theft detection method based on system state estimation relies on mathematical constraints on various electrical parameters in the power system. By observing the relationships between electrical parameters and the differences between actual measured values and estimated values, this method determines whether there is theft behavior by users. However, this approach has limitations in practical engineering applications, as it requires complete network topology information and accurate parameter data. The principle of electricity theft detection based on game theory involves using the concepts and methods of game theory to model and analyze the interaction between electricity thieves and power companies in the power system. However, this approach relies on a deep understanding of the strategies and objectives of the participants, making it challenging to find a suitable model to represent the relationship between electricity thieves and power companies [10]. With the widespread application of big data and artificial intelligence algorithms in smart grids, an increasing number of researchers are incorporating artificial intelligence algorithms into the data-driven theft detection field. The advantage of this approach is that it is less dependent on comprehensive grid information. Leveraging algorithms and data-mining techniques, it can more flexibly handle and analyze electricity-consumption data, providing a novel pathway for electricity theft detection. Regarding electricity theft detection methods based on artificial intelligence algorithms, we will provide a more detailed introduction in the following sections.

In order to improve the accuracy of electricity theft detection, this paper proposes a novel electricity theft detection strategy based on dual-time feature fusion and deep learning methods. This strategy uses temporal convolutional networks (TCN) with long short-term memory (LSTM) multi-level feature extraction module (LSTM-TCN) and deep convolutional neural network (DCNN) to extract the user′s one-dimensional power-consumption features and two-dimensional power-consumption features in parallel. Then, the power-consumption characteristics of dual-time scales are coupled, user classification is performed through a fully connected (FC) layer, and, finally, electricity theft detection is realized. In order to verify the effectiveness of the proposed method, this paper uses the State Grid Corporation of China (SGCC) dataset [11] to conduct simulation experiments and compares it with other electricity theft detection model algorithms. These comparative models include support vector machine (SVM) [12], random forest (RF), extreme gradient boosting (XGBoost) [13], gradient boosting decision tree (GBDT), 1D-CNN, 2D-CNN, DCNN, LSTM-TCN, and CNN-LSTM [14]. The results indicate that, compared to the comparative models, the proposed model outperforms in terms of accuracy, recall, F1 score, and area under the receiver operating characteristic curve (AUC), and it achieves the second-best performance in precision. The proposed electricity theft detection strategy not only exhibits higher accuracy and outstanding performance but also demonstrates superior performance under different training set proportions. In this context, the proposed electricity theft detection strategy shows excellent robustness. The main contributions of this paper are summarized as follows:

- By visualizing and introducing the Pearson correlation coefficient, we analyzed the differences in electricity-consumption characteristics between normal users and electricity theft users on a weekly scale. The conclusion drawn is that, compared to normal users, electricity theft users exhibit less apparent or irregular periodic electricity-consumption features.
- In order to integrate the electricity-consumption features at both daily and weekly scales, this paper proposes a novel theft detection strategy based on the fusion of dual-time features and deep learning methods. This strategy utilizes a hybrid model

composed of LSTM-TCN and DCNN to concurrently extract features from both scales and achieves binary classification through an FC layer. The effectiveness of the proposed theft detection strategy is validated through simulation experiments and result analysis.

The remaining sections of this paper are organized as follows. Section 2 briefly describes previous studies on electricity theft detection methods. Section 3 analyzes the difference in electricity-consumption characteristics between normal users and electricity theft users on a weekly scale through visualization and the introduction of the Pearson correlation coefficient. Section 4 introduces the overall framework of the novel theft detection strategy based on deep learning and dual-time feature fusion, with further details on the feature extraction module. Section 5 presents experimental simulations and case analyses. Section 6 discusses the electricity theft detection strategy proposed in this paper. Finally, we conclude this paper in Section 7.

## 2. Related Works

The artificial intelligence detection methods mainly include machine learning methods and deep learning methods. In previous research, various methods have been applied to theft detection. In [12], a new approach towards NTL detection in power utilities using an artificial intelligence based on SVM is proposed, and experimental results show that SVM has high discriminative ability and can accurately classify electricity customers. In [13], a metered data theft detector based on XGBoost is proposed. Simulation results indicate that, compared to other machine learning methods, this approach can detect theft behavior with higher accuracy and lower false alarm rates, demonstrating good robustness in the presence of data imbalance. However, given the limited detection performance of a single learner, an increasing number of scholars are exploring the combination of various learners to enhance the accuracy of electricity theft detection. In [15], Jindal et al. proposes a comprehensive top-down scheme based on decision tree (DT) and SVM. Experimental results demonstrate that this scheme significantly reduces false alarms and is practical enough for real-time implementation. In [16], the authors proposed a hybrid method based on deep learning and SVM for energy theft detection, and the results verified the effectiveness of the method in terms of accuracy and a small detection error. In [17], an ensemble learning-based system for detecting energy theft using a hybrid approach is proposed. The simulation results show that the proposed ensemble model outperforms the state-of-the-art methods in terms of accuracy when compared with the state-of-the-art methods. Through the studies [15–17], we can conclude that integrating multiple machine learning methods can effectively improve the performance of electricity theft detection. In the field of machine learning, deep learning methods are widely applied in areas such as load forecasting [18], energy scheduling optimization [19], and analysis of power quality disturbances [20,21]. This paper applies them to the field of electricity theft detection. Deep learning methods primarily utilize neural networks such as convolutional neural networks (CNN), recurrent neural networks (RNN), and their variants, simulating the structure and self-training ability of the human brain. They learn electricity-consumption characteristics from users' electricity-consumption data to achieve the detection of electricity theft. Zheng et al. [11] propose a hybrid CNN to improve the electricity theft detection accuracy, but the limitation of this work is that the fully connected layer that extracts one-dimensional electricity consumption cannot learn the time dependence in customer electricity-consumption time series [1]. In [14], a theft detection system that combines CNN and LSTM structures is proposed. The experimental results indicate that this approach can classify users with high accuracy. A novel theft detection method based on a convolution–non-convolution parallel deep network is proposed in [22]. This method converts load time series into two-dimensional images and utilizes neural networks to capture features at different time scales. Simulations indicate that this approach significantly improves the performance of electricity theft detection. In [23], a multi-resolution convolutional neural networks architecture for fraud detection on smart grids is proposed. This method converts electricity-consumption data into images, combines the electricity-consumption features of one-dimensional electricity consumption, and extracts features through CNN, respectively. However, the conversion of electricity-consumption data into two-dimensional images through methods such as visualization, Gram's angle fields, and Markov transition fields will complicate the design and training process of the classification model, and the visualization may lead to the loss of features between data. On the contrary, directly using electricity-consumption data as the input of the model can greatly simplify the training process and preserve the temporal characteristics between the data. Table 1 summarizes the algorithms, data sources, and the advantages and disadvantages of related works.

**Table 1.** The summary of algorithms, data source, and the advantages and disadvantages of related works.

| Refs. | Algorithms | Data Source | Advantages | Disadvantages |
|---|---|---|---|---|
| [12] | SVM | Tenaga Nasional Berhad Distribution (TNBD), Sdn.Bhd. | Detects meter tampering and meter bypassing/Detects abrupt changes in load-profile. | Readings are transformed into average, which can deviate from actual values. |
| [13] | XGBoost | Irish Smart Energy Trial Dataset | The proposed method is robust when the data are imbalanced. | The proposed method only analyzes electricity-consumption data alone, which may produce limited results. |
| [15] | Decision Tree coupled SVM | References [37–39] in [15] | The proposed scheme is capable enough to detect the thefts happening anywhere in the power network. | This scheme needs to obtain many features in advance. |
| [16] | CNN-SVM | Non-public | The proposed model can extract features automatically. | The detection performance on other data sets needs further verification. |
| [17] | ensemble-learning approach and ML-classifier training method | Mendeley datasets | This study presents an ensemble learning-based system for detecting energy theft using a hybrid approach. | The main limitation of this work is its computational complexity. |
| [11] | Hybrid CNN | SGCC Dataset | A hybrid CNN model is proposed to improve the accuracy of electricity theft detection. | The fully connected layer in the model cannot learn the temporal dependence in customer electricity-consumption time series. |
| [14] | CNN-LSTM | SGCC Dataset | The CNN is used for automatically extracting high-level features, then these features are then flattened and fed into an LSTM neural network for capturing temporal dependencies. | Requires larger computing resources. |
| [22] | Convolution + non-convolution deep network | SGCC Dataset | The proposed method comes from the CNCP structure, which can capture features of electricity-consumption time series at different scales. | The process of visualizing data is relatively complex and carries the potential risk of losing features among the data. |
| [23] | CNN | Uruguayan power generation and distribution company and CER dataset | Present a multi-resolution convolutional neural networks architecture for fraud detection on smart grids. | The process of visualizing data is relatively complex and carries the potential risk of losing features among the data. |

In comparison to previous works, this paper tackles the challenges posed by long-term dependencies in one-dimensional electricity-consumption sequences through the utilization of LSTM-TCN. Additionally, drawing on previous research findings that demonstrate a better representation of electricity-consumption patterns by converting one-dimensional usage data into a two-dimensional format on a weekly basis, this paper opts to utilize DCNN for extracting two-dimensional electricity-consumption features. Furthermore, the paper effectively addresses the issue of sample imbalance, introduces additional evaluation metrics, and incorporates comparison models to comprehensively assess the effectiveness of the proposed electricity theft detection strategy.

## 3. Analysis of User Electricity-Consumption Characteristics

After visualizing and analyzing the one-dimensional electricity-consumption data for each electricity customer, it was observed that the electricity-consumption curve of regular users exhibits a certain periodicity, while the electricity-consumption curve of theft users shows significant fluctuations and pronounced peaks and valleys, indicating sudden increases and decreases in electricity consumption during specific periods. Figure 1 illustrates the monthly electricity consumption for a normal user and an electricity theft user.

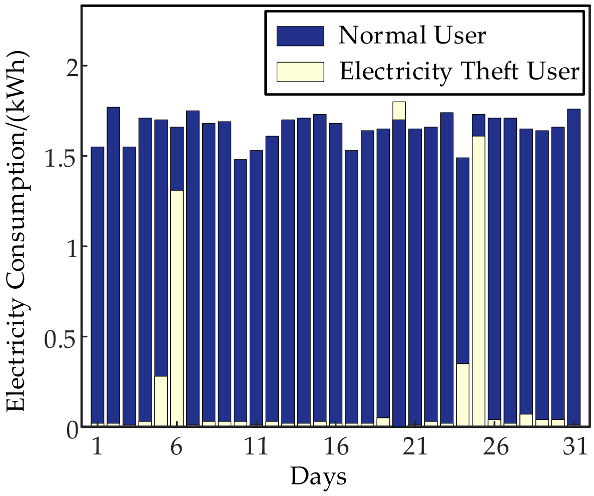


**Figure 1.** Monthly electricity consumption by normal user and electricity theft user.

Through Figure 1, it can be observed that the electricity-consumption patterns of regular users exhibit a consistent regularity each week, while there is a noticeable difference in the electricity usage of electricity theft users in the second week compared to the other three weeks. Therefore, this study considers plotting the electricity-consumption graph on a weekly basis to observe the consumption patterns of electricity customers. In the graph, each week is represented on the vertical axis, and each day within a week is represented on the horizontal axis. The electricity-consumption bar chart is generated using data from the first 28 days (4 weeks) as shown in Figure 2.

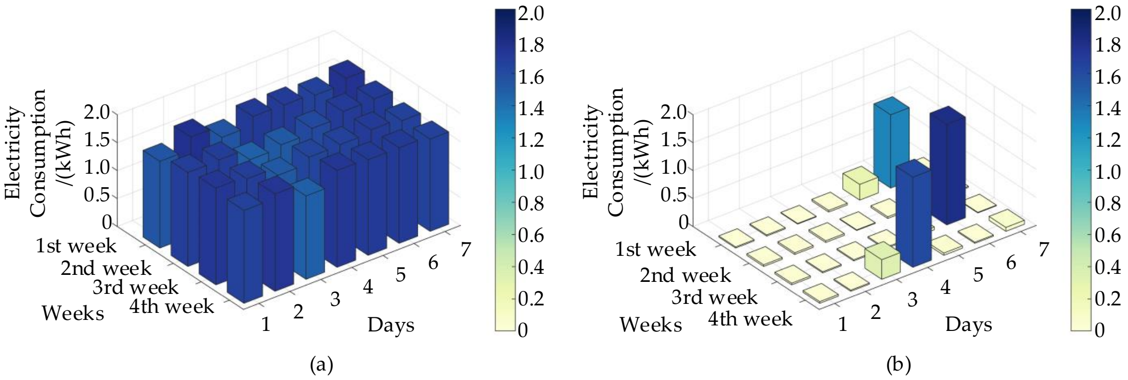


**Figure 2.** Weekly electricity consumption of electricity consumers. (**a**) Normal User. (**b**) Electricity Theft User.

Furthermore, to better analyze the periodicity of user electricity consumption, this study incorporates the Pearson correlation coefficient to measure the degree of correlation between consecutive weeks. Figure 3 depicts the correlation in weekly electricity-consumption patterns between normal users and electricity theft users.

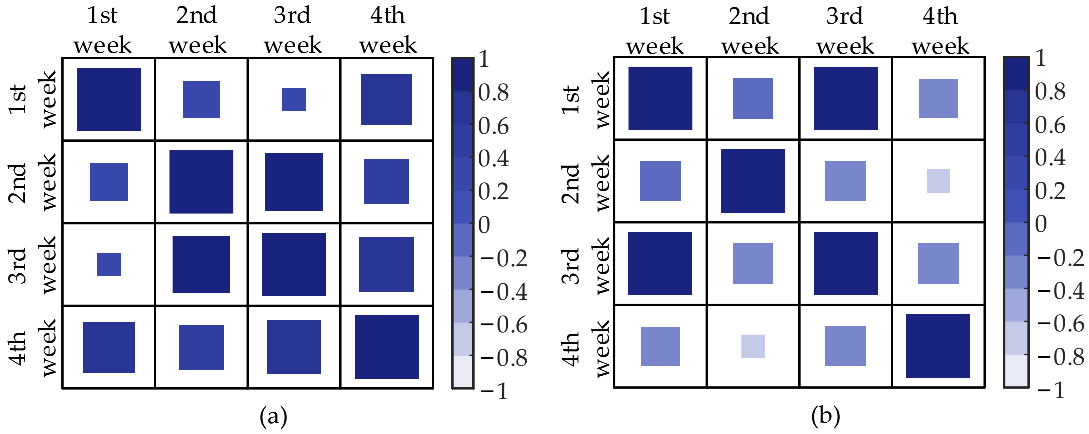


**Figure 3.** The correlation heat map of weekly electricity consumption of electricity customers. (**a**) Normal User. (**b**) Electricity Theft User.

By combining Figures 2 and 3, it can be observed that normal users exhibit a weekly electricity-consumption pattern with positive correlations between consecutive weeks, indicating a strong relationship in weekly electricity consumption. Furthermore, this normal user experiences a peak in electricity consumption every Tuesday and a trough every Wednesday within each week. The difference in electricity consumption between adjacent days does not exceed 0.3 kWh, demonstrating small and relatively stable fluctuations. In contrast, electricity theft users show mostly negative correlations between weeks, lacking clear weekly electricity-consumption patterns. These users tend to have days with nearly zero electricity consumption, occasionally experiencing sudden peaks in electricity usage on certain days. The difference in electricity consumption between adjacent days can reach a maximum of 1.75 kWh, indicating significant and erratic fluctuations in electricity consumption.

In conclusion, the analysis suggests that, in comparison to normal users, electricity theft users exhibit either minimal or no apparent weekly electricity-consumption patterns. Consequently, this study introduces weekly scale electricity-consumption features as an auxiliary tool for electricity theft detection.

## 4. Framework for Electricity Theft Detection Strategy

In this section, we will introduce in detail the new electricity theft detection framework based on dual-time feature fusion and deep learning methods proposed in this article, then further introduce each module.

### *4.1. TCN with LSTM Multi-Level Feature Extraction Module*

Given that users' electricity-consumption data constitutes a complex long sequence classification problem [24], LSTM was particularly well-suited for feature extraction in this application. LSTM is a specialized form of RNN designed for handling long-term dependencies in time series data, and its structure is shown in Figure 4. Unlike traditional RNNs, LSTM excels in processing significant events with extended intervals. It introduces memory cells, input gates, output gates, and forgetting gates to effectively capture and manage crucial long-term dependencies in sequences. This architecture overcame the gradient-related challenges associated with traditional RNNs, making it well-suited for predicting and processing events with prolonged intervals in time series data.

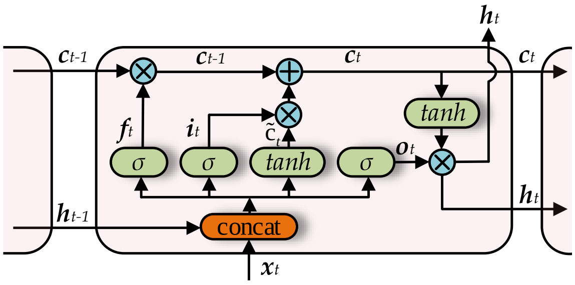


**Figure 4.** Long Short-Term Memory structure diagram.

In the Figure 4, $x_t$ represents the input information at the current moment, $h_{t-1}$ denotes the hidden state from the previous moment, $h_t$ signifies the hidden state passed to the next moment, $\sigma$ represents the sigmoid function, and *tanh* represents the hyperbolic tangent function. The components related to the forget gate, input gate, candidate cell state, new cell state, and output are, specifically, as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\widetilde{C}_t = \sigma(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3}$$

$$C_t = f_t \times C_{t-1} + i_t \times \widetilde{C}_t \tag{4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \times \tan \mathrm{h}(C_t), \tag{6}$$

where $W$ represents weight.

With tanh as the activation function, its zero-centering contributes to model stability, ensuring an output range within $[-1, 1]$. This helps prevent gradient vanishing and promotes model convergence. Its definition is as follows:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{7}$$

Concerning the network selection issue, prior research highlights the effectiveness of RNN in addressing various sequence problems [25,26], including the sequence classification challenge posed by electricity theft detection. However, practical application exposes a limitation of RNN: its inability to perform large-scale parallel computations due to its sequential processing nature, handling one time step at a time. This sequential approach leads to excessive memory consumption and prolonged training times.

To address these shortcomings, we chose TCN as one of the modules for extracting daily-scale electricity-consumption features [27]. TCN capitalizes on the parallel processing capabilities of convolutional networks within a multi-layer structure, allowing for a broader receptive field. In contrast to RNN, TCN's parallel processing nature ensures efficient computation, even with a deep network structure, resulting in substantial time savings during training. TCN is built upon CNN and incorporates enhancements inspired by the residual network.

Causal Convolution: In Figure 5a,b, the computational structures of traditional convolution and causal convolution are presented. When considering timing issues, a drawback of traditional one-dimensional convolution becomes evident—it operates bidirectionally, observing the entire input sequence during convolution. This characteristic may inadvertently disclose future information, posing challenges for time series data analysis, where temporal correlation is crucial. To mitigate this, causal convolution was introduced. Causal convolution ensures that each output time step depends solely on current and past inputs, preventing reliance on future inputs. When applied to time series data, causal convolution effectively preserves temporal correlations, mitigating the risk of revealing future information and thereby enhancing model accuracy.

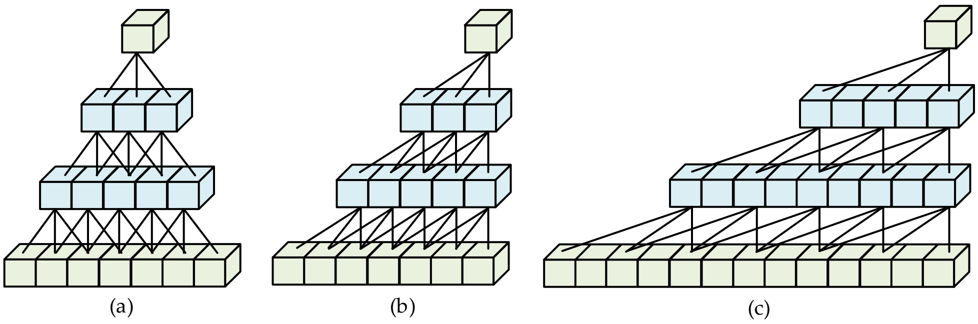


**Figure 5.** Comparison of different convolution modes. (**a**) Convolution. (**b**) Causal Convolution. (**c**) Dilated Convolution.

Dilated Convolution: In conventional convolution, a down-sampling operation is typically employed post-convolution to reduce parameters and computational load, but this comes at the cost of information loss—commonly associated with down-sampling techniques, such as pooling [28]. To address this trade-off, dilated convolution emerged as a solution. Dilated convolution (as shown in Figure 5c) achieves efficient down-sampling by increasing the distance (dilation rate) in the convolution kernel without significant information loss. By widening the interval, the receptive field of the convolution kernel on the feature map expands, allowing the convolutional neural network to extract features across a broader range. This strategic approach reduces the size of the feature map while minimizing information loss.

Given the uncertainty, randomness, and complexity of electricity users, the proposed one-dimensional feature extraction framework in this article initially concatenated features from the original input and LSTM output. Subsequently, one-dimensional feature extraction was conducted through TCN. The incorporation of multi-level interleaved connections ensured that the model captured both deep and shallow crucial information. Utilizing a neural network, tailored for sequence data in the feature extraction process, guaranteed the accurate extraction of features from the electrical power data sequence. In TCN, rectified linear unit (ReLU) is used as the activation function. Its advantages include simple computation, effective prevention of gradient vanishing, and accelerated model convergence [14]. The ReLU activation function is defined as follows:

$$\mathrm{ReLU}(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases} \tag{8}$$

Please refer to the TCN with LSTM multi-level feature fusion module in Figure 6 for the specific model structure. The parameter design of this part is shown in Table 2.

**Table 2.** Parameter settings of the TCN with LSTM multi-level feature extraction module.

| Layer Name | Parameters | Activation Function |
|---|---|---|
| LSTM | units = 1<br>return sequences = True | tanh |
| Concatenate | axis = 0 | None |
| Res Block1 | filters =32<br>kernel size = 3<br>dilation rate = 1 | ReLU |
| Res Block2 | filters =32<br>kernel size = 3<br>dilation rate = 2 | ReLU |
| Res Block3 | filters =16<br>kernel size = 3<br>dilation rate = 4 | ReLU |
| Dense | units = 128 | ReLU |

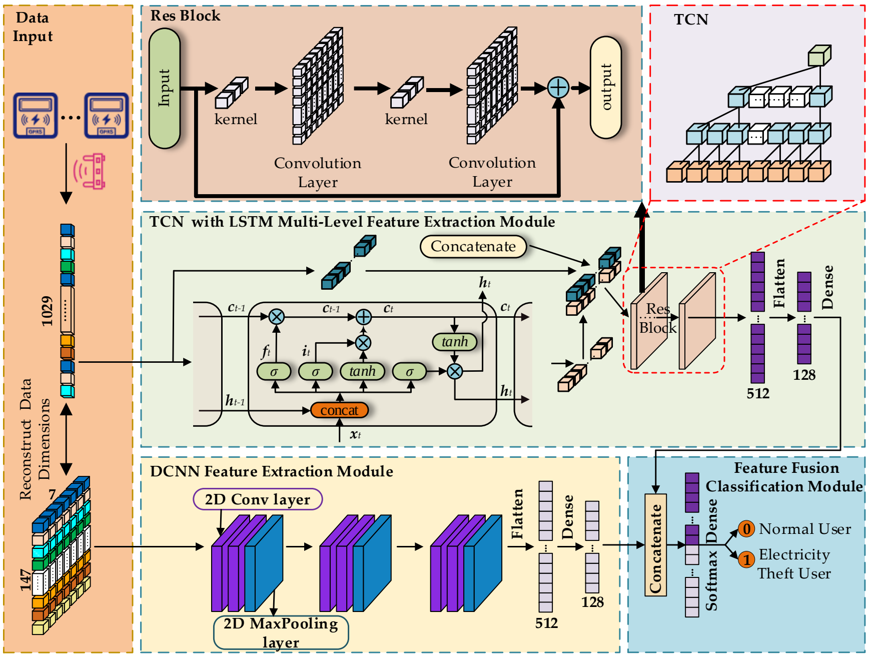


**Figure 6.** Structure diagram of electricity theft detection based on dual-time feature fusion and deep learning methods.

### 4.2. DCNN Feature Extraction Model

From the analysis in Section 2, it is evident that observing only the one-dimensional electricity-consumption data made it challenging to discern the consumption patterns of electricity users. However, transforming the one-dimensional electricity-consumption data into a weekly format can better highlight these patterns. Inspired by this insight, this study used CNN as a basis to design a two-dimensional power-consumption feature extraction module DCNN. The user's one-dimensional electricity-consumption data sequence was transformed into two-dimensional data on a weekly basis, which was then fed into this module to extract periodic electricity-consumption patterns from the two-dimensional data.

Utilizing DCNN for two-dimensional electricity-consumption feature extraction brings forth several advantages. First, DCNN excels at processing image data, and representing the weekly electricity-consumption dataset in a two-dimensional format is akin to treating it as an image data representation. This approach facilitates DCNN more effectively capturing features related to weekly-scale electricity consumption. Second, DCNN is adept at learning and extracting hierarchical features, providing valuable assistance in identifying intricate patterns within electricity-consumption data. Through the synergistic integration of multiple convolutional and pooling layers, the network can progressively abstract and comprehend high-level features of the data, thereby enhancing its ability to capture intricate electricity-consumption patterns and trends. Consequently, this paper introduced a meticulously designed DCNN aimed at efficiently extracting features associated with weekly-scale electricity consumption. The process of extracting electricity-consumption features was as follows:

$$Z_i(\mathrm{x}) = A_{i-1} * W_i + b_i \tag{9}$$

$$A_i = f_{pool}(Z_i) \tag{10}$$

In each feature extraction module mentioned above, $Z$ represents the linear output of the convolutional layer, $A$ corresponds to the activation output of the previous layer, $W$ denotes the weights associated with the convolutional kernel, $b$ is the bias of the convolutional layer, and $f_{pool}$ represents the pooling function. After the convolutional computation is completed, the result is fed through the ReLU activation function to the next layer of connected units, cycling twice before sending the features to the pooling layer.

The parameter settings for the two-dimensional electricity-consumption feature extraction module are outlined in Table 3. The module consists of three feature extraction blocks, each comprising two convolutional layers and one pooling layer. Finally, the output of the last max pooling layer is processed through a flatten layer and a dense layer to produce the two-dimensional electricity-consumption features.

**Table 3.** Parameter settings of DCNN feature extraction module.

| Layer Name | Parameters | Activation Function |
|---|---|---|
| Conv-1 | kernel_size = 3 × 3, filter = 32 | ReLU |
| Conv-2 | kernel_size = 3 × 3, filter = 32 | ReLU |
| MaxPooling-1 | kernel_size = 2×2 | None |
| Conv-3 | kernel_size = 3 × 3, filter = 64 | ReLU |
| Conv-4 | kernel_size = 3 × 3, filter = 64 | ReLU |
| MaxPooling-2 | kernel_size = 2 × 2 | None |
| Conv-3 | kernel_size = 3 × 3, filter = 128 | ReLU |
| Conv-4 | kernel_size = 3 × 3, filter = 128 | ReLU |
| MaxPooling-2 | kernel_size = 1 × 1 | None |

### *4.3. Feature Fusion Module*

Building upon the earlier discussion, this paper proposed a novel approach for detecting electricity theft. The approach integrated dual-time feature fusion and deep learning methods, as illustrated in Figure 6. It included extracting global power-consumption features from one-dimensional data using the TCN with LSTM multi-level feature extraction. Additionally, it captured periodic power-consumption patterns from two-dimensional data through the DCNN, using the concatenate module for feature fusion. The fused features undergo two-class classification via the FC layer, employing the Softmax function for activation. The resulting variable $Y$ represented the probability $P(Y_i)$ of the $i$-th category, calculated as follows:

$$P(Y_i) = \text{Softmax}(Y_i) = \mathrm{e}^{Y_i} / \sum_{k=1}^{N} \mathrm{e}^{Y_k}, \tag{11}$$

where $N$ represented the number of categories that needed to be classified, and $Y_i$ represented the output value of the $i$-th node. Finally, the maximum value of $P(Y_i)$ was selected as the final result to complete the detection of electricity theft users.

## 5. Simulation Analysis and Experimental Results

In order to verify the feasibility of the new electricity theft detection strategy based on dual-time feature fusion and deep learning methods proposed in this article, this section will conduct simulation experiments and analyze the experimental results.

The proposed model and the comparison models involved in this study are built using the TensorFlow and Keras frameworks in Python. The hardware specifications and software versions configured in the experimental environment are presented in Table 4.

**Table 4.** Hardware Models and Software Versions.

| Hardware/Software | Model/Version | Hardware/Software | Model/Version |
|---|---|---|---|
| OS | Win10 (64 bit) | Python | 3.6 |
| CPU | Intel Core i9-9820X<br>@3.3.0 GHz | Tensorflow<br>Keras | 2.0.0<br>2.3.1 |
| GPU | NVIDIA GeForce RTX 2080 | Scikit-learn | 0.24.2 |
| RAM | DDR4 32 GB | CUDA | 10.0 |
| HDD | SSD 1 TB | cuDNN | 7.6.5 |

### 5.1. Data Preparation

The dataset utilized in this paper comprises a real dataset provided by the SGCC, encompassing the daily electricity-consumption data of 42,372 users over 1034 days, from 1 January 2014, to 31 October 2016. Among them, 38,757 normal users are marked as 0, and 3615 electricity theft users are marked as 1, and electricity theft users account for 8.5% of the entire dataset.

The initial dataset is riddled with numerous missing values and outliers, potential detriments to the model's classification performance. Moreover, as mentioned earlier, the dataset showcases an imbalance in class distribution, with a tendency for the model to prioritize learning the electricity-consumption patterns of regular users, potentially overlooking those of theft users. To address these issues, this study adopts the data preprocessing method proposed in [29] to handle missing and outlier values in the raw electricity-consumption dataset. The processing formula is shown below. Additionally, a comprehensive sampling method from [29] is employed to balance the dataset by adjusting the number of normal and theft users.

The formula for handling missing values is as follows:

$$f(x_i) = \begin{cases} \frac{x_{i-1}+x_{i+1}}{2}, & x_i \in \text{NaN}, x_{i-1} \& x_{i+1} \notin \text{NaN} \\ 0, & x_i \in \text{NaN}, x_{i-1} | x_{i+1} \notin \text{NaN} \\ x_i, & x_i \notin \text{NaN} \end{cases}, \tag{12}$$

where $x_i$ represents the electricity-consumption value of the electricity customer on a certain day. If $x_i$ is a null value, it is represented by NaN.

For the treatment of outliers, this paper utilizes the "3σ rule" to handle the outliers present in the dataset. The formula is as follows:

$$f(x_i) = \begin{cases} X, \text{ if } x_i > X \\ x_i, \text{ otherwise} \end{cases}, \tag{13}$$

where $X$ is equal to avg($x$) + 2·std($x$), where avg($x$) is the average value of $x$ and std($x$) is the standard deviation of $x$. Here, $x$ represents the daily electricity-consumption set for a particular customer, $x = \{x_1, x_2, \ldots, x_n\}$ ($n$ is the total number of days a user consumes electricity).

Normalization processing can accelerate neural network convergence and prevent significant detection errors arising from vast differences in data magnitude. This article employs the maxmin method to normalize electricity-consumption data within the range of (0, 1). The formula for the max-min normalization method is as follows:

$$f(x_i) = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{14}$$

### 5.2. Evaluation Index

Electricity theft detection is essentially a two-classification problem. This paper introduces a confusion matrix to illustrate the model's classification results. The electricity theft detection confusion matrix is presented in Table 5. Model classification results are typically categorized into four situations: true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

**Table 5.** Confusion Matrix for Electricity Theft Detection.

| User Category | Actual: Abnormal | Actual: Normal |
|---|---|---|
| Predict: Abnormal | TP | FP |
| Predict: Normal | FN | TN |

According to the confusion matrix, this article introduces six indicators: accuracy (Acc), precision (Pre), recall, F1 score (F1), receiver operating characteristic curve (ROC), and area under the ROC curve (AUC) to evaluate model performance. The abscissa of the ROC curve, FPR, is the false positive rate, and TPR is the true positive rate, also known as the recall rate. The formulas of each indicator are as follows:

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{15}$$

$$\mathrm{Pre} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{16}$$

$$\mathrm{Recall} = \mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{17}$$

$$\mathrm{F1} = 2 \times \frac{\mathrm{Pre} \times \mathrm{Recall}}{\mathrm{Pre} + \mathrm{Recall}} \tag{18}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{19}$$

The AUC value represents the probability that a randomly selected positive sample ranks higher than a randomly selected negative sample. The AUC calculation formula is as follows:

$$\mathrm{AUC} = \frac{\sum_{i \in positiveclass} Rank_i - \frac{M(1+M)}{2}}{M \times N}, \tag{20}$$

where $Rank_i$ represents the sequence number of sample $i$ in ascending order of probability and $M$ and $N$ represent the number of positive samples and negative samples, respectively.

### 5.3. Hyperparameter Selection

During the model training process, this article introduces the learning rate attenuation strategy and early stopping mechanism to enhance network training performance and prevent the deepening of network overfitting. The initial learning rate is set to 0.001. In the learning rate decay strategy, if the accuracy of the validation set does not increase for three consecutive epochs, the initial learning rate is reduced to 0.1 times the original value. The early stopping mechanism is triggered when the accuracy of the training set fails to increase for five consecutive epochs, leading the network to halt training and save the optimal model.

The Adam optimizer can adaptively adjust the learning rate according to gradient changes. Therefore, this article opts for Adam as the training optimizer.

An appropriate batch size is crucial for the training of neural networks. Too small of a batch size may lead to excessively long training times and difficulty in achieving convergence, while too large of a batch size may result in a decline in the model's generalization performance. Therefore, this paper conducts a detailed comparison of batch sizes, including 16, 32, 64, 128, and 256. The comparative results are illustrated in Figure 7.

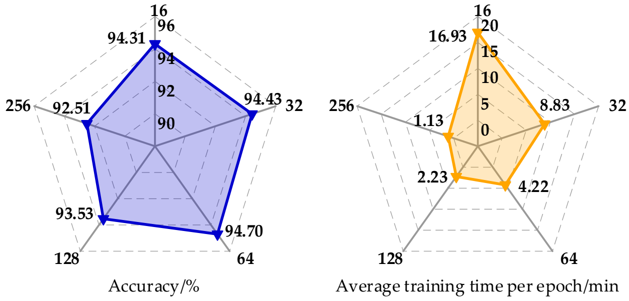


**Figure 7.** Training results for different batch sizes.

The impact of the batch size on the training time and accuracy is clearly illustrated in Figure 7. The training time exhibits a negative correlation with the batch size, decreasing as the batch size increases. Notably, the test set accuracy peaks when the batch size is 64, indicating optimal performance. For batch sizes less than 64, the test set accuracy is directly proportional to the batch size, while for batch sizes exceeding 64, the accuracy decreases with an increase in the batch size. Based on this analysis, the paper ultimately selects a batch size of 64 to achieve the best test set accuracy while maintaining efficient training.

### *5.4. Model Performance Evaluation*

In order to verify the performance of the proposed model in electricity theft detection, this paper uses the same data set for comparative testing on different models. This article selects four machine learning algorithms, SVM [12], RF, XGBoost [13], and GBDT, and five deep learning algorithms, DCNN, 1D-CNN, 2D-CNN, LSTM-TCN, and CNN-LSTM [14], as comparison models for the model proposed in this paper. During model training, all models utilize the same dataset, randomly split into training and testing sets in an 8:2 ratio. The validation set comprises 10% of the entire training set. The network parameters for the comparative models are set according to the original papers.

Figure 8 presents the training process of the proposed model alongside five comparison deep learning models, offering insights into the training performance of these models. The training set accuracy (Train_Acc) and validation set accuracy (Val_Acc) serve as key evaluation metrics throughout the training of various models. Additionally, the epochs at which each model concludes training are marked in Figure 8.

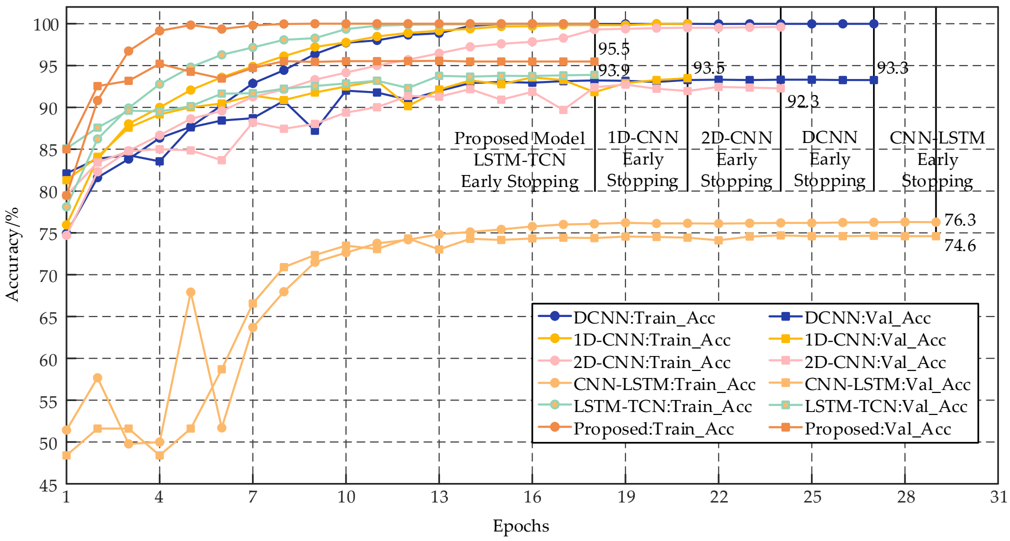


**Figure 8.** Training process of models.

It can be concluded from Figure 8 that each model triggers the conditions set by the early stopping mechanism during training and stops training early, and the model reaches convergence. By observing the training curves of each model, it is apparent that overfitting occurs across all models. Observing the training processes of various models, it is evident that CNN-LSTM exhibits relatively lower overfitting but with a lower accuracy, approximately around 76%. The proposed model, along with 1D-CNN, 2D-CNN, DCNN, and LSTM-TCN, converges to Train_Acc of around 99%, with Val_Acc all exceeding 92%. Importantly, the proposed model's Val_Acc is 3.2%, 2.2%, 2.0%, and 1.6% higher compared to 2D-CNN, DCNN, 1D-CNN, and LSTM-TCN, respectively. This indicates that the novel strategy for electricity theft detection, based on dual-time feature fusion and deep learning methods proposed in this paper, effectively improves Val_Acc, mitigates overfitting, and enhances the model's generalization performance. Notably, the two electricity-consumption feature extraction modules proposed in this paper achieve higher Val_Acc compared to 1D-CNN and 2D-CNN, further confirming the effectiveness of the designed electricity-consumption feature extraction modules. In summary, the fusion of dual-time scale electricity-consumption features leads to a higher accuracy in theft detection, validating the effectiveness of the proposed strategy. The proposed model concludes training prematurely after 18 epochs, with the Val_Acc curve consistently surpassing all other models. It stabilizes around the seventh epoch, maintaining a level of approximately 95%. This indicates that the model proposed in this study possesses a fast training speed and superior capability in extracting electrical features.

In view of the potential impact of different amounts of training data on model performance, in addition to the default training set ratio of 80%, this article also adds an additional ratio of 60% and 70%, a total of three different training set ratios, to further evaluate the performance of the model. The performance of the evaluation indicators of each model under three different training set ratios is shown in Table 6. In Table 6, Acc is the test set accuracy, and bold indicates the best effect.

**Table 6.** Comparison of evaluation indexes for various models at different training set ratios.

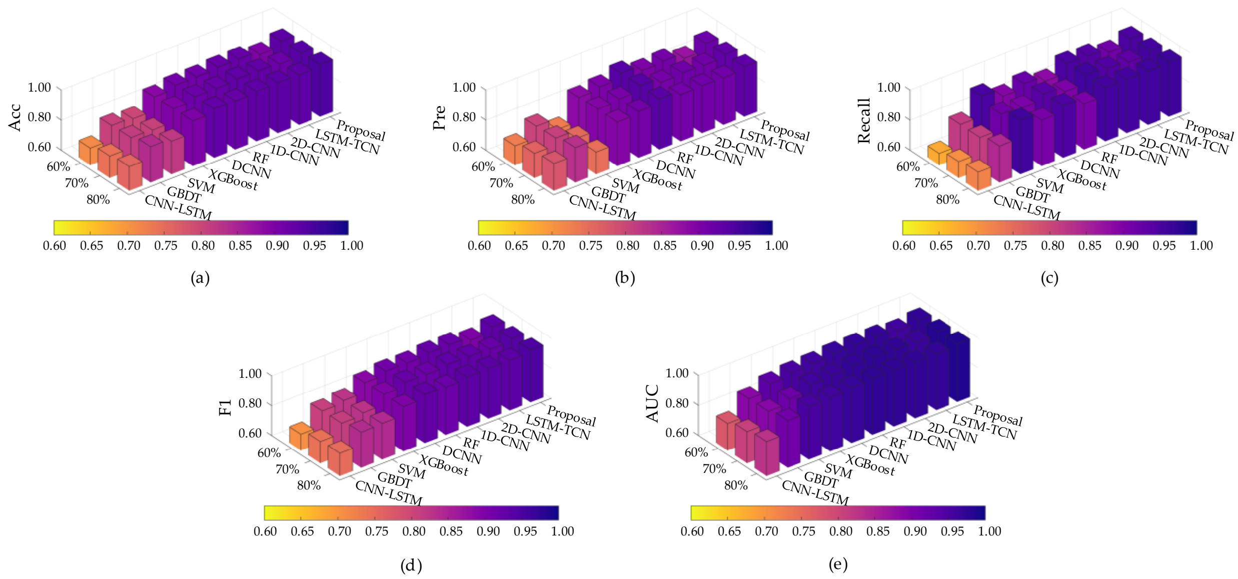
| **Models** | **Train Ratio 60%** | | | | | **Train Ratio 70%** | | | | | **Train Ratio 80%** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Acc** | **Pre** | **Recall** | **F1** | **AUC** | **Acc** | **Pre** | **Recall** | **F1** | **AUC** | **Acc** | **Pre** | **Recall** | **F1** | **AUC** |
| RF | 0.913 | **0.938** | 0.885 | 0.911 | 0.970 | 0.921 | **0.940** | 0.900 | 0.920 | 0.973 | 0.922 | **0.936** | 0.904 | 0.920 | 0.975 |
| SVM | 0.799 | 0.729 | **0.951** | 0.825 | 0.934 | 0.808 | 0.754 | 0.915 | 0.827 | 0.918 | 0.819 | 0.747 | 0.960 | 0.840 | 0.952 |
| GBDT | 0.809 | 0.804 | 0.816 | 0.810 | 0.888 | 0.813 | 0.810 | 0.814 | 0.812 | 0.890 | 0.837 | 0.831 | 0.841 | 0.836 | 0.912 |
| XGBoost | 0.889 | 0.885 | 0.894 | 0.889 | 0.952 | 0.897 | 0.894 | 0.903 | 0.898 | 0.959 | 0.900 | 0.893 | 0.905 | 0.899 | 0.958 |
| 1D-CNN | 0.923 | 0.901 | 0.949 | 0.924 | 0.971 | 0.930 | 0.907 | **0.958** | 0.931 | 0.976 | 0.932 | 0.910 | 0.957 | 0.933 | 0.975 |
| 2D-CNN | 0.924 | 0.909 | 0.943 | 0.923 | 0.972 | 0.928 | 0.915 | 0.944 | 0.929 | 0.974 | 0.934 | 0.914 | 0.957 | 0.935 | 0.974 |
| DCNN | 0.911 | 0.901 | 0.923 | 0.912 | 0.958 | 0.920 | 0.906 | 0.938 | 0.922 | 0.965 | 0.926 | 0.908 | 0.947 | 0.927 | 0.969 |
| LSTM-TCN | 0.899 | 0.873 | 0.913 | 0.900 | 0.958 | 0.919 | 0.911 | 0.932 | 0.923 | 0.966 | 0.932 | 0.912 | 0.958 | 0.934 | 0.973 |
| CNN-LSTM | 0.713 | 0.728 | 0.678 | 0.708 | 0.778 | 0.745 | 0.763 | 0.708 | 0.744 | 0.809 | 0.760 | 0.780 | 0.723 | 0.751 | 0.828 |
| Proposal | **0.933** | 0.918 | 0.949 | **0.933** | **0.974** | **0.939** | 0.930 | 0.957 | **0.940** | **0.983** | **0.947** | 0.932 | **0.964** | **0.948** | **0.986** |

Based on the data in Table 6, the performance of all models generally improves with an increase in the training set proportion. This is attributed to the increased volume of training data, which enhances the model's ability to better learn electrical features. Across different training set proportions, the proposed model consistently outperforms other comparison models on the majority of evaluation metrics, demonstrating its robustness. When the training set ratio is 80%, the Pre of the proposed model reaches the second highest among all models, 0.932, which is slightly lower than RF. In addition, it achieved the best results of 0.947, 0.964, and 0.948 in terms of Acc, recall, and F1, respectively. These three evaluation indicators exceeded the other nine comparison models. Recall increases by 12.3% compared to GBDT, and the F1 value increases by 11.2% compared to GBDT. It is worth noting that the Acc of the proposed model is improved by 18.7%, 12.8%, 11.0%, 4.7%, 2.5%, 2.1%, 1.5%, and 1.3%, respectively, compared to CNN-LSTM, SVM, GBDT, XGBoost, RF, DCNN, 1D-CNN, LSTM-TCN, and 2D-CNN. It is fully verified that the model proposed in this article can significantly improve the detection accuracy.

To further showcase the performance of the proposed model across various metrics, this article presents a histogram illustrating each model evaluation indicator under different training set proportions, as depicted in Figure 9.

From Figure 9e, it is evident that under different training set proportions, the proposed model exhibits a notable improvement in AUC compared to the comparison models, achieving the best performance. When the training set proportion is 80%, ROC curves are plotted based on the variation in TPR at different FPR levels for each model, as illustrated in Figure 10.

Combining the information from Table 6 and Figure 10, it is evident that the ROC curve of the proposed model surpasses that of other comparison models, with an AUC value reaching 0.986. Compared to CNN-LSTM, GBDT, SVM, DCNN, XGBoost, 2D-CNN, LSTM-TCN, RF, and 1D-CNN, the proposed model demonstrates improvements of 15.8%, 7.4%, 3.4%, 2.8%, 1.7%, 1.3%, 1.2%, and 1.1%, respectively. The comprehensive analysis indicates that the proposed model achieves a good balance between the true positive rate and false positive rate, establishing itself as a reliable theft detection model.



(a)

(b)

(c)

(d)

(e)

**Figure 9.** Visual comparison of model evaluation indicators under different training set ratios. (**a**) Acc. (**b**) Pre. (**c**) Recall. (**d**) F1. (**e**) AUC.

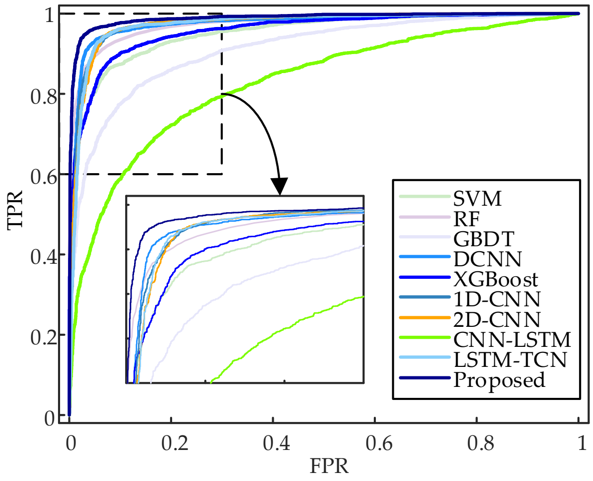


**Figure 10.** Comparison of ROC curves for various model.

After synthesizing the above analysis, it becomes apparent that the innovative electricity theft detection strategy, based on the fusion of dual-time feature and deep learning methods, can fully integrate electricity-consumption features from both time scales, unlike the approach focusing solely on a single scale. This strategy not only attains a higher detection accuracy but also showcases notably superior performance.

## 6. Discussion

To enhance the accuracy of theft detection for electricity consumers, this paper introduces a novel strategy based on the dual-time feature fusion and deep learning methods. First, we conducted an analysis of users' electricity-consumption patterns. By observing Figures 2 and 3, it is evident that normal users exhibit noticeable electricity consumption periodicity on a weekly scale, whereas electricity theft users demonstrate less apparent or irregular consumption patterns on a weekly scale. Based on this observation, we are considering incorporating weekly-scale electricity-consumption features to assist in theft detection, integrating them with the daily-scale electricity-consumption features.

Following that, we introduced a theft detection strategy based on the fusion of dual-time features and deep learning methods. The overall framework is illustrated in Figure 6, accompanied by a detailed explanation of the key modules within it. In this framework, we have devised a hybrid model composed of LSTM-TCN and DCNN to simultaneously extract dual temporal-scale electricity-consumption features. This combination was chosen because LSTM excels in capturing long-term dependencies in time series data, while TCN can capture local features at different levels, effectively avoiding the gradient vanishing issues present in traditional RNN models. Consequently, LSTM-TCN has the advantage of capturing both long-term dependencies and electricity-consumption patterns simultaneously, providing a more comprehensive understanding of users' daily-scale electricity-consumption features. On the other hand, DCNN specializes in handling image data, representing weekly electricity-consumption data in a two-dimensional format, similar to treating it as image data. This representation aids DCNN in more effectively capturing the weekly-scale power-consumption features. Furthermore, DCNN can proficiently learn and extract hierarchical features. Through the collaborative effect of multiple convolutional layers and pooling layers, the network is better equipped to capture electricity-consumption patterns and regularities. In the end, the fused electricity features are input into the FC layer for classification, accomplishing theft detection.

To validate the effectiveness of the theft detection strategy proposed in this paper, we conducted simulation experiments and performed a result analysis in Section 5. We compared the model proposed in this paper with some existing theft detection models or methods and analyzed the performance of each model under different training set ratios. The results indicate that, during the training process of deep learning models, the model proposed in this paper exhibits a significant advantage in network fitting and electricity feature extraction. As the training set ratio is adjusted, the proposed model consistently excels across most evaluation metrics, demonstrating excellent detection performance and robustness.

## 7. Conclusions

In this paper, we aim to enhance the accuracy of theft detection for electricity consumers under a data-driven approach. We consider electricity-consumption features across different time scales and leverage the advantages of feature extraction using deep learning models. Consequently, we introduce a novel theft detection strategy based on dual-time feature fusion and deep learning methods. To implement this strategy, we first utilize LSTM-TCN framework and DCNN in parallel to extract electricity-consumption features at dual temporal scales. The resulting fused features are then processed by an FC layer for classification. Subsequently, to validate the effectiveness of our proposed strategy, we conduct simulation verification using the SGCC dataset. The experimental results indicate that, during the training process, our proposed model exhibits a notable ability to alleviate overfitting compared to other models, showcasing strong capabilities in generalization and feature extraction. With an 80% training set proportion, the proposed model demonstrated optimal performance in terms of Acc, recall, F1, and AUC, achieving values of 0.947, 0.964, 0.948, and 0.986, respectively. In terms of Pre, it secured the second position among all models with a value of 0.932. With variations in the training set proportion, our model exhibited outstanding performance and robustness across evaluation metrics.

In this study, there are also some limitations, such as model computational complexity and resource constraints, diversification of application scenarios, and expansion of real-world applications. Addressing these limitations will be a key focus of our future work, and we will strive to resolve the following issues:

- We will optimize the model structure and accelerate model training to reduce the computational complexity and resource constraints of the model.
- We are considering deploying the optimized model on a Raspberry Pi. The deployment involves communication with smart meters to collect users' electricity-consumption information and automatically detect electricity theft. If the model is deployed, it can be enhanced with Intel's Neural Compute Stick. It is indeed feasible to further optimize the proposed model and integrate it into practical engineering applications.

- We will deploy Raspberry Pi embedded with the optimized model into various real-world scenarios, further investigating the model's generalization ability and robustness in different contexts as proposed in this paper.

**Author Contributions:** Methodology, Q.H., M.Y. and T.J.; Software, X.W., M.H. and F.L.; Validation, X.W.; Formal analysis, M.Y.; Investigation, F.L.; Resources, M.H.; Writing—original draft, Q.H.; Writing—review & editing, Q.H.; Supervision, Z.T. and T.J.; Project administration, Z.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/henryRDlab/ElectricityTheftDetection.

**Conflicts of Interest:** Author Zhenli Tang, Xiaofeng Weng, Min He, Fang Liu were employed by the company Fujian YILI Information Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Xia, X.; Yang, X.; Liang, W.; Cui, J. Detection Methods in Smart Meters for Electricity Thefts: A Survey. *Proc. IEEE* **2022**, *110*, 273–319. [CrossRef]
2. Ahmad, T.; Chen, H.; Wang, J.; Guo, Y. Review of various modeling techniques for the detection of electricity theft in smart grid environment. *Renew. Sustain. Energy Rev.* **2018**, *82*, 2916–2933. [CrossRef]
3. Andrey, P.; Firuz, K.; Pavel, Y.G.; Murodbek, S.; Vladislav, S.; Nikita, M.; Ismoil, O.; Inga, Z. Data-Driven Machine Learning Methods for Nontechnical Losses of Electrical Energy Detection: A State-of-the-Art Review. *Energies* **2023**, *16*, 7460.
4. Sajad, A.; Min, Y.; Wajid, A. Prevention and Detection of Electricity Theft of Distribution Network. *Sustainability* **2023**, *15*, 4868.
5. Erika, S.; Antonello, R.; Gianfranco, D.; Massimo, P.; Rodolfo, A. Systematic review of energy theft practices and autonomous detection through artificial intelligence methods. *Renew. Sustain. Energy Rev.* **2023**, *184*, 11544.
6. Xia, R.; Gao, Y.; Zhu, Y.; Gu, D.; Wang, J. An Efficient Method Combined Data-Driven for Detecting Electricity Theft with Stacking Structure Based on Grey Relation Analysis. *Energies* **2022**, *15*, 7423. [CrossRef]
7. Wang, Y.; Chen, Q.; Hong, T.; Kang, C. Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges. *IEEE Trans. Smart Grid.* **2019**, *10*, 3125–3148. [CrossRef]
8. Chen, J.; Nanehkaran, Y.; Chen, W.; Liu, Y.; Zhang, D. Data-driven intelligent method for detection of electricity theft. *Int. J. Electr. Power Energy Syst.* **2023**, *148*, 108948. [CrossRef]
9. Zheng, K.; Chen, Q.; Wang, Y.; Kang, C.; Xia, Q. A Novel Combined Data-Driven Approach for Electricity Theft Detection. *IEEE Trans. Ind. Inf.* **2019**, *15*, 1809–1819. [CrossRef]
10. Liao, W.; Yang, Z.; Bak-Jensen, B.; Pillai, R.; Krannichfeldt, L.; Wang, Y.; Yang, D. Simple Data Augmentation Tricks for Boosting Performance on Electricity Theft Detection Tasks. *IEEE Trans. Ind. Appl.* **2023**, *59*, 4846–4858. [CrossRef]
11. Zheng, Z.; Yang, Y.; Niu, X.; Dai, H.; Zhou, Y. Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids. *IEEE Trans. Ind. Inf.* **2018**, *14*, 1606–1615. [CrossRef]
12. Nagi, J.; Yap, K.; Tiong, S.; Ahmed, S.; Mohamad, M. Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines. *IEEE Trans. Power Delivery.* **2010**, *25*, 1162–1171. [CrossRef]
13. Yan, Z.; Wen, H. Electricity Theft Detection Base on Extreme Gradient Boosting in AMI. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–9. [CrossRef]
14. Hasan, M.N.; Toma, R.N.; Nahid, A.-A.; Islam, M.M.M.; Kim, J.-M. Electricity Theft Detection in Smart Grid Systems: A CNN-LSTM Based Approach. *Energies* **2019**, *12*, 3310. [CrossRef]
15. Jindal, A.; Dua, A.; Kaur, K.; Singh, M.; Kumar, N.; Mishra, S. Decision Tree and SVM-based Data Analytics for Theft Detection in Smart Grid. *IEEE Trans. Ind. Inf.* **2016**, *12*, 1005–1016. [CrossRef]
16. Haq, E.; Huang, J.; Xu, H.; Li, K.; Ahmad, F. A hybrid approach based on deep learning and support vector machine for the detection of electricity theft in power grids. *Energy Rep.* **2021**, *7*, 349–356. [CrossRef]
17. Mohammad, F.; Saleem, K.; Al-Muhtadi, J. Ensemble-Learning-Based Decision Support System for Energy-Theft Detection in Smart-Grid Environment. *Energies* **2023**, *16*, 1907. [CrossRef]
18. Tian, C.; Ma, J.; Zhang, C.; Zhan, P. A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network. *Energies* **2018**, *11*, 3493. [CrossRef]
19. Zheng, X.; Bai, F.; Zhuang, Z.; Chen, Z.; Jin, T. A new demand response management strategy considering renewable energy prediction and filtering technology. *Renewable Energy* **2023**, *211*, 656–668. [CrossRef]
20. Liu, Y.; Yuan, D.; Fan, H.; Jin, T.; Mohamed, M.A. Multidimensional Feature-Driven Ensemble Model for Accurate Classification of Complex Power Quality Disturbance. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1501613. [CrossRef]

21. Liu, Y.; Jin, T.; Mohamed, M. A novel dual-attention optimization model for points classification of power quality disturbances. *Appl. Energy* **2023**, *339*, 121011. [CrossRef]
22. Wang, Y.; Jin, S.; Cheng, M. A Convolution–Non-Convolution Parallel Deep Network for Electricity Theft Detection. *Sustainability* **2023**, *15*, 10127. [CrossRef]
23. Massaferro, P.; Martino, J.M.D.; Fernández, A. Fraud Detection on Power Grids While Transitioning to Smart Meters by Leveraging Multi-Resolution Consumption Data. *IEEE Trans. Smart Grid* **2022**, *13*, 2381–2389. [CrossRef]
24. 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 Trans. Inf. Forensics Secur.* **2023**, *18*, 2553–2568. [CrossRef]
25. Takiddin, A.; Ismail, M.; Nabil, M.; Mahmoud, M.M.E.A.; Serpedin, E. Detecting Electricity Theft Cyber-Attacks in AMI Networks Using Deep Vector Embeddings. *IEEE Syst. J.* **2021**, *15*, 4189–4198. [CrossRef]
26. Chen, Z.; Meng, D.; Zhang, Y.; Xin, T.; Xiao, D. Electricity Theft Detection Using Deep Bidirectional Recurrent Neural Network. In Proceedings of the 2020 22nd International Conference on Advanced Communication Technology (ICACT), Phoenix Park, Republic of Korea, 16–19 February 2020; pp. 401–406.
27. Lea, C.; Vidal, R.; Reiter, A.; Hager, G.D. Temporal Convolutional Networks: A Unified Approach to Action Segmentation. In Proceedings of the Computer Vision—ECCV 2016 Workshops, Pt III, Amsterdam, The Netherlands, 8–10, 15–16 October 2016; Volume 9915, pp. 47–54.
28. Stergiou, A.; Poppe, R.; Kalliatakis, G. Refining activation downsampling with SoftPool. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 10337–10346.
29. Khan, I.; Javeid, N.; Taylor, C.; Gamage, K.; Ma, X. A Stacked Machine and Deep Learning-Based Approach for Analysing Electricity Theft in Smart Grids. *IEEE Trans. Smart Grid* **2022**, *13*, 1633–1644. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.