IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

TOWARDS SUSTAINABLE ENERGY EFFICIENCY WITH INTELLIGENT ELECTRICITY THEFT DETECTION IN SMART GRIDS EMPHASISING ENHANCED NEUTRAL NETWORKS

Kartik Jain

This paper proposes a novel convolution–non-convolution parallel deep network (CNCP)-based method for electricity theft detection. First, the load time series of normal residents and electricity thieves were analyzed and it was found that, compared with the load time series of electricity thieves, the normal residents' load time series present more obvious periodicity in different time scales, e.g., weeks and years; second, the load times series were converted into 2D images according to the periodicity, and then electricity theft detection was considered as an image classification issue; third, a novel CNCP-based method was proposed in which two heterogeneous deep neural networks were used to capture the features of the load time series in different time scales, and the outputs were fused to obtain the detection result. Extensive experiments show that, compared with some state-of-the-art methods, the proposed method can greatly improve the performance of electricity theft detection.

Keywords:

Smart grid; electricity theft detection; deep learning; parallel deep network

1. Introduction

Electricity theft is the use of various methods to reduce the payment of electricity charges and achieve illegal benefits. Electricity theft detection is of great significance to ensure the safety and stability of a power system, and, at present, much attention has been paid to the modeling and detection of electricity theft [1]. Electricity theft detection can be completed by adding special hardware on the user side; however, this would significantly increase the system and maintenance cost. Meanwhile, electricity thieves can also avoid detection by tampering with the hardware. Therefore, currently, most efforts that have been made on electricity theft detection are based on electricity consumption data analysis [2,3].

Electricity theft detection is a very challenging issue as there are many ways to steal electricity; meanwhile, electricity theft may not occur continuously [4,5]. Based on the different detection principles, current electricity theft detection methods can be divided into two types: short-period-comparison-based electricity theft detection methods and whole-period-classification methods. Short-period-comparison-based electricity theft detection methods generally adopt anomaly detection methods based on a sliding time period window, and Table 1 gives some typical short-period-comparison-based electricity theft detection methods. The main assumption is that the electricity theft of an electricity thief only occurs in parts of periods, and that the electricity theft behaviors would change the distribution of electricity consumption data significantly. Therefore, for the beginning of the electricity consumption profile, if some features of electricity consumption data in the current time period window are obviously different from those in the previous time period window, it can be considered that electricity theft has occurred. For example, in [6,7], if the mean value and variance of the data in the current time period window change significantly compared with the previous window, it is considered that there is electricity theft. In [8,9], autoencoders are trained using the normal data, and if the reconstruction error of the data in the current time period window on the trained autoencoder is abnormally large, it is considered that there is electricity theft. In [10,11], if the electricity consumption data in the current time period window do not belong to the same cluster as the data in the previous time period window, it is considered that there is electricity theft. In [12], if the correlation between the electricity consumption data in the current time period window and in the previous window is low, it is considered that there is electricity theft. Jokar et al. [13] modeled a variety of typical electricity theft patterns, and if the electricity consumption data in the current time period window conform to a certain theft pattern, it is considered that there is electricity theft. The advantage of the short-period-comparison-based methods is that they can identify the occurrence time of electricity theft. However, due to the complexity of the electricity consumption data of normal users, as well as the diversity of electricity theft methods, electricity theft data and normal consumption data can generally only be effectively distinguished using long-time-period data. Because the features used in short-period-comparison-based methods may not change significantly in the actual scenes, such methods are usually verified on simulated datasets.

Due to the above reasons, current research mainly focuses on electricity theft detection based on the whole period data, which can be regarded as a binary classification problem. According to the classification process, whole-period-classification methods can be divided into two types: two-stage methods and one-stage methods, give some state-of-the-art two-stage and one-stage whole-period-classification methods, respectively. The two-stage methods consist of two procedures: feature extraction and feature-based classification [14]. For example, ref. [15] uses seven networks for feature extraction and one discrimination network for classification. Ref. [16] uses breakout detection for feature extraction and an SVM for classification. Ref. [17] uses six feature extraction methods for feature extraction and XGBoost for classification. Ref. [18] uses discrete wavelet transform for feature extraction and random undersampling boosting for classification. Ref. [19] is a special two-stage detection method. It uses the decision tree for feature extraction, but inputs all the original data together with the result of the decision tree into an SVM for theft detection; therefore it can be regarded as a 1.5-stage detection method. The procedure of feature extraction affords the two-stage whole-period-classification methods better interpretability; however, in addition to the performances of feature extraction algorithms and electricity theft classification algorithms, the effectiveness of the extracted features also determines the accuracy of theft detection [20]. Different from two-stage methods, one-stage methods identify electricity theft based on the data directly. Many one-stage methods employ a single classifier, such as support vector description [21], LightGBM [22,23], XGBoost [23,24], long short-term memory (LSTM) [25], deep vector embedding [26], Bayesian risk framework [27], linear regression [28], the black hole algorithm [29], random forest [30], feed forward neural networks [31], TripleGAN [32], etc. Note that [30,31,32] use a stacked autoencoder, deep autoencoder, and relational autoencoder for features extraction, and then use a random forest, feed forward neural networks, and TripleGAN for classification. However, because the main function of the autocoder is dimensionality and noise reduction, which can be regarded as a data preprocessing process, [30,31,32] are still regarded as one-stage algorithms in this paper. Meanwhile, some of the one-stage methods fuse a multi-classifier to further improve the performance of the detection methods. For example, ref. [33] uses LSTM and multi-layer perceptrons (MLPs), ref. [34] uses the maximum information coefficient and fast search and find of density peaks, and [35] uses a back propagation neural network (BP) and convolutional neural networks (CNNs). Compared with two-stage detection methods, one-stage whole-period-classification algorithms avoid the possibility of reducing the detection accuracy caused by inappropriate feature extraction. However, as end-to-end methods, the interpretability of one-stage whole-period-classification algorithms is reduced, and the generalization ability of the algorithm may decline.

Although currently there are many electricity theft detection methods, because of the difficulties in collecting large-scale real electricity theft data, most of these methods use simulated theft data generated according to the specific electricity theft pattern. Due to the lack of real electricity theft data, few of these works fully consider the features of normal and theft data, and most of them regard electricity theft detection as a general classification or abnormal detection problem, which limits the performance of the detectors.

This paper proposes a novel one-stage whole-period-classification method: a convolution–non-convolution parallel deep network (CNCP)-based method for electricity theft detection. First, similar to one-stage electricity theft detection algorithms, the proposed CNCP method avoids the possibility of reducing the detection accuracy caused by inappropriate feature extraction. Second, in some previous studies [36,37], it was found that, influenced by work statuses and living habits, the electricity consumption of normal residents usually fluctuates on a one-week cycle. Meanwhile, through data analysis in Section 2.3 and Section 2.4, it can be seen that the periodic feature of the electricity consumption of electricity theft users is weaker than those of normal users, and, after converting temporal data into 2D image data based on a certain period, such a difference becomes more obvious. On this basis, the proposed CNCP structure can effectively capture the features of the electricity consumption time series locally and globally simultaneously, and can therefore obtain much better detection results compared with state-of-the-art methods. The contribution of this paper can be summarized as follows:

The features of the electricity consumption time series of normal users and electricity thieves were analyzed, and it was found that compared with those of electricity thieves, the electricity consumption time series of normal users have more obvious periodic characteristics;

A novel CNCP-based method for electricity theft detection was proposed in which two heterogeneous deep neural networks were used to capture the features of the electricity consumption time series in different time scales effectively, therefore yielding better detection results.

The rest of this paper is organized as follows: Section 1 gives the related works, Section 2 gives the data set description and the data preprocessing method, Section 3 first presents the electricity consumption features of normal users and electricity thieves and then the proposed CNCP-based electricity theft detection method, and Section 4 and Section 5 are the experiments and the conclusions.

2. Materials and Methods

2.1. Dataset

In Table 1, Table 2 and Table 3, typical state-of-the-art electricity theft detection methods are summarized. It can be seen that, although currently there are many electricity theft detection methods, most of them use simulated theft data, and the available public real electricity theft dataset is the dataset released by the State Grid Corporation of China (SGCC). The reason comes from the fact that electricity data usually have strong privacy; meanwhile, it is very difficult to collect large-volume real theft data. For the study on electricity theft detection, although theft data can be simulated according to some specific features of electricity theft behaviors, these simulated data are far away from the real theft data.

In this paper, the SGCC dataset was used for electricity theft detection [38]. The SGCC dataset contains daily total electricity consumption time series of 42,373 users from 1 January 2014 to 1 October 2016, totalling 1036 days. In this dataset, 3615 are electricity thieves confirmed by on-site electricians. Each electricity consumption time series is labeled by 1 or 0 to indicate if it is from an electricity thief or not. Note that there is no label to indicate when electricity theft occurs or stops. Meanwhile, no other information is available, such as the location of the user, weather information, etc. Electricity theft detection on this dataset can be considered as a binary classification problem. It is the determination of whether a resident is an electricity thief or not only based on the daily total electricity consumption time series during the whole data collection period.

2.2. Data Pre processing

In the dataset, there are some 0 or invalid elements, which are brought in by the failure of data collection and transmission. In this paper, first, 0 or the invalid elements in each user's load profile are counted. When its ratio to the whole load profile of the specific user is greater than 30%, this load profile is discarded. Through this operation, 33,130 valid electricity consumption time series are retained, in which 2045 (6.17%) are from electricity thieves.

2.3. Analysis of Electricity Consumption Features of Normal Residents and Electricity Thieves

To understand the difference between normal residents and electricity thieves, the load profiles of normal residents and electricity thieves are clustered, respectively, to obtain the main consumption patterns of them. In detail, for the electricity consumption time series of normal residents, first the k-means method was adopted for data clustering using the k-means function in Matlab. As the number of patterns is unknown, we gradually increased the number of clusters until similar patterns were divided into two clusters. Finally, the number of clusters was set to 10, and the two clusters with smallest number of samples were considered as noises and ignored.

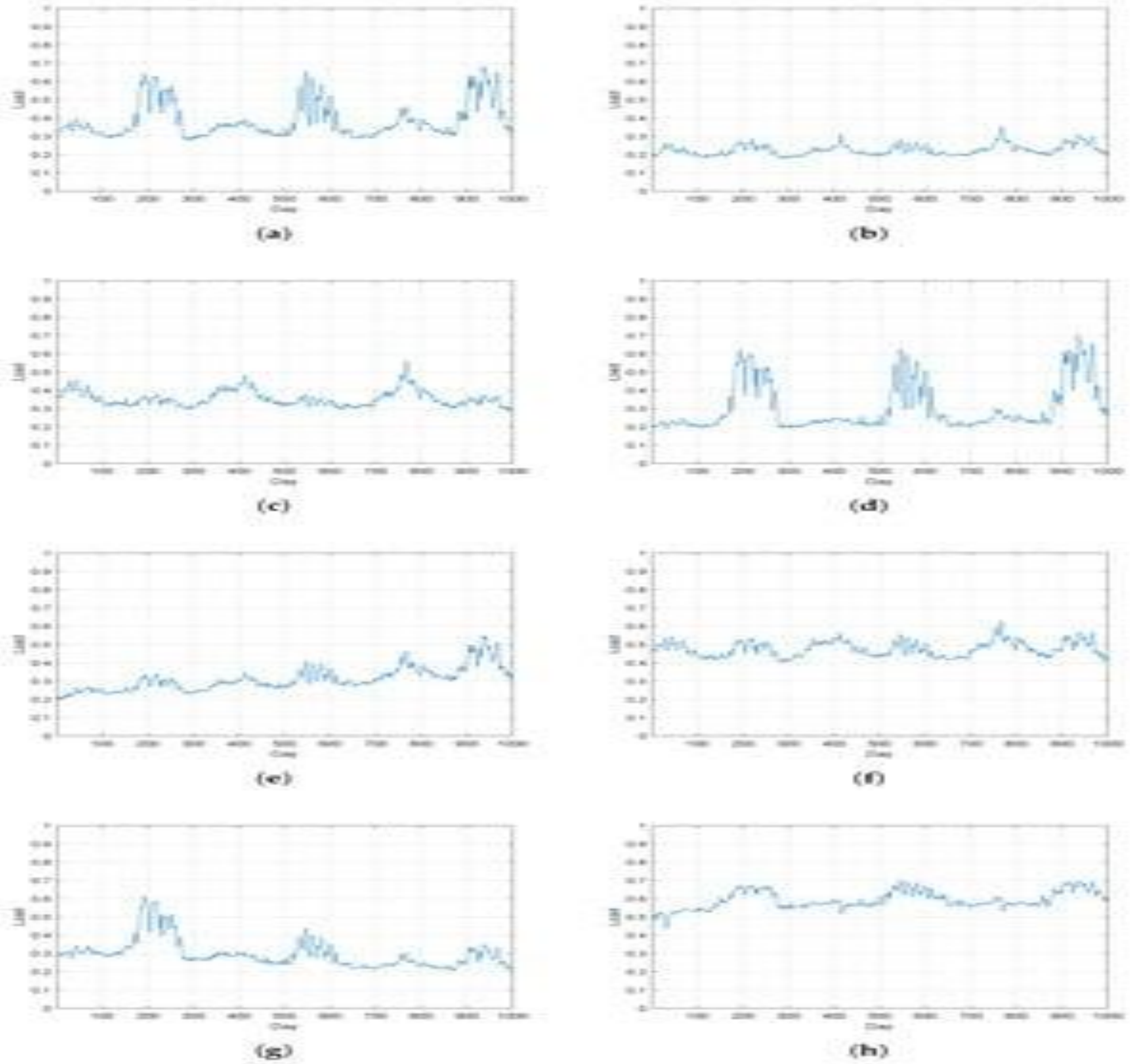


Figure 1. (a–h) Load main patterns of normal residents after clustering.

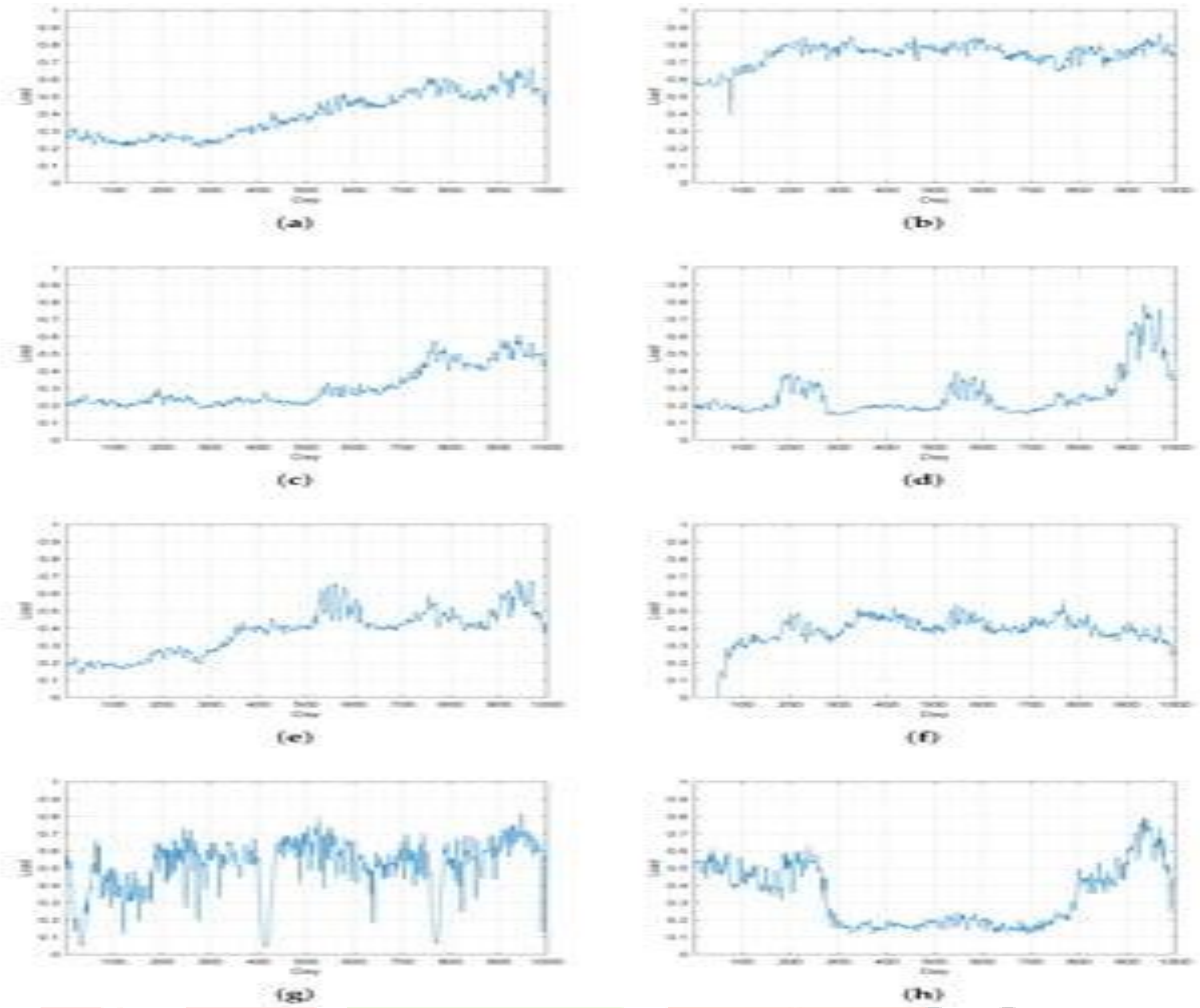


Figure 2. (a–h) Load main patterns of electricity thieves after clustering.

As shown in Figure 1, for normal residents, all load patterns present obvious periodicity, and are closely related to the outdoor temperature. Most of the patterns are relatively stable, and only a few of them show an overall upward trend. However, the trend is smooth without sudden change. For example, the peaks and valleys of pattern (a), pattern (b), pattern (c), pattern (e) and pattern (f) are stable and consistent with the outdoor temperature. And for pattern (d), pattern (g) and pattern (h), although the overall trends of them are upward, the peaks and valleys are still consistent with the outdoor temperature, indicating that these users mainly use electric equipment such as air conditioners to adjust the room temperature in summer.

However, as shown in Figure 2, for electricity thieves, most of the patterns have little correlation with the outdoor temperature, and some of them change dramatically. For example, pattern (a), pattern (c), and pattern (d) increase dramatically after day 500. Pattern (b), pattern (g), and pattern (h) are very complex and have little relationship with the outdoor temperature.

From the above analysis, it is clear that, compared with the load time series of electricity thieves, the normal residents' load time series present more obvious periodicity.

2.4. A Novel CNCP-Network-Based Detection Method

As shown in Figure 1 and Figure 2, as well as the analyses in Section 1, normal residents and electricity thieves would be classified effectively from the viewpoint of image classification. Therefore, in this paper, first, the original electricity consumption time series were converted into 2D images according to the periodicity, and then a specific image classification method was designed to distinguish normal residents and electricity thieves. Figure 3 and Figure 4 are the converted images in accordance with Figure 1 and Figure 2. In detail, first, the original electricity consumption time series (1 × 1036) was reshaped to a 37 × 28 matrix using the reshape function in Matlab, and then the matrix was rescaled into the range of 0 and 255: I. The advantages of the set of the size of the matrix are that, first, the corresponding image directly reflects the periodicity of the time series as each row of the matrix consists of successive four-week consumption data; second, the width and the height of the matrix are close to each other, which would be beneficial to the performance of the image classification method. From Figure 3 and Figure 4, it can be seen that, for normal residents, intensities of the images show relative periodicity. However, as shown in Figure 4, for electricity thieves, such periodicity is much weaker and the distributions of intensities of images are more complex.

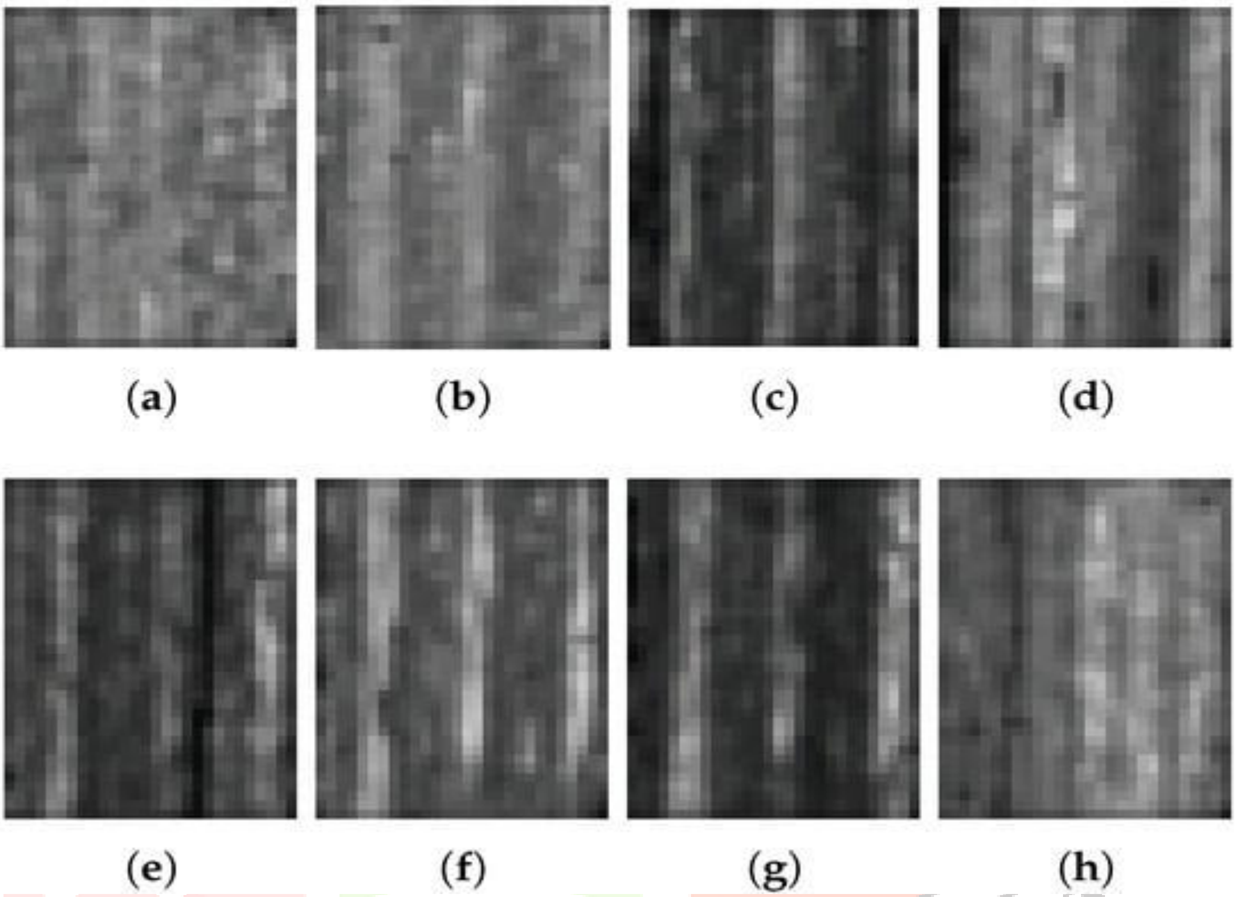


Figure 3. (a–h) Images of loads of normal residents.

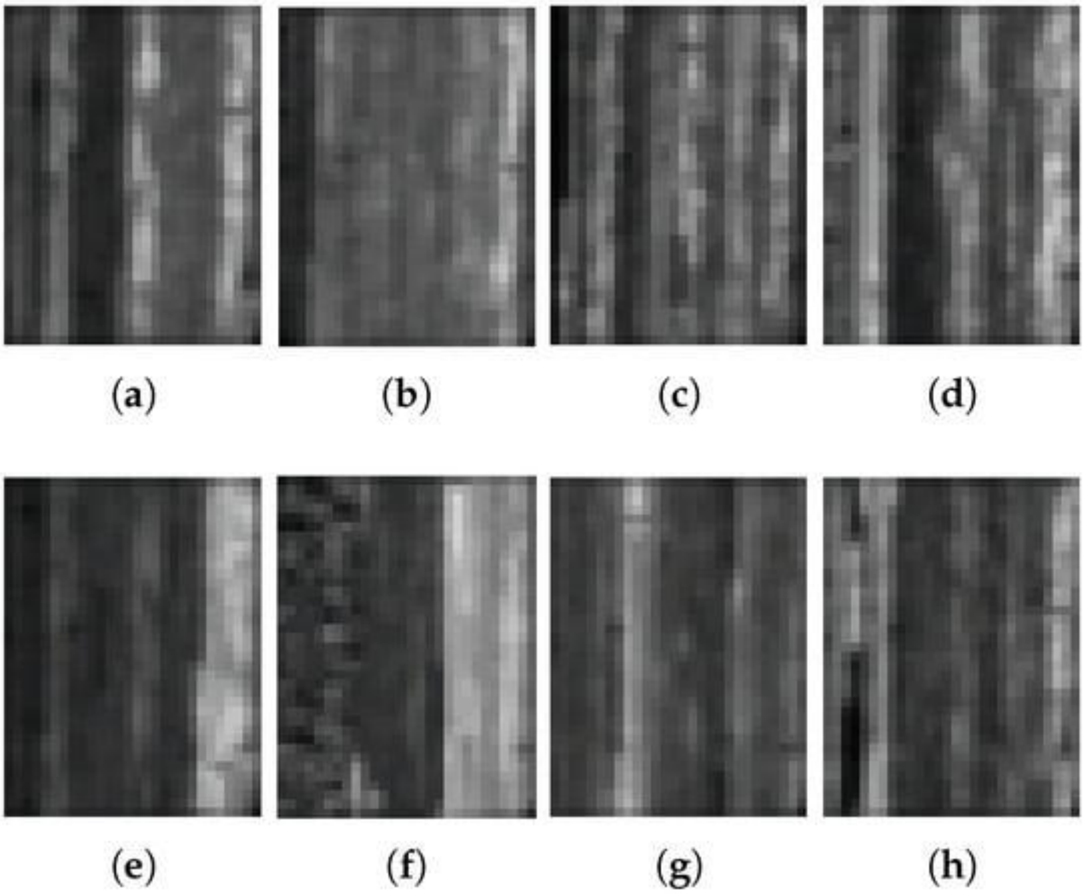


Figure 4. (a–h) Images of loads of electricity thieves.

Second, a novel parallel-network-based electricity theft detection method was proposed for electricity theft detection that consists of a convolution deep branch and a non-convolution deep branch, and the results from the two branches were finally fused by a fully connected layer and a softmax layer. The overall structure of the proposed detection method can be seen in Figure 5.

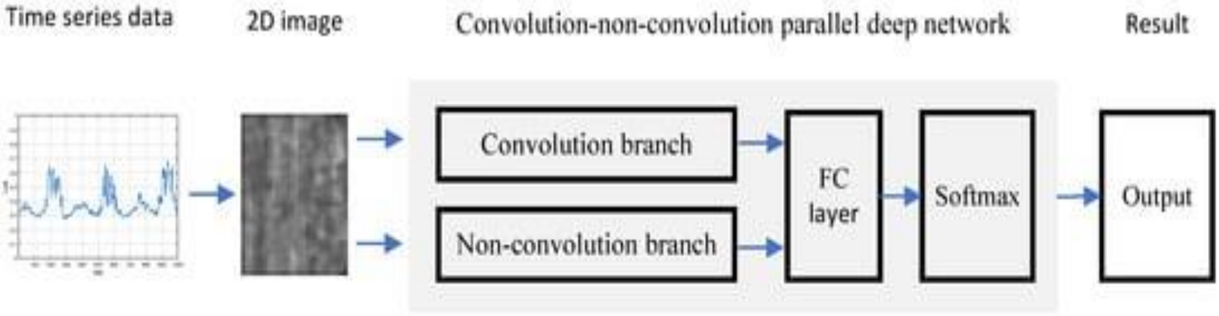


Figure 5. The overall structure of the proposed detection method.

To determine an optimal structure for the proposed method, a series of backbone networks were tested, which are listed in Table 4. This paper aimed to examine the impact of networks with different depths on classification results in order to select a network suitable for the problem of electricity theft detection. Through experiments, it was found that VGG series networks [39] have good performance for the convolutional branch. Therefore, this paper selected three types of VGG networks, namely VGG11, VGG13, and VGG16, where the number represents the depth of the network. Meanwhile, we designed fully-connected-layers-based networks with different depths for the non-convolutional branch. These networks are composed of fully connected layers and softmax layers, and the neurons in the fully connected layer increase with the number of layers. An example structure of the proposed detection method based on C1-N1 is given in Figure 6. As shown in Figure 6, there are two branches in the proposed method: one is a convolution branch, and the other is a non-convolution branch. In Figure 6, the convolution branch adopts the VGG11 structure, which consists of 7 convolutional layers, a gmpool layer, a fully connected (FC) layer, and a softmax layer. Each convolutional layer is followed by a batchnorm block and a tanh function. Meanwhile, two maxpool layers are inserted between the second and the third convolutional layers, as well as the fourth and the fifth convolutional layers, respectively. The non-convolution branch consists of two FC layers, a relu block, and a softmax layer. Finally, the two branches are fused by a FC layer, and a softmax layer is used as the output. The parameters of all blocks can be found in Figure 6.

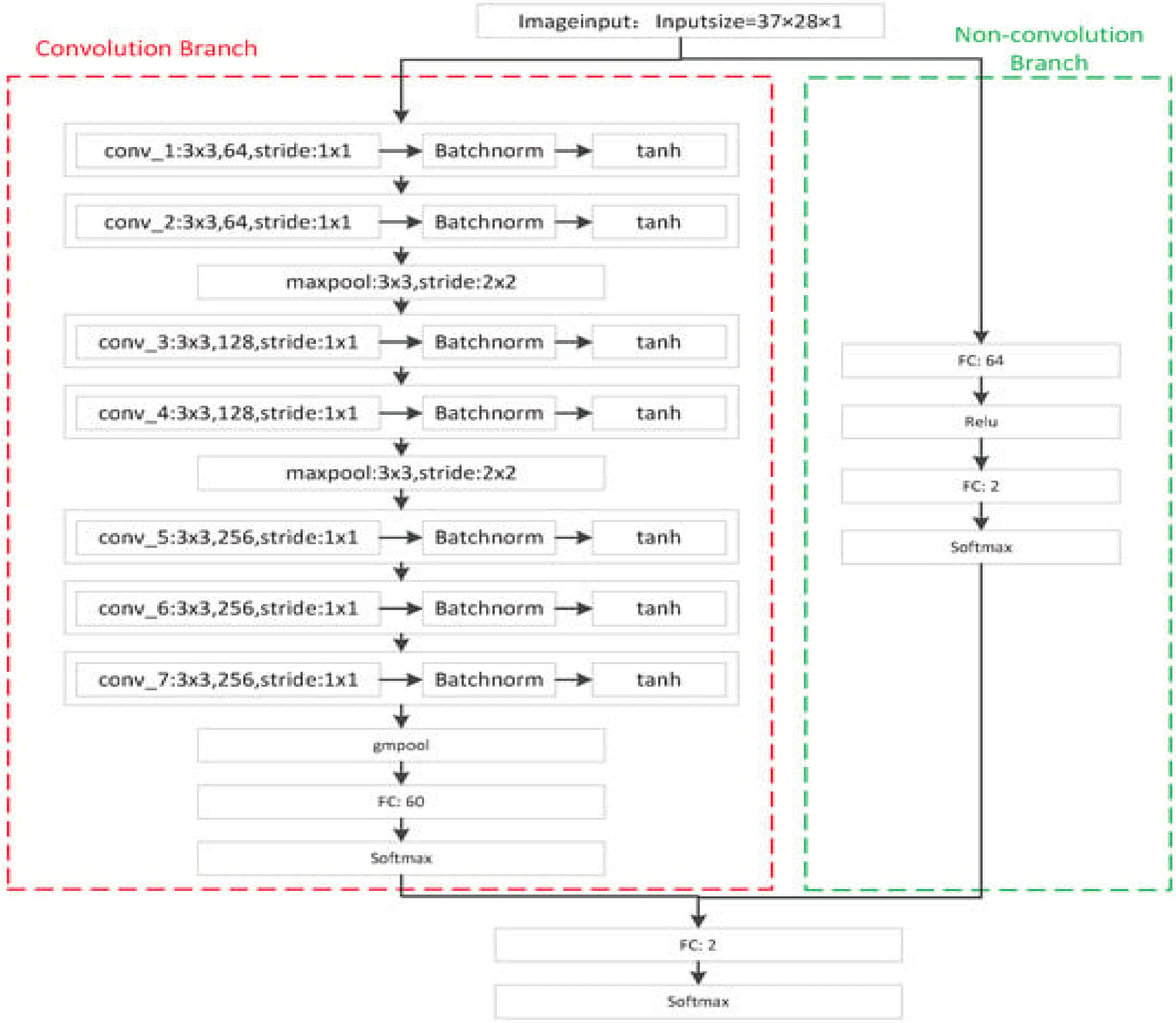


Figure 6. An example structure of the proposed method based on C1-N1.

The proposed method mainly consists of convolutional layers, pooling layers, FC layers, and the loss function. Data forward propagation between convolutional layers can be expressed by:

During the error back–forward propagation between convolutional layers process, as the parameter-updating rules of neural network are based on gradient descent methods, the gradient of any parameter can be obtained by:

2.5. Solution of Imbalance Data

After the pre-processing step in Section 3, in the whole dataset, 33,130 valid load time series remain, including 2045 (6.17%) samples of electricity theft. If using these load time series to train a classifier, the bias will be great as the items in different classes are extremely imbalanced. In this paper, a re-sampling method was designed to balance the items in different classes. In Section 3, eight main patterns of electricity thieves are obtained. To generate a new electricity thief load time series, first, a pattern P of electricity theft was randomly selected.

The above sample generation process was repeated until the number of load time series of electricity theft was equal to that of the normal residents.

3. Results

In this section, two groups of experiments are carried out. In the first group of experiments, a series of backbone networks were adopted in the proposed convolution–non-convolution parallel deep network method in order to obtain a optimal structure. In the second group of experiments, the proposed method was compared with several widely used and state-of-the-art methods. During the training process, 80% of the samples were randomly selected as the training set, and the rest were used as the test set, and the total number of training iterations was 7000. The experiments were carried out on a server with rtx2080ti, and the development environment was m. Four performance indicators were recorded, including the true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

3.1. Optimal Structure Determination

In this section, to obtain optimal structure, a series of backbone networks are tested in the proposed method, which are shown in Table 4. During the training process, binary cross entropy was used as the loss function, and He-initialization was used as the weights initializer; the optimizer was sgdm, and its momentum was 0.9; L2 regularization was 0.0001, the batch size was 128, the initial learning rate, drop factor, and drop period were 0.001, 0.1, and 10, respectively; and when the accuracy of the verification set did not increase within the last five epochs, the training was stopped. The results are shown in Table 5. In Table 5 and the tables in rest of this paper, the red color is used to indicate the best result of the corresponding evaluating indicator.

Table 5. Experiment results on different backbone structures in the proposed method.

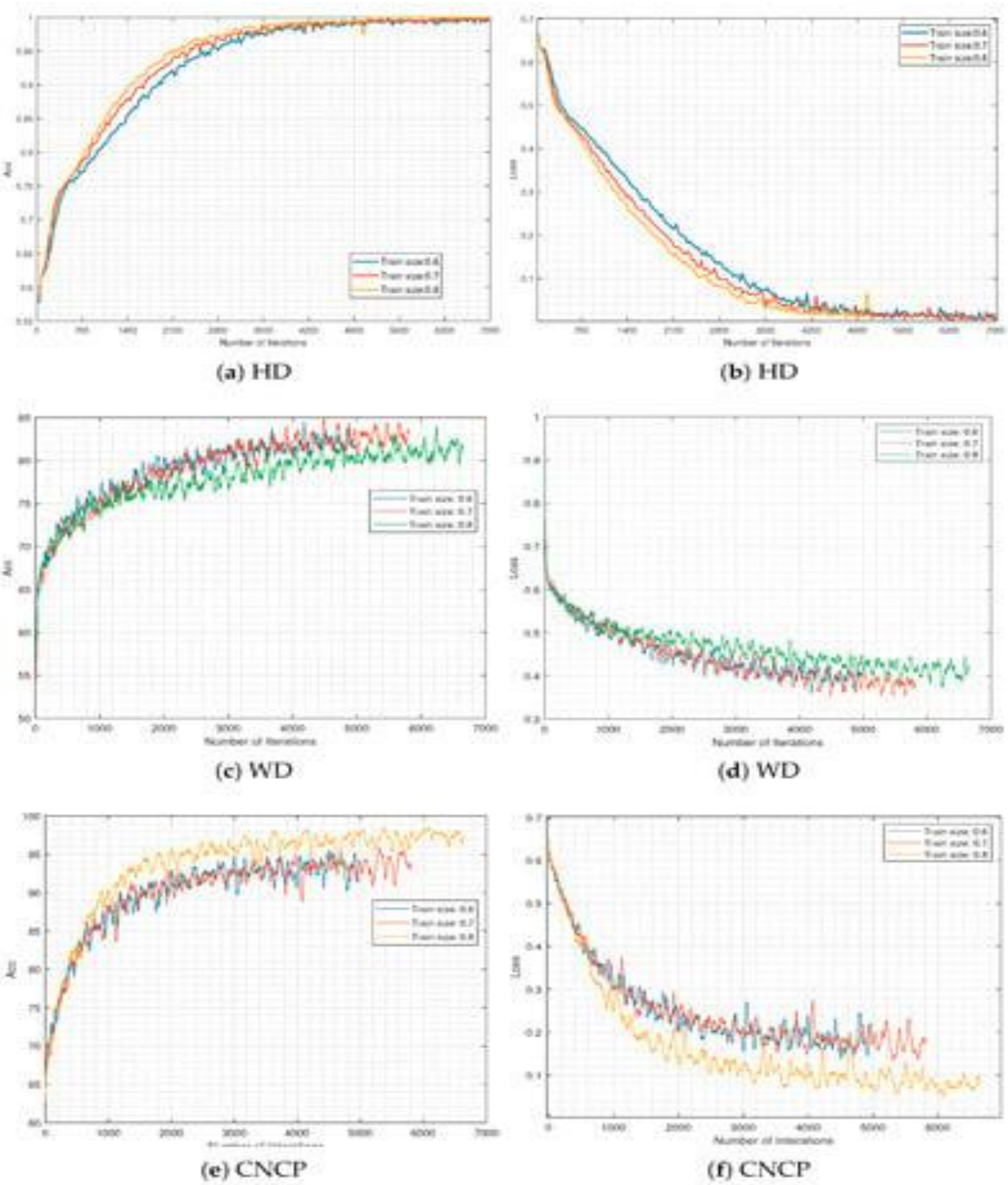
3.2. Comparison with Other Methods

In this section, some typical methods, such as SVM, logistic regression (LR), random forest (RF), extreme gradient boosted trees (XGBoost) [24], the wide-and-deep method (WD) [35], and the hybrid deep neural networks method (HD) [33] are tested as comparisons. In these methods, the SVM and LR are classical and widely used classification methods, RF and XGBoost are currently popular classification methods and win many competitions, and WD and HD are state-of-the-art methods that consist of a parallel network similar to the proposed method in this paper.

3.2.3. Parameters Setting of HD

The HD method also consists of two branches: one is an LSTM network, and the other is an MLP network. The input of LSTM is the weekly consumption data, the number of zero values, the number of missing values, and the seasonal index value. The input of the MLP network is a series of attribute information, including user information (geographical location and province), meter information (geographical location, firmware version, production year, etc.), and economic activity code. Because attribute information is usually private and not public, there was no such information in the database used in this paper. As a result, the MLP network in the HD method was ignored in the experiment. According to the suggestion in [33], the number of network layers of LSTM was set to 4, each layer contained 512 neurons, and the output layer contained a dropout layer with a parameter of 0.3. During the training process, binary cross entropy was used as the loss function; the optimizer was Adam, its one-order parameters $\beta 11$ and $\beta 22$ were 0.9 and 0.999, respectively, and the initial learning rate was 0.001; L2 regularization was 0.0001, and the batch size was 128; and when the accuracy of the verification set did not increase within the last five epochs, the training was stopped.

To better demonstrate the performance of the proposed method, in addition to the experiments in the previous section, we carried out extra experiments in which the ratio of the training data set to test data set were 6:4 and 7:3, respectively. The training processes are given in Figure 7, and the results are given in Table 6. We also give the area under the precision–recall curve (PR-AUC) and the area under the receiver operating characteristic curve (ROC-AUC) in Figure 8, as was carried out in [33], to show the relationship between precision and recall, as well as TP and FP.

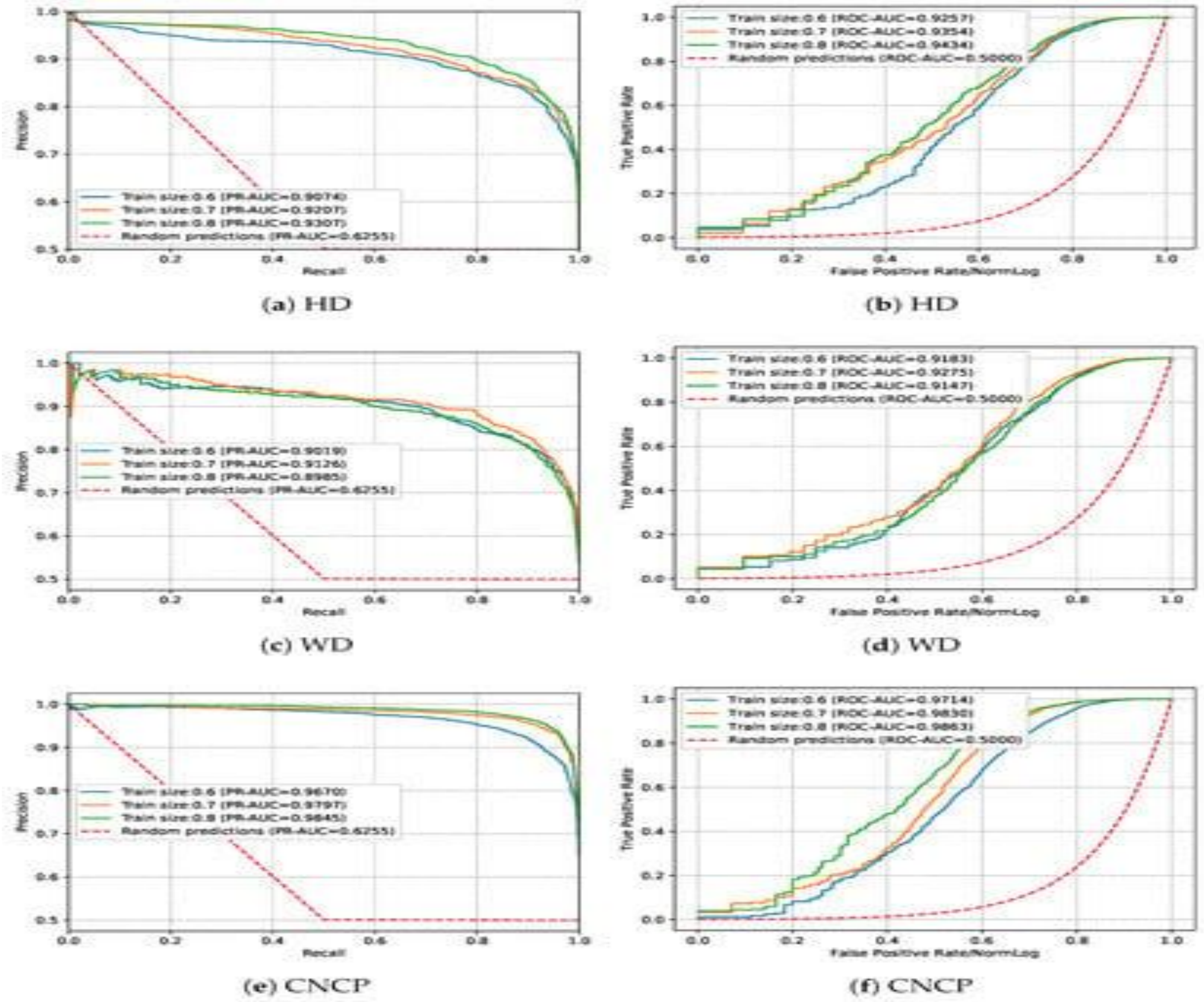


(a) HD (b) HD

(c) WD (d) WD

(e) CNCP (f) CNCP

Figure 7. The training processes.



(a) HD (b) HD

(c) WD (d) WD

(e) CNCP (f) CNCP

Figure 8. Comparisons with other methods.

Table 6. Comparisons with other methods.

Table 6 gives the test results, and Table 7 gives the sorted performance rank of nine methods, where HD-S is the result of HD with inputs 37 × 28 (e.g., inputs are four-week data), and CNCP-L is the result of the proposed method with input 148 × 7. These two methods were used to test the impact of the size of the input on the detection results. From Table 6 and Table 7, it can be seen that, in all tests, the SVM has the worst precision and random forest has the worst recall and F1-score. Meanwhile, the proposed method has the best precision, recall, and F1-score in all tests. Note that the performance of WD ranks second, indicating that the parallel network structure indeed improves the performance of the classification method; meanwhile, the performance of HD ranks third, indicating that, for time series data, the performance of the LSTM network is better than other single classifiers in this paper. Meanwhile, although the SVM has the worst precision, its recall ranks third, and its F1-score is in the middle of all nine methods. Therefore, when the performance requirement of the algorithm is not high and the complexity of the algorithm needs to be relatively low, the SVM may be a suitable algorithm. From Table 6, it can be seen that the result of HD-S is slightly lower than but similar to that of HD with the 148 × 7 matrix. The reason is that, for the LSTM method with the input of the 37 × 28 matrix, the input data can be considered as 37 × 28 vectors, meaning that its inputs are fewer but longer than the LSTM method with the 148 × 7 matrix. Although a longer input can benefit the LSTM method, fewer samples simultaneously reduce the performance of it. Meanwhile, we also tested the performance of the proposed CNCP-based method with inputs 148 × 7, which is labeled as CNCP-L. It can be seen that the performance of CNCP-L is lower than that of the CNCP. This may come from the fact that, for the CNN network, the 37 × 28 matrix has a smaller padding area than the 148 × 7 matrix during the convolution computing process, and can therefore yield better results.

3.3. Complexity Analysis

For deep neural networks, the computational complexity is usually presented by the total amount of floating-point operations per second (FLOPs) when processing a sample. The CNCP network proposed in this paper is mainly composed of a convolutional layer and fully connected layer. The FLOPs of a convolutional layer can be calculated by Equation (22):

The CNCP network proposed in this paper adopts the structure of Conv-BN-Activation, the calculation of a BN layer is included in the calculation of the convolutional layer, and the FLOPs of the BN layer are negligible relative to the convolutional layer. Table 8 gives the computational complexity of the proposed CNCP network with 3×33×3 kernels, and Table 9 gives the computational complexity of the proposed CNCP network with different kernels.

4. Discussion

The advantage of the proposed method comes from the CNCP structure, which can capture features of electricity consumption time series at different scales. From Figure 2, Figure 3, Figure 4 and Figure 5, it can be seen that the normal electricity consumption time series has more significant periodic characteristics than the electricity theft time series. The algorithm proposed in this paper includes two branches, namely the convolution branch and the non-convolution branch. Using the data reconstructed based on the periodic characteristics of normal electricity consumption time series, if the periodic distribution of the electricity theft time series is significantly different from that of the normal electricity consumption time series, the convolution branch can realize effective theft detection. When the periodic distribution difference between theft data and normal data is not significant, the efficiency of the convolution branch is reduced. However, because the first layer of the non-convolution branch is the fully connected layer, the input data are globally mapped; that is, the classification is based on the global information of the data, which overcomes the problem of the low efficiency of the convolution branch so as to effectively improve the detection accuracy. For other time series analysis algorithms such as LSTM, it can be seen from the experiment section of this paper that the detection performance is lower than the proposed method. This is due to the fact that, for the LSTM model, according to the analysis in Section 2.3 and Section 2.4, the main difference between the electricity consumption of electricity theft users and that of normal users is their periodicity significance. LSTM can infer this periodic difference through extensive training. However, in this proposed method, the temporal data are first converted into 2D images based on the periodic feature of normal user electricity data, and electricity theft detection is carried out based on image classification. It can be considered that the proposed method utilizes the periodic prior of normal electricity data; therefore, the proposed method can achieve a better classification performance.

If a smart thief makes the features of their electricity consumption time series consistent with the normal time series, such as reducing the electricity consumption record to half of their normal electricity consumption by tampering with the smart meter, the algorithm proposed in this paper would not be applicable. In this situation, the problem of electricity theft detection can be regarded as a general binary classification problem, and some methods suitable for time series classification such as LSTM should be good choices. However, it is obvious that the classification methods that fully consider the data features would generally obtain a better accuracy.

5. Conclusions

Electricity theft detection is of great significance for ensuring the safety and stability of power systems, and it is a very challenging issue as there are many ways to steal electricity and electricity theft may not occur continuously. This paper analyzed the features of the load time series of normal residents and electricity thieves and proposed a novel CNCP-based method for electricity theft detection. Meanwhile, the influence and solution of the imbalance data were discussed. Extensive experiments prove the effectiveness of the proposed method. The core idea of the proposed method is that, compared with the load time series of electricity thieves, the normal residents' load time series present more obvious periodicity. It is applicable in many practical electricity-stealing scenarios. Therefore, in the future research work, we will focus on this electricity theft scenario.