Hindawi
Journal of Electrical and Computer Engineering
Volume 2019, Article ID 4136874, 12 pages
https://doi.org/10.1155/2019/4136874

Research Article

Electricity Theft Detection in Power Grids with Deep Learning and Random Forests

Shuan Li,[1] **Yinghua Han**,[1] **Xu Yao**,[1] **Song Yingchen**,[2] **Jinkuan Wang**,[3] **and Qiang Zhao**[2]

[1]*School of Computer and Communication Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China*
[2]*School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China*
[3]*College of Information Science and Engineering, Northeastern University, Shenyang 110819, China*

Correspondence should be addressed to Yinghua Han; yhhan@neuq.edu.cn

Received 19 March 2019; Revised 30 July 2019; Accepted 14 August 2019; Published 3 October 2019

Academic Editor: Ping-Feng Pai

As one of the major factors of the nontechnical losses (NTLs) in distribution networks, the electricity theft causes significant harm to power grids, which influences power supply quality and reduces operating profits. In order to help utility companies solve the problems of inefficient electricity inspection and irregular power consumption, a novel hybrid convolutional neural network-random forest (CNN-RF) model for automatic electricity theft detection is presented in this paper. In this model, a convolutional neural network (CNN) firstly is designed to learn the features between different hours of the day and different days from massive and varying smart meter data by the operations of convolution and downsampling. In addition, a dropout layer is added to retard the risk of overfitting, and the backpropagation algorithm is applied to update network parameters in the training phase. And then, the random forest (RF) is trained based on the obtained features to detect whether the consumer steals electricity. To build the RF in the hybrid model, the grid search algorithm is adopted to determine optimal parameters. Finally, experiments are conducted based on real energy consumption data, and the results show that the proposed detection model outperforms other methods in terms of accuracy and efficiency.

1. Introduction

The loss of energy in electricity transmission and distribution is an important problem faced by power companies all over the world. The energy losses are usually classified into technical losses (TLs) and nontechnical losses (NTLs) [1]. The TL is inherent to the transportation of electricity, which is caused by internal actions in the power system components such as the transmission liner and transformers [2]; the NTL is defined as the difference between total losses and TLs, which is primarily caused by electricity theft. Actually, the electricity theft occurs mostly through physical attacks like line tapping, meter breaking, or meter reading tampering [3]. These electricity fraud behaviours may bring about the revenue loss of power companies. As an example, the losses caused by electricity theft are estimated as about $4.5 billion every year in the United States (US) [4]. And it is estimated that utility companies worldwide lose more than 20 billion every year in the form of electricity theft [5]. In addition, electricity theft behaviours can also affect the power system safety. For instance, the heavy load of electrical systems caused by electricity theft may lead to fires, which threaten the public safety. Therefore, accurate electricity theft detection is crucial for power grid safety and stableness.

With the implementation of the advanced metering infrastructure (AMI) in smart grids, power utilities obtained massive amounts of electricity consumption data at a high frequency from smart meters, which is helpful for us to detect electricity theft [6, 7]. However, every coin has two sides; the AMI network opens the door for some new electricity theft attacks. These attacks in the AMI can be launched by various means such as digital tools and cyber attacks. The primary means of electricity theft detection include humanly examining unauthorized line diversions, comparing malicious meter records with the benign ones, and checking problematic equipment or hardware. However, these methods are extremely time-consuming and costly during full verification of all meters in a system.

Besides, these manual approaches cannot avoid cyber attacks. In order to solve the problems mentioned above, many approaches have been put forward in the past years. These methods are mainly categorized into state-based, game-theory-based, and artificial-intelligence-based models [8].

The key idea of state-based detection [9–11] is based on special devices such as wireless sensors and distribution transformers [12]. These methods could detect electricity theft but rely on the real-time acquisition of system topology and additional physical measurements, which is sometimes unattainable. Game-based detection schemes formulate a game between electricity utility and theft, and then different distributions of normal and abnormal behaviours can be derived from the game equilibrium. As detailed in [13], they can achieve a low cost and reasonable result for reducing energy theft. Yet formulating the utility function of all players (e.g., distributors, regulators, and thieves) is still a challenge. Artificial-intelligence-based methods include machine learning and deep learning methods. Existing machine learning solutions can be further categorized into classification and clustering models, as is presented in [14–17]. Although aforementioned machine learning detection methods are innovative and remarkable, their performances are still not satisfactory enough for practice. For example, most of these approaches require manual feature extraction, which partly results from their limited ability to handle high-dimensional data. Indeed, traditional hand-designed features include the mean, standard deviation, maximum, and minimum of consumption data. The process of manual feature extraction is a tedious and time-consuming task and cannot capture the 2D features from smart meter data.

Deep learning techniques for electricity theft detection are studied in [18], where the authors present a comparison between different deep learning architectures such as convolutional neural networks (CNNs), long-short-term memory (LSTM) recurrent neural networks (RNNs), and stacked autoencoders. However, the performance of the detectors is investigated using synthetic data, which does not allow a reliable assessment of the detector's performance compared with shallow architectures. Moreover, the authors in [19] proposed a deep neural network- (DNN-) based customer-specific detector that can efficiently thwart such cyber attacks. In recent years, the CNN has been applied to generate useful and discriminative features from raw data and has wide applications in different areas [20–22]. These applications motivate the CNN applied for feature extraction from high-resolution smart meter data in electricity theft detection. In [23], a wide and deep convolutional neural network (CNN) model was developed and applied to analyse the electricity theft in smart grids.

In a plain CNN, the softmax classifier layer is the same as a general single hidden layer feedforward neural network (SLFN) and trained through the backpropagation algorithm [24]. On the one hand, the SLFN is likely to be overtrained leading to degradation of its generalization performance when it performs the backpropagation algorithm. On the other hand, the backpropagation algorithm is based on empirical risk minimization, which is sensitive to local minima of training errors. As mentioned above, because of the shortcoming of the softmax classifier, the CNN is not always optimal for classification, although it has shown great advantages in the feature extraction process. Therefore, it is urgent to find a better classifier which not only owns the similar ability as the softmax classifier but also can make full use of the obtained features. In most classifiers, the random forest (RF) classifier takes advantage of two powerful machine learning techniques including bagging and random feature selection which could overcome the limitation of the softmax classifier. Inspired by these particular works, a novel convolutional neural network-random forest (CNN-RF) model is adopted for electricity theft detection. The CNN is proposed to automatically capture various features of customers' consumption behaviours from smart meter data, which is one of the key factors in the success of the electricity theft detection model. To improve detection performance, the RF is used to replace the softmax classifier detecting the patterns of consumers based on extracted features. This model has been trained and tested with real data from all the customers of electricity utility in Ireland and London.

2. Overview Flow

The main aim of the methodology described in this paper is to provide the utilities with a ranked list of their customers, according to their probability of having an anomaly in their electricity meter.

As shown in Figure 1, the electricity theft detection system is divided into three main stages as follows:

(i) Data analysis and preprocess: to explain the reason of applying a CNN for feature extraction, we firstly analyse the factors that affect the behaviours of electricity consumers. For the data preprocess, we consider several tasks such as data cleaning (resolving outliers), missing value imputation, and data transformation (normalization).

(ii) Generation of train and test datasets: to evaluate the performance of the methodology described in this paper, the preprocessed dataset is split into the train dataset and test dataset by the cross-validation algorithm. The train dataset is used to train the parameters of our model, whilst the test dataset is used to assess how well the model generalizes to new, unseen customer samples. Given that electricity theft consumers remarkably outnumber non-fraudulent ones, the imbalanced nature of the dataset can have a major negative impact on the performance of supervised machine learning methods. To reduce this bias, the synthetic minority oversampling technique (SMOT) algorithm is used to make the number of electricity thefts and non-fraudulent consumers equal in the train dataset.

(iii) Classification using the CNN-RF model: in the proposed CNN-RF model, the CNN firstly is designed to learn the features between different hours of the day and different days from massive and varying smart meter data by the operations of

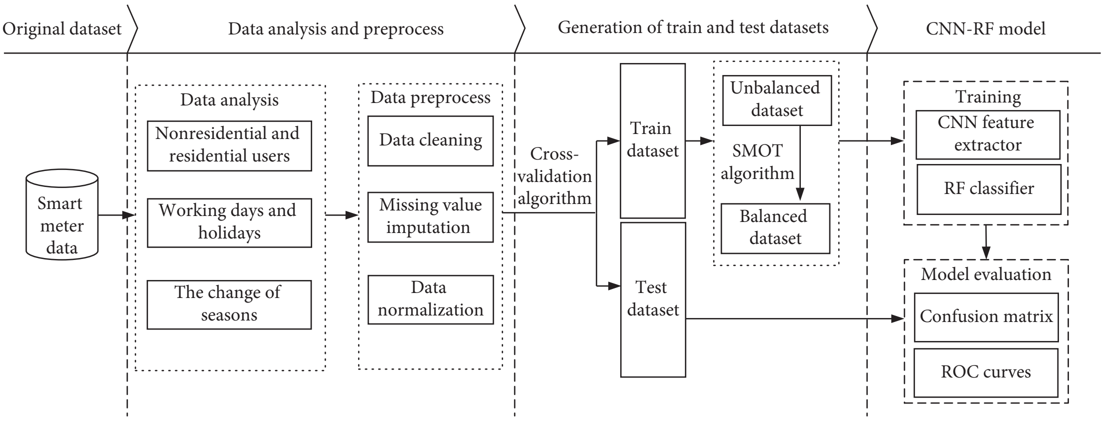


Figure 1: Flow of electricity theft detection.

convolution and downsampling. And then, RF classification is trained based on the obtained features to detect whether the consumer steals electricity. Finally, the confusion matrix and receiver-operating characteristic (ROC) curves are used to evaluate the accuracy of the CNN-RF model on the test dataset.

3. Data Analysis and Preprocess

The dataset was collected by Electric Ireland and Sustainable Energy Authority of Ireland (SEAI) in January 2012 [25], which consists of electricity usage data from over 5000 residential households and businesses. The smart meters recorded electricity consumption at a resolution of 30 min during 2009 and 2010. Customers who participated in the trial had a smart meter installed in their homes and agreed to take part in the research. Therefore, it is a reasonable assumption that all samples belong to honest users. However, malicious samples cannot be obtained since energy theft might never or rarely happen for a given customer. To address this issue, we generated 1200 malicious customers as described in [26]. The large number and a variety of customers with a long period of measurements make this dataset an excellent source for research in the area of smart meter data analysis. For each customer i, there is a half-hourly meter reporting for a 525-day period. We reduced the sampling rate to one per hour for each customer. This section firstly analyses smart meter data to describe the rationale of applying a CNN for feature extraction. Then, it describes three main procedures utilized for preprocessing the original energy consumption data: data cleaning, missing value imputation, and data normalization.

3.1. Data Analysis. To introduce the rationale for applying a CNN for feature extraction rather than other machine learning techniques, this section analyses smart meter data. There are the daily load curves of four different users (residential and nonresidential), as shown in Figure 2, and it can be seen that the daily electricity consumption of nonresidential users is significantly higher than that of residential users.

In addition, Figures 3 and 4 show the daily load profiles of a consumer during working days and holidays. The load files are highly similar but slightly shifted. In a CNN, the filter weights are uniform for different regions. Thus, the features calculated in the convolutional layer are invariant to small shifts, which means that relatively stable features can be obtained from varying load profiles. As described in [27], a deep CNN first automatically extracts features from massive load profiles and support vector machine (SVM) identifies the characteristics of the consumers. To further investigate the load series, we plot the four seasons' load profiles for a customer. It can be seen from Figure 5 that the electricity consumption of the consumer changes with the change of seasons.

In summary, the load consumption differs in both magnitude and time of use and is dependent on lifestyle, seasons, weather, and many other uncontrollable factors. Therefore, consumer load profiles are affected not only by weather and conditions but also by the types of consumers and other factors.

Based on the above analysis, extracting features from smart meter data based on experience is difficult. However, feature extraction is a key factor in the success of the detection system. Conventionally, manual feature extraction requires elaborately designed features for a specific problem that make it uneasy to adapt to other domains. In the CNN, the successive alternating convolutional and pooling layers are designed to learn progressively higher-level features (e.g., trend indicators, sequence standard deviation, and linear slope) with 2D historical electricity consumption data. In addition, highly nonlinear correlations exist between electricity consumption and these influencing factors. Since activation function has been designed on convolutional and fully connected layers, the CNN is able to model highly nonlinear correlations. In this paper, activation function named "rectified linear unit" (ReLU) is used because of its sparsity and minimizing gradient vanishing problem in the proposed CNN-RF model.

3.2. Data Preprocess. $x_{i,t}$ and $x'_{i,t}$ are defined as the energy consumption for an honest or malicious customer i at time interval t. In this section, the main procedures utilized for preprocessing raw electricity data are described as follows:

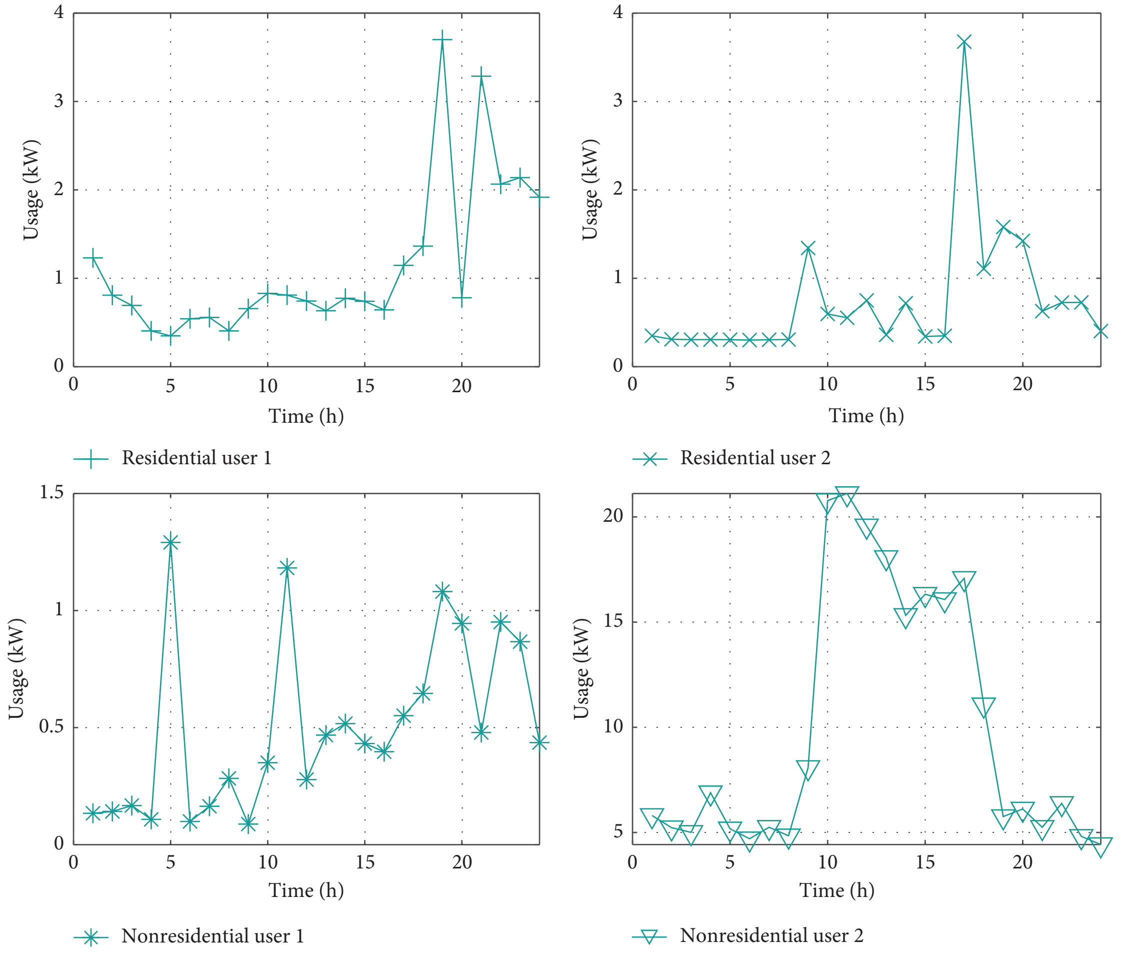


Figure 2: Load profiles for nonresidential and residential users.

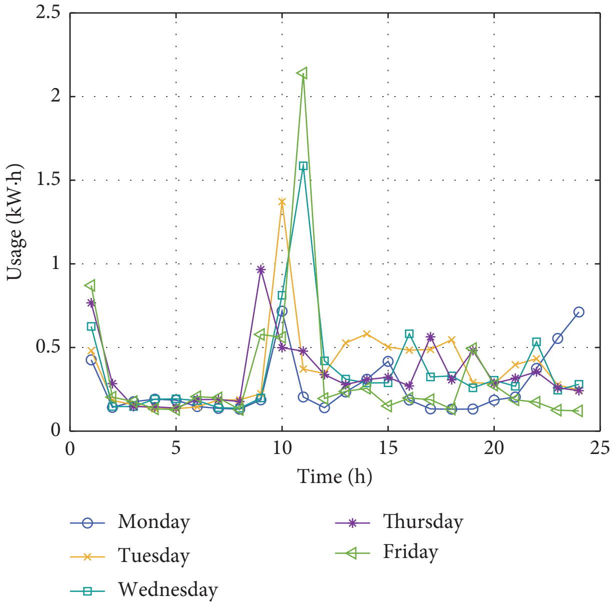


Figure 3: Daily load profiles for a customer in working days.

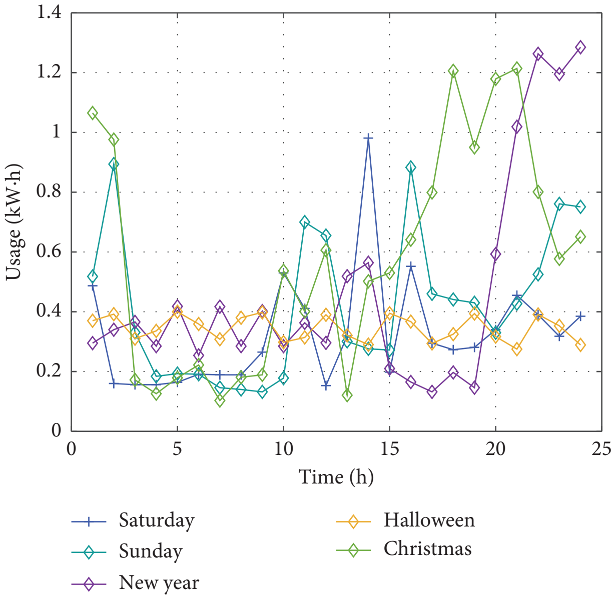


Figure 4: Daily load profiles for a customer during holidays.

(1) Data cleaning: there are erroneous values (e.g., outliers) in raw data, which correspond to peak electricity consumption activities during holidays or special occasions such as birthdays and celebrations. In this paper, the "three-sigma rule of thumb" [28] is used to restore the outliers according to the following formula:

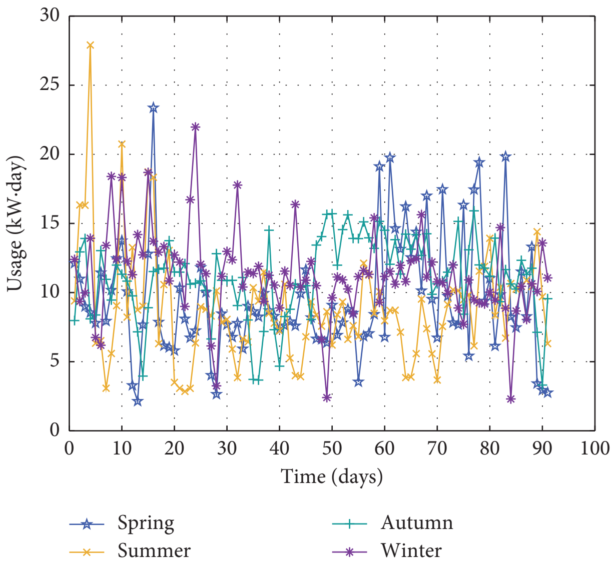


Figure 5: Four seasons' load profiles for a customer.

$$F\left(x_{i,t}\right)=\begin{cases}\operatorname{avg}\left(x_{i,t}\right)+2\sigma\left(x_{i,t}\right), & x_{i,t}>x_{i,t}^{*},\\ x_{i,t}, & \text{else,}\end{cases}\qquad(1)$$

where $x_{i,t}^{*}$ is computed by the mean avg$(\cdot)$ and standard deviation $\sigma(\cdot)$ for each time interval comprising the weekday/time pair for each month.

(2) Missing value imputation: because of various reasons such as storage issues and failure of smart meters, there are missing values in electricity consumption data. Through the analysis of original data, it is found that there are two kinds of data missing: one is the continuous missing of multiple data, and the solution is to delete the users when the number of missing values exceeds 10; the other is missing single data, which is processed by the formula (2). Thus, missing values can be recovered as follows:

$$F\left(x_{i,t}\right)=\begin{cases}\dfrac{x_{i,t-1}+x_{i,t+1}}{2}, & x_{i,t}\in \text{NaN},\\ x_{i,t}, & \text{else,}\end{cases}\qquad(2)$$

where $x_{i,t}$ stands for the electricity usage of the consumer i over a period (e.g., an hour); if $x_{i,t}$ is null, we represent it as NaN.

(3) Data normalization: finally, data need to be normalized because the neural network is sensitive to the diverse data. One of the common methods for this scope is min-max normalization, which is computed as

$$F\left(x_{i,t}\right)=\frac{x_{i,t}-\min\left(x_{i,T}\right)}{\max\left(x_{i,T}\right)-\min\left(x_{i,T}\right)},\qquad(3)$$

where min$(\cdot)$ and max$(\cdot)$ represent the min and max values over a day, respectively.

4. Generation of Train and Test Datasets

After data processing, the SEAI dataset contains the electricity consumption data of 4737 normal electricity customers and 1200 malicious electricity customers within 525 days (with 24 hours).

Our preliminary analytical results (refer to Section 3.1) reveal the electricity usage consumer load profiles are affected not only by weather and conditions but also by the types of consumers and other factors. However, it is difficult to extracting features based on experience from the 1D electricity consumption data since the electricity consumption every day fluctuates in a relatively independent way. Motivated by the previous work in [29], we design a deep CNN to process the electricity consumption data in a 2D manner. In particular, we transform the 1D electricity consumption data into 2D data according to days. We define 2D data as a matrix of actual energy consumption values for a specific customer, where the rows of 2D data represent days as D ($D=\{1,\ldots,525\}$) and columns represent the time periods as T ($T=\{1,\ldots,24\}$). It is worth mentioning that the CNN learns the features between different hours of the day and different days from 2D data.

Then, we divide the dataset into train and test sets using the cross-validation method, in which 80% are the train set and 20% are the test set. Given that electricity theft consumers remarkably outnumber nonfraudulent ones, the imbalanced nature of the dataset can have a major negative impact on the performance of supervised machine learning methods. To reduce this bias, the SMOT algorithm is used to make the number of normal and abnormal samples equal in the train set. Finally, the train dataset contains 7484 customers, in which the number of normal and abnormal samples is equal. And the test dataset consists of 1669 customers.

5. The Novel CNN-RF Algorithm

In this section, the design of the CNN-RF structure is presented in detail and some techniques are proposed to reduce overfitting. Moreover, correlative parameters including the number of filters and the size of each feature map and kernel are given. Finally, the training process of the proposed CNN-RF algorithm is introduced.

5.1. CNN-RF Architecture. The proposed CNN-RF model is designed by integrating CNN and RF classifiers. As shown in Figure 6, the architecture of the CNN-RF mainly includes an automatic feature extractor and a trainable RF classifier. The feature extractor CNN consists of convolutional layers, downsampling layers, and a fully connection layer. The CNN is a deep supervised learning architecture, which usually includes multiple layers and can be trained with the backpropagation algorithm. It also can explore the intricate distribution in smart meter data by performing the

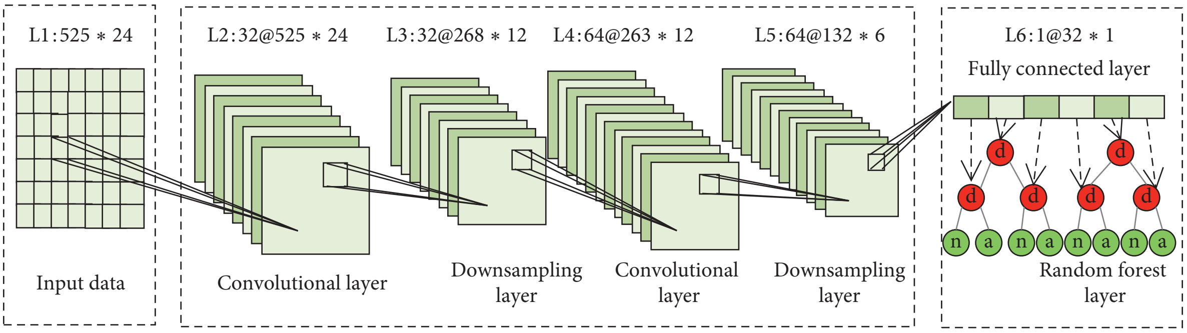


Figure 6: Architecture of the hybrid CNN-RF model.

stochastic gradient method. The RF classifier consists of a combination of tree classifiers, in which each tree contributes with a single vote to the assignment of the most frequent class in input data [30]. In the following, the design of each part of the CNN-RF structure is presented in detail.

5.1.1. Convolutional Layer. The main purpose of convolutional layers is to learn feature representation of input data and reduce the influence of noise. The convolutional layer is composed of several feature filters which are used to compute different feature maps. Specially, to reduce the parameters of networks and lower the complication of relevant layers during the process of generating feature maps, the weights of each kernel in each feature map are shared. Moreover, each neuron in the convolutional layer connects the local region of front layers, and then a ReLU activation function is applied on the convolved results. Mathematically, the outputs of the convolutional layers can be expressed as

$$y_{\text{conv}}\left(X_l^{f_l}\right) = \delta\left(\sum_{f_l=1}^{F_l} W_l^{f_l} * X_l^{f_l} + b_l^{f_l}\right), \qquad (4)$$

where δ is the activation function, $*$ is the convolution operation, and both $W_l^{f_l}$ and $b_l^{f_l}$ are the learnable parameters in the f-th feature filter.

5.1.2. Downsampling Layer. Downsampling is an important concept of the CNN, which is usually placed between two convolutional layers. It can reduce the number of parameters and achieve dimensionality reduction. In particular, each feature map of the pooling layer is connected to its corresponding feature map of the previous convolutional layer. And the size of output feature maps is reduced without reducing the number of output feature maps in the downsampling layer. In the literature, downsampling operations mainly include max-pooling and average-pooling. The max-pooling operation transforms small windows into single values by maximum which is expressed as formula 5, but average-pooling returns average values of activations in a small window. In [31], experiments show that the max-pooling operation shows better performance than average-pooling. Therefore, the max-pooling operation is usually implemented in the downsampling layer:

$$y_{\text{pool}}\left(X_l^{f_l}\right) = \max_{m\in M}\left\{X_{l,m}\right\}, \qquad (5)$$

where M is a set of activations in the pooling window and m is the index of activations in the pooling window.

5.1.3. Fully Connected Layer. With the features extracted by sequential convolution and pooling, a fully connected layer is applied for flattening feature maps into one vector as follows:

$$y_{f_l}\left(X_l\right) = \delta\left(W_l \cdot X_l + b_l\right), \qquad (6)$$

where W_l is the weight of the l-th layer and b_l is the bias of the l-th layer.

5.1.4. RF Classifier Layer. Traditionally, the softmax classifier is used for the last output layer in the CNN, but the proposed model uses the RF classifier to predict the class based on the obtained features. Therefore, the RF classifier layer can be defined as

$$y_{\text{out}}\left(X_L\right) = \text{sigm}\left(W_{\text{rf}} \cdot X_l + b_l\right), \qquad (7)$$

where $\text{sigm}(\cdot)$ is the sigmoid function, which maps abnormal values to 0 and normal values to 1. The parameter set W_{rf} in the RF layer includes the number of decision trees and the maximum depth of the tree, which are obtained by the grid search algorithm.

5.2. Techniques for Selecting Parameters and Avoiding Overfitting. The detailed parameters of the proposed CNN-RF structure are summarized in Table 1, which include the numbers of filters in each layer, filter size, and stride. Two factors are considered in determining the CNN structure. The first factor is the characteristics of consumers' electricity consumption behaviours. Since the load files are such a variable, two convolutional layers with a filter size of $3 * 3$ (stride 1) are applied to capture hidden features for detecting electricity theft. Moreover, because the dimension of input data size is 525×24, which is similar to what is used in image recognition problems, two max-pooling layers with a filter size of $2 * 2$ (stride 2) are used. The second factor is the number of training samples since the number of samples is limited; to reduce the risk of overfitting, we design a fully connected layer with a dropout rate of 0.4, whose main idea

Table 1: Parameters of the proposed CNN-RF model.

Layer	Size of the filter	Number of filters	Stride
Convolution1	3 * 3	32	1
MaxPooling1	2 * 2	—	2
Convolution2	3 * 3	64	1
MaxPooling2	2 * 2	—	2

is that dropping units randomly from the neural network during training can prevent units from coadapting too much [32] and make a neuron not rely on the presence of other specific neurons. Applying an appropriate training method is also useful for reducing overfitting. It is essentially a regularization that adds a penalty for weight update at each iteration. And we also use a binary cross entropy as the loss function. Finally, the grid search algorithm is used to optimize RF classifier parameters such as the maximum number of decision trees and features.

5.3. Training Process of CNN-RF. There is no doubt that the CNN-RF structure needs to train the CNN at first before the RF classifier is invoked. After that, the RF classifier is trained based on features learned by the CNN.

5.3.1. Training Process of CNN. With a batch size (the number of training samples in an iteration) of 100, CNNs are trained using the forward propagation algorithm and backpropagation algorithm. As shown in Figure 6, firstly, the data in the input layer are transferred to convolutional layers, pooling layers, and a fully connected layer, and the predicted value is obtained. As shown in Figure 7, the backpropagation phase begins to update parameters if the difference of the output value and target value is too large and exceeds a certain threshold. The difference between the actual output of the neural network and the expected output determines the adjustment direction and strip size of weights among layers in the network.

Given a sample set $D = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ with m samples in total, $\tilde{y}$ is obtained at first by using the feedforward process. As for all samples, the mean of difference between the actual output y_i and the expected output $\tilde{y}_{w,b}(x_i)$ can be expressed by

$$J(W, b) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{2} \left\| \tilde{y}_{w,b}(x_i) - y_i \right\|^2 \right), \tag{8}$$

where w represents the connection weight between layers of the network and b is the corresponding bias.

Then, each parameter is initialized with a random value generated by a normal distribution when the mean is 0 and variance is ϑ, which are updated with the gradient descent method. The corresponding expressions are as follows:

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b), \tag{9}$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} J(W, b), \tag{10}$$

where α represents a learning rate, $W_{ij}^{(l)}$ represents the connection weight between the i-th neuron in the l-th layer and the j-th neuron in the $(l + 1)$-th layer, and $b_i^{(l)}$ represents the bias of the i-th neuron in the l-th layer.

Finally, all parameters W and b in the network are updated according to formulas 9 and 10. And all these steps are repeated to reduce the objective function $J(W, b)$.

5.3.2. Training Process of RF Classifier. The process of the RF classifier takes advantage of two powerful machine learning techniques: bagging and random feature selection. Bagging is a method to generate a particular bootstrap sample randomly from the learnt features for growing each tree. Instead of using all features, the RF randomly selects a subset of features to determine the best splitting points, which ensures complete splitting to one classification of leaf nodes of decision trees. Specifically, the whole process of RF classification can be written as Algorithm 1.

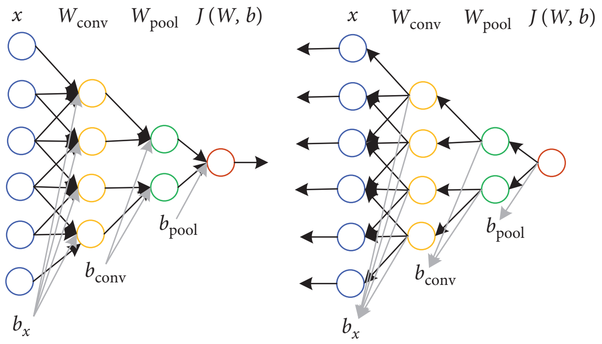
6. Experiments and Result Analysis

All the experiments are implemented using Python 3.6 on a standard PC with an Intel Core i5-7500MQ CPU running at 3.40 GHz and with 8.0 GB of RAM. The CNN architecture is constructed based on TensorFlow [33], and the interface between the CNN and the RF is programmed using scikit-learn [34].

6.1. Performance Metrics. In this paper, the problem of electricity theft is considered a discrete two-class classification task, which should assign each consumer into one of the predefined classes (abnormal or normal).

In particular, the result of classifier validation is often presented as confusion matrices. Table 2 shows a confusion matrix applied in electricity theft detection, where TP, FN, FP, and TN, respectively, represent the number of consumers that are classified correctly as normal, classified falsely as abnormal, classified falsely as normal, and classified correctly as abnormal. Although these four indices show all of the information about the performance of the classifier, more meaningful performance measures can be extracted from them to illustrate certain performance criteria such as TPR = TP/(TP + FN), FPR = FP/(FP + TN), precision = TP/(TP + FP), recall = TP/(TP + FN), and $F1$ score = 2(precision × recall)/(precision + recall), where the true positive rate (TPR) is the ability of the classifier to correctly identify a consumer as normal, also referred to as sensitivity, and the false positive rate (FPR) is the risk of positively identifying an abnormal consumer. Precision is the ratio of consumers detected correctly to the total number of detected normal consumers. Recall is the ratio of the number of consumers correctly detected to the actual number of consumers (Figure 7).

In addition, the ROC curve is introduced by plotting all TPR values on the y-axis against their equivalent FPR values for all available thresholds on the x-axis. The ROC curve at the upper left has better detection effects, where lower FPR values are caused by the same TPR. The AUC indicates the



(a) (b)

Figure 7: Process of forward propagation (a) and backpropagation (b).

```
Require:
  (1) Original training datasets
    $S = \{(x_i, y_i), i = 1, 2, 3, \ldots, m\}, (X, Y) \in R^e \times R$
  (2) Test datasets $x_j \in R^m$
  for $d = 1$ to $N_{\text{tree}}$ do
    (1) Draw a bootstrap $S_d$ from the original training data
    (2) Grow an unpruned tree $h_d$ using data $S_d$
      (a)Randomly select new feature set $M_{\text{try}}$ from original feature set $e$
      (b)Select the best features from feature set $M_{\text{try}}$ based on Gini indicator on each node
      (c)Split until each tree grows to its maximum
  end for
Ensure:
  (1) The collection of trees $\{h_d, d = 1, 2, \ldots, N_{\text{tree}}\}$
  (2) For $x_j$, the output of a decision tree is $h_d(x_j)$
    $f(x_j) = \text{majority vote}\{h_d(x_j)\}_{N_{\text{tree}}}^{d}$
return $f(x_j)$
```

Algorithm 1: Random forest algorithm.

Table 2: Confusion matrix applied in electricity theft detection.

Actual/detected	Normal	Abnormal
Normal	TP (true positive)	FN (false negative)
Abnormal	FP (false positive)	TN (true negative)

quality of the classifier, and the larger AUC represents better performance.

6.2. Experiment Based on CNN-RF Model. The experiment is conducted based on normalized consumption data. With the batch size of 100, the training procedure is stopped after 300 epochs. As shown in Figure 8, it could be seen that training and testing losses settled down smoothly. This means that the proposed model learned train dataset, and there was small bias in the test data dataset as well. Then, the RF classifier is the last layer of the CNN to predict labels of input consumers. Forty values from the fully connected layer of the trained CNN are used as a new feature vector to detect abnormal users and are fed to the RF for learning and testing. To build the RF in the hybrid model, the optimal parameters are determined by applying the grid search algorithm based on the training dataset. The grid searching range of each parameter is given as follows: numTrees $= [50, 60, \ldots, 100]$ and max depth $= [10, 20, \ldots, 40]$. $6 \times 4 = 24$ are tried with

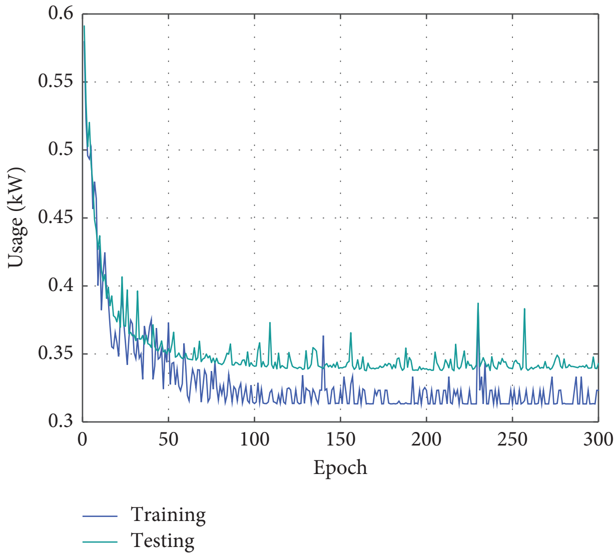


Figure 8: Loss curves of training and testing.

different combinations. The best result is achieved when numTrees = 100 and maxDepth = 30. These parameters are then used to train the hybrid model.

The classification results of the detection model proposed in this paper are as follows: the number of TPs, FNs, FPs, and TNs is 1041, 28, 23, and 577, respectively. Therefore, the precision, recall, and $F1$ score can be calculated, as shown in Table 3, where Class 0 is an abnormal pattern and Class 1 is a normal pattern. Also, the AUC value of the CNN-RF algorithm is calculated, which is 0.98 and far better than the value of the baseline model (AUC = 0.5). This means the proposed algorithm is able to classify both classes accurately, as shown in Figure 9.

Table 3: Classification score summary for CNN-RF.

	Precision	Recall	$F1$ score
Class 0	0.97	0.96	0.96
Class 1	0.98	0.98	0.98
Average/total	0.97	0.97	0.97

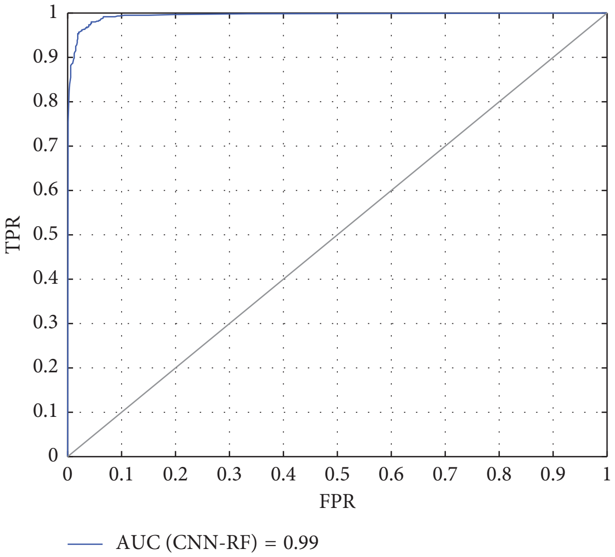


Figure 9: ROC curve for CNN-RF.

6.3. Comparative Experiments and the Analysis of Results. To evaluate the accuracy of the CNN-RF model, comparative experiments are conducted based on handcrafted features with nondeep learning methods including SVM, RF, gradient boosting decision tree (GBDT), and logistic regression (LR). Moreover, we compared the obtained classification results by various supervised classifiers: CNN features with the SVM classifier (CNN-SVM) and CNN features with the GBDT classifier (CNN-GBDT), to those obtained by the previous classification work. In the following, five methods are introduced, respectively, and then the results are analysed:

(1) Logistic regression: it is a basic model in binary classification that is equivalent to one layer of neural network with sigmoid activation function. A sigmoid function obtains a value ranging from 0 to 1 based on linear regression. Then, any value larger than 0.5 will be classified as a normal pattern and less than 0.5 will be classified as an abnormal pattern.

(2) Support vector machine: this classifier with kernel functions can transform a nonlinear separable problem into a linear separable problem by projecting data into the feature space and then finding the optimal separate hyperplane. It is applied to predict the patterns of consumers based on the aforementioned handcrafted features.

(3) Random forest: it is essentially an integration of multiple decision trees that can achieve better performance when maintaining the effective control of overfitting in comparison with a single decision tree. The RF classifier also can handle high-dimensional data while maintaining high computational efficiency.

(4) Gradient boosting decision tree: it is an iterative decision tree algorithm, which consists of multiple decision trees. For the final output, all results or weights are accumulated in the GBDT, while random forests use majority voting.

(5) Deep learning methods: softmax is used in the last layer of the CNN, CNN-GBDT, and CNN-SVM in comparison with the proposed method.

Table 4: Experimental parameters for comparative methods.

Methods	Input data	Parameters
LR	Features (1D)	Penalty: L2 Inverse of regulation strength: 1.0
SVM	Features (1D)	Penalty parameter of the error term: 10^5 Parameter of kernel function (RBF): 0.0005
GBDT	Features (1D)	The number of estimators: 200
RF	Features (1D)	The number of trees in the forest: 100 The max depth of each tree: 30
CNN	Raw data (2D)	The same as the CNN component of the proposed approach

Table 4 summarizes the parameters of comparative methods. All experiments have been done. Figure 10 shows the results of different methods and proposed hybrid CNN-RF method, and it can be seen that the AUC value of CNN-RF, CNN-GBDT, CNN-SVM, CNN, SVM, RF, LR, and GBDT is 0.99, 0.97, 0.98, 0.93, 0.77, 0.91, 0.63, and 0.77. And Figure 11 presents the results of all comparative experiments in terms of precision, recall, and $F1$ score. Among the performances of eight different electricity theft detection algorithms, the deep learning methods (such as CNN, CNN-RF, CNN-SVM, and CNN-GBDT) show better performances than machine learning methods (e.g., LR, SVM, GBDT, and RF). The reason of this result is that the CNN can learn features through a large number of electricity consumption data. Thus, the accuracy of electricity theft detection is greatly improved. In addition, it is indicated that the proposed CNN-RF model shows the best performance compared with other methods.

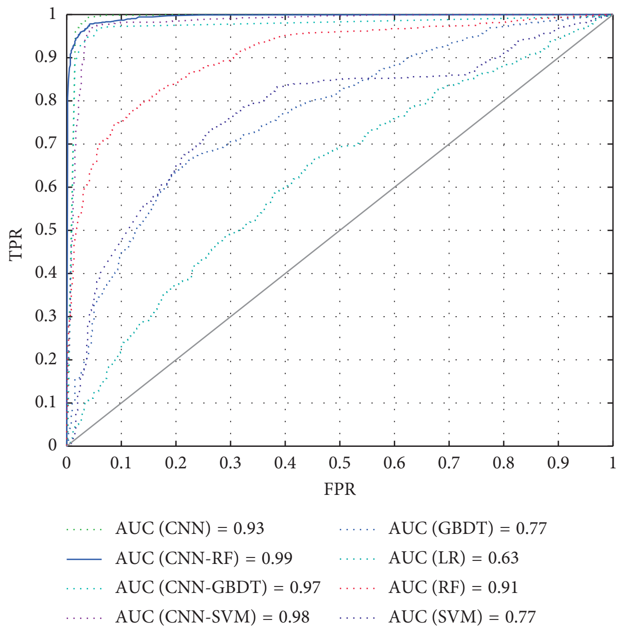


Figure 10: ROC curves for eight different algorithms based on the SEAI dataset.

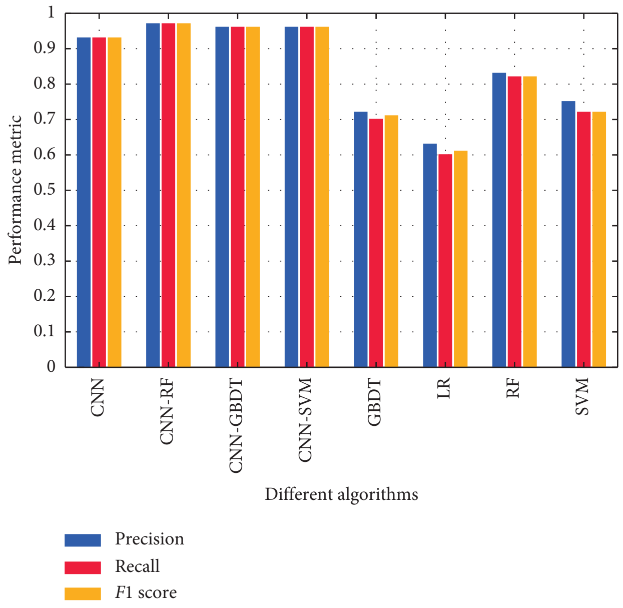


Figure 11: Comparison with other algorithms based on the SEAI dataset.

As detecting electricity theft consumers is achieved by discovering the anomalous consumption behaviours of suspicious customers, we can also regard it as an anomaly detection problem essentially. Since CNN-RF has shown superior performance on the SEAI dataset, it would be interesting to test its generalization ability on other datasets. In particular, we have further tested the performance of CNN-RF together with the above-mentioned models on another dataset: Low Carbon London (LCL) dataset [35]. The LCL dataset is composed of over 5500 honest electricity consumers in total, each of which consists of 525 days (the sampling rate is hour). Among the 5500 customers, we randomly chose 1200 as malicious customers and modified their intraday load profiles according to the fraudulent sample generation methods similar to the SEAI dataset.

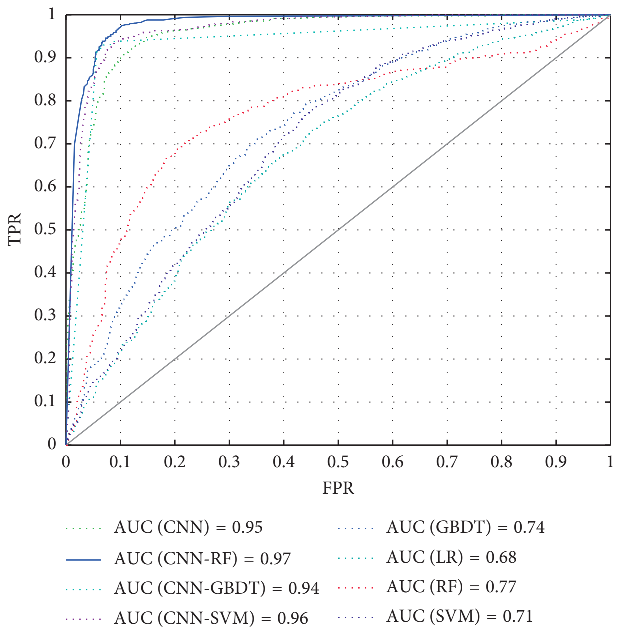


Figure 12: ROC curves for eight different algorithms based on the LCL dataset.

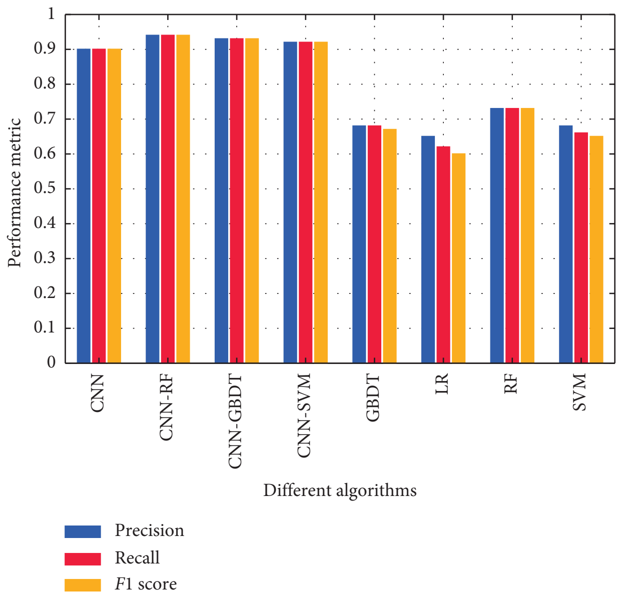


Figure 13: Comparison with other algorithms based on the LCL dataset.

After the preprocess of the LCL dataset (similar to the SEAI dataset), the train dataset contains 7584 customers, in which the number of normal and abnormal samples is equal. And the test dataset consists of 1700 customers together with corresponding labels indicating anomaly or not. Then, the training of the proposed CNN-RF model and comparative models is all done following the aforementioned procedure based on the train dataset. The performances of all the models on the test dataset are shown in Figures 12 and 13.

Figures 12 and 13 show that the performances of CNN-SVM, CNN-GBDT, and CNN-RF are better than those of other models on the whole. Since they are deep learning models, the distinguishable features can be learned directly from raw smart meter data by the CNN, which has equipped them with better generalization ability. In addition, the CNN-RF model has achieved the best performance since the RF can handle high-dimensional data while maintaining high computational efficiency.

7. Conclusions

In this paper, a novel CNN-RF model is presented to detect electricity theft. In this model, the CNN is similar to an automatic feature extractor in investigating smart meter data and the RF is the output classifier. Because a large number of parameters must be optimized that increase the risk of overfitting, a fully connected layer with a dropout rate of 0.4 is designed during the training phase. In addition, the SMOT algorithm is adopted to overcome the problem of data imbalance. Some machine learning and deep learning methods such as SVM, RF, GBDT, and LR are applied to the same problem as a benchmark, and all those methods have been conducted on SEAI and LCL datasets. The results indicate that the proposed CNN-RF model is quite a promising classification method in the electricity theft detection field because of two properties: The first is that features can be automatically extracted by the hybrid model, while the success of most other traditional classifiers relies largely on the retrieval of good hand-designed features which is a laborious and time-consuming task. The second lies in that the hybrid model combines the advantages of the RF and CNN, as both are the most popular and successful classifiers in the electricity theft detection field.

Since the detection of electricity theft affects the privacy of consumers, the future work will focus on investigating how the granularity and duration of smart meter data might affect this privacy. Extending the proposed hybrid CNN-RF model to other applications (e.g., load forecasting) is a task worth investigating.

Data Availability

Previously reported [SEAI][LCL] datasets were used to support this study and are available at [http://www.ucd.ie/issda/data/][https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=7857]. These prior studies (and datasets) are cited at relevant places within the text as references [19, 26].

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2016YFB0901900), the Natural Science Foundation of Hebei Province of China (F2017501107), Open Research Fund from the State Key Laboratory of Rolling and Automation, Northeastern University (2017RALKFKT003), the Foundation of Northeastern University at Qinhuangdao (XNB201803), and the Fundamental Research Funds for the Central Universities (N182303037).

References

[1] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Electricity theft: overview, issues, prevention and a smart meter based approach to control theft," *Energy Policy*, vol. 39, no. 2, pp. 1007–1015, 2011.

[2] J. P. Navani, N. K. Sharma, and S. Sapra, "Technical and non-technical losses in power system and its economic consequence in Indian economy," *International Journal of Electronics and Computer Science Engineering*, vol. 1, no. 2, pp. 757–761, 2012.

[3] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A multi-sensor energy theft detection framework for advanced metering infrastructures," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1319–1330, 2013.

[4] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Security & Privacy Magazine*, vol. 7, no. 3, pp. 75–77, 2009.

[5] T. B. Smith, "Electricity theft: a comparative analysis," *Energy Policy*, vol. 32, no. 1, pp. 2067–2076, 2004.

[6] J. I. Guerrero, C. León, I. Monedero, F. Biscarri, and J. Biscarri, "Improving knowledge-based systems with statistical techniques, text mining, and neural networks for non-technical loss detection," *Knowledge-Based Systems*, vol. 71, no. 4, pp. 376–388, 2014.

[7] C. C. O. Ramos, A. N. Souza, G. Chiachia, A. X. Falcão, and J. P. Papa, "A novel algorithm for feature selection using harmony search and its application for non-technical losses detection," *Computers & Electrical Engineering*, vol. 37, no. 6, pp. 886–894, 2011.

[8] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, "The challenge of non-technical loss detection using artificial intelligence: a surveyficial intelligence: a survey," *International Journal of Computational Intelligence Systems*, vol. 10, no. 1, pp. 760–775, 2017.

[9] S.-C. Huang, Y.-L. Lo, and C.-N. Lu, "Non-technical loss detection using state estimation and analysis of variance," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2959–2966, 2013.

[10] O. Rahmati, H. R. Pourghasemi, and A. M. Melesse, "Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran region, Iran," *CATENA*, vol. 137, pp. 360–372, 2016.

[11] N. Edison, A. C. Aranha, and J. Coelho, "Probabilistic methodology for technical and non-technical losses estimation in distribution system," *Electric Power Systems Research*, vol. 97, no. 11, pp. 93–99, 2013.

[12] J. B. Leite and J. R. S. Mantovani, "Detecting and locating non-technical losses in modern distribution networks," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1023–1032, 2018.

[13] S. Amin, G. A. Schwartz, A. A. Cardenas, and S. S. Sastry, "Game-theoretic models of electricity theft detection in smart utility networks: providing new capabilities with advanced

metering infrastructure," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 66–81, 2015.

[14] A. H. Nizar, Z. Y. Dong, Y. Wang, and A. N. Souza, "Power utility nontechnical loss analysis with extreme learning machine method," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 946–955, 2008.

[15] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 1162–1171, 2010.

[16] E. W. S. Angelos, O. R. Saavedra, O. A. C. Cortés, and A. N. de Souza, "Detection and identification of abnormalities in customer consumptions in power distribution systems," *IEEE Transactions on Power Delivery*, vol. 26, no. 4, pp. 2436–2442, 2011.

[17] K. A. P. Costa, L. A. M. Pereira, R. Y. M. Nakamura, C. R. Pereira, J. P. Papa, and A. Xavier Falcão, "A nature-inspired approach to speed up optimum-path forest clustering and its application to intrusion detection in computer networks," *Information Sciences*, vol. 294, pp. 95–108, 2015.

[18] R. R. Bhat, R. D. Trevizan, X. Li, and A. Bretas, "Identifying nontechnical power loss via spatial and temporal deep learning," in *Proceedings of the IEEE International Conference on Machine Learning and Applications*, Anaheim, CA, USA, December 2016.

[19] M. Ismail, M. Shahin, M. F. Shaaban, E. Serpedin, and K. Qaraqe, "Efficient detection of electricity theft cyber attacks in ami networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference*, Barcelona, Spain, April 2018.

[20] R. Mehrizi, X. Peng, X. Xu, S. Zhang, D. Metaxas, and K. Li, "A computer vision based method for 3D posture estimation of symmetrical lifting," *Journal of Biomechanics*, vol. 69, no. 1, pp. 40–46, 2018.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE Conference on Computer for Sustainable Global Development*, pp. 1026–1034, New Delhi, India, March 2015.

[22] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," *Astrophysical Journal Supplement Series*, vol. 221, no. 1, pp. 806–813, 2014.

[23] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1606–1615, 2018.

[24] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 43–62, 1997.

[25] Irish social science data archive: http://www.ucd.ie/issda/data/commissionforenegryregulationcer/.

[26] P. Jokar, N. Arianpoo, and V. C. M. Leung, "Electricity theft detection in AMI using customers' consumption patterns," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 216–226, 2016.

[27] Y. Wang, Q. Chen, D. Gan, J. Yang, D. S. Kirschen, and C. Kang, "Deep learning-based socio-demographic information identification from smart meter datafication from smart meter data," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2593–2602, 2019.

[28] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," *ACM Computing Surveys*, vol. 41, no. 3, 2009.

[29] H.-T. Cheng, L. Koc, J. Harmsen et al., "Wide & deep learning for recommender systems," in *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems—DLRS 2016*, pp. 7–10, Boston, MA, USA, September 2016.

[30] E. Feczko, N. M. Balba, O. Miranda-Dominguez et al., "Subtyping cognitive profiles in autism spectrum disorder using a functional random forest algorithmfiles in autism spectrum disorder using a functional random forest algorithm," *NeuroImage*, vol. 172, pp. 674–688, 2018.

[31] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proceedings of the 27th International Conference on Machine Learning*, pp. 111–118, Haifa, Israel, June 2010.

[32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simply way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

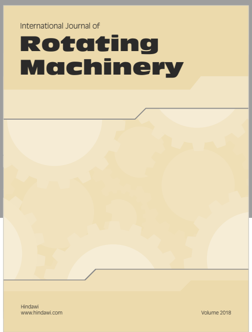
[33] M. Abadi, A. Agarwal, P. Barham et al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," 2016, https://arxiv.org/abs/1603.0446.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: machine learning in python," *Journal of Machine Learning Research*, vol. 12, no. 4, pp. 2825–2830, 2011.

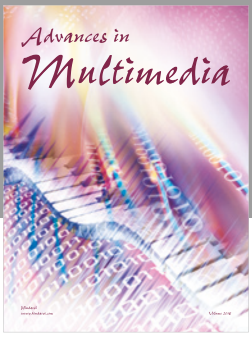
[35] J. R. Schofield, "Low carbon london project: data from the dynamic time-of-use electricity pricing trial," 2017, https://discover.ukdataservice.ac.uk//catalogue?sn=7857.

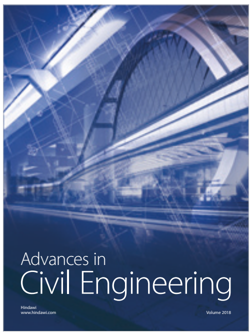
Journal of
Engineering
Hindawi
www.hindawi.com
Volume 2018

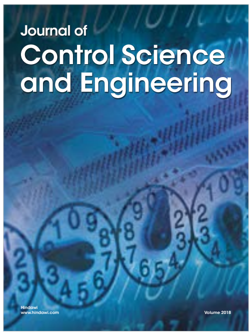
The Scientific
World Journal
Hindawi
www.hindawi.com
Volume 2018


International Journal of
Rotating
Machinery
Hindawi
www.hindawi.com
Volume 2018

Journal of
Sensors
Hindawi
www.hindawi.com
Volume 2018

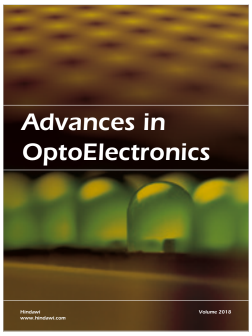

Advances in
Multimedia
Hindawi
www.hindawi.com
Volume 2018


Advances in
Civil Engineering
Hindawi
www.hindawi.com
Volume 2018


Journal of
Control Science
and Engineering
Hindawi
www.hindawi.com
Volume 2018

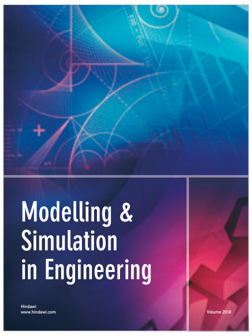
Journal of
Robotics
Hindawi
www.hindawi.com
Volume 2018

Journal of
Electrical and Computer
Engineering
Hindawi
www.hindawi.com
Volume 2018

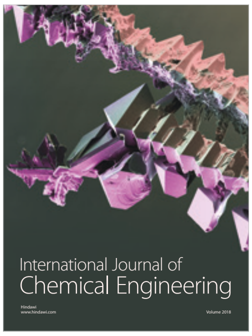

Advances in
OptoElectronics
Hindawi
www.hindawi.com
Volume 2018

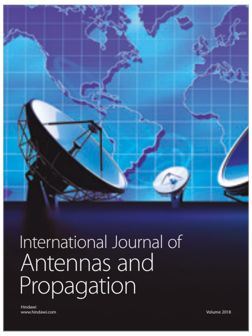
VLSI Design
Hindawi
www.hindawi.com
Volume 2018

International Journal of
Navigation and
Observation
Hindawi
www.hindawi.com
Volume 2018

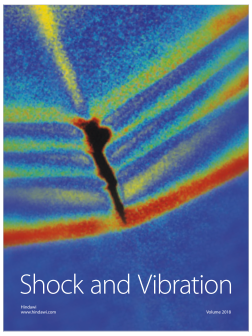

Modelling &
Simulation
in Engineering
Hindawi
www.hindawi.com
Volume 2018

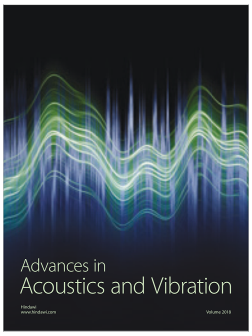
International Journal of
Aerospace
Engineering
Hindawi
www.hindawi.com
Volume 2018


International Journal of
Chemical Engineering
Hindawi
www.hindawi.com
Volume 2018


International Journal of
Antennas and
Propagation
Hindawi
www.hindawi.com
Volume 2018

Active and Passive
Electronic Components
Hindawi
www.hindawi.com
Volume 2018


Shock and Vibration
Hindawi
www.hindawi.com
Volume 2018


Advances in
Acoustics and Vibration
Hindawi
www.hindawi.com
Volume 2018