IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON EVOLVING TECHNOLOGIES IN ENERGY STORAGE SYSTEMS FOR ENERGY SYSTEMS APPLICATIONS

Received February 13, 2021, accepted February 23, 2021, date of publication March 1, 2021, date of current version March 31, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3062675*

# Abnormal Detection of Electricity Consumption of User Based on Particle Swarm Optimization and Long Short Term Memory With the Attention Mechanism

**JIAHAO BIAN[1], LEI WANG[1,4], RAFAŁ SCHERER[2], (Member, IEEE), MARCIN WOŹNIAK[3], PENGCHAO ZHANG[1], AND WEI WEI[4], (Senior Member, IEEE)**

[1]Shaanxi Key Laboratory of Industrial Automation, Shaanxi University of Technology, Hanzhong 723001, China
[2]Institute of Computational Intelligence, Czestochowa University of Technology, 42-200 Czestochowa, Poland
[3]Faculty of Applied Mathematics, Silesian University of Technology, 44-100 Gliwice, Poland
[4]Shaanxi Key Laboratory of Network Computing and Security Technology, Xi'an University of Technology, Xi'an 710048, China

Corresponding author: Lei Wang (leiwang@xaut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61773314, in part by the Shaanxi Provincial Natural Science Basic Research Program under Grant 2019JZ-11, in part by the Scientific Research Project of Education Department of Shaanxi Provincial Government under Grant 19JC011, in part by the Key Research and Development Program of Shaanxi Province under Grant 2018ZDXM-GY-036, and in part by the Shaanxi Key Laboratory of Intelligent Processing for Big Energy Data under Grant IPBED7.

**ABSTRACT** In the process of power transmission and distribution, non-technical losses are usually caused by users' abnormal power consumption behavior. It will not only affect the dispatch and operation of the distribution network, bring hidden dangers to the security of the power grid, but also damage the operating costs of power companies and disrupt the operation of the power market. Aiming at users' abnormal electricity consumption behavior, this paper proposes a model based on particle swarm optimization and long-short term memory with the attention mechanism (PSO-Attention-LSTM). Firstly, according to the actual electricity theft behavior, six typical electricity theft modes are summarized, and 4 composite modes are obtained by combining them, so as to comprehensively test the detection performance of the model for various electricity theft behaviors. Secondly, a detection model based on PSO-Attention-LSTM is proposed, and the model is built using the TensorFlow framework. The model uses the attention mechanism to give different weights to the hidden state of LSTM, which reduces the loss of historical information, strengthens important information and suppresses useless information. Use PSO to solve the difficult problem of model parameter selection, and optimize the hyperparameters to improve the model performance. Finally, the data set of the University of Massachusetts was used for simulation and compared with convolutional neural network-long short term memory (CNN-LSTM), attention mechanism-based long short term memory (Attention-LSTM), LSTM, gated recurrent unit (GRU), support vector regression (SVR), random forest (RF) and linear regression (LR) to verify the effectiveness and accuracy of the method used in this article. In this paper, Matlab software is used to analyze and visualize the detection result data.

**INDEX TERMS** Abnormal detection, LSTM, particle swarm optimization, attention mechanism, electricity theft modes.

## I. INTRODUCTION

There are often losses in power grid transmission and distribution transmission. The losses are mainly divided into technical loss (TL) and non-technical loss (NTL). The main cause of non-technical loss is abnormal electricity consumption by users' behaviors such as stealing electricity, fraud, etc [1]. Abnormal electricity consumption behavior of users will affect the dispatch operation of the regional power grid, interfere with the safety management of power supply, and bring hidden dangers to the security of the power grid. At the same time, with the reform of the national power market,

The associate editor coordinating the review of this manuscript and approving it for publication was Eklas Hossain.

independent power sales and distribution companies appear, and users' power theft would directly damage the profits of the power companies and disrupt the operation of the power market [2]. Abnormal electricity consumption behaviors of users are general in various countries. According to incomplete statistics, non-technical losses in China, India, Brazil and Mexico account for 6.42%, 26.2%, 16.85% and 15.83% of the national power consumption [3]. China has a large population base, a large demand for electricity, and increasing electricity consumption. The annual economic loss due to electricity theft across the country can reach several billion yuan [4]. Therefore, the research on the detection of abnormal electricity consumption of users is an urgent problem to be solved.

With the updation and development of smart grid and advanced metering infrastructure (AMI), smart meters have also begun to be widely used. Compared with traditional electricity meters, smart electricity meters can collect user data more accurately and efficiently, and can obtain massive amounts of user electricity consumption data, providing sample support for the analysis and detection of users' abnormal electricity consumption behavior. In contrast, using smart meters for data transmission is also vulnerable to attacks in communications and networks. Nowadays, electricity thief can use digital storage technology and network communication technology to attack smart meters, thereby tampering with data to reduce electricity bills, and attacks on smart meters are more ''invisible'' [5]. The traditional methods for detecting abnormal electricity consumption of users mainly use manual on-site verification of electricity meter information, judgment through experience, installation of monitoring equipment, etc [6]. These methods have low detection efficiency and high cost, cannot quickly and accurately determine abnormal electricity consumption behavior, and it is difficult to detect tampering in communications and networks [7].

In view of the shortcomings of traditional methods, some scholars have carried out a series of studies. At present, the detection methods for abnormal electricity consumption mainly include statistical models [8], [9], data-driven methods, and game theory-based methods [10]. With the improvement of user-side online monitoring systems and power information management systems, data-driven detection methods have received attention [11]. For power grid lines, Shah *et al.* in [12] proposed an algorithm that uses smart meter measurement technology to update the network cable impedance, detects and classifies technical losses and non-technical losses when errors occur in smart meters, and estimates the resulting losses. Wang *et al.* in [13] uses different clustering algorithms to detect 10kV non-technical losses based on the average loss, line loss variation coefficient and ammeter open-circuit records collected by the meter, and finally analyzes and compares the detection effects of various clustering algorithms. For electricity users, the data-driven detection method is mainly based on the difference between abnormal users' electricity consumption behavior characteristics and normal users, and analyzes and judges abnormal points through massive electricity consumption data. Zhang *et al.* in [14] proposed a detection model based on real-valued deep belief network (RDBN), which uses the firefly algorithm (FFA) to solve the local optimum, and uses undersampling and lasso algorithm to solve the data imbalance problem. Finally, The detection model achieves higher accuracy. Xu *et al.* in [15] constructed a random forest model, using sparseness combined with anomalous cumulant index to judge the anomaly of the sample, and the detection effect is good. Zhao *et al.* in [16] uses long short-term memory networks to extract sequence features, classifies sequence features through a full connected network (FCN), and judges abnormal users through classification. Buzau *et al.* in [17] uses all data recorded by smart meters (energy consumption, alarms and electrical magnitudes) to detect non-technical losses through supervised learning. The method has been developed and tested based on real smart meter data from Endesa's industry and customers. Buzau *et al.* in [18] uses a long and short-term memory network and a multi-layer perceptron hybrid deep neural network to detect anomalies and frauds in the electricity meter by analyzing daily energy consumption and geographic information, and finally tested it in the Spanish power company. Ghori *et al.* in [19] evaluated 9 types and 15 existing machine learning classifiers, and analyzed and compared the detection performance of various classifiers using the Pakistan Power Supply Company data set.

In the data-driven method, the regression method for anomaly detection method can consider user consumption behavior, and has a good detection effect for users with different consumption behaviors. The accuracy of the regression-based anomaly detection method is affected by the accuracy of user power consumption prediction. Compared with other methods, the LSTM model can better model dynamic time series data and has advantages for time series forecasting. But for too long time series, the LSTM model is easy to lose the sequence information. In addition, the adjustment of model hyperparameters depends on experience, which is complicated to adjust parameters and affects model prediction and detection accuracy.

Based on the above considerations, this paper introduces a PSO-Attention-LSTM model abnormal user power consumption detection. Based on the electricity consumption data of normal users and the corresponding weather characteristics, this paper uses Tensor Flow to establish an LSTM prediction model based on the attention mechanism, and predicts the future electricity consumption of normal users with a sliding window of 24 unit steps. According to the 6 kinds of electricity theft modes and the combination in reality, we obtain 4 kinds of compound electricity theft modes, and the normal user power consumption data is simulated as abnormal user data. We judge the time period of abnormal behavior according to the abnormality of the user power curve and the detection threshold. This paper uses the electricity data set of the University of Massachusetts as a simulation

example, and combines the PSO-optimized LSTM model based on the attention mechanism (PSO-Attention-LSTM) with CNN-LSTM, Attention-LSTM, LSTM, GRU, SVR, RF and LR Compare. Finally, confusion matrix, detection rate, false detection rate and radar chart are used to evaluate detection accuracy. The result proves that the detection model can accurately detect the user's abnormal behavior time period, which verifies the feasibility and accuracy of this method. This method can more accurately predict the change of user power consumption, and has certain reference value for the research of abnormal power consumption detection of users [20]–[31]. The main contributions to this paper are as follows.

1)According to the actual law of stealing electricity, we have constructed 6 stealing modes. And through the combination of these 6 power stealing modes, 4 composite modes are obtained. According to these 10 power stealing patterns, a data set of abnormal users is generated. Experiments have proved that the PSO-Attention-LSTM model has good detection performance for various power theft modes.

2)Using the LSTM model can fully consider the time series characteristics of user power consumption, and has a good time series data fitting regression ability. At the same time, the Attention mechanism is introduced to give different probability weights to the hidden states of LSTM, and strengthen the influence of important information, so that the model has better prediction accuracy and detection effect.

3)Use PSO optimized model hyperparameters and obtain optimal parameters, so that the model has more accurate detection performance. Through the establishment of comparative experiments, comparing with CNN-LSTM, Attention-LSTM, LSTM, GRU, SVR, RF and LR, it is proved that the model used in this paper has higher anomaly detection ability.

The rest of this paper is organized as follows. In Section II, the related work of the paper is introduced. Section III the principle of abnormal power consumption detection and the principle of LSTM sliding window prediction are introduced. Section IV introduces LSTM network, attention mechanism and particle swarm algorithm. Analyze and summarize 10 abnormal power consumption patterns. Section V introduces the simulation data and simulation settings, and conducts simulations to verify the accuracy and validity of the model. Finally, Section VI summarizes the work of this article and future work.

## II. RELATED WORK

### A. DETECTION METHOD BASED ON STATISTICS MODEL

The statistical model-based detection method mainly combines user-side smart meter electricity data, distribution network voltage, current, power and other network status data and network topology to establish a statistical model for abnormal electricity use detection. It is difficult for most users to perform data tampering to achieve data coordination to detect abnormalities. For example, Lo *et al.* in [32] uses the weighted least squares method to estimate the system state through the topology of the distribution system, the voltage of each node and the reactive power, and establishes the system objective function for anomaly detection. This method has high detection accuracy and low false detection rate. However, this method depends on the topology and parameters of the distribution network, and the topology and parameters of the distribution network are not constant. It is noted that, the functional relationship of the parameters changes after data tampering, and the model may have convergence problems [4].

### B. DETECTION METHOD BASED ON GAME THEORY

This method assumes that each user's decision-making behavior is to maximize their own interests, and detects abnormal users based on the difference between the decision set of the stealing user and the normal user. Saurabh *et al.* in [10] established a user-distribution company game model for anti-theft. Amin *et al.* in [33] applied the likelihood ratio test to anomaly detection, and established a game model by discussing more parameters such as electricity price and the proportion of stealers. This method has only undergone theoretical simulation, and has not yet been verified.

### C. DATA-DRIVEN DETECTION METHOD

With the improvement of user-side online monitoring systems and power information management systems, data-driven detection methods have received more and more attention. Data-driven detection methods are mainly divided into three types: classification-based, clustering-based, and regression-based.

1)Classification-based approach: The classification-based method uses the characteristics of electricity consumption data to classify normal and abnormal. For example, Zhang *et al.* in [14] uses the detection model of real-valued deep confidence network, uses the firefly algorithm to solve the local optimum, and finally detects anomalies through classification. Xu *et al.* in [15] uses a sparse random forest model to classify the electricity consumption behavior on the electricity consumption side and detect abnormal behavior. Zhao *et al.* in [16] used LSTM for feature extraction and classified it through a fully connected network (FCN). Buzau *et al.* in [17], the XGBoost classifier is used for non-technical loss detection. Buzau *et al.* in [18] uses long-term and short-term memory networks and multilayer perceptron hybrid deep neural networks, which have higher accuracy than other classifiers. Ghori *et al.* in [19], using the data set of Pakistan Electric Power Company, the detection performance of 15 classifiers was tested. Zheng *et al.* in [34] used deep convolutional neural network (CNN) to classify data sets to detect abnormal electricity consumption. Ibrahem *et al.* in [35] proposed the ETDFE scheme, which can use the machine learning classifier model for electricity theft detection while protecting user privacy. This method exploited the inner-product operations on encrypted readings to evaluate

a machine-learning model to detect fraudulent consumers. This method has good detection accuracy, but the modeling is complicated and requires label data.

2)Cluster-based method: This method mainly divides the data set into different sub-data sets according to the characteristics through a specific algorithm. Wang *et al.* in [13] uses different clustering algorithms to detect 10kV non-technical losses, and compares the detection effects of various algorithms. Passos Júnior *et al.* in [36] use an optimal path forest clustering method for detection. Tian *et al.* in [37] used the density-based spatial clustering of applications with noise (DBSCAN) clustering algorithm to cluster the fluctuation interval of the user's electricity load curve, and divided all data points into core points, reachable points and abnormalities point. This method does not require label data and is widely used, but there are problems in parameter selection, the detection performance is poor.

3)Regression-based method: This method mainly uses short-term load forecasting for users, and judges abnormal points based on the deviation between the actual power consumption and the predicted amount. Liu *et al.* in [38] uses an attention mechanism-based convolutional neural network-long short-term memory model for abnormality detection. In the case of protecting user privacy, feature extraction is used to predict time series data to detect abnormalities such as failure or shutdown of the electricity meter. This model conforms to the timeliness of industrial anomaly detection, and can quickly detect the failure or shutdown of edge devices. This method optimizes the feature extraction of the convolutional neural network through the attention mechanism, and uses the long and short-term memory model to learn the consumption characteristics, so as to accurately detect the abnormality of the industrial edge equipment. This article focuses on the detection of abnormal electricity consumption behavior of residential users. The attention mechanism is used to strengthen the long and short-term memory model's learning of important information, suppress useless information, and improve prediction accuracy. The particle swarm algorithm is used to optimize the hyperparameters of the long and short-term memory model, which solves the complexity of manual parameter adjustment and further improves the prediction accuracy. This method can more accurately predict the consumption behavior of residential users and accurately detect abnormal behaviors of users.

## III. PROBLEM DESCRIPTION

The abnormal user electricity consumption detection mainly judges abnormal users through the abnormal degree of the user power consumption data. A user's abnormal electricity consumption behavior will cause differences in electricity consumption data from the electricity consumption data of normal users of the same type. The electricity consumption curve, current curve, voltage curve, power curve and other electricity consumption data during these abnormal behavior periods will all change. The abnormality can be judged by the degree of difference between the abnormal electricity usage data and the normal electricity usage data. In this paper, anomaly detection is carried out through user electricity consumption. In order to reduce electricity costs, abnormal users adopt various methods of stealing electricity to tamper with the user's electricity consumption. Therefore, the user electricity consumption data is abnormal, and the user electricity consumption is available, which can effectively detect abnormal users.

Nowadays, the number of electricity users is huge, so the research on abnormal electricity consumption detection mainly focuses on the high-efficiency detection ability of large-scale users, which can quickly and effectively screen out suspected abnormal users among the same types of users. However, there are some users with relatively large electricity consumption characteristics among normal users, and their detection methods are easy to ignore the individual user's own power consumption habits, which are prone to misjudgment, and rapid investigation of anti-electricity stealing personnel increases difficulty. Therefore, this article considers the user's electricity consumption habits, customizes the model for the user, and detects abnormal users by predicting the degree of deviation between the user's electricity consumption data for a period of time in the future and the actual electricity consumption. This method can make up for each other with the abnormal detection of large-scale users, detect the suspected abnormal users after large-scale detection, and eliminate the misdetection caused by special electricity consumption habits. At the same time, the prediction performance of the attention mechanism model is adopted, and the particle swarm optimization is used to optimize the adjustment of model hyperparameters to improve the accuracy of anomaly detection.

For abnormal user detection, first of all, it is necessary to accurately predict the user's electricity consumption data in the future period. For this time series with massive data, LSTM has certain advantages. LSTM can dig out the user's electricity consumption characteristics and laws from historical information, has a long-term memory function, and more accurately predicts the electricity consumption data in the future when considering the user's electricity consumption habits. At the same time, particle swarm optimization and attention mechanism are used to improve the prediction accuracy and detection performance of the LSTM model. According to the periodicity and trend of users' electricity consumption behavior [39], [40], this paper uses one-day data to do window sliding processing, that is, the data of the previous 24 time steps predict the data of the next time point. The specific prediction principle is shown in Figure 1.

## IV. THE ANOMALY DETECTION MODEL

### A. INTRODUCTION TO LSTM

LSTM network is a special recurrent neural network(RNN). By adding a new memory unit $c$, it solves the problems of gradient disappearance and gradient explosion in RNN, and improves the reliability of the model [41]. It has a feedback

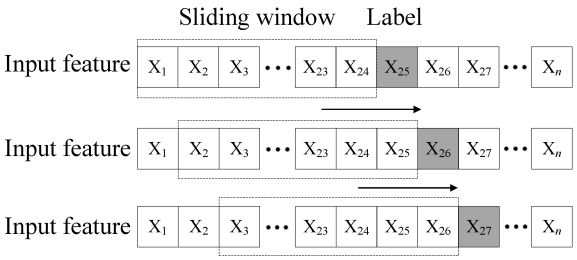


**FIGURE 1.** LSTM sliding window prediction principle.

structure, can mine data characteristics in historical information, and has advantages in time series forecasting. The LSTM network structural unit is shown in Figure 2.

LSTM is mainly divided into input gate, forget gate and output gate. The input gate controls the input at the current moment, the output gate controls the output at the current moment, and the forget gate controls the state at the previous moment. Formula (1) is the specific calculation of the forget gate, formulas (2)-(3) is the specific calculation of the input gate, formula (4) is the update formula of the memory unit, and formula (5)-(6) is the specific calculation of the output gate [42].

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{1}$$
$$i_t = Sigmoid(W_i[h_{t-1}, x_t] + b_i) \tag{2}$$
$$c'_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{3}$$
$$c_t = c_{t-1} \cdot f_t + i_t \cdot c'_t \tag{4}$$
$$o_t = Sigmoid(W_o[h_{t-1}, x_t] + b_o) \tag{5}$$
$$h_t = \tanh(c_t) \cdot o_t \tag{6}$$

In the formula, $h_{t-1}$ and $h_t$ are the output at the previous moment and the current moment respectively, $x_t$ is the input at the current moment, $W$ and $b$ are the weight matrix and bias in the network respectively, $f_t$ is the output of the forget gate, and $i_t$ is the input gate output, $o_t$ is the input of the output gate, $c'_t$ and $c_t$ are the current state of the input and output memory unit.

### B. ATTENTION-LSTM MODEL

The attention mechanism [43] is a resource allocation mechanism that simulates the attention of the human brain. It mainly changes the attention to information, thereby increasing useful information and ignoring useless information. Focus on important information, get more detailed information, suppress and ignore useless information. The attention mechanism is used to effectively highlight the key features that affect the user's power consumption in the prediction results of the LSTM layer, and improve the prediction performance and detection effect of the model.

The Attention-LSTM model mainly includes the input layer, LSTM layer, Attention layer, and Output layer. In this paper, the Attention layer is added behind the LSTM layer, and the input layer of the Attention layer is the feature vector output by the LSTM layer. The probability distribution value

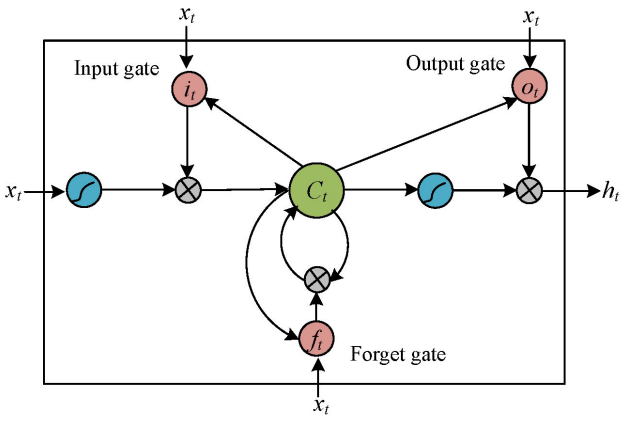


**FIGURE 2.** LSTM network structure unit.

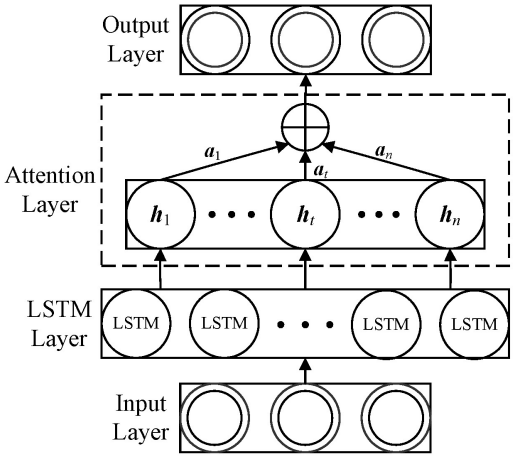


**FIGURE 3.** Attention-LSTM model structure.

of the feature vector is calculated by the features learned by the LSTM layer according to the weight distribution principle, and better weight parameters are obtained by updating iteratively. Finally, through the fully connected layer, the final user power consumption forecast value is output. The structure of the Attention-LSTM model is shown in Figure 3.

Among them, the calculation formula of Attention's weight coefficient is as follows:

$$e_t = u\tanh(wh_t + b) \tag{7}$$
$$a_t = \frac{\exp(e_t)}{\sum_{i=1}^{n}\exp(e_i)} \tag{8}$$
$$s_t = \sum_{t=1}^{n} e_t a_t \tag{9}$$

In the formula, $e_t$ is the important feature of the LSTM layer output vector $h_t$ at the $t$-th time. $u$ and $w$ are the weight coefficients, and $b$ is the bias coefficient. $s_t$ is the output of attention at time $t$.

### C. INTRODUCTION TO PSO

The particle swarm optimization mainly seeks the optimal solution through information exchange and mutual cooperation between individuals among groups. In the algorithm,

each individual is called a particle. All particles have their fitness value, and the quality of the particles is judged based on the fitness value. Each particle has two variables: speed and position, which are mainly updated based on the optimal position in the group and the optimal position in the individual history.

Suppose there are m particles in a d-dimensional space, and the update formula for the velocity and position of each particle is:

$$v_{id}^{n+1} = \omega v_{id}^{n} + c_1 r_1 (p_{id}^{n} - x_{id}^{n}) + c_2 r_2 (p_{gd}^{n} - x_{id}^{n}) \quad (10)$$

$$x_{id}^{n+1} = x_{id}^{n} + v_{id}^{n+1} \quad (11)$$

where $v_{id}^{n}$ is the velocity of the $d$-th dimension component of the $i$-th particle in the $n$-th iteration. $x_{id}^{n}$ is the position of the $d$-dimensional component of the $i$-th particle in the $n$-th iteration. $p_{id}^{n}$ is the individual optimal $d$-th dimension component of the $i$-th particle in the $n$-th iteration. $p_{gd}^{n}$ is the $d$-th dimension component of the optimal population of the $i$-th particle in the $n$-th iteration. $n$ is the number of iterations. $c_1$ and $c_2$ are learning factors. $r_1$ and $r_2$ are random numbers between 0 and 1. $\omega$ is the weight of inertia.

### D. THE OVERALL FRAMEWORK OF ANOMALY DETECTION MODEL

This paper is based on the PSO-Attention-LSTM model to detect abnormal electricity consumption of users. Firstly, predict the user's electricity consumption curve in the future based on the user's historical electricity consumption data and weather attributes, and then calculate the degree of abnormality between the actual electricity consumption curve and the predicted electricity consumption curve, and finally determine abnormal users according to the threshold. The overall framework of the specific user abnormal electricity consumption detection model is shown in Figure 4:

The main steps of the abnormal power consumption detection model based on PSO-Attention-LSTM are as follows:

Step 1: Construct a normal user data set with the processed user electricity consumption and corresponding weather data as the input data of the PSO-Attention-LSTM model. According to the 6 types of electricity theft modes and the characteristics of abnormal users' electricity consumption curves, 10 electricity theft patterns are combined and established. Finally, the data set simulated according to the electricity stealing patterns is used as the actual output data of abnormal users.

Step 2: According to the normal user data set output in step 1, build an PSO-Attention-LSTM model. The particle swarm algorithm is used to optimize model hyperparameters. The attention mechanism retains important information in the data, suppresses useless information, and improves model performance.

Step 3: Calculate the degree of deviation between the electricity consumption prediction curve of the PSO-Attention-LSTM model and the actual electricity consumption curve in the abnormal user data set, and finally detect whether the user has abnormal behavior.

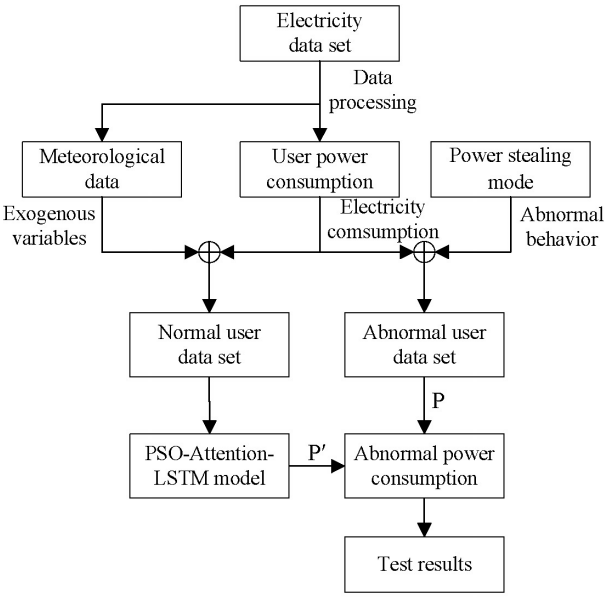


**FIGURE 4.** Overall flow chart of anomaly detection.

### E. ABNORMAL USER DATA SET BASED ON ELECTRICITY THEFT MODE

Because the data set used in the article is the electricity consumption data of normal users, the data needs to be transformed according to the electricity theft mode to simulate and generate the electricity consumption data set of abnormal users. Abnormal power usage behavior of users is not stochastic, and its ultimate goal is to reduce the electricity bill that needs to be paid, so its abnormal behavior has a certain law. The typical electricity consumption curve of a user who steals electricity is a small amount of electricity used continuously for a long time, and the curve is stable [44], that is, it has the characteristics of continuous, small amount, and stability.

Based on the characteristics of the electricity consumption curve of users who steal electricity and the fact that there are 6 typical electricity theft modes [45], [46], this paper establishes a data set of abnormal users under typical electricity theft modes. Through the analysis and combination of 6 typical electricity theft modes, a data set of abnormal users under 4 compound modes is finally established. Typical electricity theft modes are shown in formulas (7)-(12). Among them, $P_{i:j}$ and $P'_{i:j}$ are the electricity consumption data of the normal user and the simulated abnormal user from the $i$-th hour to the $j$-th hour, and $\bar{P}$ is the average electricity consumption of the normal user in the previous month. In mode 2 and mode 3, $\min(P_{i:j}) < P < \max(P_{i:j})$. In mode 4, $i < m < n < j$. In mode 5 and mode 6, $0 < \alpha(t) < 1$, $i \le t \le j$.

$$\text{Mode 1: } P'_{i:j} = \alpha \cdot P_{i:j}, \{0 < \alpha < 1\} \quad (12)$$

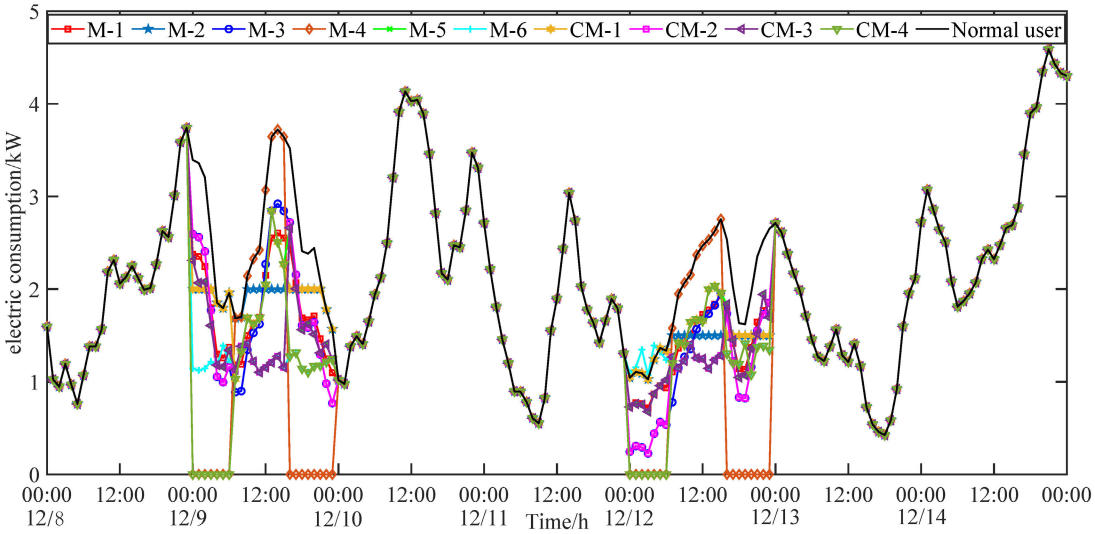


**FIGURE 5.** Comparison chart of 10 power stealing modes.

$$\text{Mode 2: } P'_{i:j} = \begin{cases} P_{i:j} & P_{i:j} \le P \\ P & P_{i:j} > P \end{cases} \tag{13}$$

$$\text{Mode 3: } P'_{i:j} = \max(P_{i:j} - P, 0) \tag{14}$$

$$\text{Mode 4: } P'_{i:j} = f(t) \cdot P_{i:j},$$

$$f(t) = \begin{cases} 0 & t \in [m, n] \\ 1 & t \in (i, m) \cup (n, j) \end{cases} \tag{15}$$

$$\text{Mode 5: } P'_{i:j} = P_{i:j} \cdot \alpha(t) \tag{16}$$

$$\text{Mode 6: } P'_{i:j} = \bar{P} \cdot \alpha(t) \tag{17}$$

Among the above-mentioned 6 typical electricity theft modes, mode 1 and mode 5 have similarities, and it is of little significance to combine with each other, so only one of them is selected to be combined with other modes. Modes 2, 3, and 4 also have similarities, and only one of them is selected for combination. Therefore, the 6 typical electricity theft modes can be divided into mode 1, mode 5, mode 2, mode 3, mode 4, mode 6 three types, each of which selects one mode and combines them to create a total of 4 composite modes. Among them, composite mode 1 is a combination of mode 2 and mode 5, composite mode 2 is a combination of mode 3 and mode 6, composite mode 3 is a combination of mode 5 and mode 6, and composite mode 4 is a combination of mode 4, mode 5 and mode 6. The compound mode is shown in formulas (13)-(16).

$$\text{CM 1: } P'_{i:j} = \begin{cases} P(t) & P(t) \le P \\ P & P(t) > P \end{cases} \tag{18}$$

where $P(t)$ is consistent with mode 5, $0<\alpha(t)<1$, $i \le t \le j$ and $\min(P(t))<P<\max(P(t))$.

$$\text{CM 2: } P'_{i:j} = \max(P(t) - P, 0), \tag{19}$$

where $P(t)$ is consistent with mode 6, $0<\alpha(t)<1$, $i \le t \le j$ and $\min(P(t))<P<\max(P(t))$.

$$\text{CM 3: } P'_{i:j} = \begin{cases} \bar{P} \cdot \alpha(t) & t \in [m, n] \\ P_{i:t+t:j} \cdot \alpha(t) & t \in [i, m) \cup (n, j] \end{cases} \tag{20}$$

Among them, $0.6<\alpha(t)<0.8$, $i<m<n<j$.

$$\text{CM 4: } P'_{i:j} = \begin{cases} P_{i:m} \cdot f(t) & t \in [i, m) \\ P_{m:n} \cdot \alpha(t) & t \in [m, n] \\ \bar{P} \cdot \alpha(t) & t \in (n, j] \end{cases} \tag{21}$$

where $P(t)$ is consistent with mode 4, $i \le t < m$, $i < m < n < j$ and $0.6<\alpha(t)<0.8$.

According to the above-mentioned electricity theft mode, the power consumption data of abnormal users is simulated. Figure 5 shows the selected normal power consumption curve from Dec $8^{th}$, 2016 to Dec $14^{th}$, 2016, and the abnormal electricity consumption curve under 6 electricity theft modes and 4 combined modes.

### F. ANOMALY DETECTION MODEL CONSTRUCTION

The data set is divided into historical electricity consumption data and current electricity consumption data. The historical electricity consumption data is used to train the model, and the PSO-Attention-LSTM prediction model is established to predict the current electricity consumption data of the user. The abnormal user electricity consumption data simulated above is used as the actual electricity consumption data at the current moment, and compared with the predicted current electricity consumption data, the abnormal point is judged by the threshold. The specific construction process of the detection model based on PSO-Attention-LSTM is shown in Figure 6.

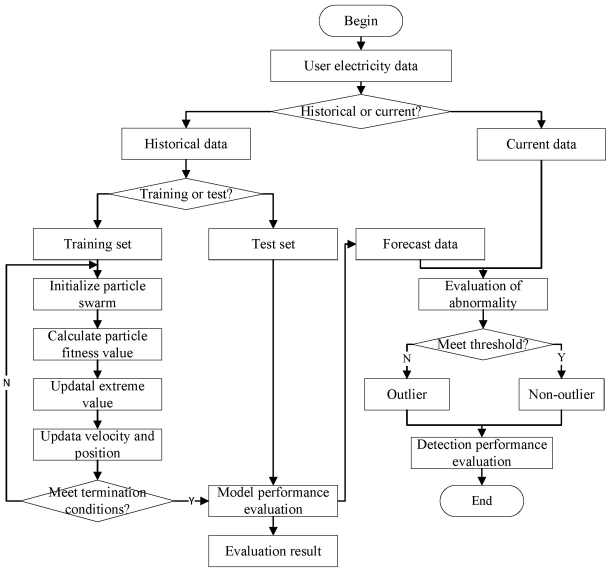


**FIGURE 6.** PSO-Attention-LSTM model construction process.

## V. EXPERIMENTAL RESULTS

### A. DATA ACQUISITION AND PROCESSING

The data used in this article comes from the public data set of the University of Massachusetts [47], which records user electricity consumption records from Oct $7^{th}$, 2014 to Dec $15^{th}$, 2016. The weather data record corresponds to the weather changes from Jan $1^{st}$, 2014 to Dec $31^{st}$, 2016, mainly hourly temperature, humidity, body temperature, etc.

The original data comes from the electricity consumption data of real users, so the data will be affected by external influences and there will be missing and ''noise''. In order to ensure the accuracy and reliability of the simulation, for the existing missing data and wrong data, linear interpolation is used to fill in the data. The paper selects hourly electricity consumption data for a total of 793 days from Oct $14^{th}$, 2014 to Dec $14^{th}$, 2016. The weather data is selected to correspond to the user's temperature, humidity, and body temperature, totaling 19,932 data.

In order to ensure the stability and speed of model training, the data stochastic is normalized. The specific normalization formula is shown in formula (17).

$$\tilde{P} = \frac{P - P_{\min}}{P_{\max} - P_{\min}} \tag{22}$$

In the formula, $\tilde{P}$ is the user's electricity consumption after normalization, $P$ is the user's electricity consumption before normalization, and $P_{\max}$ and $P_{\min}$ are the user's maximum and minimum electricity consumption respectively.

### B. SIMULATION SETTINGS

In the data set used in this article, 18,696 pieces of data with a total of 779 days from Oct $14^{th}$, 2014 to Dec $1^{st}$, 2014 are used as the training set for PSO-Attention-LSTM model training. The remaining 168 pieces of data for a total of 7 days from Dec $1^{st}$, 2016 to Dec $7^{th}$, 2016 are used as the test set. Use the data from Dec $8^{th}$, 2016 to Dec $14^{th}$, 2016 to simulate abnormal user electricity consumption data. Among them, select Dec $9^{th}$ where the curve ''peak'' and ''valley'' are obvious and Dec $12^{th}$ where the curve is relatively stable perform abnormal behavior simulation to test the model's ability to detect abnormal users in different states. Compare the electricity consumption curve predicted by the model with the simulated abnormal electricity consumption curve to determine the abnormal point. This article uses Tensor Flow to build a detection model based on PSO-Attention-LSTM. The model uses MSE loss function and Adam optimizer, and uses particle swarm optimization to optimize the number of hidden layers, iterations, and batch samples. In partical swarm optimization, the inertia weight $\omega$ is set to 0.5, the learning factors $c_1$ and $c_2$ are both set to 2, the population size is set to 20, and the maximum number of evolutionary iterations is set to 100.

### C. SIMULATION RESULTS AND ANALYSIS

#### 1) PSO-ATTENTION-LSTM MODEL PREDICTION RESULTS

To accurately detect abnormal electricity consumption behavior of users, first accurately predict the electricity consumption of normal users, and judge the abnormal points based on the deviation between the predicted electricity consumption data and the actual abnormal behavior electricity consumption data. This paper uses the historical usage data of normal users and the corresponding weather attributes (air temperature, humidity, apparent temperature) to predict the electricity consumption curve of users in the future. In this paper, root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and absolute error (AE) are used to evaluate model performance. Confusion matrix, positive rate (PR) and false positive rate (FPR) are used to evaluate detection performance. The specific evaluation formula is as follows.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|P'_i - P_i\right|^2} \tag{23}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|P'_i - P_i\right| \tag{24}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|P'_i - P_i\right|}{|P_i|} \tag{25}$$

$$AE = \left|P'_i - P_i\right| \tag{26}$$

$$M = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \tag{27}$$

$$PR = \frac{TP}{TP + FN} \tag{28}$$

$$FPR = \frac{FP}{FP + TN} \tag{29}$$

where, $i$ is the label of the data sample, that is, hourly time; $n$ is the total number of data samples; $P_i$ is the actual electricity consumption of normal users; $P'_i$ is the normal user electricity

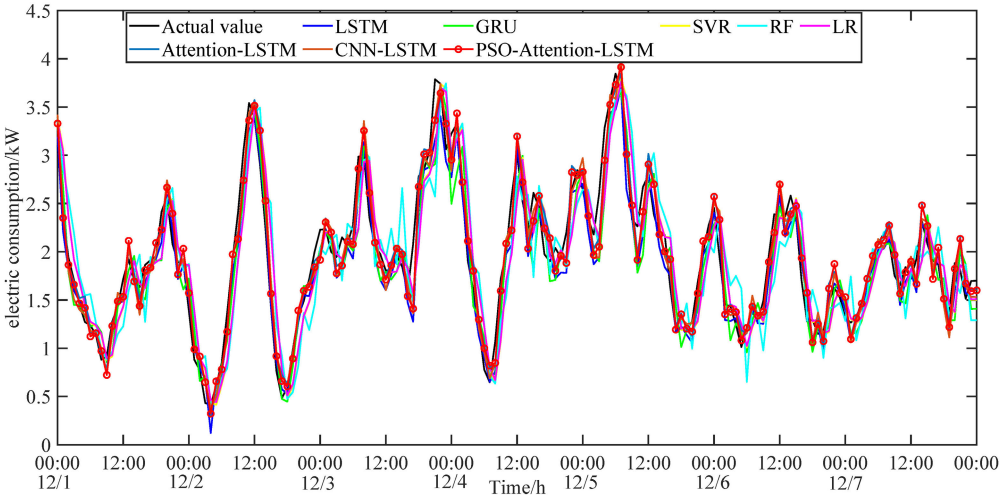


**FIGURE 7.** Forecast value of electricity consumption of 8 models.

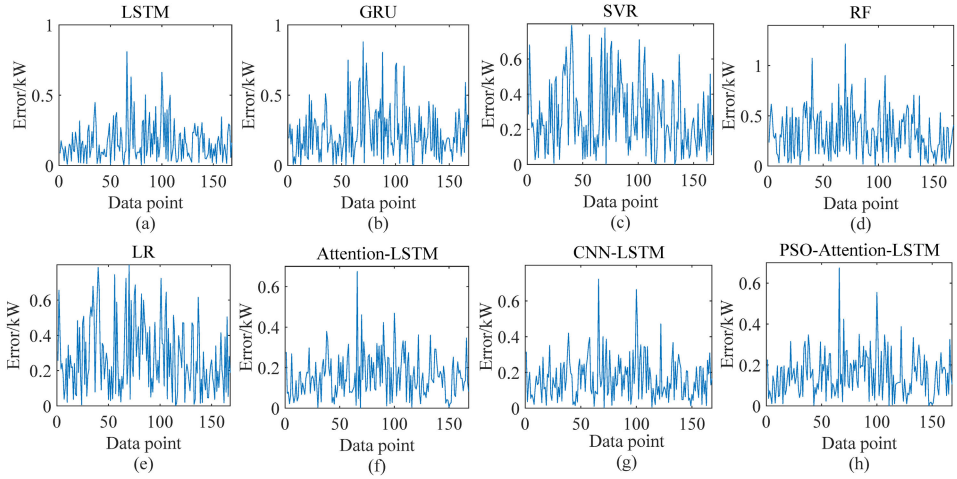


**FIGURE 8.** Comparison chart of absolute error of 8 models.

consumption predicted by the model; *TP* is the actual abnormal electricity consumption data is detected as an abnormal point; *FN* is the actual abnormal electricity consumption data is detected as a normal point; *FP* is the actual normal electricity consumption data is detected as an abnormal point; *TN* is the actual normal electricity consumption data is detected as a normal point.

This article uses SVR, RF, LR, GRU, LSTM, CNN-LSTM and Attention-LSTM for comparison. Figure 7 shows the predicted curves and actual electric consumption curves of the five methods. Figure 8 shows the absolute error curves of the five methods. Table 1 shows the error indicators of the five methods.

Figure 7 shows the actual electricity consumption curve from December $1^{st}$ to December $7^{th}$, 2016, and the forecast curves of the five models. Table 1 and Figure 8 show the corresponding error curves and error indicators. Table 1 shows the MAE, MRE and RMSE of the 8 methods. The three error indicators of PSO-Attention-LSTM are all the smallest, and RF has the largest error. It can be seen in Figs. 7 and 8 that PSO-Attention-LSTM has high individual point errors, but the error curve is relatively lowest, the prediction curve is the best, and the RF error curve is the highest. The data used in this article are residential electricity consumption data, the electricity consumption is small, the fluctuation is large, and there are certain errors in data collection and transmission, so there is a certain ''burr'' in the error curve. Most of the ''burrs'' in the PSO-Attention-LSTM error curve are the errors of the ''peak'' and ''valley'' points of the electricity consumption curve, and the overall error is small. In summary, the PSO-Attention-LSTM model has better accuracy than other models and can accurately predict the electricity consumption curve.

### 2) SIMULATION RESULTS AND ANALYSIS OF ABNORMAL ELECTRICITY CONSUMPTION DETECTION MODEL

The electricity consumption of users in the future period predicted by the model is compared with the actual value of abnormal electricity consumption in 10 simulated power theft modes, and abnormality is evaluated through absolute error

**TABLE 1. 8 methods of prediction error indicators.**

| | LSTM | GRU | SVR | RF | LR | Attention-LSTM | CNN-LSTM | PSO-Attention-LSTM |
|---|---|---|---|---|---|---|---|---|
| MAPE | 0.0932 | 0.1259 | 0.1588 | 0.191 | 0.1581 | 0.0904 | 0.920 | 0.0866 |
| MAE/kW·h | 0.1663 | 0.2324 | 0.2752 | 0.3244 | 0.2733 | 0.1576 | 0.1652 | 0.1522 |
| RMSE/kW·h | 0.2162 | 0.2949 | 0.3399 | 0.3957 | 0.337 | 0.1905 | 0.2047 | 0.1861 |

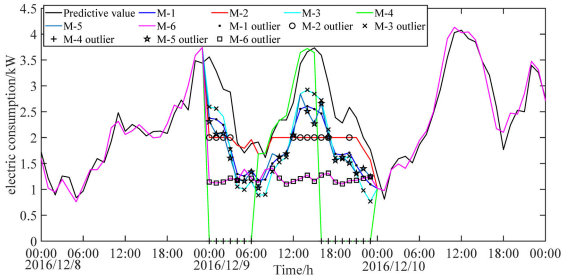


**FIGURE 9. 6 electricity consumption curve test results of electricity theft modes (Dec 9$^{th}$, 2016).**

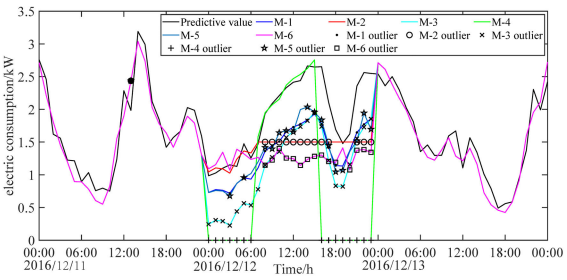


**FIGURE 10. 6 electricity consumption curve test results of electricity theft modes (Dec 12$^{th}$, 2016).**

and relative error, and finallythe abnormal point is judged according to the threshold. For the setting of the threshold, setting too small will cause frequent false detections, and setting too large will result in failure to detect. In order to prevent as much as possible the misdetection caused by model errors caused by the randomness of users' electricity consumption, the threshold should not be set too small, and users' electricity consumption habits need to be considered. At the same time, in order to detect as many abnormal user behaviors as possible, the threshold setting should not be too large, and the threshold needs to be set according to the user type. Combining the proportion of residents' electricity stealing and the average electricity consumption of residents, the relative error threshold is set to 0.2, and the absolute error threshold is set to 0.39kW·h. There is a certain error in the detection of the ''peak'' value generated by the user's randomness in the detection model, that is, the ''burr'' in the error curve, and only using the absolute error threshold is prone to misjudgment. At the same time, since the ''valley'' value of the electricity consumption curve is less than 1, the relative error is relatively large. Only using the relative error threshold is easy to misjudge the ''valley'' value of the electricity consumption curve. In summary, two thresholds are used in the article to prevent false detections as much as possible. Figures 9 and 10 show the detection results of the 6 single power theft modes of the LSTM detection model from Dec 8$^{th}$ to Dec 10$^{th}$, 2016 and from Dec 11$^{th}$ to Dec 13$^{th}$, 2016. Figures 11 and 12 show the relative error curves and absolute error curves of the 6 power theft modes from Dec 8$^{th}$ to Dec 14$^{th}$, 2016.

Figures 9 and 10 respectively select Dec 9$^{th}$ and Dec 12$^{th}$ for electricity theft. It can be seen from Figure 9-10 that the method used in the article can well detect the time period of abnormal behavior for the abnormal electricity consumption curve under the 6 electricity theft modes. The misjudgment points in the figure are mainly due to the sharp rise and fall of the curve, which makes the LSTM unable to accurately

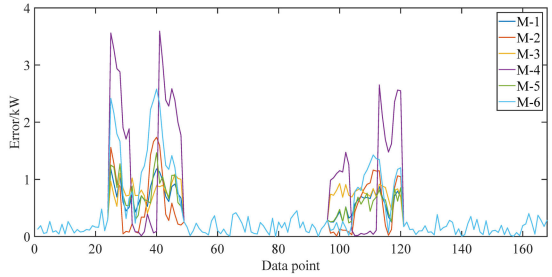


**FIGURE 11. Absolute error curve of 6 electricity theft modes.**

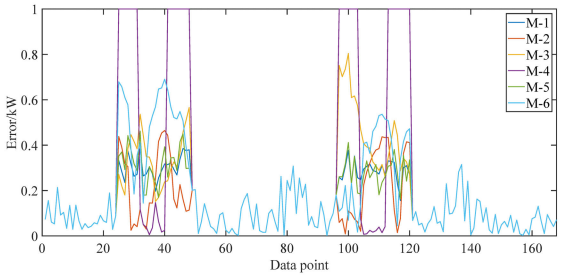


**FIGURE 12. Relative error curve of 6 electricity theft modes.**

predict, so they are judged as abnormal points. But overall, the method used in this article can accurately detect abnormal behavior. Figures 11 and 12 are the absolute and relative error curves of the six electricity theft modes. The two error curves of Mode 4 have the highest degree of abnormality, so the detection effect is the best. In mode 4, the electricity consumption is randomly set to zero, resulting in a relative error curve of 1 in the abnormal period, and a larger absolute error curve, so the detection accuracy of abnormal points is the highest. The error curve of mode 5 is relatively low, and its anomaly detection effect is the worst. Mode 5 modifies the current month's electricity consumption based on the average electricity consumption of the previous month. Therefore, its electricity consumption curve is relatively flat and fluctuates above and below the average value. The error is relatively low, which makes it impossible to accurately detect abnormalities.

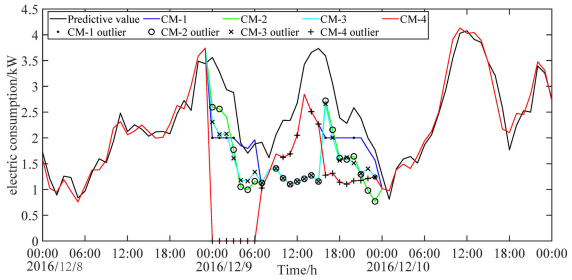


**FIGURE 13.** 4 kinds of composite mode electricity consumption curve test results(Dec 9$^{th}$, 2016).

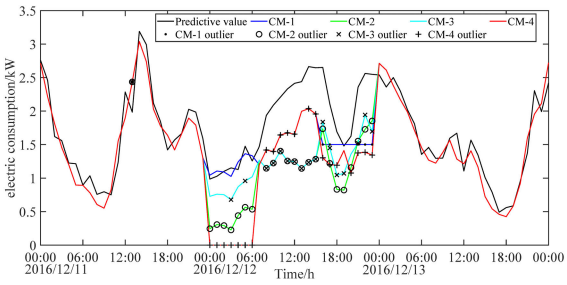


**FIGURE 14.** 4 kinds of composite mode electricity consumption curve test results(Dec 12$^{th}$, 2016).

Figure 13 and Figure 14 show the detection results of the four composite electric stealing modes of the PSO-Attention-LSTM detection model from Dec 8$^{th}$, 2016 to Dec 10$^{th}$, 2016 and from Dec 11$^{th}$ to Dec 13$^{th}$, 2016. Figures 15 and 16 show the relative error curves and absolute error curves of the four power theft modes from Dec 8$^{th}$ to Decr 14$^{th}$, 2016.

It can be seen from Figure 15 and 16 that the two abnormality curves during the period of compound mode 1 abnormal behavior are the lowest. In Figure 13 and 14, it can also be seen that composite mode 1 detects the least abnormal points. Therefore, composite mode 1 has the worst anomaly detection effect. It can be seen from the above that the absolute error of mode 5 at the ''valley'' point is relatively small, and the detection effect of mode 5 is the worst compared to other modes. The mode 2 is mainly to steal electricity by modifying the ''peak'' point, so the detection effect of the ''valley'' point of the compound mode 1 is poor. The confusion matrix was used to count the detection results of PSO-Attention-LSTM and compared with the detection results of GRU, SVR, RF, LR, CNN-LSTM and Attention-LSTM. Table 2 is the confusion matrix of the detection results of the abnormal points of the 6 power theft modes, and Table 3 is the confusion matrix of the detection results of the abnormal points of the 4 composite modes. According to the confusion matrix, the detection rate and false detection rate are used to evaluate the detection accuracy. Figure 17 is a radar chart of the positive rate and false positive rate of the 10 electricity theft modes for 5 detection methods.

It can be seen from Table 2 and Table 3 that the LSTM detection model in the electricity stealing mode and the compound stealing mode both detects more abnormal points and fewer false detection points. Figure 17 also shows that the

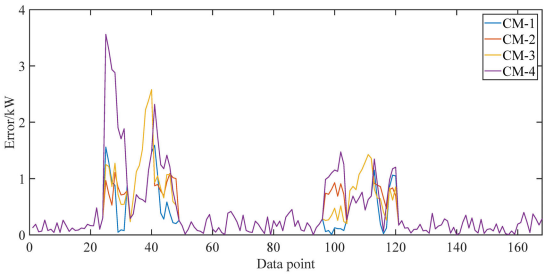


**FIGURE 15.** Absolute error curve of 4 compound modes.

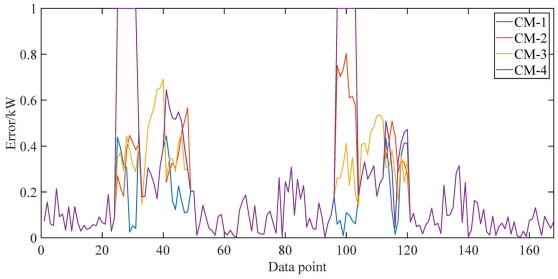


**FIGURE 16.** Relative error curve of 4 compound modes.

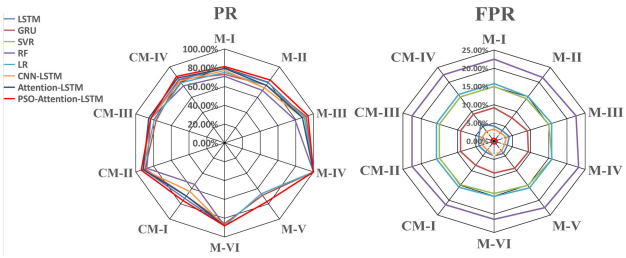


**FIGURE 17.** Radar chart of model positive rate and false positive rate.

positive rate is the largest and the false positive rate is the smallest. LSTM has advantages in processing time series data, and can learn users' electricity consumption habits, so the forecasting electricity consumption curve is accurate and the detection performance is good. Compared with the LSTM model, the PSO-Attention-LSTM model further optimizes the parameters, and the detection effect is the best. Relatively speaking, the positive rate of RF is relatively the smallest, the false positive rate is the largest, and the detection effect is the worst. Electricity consumption data has noise due to the randomness of electricity consumption, and RF is prone to over-fitting, resulting in the worst detection effect. GRU also has advantages in processing time series data, so the detection effect is stronger than SVR, RF and LR. It can also be seen in Figure 17 that the positive rates of SVR and LR are basically the same, and the false positive rate of SVR is slightly lower than that of LR, and the detection effect is similar. The positive rate of CNN-LSTM and Attention-LSTM is similar to that of LSTM, but the false positive rate is lower than that of the LSTM model, which reduces false detections. In summary, the detection model based on PSO-Attention-LSTM has better anomaly detection capabilities than the model. The detection results of this

**TABLE 2.** 6 kinds of typical electricity theft mode abnormal point detection results.

| | | M-1 | | M-2 | | M-3 | | M-4 | | M-5 | | M-6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Outlier | Normal point | Outlier | Normal point | Outlier | Normal point | Outlier | Normal point | Outlier | Normal point | Outlier | Normal point |
| LSTM | Outlier | 38 | 5 | 23 | 5 | 43 | 5 | 30 | 5 | 37 | 5 | 37 | 5 |
| | Normal point | 10 | 115 | 6 | 134 | 5 | 115 | 0 | 133 | 11 | 115 | 5 | 121 |
| GRU | Outlier | 35 | 11 | 22 | 11 | 44 | 11 | 30 | 13 | 33 | 11 | 36 | 11 |
| | Normal point | 13 | 109 | 7 | 128 | 4 | 109 | 0 | 125 | 15 | 109 | 6 | 115 |
| SVR | Outlier | 36 | 18 | 22 | 20 | 43 | 18 | 30 | 22 | 32 | 18 | 37 | 18 |
| | Normal point | 12 | 102 | 7 | 119 | 5 | 102 | 0 | 116 | 16 | 102 | 5 | 108 |
| RF | Outlier | 34 | 27 | 20 | 30 | 38 | 27 | 30 | 32 | 32 | 27 | 37 | 27 |
| | Normal point | 14 | 93 | 9 | 109 | 10 | 93 | 0 | 106 | 16 | 93 | 5 | 99 |
| LR | Outlier | 36 | 19 | 22 | 21 | 43 | 19 | 30 | 22 | 32 | 19 | 37 | 19 |
| | Normal point | 12 | 101 | 7 | 118 | 5 | 101 | 0 | 116 | 16 | 101 | 5 | 101 |
| CNN-Attention | Outlier | 37 | 4 | 21 | 4 | 42 | 4 | 30 | 4 | 37 | 4 | 37 | 5 |
| | Normal point | 11 | 116 | 8 | 135 | 6 | 116 | 0 | 135 | 11 | 117 | 5 | 121 |
| Attention-LSTM | Outlier | 38 | 3 | 22 | 3 | 42 | 3 | 30 | 3 | 37 | 3 | 37 | 5 |
| | Normal point | 10 | 117 | 7 | 136 | 6 | 117 | 0 | 135 | 11 | 117 | 5 | 121 |
| PSO-Attention-LSTM | Outlier | 39 | 1 | 24 | 1 | 45 | 1 | 30 | 1 | 37 | 1 | 37 | 1 |
| | Normal point | 9 | 119 | 5 | 138 | 3 | 119 | 0 | 137 | 11 | 119 | 3 | 125 |

**TABLE 3.** 4 kinds of compound mode abnormal point detection results.

| | | CM-1 | | CM-2 | | CM-3 | | CM-4 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Outlier | Normal point | Outlier | Normal point | Outlier | Normal point | Outlier | Normal point |
| LSTM | Outlier | 24 | 5 | 44 | 5 | 40 | 5 | 40 | 7 |
| | Normal point | 9 | 130 | 4 | 115 | 8 | 115 | 8 | 113 |
| GRU | Outlier | 23 | 11 | 43 | 11 | 37 | 11 | 40 | 11 |
| | Normal point | 10 | 124 | 5 | 109 | 11 | 109 | 8 | 109 |
| SVR | Outlier | 21 | 20 | 44 | 18 | 38 | 18 | 38 | 18 |
| | Normal point | 12 | 115 | 4 | 102 | 10 | 102 | 10 | 102 |
| RF | Outlier | 18 | 29 | 42 | 27 | 40 | 27 | 36 | 27 |
| | Normal point | 15 | 106 | 6 | 93 | 8 | 93 | 12 | 93 |
| LR | Outlier | 21 | 21 | 44 | 19 | 38 | 19 | 38 | 19 |
| | Normal point | 12 | 114 | 4 | 101 | 10 | 101 | 10 | 101 |
| CNN-LSTM | Outlier | 21 | 4 | 44 | 4 | 40 | 4 | 39 | 4 |
| | Normal point | 12 | 131 | 4 | 116 | 8 | 116 | 9 | 116 |
| Attention-LSTM | Outlier | 23 | 3 | 44 | 3 | 41 | 3 | 41 | 3 |
| | Normal point | 10 | 132 | 4 | 117 | 7 | 117 | 7 | 117 |
| PSO-Attention-LSTM | Outlier | 25 | 1 | 45 | 1 | 40 | 1 | 42 | 1 |
| | Normal point | 8 | 134 | 3 | 119 | 8 | 119 | 6 | 119 |

model can provide auxiliary decision-making functions for power grid anti-stealing personnel to find abnormal users, and also provide a certain reference for the research on reducing non-technical losses and grid abnormal detection.

## VI. CONCLUSION

In order to investigate suspicious abnormal users and reduce the non-technical losses of the power grid, this paper proposes an abnormal user electricity detection method based on the PSO-Attention-LSTM model. Firstly, we establish a power theft mode, and generate a data set of abnormal users, then build a PSO-Attention-LSTM prediction model based on historical power consumption data and corresponding weather characteristics, and finally determine the detection performance of the model through the set threshold and abnormal user data set. This method has the following advantages:

1)Constructed 6 actual electric stealing modes and 4 composite modes, the detection performance of the detection model for different kinds of abnormalities can be more comprehensively evaluated through these 10 electric stealing modes.

2)Using the LSTM network can fully consider the time series characteristics of the user's power consumption, has a good time series data fitting regression ability. At the same time, the attention mechanism is used to solve the problem of long-sequence data information loss, and it can also enhance important information and suppress useless information. Use the PSO to optimize the hyperparameters of the Attention-LSTM model, obtain the optimal parameters, and further improve the performance of the model.

3)By setting up comparative experiments, comparing with LSTM, GRU, SVR, RF, LR, CNN-LSTM and Attention-LSTM, it is verified that the PSO-Attention-LSTM model has advantages in positive rate and false positive rate, and has stronger anomaly detection ability.

The detection model established in this paper has a poor detection effect on "burr" points caused by large noise and strong randomness, and does not consider the related problems of practical application. Therefore, in future work, this paper will eliminate the bad detection effect of "burr" points, further improve the detection accuracy of the model, and study the application of the detection model in practice.

## REFERENCES

[1] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. Shen, "Energy-theft detection issues for advanced metering infrastructure in smart grid," *Tsinghua Sci. Technol.*, vol. 19, no. 2, pp. 105–120, Apr. 2014, doi: 10.1109/TST.2014.6787363.

[2] National Development and Reform Commission. *Implementation Opinions on Promoting the Reform of the Electricity Sales Side*. Accessed: Nov. 26, 2015. [Online]. Available: http://sdb.nea.gov.cn/doc/201511305.pdf

[3] 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 Trans. Power Del.*, vol. 26, no. 4, pp. 2436–2442, Oct. 2011, doi: 10.1109/TPWRD.2011.2161621.

[4] Q. Chen, K. Zheng, C. Kang, and F. Huangfu, "Detection methods of abnormal electricity consumption behaviors: Review and prospect," *Automat. Electr. Power Syst.*, vol. 42, no. 17, pp. 189–199, 2018, 10.7500/AEPS20171128013.

[5] E. S. Mclaughlin, D. Podkuiko, and P. Mcdaniel, "Energy theft in the advanced metering infrastructure," in *Proc. CRITIS*, Bonn, Germany, 2010, pp. 176–187.
[6] T. T. Zhang, "Research on application of machine learning in electric consumption analysis of customers," M.S. thesis, Dept. Comput. Tech., Xi'an Shiyou Univ., Xi'an, China, 2019.
[7] G. M. Messinis, A. E. Rigas, and N. D. Hatziargyriou, "A hybrid method for non-technical loss detection in smart distribution grids," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6080–6091, Nov. 2019, doi: 10.1109/TSG.2019.2896381.
[8] Y. Sun, S. H. Li, C. Cui, B. Lin, S. S. Chen, and G. Y. Cui, "Improved outlier detection method of power consumer data based on Gaussian kernel function," *Power Syst. Technol.*, vol. 42, no. 5, pp. 1595–1606, May 2018, doi: 10.13335/j.1000-3673.pst.2017.1586.
[9] I. Monedero, F. Biscarri, C. León, J. I. Guerrero, J. Biscarri, and R. Millán, "Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees," *Int. J. Electr. Power Energy Syst.*, vol. 34, no. 1, pp. 90–98, Jan. 2012, doi: 10.1016/j.ijepes.2011.09.009.
[10] S. Amin, G. A. Schwartz, and H. Tembine, *Incentives and Security in Electricity Distribution Networks*. Berlin, Germany: Springer, 2012, pp. 264–280.
[11] P. Wang, J. Y. Lin, S. Guo, W. P. Luan, and T. Lin, "Distribution system data analytics and applications," *Power Syst. Technol.*, vol. 41, no. 10, pp. 3333–3340, Oct. 2017, doi: 10.13335/j.1000-3673.pst.2017.1624.
[12] A. L. Shah, W. Mesbah, and A. T. Al-Awami, "An algorithm for accurate detection and correction of technical and nontechnical losses using smart metering," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8809–8820, Nov. 2020, doi: 10.1109/TIM.2020.2999175.
[13] Z. Wang, G. Li, X. Wang, C. Chen, and H. Long, "Analysis of 10kV non-technical loss detection with data-driven approaches," in *Proc. ISGT Asia*, Chengdu, China, May 2019, pp. 4154–4158.
[14] C. Z. Zhang, X. Y. Xiao, and Z. X. Zheng, "Electricity theft detection for customers in power utility based on real-valued deep belief network," *Power Syst. Technol.*, vol. 43, no. 3, pp. 1083–1091, Mar. 2019, doi: 10.13335/j.1000-3673.pst.2018.1045.
[15] G. Xu, Y. P. Tan, and T. H. Dai, "Sparse random forest based abnormal behavior pattern detection of electric power user side," *Power Syst. Technol.*, vol. 41, no. 6, pp. 1964–1973, Jun. 2017, doi: 10.13335/j.1000-3673.pst.2016.2065.
[16] W. Q. Zhao, Z. J. Shen, and G. Ji, "Anomaly detection for power consumption pattern based on deep learning," *Electr. Power Autom. Eq.*, vol. 38, no. 9, pp. 34–38, Sep. 2018, doi: 10.16081/j.issn.1006-6047.2018.09.006.
[17] M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, "Detection of non-technical losses using smart meter data and supervised learning," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2661–2670, May 2019, doi: 10.1109/TSG.2018.2807925.
[18] M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, "Hybrid deep neural networks for detection of non-technical losses in electricity smart meters," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1254–1263, Mar. 2020, doi: 10.1109/TPWRS.2019.2943115.
[19] K. M. Ghori, R. A. Abbasi, M. Awais, M. Imran, A. Ullah, and L. Szathmary, "Performance analysis of different types of machine learning classifiers for non-technical loss detection," *IEEE Access*, vol. 8, pp. 16033–16048, Dec. 2020, doi: 10.1109/ACCESS.2019.2962510.
[20] W. Wei, B. Zhou, D. Połap, and M. Woźniak, "A regional adaptive variational PDE model for computed tomography image reconstruction," *Pattern Recognit.*, vol. 92, pp. 64–81, Aug. 2019, doi: 10.1016/j.patcog.2019.03.009.
[21] W. W, X. Xia, M. Wozniak, X. Fan, R. Damaševičius, and Y. Li, "Multi-sink distributed power control algorithm for cyber-physical-systems in coal mine tunnels," *Comput. Netw.*, vol. 161, pp. 210–219, Oct. 2019, doi: 10.1016/j.comnet.2019.04.017.
[22] W. Wei, H. Song, W. Li, P. Shen, and A. Vasilakos, "Gradient-driven parking navigation using a continuous information potential field based on wireless sensor network," *Inf. Sci.*, vol. 408, pp. 100–114, Oct. 2017, doi: 10.1016/j.ins.2017.04.042.
[23] W. Wei, Q. Xu, L. Wang, X. H. Hei, P. Shen, W. Shi, and L. Shan, "GI/Geom/1 queue based on communication model for mesh networks," *Int. J. Commun. Syst.*, vol. 27, no. 11, pp. 3013–3029, Apr. 2013, doi: 10.1002/dac.2522.
[24] Q. Ke, J. Zhang, H. Song, and Y. Wan, "Big data analytics enabled by feature extraction based on partial independence," *Neurocomputing*, vol. 288, pp. 3–10, May 2018, doi: 10.1016/j.neucom.2017.07.072.
[25] Q. Ke, J. Zhang, W. Wei, D. Połap, M. Woźniak, L. Kośmider, and R. Damaševičius, "A neuro-heuristic approach for recognition of lung diseases from X-ray images," *Expert Syst. Appl.*, vol. 126, pp. 218–232, Jul. 2019, doi: 10.1016/j.eswa.2019.01.060.
[26] Q. Ke, J. Zhang, W. Wei, R. Damaševičius, and M. Woźniak, "Adaptive independent subspace analysis of brain magnetic resonance imaging data," *IEEE Access*, vol. 7, pp. 12252–12261, Jan. 2019, doi: 10.1109/ACCESS.2019.2893496.
[27] Q. Ke, J. Zhang, M. Wozniak, and W. Wei, "The phase and shift-invariant feature by adaptive independent subspace analysis for cortical complex cells," *Inf. Technol. Control*, vol. 48, no. 1, pp. 58–70, Mar. 2019, doi: 10.5755/j01.itc.48.1.21706.
[28] S. Qi, Y. Q. Zheng, X. F. Chen, and W. Wei, "Ants can carry cheese: Secure and private RFID-enabled third-party distribution," *IEEE Trans. Dependable Secur. Comput.*, early access, Sep. 25, 2020, doi: 10.1109/TDSC.2020.3026191.
[29] W. Wei, B. Zhou, D. Połap, and M. Woźniak, "A regional adaptive variational PDE model for computed tomography image reconstruction," *Pattern Recognit.*, vol. 92, pp. 64–81, Aug. 2019, doi: 10.1016/j.patcog.2019.03.009.
[30] W. W, X. Xia, M. Woźniak, X. Fan, R. Damaševičius, and Y. Li, "Multi-sink distributed power control algorithm for Cyber-physical-systems in coal mine tunnels," *Comput. Netw.*, vol. 161, pp. 210–219, Oct. 2019, doi: 10.1016/j.comnet.2019.04.017.
[31] S. Y. Qi, Y. S. Lu, W. Wei, and X. F. Chen, "Efficient data access control with fine-grained data protection in cloud-assisted IIoT," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2886–2899, Feb. 2020, doi: 10.1109/JIOT.2020.3020979.
[32] Y.-L. Lo, S.-C. Huang, and C.-N. Lu, "Non-technical loss detection using smart distribution network measurement data," in *Proc. ISGT Asia*, Tianjin, China, May 2012, pp. 1–5.
[33] S. Amin, G. A. Schwartz, and A. A. Cardenas, "Game-theoretic models of electricity theft detection in smart utility networks: Providing new capabilities with advanced metering infrastructure," *IEEE Contr. Syst. Mag.*, vol. 35, no. 1, pp. 66–81, Feb. 2015, doi: 10.1109/MCS.2014.2364711.
[34] 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 Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1606–1615, Apr. 2018, doi: 10.1109/TII.2017.2785963.
[35] M. I. Ibrahem, M. Nabil, M. M. Fouda, M. Mahmoud, W. Alasmary, and F. Alsolami, "Efficient privacy-preserving electricity theft detection with dynamic billing and load monitoring for AMI networks," *IEEE Internet Tings J.*, vol. 8, no. 2, pp. 1243–1258, Jan. 2021, doi: 10.1109/JIOT.2020.3026692.
[36] L. A. Passos Júnior, C. C. O. Ramos, D. Rodrigues, D. R. Pereira, A. N. de Souza, K. A. P. da Costa, and J. P. Papa, "Unsupervised non-technical losses identification through optimum-path forest," *Electr. Power Syst. Res.*, vol. 140, pp. 413–423, Nov. 2016, doi: 10.1016/j.epsr.2016.05.036.
[37] L. Tian and M. Xiang, "Abnormal power consumption analysis based on density-based spatial clustering of applications with noise in power systems," *Autom. Electr. Power Syst.*, vol. 41, no. 5, pp. 64–70, Mar. 2017, doi: 10.7500/AEPS20160510003.
[38] Y. Liu, S. Garg, J. Nie, Y. Zhang, Z. Xiong, J. Kang, and M. S. Hossain, "Deep anomaly detection for time-series data in industrial IoT: A communication-efficient on-device federated learning approach," *IEEE Internet Things J.*, early access, Jul. 24, 2020, doi: 10.1109/JIOT.2020.3011726.
[39] X. Q. Wang, Y. L. Chen, Q. Yang, and H. C. Liu, "Analysis and prediction of user electricity consumption based on time series decomposition," *Comput. Eng. Appl.*, vol. 38, no. 9, pp. 230–236, Sep. 2018, doi: 10.3778/j.issn.1002-8331.1801-0173.
[40] D. Y. Deng, J. Li, Z. Y. Zhang, Y. F. Teng, and Q. Hhuang, "Short-term electric load forecasting based on EEMD-GRU-MLR," *Power Syst. Technol.*, vol. 44, no. 2, pp. 593–602, Feb. 2020, doi: 10.13335/j.1000-3673.pst.2019.0113.
[41] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," *J. Mach. Learn. Res.*, vol. 3, no. 1, pp. 115–143, Jan. 2003, doi: 10.1162/153244303768966139.
[42] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000, doi: 10.1162/089976600300015015.
[43] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreeit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. NIPS*, Red Hook, NY, USA, 2017, pp. 5999–6009.

[44] 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 Trans. Power Del.*, vol. 25, no. 2, pp. 1162–1171, Apr. 2010, doi: 10.1109/TPWRD.2009.2030890.
[45] M. Zanetti, E. Jamhour, M. Pellenz, M. Penna, V. Zambenedetti, and I. Chueiri, "A tunable fraud detection system for advanced metering infrastructure using short-lived patterns," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 830–840, Jan. 2019, doi: 10.1109/TSG.2017.2753738.
[46] P. Jokar, N. Arianpoo, and V. C. M. Leung, "Electricity theft detection in AMI using Customers' consumption patterns," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 216–226, Jan. 2016, doi: 10.1109/TSG.2015.2425222.
[47] Umass Smart Data Set. *Smart Data Set for Sustainability*. Accessed: Sep. 15, 2020. [Online]. Available: http://traces.cs.umass.edu/index.php/Smart/Smart

**JIAHAO BIAN** is currently pursuing the degree with the School of Electrical Engineering, Shaanxi University of Technology. His research interests include artificial intelligence and power big data.

**LEI WANG** received the B.S. and M.S. degrees in computer science and technology from the Xi'an University of Technology, Xi'an, China, in 1994 and 1997, respectively, and the Ph.D. degree in electronic science and technology from Xidian University, Xi'an, in 2001. He is currently a Professor with the Faculty of Shaanxi Key Laboratory of Industrial Automation, Shaanxi University of Technology, Hanzhong, Shaanxi, China. His current research interests include evolutionary algorithms, neural networks, and big data.

**RAFAŁ SCHERER** (Member, IEEE) received the M.S. degree in electrical engineering from the Department of Electrical Engineering and the Ph.D. degree in computer science (Methods of Classification Using Neuro-Fuzzy Systems) from the Department of Mechanical Engineering and Computer Science, Czestochowa University of Technology. He is currently an Associate Professor with the Institute of Computational Intelligence, Czestochowa University of Technology. He was a Principal Investigator of the Polish Ministry of Science and Higher Education project Computational Intelligence Methods in Data Mining and a Researcher in the Polish-Singapore Research Project (Development of Intelligent Techniques for Modeling, Controlling and Optimizing Complex Manufacturing Systems). He is a Co-coordinator of the Microsoft Dynamics Academic Alliance Program, Czestochowa University of Technology. He authored a book on multiple classification techniques published in Springer. He authored more than 80 research articles. His research interests include developing new methods in computational intelligence and data mining, ensembling methods in machine learning, and content-based image indexing. He was a Reviewer for major computational intelligence journals. He co-organizes every year or two years the International Conference on Artificial Intelligence and Soft Computing in Zakopane (http://www.icaisc.eu/) which is one of the major events on computational intelligence. He is also a Co-Editor of the *Journal of Artificial Intelligence and Soft Computing Research* (http://jaiscr.eu/).

**MARCIN WOŹNIAK** received the M.Sc. degree in applied mathematics from the Silesian University of Technology, Gliwice, Poland, in 2007, and the Ph.D. degree in computational intelligence and the D.Sc. degree in computational intelligence from the Czestochowa University of Technology, Czestochowa, Poland, in 2012 and 2019, respectively. He is currently an Associate Professor with the Faculty of Applied Mathematics, Silesian University of Technology. He is a Scientific Supervisor in editions of The Diamond Grant and The Best of the Best programs for highly talented students from the Polish Ministry of Science and Higher Education. He participated in various scientific projects (as Lead Investigator, Scientific Investigator, Manager, or Participant) at Polish and Italian universities. He was a Visiting Researcher with universities in Italy, Sweden, and Germany. He has authored/coauthored more than 100 research papers in international conferences and journals. His current research interests include neural networks with their applications together with various aspects of applied computational intelligence. He was a Session Chair at various international conferences and symposiums, including the IEEE Symposium Series on Computational Intelligence and the IEEE Congress on Evolutionary Computation. He was the Editorial Board member or an Editor of *Sensors*, IEEE Access, *Frontiers in Human Neuroscience*, *PeerJ CS*, the *International Journal of Distributed Sensor Networks*, *Computational Intelligence and Neuroscience*, the *Journal of Universal Computer Science*, and so on.

**PENGCHAO ZHANG** received the B.Eng. degree in automation from the Shaanxi University of Technology (SNUT), Hanzhong, China, and the M.Eng. degree in traffic control engineering from Northwestern Polytechnical University (NPU), Xi'an, China, where he is currently pursuing the Ph.D. degree. He is also an Associate Professor with SNUT. His current research interests include industrial robot and mobile robotics.

**WEI WEI** (Senior Member, IEEE) received the M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 2005 and 2011, respectively. He is currently an Associate Professor with the School of Computer Science and Engineering, Xi'an University of Technology, Xi'an. He ran many funded research projects as a principal investigator and technical members. He has published around 100 research papers in international conferences and journals. His current research interests include wireless networks, wireless sensor networks application, image processing, mobile computing, distributed computing, pervasive computing, the Internet of Things, and sensor data clouds. He is a Senior Member of the China Computer Federation. He is a TPC member of many conferences. He is an Editorial Board Member of the *Future Generation Computer System*, IEEE Access, *Ad Hoc & Sensor Wireless Sensor Network*, the *Institute of Electronics, Information and Communication Engineers*, and *KSII Transactions on Internet and Information Systems*. He is a Regular Reviewer of the IEEE Transactions on Parallel and Distributed Systems, the IEEE Transactions on Image Processing, the IEEE Transactions on Mobile Computing, the IEEE Transactions on Wireless Communications, the *Journal of Network and Computer Applications*, and many other Elsevier journals.