

Research Article

Nonintrusive Industrial Load Identification Combined with Random Forests and Steady-State Waveform

Shuhui Yi ¹, Hongxia Zhu,² Junjie Liu,¹ and Junnan Li³

¹China Electric Power Research Institute, Wuhan 430070, China

²State Grid Shandong Electric Power Company Marketing Service Center, Jinan 250001, China

³State Grid Henan Electric Power Company Marketing Service Center, Zhengzhou 450015, China

Correspondence should be addressed to Shuhui Yi; yi430070@163.com

Received 6 August 2021; Revised 3 November 2021; Accepted 8 November 2021; Published 29 November 2021

Academic Editor: Yang Li

Copyright © 2021 Shuhui Yi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nonintrusive industrial load identification can accurately acquire the operation data of each load in the plant, which is the benefit of intelligent power management. The identification method of the industrial load is complicated and difficult to be realized due to the difficulty in collecting transient data for modeling, and high-precision measuring equipment is required. Aiming at this situation, the article proposes a nonintrusive industrial load identification method using a random forest algorithm and steady-state waveform. Firstly, by monitoring the change of the industrial load power state, when the load changes and becomes stable, the steady-state waveform is extracted. Due to different electrical characteristics of industrial loads, the current waveform of loads is different to some extent. We can construct characteristic data for each industrial load to construct its own current steady-state waveform. Then, using the high-dimensional data of the steady-state waveform as the sample data, the bootstrap sampling method and the CART algorithm in the random forest algorithm are used to generate multiple decision trees. Finally, the industrial load types are identified by voting multiple decision trees. The actual operating load data of a factory are used as the sample data in the simulation, and the effectiveness and rapidity of the proposed identification algorithm are verified by the combined load method simulation comparison. The simulation results show that the accuracy of the proposed identification algorithm is more than 99%, the identification time is 3.36 s, which is much higher than that of other methods, and the operation time is less than that of other methods. Therefore, the proposed identification algorithm can effectively realize the nonintrusive industrial load identification.

1. Introduction

With the continuous development of smart grids, guided by big data and the concept of a conservation-oriented society, the collection of various load power data and power optimization management in the power industry have become increasingly important. The research of nonintrusive load identification methods is very necessary [1]. Traditional load monitoring is generally installed on the equipment side, and the equipment operation status can be collected without load identification, but the cost is high, and the installation is inconvenient [2]. At the same time, industrial loads have continuous production and high safety requirements [3]. Therefore, the advantages of nonintrusive industrial load identification with low cost, convenient installation, and easy promotion play an important role. It can not only

provide an important basis for the fine management of the power grid but also provide a reference for users to use electricity reasonably [4, 5]. Nonintrusive industrial load identification can improve the overall reliability and profitability of factories and related industrial systems [6].

At present, the focus and difficulty of nonintrusive load monitoring are load identification, among which there are many related research studies on household load identification [7, 8]. Nonintrusive load identification methods focus on the clustering algorithm [9, 10], V-I trajectory [11], artificial intelligence algorithm [12, 13], etc. In terms of the selection of electrical characteristics, the study is divided into transient [14–16] and steady state [17]. Xu et al. [18] established an applied deep confidence network method, which used multiple restricted Boltzmann machines and a layer backward propagation neural network to carry out load

identification. Wu et al. [19] proposed monitoring at the metering end of the total low-voltage electricity consumption and studied a waveform analytic identification method at the edge of the event. Gillis et al. [20] presented a new concept based on wavelet design and machine learning applied to nonintrusive load monitoring that can improve the prediction accuracy. Fang et al. [21] proposed a novel NILM method that leverages advances in statistical learning achieving superior performance. Zhou et al. [22] proposed a nonintrusive load decomposition method based on a hybrid deep learning model that improves the performance of the whole network system.

However, most of the current algorithms have complex extraction process, slow recognition speed, and limited application scope. And there are relatively few studies on industrial load identification. Since it is not convenient to collect transient characteristics for modeling of industrial load due to production demand, how to extract data for modeling identification without affecting the operation of industrial load becomes very critical [23, 24]. In this paper, the operation of industrial load is studied. The proposed method requires few electrical characteristic data, has a simple training dataset, and is more convenient for sampling, which improves the accuracy of identification and operation efficiency. The contributions are as follows:

- (1) Integrated with the steady-state current waveform of the industrial load, here, the presented feature extraction method described the characteristics of the industrial load
- (2) A nonintrusive industrial load identification method is proposed based on the random forest algorithm and voting method
- (3) Based on the actual factory's operating load data from Henghe DL850 acquisition equipment, plenty of simulations are carried out to verify the effectiveness and rapidity of the proposed method
- (4) The flexibility, practicality, and scalability of the proposed identification algorithm are further analyzed and discussed according to the simulations and comparisons

2. Principle of Nonintrusive Industrial Load Identification

The principle of nonintrusive load identification is shown in Figure 1, including five links of data collection, data processing, state change, feature selection, and load identification [25].

2.1. Data Collection and Processing. For the collection of industrial load characteristics, due to the high voltage of the industrial load, electrical data are mainly obtained by using smart meters or professional collection equipment through the low-voltage side. The collected data include active power, reactive power, voltage, and current. The collected data need to be preprocessed to meet the requirements of the algorithm program parameter input.

2.2. State Change. Before judging the load type, it is necessary to judge in advance whether there is a change in the state of the load starting and stopping or the same load multistate change. During the load status change, the load power changes more clearly. Generally, by observing the changed power value within a specified time and comparing it with the threshold value, it is further judged whether there is a load switching event.

2.3. Feature Selection. This paper uses the load power as the load feature extraction for comparison when the state changes and uses the steady-state current data as the load feature for calculation during the load identification process.

2.4. Load Identification. After judging the load status change, this article collects steady-state current waveform data, uses the random forest algorithm to identify the load, and judges the load type and load state.

3. Industrial Load Feature Extraction

In this paper, the industrial load feature extraction is based on the monitoring of the industrial load power state changes through events, and then the steady-state current waveform is extracted.

3.1. Event-Based Monitoring of Industrial Load Power. During the normal operation of industrial loads, it is generally not possible to repeat the start and stop operations. It is more convenient to collect data when the load is working stably. The voltage and current waveforms of the load are generally sine waves, so when the load status changes, it is mainly reflected in the changes in power and current. Usually, the change point detection algorithm is used for the active data to capture the moment when the power load status changes, and then the relevant event characteristics are extracted.

Assume that, during the period of t_s (t_s is an integer multiple of the sampling period T), the state of the electrical load transitions from one stable state to another stable state. When the load state changes, the power of the load also changes. So, the change of load state can be identified by the step characteristic of active power. We set P_s as the mutation amount of state change and H as the threshold value, and the calculation formula is shown in formula (1). When P_s exceeds the threshold H , it can be considered that a load event occurs at time t .

$$P_s = \left| P_{(t+t_s)} - P_{(t)} \right|, \quad (1)$$

$$P_s > H. \quad (2)$$

Due to the nature of the industrial load, low-frequency sampling is adopted for identification. First, the active power data are read into the database through professional acquisition equipment, and then the active power changes are monitored in real time. Finally, when the active power changes meet formula (2), the load status changes. We can

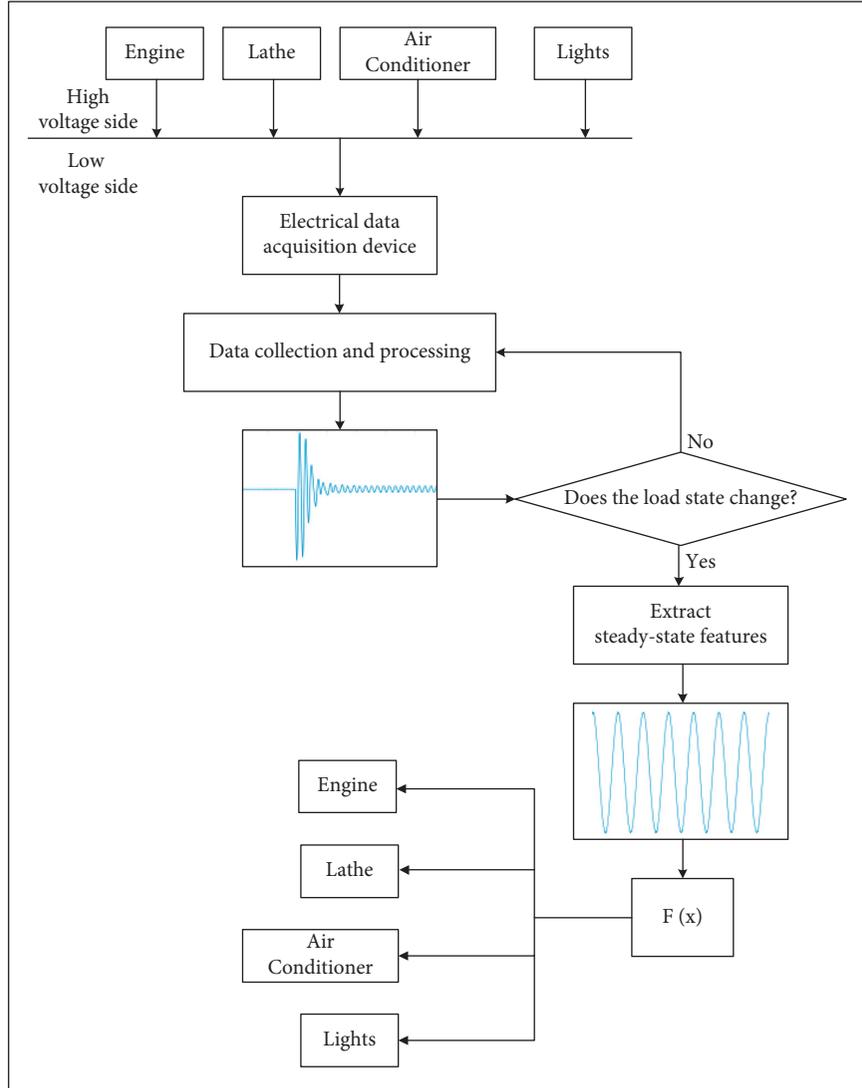


FIGURE 1: Nonintrusive industrial load identification.

also use the steady-state electrical component difference method to extract the electrical features. As shown in Figure 2, the current and active power change after the load status changes.

3.2. Extract Industrial Load Characteristics Based on Steady-State Waveforms. When event A is detected, we need to extract the load waveform. Usually, the steady-state waveform can characterize the basic state of the load during normal operation. In this paper, the steady-state waveform of the event is extracted as the event waveform for analysis. Therefore, it is necessary to judge when the event is in a steady state before and after the occurrence. When there is no significant change in active power for multiple cycles, the current signal is considered in a stable state.

The fundamental phase angle of the steady-state current is determined by the initial phase of the voltage during measurement. It is guaranteed that the steady-state current can be measured under the same initial phase angle voltage to meet the current superposition. Therefore, when the

current is judged to be in a steady state, the waveform can be extracted by detecting the zero-crossing point of the corresponding voltage. The voltage corresponding to the steady-state current I_{m-1} before the event is U_{m-1} . U_{m-1} sampling point j corresponds to $U_{m-1,j}$. When $U_{m-1,j}$ satisfies formula (3), this point is considered to be voltage crossing zero of the steady-state voltage waveform U_{m-1} before the event occurs.

$$\begin{cases} U_{m-1,j-1} < U_{m-1,j} < U_{m-1,j+1}, \\ |U_{m-1,j}| < |U_{m-1,j-1}|, \\ |U_{m-1,j-1}| < |U_{m-1,j+1}|, \end{cases} \quad (3)$$

where j is the serial number of the sampling point corresponding to the voltage zero-crossing point showing an upward trend. Starting from the moment corresponding to the zero-crossing point, one cycle can be intercepted to extract the steady-state current I_{m-1} before the occurrence of event A .

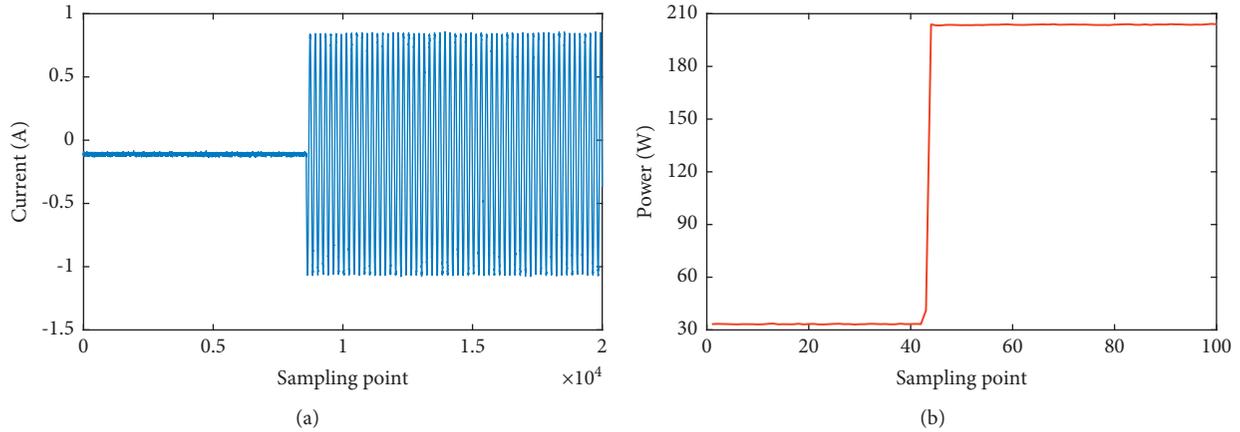


FIGURE 2: Load step change. (a) Current. (b) Power.

In the same way, the above method is used to detect the voltage zero-crossing point of the steady-state voltage waveform U_s after the occurrence of event A , and then the steady-state periodic current I_s is extracted. According to the principle of current superposition, we use formula (4) to obtain the event current waveform I_v of A and the corresponding voltage waveform U_v of A .

$$\begin{cases} I_v = I_m - I_{m-1}, \\ U_v = U_m. \end{cases} \quad (4)$$

The detailed description of steady-state current waveform characteristics can provide an important basis for load identification. Therefore, the feature extraction of the steady-state current of the load in this paper includes the peak, peak valley, root mean square, and standard deviation of the steady-state current. At the same time, the steady-state current waveform is used as an industrial load imprint to identify various loads. Figure 3 shows the steady-state current waveform characteristics of industrial loads.

4. Industrial Load Identification Based on the Random Forest Algorithm

Random forest is characterized by strong antinoise and randomness [26]. It is an ensemble learning model with good predictive ability. Random forest is a kind of combined classifier, which uses a decision tree as the base classifier, which reduces the limitation of the single classifier to a certain extent and can achieve higher prediction accuracy. At the same time, the randomness of the random forest itself can tolerate the effects of abnormal points and noise to a greater extent, thereby reducing the overfitting phenomenon in the decision tree algorithm to a certain extent and improving the generalization ability. In this paper, the specific steps of the process of industrial load identification based on the random forest are as follows, as shown in Figure 4.

4.1. Industrial Load Sample Dataset. The collected steady-state current waveform data are preprocessed by a linear regression algorithm to remove abnormal values and

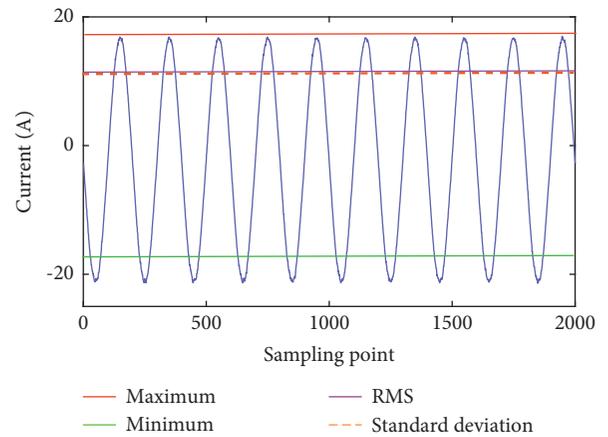


FIGURE 3: Current characteristics of the load.

irrelevant data. Then, the current waveforms are grouped into k pieces to form a sample dataset.

4.2. Bootstrap Algorithm. Bootstrap is a sampling method with replacement. First, a certain amount of sample data needs to be collected, then the sample after each sampling needs to be put back into the sample dataset, and the next sampling can be selected again. Finally, after sampling the sample data N times, N training sets are obtained, and each training set can form a decision tree, as shown in Figure 5.

4.3. Based on the CART Decision Tree. Random forest is a combination classifier that uses a decision tree as the base classifier. The decision tree is the process of continuously splitting the root node to generate leaf nodes. The leaf nodes represent the classification results, and there is no need to continue splitting [27]. Each training set generates a decision tree according to the CART algorithm. In the process of producing the decision tree, m ($m \leq M$) feature variables are randomly selected from the total M attributes in the data sample set, and the CART algorithm is used to calculate the m features which obtain an optimal attribute, proceed to the next step of splitting the attributes, and construct N decision trees accordingly, as shown in Figure 6.

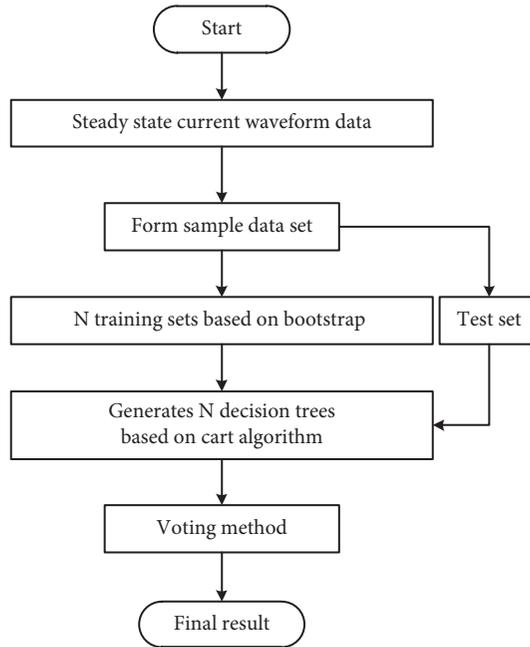


FIGURE 4: Flowchart of industrial load identification based on the random forest.

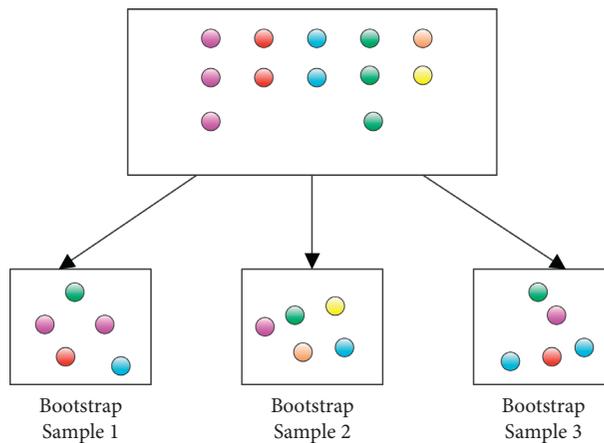


FIGURE 5: Resampling method based on bootstrap.

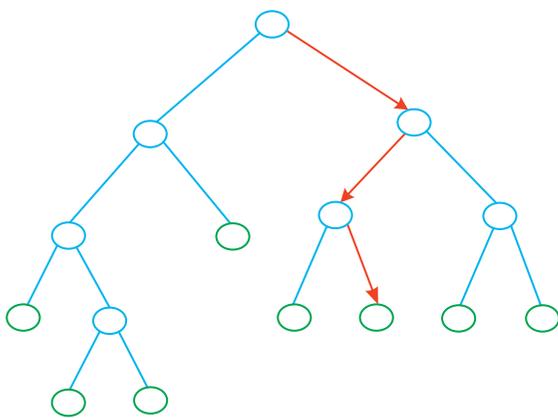


FIGURE 6: Decision tree.

This paper adopts the CART algorithm and selects attributes based on the principle of minimum Gini coefficient to split the binary tree of nodes. The exponential formula of the distribution is as follows [27]:

$$\text{Gini}(p) = 1 - \sum_{k=1}^k p_k^2, \quad (5)$$

where k is the total number of types of feature samples in the node and p_k is the probability of the K type of feature samples in the node in formula (5).

The Gini index in sample set D can be calculated using the following formula:

$$\text{Gini}(D) = 1 - \sum_{k=1}^k \left(\frac{D_K}{D}\right)^2, \quad (6)$$

where D_k is the subset of samples belonging to the K category in the sample set D .

The Gini index involves the binary division of different attributes. If a certain attribute A is a discrete value, A needs to be divided into two parts D_1 and D_2 . The Gini index divided by each node can be calculated by the following formula:

$$\text{Gini}_A(D) = \frac{D_1}{D} \text{Gini}(D_1) + \frac{D_2}{D} \text{Gini}(D_2). \quad (7)$$

D_1 and D_2 are two subsets of the sample set D divided into. Therefore, the impurity reduction of A is as shown in the following formula:

$$\Delta \text{Gini}_A(D) = \text{Gini}(D) - \text{Gini}_A(D). \quad (8)$$

It can be seen that, in each binary division, when the Gini index is the smallest, this attribute will be selected. Therefore, it is necessary to reduce its impurity, Gini_A larger and $\text{Gini}_A(D)$ smaller, so that the splitting effect of attribute A will be the best.

4.4. Identification Results Based on the Voting Method. The dataset to be tested is input into the N decision tree models that have been trained, and the decision tree model calculates each type based on the parameters trained in the sample dataset. The results obtained by N decision trees are voted to obtain the final result. The highest proportion of results in the decision tree is the result of industrial load identification.

5. Case Analysis

In this experiment, the Henghe DL850 collection device was used to collect industrial load operating data on the low-voltage side of the bus side for load analysis and identification research. The algorithm is run on a computer with Intel Core i5-3470, 4 G memory, CPU clocked at 3.2 GHz, and a 64 bit operating system. The operating environment used is Matlab2019a. Compare by setting different decision trees, record the running time of the program, and take the average result of 20 tests for each type of decision tree.

5.1. Data Collection. After the device starts and stabilizes, it starts to collect voltage and current data by the method of steady-state waveform extraction, the sampling frequency is set to 10 kHz, and the acquisition time is set to 20 min. The fundamental frequency of the domestic load is 50 Hz (the period is 0.02 s), 200 points are sampled in one cycle, and 60,000 cycles are sampled in 20 min. The sampled load types include motor A, motor B, heater, electric lamp, the specific collection load combination type, and label which are shown in Table 1, and the collected waveform is shown in Figure 7.

TABLE 1: Load type.

Label	Load type
M1	Electric machinery 1
M2	Electric machinery 2
M3	Air heater
M4	Light
M5	Electric machinery 1 + air heater
M6	Electric machinery 1 + light
M7	Electric machinery 2 + air heater
M8	Electric machinery 2 + light

5.2. Formation of Sample Data. To improve the quality of the sample data, the collected data are imported into MATLAB data using XView software, and the sample data are obtained after preprocessing. The main steps are as follows:

- (1) XView software decomposes the sampled data to obtain data containing time, current, voltage, and other information, which is the decomposition process of the collected signal.
- (2) Process in MATLAB software, intercept 10,000 stable cycle data, and convert them to 10 cycles per line, 1000 lines of data.
- (3) Extract the single-phase current signal for load identification and analysis. Therefore, extract the sample data of each load with a row of 10 cycles, a total of 2000 points, and a matrix of 1000×2000 in size.
- (4) Choose 700 sets of signals as test data and the remaining 300 sets of signals as training data.
- (5) The combined training data are 5600 groups, the test data are 2400 groups, and the random forest algorithm is used to train and test the sample data.

5.3. Experimental Results and Analysis. Using the random forest model algorithm to increase the number of decision trees, the experimental results are shown in Table 2 and Figure 8. When the number of decision trees is more than 15 trees, the load identification accuracy rate has reached 100%, and the average running time is only 3.36 s.

The random forest algorithm with 15 decision trees is compared with the traditional Bayes algorithm and k -adjacent algorithm, using the same sample set, and the running results are shown in Table 3. The recognition rate and running time calculated by the random forest model have exceeded the Bayes algorithm, which is suitable for the needs of industrial load identification.

From Tables 2 and 3, it can be seen that the industrial load identification based on the random forest has a good effect. The average identification rate of the sampled 8 kinds of power loads reaches 100%. Compared with other algorithms, the industrial load identification rate based on the random forest has certain superiority.

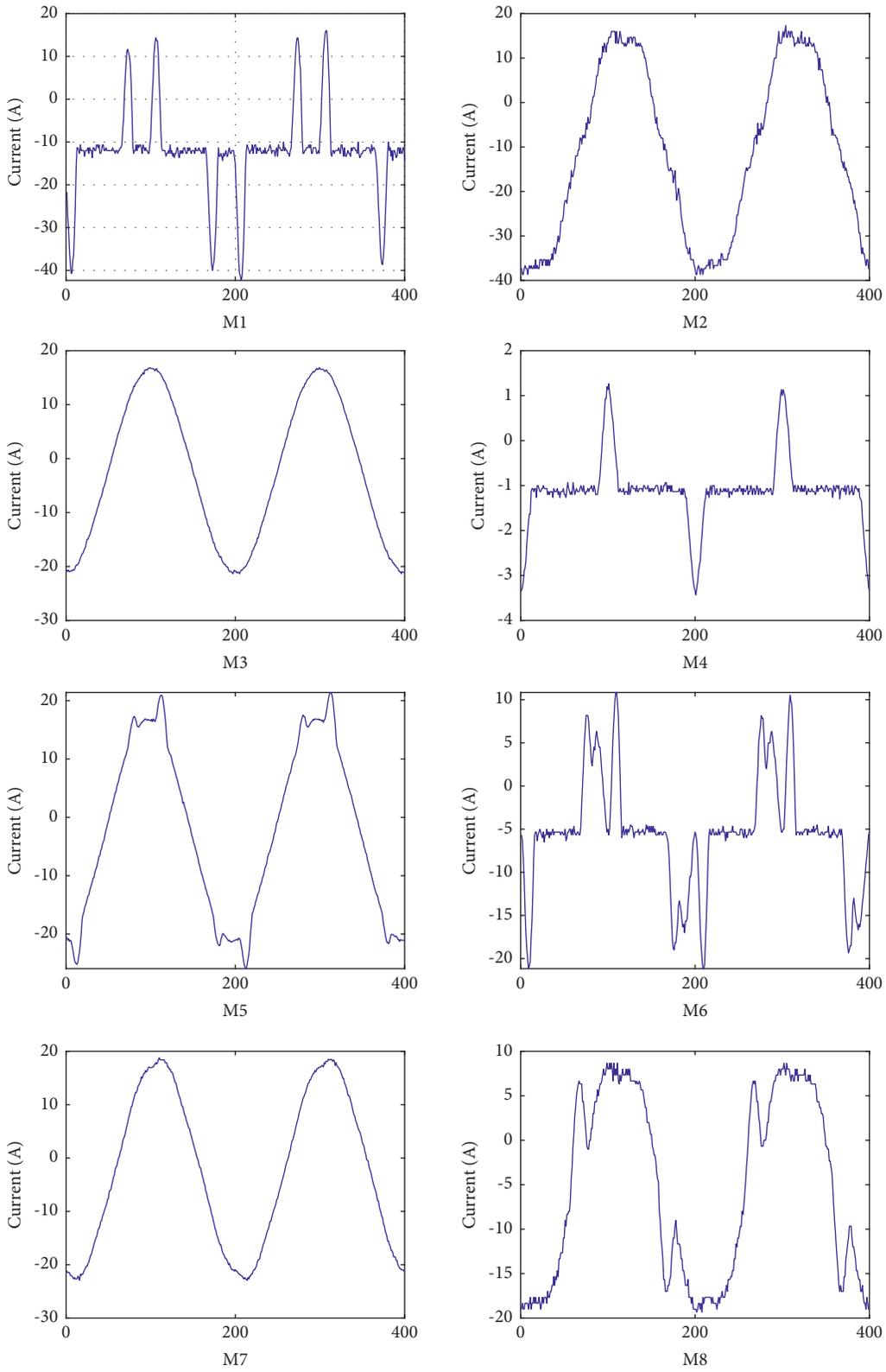


FIGURE 7: Steady-state current waveforms of different types of equipment.

TABLE 2: Load types.

The number of decision trees	Discrimination (%)	Running time (s)
5	99.89	2.04
10	99.99	2.62
15	100	3.36
20	100	3.91
100	100	14.22

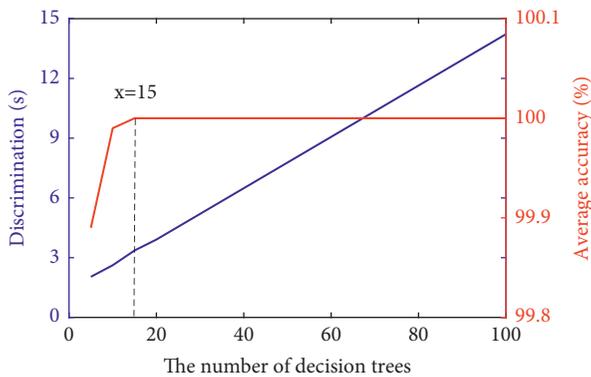


FIGURE 8: Running time of different decision trees.

TABLE 3: Comparison of identification results of different models.

	Random forest	Bayes	k -adjacent
Recognition rate (%)	100	63.8	100
Running time (s)	3.36	6.15	34.43

6. Conclusion

The industrial load identification algorithm proposed in this paper is based on the random forest and steady-state waveform. The random forest algorithm has a good prediction and classification effect, but it has not been applied to this field. The random forest algorithm improves the accuracy and efficiency of industrial load identification on the whole. The algorithm in this paper mainly uses current waveform as the recognition condition to classify and identify the electricity information of the industrial load in a nonintervening way. The electrical characteristic information is obvious, and the extraction process is simple. The random forest algorithm has good randomness and classification effect and high classification efficiency. The simulation results show that the identification of the industrial load based on the random forest model can overcome the difficulty of industrial load information collection, and the recognition rate reaches 100%, which is better than other classification algorithms in terms of accuracy and running time. Noninvasive industrial load identification based on the random forest and steady-state waveform can be further applied to other large factories and homes to improve identification accuracy and operational efficiency. However, the identification method of the continuous state load model is not considered, which needs further study.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request (yi430070@163.com).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] H. Guo, J. Lu, P. Yang, and Z. Liu, "Review on key techniques of non-intrusive load monitoring," *Electric Power Automation Equipment*, vol. 41, no. 1, pp. 135–146, 2021.
- [2] Y. Yu and Y. Liu, "Automation of electric power systems," *Automation of Electric Power Systems*, vol. 39, no. 2, pp. 1–5, 2015.
- [3] P. Dai, J. Zou, J. Tian, T. Liu, and H. Zhou, "Integrated optimization of CO₂ emission mitigation in China power sector," *Automation of Electric Power Systems*, vol. 37, no. 14, pp. 1–6, 2013.
- [4] C. W. Gellings, "Evolving practice of demand-side management," *Journal of Modern Power Systems and Clean Energy*, vol. 5, no. 1, pp. 1–9, 2017.
- [5] H. Yan, B. Li, S. Chen et al., "Future evolution of automated demand response system in smart grid for low-carbon economy," *Journal of Modern Power Systems and Clean Energy*, vol. 3, no. 1, pp. 72–81, 2015.
- [6] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.
- [7] C. Liu, A. Akintayo, Z. Jiang, G. P. Henze, and S. Sarkar, "Multivariate exploration of non-intrusive load monitoring via spatiotemporal pattern network," *Applied Energy*, vol. 211, no. 12, pp. 1106–1122, 2018.
- [8] T. Hassan, F. Javed, and N. Arshad, "An empirical investigation of V-I trajectory based load signatures for non-intrusive load monitoring," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 870–878, 2014.
- [9] W. Liu, X. Tan, K. Zhou, and J. Ma, "Non-intrusive load monitoring method based on SAGA-FCM algorithm," *Modern Electronics Technique*, vol. 42, no. 23, pp. 72–76, 2019.
- [10] D. Zhou, H. Zhang, H. Zhou, and W. Hu, "Non-intrusive load event detection method based on state feature clustering," *Transactions of China Electrotechnical Society*, vol. 35, no. 21, pp. 4565–4575, 2020.
- [11] Y. Yang, Y. Huang, and W. Chen, "Non-intrusive load monitoring algorithm based on V-I trajectory feature," *Computer and Digital Engineering*, vol. 48, no. 10, pp. 2353–2358, 2020.
- [12] R. Li, P. Zhang, Y. Liu, H. Zhang, and D. Zhou, "Non-intrusive household load identification method based on random forest," *Electrical Measurement and Instrumentation*, vol. 58, no. 4, pp. 9–16, 2021.
- [13] X. Liu, M. Cao, and Y. Xu, "Non-intrusive load monitoring based on improved chicken swarm optimization algorithm," *Electric Power Automation Equipment*, vol. 38, no. 5, pp. 235–240, 2018.
- [14] L. Niu and H. Jia, "A transient event detection algorithm for nonintrusive load monitoring," *Automation of Electric Power Systems*, vol. 35, no. 9, pp. 30–35, 2011.
- [15] R. Cox, S. B. Leeb, S. R. Shaw, and L. K. Norford, "Transient event detection for nonintrusive load monitoring and demand side management using voltage distortion," in *Proceedings of*

- the Twenty-First Annual IEEE Applied Power Electronics Conference and Exposition*, Dallas, TX, USA, March 2006.
- [16] W. Zhao, S. Zhang, and G. Li, "Non-intrusive load decomposition of residential users based on cluster and association analysis," *Electric Power Automation Equipment*, vol. 40, no. 6, pp. 8–19, 2020.
- [17] B. Qi, C. Dong, X. Wu, and G. Cui, "Non-intrusive load identification method based on DTW algorithm and steady-state current waveform," *Automation of Electric Power Systems*, vol. 42, no. 3, pp. 70–76, 2018.
- [18] C. Xu, K. Chen, J. Ma, J. Liu, and J. Wu, "Recognition of power loads based on deep belief network," *Transactions of China Electrotechnical Society*, vol. 34, no. 19, pp. 4135–4142, 2019.
- [19] X. Wu, J. Yu, L. Peng, and Q. Wang, "Non-intrusive perception and identification of industrial power load based on event analysis user edge," *Automation of Electric Power Systems*, vol. 45, no. 4, pp. 29–37, 2021.
- [20] J. M. Gillis, S. M. Alshareef, and W. G. Morsi, "Nonintrusive load monitoring using wavelet design and machine learning," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 320–328, 2016.
- [21] Z. Fang, D. Zhao, C. Chen, Y. Li, and Y. Tian, "Non-intrusive appliance identification with appliance-specific networks," *IEEE Transactions on Industry Applications*, vol. 56, no. 4, p. 1, 2020.
- [22] X. Zhou, J. Feng, and Y. Li, "Non-intrusive load decomposition based on CNN-LSTM hybrid deep learning model," *Energy Report*, vol. 7, pp. 5762–5771, 2021.
- [23] E. Holmegaard and M. K. Baun, "NILM in an industrial setting: a load characterization and algorithm evaluation," in *Proceedings of the 2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, IEEE, St Louis, MO, USA, May 2016.
- [24] S. Yi, X. Yin, and Y. Diao, "A new event-detection method based on composite windows in nilm for industrial settings," in *Proceedings of the 2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, IEEE, Beijing, China, November 2019.
- [25] Z. Cao, B. Xie, and H. Ji, "A hybrid algorithm of random forest," *Statistics & Decisions*, vol. 4, pp. 7–9, 2014.
- [26] Q. Huang, "Research on Short Term Load Forecasting of Power System Based on Random Forest," North China Electric Power University, Beijing, China, 2018.
- [27] L. Qin, *Research on Non-intrusive Household Load Identification Based on Wavelet Transform and Random Forest*, Chongqing University, Chongqing, China, 2018.