Load Identification Method Based on ISMA-GRNN

Pengfei Yang, Xianbo Sun, Li Zhu, Yuhan Wu, and Baofu Dai

School of Information Engineering, Hubei Minzu University, Enshi, China

Correspondence should be addressed to Xianbo Sun; sunxianbo@whut.edu.cn

Received 30 November 2021; Accepted 21 February 2022; Published 30 March 2022

Copyright © 2022 Pengfei Yang 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.

The noninvasive load monitoring method carries out load identification after event detection and feature extraction of load data. At present, nonload intrusive load monitoring faces the problems of low load identification accuracy and long load identification time. In order to solve these problems, a load identification method based on the improved slime mould algorithm-generalized regression neural network (ISMA-GRNN) is proposed. Firstly, by adding mutation operation in slime mould algorithm (SMA) position update, the global optimization ability of SMA is improved. Then, the improved slime mould algorithm (ISMA) is used to optimize the smoothing factor of GRNN and find the best smoothing factor. Finally, the best smoothing factor is input into GRNN for load identification, and the load identification results are output. To measure the effect of load identification, load identification precision, load identification accuracy, and load identification time are used as evaluation indicators. The simulation results show that compared with HHO-GRNN and WOA-GRNN, the load identification time of SMA-GRNN is greatly shortened, but the results are not satisfactory. On the basis of SMA-GRNN, ISMA-GRNN has significantly improved the accuracy and precision of load identification. In conclusion, ISMA-GRNN can better adapt to the load identification of multiple electrical equipment scenes.

1. Literature Review

With the in-depth research on smart grid and ubiquitous power Internet of things, NILM, as one of the key technologies for building ubiquitous power Internet of things, has become a research hotspot [1–3]. NILM technology can not only realize energy consumption management through the identification of electric equipment on the user side [4], but also predict power demand through load data to realize good interaction between power generation and power consumption [5]. The four steps of data measurement, event detection, feature extraction, and load identification form a NILM. As the core step of noninvasive load monitoring, the load identification method directly affects the effect of load identification [6].

The essence of load identification is load classification according to the load characteristics of electrical equipment [7]. Many algorithms such as SVM, Bayesian algorithm, and GRNN are used to deal with classification problems. SVM classifies data by constructing classification hyperplane, which is mainly used to deal with the classification problems of small- and medium-sized data samples, nonlinear and high-dimensional. Zhao et al. [8] used SVM to build a bearing fault diagnosis model and verified the recognition ability of SVM method to a variety of rolling bearing fault types. Bayesian algorithm calculates the probability of different events and selects the one with the greatest probability as the classification result. Jake et al. [9] used Bayesian algorithm to build bipolar short circuit fault location model. This method has the advantages of high precision, few measuring points, and high location accuracy. GRNN is an artificial neural network that uses a radial basis function as an activation function. Yao et al. [10] used clustering algorithm to cluster weather characteristic quantities and GRNN to build photovoltaic power generation prediction model. This method can take into account small sample usage scenarios and has high generalization to realize high-precision prediction of photovoltaic power generation. In the field of load identification, the above-mentioned SVM and Bayesian algorithm have been applied, but there is almost no literature based on the GRNN method.

In the field of load identification, many experts and scholars have also done a lot of research. Du et al. [11] selected the fundamental power factor and the third harmonic content difference of voltage current during equipment operation as the
load characteristics and combined the improved Gray Wolf Algorithm with FCM clustering algorithm. This method has a high recognition rate for small power loads, but it is prone to feature overlap and local optimal solution when dealing with multifluid classification problems. Qi and et al. [12] transformed V-I image features into track coding through DPHSH model and used them as load features and then built a load recognition model using CNN. This method can solve the problem of reducing the load identification rate when unknown equipment is added, but the overall accuracy of this method is not high and the load identification time is too long. Zhou et al. [13] used Fourier transform to decompose current data, selected current odd harmonics as load characteristics, and then used AdaBoost algorithm combined with BP neural network to build load identification model. Although this method can effectively improve the identification ability of easily misjudged load, the model training time becomes longer.

To sum up, in order to solve the problems of low load identification accuracy and long load identification time in the above methods, a noninvasive load monitoring method based on improved slime mould algorithm-generalized regression neural network (ISMA-GRNN) is proposed. Firstly, the method carries out data processing operations such as denoising, eliminating outliers, event detection, and feature extraction. Then, aiming at the problem that SMA algorithm is easy to fall into local optimization, the smoothing factor of GRNN is optimized after adding mutation disturbance on the basis of SMA, the optimized optimal smoothing factor is input into GRNN for load identification, and the load identification results are output. Finally, the load identification accuracy, load identification precision, and load identification time are compared with HHO-GRNN, WOA-GRNN, and SMA-GRNN, in order to reflect the advantages of ISMA-GRNN in load identification.

The rest of this study is organized as follows: Section 2 introduces the methods of event detection and feature extraction. Section 3 introduces the theory of SMA and GRNN and introduces the disturbance formula to improve the SMA method. Section 4 establishes the ISMA-GRNN model and describes its process in detail. Multiple indicators are used to measure ISMA-GRNN in Section 5 followed by the conclusion in Section 6.

2. Event Detection and Feature Extraction

Event detection scans the load data for a period of time by using the algorithm to detect the connection and disconnection system of electrical equipment, which is the premise of noninvasive load monitoring [14]. Niu and Jia [15] proposed a bilateral cumulative sum load event detection method based on sliding window, which can effectively detect high-power load events according to power load data, and the detection of low-power load events is inaccurate. On this basis, the improved mean difference accumulation method can improve the event detection sensitivity of the algorithm and effectively detect low-power load events.

In this paper, the sliding window mean difference cumulative sum algorithm is used for load event detection. By setting the load event threshold, when the electrical equipment is connected to the system, the power data fluctuates greatly. After the mean difference accumulation calculation, the threshold will be triggered, and the event detection algorithm can detect the access or exit of the electrical equipment from the system. Define positive load event threshold \( h^+ \), negative load event threshold \( h^- \), the cumulative function \( g \), and the step size of sliding window \( sp \). Characteristic power sequence during load operation is \( P_1 = [ p_1, p_2, \ldots, p_n ] \). The matrix after sliding window mean difference processing is \( X \), and its expression is shown in the following equation:

\[
X = \begin{bmatrix} x_1 & x_2 & \cdots & x_l \end{bmatrix}^T_{l(n-sp)},
\]

where \( l = n-sp \). When the electric equipment is connected, the load data has a positive offset; that is, \( x_k \geq x_{k-1} \), and the power sequence increases. Then, the positive offset will be accumulated into \( g_k^+ \) and \( g_k^- \) will continue to increase. On the contrary, if the load data has a negative offset when the electric equipment exits the system, then, \( x_k < x_{k-1} \), \( g_k^- \) will continue to increase. When \( g_k^- \geq h^- \) or \( g_k^+ < h^- \), the system will detect the occurrence of load events. When \( g_k^+ \geq h^+ \), positive load events are recorded. Similarly, when \( g_k^- < h^- \), negative load events are recorded.

After data processing operations such as denoising and eliminating outliers, the collected load data is input into the sliding window mean difference cumulative sum algorithm for event detection. Then, the load event data will be input into the feature extraction method for feature extraction.

The load characteristics of electrical equipment can generally be divided into transient characteristics and steady-state characteristics. The steady-state characteristics are widely used because of their advantages such as strong repeatability, low cost, and simple acquisition [16]. Hilbert Huang transform (HHT) can transform one-dimensional signal into a two-dimensional complex plane signal, which can better display signal features, and is often used for feature extraction [17]. The improved CEEMDAN algorithm is used for feature extraction [18, 19]. CEEMDAN method uses the estimation of local mean to replace the estimation of mode and uses the local mean of signal to extract k-order mode, which can effectively solve the problem of mode aliasing. Therefore, we use CEEMDAN + HT transform to decompose the current data of load events and extract instantaneous amplitude, instantaneous frequency, and instantaneous phase as the load characteristics of electrical equipment. The digital model is derived as follows.

Assuming that the current characteristic of the load event is \( I(t) \), the first intrinsic mode component is defined as \( I_1(t) \), \( \epsilon \) is the adaptive coefficient, \( a \) is the Gaussian white noise added during decomposition, and \( N \) is the number of EMD decompositions of the signal. The expressions of \( I_1 \) and \( R_k(t) \) are as follows:

\[
I_1 = \frac{1}{N} \sum_{i=1}^{N} I_k^i(t),
\]

\[
R_k(t) = I(t) - IMF_k.
\]

By repeating the above operation, we can get \( I_1, I_2, I_3, \ldots, I_{K+1} \), where the expressions of \( R_k(t) \) and \( I_{K+1} \) are as follows:
\[ R_k(t) = R_{k-1}(t) - I_k, \]
\[ I_{k+1} = \frac{1}{N} \sum_{i=1}^{N} E_i \left( R_k(t) + \varepsilon_k E_K(\omega_i(t)) \right). \]

If the number of extreme points of residual sequence \( R_k(t) \) is less than 2, CEEMDAN decomposition is stopped. The final decomposition of load event current characteristic \( I(t) \) is shown in the following equation:

\[ I(t) = \sum_{i=1}^{K} I_i + R_1(t). \]

Calculate the correlation parameters of \( I_1, I_2, I_3, \ldots, I_{K+1} \) according to the Pearson correlation, sort the correlation according to the value, and select the larger first \( D \) intrinsic mode components to construct the analytical signal \( I_i(t) = c_i(t) + j\phi(t) = a_i(t)e^{j\phi_i(t)} \). Then, HT transformation is performed on the selected intrinsic mode components to obtain the characteristics of instantaneous amplitude \( a_i(t) \), instantaneous phase \( \phi_i(t) \), and instantaneous frequency \( f(t) \). The obtained formula is as follows:

\[
\begin{align*}
   a_i(t) &= \sqrt{c_i^2(t) + \tilde{c}_i^2(t)}, \\
   \phi_i(t) &= \arctan \frac{\tilde{c}_i(t)}{c_i(t)} , \\
   f(t) &= \frac{1}{2\pi} \frac{d\phi_i(t)}{dt} .
\end{align*}
\]

The load data is subject to the sliding window mean difference cumulative sum algorithm for event detection and CEEMDAN + HT transformation for feature extraction, and the instantaneous amplitude, instantaneous frequency, and instantaneous phase of electrical equipment during operation are extracted as load features. Then, the training set and test set are divided by combining the load label, and the load identification model is input for training and identification.

3. ISMA-GRNN Load Identification Model

3.1. GRNN. GRNN is mainly composed of four layers of neurons: input layer, mode layer, summation layer, and output layer [19] are two transfer functions in the sum. Assuming the input data \( X = [x_1, x_2, x_3, \ldots, x_n] \), the calculation formula of mode layer data is shown in (6), where \( \sigma \) is the smoothing factor of GRNN.

\[ p_i = \exp \left[ \frac{(x - x_i)^T(x - x_i)}{2\sigma^2} \right], \quad i = 1, 2, \ldots, n. \]

There are two transfer functions in the summation layer [20]. One is arithmetic summation of the outputs of all mode neurons, and its transfer function is \( S_D = \sum_{i=1}^{n} p_i \). The other is weighted summation of all mode neurons, and its transfer function is \( S_{nj} = \sum_{i=1}^{n} y_{ij} p_i \). The output layer neuron is equal to the output data dimension, and its transfer function is a proportional function: \( y_j = S_{nj}/S_D \), \( j = 1, 2, \ldots, m \). Its structure is shown in Figure 1.

3.2. SMA. SMA is an optimization algorithm based on the foraging behavior of slime moulds. Slime moulds determine the regional location of food through diffusion behavior and foraging behavior. SMA first updates the location through three steps: approaching food, wrapping food, and obtaining food, and then, it finds the global optimal location by comparing fitness. It has the characteristics of fast convergence speed and strong optimization ability [21].

At the stage of approaching food, the approaching behavior of slime moulds is approximate to a mathematical equation, and its contraction rule mode is as follows:

\[ X(t+1) = \begin{cases} X_0(t) + \nu \cdot X_0(t) - X_a(t), & r < p, \\ \nu \cdot X_0(t), & r \geq p, \end{cases} \]

where \( \nu b \) is the random number between \([-a, a] \), \( \nu c \) is a parameter that is linearly decreasing from 1 to 0, \( t \) is the number of current iterations, \( X_0(t) \) represents the individual position with the best current fitness, \( X(t) \) represents the individual position with current slime moulds, \( X_a(t) \) and \( X_b(t) \) are two random individual positions, and \( W \) represents the weight coefficient of the slime moulds.

The control parameter \( p \) and parameter \( a \) with the weight coefficient \( W \) are updated using

\[ p = \tanh|S(i) - DF|, \]
\[ a = \arctan h\left( -\frac{t}{\max_x} \right) + 1 \],
\[ W(\text{smellIndex}(i)) = \begin{cases} 1 + r \cdot \log \frac{bF - S(i)}{bF - wF} + 1, & \text{condition}, \\ 1 - r \cdot \log \frac{bF - S(i)}{bF - wF} + 1, & \text{others}, \end{cases} \]
\[ \text{SmellIndex} = \text{sort}(S), \]
where \( i \in 1, 2, \ldots, n, n \) is the number of populations, \( S(i) \) represents the fitness value of the \( i \)-th slime moulds individual, and \( DF \) is the optimal fitness value achieved. Condition denotes the first half of individuals with fitness in the population, others denote the remaining individuals, \( r \) denotes the random number between \([0, 1]\), \( bF \) denotes the best fitness value acquired for the current iteration, \( WF \) denotes the worst fitness value for the current iteration. SmellIndex\((i)\) is the fitness value sequence (the minima problem is an increasing sequence).

During the food-wrapping phase, the location of slime moulds individuals is updated as follows:

\[
X^t = \begin{cases} 
\text{rand} \cdot (\text{UB} - \text{LB}) + \text{LB}, & \text{rand} < z, \\
\frac{X_B(t) + v_b \cdot (w \cdot X_A(t) - X_B(t))}{r < p}, & \text{otherwise}, \\
v_b \cdot X(t), & r \geq p,
\end{cases}
\]

where \( \text{UB} \) and \( \text{LB} \) are upper and lower bounds, random numbers are evenly distributed between 0 and 1, and \( z \) is a custom parameter value of 0.03.

In the stage of obtaining food, the value of \( v_b \) oscillates randomly between \([-a, a] \) and gradually approaches zero with the increase of the number of iterations. The value of \( vc \) oscillates between \([-1, 1] \) and finally tends to 0.

3.3. ISMA. SMA can only alleviate the problem of falling into local optimal solution to a certain extent. By adding the method of variation disturbance, the variation disturbance is added in the iterative process to make the slime moulds individual jump from local optimal solution, which can improve its global optimization ability, help GRNN find the best smoothing factor, and improve the identification accuracy of load identification method. The added variation disturbance formula is as follows:

\[
x^t_i = x^{t-1}_i \cdot [1 + A(0.5 - \text{rand})], \quad t = nT, n = 1, 2, \ldots
\]

where \( A \) is the coefficient of variation, which is mainly used to control the variation amplitude, \( T \) is the variation period, and \( \text{rand} \) is a random number evenly distributed between \([0, 1]\).

The main steps for ISMA to optimize GRNN smoothing factor parameters are as follows:

Step 1: SMA population initialization and mutation disturbance parameter initialization. After many experiments, the improvement effect is the best when \( t = 5 \) and \( a = 1 \).

Step 2: Sorting the initial fitness parameters after calculating the initial fitness parameters according to the initial value of slime moulds position, and selecting the larger value as the initial parameter.

Step 3: Calculating the control parameters, \( \alpha \) parameters and weight parameters according to equations (10) to (12).

Step 4: Updating the individual position of slime moulds by using equations (9) and (14), calculating the fitness value, comparing it with the initial fitness value, and updating the fitness value.

Step 5: Using equation (15) to update the position of variation disturbance, and using the fitness value to update the position of slime moulds and the global optimal value.

Step 6: Judging whether the number of iterations reaches the maximum number of iterations. If the number of iterations has been reached, output the best smoothing factor. Otherwise, go to step 3 to continue optimization.

4. Load Identification Model Based on ISMA-GRNN

4.1. Input and Output of ISMA-GRNN. Load data such as voltage, current, and power are input into NILM method after denoising and eliminating abnormal values. The load events are detected by the sliding window mean difference cumulative sum algorithm, and the instantaneous amplitude, instantaneous phase, and instantaneous frequency are extracted as load characteristics by CEEMDAN + HT transformation. Through many experiments, we find that the load identification accuracy is the highest when extracting the first five-dimensional intrinsic mode functions, so the input dimension of the load identification method is 15.

The load data used in this paper includes 9 kinds of electrical equipment, a total of 17 load states. The output of multiclass classification problems generally includes binary coded output and direct output. Binary encoded output is often used to identify situations where there are fewer electrical devices. Because there are many load states corresponding to the load data used in this paper, the direct output label is adopted, and the output dimension of the load identification method is 1. In order to facilitate the subsequent representation of electrical equipment and status, the electrical load and its status are numbered, and the corresponding labels are shown in Table 1.

4.2. ISMA-GRNN Implementation Process. After the load data is input into the NILM method, data preprocessing operations such as denoising and eliminating outliers are carried out in the first stage to reduce the impact of load data acquisition error on load identification results. In the second stage, the sliding window mean difference cumulative sum algorithm is used to detect the load event of power load data, and the access state of useless electrical equipment in the load process is eliminated. In the third stage, CEEMDAN + HT transform is used to decompose the current data of load events, extract instantaneous amplitude, instantaneous frequency, and instantaneous phase as load characteristics, and then the extracted load characteristic data set is
divided into training set and test set. In the fourth stage, ISMA-GRNN is used for load identification, and then the load identification results are output. The main process of ISMA-GRNN is shown in Figure 2.

The fitness function of ISMA is shown in the following equation:

$$\text{Acc} = \frac{C}{Z} \quad (14)$$

where $C$ represents the number of correct classification groups in the load identification results and $Z$ represents the total number of load data groups input into the load identification method.

In ISMA-GRNN, firstly initialize SMA and variation disturbance, then calculate the initial fitness value, control parameters, $\alpha$, and weight parameters, and update the individual position and fitness of slime mould by using equations (9), (14), and (10). Finally, judge whether the preset number of iterations is reached. If the preset number of iterations is reached, output the best smoothing factor. If the preset number of iterations is not reached, repeat equations (9), (14), and (10) to update the slime mould position.

4.3. Evaluation Index. Load identification is essentially a classification problem. On the binary classification problem, the classification methods are mainly evaluated by accuracy, precision ($P$), recall ($R$), and $F1$ [22]. For multiple-class classification problems, the average values of accuracy, precision, recall, and $F1$ are generally used to measure the performance of the algorithm. The average value is mainly calculated by micro average and macro average. Macro average calculates the accuracy, precision, recall, and $F1$ of each category and then calculates the average value. This calculation method can treat each category equally. The micro average is calculated by combining the contribution value of the category [23]. In NILM, the category of each piece of electrical equipment exists independently, so it is more reasonable to use Macro-$P$, Macro-$R$, accuracy, and Macro-$F$ measurement models. The macro parameters are calculated as follows:

<table>
<thead>
<tr>
<th>Load type</th>
<th>Microwave oven</th>
<th>Electric fan</th>
<th>Kettle</th>
<th>Incandescent lamp</th>
<th>Energy saving lamp</th>
<th>Water dispenser</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load status</td>
<td>Heating</td>
<td>Heat</td>
<td>1st</td>
<td>Turn on</td>
<td>Turn on</td>
<td>Heating</td>
</tr>
<tr>
<td>Label</td>
<td>1</td>
<td>preservation</td>
<td>gear</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Load type</td>
<td>Water dispense</td>
<td>Air-conditioner</td>
<td>2nd</td>
<td>Turn on</td>
<td>Turn on</td>
<td>8</td>
</tr>
<tr>
<td>Load status</td>
<td>Refrigeration</td>
<td>Refrigeration</td>
<td>Turn on</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Label</td>
<td>9</td>
<td>Heat</td>
<td>1-hot</td>
<td>16</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Refrigeration</td>
<td>Auxiliary heat</td>
<td>2-cold</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Load status    | Heating and cooling | 1-cold | 2-cold | Turn on |
| Label          | Refrigeration | 11     | 12     | 15     |
| Load type      | Refrigeration | 13     | 14     | 16     |
| Load status    | Refrigeration | 14     | 15     | 17     |
| Label          | Refrigeration | 15     | 16     | 17     |

Table 1: Correspondence table of electrical equipment and status label.
The confusion matrix [24] is the most intuitive and simplest way to measure the results of a multiple-class classification method. References [25, 26] used a confusion matrix to measure the load identification results, but due to the inconsistency in the number of sample groups for multiple classified loads, their consistency needed to be tested using the kappa. Kappa is a statistical index to measure classification accuracy based on confusion matrix calculation. In this paper, we use Kappa to analyze the consistency of load identification accuracy of different power devices.

Load identification time is an important index to measure the performance of load identification methods. In engineering applications, generally by comparing the recognition accuracy and running time of different algorithms, the algorithm with high recognition accuracy and short recognition time is selected as load identification method [27, 28].

In summary, we use the confusion matrix and kappa to represent the accuracy of load identification and use the accuracy of load identification and load identification time to evaluate the quality of load identification methods.

5. Experiments and Analysis

The software simulation platform used in this paper is Matlab 2018b, the hardware platform is intel i7-9700 3.0 GHz, and the RAM is 16 g. A total of 2743 groups of load
Table 2: Data group number table corresponding to electrical equipment and load status.

<table>
<thead>
<tr>
<th>Label</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group number</td>
<td>393</td>
<td>233</td>
<td>58</td>
<td>58</td>
<td>74</td>
<td>63</td>
<td>63</td>
<td>195</td>
<td>95</td>
<td>220</td>
<td>62</td>
<td>58</td>
<td>135</td>
<td>63</td>
<td>106</td>
<td>68</td>
<td>44</td>
</tr>
</tbody>
</table>

Figure 3: Confusion matrix results of four load identification methods. (a) Confusion matrix of HHO-GRNN. (b) Confusion matrix of WOA-GRNN. (c) Confusion matrix of SMA-GRNN. (d) Confusion matrix of ISMA-GRNN.

The data are collected. After data preprocessing, event detection, and feature extraction, the number of data groups input into the load identification method is 1988. The number of data groups corresponding to electrical equipment and load status is shown in Table 2.

We divide the processed feature dataset into training set and test set, then optimize it in the GRNN processed by the input optimization algorithm, output the best smoothness factor, and finally test the optimized GRNN using the test set data to get the load identification results. To highlight the superiority of ISMA-GRNN, the advantages of load identification in scenes of multiple electrical equipment are verified by comparing it with HHO-GRNN, WOA-GRNN, and SMA-GRNN using confusion matrix, precision (P), recall (R), F1, and load identification time as evaluation indexes.

In order to reflect the fairness of the result comparison, the initial population number of the four optimization algorithms is set to 30, the maximum number of iterations is set to 100, the upper limit of the smoothing factor is set to 1000, and the lower limit is set to 0.1. The simulation result confusion matrix is shown in Figure 3.

In Figure 4, we can draw the following conclusions: in the load identification of multiple electrical equipment use scenes, the load identification effect of ISMA-GRNN is significantly better than the other three load identification
methods. In Figure 4, the other three methods do not have high recognition accuracy for label 16, which is easy to misidentify as label 17. At the same time, WOA-GRNN can easily misidentify label 17 as label 16. The main reason is that these two kinds of electrical equipment have similar load data characteristics and have similar intrinsic mode functions when extracting features, so they are easy to misidentify. But ISMA-GRNN alleviates this problem to some extent, and its load identification accuracy for labels 16 and 17 is higher than 90%.

The main reason for misidentification of SMA-GRNN is that there are similar loads on the four load devices (Label 3, Label 6, Label 7, and Label 11), which leads SMA to fall in locally optimal solution when searching for smoothing factor, resulting in different degrees of misidentification in load identification. ISMA improves the accuracy of load identification for Label 3, Label 6, and Label 7 by increasing the update of mutation operation, taking into account the global optimum. In ISMA-GRNN, only the identification accuracy of tag 11 is 87%, and the load identification rate of other electrical equipment and states is more than 90%.

However, because the number of load data groups of different load labels is unbalanced, we need to use kappa to check the consistency of confusion matrix. Kappa under the four methods is shown in Table 3.

It is obvious from Table 3 that the kappa of ISMA-GRNN is the highest, indicating that, in the confusion matrix obtained by ISMA-GRNN load identification method, the prediction results are almost the same as the actual classification results, and the results have high credibility. At the same time, the kappa of the four load identification methods is greater than 0.93, which shows that the identification accuracy is high.
results of the four load identification methods have good repeatability, and their prediction results are highly consistent with the actual classification results. When the number of load data groups of different electric equipment is uneven, the confusion matrix cannot fully reflect the identification ability of the load identification method in the scenes of multiple electrical equipment. In order to reflect the rationality of the comparison of different load identification methods, we use precision \((P)\), recall \((R)\), and \(F_1\) to comprehensively evaluate the algorithm performance of the four load identification methods. The test results of the four load identification methods are shown in Table 4.

In Table 4, after comparing the identification effects of the four load identification methods on 17 load states, it is found that the identification effect of ISMA-GRNN is better, and its precision, recall, and \(F_1\) are higher than those of the other three methods. The minimum value of the classification index is usually used to measure the identification effect of the load identification method. The greater the minimum value, the better the performance of the classification model and the more applicable to different use scenarios.

By comparing the minimum values of the three classification indexes under the four load identification methods, it is found that the recall and \(F_1\) of ISMA-GRNN are higher than those of the other three methods, and the precision is only slightly lower than that of HHO-GRNN, indicating that ISMA-GRNN is more suitable for scenes of multiple electrical equipment use than the other three methods.

In engineering application, the less time it takes for load identification, the faster the iterative optimization process of the algorithm, and the stronger the performance of the algorithm. In addition, the parameters in Table 4 are mainly used to describe the identification effect of the load identification method on a single electrical equipment, which cannot well reflect the overall effect of the load identification model. Therefore, we use macro average and load identification time to comprehensively describe the overall performance of the load identification model in the scenes of multiple electrical equipment. The macro average and load identification time under the four load identification methods are shown in Table 5.

It can be seen from Table 5 that, among the four methods, there is little difference in macro average parameters between HHO-GRNN and WOA-GRNN, but there is a great difference in load identification time. The main reason is that the update strategy of HHO is more complex, which results in longer updates per iteration, so HHO-GRNN takes longer to reach 376.02s. Compared with HHO-GRNN and WOA-GRNN, SMA-GRNN improves the accuracy and time of load identification, but other indicators decrease accordingly. ISMA-GRNN, on the basis of SMA-GRNN, makes SMA take into account both global optimization and local development capabilities by adding mutation operation to the location update, which improves the macro average greatly. However, due to the addition of location mutation operation, the iteration time becomes longer, resulting in a load identification time of 1.63s longer than SMA-GRNN.

The load identification result figure is represented by comparing the load identification output label with the test set label, which can intuitively represent the identification effect of the load identification method. The identification results of the four load identification methods are shown in Figure 4. If the characteristics of two pieces of electric equipment are similar, the load identification method is easy to misjudge them, and both of them are misjudged as each other. It
can be seen from Figure 4 that the main misjudgment of HHO-GRNN is on label 6, label 7, and label 12, but we believe that the reason for the misjudgment is that the possibility of misidentification caused by similar load characteristics is very small. The main reason for the misjudgment of label 8 and label 13 in WOA-GRNN is that the number of data sample groups of tag 8 accounts for a large proportion, resulting in the misidentification of label 13. The results of SMA-GRNN show complex false recognition results. The main reason is that the algorithm falls into local optimum and cannot find the best smoothing factor, resulting in false recognition. Compared with the other three methods, the load identification result of ISMA-GRNN is the best. There is only misidentification of individual electrical equipment, and the identification error range is within the acceptable range.

6. Conclusion

In this paper, a noninvasive load monitoring load identification method based on ISMA-GRNN is proposed. Firstly, load event monitoring and feature extraction are carried out for load data, then mutation operation is added in the position update process of SMA, ISMA is used to optimize the smoothing factor of GRNN, and finally ISMA-GRNN is used for load identification, and the identification results are output. The load data we use includes 9 kinds of electrical equipment and a total of 17 power consumption states. The load identification accuracy, load identification precision, and load identification time are used as evaluation indexes. Through simulation experiments, the following conclusions are obtained:

1. Compared with HHO-GRNN and WOA-GRNN, SMA-GRNN has the highest load identification accuracy and the shortest load identification time, but the load identification precision is reduced.

2. After adding mutation operation, ISMA can move off the local optimal solution to a certain extent and find the best smoothing factor. Based on SMA-GRNN, the load identification accuracy of ISMA-GRNN is improved by 3.52%, the load identification precision is improved by 5.15%, and the load identification time is only increased by 1.63s.

3. For label 11 and label 12, the four methods have misidentification. The possible reason is that there is a certain time difference between the collection of operation data of electric equipment and the recording of load label when collecting load data for different power consumption states of the same electric equipment, resulting in load misidentification. Next, we will consider using the algorithm to record the load tag to reduce the time difference between running data collection and label recording.

Data Availability

Data sets cost a lot of energy to measure in the laboratory, so we choose not to disclose them.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


