

Article

An Industrial Load Classification Method Based on a Two-Stage Feature Selection Strategy and an Improved MPA-KELM Classifier: A Chinese Cement Plant Case

Mengran Zhou, Ziwei Zhu *, Feng Hu , Kai Bian and Wenhao Lai 

School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232001, China; mrzhou@aust.edu.cn (M.Z.); hufeng0106@163.com (F.H.); kbian92@163.com (K.B.); whlai9@163.com (W.L.)

* Correspondence: zwzhu0039@aust.edu.cn

Abstract: Accurately identifying industrial loads helps to accelerate the construction of new power systems and is crucial to today's smart grid development. Therefore, this paper proposes an industrial load classification method based on two-stage feature selection combined with an improved marine predator algorithm (IMPA)-optimized kernel extreme learning machine (KELM). First, the time- and frequency-domain features of electrical equipment (active and reactive power) are extracted from the power data after data cleaning, and the initial feature pool is established. Next, a two-stage feature selection algorithm is proposed to generate the smallest features, leading to superior classification accuracy. In the initial selection phase, each feature weight is calculated using ReliefF technology, and the features with smaller weights are removed to obtain the candidate feature set. In the reselection stage, the k-nearest neighbor classifier (KNN) based on the MPA is designed to obtain the superior combination of features from the candidate feature set concerning the classification accuracy and the number of feature inputs. Third, the IMPA-KELM classifier is developed as a load identification model. The MPA improvement strategy includes self-mapping to generate chaotic sequence initialization and boundary mutation operations. Compared with the MPA, IMPA has a faster convergence speed and more robust global search capability. In this paper, actual data from the cement industry within China are used as a research case. The experimental results show that after two-stage feature selection, the initial feature set reduces the feature dimensionality from 58 dimensions to 3 dimensions, which is 5.17% of the original. In addition, the proposed IMPA-KELM has the highest overall recognition accuracy of 93.39% compared to the other models. The effectiveness and feasibility of the proposed method are demonstrated.

Keywords: machine learning; industrial load identification; ReliefF; kernel extreme learning machine; marine predator algorithm; two-stage feature selection



Citation: Zhou, M.; Zhu, Z.; Hu, F.; Bian, K.; Lai, W. An Industrial Load Classification Method Based on a Two-Stage Feature Selection Strategy and an Improved MPA-KELM Classifier: A Chinese Cement Plant Case. *Electronics* **2023**, *12*, 3356. <https://doi.org/10.3390/electronics12153356>

Academic Editor: Nikolay Hinov

Received: 11 July 2023

Revised: 2 August 2023

Accepted: 4 August 2023

Published: 5 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the development of power demand-side management and smart grids worldwide [1], applying load monitoring and identification technologies for energy management has received increasing attention [2]. Industrial loads account for a large percentage of energy consumption and are considered an essential demand-side resource [3,4]. Achieving accurate and reliable identification of industrial loads is necessary for users to manage their electrical loads effectively [5]. Valid industrial load identification helps users to keep abreast of electricity consumption and adjust the production method according to the demand, save electricity, and optimize the industrial structure. At the same time, it helps power suppliers to adjust the electricity consumption structure and promote the development of an intelligent grid [6]. Therefore, it is crucial to develop an effective industrial load identification method.

The initial idea of load monitoring and identification was proposed in the 1980s [7]. Up to now, many methods have been proposed by domestic and foreign scholars [8–10]. Tao et al. [11] proposed a two-layer classification load identification and decomposition method based on the k-NN algorithm. The first level uses PQ load characteristics, and the second level uses the third and fifth harmonics of the current as the load characteristics. Considering that there is feature overlap when using a single load feature to identify devices, the requirements for the fine-grained classification of devices cannot be met. Zhang et al. [12] proposed a feature fusion using RGB color coding and implemented a load recognition method by improving the region-based fully convolutional networks (R-FCN) model. It is worth noting that the above studies are all based on the premise of high-frequency sampling data. However, high-frequency acquisition has high requirements for sampling hardware equipment, requires a large amount of data to be stored, and is challenging to promote. Therefore, load identification under low-frequency sampling has become a hot spot for domestic and international research. With the rapid development of deep learning techniques, some new solutions for load recognition have been provided in recent years. Kim et al. [13] proposed a method for load recognition using advanced deep learning and the long short-term memory recurrent neural network (LSTM-RNN) to improve the model's performance. Mukaroh et al. [14] used the generative adversarial network (GAN) to generate the noise distribution of the background load while constructing a convolutional neural network (CNN) which is implemented as a load classifier for the energy consumption analysis of device loads. This method avoids confusing actual load features and background load in Dorf in recognition, with a load recognition accuracy of 92.04%. Despite the advantages of deep learning techniques in the load recognition field, their performance needs to be improved by further developing the architecture of neural networks and training the model's parameters on large-scale datasets [15]. Finally, it is noteworthy that while load identification techniques for residential scenarios have been heavily researched, they are still an open challenge for industrial scenarios. This has a lot to do with the difficulty of data collection for industrial equipment and knowledge migration for industrial customer load characteristics [16].

Traditional low-frequency power load identification methods are mainly implemented through data processing, feature extraction, feature selection, and load classification. The effectiveness of feature extraction plays a vital role in the accuracy of load recognition [17]. Given this, time- and frequency-domain features of active and reactive power are extracted in this paper and combined as an initial feature pool. However, not all of these features help to identify the load type. In this case, feature selection is an effective technique to deal with these problems [18]. It finds the most efficient subset of features from the original feature set, thus achieving shorter model training time, avoiding dimensional disasters, and enhancing generalization by reducing overfitting [19]. Feature selection is divided into two main categories: filter and wrapper. Filtering methods usually rely on the statistical properties of the training data to evaluate the merits of feature subsets and are less time-consuming. However, since data mining algorithms are not involved, the results of filtering methods are only sometimes satisfactory [20]. Wrapper methods select the optimal feature subset based on the classifier's performance evaluation. Therefore, wrapper methods are more efficient in classification accuracy but more time-consuming [21].

In conclusion, the filter and wrapper methods have advantages and disadvantages. By combining the benefits of each type of method, hybrid selection methods have broad application potential [22]. ReliefF is a filter method that assigns a corresponding weight value to the feature [23]. The higher the weight value, the stronger the discriminative power of the feature. Therefore, some features with high weights can be selected for the wrapper method. The dimensionality of the original feature set is thus reduced. Today, researchers are focusing more on using metaheuristic algorithms in the form of wrapper models applied to wrapper feature selection [24]. Some of these examples include particle swarm optimization (PSO) [25], gray wolf optimization (GWO) [26], and the whale optimization algorithm (WOA) [27]. The marine predator algorithm (MPA) is a novel

metaheuristic algorithm. Compared with other metaheuristic algorithms, it performs well in finding solutions to optimization problems [28]. Therefore, this paper uses MPA as the second-stage wrapper feature selection method to select the best feature subset from the candidate features.

With the same feature set, choosing different classifiers is crucial to whether the load can be accurately identified [29]. As a classical algorithm in machine learning, the extreme learning machine (ELM) has been applied to many problems since its introduction [30,31]. ELM randomly generates input weights and hidden layer biases, leading to unstable algorithms. To solve this problem, a kernel function is introduced based on ELM to obtain the kernel extreme learning machine (KELM). In solving some practical engineering problems, KELM shows high classification accuracy and good generalization performance [32]. However, the performance of KELM is affected mainly by the kernel function parameter γ and the penalty factor C . Therefore, many optimization algorithms are used to optimize its parameters to improve its performance [33–35]. Given the excellent merit-seeking ability of the MPA algorithm [36], this paper utilizes it to optimize the critical parameters of the KELM. In addition, to further improve the solution's quality, this study presents the design of an improved marine predator algorithm (IMPA). A logical self-mapping is used to generate chaotic sequences for initialization and to improve population diversity. The boundary mutation operation also performs the optimization of the MPA merit-seeking process. The experimental results show that the proposed IMPA obtains better model parameters than other heuristic algorithms. The IMPA-KELM model achieves a higher accuracy of load identification.

In summary, this paper proposes a new industrial load classification method based on the significant advantages of each technique, combining time–frequency-domain features, novel two-stage feature selection, and the IMPA-KELM classifier. Firstly, the time–frequency-domain features of the PQ of cement plant equipment are extracted as the initial feature set. Then, a two-stage feature selection method based on ReliefF, MPA, and KNN classifiers is proposed to select a combined feature set that considers the feature dimension and classification accuracy. A candidate feature set is selected from all features in the initial selection stage using the ReliefF technique. Then, an MPA-based KNN classifier is designed to find a better-combined feature set in the candidate feature set. Next, the MPA algorithm is improved by designing chaotic population initialization and boundary mutation strategies. Finally, the new reduced feature set is fed into the improved MPA-KELM classifier to achieve high-accuracy industrial load identification. The proposed method is tested for different equipment types in cement plants within China. The results show the superiority of the proposed method.

The remainder of this paper is organized as follows: Section 2 summarizes the theories used in the proposed method. Sections 3 and 4 prove that the proposed method has high effectiveness and superiority through a large number of experiments. Section 5 discusses the conclusions and future directions of this work.

2. Methodology

2.1. Framework of the Proposed Method

Based on the above methods, a new load identification method is proposed in this paper. It mainly includes power data preprocessing and feature extraction, two-stage feature selection combining filter and wrapper, and a classification model based on an improved MPA algorithm with optimized KELM. The framework of the proposed method is shown in Figure 1.

Firstly, the collected power data of cement plant equipment are cleaned, including median filtering. Then, the time- and frequency-domain features of the load data PQ are extracted as the original feature set. In the two-stage feature selection process, the candidate features are first selected from the original feature set using the ReliefF technique. Then, the MPA-based KNN classifier is designed for feature optimization to filter features further and obtain the optimal feature set. Next, MPA's convergence speed and optimization-

seeking accuracy are improved by introducing a chaotic initialization strategy and boundary mutation operation. Finally, the improved MPA algorithm is used to optimize the kernel parameters and penalty factors of KELM to identify industrial loads accurately.

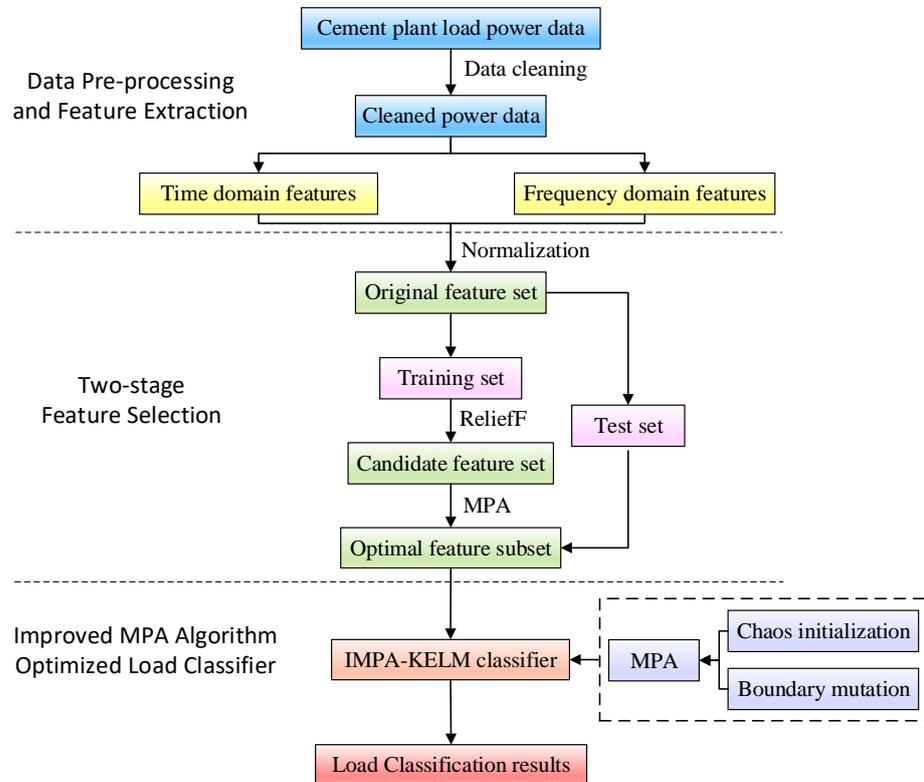


Figure 1. The framework of the proposed method.

2.2. Time–Frequency-Domain Features

The power signal’s time- and frequency-domain characteristics vary for different loads in steady-state operation [37]. The power signal characteristics for different loads can be obtained by analyzing the power signal’s time-domain waveform and frequency spectrum [38]. In this paper, 16 time-domain features (mean, root mean square, square root mean, absolute mean, and skewness, etc.) are extracted from the power signal of the load and are shown in Table 1, where $x(n)$ is a signal series for $n = 1, 2, \dots, N$ and N is the number of data points.

Table 1. Features extraction of the time domain.

Feature	Equation	Feature	Equation
Mean value	$TD_1 = \frac{\sum_{n=1}^N x(n)}{N}$	Minimum value	$TD_9 = \min\{x(n)\}$
Root mean square	$TD_2 = \sqrt{\frac{\sum_{n=1}^N (x(n))^2}{N}}$	Peak-to-peak value	$TD_{10} = TD_8 - TD_9$
Square mean root	$TD_3 = \left(\frac{\sum_{n=1}^N \sqrt{ x(n) }}{N}\right)^2$	Waveform index	$TD_{11} = \frac{TD_2}{TD_4}$
Absolute mean	$TD_4 = \frac{\sum_{n=1}^N x(n) }{N}$	Peak index	$TD_{12} = \frac{TD_8}{TD_2}$
Skewness	$TD_5 = \frac{\sum_{n=1}^N (x(n) - T_1)^3}{N}$	Pulse index	$TD_{13} = \frac{TD_8}{TD_4}$
Kurtosis	$TD_6 = \frac{\sum_{n=1}^N (x(n) - TD_1)^4}{N(TD_2^2)}$	Margin index	$TD_{14} = \frac{TD_8}{TD_3}$
Variance	$TD_7 = \frac{\sum_{n=1}^N (x(n) - T_1)^2}{N}$	Skewness index	$TD_{15} = \frac{TD_5}{(\sqrt{TD_7})^3}$
Maximum value	$TD_8 = \max\{x(n)\}$	Kurtosis index	$TD_{16} = \frac{TD_6}{(\sqrt{TD_7})^4}$

The frequency domain can disclose the data that cannot be discovered in the time domain. In this work, 13 frequency-domain features (mean, frequency center, root variance, mean square deviation, kurtosis, etc.) are extracted using FFT, as shown in Table 2, where $s(k)$ is a spectrum for $k = 1, 2, \dots, K$; K is the number of spectrum lines; and f_k is the frequency value of the k -th spectrum line.

Table 2. Features extraction of the frequency domain.

Feature	Equation	Feature	Equation
Mean	$FD_1 = \frac{\sum_{k=1}^K s(k)}{K}$	Kurtosis	$FD_8 = \frac{\sum_{k=1}^K (f_k - FF_{10})^4 s(k)}{K(FF_6^4)}$
Variance of mean frequency	$FD_2 = \frac{\sum_{k=1}^K (s(k) - FD_1)^2}{K}$	Root mean square ratio	$FD_9 = \frac{\sum_{k=1}^K \sqrt{(f_k - FD_{10})s(k)}}{K\sqrt{FD_5}}$
Skewness power spectrum	$FD_3 = \frac{\sum_{k=1}^K (s(k) - FD_1)^3}{K(\sqrt{FD_2})^3}$	Frequency center	$FD_{10} = \frac{\sum_{k=1}^K f_k s(k)}{\sum_{k=1}^K s(k)}$
Kurtosis power spectrum	$FD_4 = \frac{\sum_{k=1}^K (s(k) - FD_1)^4}{K(FD_2^2)}$	Root mean square	$FD_{11} = \sqrt{\frac{\sum_{k=1}^K f_k^2 s(k)}{\sum_{k=1}^K s(k)}}$
Root variance	$FD_5 = \sqrt{\frac{\sum_{k=1}^K (f_k - FD_{10})^2 s(k)}{K}}$	Mean frequency that crosses the mean of the time-domain signal	$FD_{12} = \sqrt{\frac{\sum_{k=1}^K f_k^4 s(k)}{\sum_{k=1}^K f_k^2 s(k)}}$
Coefficient of variability	$FD_6 = \frac{FD_5}{FD_{10}}$	Stabilisation factor	$FD_{13} = \frac{\sum_{k=1}^K f_k^2 s(k)}{\sqrt{\sum_{k=1}^K s(k) \sum_{k=1}^K f_k^4 s(k)}}$
Skewness	$FD_7 = \frac{\sum_{k=1}^K (f_k - FD_{10})^3 s(k)}{K(FD_5^3)}$		

2.3. ReliefF

The ReliefF algorithm was developed by Kononeill in 1994 to extend the functionality of the Relief algorithm to deal with multi-category problems [39]. The ReliefF algorithm evaluates attributes' quality based on attribute values' ability to distinguish samples close to each other [40]. In dealing with multi-class problems, the R -th sample is randomly removed from the training data one at a time, and then the k -nearest neighbor samples (Near Hits) of the same kind as R are found in the training sample set, denoted as H_j . The k -nearest neighbor samples (Near Misses), marked as $M_j(C)$ ($j = 1, 2, \dots, k, C \neq \text{class}(R)$), are identified from the set of samples that differ from R . The update of the weight $W(A)$ of the attribute A depends on the sample R , the nearest neighbor samples H_j of the same kind of R , and the samples M that are different from the class of R , as shown in Equation (1):

$$W(A) = W(A) - \sum_{j=1}^k \frac{\text{diff}(A, R, H_j)}{mk} + \sum_{C \neq \text{class}(R)} \left[\frac{P(C)}{1 - P(\text{class}(R))} \sum_{j=1}^k \text{diff}(A, R, M_j(C)) \right] / (mk) \tag{1}$$

where $\text{diff}(A, R_1, R_2)$ denotes the distance between samples R_1 and R_2 on feature A ; $P(C)$ denotes the probability of the class C target; and $M_j(C)$ denotes the j -th nearest neighbor sample in the class C target.

2.4. Marine Predator Algorithm

The marine predator algorithm is a new metaheuristic algorithm inspired by nature and proposed by Faramarzi et al. [41]. It relies on the movements of marine predators in search of prey, namely Levy and Brownian stochastic activities, respectively. When prey is scarce, predators use Levy motion, while when prey is abundant, they use Brownian motion. The mathematical formulation of the MPA algorithm is as follows.

1. Initialization: construct the *Prey* matrix and *Elite* matrix. They contain the position vectors of variable random positions and repeated best fitness function in the proposed domain, respectively.

- Phase 1 [while $Iter < \frac{1}{3}Iter_{max}$]: The first one-third of the iterations are dedicated to this phase, when the prey moves using the Brownian strategy and the predator is stationary. Equation (2) reflects this phase.

$$\begin{cases} \vec{S}_i = \vec{R}_B \otimes (\vec{Elite}_i - (\vec{R}_B \otimes \vec{Prey}_i)) \\ \vec{Prey}_i = \vec{Prey}_i + (0.5\vec{R} \otimes \vec{S}_i) \end{cases} \quad i = 1, 2, \dots, n \quad (2)$$

where \vec{S}_i is the step size, \vec{R}_B is a vector with random numbers based on Brownian motion normal distribution and \vec{R} is a random uniform vector between [0, 1].

- Stage 2 [while $\frac{1}{3}Iter_{max} < Iter < \frac{2}{3}Iter_{max}$]: The predator uses Brownian movement and the prey uses Levy movement. The first half of the population was updated using Equation (3).

$$\begin{cases} \vec{S}_i = \vec{R}_L \otimes (\vec{Elite}_i - (\vec{R}_L \otimes \vec{Prey}_i)) \\ \vec{Prey}_i = \vec{Prey}_i + (0.5\vec{R} \otimes \vec{S}_i) \end{cases} \quad i = 1, 2, \dots, n \quad (3)$$

where the vector \vec{R}_L contains random values based on the Levy motion normal distribution. In addition, the other half of the population is updated using Equation (4).

$$\begin{cases} \vec{S}_i = \vec{R}_B \otimes ((\vec{R}_B \otimes \vec{Elite}_i) - \vec{Prey}_i) \\ \vec{Prey}_i = \vec{Elite}_i + (0.5C_f \otimes \vec{Step}_i) \end{cases} \quad i = 1, 2, \dots, n \quad (4)$$

where C_f controls the predator step length and is calculated as in Equation (5).

$$C_f = [1 - (Iter / Iter_{max})]^{(2 \times Iter / Iter_{max})} \quad (5)$$

- Phase 3 [While $Iter > \frac{2}{3}Iter_{max}$]: This occurs in the last third of the iteration. In this phase, the predator moves using Levy motion and the prey is updated using Equation (6).

$$\begin{cases} \vec{S}_i = \vec{R}_L \otimes ((\vec{R}_L \otimes \vec{Elite}_i) - \vec{Prey}_i) \\ \vec{Prey}_i = \vec{Elite}_i + (0.5C_f \otimes \vec{S}_i) \end{cases} \quad i = 1, 2, \dots, n \quad (6)$$

- The effects of FADs: Environmental issues can also cause changes in the behavior of marine predators. One example is the effects of fish-aggregating devices (FADs), also known as eddy formation. The mathematical model of FAD's effect is defined in Equation (7):

$$\vec{Prey}_i = \begin{cases} \vec{Prey}_i + C_f [X_{min} + \vec{R}_L \otimes (X_{max} - X_{min})] \otimes \vec{U}, r \leq FADs \\ \vec{Prey}_i + [FADs(1 - r) + r](\vec{Prey}_{r1} - \vec{Prey}_{r2}), r > FADs \end{cases} \quad (7)$$

where $FADs = 0.2$ denotes the probability of being affected by FADs during optimization, and U is constructed by generating a random vector in an array of binary vectors containing 0 and 1. If the array is less than 0.2, the array is changed to 0; if it is greater than 0.2, it is changed to 1. r is a uniform random number in [0, 1], and the r_1 and r_2 subscripts denote the random indexes of the prey matrix.

2.5. MPA Feature Selection

In this feature selection phase, MPA aims to extract the candidate features filtered by ReliefF. By designing the fitness function, the purpose of considering higher classification accuracy and fewer feature numbers is achieved. The necessary steps for MPA feature selection are summarized as follows:

MPA creates N predators in the initialization phase, with each agent representing a set of features to be evaluated. Before evaluating the fitness values, each solution X_i is translated into X_i^{SF} using the binary operator using Equation (8). With this step, all features are encoded as [0, 1]. If the value of an element is within [0.5, 1], the feature is retained; if not, it is excluded.

$$X_i^{SF} = \begin{cases} 1 & \text{if } X_i > 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

The fitness function is determined for each agent X_i^{SF} to determine the quality of these features after evaluating the subset of selected features. Classification accuracy and feature cost are the two key factors used to design the fitness function, where feature cost is obtained from the ratio of selected features to the total number of features. The smaller the feature cost (the fewer the selected features), the higher the classification accuracy, indicating that the subset of features is superior. Therefore, the i th solution's fitness is determined by Equation (9):

$$FS\text{-fitness} = \mu(1 - Acc) + (1 - \mu) \frac{N_S}{N_F} \quad (9)$$

where $(1 - Acc)$ represents the error rate of KNN classification, selected features are denoted by N_S , N_F indicates the total number of features, μ denotes the classification error weight, $(1 - \mu)$ indicates the feature selection quality weights, and μ can range between 0 and 1. In our experiments, μ is set to 0.9.

Next, we enter the update phase, in which the process of updating the solution of MPA introduced in Section 2.4 is executed in an orderly manner. The process is reproduced in the case that the termination condition is satisfied. The termination condition corresponds to the maximum number of iterations that allow us to evaluate the performance of the MPA algorithm. Then, the best solution X_{best} is returned and transformed to determine the number of relevant features. The specific MPA feature selection steps are shown in Figure 2.

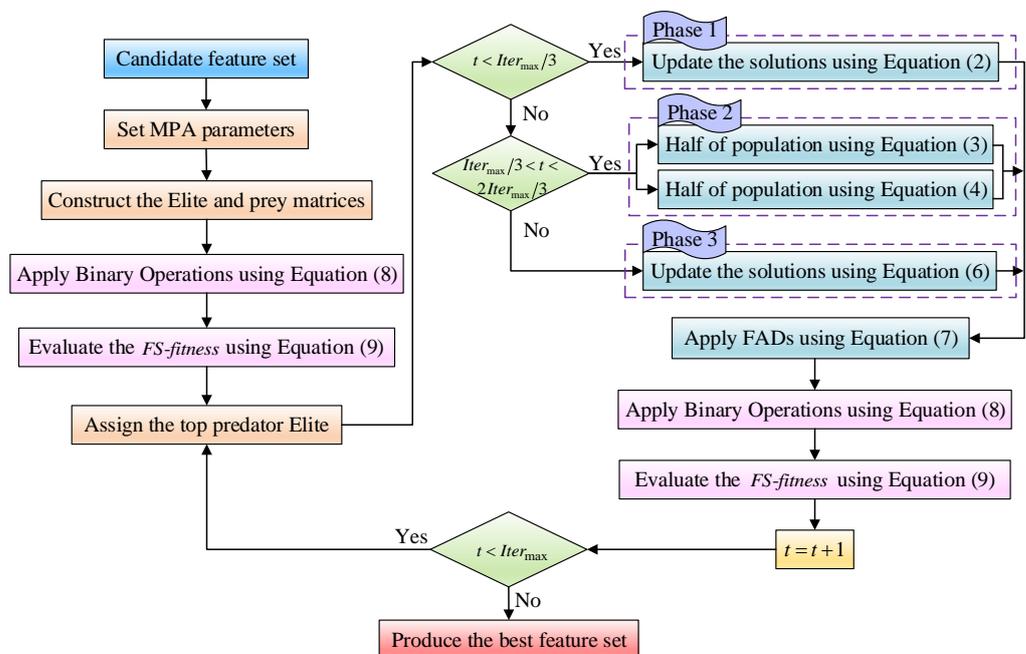


Figure 2. MPA feature selection flowchart.

2.6. Kernel Extreme Learning Machine Optimized by the Improved Marine Predator Algorithm (IMPA-KELM)

2.6.1. Kernel Extreme Learning Machine (KELM)

KELM is one of the most popular learning techniques. It is derived from the traditional ELM [42]. However, it can obtain better results than ELM. In KELM, different kernel functions are used for various applications. The kernel mapping function of ELM can be written as

$$\Omega_k = HH^T, \Omega_{k,i,j} = h(p_i)h(p_j) = K(p_i, p_j) \tag{10}$$

where H represents the output matrix of the hidden layer.

In addition, the generalized output function of KELM can be defined as

$$F(\lambda) = \begin{bmatrix} K(p, p_1) \\ \dots \\ K(p, p_N) \end{bmatrix}^T \left(\frac{1}{\lambda} + \Omega_k \right)^{-1} Q \tag{11}$$

where Q is the output vector and λ denotes the penalty parameter.

Among the many kernel functions, the radial basis function (RBF) is one of the most widely used functions (this paper also uses RBF as the kernel function). The RBF kernel can be defined as

$$K(p, p_i) = \exp(-\gamma \|p - p_i\|^2) \tag{12}$$

where γ is a kernel parameter.

2.6.2. Improved Marine Predator Algorithm

a. Chaos initialization strategy

Using the randomness, ergodicity, and initial value sensitivity of chaotic motion to improve the efficiency of stochastic optimization algorithms is called chaos optimization. Among various models of chaotic sequences, it is shown that the chaotic sequences obtained using the logistic self-mapping function are better than the logistics mapping [43]. A chaotic series is generated using the logistic self-mapping function, as shown in Equation (13).

$$cx_j^{k+1} = 1 - 2 \times (cx_j^k)^2, cx_j^0 \in (-1, 1), j = 1, 2, \dots, d \tag{13}$$

In this case, the initial values cannot be 0 and 0.5 in order to avoid chaotic sequences with 1 s or 0.5 s intervals that are all 1 or 0.5. cx_j^k represents the j -th dimensional component of a chaotic variable, and k is the number of iteration steps.

Step 1: For M prey individuals in D -dimensional space, the initial variables are generated randomly in the $(-1, 1)$ interval by initializing the chaotic variables according to the nature of the logical self-map function.

Step 2: Iterating according to Equation (13), the $MaxGeneration * D - 1$ chaotic variables generated by the logical self-mapping, together with the initial chaotic variables, correspond to all $MaxGeneration * D$ prey individuals.

Step 3: The resulting sequence of chaotic variables is transformed into the search space of the objective function according to Equation (14) to generate the prey matrix for the initial population of M individuals with the following equation.

$$x_{i,d} = Lb + (Ub - Lb) * y_{i,d} \tag{14}$$

where Lb and Ub denote the lower and upper limits of the d -th dimension of the search space, respectively. $y_{n,d}$ is the d -th dimensional chaotic variable corresponding to the i -th prey generated according to Equation (14). $x_{i,d}$ denotes the coordinate value of the i -th prey in the d -th dimension of the search space.

b. Boundary mutation strategy

In the process of optimization with the marine predator algorithm, the location of the prey is likely to break its boundary value. In this case, the traditional approach replaces the individual's position with the boundary value that is exceeded by the domain constraint formula shown in Equation (15).

$$x_i = \begin{cases} x_{\max}, & x_i > x_{\max} \\ x_{\min}, & x_i < x_{\min} \end{cases} \quad (15)$$

where x_{\max} is the upper limit of the search space and x_{\min} is the lower limit of the search space. This boundary control strategy tends to cause the algorithm to fall into the local optimum, and the points beyond the boundary are all constrained at the boundary, which may cause the algorithm to converge prematurely at the boundary and reduce the search rate of the algorithm.

Therefore, a boundary mutation strategy is introduced to solve this problem when the domain constraint formula is shown in Equation (16).

$$x_i = \begin{cases} 2x_{\max} - x_i, & x_i > x_{\max} \\ 2x_{\min} - x_i, & x_i < x_{\min} \end{cases} \quad (16)$$

The boundary mutation operation can keep the position of the searching individuals in the population within the feasible domain at all times, preventing the marine predator algorithm from falling into the local optimum at the boundary. At the same time, the diversity of the population is improved to a certain extent, and the boundary variation operation effectively improves the optimal search performance of the marine predator algorithm.

2.6.3. Steps for IMPA to Optimize KELM

The above describes the principles of the KELM and IMPA algorithms, respectively. In the implementation of the KELM learning algorithm, the performance of the model will be affected if the penalty coefficient C and the kernel parameter γ are not selected properly. Therefore, in this paper, the IMPA optimization algorithm is combined with KELM to optimize C and γ to improve the performance of the model to some extent. The classification error rate is used as the fitness function *CE-fitness*. The smaller the value of *CE-fitness*, the lower the error rate and the better the classification effect. Based on this, the IMPA-KELM classifier can be obtained. The detailed optimization steps are as follows:

- (1) The load features after two-stage feature selection are divided into the training and test samples at a ratio of 8:2.
- (2) The prey matrix is designed according to the chaotic optimization strategy (Equations (13) and (14)), while the IMPA parameters are initialized, the number of populations is set to 30, and the maximum number of iterations is 50. The upper and lower bounds for the KELM penalty coefficient C and the kernel parameter γ are set to 0.001 and 1000, respectively, for the optimization search.
- (3) Calculate the fitness value of prey, and update the best fitness value.
- (4) Update the positions of predators and prey at different iteration periods according to Equations (2)–(6). Calculate and update the optimal fitness values again. The prey moves according to the FADs in Equation (7), thus changing the predator's behavior. In this case, the position correction is performed using Equation (16) for prey that are beyond the search boundary.
- (5) Steps (3) and (4) are repeated until the maximum number of iterations is reached, and the predator position with the best adaptation is obtained and retained as the best C and γ parameters for KELM.
- (6) KELM classification model is developed using the obtained parameters C and γ . Implement load identification of the test set samples using IMPA-KELM.

2.7. Performance Metrics

To verify the effectiveness of the proposed method for load identification in this paper, several multi-class classification evaluation metrics are used to evaluate models. The performance metrics used in the paper include accuracy, precision, recall, and F1-score, in addition to classification accuracy.

$$Accuracy = \frac{TP + FP}{TP + TN + FP + FN} \quad (17)$$

$$Precision = \frac{TP}{TP + FP} \quad (18)$$

$$Recall = \frac{TP}{TP + FN} \quad (19)$$

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (20)$$

where true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are used to show how much the applied model correctly and wrongly classified load category.

3. Experiments and Results

In this chapter, the performance of the proposed method is verified using the autonomously collected load dataset. The operation process of the whole procedure and the functions of its various parts are also explained in detail. A desktop computer with an Inter Core i7 (10700) processor and 32 GB RAM was used, where MATLAB R2022a was employed as a simulation tool.

In this paper, a cement plant in East China was used as a research case to verify the effectiveness of the proposed method. To ensure the diversity of electrical loads, equipment in different workshops was selected considering the various operating characteristics of the loads in different workshops. These pieces of equipment were a raw material mill (RMM), a kiln tail high-temperature fan (KTHTF), an exhaust gas treatment fan (EGTF), a coal mill fan (CMF), a coal mill motor (CMM), a kiln head induced draft fan (KHIDF), a fixed roller press (FRP) and a dynamic roller press (DMP), providing a total of eight types of motor equipment. The electrical parameters collected were active power P (kW) and reactive power Q (kVar). The electrical data time range was from 12 October 2021 to 30 December 2021, recorded every 5 min and sampled at low frequency. It is worth noting that there were missing cases during the data collection and entry process. For example, the FRP's reactive power for 23 October and 3 November was missing. Figures 3 and 4 show the active and reactive power of different loads. Figures 3a and 4a show the power state diagrams of the eight pieces of electrical equipment, and it can be seen that there are differences in the power characteristics of the operation of different pieces of power equipment. Some of the loads operate with similar power variation trends, especially the two loads, FRP and DRP. This phenomenon can be seen in Figures 3b and 4b. This makes it more difficult to distinguish between the two loads accurately.

After acquiring and analyzing the load dataset above, the data were pre-processed. This step was essential. This paper used the P and Q of electrical appliances as input information. Firstly, the power data of each electrical device were divided as a data sample, with the length of a day as the basic unit. Then, the samples with all zero power (i.e., the electrical equipment is not used) were excluded. In the case of missing data in the dataset, if there were a few missing values in the whole day, linear interpolation was used to fill them. If there were more missing values for that day, the sample was discarded directly, and no further interpolation was performed. In addition, it can be seen from Figures 3 and 4 that some load power curves have obvious burrs due to outliers in the obtained electric load operating power. Given this, this paper adopted the median filtering method to reduce the noise of the power data. As shown in Figure 5, the median filtering

effect of the reactive power curve from 24 December to 30 December is plotted for the raw material mill (RMM) as an example.

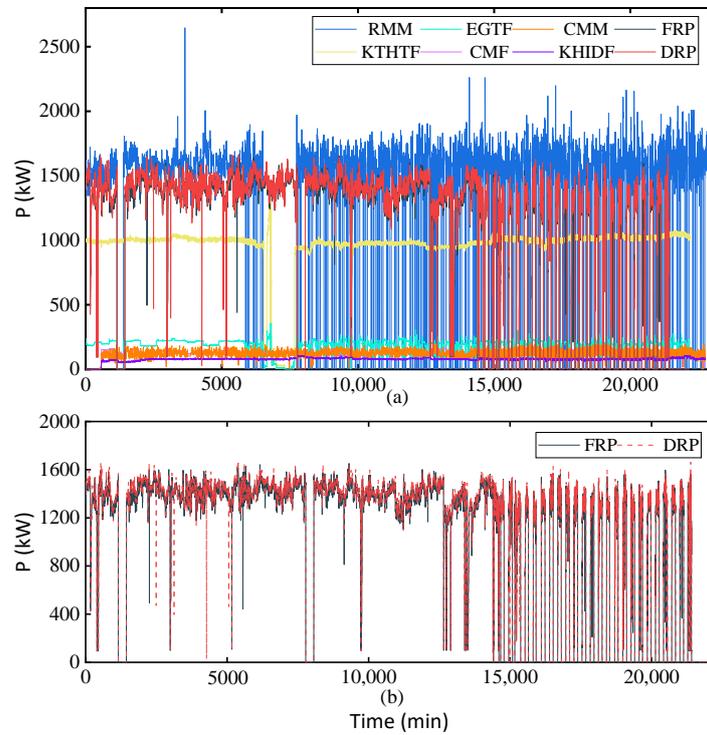


Figure 3. Active power curve of different electric loads. (a) Active power diagrams for eight types of electrical equipment. (b) Comparison of active power between FRP and DRP loads.

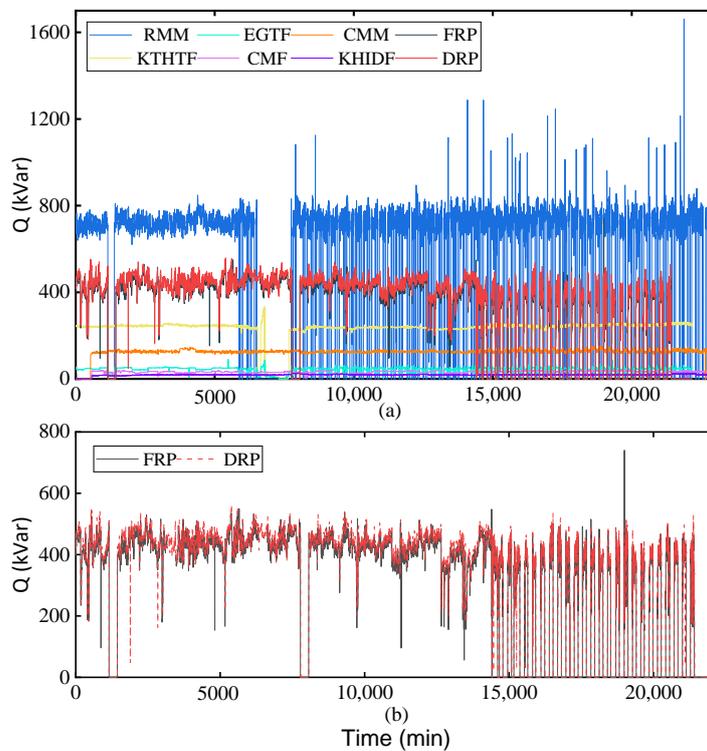


Figure 4. Reactive power curve of different electric loads. (a) Reactive power diagrams for eight types of electrical equipment. (b) Comparison of reactive power between FRP and DRP loads.

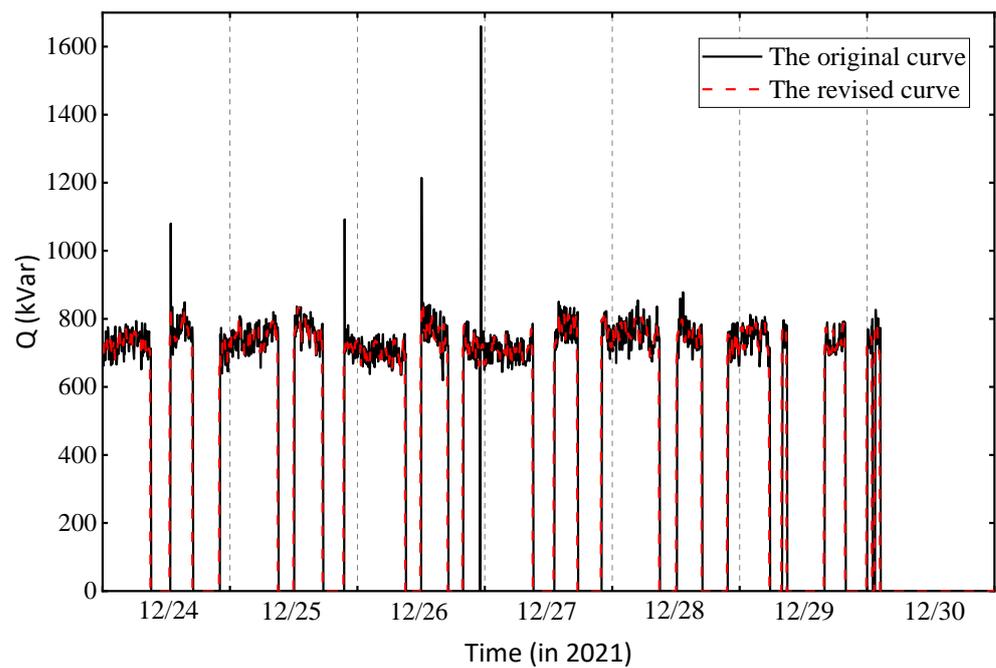


Figure 5. Raw material mill (RMM) median filtering effect graph (24 December–30 December).

After pre-processing, the active and reactive power data for the eight types of cement plant equipment are shown in Table 3. The missing data for each load collected differed, resulting in a slightly different final sample size for them. As can be seen from the table, the RMM has the most significant operating power consumption among the eight loads. For the two motors, CMF and CMM, the active power operating state parameters do not differ much, but the reactive power has a more noticeable difference. The KHIDF is the only one working during the whole data collection period. The FRP and DRP two electric loads operate with similar power parameters, corresponding to the above analysis.

The above pre-processing process obtains “clean” raw power data. At this point, the data could not be directly input into the classifier. It was necessary to extract useful information using a feature extraction technique that converts the raw input data into simplified information or encodes the relevant information to identify the load class. Sixteen time-domain features and thirteen frequency-domain features were extracted from the denoised power data (active and reactive) to generate an initial load feature set that contained rich information about the load identification system. Finally, a feature pack containing $2 \times (16 + 13) = 58$ features was obtained to represent the load operation state.

After feature extraction, the most critical work was selecting compelling features and removing redundant ones. This study employed a two-stage feature selection strategy, with the ReliefF filtering method being used in the initial screening.

The 58 joint time–frequency-domain feature weights calculated using ReliefF are shown in Figure 6. The weight of each feature varies, and a larger weight indicates the corresponding feature is more efficient. The frequency-domain features generally have a higher weighting factor than the time-domain features. The weights of the time-domain features vary widely for both active and reactive power, and the first-time domain feature TD_1 , which represents the mean value (Mean value), has a high weight. The time-domain features TD_5 , TD_6 and $TD_{11} \sim TD_{16}$ have less weight. In this paper, the candidate feature set comprises the top 80% of the weighted features. Through this feature selection stage, the number of features in the candidate feature set is reduced from the original 58 to 46.

Table 3. Load information table.

No.	Device ID	Device Type	P/kW				Q/kVar				Sample Size
			Max	Min	Mean	Variance	Max	Min	Mean	Variance	
1	RMM	Raw material mill	2.6467×10^3	0	1.1478×10^3	713.4977	1.0959×10^3	0	514.6201	329.7208	76
2	KTHTF	kiln tail high temperature fan	1.1358×10^3	0	953.7741	199.1667	333.5100	0	232.5849	48.6378	77
3	EGTF	Exhaust gas treatment fan	354.8300	0	173.9166	56.7118	84.2600	0	41.2736	12.1320	77
4	CMF	Coal mill fan	165.3700	0	133.2663	22.7862	40.6000	0	31.7787	5.7884	78
5	CMM	Coal mill motor	195.4000	0	121.2151	27.3095	148.4600	0	125.8172	14.7483	78
6	KHIDF	Kiln head induced draft fan	105.9500	49.4900	79.3093	7.6361	26.6500	12.0500	19.1434	1.9674	77
7	FRP	Fixed roller press	1.6523×10^3	0	1.0958×10^3	557.2647	549.7500	0	335.6737	169.0129	71
8	DRP	Dynamic roller press	1.6599×10^3	0	1.1245×10^3	561.6063	558.8900	0	351.2886	173.1773	73

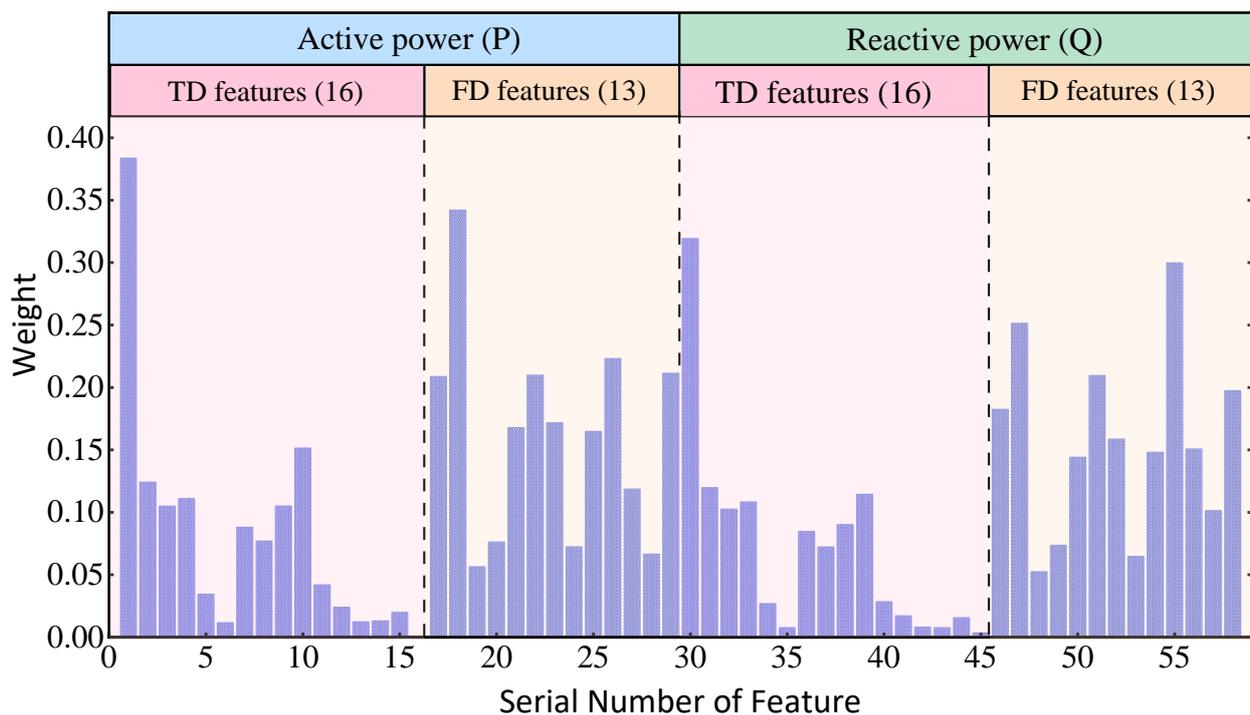


Figure 6. ReliefF weights of 58 joint time–frequency-domain features.

The best subset of features must maximize the classification accuracy and minimize the classification error rate. In addition, it must have a minimum number of selected features. Therefore, MPA is used as a re-feature selection method to perform a screening of the candidate feature set and maintain a balance between a high classification accuracy and a small number of selected features. In the MPA, the number of populations and iterations is set to 10 and 100, respectively. In the expression of Equation (9), the error rate weight μ is set to 0.9 to seek a lower error rate for the classification model. The feature selection quality weight is 0.1 to ensure the quality of the selected features while pursuing classification efficiency. Finally, the dimension of the preselected features can be further reduced from 46 to 3. Two-step feature selection reduces the number of features from 58 to 3. The three features selected after the feature selection are the frequency center (FD_{10}), coefficient of variation (FD_6) and variance (TD_6) of reactive power. The final number of features is only 5.17% of the initial number, significantly reducing the number of features.

After feature extraction and selection, the processed data were fed into the classifier for load identification. The classifier used in this paper was IPMA-KELM, where IPMA was derived from the chaos strategy and the boundary mutation strategy to improve MPA. IMPA was used to optimize the kernel parameters γ and penalty coefficient C of KELM. The classification error rate of KELM was the fitness function. The initial population number and the maximum number of iterations of IMPA were set to 30 and 50, respectively. The

upper and lower bounds of γ and C seeking of KELM were 0.001 and 1000. The number of samples in the training and test sets in the experiment was 486 and 121, respectively.

The confusion matrix of the experimental results of the proposed model is shown in Figure 7. Where, the integers in the figure represent the number of samples of devices in the test set. Blue represents the number of correctly classified samples, and orange represents the number of incorrect samples. It was obtained by identifying eight different types of cement plant equipment. There were 16 samples of CMF in the test set, among which the proposed model misclassified only one sample as an EGTF. The recognition accuracy of the coal mill fans is 93.75%. From the analysis in Section 2.3, it is clear that the power data of the dynamic roller press and the fixed roller press are highly similar, with little differentiation. This makes it more difficult for the model to identify both accurately. Even so, the proposed model achieves 71.43% and 80% accuracy for both, respectively. The recognition accuracy of the other five devices is 100%. The excellent classification performance of the proposed model is demonstrated.

True Class	DRP						3	12	80.0%	20.0%	
	FRP						10	4	71.4%	28.6%	
	KHIDF					15			100%		
	CMM					16			100%		
	CMF			1	15				93.8%	6.2%	
	EGTF			15					100%		
	KTHTF		15						100%		
	RMM	15							100%		
		RMM	KTHTF	EGTF	CMF	CMM	KHIDF	FRP	DRP		
		Predicted Class									

Figure 7. Confusion matrix of the proposed method for eight different loads.

4. Comparative Study

In this chapter, a large number of comparative experiments are designed to demonstrate the superiority of the proposed method in this paper. Ablation experiments are also set up to verify the necessity of improvements in the proposed method and some steps used.

4.1. Comparison of Different Feature Extraction

To verify the superiority of time–frequency-domain feature sets, this paper directly inputs time–frequency-domain feature sets, time-domain feature sets, and frequency-domain feature sets into the KELM model without any feature selection step. The effects of the initial feature set selection on the model performance are compared. In this case, the kernel parameter γ and the penalty coefficient C of KELM are set to 1. The classification accuracies of different initial feature sets are shown in Table 4. It can be seen that the classification accuracy of the load corresponding to the time–frequency-domain feature pool is the highest among the three initial feature sets, reaching 0.8930 and 0.8264, respectively. The classification accuracy of the training set of the time–frequency-domain feature set is improved by 8.23% and 5.08% compared to the time-domain and frequency-domain feature sets, respectively—the accuracy of the test set was enhanced by 7.52% and 4.16%, respectively. The classification effect is ranked as follows: time–frequency domain > frequency domain > time domain. As can be seen from Table 4, the load recog-

nition efficiency of the time-domain feature set with 32 features is worse than that of the frequency-domain feature set with 26 features. In addition, the time–frequency-domain feature set that combines both features has the highest classification accuracy. This illustrates that the selection of the initial feature set can affect the model’s performance to some extent. In addition, the number of features does not determine the good or bad final classification results. This further shows the importance of feature selection.

Table 4. Comparison of classification accuracy with different initial feature sets.

Feature Set	Number of Features	Training Set	Test Set
Time-domain features	32	0.8251 (401/486)	0.7686 (93/121)
Frequency-domain features	26	0.8498 (413/486)	0.7934 (96/121)
Time–frequency-domain features	58	0.8930 (434/486)	0.8264 (100/121)

4.2. Comparison of Feature Selection Methods

4.2.1. Comparison of First-Stage Filtered Feature Selection

Preliminary feature screening is especially important for the overall two-stage feature selection. The common filtered feature selection methods in machine learning are selected for comparison experiments in this paper. These are Pearson’s correlation coefficient (PCC), Spearman’s coefficient (SCC) and the maximum information coefficient (MIC). The experimental results are shown in Figure 8. The change in load recognition accuracy with the gradual increase in features under different filtered feature selection methods is depicted.

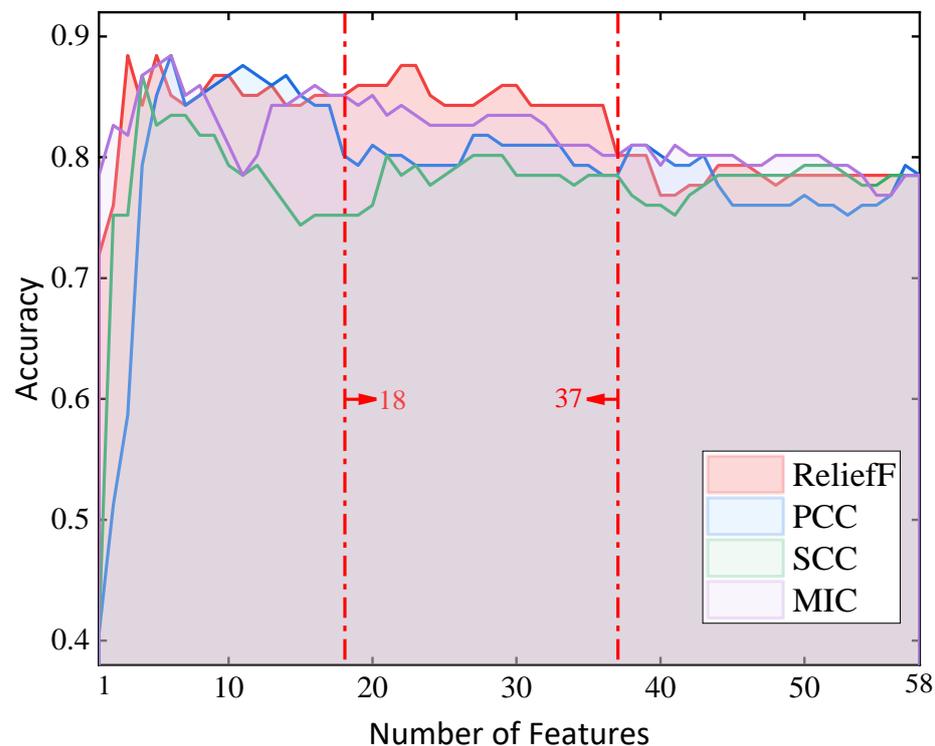


Figure 8. Performance comparison of common filtered feature selection methods.

From Figure 8, it can be seen that the model’s classification accuracy increases faster at the beginning when the number of feature dimensions grows. When the number of feature dimensions exceeds a certain threshold, the model performance will no longer increase or gradually decrease. Reflecting the importance of feature selection, there are redundant bad features among the 58 time–frequency-domain features. Among the four filtered feature selection methods, the overall performance of SCC is the worst in relative terms. The

number of features ranges from 18 to 37, and the accuracy of features selected using the ReliefF method significantly exceeds that of the other filtered methods. Meanwhile, the number of features ranges from 1 to 18, and the performance of the ReliefF feature selection model is equally good. Although there is no method that can beat other methods in all cases, overall, ReliefF performs better than other single-feature selection methods. The superiority of ReliefF in the first stage of feature selection in this paper is verified.

4.2.2. Comparison of Second-Stage Heuristic Feature Selection

To show the superiority of MPA as a second-stage feature selection technique, MPA, the firefly algorithm (FA) [44], and the slime mold algorithm (SMA) [45] are compared under the premise that the first step of feature screening is the ReliefF method. The convergence plots of the *FS-fitness* convergence of the fitness function of MPA and different intelligent optimization algorithms are shown in Figure 9. From Equation (9), the smaller the *FS-fitness* value, the better the selected features, i.e., the lower the recognition error rate and the smaller the number of feature dimensions. The adaptation values of ReliefF-FA, ReliefF-SMA and ReliefF-MPA are 0.1662, 0.1032 and 0.0809, respectively. Lower adaptation values mean that the selected features are more effective. Compared to FA and SMA as the second-stage feature selection, the fitness values of MPA decreased by 51.32% and 21.61%, respectively. This proves the superiority of MPA as the second-stage feature selection method.

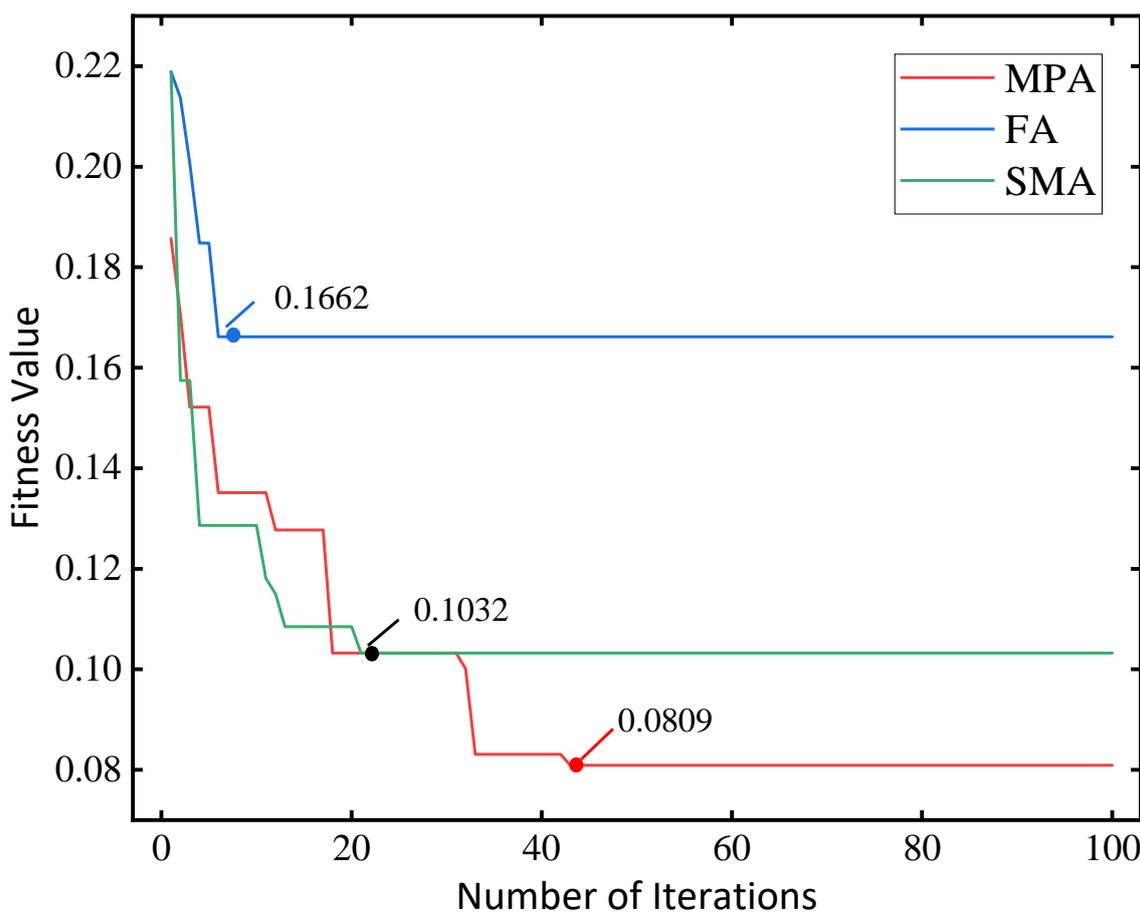


Figure 9. Adaptation curves of different heuristic feature selection methods under ReliefF.

To observe the effect of different feature choices on the actual electrical load classification more clearly, the KNN algorithm was used to classify the test set, and the results are shown in Figure 10. The features selected by the two-stage feature selection method proposed in this paper have the best effect. Even if the feature pre-selection stage is all Reli-

eff, the different heuristic algorithms used in the reselection stage can significantly affect the final classification results. If FA is selected in the reselection stage, the classification algorithm will misclassify five types of equipment (KTHTF, CME, CMM, FRP, and DRP), with 20 misidentified samples and 0.8347 classification accuracy. If SMA is selected in the reselection stage, four types of electrical equipment will be misidentified (KTHTF, CME, FRP, and DRP), with 13 misidentified samples and 0.8926 classification accuracy.

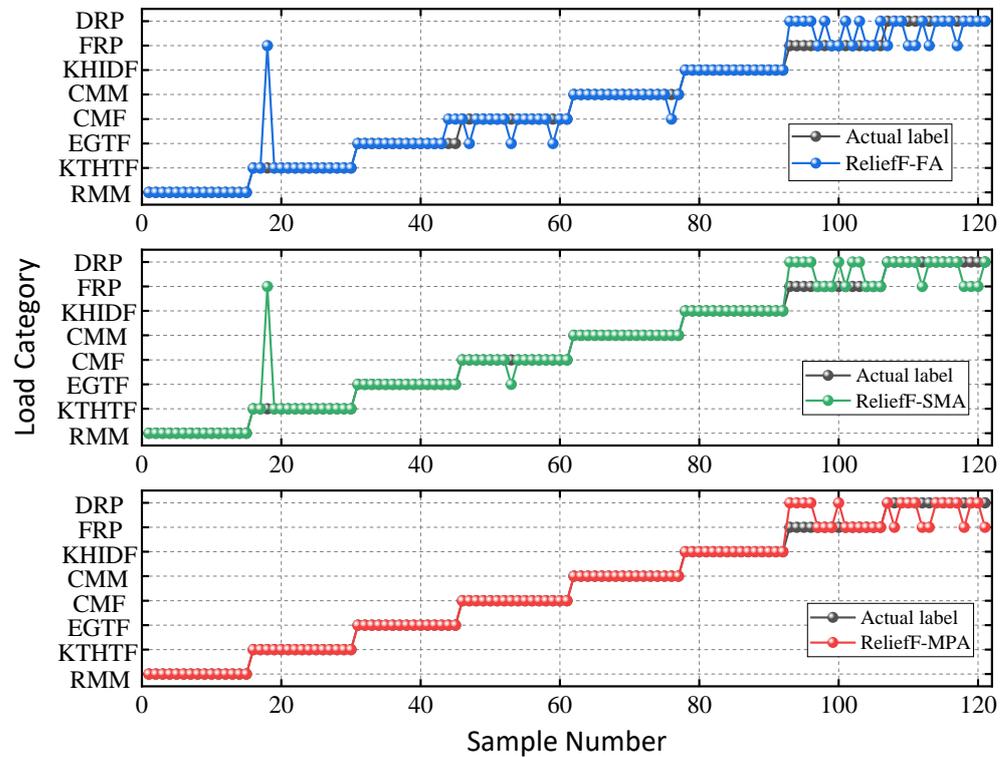


Figure 10. Classification results of KNN under different wrapper methods.

In contrast, the MPA feature selection method proposed is used. Only the DRP and FRP of the dynamic roller press are incorrectly identified, with an accuracy of 0.9174. The superiority of MPA as the reselection stage of the two-stage feature selection is reflected in the number of equipment types identified correctly and the total number of correctly identified samples.

4.3. Comparison of Different Classifiers and Optimization Algorithms

To verify the superiority of the IMPA-KELM classifier proposed in this paper, it is compared with support vector machine (SVM), extreme learning machine (ELM), and classifiers optimized by the FA and the SMA without improvements. The performance metrics of each classifier obtained from the experimental results are recorded in Table 5. Among them, regarding SVM, this paper uses the LIBSVM3.1 toolbox to implement it. The kernel function is RBF. The number of nodes in the hidden layer of ELM is set to 8. The kernel parameters and penalty coefficients of KELM are set to 1. To ensure the fairness of comparison among the algorithms, the number of populations and the maximum number of iterations of each optimization algorithm are set to 30 and 50, respectively. The optimal value of each evaluation metric is marked in bold font. Rank is the sum of the ranking of the four metrics of the algorithm among all models. These results prove the efficiency of the IMPA-KELM.

Table 5. Load classification performance comparison.

Classification Model	Accuracy	Precision	Recall	F1-Score	Rank
SVM	0.7769	/	0.7714	/	8
ELM	0.7273	0.7432	0.7239	0.6995	6
KELM	0.7355	0.7445	0.7342	0.7124	10
FA-SVM	0.9008	0.8975	0.8975	0.8971	36
FA-ELM	0.8347	0.8312	0.8306	0.8243	22
FA-KELM	0.9256	0.9242	0.9225	0.9222	55
SMA-SVM	0.9091	0.9069	0.9065	0.9056	43
SMA-ELM	0.8182	0.8123	0.8145	0.8108	16
SMA-KELM	0.8926	0.8886	0.8880	0.8874	32
MPA-SVM	0.9091	0.9064	0.9053	0.9048	40
MPA-ELM	0.8347	0.8372	0.8300	0.8186	21
MPA-KELM	0.9256	0.9225	0.9220	0.9220	52
IMPA-SVM	0.9174	0.9148	0.9148	0.9143	48
IMPA-ELM	0.8430	0.8613	0.8436	0.8375	28
IMPA-KELM	0.9339	0.9321	0.9315	0.9313	60

From the table, it can be seen that the load recognition performance of the classifier optimized by the optimizer is superior to that of the classifier without hyperparameter tuning. This reflects the importance of the setting of hyperparameters for the classification performance of the algorithm. In addition, even for the same classifier, there are significant differences in the results obtained using different optimization algorithms. This is due to the differences in the optimal parameters set by the optimization algorithms for SVM, ELM and KELM. As can be seen from the table, the IMPA-KELM model obtained better results than the other models in all four evaluation metrics, all of which exceeded 0.93. The introduction of IMPA improved the classification accuracy of SVM, ELM and KELM by 18.08%, 15.91% and 26.97%, respectively. This is unmatched by other optimization algorithms, which indicates that IMPA always finds more suitable parameters for SVM, ELM and KELM than other optimization algorithms. IMPA satisfies their potential classification performance and demonstrates solid parameter-finding ability.

The classification performance of metaheuristic-based algorithms is further investigated using Table 5’s ranking technique for each benchmark. In Figure 11, stacked bars depict the total ranking results. As can be seen from the figure, overall, KELM has the highest ranking score, followed by SVM, and ELM has the lowest. In addition, the performance scores are greatly improved after the metaheuristic algorithm optimization, regardless of the algorithm. It is also worth noting that the IMPA algorithm finds more suitable parameters for each algorithm.

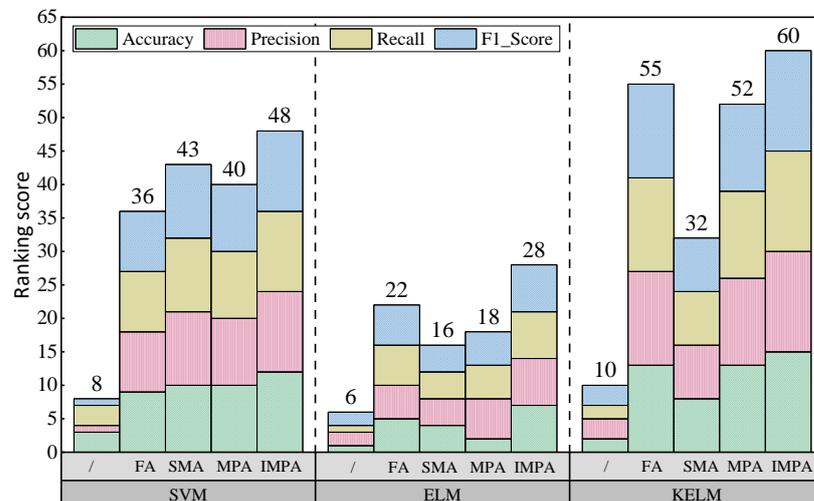


Figure 11. Overall stacked ranking results for different classifiers.

The convergence curves of several metaheuristic KELM-based models are shown in Figure 12. The figure shows that the KELM model optimized by IMPA has the lowest recognition error rate for power equipment. This indicates that the improved MPA has a stronger global search capability. In addition, it can be seen that the convergence speed of IMPA is faster compared to MPA. This proves the superiority of IMPA proposed in this paper.

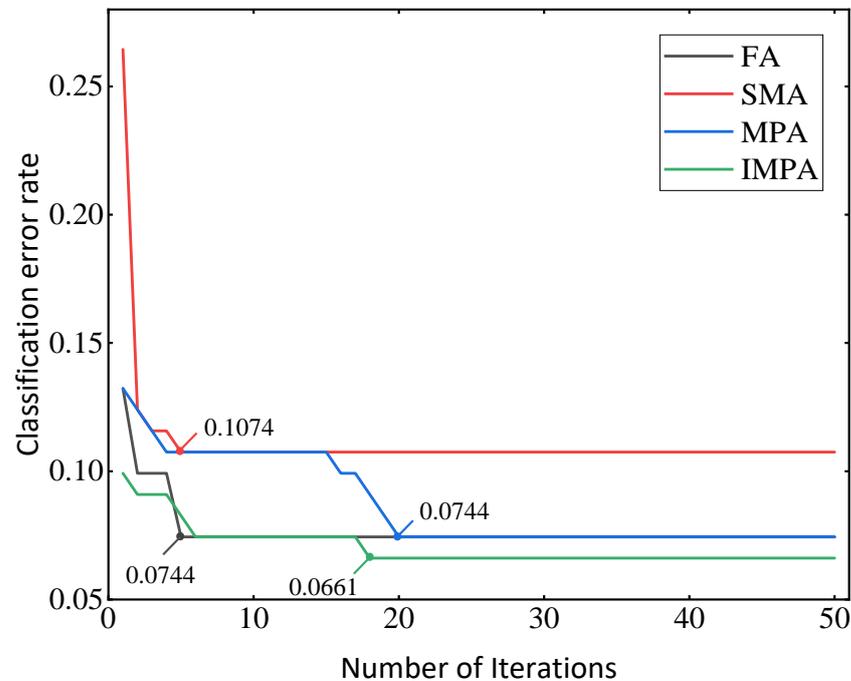


Figure 12. Comparison of convergence curves of various meta-heuristics-based KELM models.

5. Conclusions

In this paper, a classification method based on a two-stage feature selection strategy and IMPA-KELM model is proposed for the industrial power load identification problem. The superiority and feasibility of the proposed method are verified by using the collected power data from different workshops of a cement plant in China. The experimental results show the advantages of the proposed method as follows.

- (1) The operating information of each load is extracted as much as possible by extracting the time-domain features and frequency-domain features of P and Q of electrical equipment. A rich set of combined features is generated.
- (2) This paper proposes a new method combining ReliefF and MPA feature selection. ReliefF is used as a pre-feature filter, and MPA is used to optimize the features again. Feature dimensionality and recognition accuracy are considered by setting the fitness function equation. This method reduces the original time–frequency-domain feature set of 58 features to 3, which is 5.17% of the original feature dimensionality. While ensuring high classification accuracy, features with redundant information are eliminated as much as possible, reducing the complexity of the classification process, the computational cost and storage requirements.
- (3) A new and improved marine predator algorithm is proposed. This algorithm introduces a chaotic initialization strategy and boundary variation operation to improve MPA's convergence speed and global search capability. In this paper, the improved MPA is applied to the parameter optimization process of KELM to obtain the new classifier, IMPA-KELM, which achieves the optimal selection of KELM kernel parameters and penalty coefficients and thus improves the accuracy of its electric load identification.

- (4) A valuable reference for future research is provided. A large number of comparative experiments are set up to summarize and verify the determination of the initial feature set, the application effect of the dimensionality reduction methods, the classification performance of different classifiers and the impact of some optimization algorithms. The datasets are obtained through the actual cement plant's record collection, which avoids the performance saturation of some methods and can effectively distinguish the effects in different ways.

Although the proposed method has a superior performance, it still presents some misclassifications that must be further studied.

Author Contributions: M.Z.: Methodology, Formal analysis, Writing—original draft. Z.Z.: Software, Supervision, Project administration. F.H.: Data curation, Formal analysis. K.B.: Conceptualization, Resources, Investigation. W.L.: Writing—review and editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China under Grant No. 2018YFC0604503, the Energy Internet Joint Fund Project of Anhui Province, China under Grant No. 2008085UD06, the Major Science and Technology Program of Anhui Province, China under Grant No. 201903a07020013, and the Graduate Innovation Fund Project of Anhui University of Science and Technology, China under Grant No. 2021CX1009.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no known competing financial interest or personal relationship that could have appeared to influence the work reported in this paper.

References

1. Lamnatou, C.; Chemisana, D.; Cristofari, C. Smart grids and smart technologies in relation to photovoltaics, storage systems, buildings and the environment. *Renew. Energy* **2022**, *185*, 1376–1391. [[CrossRef](#)]
2. Wu, X.; Han, X.; Liang, K.X. Event-based non-intrusive load identification algorithm for residential loads combined with underdetermined decomposition and characteristic filtering. *IET Gener. Transm. Distrib.* **2019**, *13*, 99–107. [[CrossRef](#)]
3. Zator, S. Power scheduling scheme for DSM in smart homes with photovoltaic and energy storage. *Energies* **2021**, *14*, 8571. [[CrossRef](#)]
4. Yu, J.Y.; Liu, W.N.; Wu, X. Noninvasive industrial power load monitoring based on collaboration of edge device and edge data center. In Proceedings of the 2020 IEEE International Conference on Edge Computing (EDGE), Beijing, China, 19–23 October 2020; pp. 23–30.
5. Du, Y.; Du, L.; Lu, B.; Harley, R.; Habetler, T. A review of identification and monitoring methods for electric loads in commercial and residential buildings. In Proceedings of the 2010 IEEE Energy Conversion Congress and Exposition, Atlanta, GA, USA, 12–16 September 2010; pp. 4527–4533.
6. Huang, Y.; Yang, H. A Method based on K-Means and fuzzy algorithm for industrial load identification. In Proceedings of the 2012 Asia-Pacific Power and Energy Engineering Conference, Shanghai, China, 27–29 March 2012; pp. 1–4.
7. Hart, G.W.; Kern, E.C., Jr.; Schweppe, F.C. Non-Intrusive Appliance Monitor Apparatus. U.S. Patent 4,858,141, 15 August 1989.
8. Heo, S.; Kim, H. Toward load identification based on the hilbert transform and sequence to sequence long short-term memory. *IEEE Trans. Smart Grid* **2021**, *12*, 3252–3264.
9. Yu, M.; Wang, B.; Lu, L.; Bao, Z.; Qi, D. Non-intrusive adaptive load identification based on siamese network. *IEEE Access* **2022**, *10*, 11564–11573. [[CrossRef](#)]
10. Bucci, G.; Ciancetta, F.; Fiorucci, E.; Mari, S.; Fioravanti, A. A non-intrusive load identification system based on frequency response analysis. In Proceedings of the 2021 IEEE International Workshop on Metrology for Industry 4.0&IoT (MetroInd4.0&IoT), Rome, Italy, 7–9 June 2021; pp. 254–258.
11. Tao, P.; Liu, X.A.; Zhang, Y.; Li, C.; Ding, J. Multi-level non-intrusive load identification based on k-NN. In Proceedings of the 2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2), Changsha, China, 8–10 November 2019; pp. 1905–1910.
12. Zhang, R.; Song, Y. Non-intrusive load identification method based on color encoding and improve R-FCN. *Sustain. Energy Technol. Assess.* **2022**, *53*, 102714. [[CrossRef](#)]
13. Kim, J.; Le, T.T.; Kim, H. Nonintrusive load monitoring based on advanced deep learning and novel signature. *Comput. Intell. Neurosci.* **2017**, *2017*, 4216281. [[CrossRef](#)]
14. Mukaroh, A.; Le, T.T.; Kim, H. Background load denoising across complex load based on generative adversarial network to enhance load identification. *Sensors* **2020**, *20*, 5674. [[CrossRef](#)]

15. Yan, L.; Sheikholeslami, M.; Gong, W.; Tian, W.; Li, Z. Challenges for real-world applications of nonintrusive load monitoring and opportunities for machine learning approaches. *Electr. J.* **2022**, *35*, 107136. [[CrossRef](#)]
16. Luan, W.; Yang, F.; Zhao, B.; Liu, B. Industrial load disaggregation based on hidden Markov models. *Electr. Power Syst. Res.* **2022**, *210*, 108086. [[CrossRef](#)]
17. Wei, W.; Peng, T.; Ye, L.; Feng, X.; Jiang, Z.; Yu, H. A Feature Extraction Method for Non-intrusive Load Identification. *J. Phys. Conf. Ser.* **2022**, *2195*, 012008. [[CrossRef](#)]
18. Huang, Z.; Yang, C.; Zhou, X.; Huang, T. A hybrid feature selection method based on binary state transition algorithm and ReliefF. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 1888–1898. [[CrossRef](#)]
19. Yan, C.; Liang, J.; Zhao, M.; Zhang, X.; Zhang, T.; Li, H. A novel hybrid feature selection strategy in quantitative analysis of laser-induced breakdown spectroscopy. *Anal. Chim. Acta* **2019**, *1080*, 35–42. [[CrossRef](#)]
20. Cao, Y.; Sun, Y.; Xie, G.; Li, P. A sound-based fault diagnosis method for railway point machines based on two-stage feature selection strategy and ensemble classifier. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 12074–12083. [[CrossRef](#)]
21. Wu, S.D.; Wu, P.H.; Wu, C.W.; Ding, J.J.; Wang, C.C. Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy* **2012**, *14*, 1343–1356. [[CrossRef](#)]
22. Yun, Y.H.; Li, H.D.; Deng, B.C.; Cao, D.S. An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC Trends Anal. Chem.* **2019**, *113*, 102–115. [[CrossRef](#)]
23. Sun, L.; Yin, T.; Ding, W.; Qian, Y.; Xu, J. Multilabel feature selection using ML-Relieff and neighborhood mutual information for multilabel neighborhood decision systems. *Inf. Sci.* **2020**, *537*, 401–424. [[CrossRef](#)]
24. Varzaneh, Z.A.; Hossein, S.; Mood, S.E.; Javidi, M.M. A new hybrid feature selection based on Improved Equilibrium Optimization. *Chemom. Intell. Lab. Syst.* **2022**, *228*, 104618. [[CrossRef](#)]
25. Kılıç, F.; Kaya, Y.; Yildirim, S. A novel multi population based particle swarm optimization for feature selection. *Knowl. Based Syst.* **2021**, *219*, 106894. [[CrossRef](#)]
26. Hu, P.; Pan, J.S.; Chu, S.C. Improved binary grey wolf optimizer and its application for feature selection. *Knowl. Based Syst.* **2020**, *195*, 105746. [[CrossRef](#)]
27. Nadimi-Shahraki, M.H.; Zamani, H.; Mirjalili, S. Enhanced whale optimization algorithm for medical feature selection: A COVID-19 case study. *Comput. Biol. Med.* **2022**, *148*, 105858. [[CrossRef](#)] [[PubMed](#)]
28. Faramarzi, A.; Heidarinejad, M.; Mirjalili, S.; Gandomi, A.H. Marine Predators Algorithm: A nature-inspired metaheuristic. *Expert Syst. Appl.* **2020**, *152*, 113377. [[CrossRef](#)]
29. Wang, Z.; Zheng, L.; Wang, J.; Du, W. Research on novel bearing fault diagnosis method based on improved krill herd algorithm and kernel extreme learning machine. *Complexity* **2019**, *2019*, 4031795. [[CrossRef](#)]
30. Eshtay, M.; Faris, H.; Obeid, N. Metaheuristic-based extreme learning machines: A review of design formulations and applications. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 1543–1561. [[CrossRef](#)]
31. Wang, J.; Lu, S.; Wang, S.H.; Zhang, Y.D. A review on extreme learning machine. *Multimed. Tools Appl.* **2021**, *81*, 41611–41660. [[CrossRef](#)]
32. Cai, Z.; Gu, J.; Luo, J.; Zhang, Q.; Chen, H.; Pan, Z.; Li, Y.; Li, C. Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf optimization strategy. *Expert Syst. Appl.* **2019**, *138*, 112814. [[CrossRef](#)]
33. Lu, H.; Du, B.; Liu, J.; Xia, H.; Yeap, W.K. A kernel extreme learning machine algorithm based on improved particle swarm optimization. *Memetic Comput.* **2017**, *9*, 121–128. [[CrossRef](#)]
34. Wang, M.; Chen, H.; Yang, B.; Zhao, X.; Hu, L.; Cai, Z.; Huang, H.; Tong, C. Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses. *Neurocomputing* **2017**, *267*, 69–84. [[CrossRef](#)]
35. Zhu, W.; Ma, C.; Zhao, X.; Wang, M.; Heidari, A.A.; Chen, H.; Li, C. Evaluation of sino foreign cooperative education project using orthogonal sine cosine optimized kernel extreme learning machine. *IEEE Access* **2020**, *8*, 61107–61123. [[CrossRef](#)]
36. Abdel-Basset, M.; El-Shahat, D.; Chakraborty, R.K.; Ryan, M. Parameter estimation of photovoltaic models using an improved marine predators algorithm. *Energy Convers. Manag.* **2021**, *227*, 113491. [[CrossRef](#)]
37. Wen, T.; Dong, D.; Chen, Q.; Chen, L.; Roberts, C. Maximal information coefficient-based two-stage feature selection method for railway condition monitoring. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 2681–2690. [[CrossRef](#)]
38. Helmi, H.; Forouzantabar, A. Rolling bearing fault detection of electric motor using time domain and frequency domain features extraction and ANFIS. *IET Electr. Power Appl.* **2019**, *13*, 662–669. [[CrossRef](#)]
39. Praveena, H.D.; Subhas, C.; Naidu, K.R. Retracted article: Automatic epileptic seizure recognition using relief feature selection and long short term memory classifier. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 6151–6167. [[CrossRef](#)]
40. Zhang, B.; Li, Y.; Chai, Z. A novel random multi-subspace based ReliefF for feature selection. *Knowl. Based Syst.* **2022**, *252*, 109400. [[CrossRef](#)]
41. Zhong, K.; Zhou, G.; Deng, W.; Zhou, Y.; Luo, Q. MOMPA: Multi-objective marine predator algorithm. *Comput. Methods Appl. Mech. Eng.* **2021**, *385*, 114029. [[CrossRef](#)]
42. Gaspar, A.; Oliva, D.; Hinojosa, S.; Aranguren, I.; Zaldivar, D. An optimized Kernel Extreme Learning Machine for the classification of the autism spectrum disorder by using gaze tracking images. *Appl. Soft Comput.* **2022**, *120*, 108654. [[CrossRef](#)]
43. Hu, W. An improved flower pollination algorithm for optimization of intelligent logistics distribution center. *Adv. Prod. Eng. Manag.* **2019**, *14*, 177–188. [[CrossRef](#)]

44. Kumar, V.; Kumar, D. A systematic review on firefly algorithm: Past, present, and future. *Arch. Comput. Methods Eng.* **2021**, *28*, 3269–3291. [[CrossRef](#)]
45. Liu, Y.; Heidari, A.A.; Ye, X.; Liang, G.; Chen, H.; He, C. Boosting slime mould algorithm for parameter identification of photovoltaic models. *Energy* **2021**, *234*, 121164. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.